Heterogeneity, Uncertainty and Learning: Semiparametric Identification and Estimation

Jackson Bunting — U. Washington
Paul Diegert — TSE
Arnaud Maurel — Duke U. and NBER

DSE Conference, UCL

July 3, 2025

Motivation

- Panel data with continuous outcomes and discrete choices (Y_{it}, D_{it}).
- (Y_{it}, D_{it}) depend on multidimensional (time invariant) latent variable, X_i^*.
- Some components of X_i^* initially unknown by the agents.
- Arises frequently in applications.
- E.g.: productivity partly unknown to workers when they enter the workforce.
- Workers update their beliefs about their productivity as they observe their wages.

Motivation

- Panel data with continuous outcomes and discrete choices (Y_{it}, D_{it}).
- (Y_{it}, D_{it}) depend on multidimensional (time invariant) latent variable, X_i^*.
- Some components of X_i^* initially unknown by the agents.
- Arises frequently in applications.
- E.g.: productivity partly unknown to workers when they enter the workforce.
- Workers update their beliefs about their productivity as they observe their wages.

- Learning models are popular in economics, empirical micro in particular.
- Used in various fields, including Labor (Miller, 1984; Antonovics and Golan, 2012; Pastorino, 2024), Education (Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2012; Arcidiacono et al., 2025); IO/Health (Ackerberg, 2003; Crawford and Shum, 2005; Aguirregabiria and Jeon, 2020).
- However, much remains to be known about identification.
- Standard to assume parametric outcome model (Gaussian), and specific learning process (RE with Bayesian updating).
- Even with these assumptions, no general identification result in the literature.

Motivation (Cont'd)

- Learning models are popular in economics, empirical micro in particular.
- Used in various fields, including Labor (Miller, 1984; Antonovics and Golan, 2012; Pastorino, 2024), Education (Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2012; Arcidiacono et al., 2025); IO/Health (Ackerberg, 2003; Crawford and Shum, 2005; Aguirregabiria and Jeon, 2020).
- However, much remains to be known about identification.
- Standard to assume parametric outcome model (Gaussian), and specific learning process (RE with Bayesian updating).
- Even with these assumptions, no general identification result in the literature.

This paper

- We provide conditions for identification of a learning model (with private information) from sequence of individual choices and outcomes:
  - Maintain some structure on the outcome model (linearity and normality of the errors and unknown factor).
  - Very few restrictions on choice process and learning rule.
  - Do not rely on measurements of unobserved heterogeneity.
- In a learning model with only unknown heterogeneity, we establish identification without assuming normality.
- Computationally tractable semi-nonparametric estimator.
- Application to ability learning in the context of occupational choice.

This paper

- We provide conditions for identification of a learning model (with private information) from sequence of individual choices and outcomes:
  - Maintain some structure on the outcome model (linearity and normality of the errors and unknown factor).
  - Very few restrictions on choice process and learning rule.
  - Do not rely on measurements of unobserved heterogeneity.
- In a learning model with only unknown heterogeneity, we establish identification without assuming normality.
- Computationally tractable semi-nonparametric estimator.
- Application to ability learning in the context of occupational choice.

This paper

- We provide conditions for identification of a learning model (with private information) from sequence of individual choices and outcomes:
  - Maintain some structure on the outcome model (linearity and normality of the errors and unknown factor).
  - Very few restrictions on choice process and learning rule.
  - Do not rely on measurements of unobserved heterogeneity.
- In a learning model with only unknown heterogeneity, we establish identification without assuming normality.
- Computationally tractable semi-nonparametric estimator.
- Application to ability learning in the context of occupational choice.

This paper

- We provide conditions for identification of a learning model (with private information) from sequence of individual choices and outcomes:
  - Maintain some structure on the outcome model (linearity and normality of the errors and unknown factor).
  - Very few restrictions on choice process and learning rule.
  - Do not rely on measurements of unobserved heterogeneity.
- In a learning model with only unknown heterogeneity, we establish identification without assuming normality.
- Computationally tractable semi-nonparametric estimator.
- Application to ability learning in the context of occupational choice.

- We provide conditions for identification of a learning model (with private information) from sequence of individual choices and outcomes:
  - Maintain some structure on the outcome model (linearity and normality of the errors and unknown factor).
  - Very few restrictions on choice process and learning rule.
  - Do not rely on measurements of unobserved heterogeneity.
- In a learning model with only unknown heterogeneity, we establish identification without assuming normality.
- Computationally tractable semi-nonparametric estimator.
- Application to ability learning in the context of occupational choice.

Outline

1. Model
2. Identification
3. Estimation
4. Simulations
5. Application

Setup

- We observe a short panel of discrete choices, continuous outcomes and covariates:

$$(D_{it}, Y_{it}, X_{it})_{t=1,2,\ldots,T}$$

- Two types of latent variables:
  - $X^*_{i,k} \in \mathbb{R}$: *known* by the agent.
  - $X^*_{i,u} \in \mathbb{R}^p$: initially *unknown* by the agent.

- $\dim(X^*_{i,u}) = p \geq 2$: non-diagonal covariance matrix $\rightarrow$ correlated learning.

- Idiosyncratic shocks affecting the outcomes, ϵ_{it}.

Setup

- We observe a short panel of discrete choices, continuous outcomes and covariates:

$$(D_{it}, Y_{it}, X_{it})_{t=1,2,\ldots,T}$$

- Two types of latent variables:
  - $X^*_{i,k} \in \mathbb{R}$: *known* by the agent.
  - $X^*_{i,u} \in \mathbb{R}^p$: initially *unknown* by the agent.

- $\dim(X^*_{i,u}) = p \geq 2$: non-diagonal covariance matrix $\rightarrow$ correlated learning.

- Idiosyncratic shocks affecting the outcomes, ϵ_{it}.

Setup

- We observe a short panel of discrete choices, continuous outcomes and covariates:

$$(D_{it}, Y_{it}, X_{it})_{t=1,2,\ldots,T}$$

- Two types of latent variables:
  - $X^*_{i,k} \in \mathbb{R}$: *known* by the agent.
  - $X^*_{i,u} \in \mathbb{R}^p$: initially *unknown* by the agent.
- $\dim(X^*_{i,u}) = p \geq 2$: non-diagonal covariance matrix $\rightarrow$ correlated learning.
- Idiosyncratic shocks affecting the outcomes, ϵ_{it}.

Setup - potential outcomes and choices

- Interactive fixed effect model for potential outcomes (omitting individual subscript i):

$$Y_t(d) = X_t^\intercal \beta_{t,d} + X_k^* \lambda_{t,d}^k + (X_u^*)^\intercal \lambda_{t,d}^u + \epsilon_t(d).$$

- Key assumption on choices: do not directly depend on X_u^*.

  Specifically:

$$D_t \perp\!\!\!\perp X_u^* \mid X^t, Y^{t-1}, D^{t-1}, X_k^*.$$

- Idiosyncratic shocks $\epsilon_t(d)$ are independent of $(X^t, Y^{t-1}, D^{t-1}, X^*)$ (with $X^* = (X_k^*, X_u^*)$).

- Interactive fixed effect model for potential outcomes (omitting individual subscript i):

$$Y_t(d) = X_t^\intercal \beta_{t,d} + X_k^* \lambda_{t,d}^k + (X_u^*)^\intercal \lambda_{t,d}^u + \epsilon_t(d).$$

- Key assumption on choices: do not directly depend on X_u^*.

  Specifically:

$$D_t \perp\!\!\!\perp X_u^* \mid X^t, Y^{t-1}, D^{t-1}, X_k^*.$$

- Idiosyncratic shocks $\epsilon_t(d)$ are independent of $(X^t, Y^{t-1}, D^{t-1}, X^*)$ (with $X^* = (X_k^*, X_u^*)$).

Setup - potential outcomes and choices (Cont'd)

Together with a law of motion for X_t, this can be summarized in the following assumption:

Assumption 1 (Conditional Independence)

For any $t \geq 2$ and $d \in Supp(D_t)$,

$$F_{\epsilon_t(d), D_t, X_t | Y^{t-1}, D^{t-1}, X^{t-1}, X^*} = F_{\epsilon_t(d)} F_{D_t | X^t, Y^{t-1}, D^{t-1}, X^*_k} F_{X_t | Y^{t-1}, D^{t-1}, X^{t-1}}.$$

For any $d \in Supp(D_1)$, $F_{\epsilon_1(d), D_1, X_1 | X^*} = F_{\epsilon_1(d)} F_{D_1 | X_1, X^*_k} F_{X_1 | X^*}$.

Connection with learning

- Denote by $\mathcal{I}_t$ the agent's information set in period t. For $t > 1$:

$$\mathcal{I}_t = (Y_1, \ldots, Y_{(t-1)}, D_1, \ldots D_{(t-1)}, X_1, \ldots, X_t, X_k^*)$$

and, for $t = 1$,

$$\mathcal{I}_1 = (X_1, X_k^*)$$

- Model consistent with agents forming their beliefs about X_u^* based on (subsets of) $\mathcal{I}_t$.
- Accommodates Bayesian updating under rational expectations.
- Also allows for various other models of expectations formation, including biased beliefs, or myopic expectations.

Connection with learning

- Denote by $\mathcal{I}_t$ the agent's information set in period t. For $t > 1$:

$$\mathcal{I}_t = (Y_1, \ldots, Y_{(t-1)}, D_1, \ldots D_{(t-1)}, X_1, \ldots, X_t, X_k^*)$$

and, for $t = 1$,

$$\mathcal{I}_1 = (X_1, X_k^*)$$

- Model consistent with agents forming their beliefs about X_u^* based on (subsets of) $\mathcal{I}_t$.
- Accommodates Bayesian updating under rational expectations.
- Also allows for various other models of expectations formation, including biased beliefs, or myopic expectations.

Outline

Model parameters (θ):

- Outcome equation parameters (β, λ).
- Distribution of the unobservables $(X_k^*, X_u^*, \epsilon_t)$.
- Conditional choice probabilities $\mathbb{P}(D_t = d | \mathcal{I}_t)$ (CCP).
- Covariate process $(X_t | Y^{t-1}, D^{t-1}, X^{t-1})$.

Two cases:

1. Unknown unobserved heterogeneity only (*Pure Learning*).
2. Known and unknown unobserved heterogeneity (*Learning with Private Information*).

Model parameters (θ):

- Outcome equation parameters (β, λ).
- Distribution of the unobservables $(X_k^*, X_u^*, \epsilon_t)$.
- Conditional choice probabilities $\mathbb{P}(D_t = d|\mathcal{I}_t)$ (CCP).
- Covariate process $(X_t|Y^{t-1}, D^{t-1}, X^{t-1})$.

Two cases:

1. Unknown unobserved heterogeneity only (*Pure Learning*).
2. Known and unknown unobserved heterogeneity (*Learning with Private Information*).

Intuition

- Selection problem: identify the interactive fixed effects model

$$Y_t(d) = \lambda_{t,d}^{\intercal} X^* + \epsilon_t(d)$$

from the distribution of (Y^T, D^T).

- Key idea: use CCPs to adjust for selection.
- Suppose $\mathrm{supp}(D_t) = \{0, 1\}$, and consider the distribution of $Y^T(1) \equiv (Y_1(1), Y_2(1), \ldots, Y_T(1))$.

Intuition

- Selection problem: identify the interactive fixed effects model

$$Y_t(d) = \lambda_{t,d}^{\intercal} X^* + \epsilon_t(d)$$

  from the distribution of (Y^T, D^T).
- Key idea: use CCPs to adjust for selection.
- Suppose $\mathrm{supp}(D_t) = \{0, 1\}$, and consider the distribution of $Y^T(1) \equiv (Y_1(1), Y_2(1), \ldots, Y_T(1))$.

Intuition (Cont'd)

- First, note that

$$f_{Y^T|D^T}(y^T|1)\frac{f_{D^T}(1)}{f_{D^T|Y^T(1)}(1|y^T)} = f_{Y^T(1)}(y^T).$$

- With no covariates and $X^* = X_u^*$, it follows from Assumption 1 that the inverse selection weight, $f_{D^T|Y^T(1)}$, is identified as follows:

$$f_{D^T|Y^T(1)}(1|y^T) = f_{D_t|Y^{t-1},D^{t-1}}(1|y^{t-1},1)f_{D^{t-1}|Y^{t-2},D^{t-2}}(1|y^{t-2},1)...f_{D_1}(1).$$

- Identifies in turn the distribution of $Y^T(1)$.
- Back to standard interactive fixed effects setup (Freyberger, 2018).

Intuition (Cont'd)

- First, note that

$$f_{Y^T|D^T}(y^T|1)\frac{f_{D^T}(1)}{f_{D^T|Y^T(1)}(1|y^T)} = f_{Y^T(1)}(y^T).$$

- With no covariates and $X^* = X_u^*$, it follows from Assumption 1 that the inverse selection weight, $f_{D^T|Y^T(1)}$, is identified as follows:

$$f_{D^T|Y^T(1)}(1|y^T) = f_{D_t|Y^{t-1},D^{t-1}}(1|y^{t-1},1)f_{D^{t-1}|Y^{t-2},D^{t-2}}(1|y^{t-2},1)...f_{D_1}(1).$$

- Identifies in turn the distribution of $Y^T(1)$.
- Back to standard interactive fixed effects setup (Freyberger, 2018).

Pure Learning

- Generalize the previous ideas to include covariates and multiple potential outcomes.
- Since $X^* = X_u^*$, $\mathcal{I}_t = \{Y^{t-1}, D^{t-1}, X^t\}$ and $f_{D_t,X_t|\mathcal{I}_t}$ and $f_{D_1|X_1}$ are identified from the data (selection on obs.).
- Assume for simplicity that $X_t \perp\!\!\!\perp X^* | X_{t-1}$ (relaxed in the paper).
- It follows that $f_{Y^T(d^T)|X^T}(y^T; x^T)$ is identified:

$$\frac{f_{Y^T,D^T,X^T}(y^T,d^T,x^T)}{f_{D_1,X_1}(d_1,z_1)\prod_{t=2}^T f_{D_t,X_t|\mathcal{I}_t}(d_t,x_t;h_t)} = f_{Y^T(d^T)|X^T}(y^T;x^T)$$

- One can then use results from the interactive fixed effect literature (e.g., Freyberger 2018) to identify the outcome eq. parameters and the distributions of (X^*, ϵ_{it}).

Pure Learning

- Generalize the previous ideas to include covariates and multiple potential outcomes.
- Since $X^* = X_u^*$, $\mathcal{I}_t = \{Y^{t-1}, D^{t-1}, X^t\}$ and $f_{D_t, X_t | \mathcal{I}_t}$ and $f_{D_1|X_1}$ are identified from the data (selection on obs.).
- Assume for simplicity that $X_t \perp\!\!\!\perp X^* | X_{t-1}$ (relaxed in the paper).
- It follows that $f_{Y^T(d^T)|X^T}(y^T; x^T)$ is identified:

$$\frac{f_{Y^T, D^T, X^T}(y^T, d^T, x^T)}{f_{D_1, X_1}(d_1, z_1) \prod_{t=2}^{T} f_{D_t, X_t | \mathcal{I}_t}(d_t, x_t; h_t)} = f_{Y^T(d^T)|X^T}(y^T; x^T)$$

- One can then use results from the interactive fixed effect literature (e.g., Freyberger 2018) to identify the outcome eq. parameters and the distributions of (X^*, ϵ_{it}).

- With private information, $X^* = (X_k^*, X_u^*)$, $\mathcal{I}_t = \{Y^{t-1}, D^{t-1}, X^t, X_k^*\}$.
- Key difference: CCPs are no longer identified directly from the data.
- Under a normality assumption, we show identification of the joint distribution of (Y^T, D^T, X^T, X_k^*).
- Having recovered the distribution of $\mathcal{I}_t$, identification follows from similar arguments as in the pure learning case.

- With private information, $X^* = (X_k^*, X_u^*)$, $\mathcal{I}_t = \{Y^{t-1}, D^{t-1}, X^t, X_k^*\}$.
- Key difference: CCPs are no longer identified directly from the data.
- Under a normality assumption, we show identification of the joint distribution of (Y^T, D^T, X^T, X_k^*).
- Having recovered the distribution of $\mathcal{I}_t$, identification follows from similar arguments as in the pure learning case.

Learning with private information - Normality

Assumption 2 (Normality)

(X_u^*, ϵ) are distributed according to

$$X_u^* \mid (X_1 = x_1, X_k^* = x_k^*) \sim \mathcal{N}\left(0, \Sigma_u(x_1)\right)$$
$$\epsilon_t(d) \sim \mathcal{N}(0, \sigma_t(d)^2)$$

- Standard in applied learning papers, which also use this assumption to specify a learning and choice model (e.g. Thomas, 2019; Arcidiacono et al., 2025).
- We maintain the restriction on the outcome model, but:
  - Remain flexible on the learning and choice process.
  - Do not assume a specific distribution for the known heterogeneity component (X_k^*).

Learning with private information - Normality

Assumption 2 (Normality)

(X_u^*, ϵ) are distributed according to

$$X_u^* \mid (X_1 = x_1, X_k^* = x_k^*) \sim \mathcal{N}\left(0, \Sigma_u(x_1)\right)$$
$$\epsilon_t(d) \sim \mathcal{N}(0, \sigma_t(d)^2)$$

- Standard in applied learning papers, which also use this assumption to specify a learning and choice model (e.g. Thomas, 2019; Arcidiacono et al., 2025).
- We maintain the restriction on the outcome model, but:
  - Remain flexible on the learning and choice process.
  - Do not assume a specific distribution for the known heterogeneity component (X_k^*).

Learning with private information - Normality

Assumption 2 (Normality)

(X_u^*, ϵ) are distributed according to

$$X_u^* \mid (X_1 = x_1, X_k^* = x_k^*) \sim \mathcal{N}\left(0, \Sigma_u(x_1)\right)$$
$$\epsilon_t(d) \sim \mathcal{N}(0, \sigma_t(d)^2)$$

- Standard in applied learning papers, which also use this assumption to specify a learning and choice model (e.g. Thomas, 2019; Arcidiacono et al., 2025).
- We maintain the restriction on the outcome model, but:
  - Remain flexible on the learning and choice process.
  - Do not assume a specific distribution for the known heterogeneity component (X_k^*).

Learning with private information - Distribution of X_k^*

Assumption 3 (Compact support)

The support of X_k^* is compact.

- Structural models typically assume the existence of a finite (and known) number of unobserved heterogeneity types.
- Allows for discrete or continuous distribution of X_k^*.
- Compactness plays an important role to identify $f_{X_k^*}$.

Assumption 3 (Compact support)

The support of X_k^* is compact.

- Structural models typically assume the existence of a finite (and known) number of unobserved heterogeneity types.
- Allows for discrete or continuous distribution of X_k^*.
- Compactness plays an important role to identify $f_{X_k^*}$.

Additional assumptions

(C) *Rank conditions*: For any d^T, all $p \times p$ submatrices of $[\lambda^u_{1,d_1} \cdots \lambda^u_{T,d_T}]$ are full rank; $f_{X_k^*|Y^{t-1},D^t,X^t} > 0$ almost surely.

(R) *Regularity conditions*, mostly ruling out knife-edge cases because of linearity, e.g., for all d, $\lambda^k_{t,d} \neq (\lambda^u_{t,d})^\intercal \Sigma_t \sum_{s=1}^{t-1} \lambda^k_{s,d_s} \frac{\lambda^k_{s,d_s}}{\sigma^2_{s,d_s}}$. $\rightarrow$ Aggregate effect of X_k^* on outcomes is non-zero.

(N) *Normalizations*: there is a d s.t. $\lambda^k_{1,d} = 1$; there is a sequence d^p such that $[\lambda^u_{t_1,d_1} \cdots \lambda^u_{t_p,d_p}] = I_p$.

▸ Assumptions

Main Identification Result

Theorem 1

Suppose the distribution of $(Y_t, D_t, X_t)_{t=1}^{T}$ is known for $T = 2p + 1$ and Assumptions 1-3, and (C), (R), (N) hold. Then θ is point identified.

- As in the model without private information, no assumption placed on the learning process.
- Existing identification results (Hotz and Miller, 1993; Arcidiacono and Miller, 2011) can then be applied to establish identification of particular choice models from the CCPs ($f_{D_t|\mathcal{I}_t}$).
- Proof uses implication of Assumptions 1-2 that cross-sectional wages are a Gaussian mixture conditional on history; use Bruni and Koch (1985) to identify conditional distributions of X_k^*. Proof Sketch

Main Identification Result

Theorem 1

Suppose the distribution of $(Y_t, D_t, X_t)_{t=1}^{T}$ is known for $T = 2p + 1$ and Assumptions 1-3, and (C), (R), (N) hold. Then θ is point identified.

- As in the model without private information, no assumption placed on the learning process.
- Existing identification results (Hotz and Miller, 1993; Arcidiacono and Miller, 2011) can then be applied to establish identification of particular choice models from the CCPs ($f_{D_t|\mathcal{I}_t}$).
- Proof uses implication of Assumptions 1-2 that cross-sectional wages are a Gaussian mixture conditional on history; use Bruni and Koch (1985) to identify conditional distributions of X_k^*. Proof Sketch

Main Identification Result

Theorem 1

Suppose the distribution of $(Y_t, D_t, X_t)_{t=1}^{T}$ is known for $T = 2p + 1$ and Assumptions 1-3, and (C), (R), (N) hold. Then θ is point identified.

- As in the model without private information, no assumption placed on the learning process.
- Existing identification results (Hotz and Miller, 1993; Arcidiacono and Miller, 2011) can then be applied to establish identification of particular choice models from the CCPs ($f_{D_t|\mathcal{I}_t}$).
- Proof uses implication of Assumptions 1-2 that cross-sectional wages are a Gaussian mixture conditional on history; use Bruni and Koch (1985) to identify conditional distributions of X_k^*. Proof Sketch

Main Identification Result

Theorem 1

Suppose the distribution of $(Y_t, D_t, X_t)_{t=1}^{T}$ is known for $T = 2p + 1$ and Assumptions 1-3, and (C), (R), (N) hold. Then θ is point identified.

- As in the model without private information, no assumption placed on the learning process.
- Existing identification results (Hotz and Miller, 1993; Arcidiacono and Miller, 2011) can then be applied to establish identification of particular choice models from the CCPs ($f_{D_t|\mathcal{I}_t}$).
- Proof uses implication of Assumptions 1-2 that cross-sectional wages are a Gaussian mixture conditional on history; use Bruni and Koch (1985) to identify conditional distributions of X_k^*. Proof Sketch

Outline

Sieve MLE estimation

- We focus here on learning under private information.
- Given an i.i.d. sample of data $(Y_{it}, D_{it}, X_{it} : t = 1, 2, \ldots, T)_{i=1}^{N}$, we estimate the model parameters via sieve MLE. Likelihood
- Nonparametric objects include the distribution of X_k^* and the CCPs.
- Establish consistency for fixed $T \geq 2p + 1$.

Sieve MLE estimation

- We focus here on learning under private information.
- Given an i.i.d. sample of data $(Y_{it}, D_{it}, X_{it} : t = 1, 2, \ldots, T)_{i=1}^{N}$, we estimate the model parameters via sieve MLE. ▸▸ Likelihood
- Nonparametric objects include the distribution of X_k^* and the CCPs.
- Establish consistency for fixed $T \geq 2p + 1$.

- We consider a sieve space for $F_{X_k^*}$ based on Koenker and Mizera (2014):

$$\mathcal{F}_n = \left\{ v \mapsto \sum_{s=1}^{q_n} \omega_s \mathbf{1}\{v \leq \bar{v}_{sn}\} \middle| \sum_{s}^{q_n} \omega_s = 1 \right\}$$

where $\mathcal{S}_n = \{\bar{v}_{1n}, \ldots, \bar{v}_{q_n n}\}$, for some q_n, as a grid of support points for X_k^*.

- Profile likelihood estimation, where we maximize over $F_{X_k^*}$ given $\theta \setminus F_{X_k^*}$ in an inner step.
- Fixing the other parameters, this is a convex optimization problem that can be solved very efficiently.
- Can be implemented using our Python package `spmlex`.

- We consider a sieve space for $F_{X_k^*}$ based on Koenker and Mizera (2014):

$$\mathcal{F}_n = \left\{ v \mapsto \sum_{s=1}^{q_n} \omega_s \mathbf{1}\{v \leq \bar{v}_{sn}\} \middle| \sum_{s}^{q_n} \omega_s = 1 \right\}$$

  where $\mathcal{S}_n = \{\bar{v}_{1n}, \ldots, \bar{v}_{q_n n}\}$, for some q_n, as a grid of support points for X_k^*.
- Profile likelihood estimation, where we maximize over $F_{X_k^*}$ given $\theta \setminus F_{X_k^*}$ in an inner step.
- Fixing the other parameters, this is a convex optimization problem that can be solved very efficiently.
- Can be implemented using our Python package `spmlex`.

Sieve MLE estimation (Cont'd)

- Often interested in particular functions of the model parameters.
- Example: decomposition of discounted lifetime earnings into a predictable and unpredictable components (Cunha et al., 2005; Cunha and Heckman, 2008, 2016).
- We consider a plug-in sieve estimator, and provide sufficient conditions for consistency and asymptotic normality for a family of functionals of the model parameters.
- Special case: variances of the predictable and unpredictable components of outcomes.

Family of functionals

Sieve MLE estimation (Cont'd)

- Often interested in particular functions of the model parameters.
- Example: decomposition of discounted lifetime earnings into a predictable and unpredictable components (Cunha et al., 2005; Cunha and Heckman, 2008, 2016).
- We consider a plug-in sieve estimator, and provide sufficient conditions for consistency and asymptotic normality for a family of functionals of the model parameters.
- Special case: variances of the predictable and unpredictable components of outcomes.

Family of functionals

Outline

Simulation design

- $T = 3$, $\dim(X_u^*) = 1$, binary choice $d \in \mathcal{D}_{it} = \{1, 2\}$, 2 covariates.
- Biased beliefs:

$$\mathcal{E}_i(Y_{it}(d)|\mathcal{I}_{it}) = \mathbb{E}(Y_{it}(d)|\mathcal{I}_{it}) + \gamma_d X_{k,i}^*$$

- Choice process:

$$D_{it} = \text{argmax}_d u_{it}(d|\mathcal{I}_{it}) + \eta_{it}(d)$$

  where $u_{it}(d|\mathcal{I}_{it}) = \rho \mathcal{E}_i\left(Y_{it}(d)|\mathcal{I}_{it}\right)$ and $\eta_{it}(d)$ is i.i.d. Type 1 EV.
- $X_{k,i}^*$ is a finite mixture of truncated normal r.v.'s., while $X_{u,i}^* \sim N(0, \sigma_u^2)$ and $\epsilon_{it}(d) \sim N(0, \sigma^2(d))$.

Simulation design

- $T = 3$, $\dim(X_u^*) = 1$, binary choice $d \in \mathcal{D}_{it} = \{1, 2\}$, 2 covariates.
- Biased beliefs:
$$\mathcal{E}_i(Y_{it}(d)|\mathcal{I}_{it}) = \mathbb{E}(Y_{it}(d)|\mathcal{I}_{it}) + \gamma_d X_{k,i}^*$$
- Choice process:
$$D_{it} = \text{argmax}_d \, u_{it}(d|\mathcal{I}_{it}) + \eta_{it}(d)$$
where $u_{it}(d|\mathcal{I}_{it}) = \rho \mathcal{E}_i\left(Y_{it}(d)|\mathcal{I}_{it}\right)$ and $\eta_{it}(d)$ is i.i.d. Type 1 EV.
- $X_{k,i}^*$ is a finite mixture of truncated normal r.v.'s., while $X_{u,i}^* \sim N(0, \sigma_u^2)$ and $\epsilon_{it}(d) \sim N(0, \sigma^2(d))$.

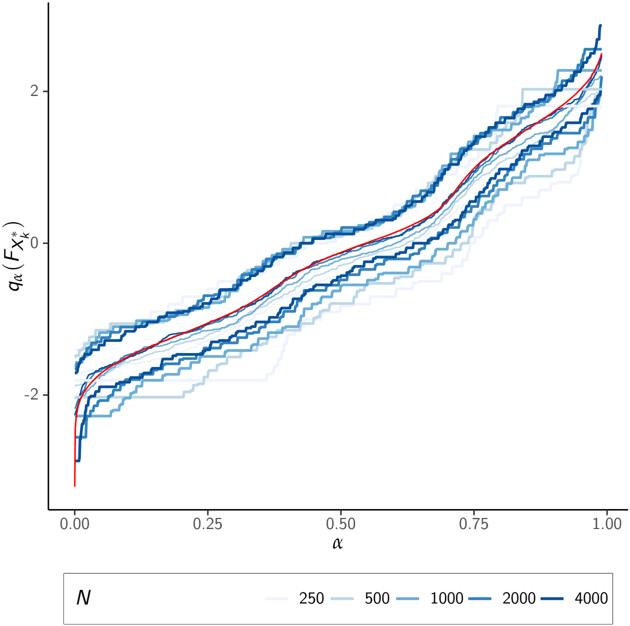
Simulation design

- $T = 3$, $\dim(X_u^*) = 1$, binary choice $d \in \mathcal{D}_{it} = \{1, 2\}$, 2 covariates.
- Biased beliefs:
$$\mathcal{E}_i(Y_{it}(d)|\mathcal{I}_{it}) = \mathbb{E}(Y_{it}(d)|\mathcal{I}_{it}) + \gamma_d X_{k,i}^*$$
- Choice process:
$$D_{it} = \text{argmax}_d u_{it}(d|\mathcal{I}_{it}) + \eta_{it}(d)$$
where $u_{it}(d|\mathcal{I}_{it}) = \rho \mathcal{E}_i\left(Y_{it}(d)|\mathcal{I}_{it}\right)$ and $\eta_{it}(d)$ is i.i.d. Type 1 EV.
- $X_{k,i}^*$ is a finite mixture of truncated normal r.v.'s., while $X_{u,i}^* \sim N(0, \sigma_u^2)$ and $\epsilon_{it}(d) \sim N(0, \sigma^2(d))$.

Monte Carlo results: squared bias and variance ($\times 1,000$)

	N = 250		N = 500		N = 1,000		N = 2,000		N = 4,000	
	sq bias	var	sq bias	var	sq bias	var	sq bias	var	sq bias	var
$\lambda^k_{1,1}$	2.746	27.521	1.701	12.890	0.624	7.268	0.009	3.684	0.001	1.468
$\lambda^k_{2,1}$	1.148	25.977	0.558	10.825	0.226	4.777	0.004	2.594	0.000	1.089
$\lambda^k_{2,2}$	0.869	10.978	0.254	5.825	0.073	2.654	0.007	1.383	0.000	0.743
$\lambda^k_{3,1}$	3.986	33.663	0.873	13.719	0.178	5.683	0.003	3.069	0.000	1.330
$\lambda^k_{3,2}$	5.702	36.861	0.674	12.556	0.224	5.305	0.011	2.408	0.005	1.080
$\lambda^u_{1,2}$	0.979	13.945	0.306	4.733	0.170	2.438	0.015	1.330	0.000	0.605
$\lambda^u_{2,1}$	0.040	8.317	0.027	5.139	0.036	1.947	0.014	1.003	0.002	0.481
$\lambda^u_{2,2}$	1.478	14.880	0.494	6.218	0.130	3.324	0.009	1.522	0.004	0.643
$\lambda^u_{3,1}$	0.446	9.912	0.093	5.003	0.062	2.187	0.030	0.968	0.023	0.469
$\lambda^u_{3,2}$	0.106	21.919	0.101	8.903	0.112	4.148	0.004	2.140	0.005	0.936
$\sigma^2(1)$	0.453	2.477	0.091	1.241	0.030	0.672	0.007	0.298	0.001	0.136
$\sigma^2(2)$	1.228	4.449	0.242	2.237	0.027	1.058	0.016	0.701	0.008	0.332

Quantiles of distribution of X_k^*



Outline

1. Model
2. Identification
3. Estimation
4. Simulations
5. **Application**

Occupational choice and learning

- We apply our framework to study occupational choice.
- Focus on the role of uncertainty vs. heterogeneity, in a context where agents may learn over time.
- We do not rely on an auxiliary measurement system, unlike much of the empirical literature (e.g. Cunha and Heckman, 2008; Arcidiacono et al., 2025).
- Data: National Longitudinal Survey of Youth 1997 (NLSY97).
  - Sample of 2,453 white men born between 1980 and 1984.
  - Restrict to full-time workers between ages 27 and 32.
  - Outcome variable (Y): Log hourly wage; choice variable (D): high- or low-skill occupation.

Occupational choice and learning

- We apply our framework to study occupational choice.
- Focus on the role of uncertainty vs. heterogeneity, in a context where agents may learn over time.
- We do not rely on an auxiliary measurement system, unlike much of the empirical literature (e.g. Cunha and Heckman, 2008; Arcidiacono et al., 2025).
- Data: National Longitudinal Survey of Youth 1997 (NLSY97).
    - Sample of 2,453 white men born between 1980 and 1984.
    - Restrict to full-time workers between ages 27 and 32.
    - Outcome variable (Y): Log hourly wage; choice variable (D): high- or low-skill occupation.

Occupational choice and learning

- We apply our framework to study occupational choice.
- Focus on the role of uncertainty vs. heterogeneity, in a context where agents may learn over time.
- We do not rely on an auxiliary measurement system, unlike much of the empirical literature (e.g. Cunha and Heckman, 2008; Arcidiacono et al., 2025).
- Data: National Longitudinal Survey of Youth 1997 (NLSY97).
  - Sample of 2,453 white men born between 1980 and 1984.
  - Restrict to full-time workers between ages 27 and 32.
  - Outcome variable (Y): Log hourly wage; choice variable (D): high- or low-skill occupation.

Set-up and implementation

- $T = 3$, $\dim(X_u^*) = 1$, $\mathcal{D}_t = \{0, 1\}$ for $t \in \{1, 2, 3\}$ (low- or high-skill sector).
- Potential outcomes ($Y_t(d)$): log hourly wages in period t (two-year average).
- Where, for $d \in \{0, 1\}$:

$$Y_t(d) = \beta_{t,d} + X_k^* \lambda_{t,d}^k + X_u^* \lambda_{t,d}^u + \epsilon_t(d).$$

- CCPs:
    - $h_t(X_k^*, Y^{t-1}, D^{t-1}) := P(D_t = 1 | X_k^*, Y^{t-1}, D^{t-1})$
- Estimation via sieve MLE.
- We use a flexible logit for the CCPs h_t, and the sieve space $\mathcal{F}_n$ for $F_{X_k^*}$ (with 56 eq. spaced support points).

Set-up and implementation

- $T = 3$, $\dim(X_u^*) = 1$, $\mathcal{D}_t = \{0, 1\}$ for $t \in \{1, 2, 3\}$ (low- or high-skill sector).
- Potential outcomes ($Y_t(d)$): log hourly wages in period t (two-year average).
- Where, for $d \in \{0, 1\}$:

$$Y_t(d) = \beta_{t,d} + X_k^* \lambda_{t,d}^k + X_u^* \lambda_{t,d}^u + \epsilon_t(d).$$

- CCPs:
  - $h_t(X_k^*, Y^{t-1}, D^{t-1}) := P(D_t = 1 | X_k^*, Y^{t-1}, D^{t-1})$
- Estimation via sieve MLE.
- We use a flexible logit for the CCPs h_t, and the sieve space $\mathcal{F}_n$ for $F_{X_k^*}$ (with 56 eq. spaced support points).

Set-up and implementation

- $T = 3$, $\dim(X_u^*) = 1$, $\mathcal{D}_t = \{0,1\}$ for $t \in \{1,2,3\}$ (low- or high-skill sector).
- Potential outcomes $(Y_t(d))$: log hourly wages in period t (two-year average).
- Where, for $d \in \{0,1\}$:

$$Y_t(d) = \beta_{t,d} + X_k^* \lambda_{t,d}^k + X_u^* \lambda_{t,d}^u + \epsilon_t(d).$$

- CCPs:
  - $h_t(X_k^*, Y^{t-1}, D^{t-1}) := P(D_t = 1 | X_k^*, Y^{t-1}, D^{t-1})$
- Estimation via sieve MLE.
- We use a flexible logit for the CCPs h_t, and the sieve space $\mathcal{F}_n$ for $F_{X_k^*}$ (with 56 eq. spaced support points).

Set-up and implementation

- $T = 3$, $\dim(X_u^*) = 1$, $\mathcal{D}_t = \{0, 1\}$ for $t \in \{1, 2, 3\}$ (low- or high-skill sector).
- Potential outcomes ($Y_t(d)$): log hourly wages in period t (two-year average).
- Where, for $d \in \{0, 1\}$:

$$Y_t(d) = \beta_{t,d} + X_k^* \lambda_{t,d}^k + X_u^* \lambda_{t,d}^u + \epsilon_t(d).$$

- CCPs:
  - $h_t(X_k^*, Y^{t-1}, D^{t-1}) := P(D_t = 1 | X_k^*, Y^{t-1}, D^{t-1})$
- Estimation via sieve MLE.
- We use a flexible logit for the CCPs h_t, and the sieve space $\mathcal{F}_n$ for $F_{X_k^*}$ (with 56 eq. spaced support points).

Model fit

	Y_1 Est.	Y_1 Data	Y_2 Est.	Y_2 Data	Y_3 Est.	Y_3 Data
A. No periods in high-skill occupation						
Mean						
	2.45	2.45	2.51	2.52	2.57	2.57
Covariance Matrix						
Y_1	0.18	0.17	0.15	0.14	0.14	0.13
Y_2	—	—	0.18	0.19	0.17	0.17
Y_3	—	—	—	—	0.22	0.21
B. Some periods in high-skill occupation						
Mean						
	2.58	2.58	2.65	2.68	2.82	2.80
Covariance Matrix						
Y_1	0.18	0.21	0.12	0.14	0.13	0.12
Y_2	—	—	0.18	0.20	0.15	0.13
Y_3	—	—	—	—	0.22	0.19
C. All periods in high-skill occupation						
Mean						
	2.78	2.76	2.91	2.91	3.01	3.00
Covariance Matrix						
Y_1	0.24	0.26	0.16	0.16	0.16	0.17
Y_2	—	—	0.23	0.21	0.16	0.19
Y_3	—	—	—	—	0.25	0.26

Selection patterns

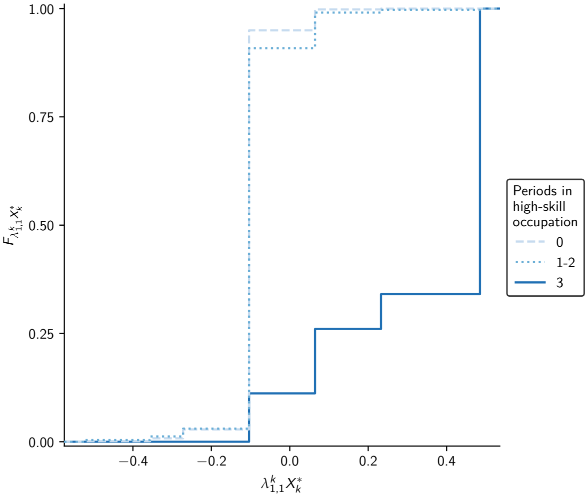


Figure: Selection into high-skill occupation.

Heterogeneity vs. Uncertainty

- We estimate the share of the variance of future wages that is due to heterogeneity (forecastable) vs. uncertainty (unforecastable).
- Focus on the discounted value of wages of the later two periods: $\overline{Y}(d_2) = \sum_{t=2}^{3}(1-\rho)^{t-2} Y_t(d_2)$, for $d_2 \in \{0,1\}$, where $\rho = .05$
- Compute the share of variance that is forecastable before ($t = 0$) and after ($t = 1$) the first period of labor market experience.

- We estimate the share of the variance of future wages that is due to heterogeneity (forecastable) vs. uncertainty (unforecastable).
- Focus on the discounted value of wages of the later two periods: $\overline{Y}(d_2) = \sum_{t=2}^{3} (1-\rho)^{t-2} Y_t(d_2)$, for $d_2 \in \{0, 1\}$, where $\rho = .05$
- Compute the share of variance that is forecastable before $(t = 0)$ and after $(t = 1)$ the first period of labor market experience.

Heterogeneity vs. Uncertainty (Cont'd)

	$\overline{Y}(1)$		$\overline{Y}(0)$	
(t, D^t)	Total Variance	Share Forecastable	Total Variance	Share Forecastable
$(0, \emptyset)$	0.66	0.12	1.07	0.43
	(0.52, 0.77)	(0.07, 0.23)	(0.73, 3.93)	(0.20, 0.85)
$(1, 0)$	0.57	0.65	0.64	0.80
	(0.44, 0.69)	(0.59, 0.74)	(0.56, 0.75)	(0.76, 0.85)
$(1, 1)$	0.70	0.53	1.27	0.76
	(0.57, 0.81)	(0.41, 0.69)	(0.80, 5.40)	(0.64, 0.96)

Table: Decomposition of variance of wages into forecastable and unforecastable components. Note: Each row reports the variance decomposition conditional on a sequence of prior choices. The first row is the variance decomposition at period 0 before the first choices are made. The second and third rows are the variance decomposition conditional on the first occupational choice. The total variance is the variance of $\overline{Y}(d_2)$, conditional having made the choice D^t, which can therefore be a selected sample. The share forecastable is the ratio of the forecastable variance (including both the variance coming from X_k^* and the posterior mean of X_u^* after observing D^t) to the total variance. Bootstrap 95% confidence intervals are given in parentheses.

Concluding remarks

- We establish identification of a potential outcomes model where a portion of the unobserved heterogeneity may be unknown at the time of decisions.
    - Setup with normally distributed errors and unknown heterogeneity, but distribution-free on the known heterogeneity component.
    - Very few restrictions on decision process and learning rule.
    - No auxiliary measurements.
- When individuals do not have private information, relax normality.
- Computationally tractable semi-nonparametric estimation procedure that performs well in finite samples (Python package `spmlex` available on GitHub).
- Illustration to uncertainty and learning in occupational choice.
- Evidence of fast ability revelation based on realized wages.

Thank you!

Concluding remarks

- We establish identification of a potential outcomes model where a portion of the unobserved heterogeneity may be unknown at the time of decisions.
    - Setup with normally distributed errors and unknown heterogeneity, but distribution-free on the known heterogeneity component.
    - Very few restrictions on decision process and learning rule.
    - No auxiliary measurements.
- When individuals do not have private information, relax normality.
- Computationally tractable semi-nonparametric estimation procedure that performs well in finite samples (Python package `spmlex` available on GitHub).
- Illustration to uncertainty and learning in occupational choice.
- Evidence of fast ability revelation based on realized wages.

Thank you!

Concluding remarks

- We establish identification of a potential outcomes model where a portion of the unobserved heterogeneity may be unknown at the time of decisions.
    - Setup with normally distributed errors and unknown heterogeneity, but distribution-free on the known heterogeneity component.
    - Very few restrictions on decision process and learning rule.
    - No auxiliary measurements.
- When individuals do not have private information, relax normality.
- Computationally tractable semi-nonparametric estimation procedure that performs well in finite samples (Python package `spmlex` available on GitHub).
- Illustration to uncertainty and learning in occupational choice.
- Evidence of fast ability revelation based on realized wages.

Thank you!

Concluding remarks

- We establish identification of a potential outcomes model where a portion of the unobserved heterogeneity may be unknown at the time of decisions.
  - Setup with normally distributed errors and unknown heterogeneity, but distribution-free on the known heterogeneity component.
  - Very few restrictions on decision process and learning rule.
  - No auxiliary measurements.
- When individuals do not have private information, relax normality.
- Computationally tractable semi-nonparametric estimation procedure that performs well in finite samples (Python package `spmlex` available on GitHub).
- Illustration to uncertainty and learning in occupational choice.
- Evidence of fast ability revelation based on realized wages.

Thank you!

Assumptions

KL1 For any $d \in \text{Supp}(D_t)$

$$F_{\epsilon_t(d),D_t,Z_t|Y^{t-1},D^{t-1},Z^{t-1}X^*} = F_{\epsilon_t(d)} F_{D_t|Y^{t-1},D^{t-1},Z^t,X_k^*} F_{Z_t|Y^{t-1},D^{t-1},Z_{t-1}}.$$

KL2 $(\lambda_u \mid Z_1 = z_1, X_k^* = v_k) \sim N\left(0, \Sigma_u(z_1, v_k)\right)$ and $\epsilon_t(d) \sim N(0, \sigma_t(d)^2)$.

Assumptions (Cont'd)

KL3 (A) For some d_1, $\alpha_1(d_1) = 0$, $F_{k1}(d_1) = 1$. (B) For some $(d_1, d_2, \ldots, d_p)$, $(F_{u1}(d_1) F_{u2}(d_2) \ldots F_{up}(d_p)) = I_{p \times p}$.

KL4 (A) Θ_1 is a compact set. (B) Supp(X_k^*) is a compact set. (C) For each t, $F_{ut}^{\mathsf{T}}(d_t) \Sigma_t F_{ut}(d_t) + \sigma_t^2(d_t) \neq 0$, $\sigma_t(d_t) \neq 0$ and $\Sigma_u(z_1, v_k)$ is non-singular. (D) $dF_{X_k^* | Y^{t-1}, Z^t, D^t}(v_k; y^{t-1}, z^t, d^t) > 0$ for all for all t and v_k in the support of X_k^*. (E) For each t, the variance-covariance matrix of $(1_n, Z_{it})$ is non-singular.

Assumptions (Cont'd)

KL5 (A) For each d_t there are sequences $d^{t-1}, \tilde{d}^{t-1}$ such that $F_{ut}(d_t)^{\intercal}\Sigma_t \sum_{s=1}^{t-1} \left(F_{us}(d_s)\frac{F_{ks}(d_s)}{\sigma_s^2(d_s)} - F_{us}(\tilde{d}_s)\frac{F_{ks}(\tilde{d}_s)}{\sigma_s^2(\tilde{d}_s)} \right) \neq 0$. (B) For all d_t, $F_{kt}(d_t) \neq 0$. (C) For all d_t, $F_{kt}(d_t) - F_{ut}(d_t)^{\intercal}\Sigma_t \sum_{s=1}^{t-1} F_{us}(d_s)\frac{F_{ks}(d_s)}{\sigma_s^2(d_s)} \neq 0$. (D) For each (d_2, d_1), $F_{u2}(d_2)^{\intercal}\Sigma_2(\lambda_{ui})F_{u1}(d_1)\frac{F_{k1}(d_1)}{\sigma_1^2(d_1)} \neq 0$ (E) There are sets $\{d_{2,i} : i = 1, 2, \ldots, k\}$, $\{\tilde{d}_{2,i} : i = 1, 2, \ldots, k\}$ which are subsets of $\mathrm{Supp}(D_2)$ and satisfy

$$
\begin{aligned}
&(F_{u2}(d_{2,1})F_{u2}(d_{2,2})\ldots F_{u2}(d_{2,k}))^{-\intercal}\,\mathrm{vec}(F_{k2}(d_{2,1}),\ldots,F_{k2}(d_{2,k})) \\
&\neq \left(F_{u2}(\tilde{d}_{2,1})F_{u2}(\tilde{d}_{2,2})\ldots F_{u2}(\tilde{d}_{2,k})\right)^{-\intercal}\,\mathrm{vec}(F_{k2}(\tilde{d}_{2,1}),\ldots,F_{k2}(\tilde{d}_{2,k})).
\end{aligned}
$$

(F) Any $p \times p$ submatrix of $(F_{u1}(d_1)F_{u2}(d_2)\ldots F_{uT}(d_T))$ has full rank Back

Unknown factor only: assumptions

L1 For any $d \in \text{Supp}(D_t)$

$$F_{\epsilon_t(d),D_t,Z_t|Y^{t-1},D^{t-1},Z^{t-1},X_u^*} = F_{\epsilon_t(d)} F_{D_t|Y^{t-1},D^{t-1},Z^t} F_{Z_t|Y^{t-1},D^{t-1},Z^{t-1}}.$$

L2 (A) The joint PDF of (Y, X_u^*) conditional on Z is bounded and continuous, as are all its marginal and conditional densities. (B) $X_u^* \mid Z$ has full support. (C) The characteristic function of $\epsilon_t(d)$ is non-vanishing, and $\mathrm{E}[\epsilon_t|Z, X_u^*] = 0$.

Unknown factor only: assumptions (Cont'd)

L3 For some choice sequence $(d_t \colon t = 1, 2, \ldots, p)$, (A) $(F_1(d_1) \ldots F_p(d_p)) = I_{p \times p}$ and (B) $\alpha_t(d_t) = 0$ for each $t = 1, 2, \ldots, p$.

L4 (A) $f_{Y^{t-1}, Z^t, D^t}(y^{t-1}, z^t, d^t) > 0$ for all t. (B) The variance-covariance matrix of $X_u^* \mid Z$ is full rank.

L5 Any $p \times p$ sub-matrix of $F(d) = (F_1(d_1) F_2(d_2) \ldots F_T(d_T))$ is full rank.

Back

Proof outline

1. $Y_t | Y^{t-1}, D^t, X^t$ is a mixture of normal distributions, with weights $f_{X_k^* | D^t, Y^{t-1}, X^t}(x_k)$ (Lemma 1).

2. Under compact support and regularity conditions, rely on Bruni and Koch (1985, Th.3) to identify the distribution $Y_t | Y^{t-1}, D^t, X^t, X_k^*$ and mixture weights, up to a one-to-one transformation π of X_k^*.

3. $E(Y_t | Y^{t-1}, D^t, X^t, X_k^*)$ is linear in $X_k^* \to \pi$ linear.

4. Rank conditions and factor normalizations $\to \pi = \mathrm{Id} \Rightarrow$ Identification of the distribution of (Y^T, D^T, X^T, X_k^*).

5. Outcome eq. parameters, distribution of unknown component X_u^* and id. shocks: from the distribution of $(Y^T, D^T, X^T | X_k^*)$, weighted by conditional choice probability.

Return

Proof outline

1. $Y_t | Y^{t-1}, D^t, X^t$ is a mixture of normal distributions, with weights $f_{X_k^* | D^t, Y^{t-1}, X^t}(x_k)$ (Lemma 1).

2. Under compact support and regularity conditions, rely on Bruni and Koch (1985, Th.3) to identify the distribution $Y_t | Y^{t-1}, D^t, X^t, X_k^*$ and mixture weights, up to a one-to-one transformation π of X_k^*.

3. $E(Y_t | Y^{t-1}, D^t, X^t, X_k^*)$ is linear in $X_k^* \rightarrow \pi$ linear.

4. Rank conditions and factor normalizations $\rightarrow \pi = \mathrm{Id} \Rightarrow$ Identification of the distribution of (Y^T, D^T, X^T, X_k^*).

5. Outcome eq. parameters, distribution of unknown component X_u^* and id. shocks: from the distribution of $(Y^T, D^T, X^T | X_k^*)$, weighted by conditional choice probability.

Return

Proof outline

1. $Y_t | Y^{t-1}, D^t, X^t$ is a mixture of normal distributions, with weights $f_{X_k^* | D^t, Y^{t-1}, X^t}(x_k)$ (Lemma 1).

2. Under compact support and regularity conditions, rely on Bruni and Koch (1985, Th.3) to identify the distribution $Y_t | Y^{t-1}, D^t, X^t, X_k^*$ and mixture weights, up to a one-to-one transformation π of X_k^*.

3. $E(Y_t | Y^{t-1}, D^t, X^t, X_k^*)$ is linear in $X_k^* \rightarrow \pi$ linear.

4. Rank conditions and factor normalizations $\rightarrow \pi = \mathrm{Id} \Rightarrow$ Identification of the distribution of (Y^T, D^T, X^T, X_k^*).

5. Outcome eq. parameters, distribution of unknown component X_u^* and id. shocks: from the distribution of $(Y^T, D^T, X^T | X_k^*)$, weighted by conditional choice probability.

Return

Proof outline

1. $Y_t | Y^{t-1}, D^t, X^t$ is a mixture of normal distributions, with weights $f_{X_k^* | D^t, Y^{t-1}, X^t}(x_k)$ (Lemma 1).

2. Under compact support and regularity conditions, rely on Bruni and Koch (1985, Th.3) to identify the distribution $Y_t | Y^{t-1}, D^t, X^t, X_k^*$ and mixture weights, up to a one-to-one transformation π of X_k^*.

3. $E(Y_t | Y^{t-1}, D^t, X^t, X_k^*)$ is linear in $X_k^* \rightarrow \pi$ linear.

4. Rank conditions and factor normalizations $\rightarrow \pi = \mathrm{Id} \Rightarrow$ Identification of the distribution of (Y^T, D^T, X^T, X_k^*).

5. Outcome eq. parameters, distribution of unknown component X_u^* and id. shocks: from the distribution of $(Y^T, D^T, X^T | X_k^*)$, weighted by conditional choice probability.

Return

Likelihood

The individual log-likelihood contribution is:

$$\log \int \prod_{t=1}^{T} \left(\frac{1}{\sigma_t\left(d_t\right)} \phi_1 \left(\frac{y_t - \alpha_t\left(d_t\right) - z_t' \beta_t\left(d_t\right) - v_u' F_{ut}\left(d_t\right) - v_k F_{kt}\left(d_t\right)}{\sigma_t\left(d_t\right)} \right) \right.$$

$$\left. \times \, \bar{h}_t(d_t, z_t, v_k, x_t) \right) \times \prod_{t=1}^{T-1} g_t(z_{t+1} \mid z_t, y_t, d_t)$$

$$\times \frac{1}{\sqrt{\left|\Sigma_u\left(z_1, v_k\right)\right|}} \phi_p \left(\Sigma_u^{-\frac{1}{2}}\left(z_1, v_k\right) v_u \right) \times dF_{X_k^*}\left(v_k; z_1\right) dv_u$$

▸▸ Back

CCP sieve space

Utility:

- Linear index in $(Y_1, \ldots, Y_{t-1}, X_k^*)$ where coefficients are conditional on full choice sequence.
- Nonlinear transformation of the linear index by a polynomial of order $m_n \to \infty$.

$$u_t(d|h) = f^{m_n}(\pi_{0,t}(d^{t-1}) + \pi_{1,t}(d^{t-1})y_1 + \ldots + \pi_{t-1,t}(d^{t-1})y_{t-1} + \pi_{k,t}(d^{t-1})x_{ik}^*$$

$$f^{m_n}(y) = \sum_{s=1}^{m_n} f_s y^{s-1}$$

Probabilities obtained by assuming an EV1 additive choice shock

$$p_t(d|h) = \frac{\exp(u_t(d|h))}{\sum_{d'} \exp(u_t(d'|h))}$$

Back

Sieve MLE estimation - functionals

- Simple example with static choice (e.g. occupational choice). For a rate of time preference ρ, the present value of lifetime earnings is:

$$\widetilde{Y}_{t_0}(d) = \sum_{t=t_0}^{T} \frac{Y_t(d)}{(1+\rho)^{t-t_0}}$$

- Predictable component is given by, denoting by $\mathcal{I}_{t_0}$ the information set at time $t = t_0$:

$$E(\widetilde{Y}_{t_0}(d)|\mathcal{I}_{t_0})$$

where we assume that $\mathcal{I}_{t_0} = \{X_k^*, W^{t_0}\}$ with $W^t = (Y^{t-1}, D^{t-1}, Z^t)$.

Sieve MLE estimation - functionals

- Simple example with static choice (e.g. occupational choice). For a rate of time preference ρ, the present value of lifetime earnings is:

$$\widetilde{Y}_{t_0}(d) = \sum_{t=t_0}^{T} \frac{Y_t(d)}{(1+\rho)^{t-t_0}}$$

- Predictable component is given by, denoting by $\mathcal{I}_{t_0}$ the information set at time $t = t_0$:

$$E(\widetilde{Y}_{t_0}(d)|\mathcal{I}_{t_0})$$

where we assume that $\mathcal{I}_{t_0} = \{X_k^*, W^{t_0}\}$ with $W^t = (Y^{t-1}, D^{t-1}, Z^t)$.

Sieve MLE estimation - functionals (Cont'd)

Variance of the predictable and unpredictable components of $\widetilde{Y}_{t_0}(d)$ are given by:

$$\sigma^2_{k,t_0}(d) = \int \left(E(\widetilde{Y}_{t_0}(d)|\mathcal{I}_{t_0}) - E(\widetilde{Y}_{t_0}(d)) \right)^2 dF_{X^*_k, W^{t_0}}(x^*_k, w^{t_0})$$

$$\sigma^2_{u,t_0}(d) = \int \mathrm{Var}(\widetilde{Y}_{t_0}(d)|\mathcal{I}_{t_0}) dF_{X^*_k, W^{t_0}}(x^*_k, w^{t_0})$$

- As agents learn and update their beliefs about $X^*_u \Rightarrow$ Evolution over time of share of predictable/unpredictable earnings variance.
- We wish to estimate and conduct inference on these types of parameters.

Sieve MLE estimation - functionals (Cont'd)

- We provide in the paper general results for the following class of functionals. Namely, consider a function f_1 which maps (θ, w, x_k^*) to $\mathbb{R}$ such that $f_1(\theta, w, x_k^*)$ is given by:

$$g\left(\mathrm{E}\left[\sum_{i \in I_t} \omega_i Y_{t_i}(d_i) \mid W^t = w, X_k^* = x_k^*\right], \mathrm{Var}\left[\sum_{i \in I_t} \omega_i Y_{t_i}(d_i) \mid W^t = w, X_k^* = x_k^*\right]\right)$$

- We define the functional of θ as

$$f(\theta) = \int f_1(\theta, w, x_k^*) dF_{W^t, X_k^*}(w, x_k^*).$$

$\Rightarrow$ We propose to estimate $f(\theta^*)$ via plug-in sieve MLE, and establish consistency and asymptotic normality.

Back

Sieve MLE estimation - functionals (Cont'd)

- We provide in the paper general results for the following class of functionals. Namely, consider a function f_1 which maps (θ, w, x_k^*) to $\mathbb{R}$ such that $f_1(\theta, w, x_k^*)$ is given by:

$$g\left(\mathrm{E}\left[\sum_{i\in I_t}\omega_i Y_{t_i}(d_i) \mid W^t = w, X_k^* = x_k^*\right], \mathrm{Var}\left[\sum_{i\in I_t}\omega_i Y_{t_i}(d_i) \mid W^t = w, X_k^* = x_k^*\right]\right)$$

- We define the functional of θ as

$$f(\theta) = \int f_1(\theta, w, x_k^*) dF_{W^t, X_k^*}(w, x_k^*).$$

$\Rightarrow$ We propose to estimate $f(\theta^*)$ via plug-in sieve MLE, and establish consistency and asymptotic normality.

Back