Lecture 18

Marginal and Distributional Treatment Effects

Tyler Ransom
ECON 6343, University of Oklahoma

Attribution

Today's material is based on lecture notes from Arnaud Maurel (Duke University)

I have adjusted his materials to fit the needs and goals of this course

Plan for the Day

1. Marginal Treatment Effects

2. Policy-Relevant Treatment Effects

3. Distributional Treatment Effects

Marginal Treatment Effects

- **Essential heterogeneity:** the standard IV approach does not identify the ATE, nor any standard average treatment effect parameter
- Heckman and Vytlacil (2005): under some restrictions on the selection rule and with an instrument, one can identify **Marginal Treatment Effects** (MTE, Bjorklund and Moffitt (1987)) for individuals at certain margins
- **Key insight:** all the standard TE parameters can be expressed as weighted averages of the MTE parameters
- Useful framework that highlights the conditions under which standard, or policy relevant TE parameters, are identified from the data

Econometric framework

- Nonseparable outcome model:

$$
\begin{aligned}
Y_1 &= \mu_1(X, U_1) \\
Y_0 &= \mu_0(X, U_0)
\end{aligned}
$$

- Decision rule for selection into treatment:

$$
D = 1[\mu_D(Z) - U_D \geq 0]
$$

Additive separability on the index is key here

- (Z, X) are observed, (U_0, U_1, U_D) unobserved
- Exclusion restriction: Z contains all the X plus one element not in X

Econometric framework (Continued)

Assumptions maintained throughout the analysis:

1. $\mu_D(Z)$ is a nondegenerate random variable conditional on X (Relevance assumption)
2. (U_1, U_D) and (U_0, U_D) are independent of Z conditional on X (Exclusion restriction assumption)
3. The distribution of U_D is absolutely continuous w.r.t. the Lebesgue measure
4. $E|Y_1|$ and $E|Y_0|$ are finite
5. $0 < P(D = 1|X) < 1$
6. No feedback condition: $X_{D=1} = X_{D=0}$ almost surely

- Overall, the model is quite general

- The availability of a valid instrument (conditions 1-2) is most restrictive in practice

Econometric framework (Continued)

- Selection probability: $P(Z) = P(D = 1|Z) = F_{U_D|X}(\mu_D(Z))$
- Normalization: $U_D \sim U_{[0,1]}$ and $\mu_D(Z) = P(Z)$
- Vytlacil (2002): this model, with the assumptions 1-5, is equivalent to the LATE model of Imbens and Angrist (1994)

The Marginal Treatment Effect

$$\Delta^{MTE}(x, u_d) = E(Y_1 - Y_0 | X = x, U_d = u_d)$$

- This is the mean effect of treatment conditional on the observable characteristics X and the unobservables U_D from the selection equation

- Other interpretation: mean effect of treatment for individuals with observable characteristics X who would be indifferent between treatment or not if exogenously assigned a value Z (z) such that $\mu_D(z) = u_d$

- Essential heterogeneity: $\Delta^{MTE}(x, u_d)$ varies with u_d

Connection between MTE and LATE

- Consider the case of a binary instrument $Z \in \{0, 1\}$
- The LATE writes:

$$E(Y_1 - Y_0|X = x, D_0 = 0, D_1 = 1) =$$
$$E(Y_1 - Y_0|X = x, u_D' \leq U_D \leq u_D)$$

where $u_D' = P(D_0 = 0) = P(0)$ and $u_D = P(D_1 = 1) = P(1)$

- For $u_D' \longrightarrow u_D$: $\Delta^{LATE}(x, u_d', u_d) \longrightarrow \Delta^{MTE}(x, u_d)$

MTE and ATE, ATT, ATU

- Key result: all the standard treatment effect parameters are obtained as weighted averages of MTE

- ATE:

$$
\begin{aligned}
E(Y_1 - Y_0 | X = x) &= E_{U_d|X=x}(E(Y_1 - Y_0 | X = x, U_d = u_d)) \\
&= \int_0^1 \Delta^{MTE}(x, u_d) du_d
\end{aligned}
$$

MTE and ATT

Similarly for the ATT:

$$E(Y_1 - Y_0|X = x, D = 1) = \int_0^1 \Delta^{MTE}(x, u_d)h_{TT}(x, u_d)du_d$$

with the weights:

$$h_{TT}(x, u_D) = \frac{1 - F_{P|X=x}(u_D)}{\int_0^1 (1 - F_{P|X=x}(t))dt}$$

Interpretation: the TT oversamples the MTE for the individuals with low values of U_D

MTE and IV estimator

- In the presence of essential heterogeneity, the standard IV approach in general does not identify any of the standard mean treatment effect parameters

- Still, the IV estimand can also be obtained as a weighted sum of the MTE parameters. Consider a function $J(Z)$ as an instrument, the IV estimand writes:

$$\beta_{IV}(x, J) = \frac{Cov(J(Z), Y|X = x)}{Cov(J(Z), D|X = x)}$$

- Denoting $\tilde{J}(Z) = J(Z) - E(J(Z)|X = x)$, we can show that the numerator writes:

$$\int_0^1 \Delta^{MTE}(x, u_d)E(\tilde{J}(Z)|X = x, P(Z) \geq u_D)P(P(Z) \geq u_d|X)du_d$$

MTE and IV estimator (Continued)

- And from the law of iterated expectations, we have: $Cov(J(Z), D|X = x) = Cov(J(Z), P(Z)|X = x)$
- This implies that $\beta_{IV}(x, J)$ is a weighted average of the MTE, with the weighting function $h_{IV}(u_D|x, J)$:

$$\frac{E(\tilde{J}(Z)|X = x, P(Z) \geq u_D)P(P(Z) \geq u_d|X)}{Cov(J(Z), P(Z)|X = x)}$$

- $P(Z)$ plays a special role here. In the case where $J(Z) = P(Z)$, the weights are nonnegative for all evaluation points
- In general this is not the case
- This may lead to counterintuitive interpretations of the IV estimator: e.g. all the MTEs are positive, but the IV estimate is negative!

Identification of the MTE

- Key question given the central role played by the MTE in policy analysis
- Idea: use $J(Z) = P(Z)$ as an instrument (**Local Instrumental Variable**, LIV) approach
- LIV is defined (and identified) by:

$$\Delta^{LIV}(x,p) = \frac{\partial E(Y|X = x, P(Z) = p)}{\partial p}$$

- Note that:

$$\begin{aligned} E(Y|X = x, P(Z) = p) &= E(Y_0|X = x, P(Z) = p) + E(D(Y_1 - Y_0)|X = x, P(Z) = p) \\ &= E(Y_0|X = x) + E(Y_1 - Y_0|X = x, P(Z) = p, D = 1)p \\ &= E(Y_0|X = x) + E(Y_1 - Y_0|X = x, p \geq U_D)p \end{aligned}$$

Identification of the MTE (Continued)

- Defining U_k by: $Y_k = E(Y_k|X) + U_k$ (for $k \in \{0, 1\}$), we have:

$$E(Y|X = x, P(Z) = p) = E(Y_0|X = x) + \Delta^{ATE}(x)p + \int_0^p E(U_1 - U_0|X = x, U_D = u_D)du_D$$

- Taking the derivative with respect to p (the LIV) directly yields the MTE evaluated in $U_D = p$:

$$\frac{\partial E(Y|X = x, P(Z) = p)}{\partial p} = \Delta^{ATE}(x) + \underbrace{E(U_1 - U_0|X = x, U_D = p)}_{\text{(Selection bias)}} = \Delta^{MTE}(x, p)$$

Identification of the MTE (Continued)

- This suggests a simple test for essential heterogeneity, which corresponds to the nonlinearity of $E(Y|X = x, P(Z) = p)$ as a function of p
- Note this only identifies the MTE on the support of P conditional on $X = x$
- Thus in general, the standard treatment effect parameters are **not** identified, unless $P(Z)$ has full support (conditional on X)
- Of course, also need a continuous instrument to identify the partial derivative w.r.t. p
- See Brinch, Mogstad, and Wiswall (2017) who study identification with a discrete instrument for a linear MTE model

Policy-Relevant Treatment Effects

Policy Relevant Treatment Effects

- In the heterogeneous case, different TE parameters answer different policy questions
- Another look at the situation: define a treatment parameter that answers a given policy question (**Policy Relevant Treatment Effect**, PRTE)
- Consider a class of policies that affect P, but not the MTE
- i.e. if the policy does not change the potential outcomes (no GE consideration here), covariates X or unobservables U_D
    - Think about tuition fees in the context of schooling

Policy Relevant Treatment Effects (Continued)

- Denoting by a and a' two potential policies, the PRTE writes:

$$E(Y_a|X = x) - E(Y_{a'}|X = x)$$
$$= \int_0^1 E(Y_1 - Y_0|X = x, U_D = u_D) h_{PRTE}(x, u_D) du_D$$

- with the weights:

$$h_{PRTE}(x, u_D) = F_{P_{a'}|X=x}(u_D) - F_{P_a|X=x}(u_D)$$

- PRTE weights for those affected by the policy:

$$\frac{h_{PRTE}(x, u_D)}{(E(P_a|X = x) - E(P_{a'}|X = x))}$$

(difference in conditional selection probability)

Policy Relevant Treatment Effects (Continued)

- Heckman and Vytlacil (2005)
  - for a given policy, possible to construct an IV such that the standard IV approach yields the PRTE
  - … under certain regularity conditions
- Another natural question is the following:
  - for a given instrument, is it always possible to find a policy such that the IV identifies the corresponding PRTE?
  - The answer is no in general

Carneiro, Heckman, and Vytlacil (2011)

- Apply the MTE-Local IV approach to estimate the returns to college for individuals induced to attend college by a marginal policy change
- **Marginal** benefits (and costs) are the relevant units to assess the optimality of a policy
- Their results provide evidence (from the NLSY79) of essential heterogeneity in the returns to college
  - individuals with higher returns are more likely to attend college
- Comparison with standard IV estimates of the returns to college
  - Findings: IV estimate is 9.5% but MTE is only 1.5%!

Econometric framework

- Model similar to Heckman and Vytlacil (2005)
- Y_1 (or Y_0) denotes the potential log-wage if the individual were to attend college (or not attend college)
- $S \in \{0, 1\}$ is a dummy for college attendance (treatment)

Econometric framework (Continued)

- Additive separability is assumed for the outcome equations:

$$Y_1 = \mu_1(X) + U_1$$
$$Y_0 = \mu_0(X) + U_0$$

  with $E(Y_S|X = x) = \mu_S(x)$
- Decision to enroll in college:

$$S = 1\{\mu_S(Z) - V \geq 0\}$$

Econometric framework (Continued)

- Key assumption: at least one exclusion restriction between μ_S and (μ_0, μ_1), with one component of Z being excluded from the outcome equations
- Exogeneity condition: $Z \perp (U_0, U_1, V)|X$
- Variables such as tuition fees or distance to college are natural candidates for this exclusion restriction in this context
- As above, the selection equation can be rewritten as:

$$S = 1\{P(Z) - U_S \geq 0\}$$

with $P(Z) = P(S = 1|Z)$ and U_S uniformly distributed

Effect of a marginal policy change: MPRTE

- New treatment effect parameter: **Marginal Policy Relevant Treatment Effect (MPRTE)**
- Marginal version of the PRTE parameter
- PRTE: in general only identified if the selection probability $P(Z)$ has full support
- MPRTE does not require this support condition
- The MPRTE answers the following question: what is the effect of a marginal change from a baseline policy?
- In the context of schooling, what are the wage returns of a marginal decrease in tuition fees?

Effect of a marginal policy change: MPRTE (Continued)

More formal definition of the MPRTE from the PRTE:

- Consider a sequence of policies indexed by α, with $\alpha = 0$ denoting status quo
- Each policy is associated with a selection probability P_α and a $PRTE_\alpha$ corresponding to the effect of going from the baseline policy to policy α
- The MPRTE is defined as the limit of the parameters $(PRTE_\alpha)_\alpha$ for $\alpha \longrightarrow 0$

Effect of a marginal policy change: MPRTE (Continued)

- Carneiro, Heckman, and Vytlacil (2010) provide a detailed analysis of the properties of the MPRTE
- Important point: MPRTE is not unique in the sense that it depends on the nature of the marginal policy change which is considered
- Context of college attendance: effect of a marginal policy change subsidizing tuitions by a fixed amount will be different from the one of a policy change subsidizing tuitions by a fixed proportion

Effect of a marginal policy change: MPRTE (Continued)

- Carneiro, Heckman, and Vytlacil (2010) show that the MPRTE can also be written as weighted sum of MTE parameters
- Key result: the weighting function for the MPRTE is always zero outside of the support of the selection probability $P(Z)$
- A valid continuous instrument Z for selection is sufficient to pin down the MPRTE, no strong support conditions on $P(Z)$
    - Continuity of the distribution of the instrument remains important to identify the MTE
- Overall, this is an advantage of **marginal** vs. **average** TE parameters

Another interpretation of the MPRTE

- The MPRTE has another nice interpretation
- It corresponds to the average treatment effect for the individuals who are indifferent between being treated or not (the Average Marginal Treatment Effect, AMTE)
- Idea: consider the average treatment effect for individuals arbitrarily close to the margin of indifference (for a certain metric)
- Carneiro, Heckman, and Vytlacil (2010) show that the MPRTE associated with different policy shifts correspond to AMTE associated with different metrics

Application to the returns to college

- Carneiro et al. apply their method to the estimation of the returns to college from the NLSY79 data
- They consider a sample of white males, who attended college $(S = 1)$, dropped out from high school or entered the labor market after graduating from high school $(S = 0)$

Estimation of MPRTE

Semiparametric estimation of the MPRTE:

1. Estimation of the MTE over the support of the selection probability $P(Z)$ using a local IV approach
2. Estimation of the MPRTE (weighted sum of the MTE over the support of $P(Z)$)

Step 1: Estimation of the MTE

Under the previous assumptions:

$$\Delta^{MTE}(x,p) = \frac{\partial E(Y|X=x, P(Z)=p)}{\partial p}$$

- $\Delta^{MTE}(x,p)$ can be estimated nonparametrically for all p in the support of $P(D=1|Z,X=x)$ using a nonparametric regression (Nadaraya-Watson)
- Problem: the estimation becomes very imprecise if the dimension of the vector of covariates X is large
- This is typically the case in practice (**curse of dimensionality**)

Step 1: Estimation of the MTE (Continued)

- More convenient to work with the partially linear regression considered above
- Three steps:

1. Estimation of the selection probability $P(Z)$ using a parametric specification
2. Estimation of (δ_0, δ_1) by regressing $Y - E(Y|P)$ on $X - E(X|P)$ and $P(X - E(X|P))$ (Robinson, 1988)
3. Estimation of $K(P)$ and $K'(P)$

- $\Delta^{MTE}(x, p)$ may be estimated for all P in the support of $P(Z)$

Step 2: Estimation of the MPRTE

1. Estimate nonparametrically the weights $h_{MPRTE}(X, p)$ (e.g. $h(X, p) = f_{P|X}(p)$ for the case of a policy of the form $P_\alpha = P + \alpha$) associated with the different MTE

2. Then integrate over the distribution of X

Distributional Treatment Effects

The evaluation problem

Recall that this is one of the two main problems related to the identification of treatment effects:

- For each individual we only observe either Y_0 or Y_1, but never both
- This is true even in the presence of a perfect randomization, where $(Y_0, Y_1) \perp D$
- In the absence of selection, marginal distributions of potential outcomes are identified
- However, this is not the case of the joint distribution of (Y_0, Y_1)
- In particular, the distribution of the treatment effects $Y_1 - Y_0$ is not identified without additional restrictions

The evaluation problem (Continued)

- We will focus on the evaluation problem, assuming that the selection problem has been dealt with using one of the methods previously discussed

- In other words, let us assume that we have identified the marginal distributions of the potential outcomes (conditional on observables X), $F_{Y_0|X}$ and $F_{Y_1|X}$

- How can we identify the joint distribution of potential outcomes, $F_{Y_0,Y_1|X}$?

Why should we care about the distribution of TE's?

Identifying the joint distribution of counterfactual outcomes is necessary to know how the benefits to treatment are distributed in the population. Needed to:

- Identify the proportion of individuals who benefit from the treatment, $P(Y_1 > Y_0|X)$
- Identify the dispersion of the benefits (or losses) from treatment
- Compare the distribution of *ex ante* (at the time of the treatment decision) and *ex post* treatment effects

Identifying the joint distribution of (Y_0, Y_1)

Econometric methods allowing to identify the joint distribution of potential outcomes are typically based on assumptions on the structure of dependence between Y_0 and Y_1

Partial identification

One can derive "no-assumption" bounds on the joint distribution of (Y_0, Y_1) from the marginal distributions of Y_0 and Y_1. Classical result in statistics (**Frechet-Hoeffding bounds**)

$$F(y_0, y_1|X) \in [\underline{F}(y_0, y_1, X), \overline{F}(y_0, y_1, X)]$$

with:

$$\underline{F}(y_0, y_1, X) = \max\left(F_{Y_0|X}(y_0) + F_{Y_1|X}(y_1) - 1, 0\right)$$

and:

$$\overline{F}(y_0, y_1, X) = \min\left(F_{Y_0|X}(y_0), F_{Y_1|X}(y_1)\right)$$

Partial identification (Continued)

- Both the upper and lower bound are proper probability distribution (Mardia, 1970), and these bounds are sharp (Rueschendorf, 1981)
- However, in practice these bounds are most often quite wide
- See Heckman and Smith (1993) and Heckman, Smith, and Clements (1997) for applications of these bounds

Dependence assumptions

(Point) identification of the joint distribution of (Y_0, Y_1) can be obtained by restricting the dependence between the potential outcomes. Extreme case: homogeneous treatment effects $(Y_1 - Y_0 = C)$

- Under the assumption that $Y_1 - Y_0 = \Delta = C$, the joint distribution $F(y_0, y_1|X)$ reduces to a one-dimensional distribution. For all (y_0, y_1) such that $y_1 - y_0 \geq \Delta$, we have:

$$\begin{aligned} F(y_0, y_1|X) &= F(y_0, y_0 + \Delta|X) \\ &= F_{Y_0|X}(y_0) \end{aligned}$$

- And $F(y_0, y_1|X) = F_{Y_0|X}(y_1 - \Delta)$ otherwise. Thus the joint distribution is identified from the marginal $F_{Y_0|X}$

Dependence assumptions (Continued)

The joint distribution of (Y_0, Y_1) can also be identified by extending the matching Conditional Independence Assumption to the dependence *between* Y_0 and Y_1

- Consider the standard matching CIA assumption: $Y_0 \perp D|X$ and $Y_1 \perp D|X$
- Augmented by the following conditional independence assumption: $Y_0 \perp Y_1|X$
- Then it directly follows that:

$$\begin{aligned} F(y_0, y_1|X) &= F_{Y_0|X}(y_0)F_{Y_1|X}(y_1) \\ &= F_{Y|X,D=0}(y_0)F_{Y|X,D=1}(y_1) \end{aligned}$$

  which identifies the joint distribution $F(y_0, y_1|X)$

Dependence assumptions (Continued)

Another approach: assume rank invariance (or rank reversal) of the positions of individuals in the marginal distributions $F_{Y_0|X}$ and $F_{Y_1|X}$

- See, e.g., Robins (1997), Heckman, Smith, and Clements (1997)
- Rank invariance:

$$Y_1 = F^{-1}_{Y_1|X}(F_{Y_0|X}(Y_0))$$

Dependence assumptions (Continued)

- Under rank invariance, which generalizes the assumption of homogeneous treatment effects, we can show that:

$$F(y_0, y_1|X) = \min\left(F_{Y_0|X}(y_0), F_{Y_1|X}(y_1)\right)$$

  This corresponds to the upper Frechet bound

- Alternatively, assuming rank reversal

$$Y_1 = F^{-1}_{Y_1|X}(1 - F_{Y_0|X}(y_0))$$

the joint distribution is equal to the lower Frechet bound

Restrictions on selection and dependence

More recently papers have imposed restrictions on the dependence between the unobservables of the model, combined with assumptions on the selection rule: Carneiro, Hansen, and Heckman (2003); Aakvik, Heckman, and Vytlacil (2005); Cunha and Heckman (2007)

- These papers use a **factor model** to restrict the dependence between the unobservables of the model and pin down the joint distribution of the potential outcomes
- Key idea: dependence across the unobservables is generated by a low-dimensional set of mutually independent random variables (factors)

Restrictions on selection and dependence (Continued)

- Main idea: express the joint distribution of the counterfactuals as a function of the marginals, and the distribution of the factors
- Access to measurements of these factors allows to identify their distribution
- As in Heckman and Vytlacil (2005), additively separable latent selection rule
- Exclusion restrictions between the selection and outcome equations are used to circumvent the selection problem

Factor models

- Key identifying assumptions: there exists a vector of unobserved factors θ such that

$$(Y_0, Y_1) \perp D|X, \theta$$

- Extension of the matching CIA, where matching is done on both observed X and unobserved θ variables

- The factor approach also builds on the second conditional independence assumption:

$$Y_0 \perp Y_1|X, \theta$$

  Identification of the distribution of the factors relies on the existence of a "sufficient" number of proxies for those factors

Factor models: set-up

- Multiple treatments: s treatment states, with $s \in \{1, \dots, S\}$
- A vector of (potential) outcomes $Y(s, X) = (Y(k, s, X))_{k \in \{1,\dots,N_Y\}}$ for each s
- A vector of $M(X) = (M(k, X))_{k \in \{1,\dots,N_M\}}$ measurements which **do not** depend on the treatment state s
- The outcomes and measurements may be discrete or continuous. We assume continuity here, and the following additively separable forms:

$$M(k, X) = \mu_M(k, X) + U_M(k)$$
$$Y(k, s, X) = \mu(k, s, X) + U(k, s)$$

where $(U(k, s), U_M(k))$ are unobservable (to the econometrician) random components

Factor models: selection model

- Utility of each treatment state:

$$R(s, Z) = \mu_R(s, Z) - V(s)$$

where Z is a vector of observed covariates affecting the choice of treatment, and $V(s)$ the unobserved component of utility

- Selection into a treatment status s:

$$s = \arg\max_k \{R(k, Z)\}$$

Factor models (Continued)

Linear factor structure for the unobservables:

$$
\begin{aligned}
U(k,s) &= \alpha'_{k,s}\theta + \varepsilon_k(s) \\
U_M(j) &= \alpha'_{j,M}\theta + \varepsilon_{j,M} \\
V(s') &= \alpha'_{V(s')}\theta + \varepsilon_V(s')
\end{aligned}
$$

where we assume mutual independence between the components of θ, between θ and the vector of the idiosyncratic shocks $\varepsilon = (\varepsilon_k(s), \varepsilon_{j,M}, \varepsilon_V(s'))_{k,j,s,s'}$, and between the idiosyncratic shocks

- We also assume independence between the covariates (X, Z) and the unobservables (θ, ε)

Factor models: Identifying the joint distribution of counterfactuals

To identify the joint distribution of counterfactual outcomes for each pair of treatment alternatives (s, s') (with $s \neq s'$), we need to identify the joint distribution of the unobservables $(U(s), U(s'))$, where $U(s) \equiv (U(1, s), \ldots, U(k, s), \ldots, U(N_Y, s))'$

- First assume that the **factor loadings** $(\alpha_{k,s})_{k,s}$ are known. Then the joint distribution of $(U(s), U(s'))$ only depends on the marginal distributions of θ, $(\varepsilon_k(s))_k$ and $(\varepsilon_k(s'))_k$
- Selection has been addressed in a first step: the effects of the covariates X on the outcomes, $(\mu(k, s, X))_{k,s}$, as well as the marginal distributions of the unobservables $(U(k, s))_{k,s}$ are identified

Kotlarski's theorem

Key result: Kotlarski's theorem (Kotlarski, 1967)

Consider the following decomposition:

$$T_1 = \theta + \varepsilon_1$$
$$T_2 = \theta + \varepsilon_2$$

where $(\theta, \varepsilon_1, \varepsilon_2)$ are mutually independent, the means of θ, ε_1 and ε_2 are finite and $E(\varepsilon_1) = E(\varepsilon_2) = 0$, the random variables possess nonvanishing characteristic functions and the joint distribution of (T_1, T_2) is identified

Then the distributions of $(\theta, \varepsilon_1, \varepsilon_2)$ are identified

Kotlarski's theorem (Continued)

Proof: Let us consider three mutually independent random variables (U_1, U_2, U_3) such that:

$$T_1 = U_1 + U_2$$
$$T_2 = U_1 + U_3$$

with $E(U_2) = E(U_3) = 0$. Denoting by $\phi(t_1, t_2)$ the characteristic function of (T_1, T_2), $(\phi_1(t), \phi_2(t), \phi_3(t))$ the characteristic functions of $(\theta, \varepsilon_1, \varepsilon_2)$ and $\psi_k(t)$ the characteristic functions of U_k, we have, for all $(t_1, t_2) \in R^2$:

$$\begin{aligned}\phi(t_1, t_2) &= E(\exp(i(t_1 T_1 + t_2 T_2))) \\ &= E(\exp(i(t_1 \varepsilon_1 + t_2 \varepsilon_2 + (t_1 + t_2)\theta))) \\ &= \phi_2(t_1)\phi_3(t_2)\phi_1(t_1 + t_2)\end{aligned}$$

Kotlarski's theorem (Continued)

Proof (Continued): Similarly, it follows from the mutual independence of (U_1, U_2, U_3) that:

$$\phi(t_1, t_2) = \psi_2(t_1)\psi_3(t_2)\psi_1(t_1 + t_2)$$

Defining $r_k(t)$ as the ratio of the characteristic functions $\frac{\psi_k}{\phi_k}(t)$, this implies:

$$r_2(t_1)r_3(t_2)r_1(t_1 + t_2) = 1$$

The $r_k(\cdot)$ are unknown, complex functions, continuous over the real line and such that $r_k(0) = 1$

Kotlarski's theorem (Continued)

Proof (Continued): Setting respectively $t_2 = 0$ and $t_1 = 0$, we have the following equalities:

$$r_2(t_1)r_1(t_1) = 1$$
$$r_3(t_2)r_1(t_2) = 1$$

And substituting into the equality from the previous slide, this yields, for all (t_1, t_2):

$$r_1(t_1 + t_2) = r_1(t_1)r_1(t_2)$$

This implies, with $r_1(0) = 1$, that: $r_1(t) = \exp(\alpha t)$ (where $\alpha \in C$). It follows that: $r_2(t) = r_3(t) = \exp(-\alpha t)$

Kotlarski's theorem (Continued)

Proof (Continued): Finally, considering the case of $r_1(\cdot)$ (same for r_2 and r_3), we have:

$$\psi_1(t) = \exp(\alpha t)\phi_1(t)$$

Using the fact that $\psi_1(-t) = \overline{\psi_1(t)}$ (characteristic function), it follows that $\exp(-\alpha t) = \overline{\exp(\alpha t)}$, which implies:

$$\psi_1(t) = \exp(i\beta t)\phi_1(t)$$

where $\beta \in \mathbb{R}$. Hence the distribution of θ is identified up to the location β (resp. $-\beta$ for ε_1 and ε_2). The zero mean condition identifies β, which finally yields full identification of the distribution of $(\theta, \varepsilon_1, \varepsilon_2)$

Identifying the factor loadings

We now turn to the identification of the factor loadings. Consider the system of factor decompositions for all the unobservables of the model, for all s:

$$U(s) = \Lambda(s)\theta + \varepsilon(s)$$

where we assume that there are K factors and $L(s)$ components of $U(s)$ From second-order moments, we identify, denoting by $V(X)$ the covariance matrix of a random vector X:

$$V(U(s)) = \Lambda(s)V(\theta)\Lambda'(s) + V(\varepsilon(s))$$

Some restrictions are necessary to identify $\Lambda(s)$ from the second-order moments. Aside from the independence between the factors, we also assume that one factor loading for each factor is equal to one (innocuous scale normalization)

Identifying the factor loadings (Continued)

To be able to recover the factor loadings from the second-order moments, we can only use the $\frac{L(s)(L(s)-1)}{2}$ non-diagonal terms of $V(U(s))$ (the covariances), which both depend on the factor loadings and the variance of the factors. Rank condition:

$$\begin{aligned} \frac{L(s)(L(s)-1)}{2} &\geq \quad (L(s)K-K)+K \\ &\Leftrightarrow \quad L(s) \geq 2K+1 \end{aligned}$$

where $(L(s)K-K)$ is the number of unrestricted factor loadings and K the number of (unknown) factor variances. Note that this is only a **necessary** condition

Identifying the factor loadings (Continued)

- Most of the factor models used in the empirical literature only make use of the first and second-order moments, leading to the preceding necessary condition for identification
- In the case where the measurements are **not** normally distributed, one may also use higher-order moments to obtain more information on the factor loadings and the distributions of the factors. This allows to identify the factor loadings, the distribution of the factors and therefore the joint distribution of the outcomes under weaker conditions
- See Bonhomme and Robin (2009); Bonhomme and Robin (2010). In particular, they show that if all the factors display kurtosis, one may identify the factor loadings for up to $K = L$ factors
- The idea dates back to Geary (1942) and Reiersol (1950), who have shown that factor loadings are identified in a measurement error model if the factor is **not** Gaussian

References

Aakvik, A, J. J. Heckman, and E. J. Vytlacil (2005). "Estimating Treatment Effects for Discrete Outcomes When Responses to Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs". In: *Journal of Econometrics* 125.1, pp. 15-51. DOI: 10.1016/j.jeconom.2004.04.002.

Bjorklund, A. and R. Moffitt (1987). "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models". In: *Review of Economics and Statistics* 69.1, pp. 42-49. DOI: 10.2307/1937899.

Bonhomme, S. and J. Robin (2009). "Consistent Noisy Independent Component Analysis". In: *Journal of Econometrics* 149.1, pp. 12-25. DOI: 10.1016/j.jeconom.2008.12.019.

Bonhomme, S. and J. Robin (2010). "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics". In: *Review of Economic Studies* 77.2, pp. 491-533. DOI: 10.1111/j.1467-937X.2009.00577.x.

Brinch, C. N, M. Mogstad, and M. Wiswall (2017). "Beyond LATE with a Discrete Instrument". In: *Journal of Political Economy* 125.4, pp. 985-1039. DOI: 10.1086/692712.

Carneiro, P, K. T. Hansen, and J. J. Heckman (2003). "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice". In: *International Economic Review* 44.2, pp. 361-422. DOI: 10.1111/1468-2354.t01-1-00074.

Carneiro, P, J. J. Heckman, and E. Vytlacil (2010). "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin". In: *Econometrica* 78.1, pp. 377-394. DOI: 10.3982/ECTA7089.

Carneiro, P, J. J. Heckman, and E. J. Vytlacil (2011). "Estimating Marginal Returns to Education". In: *American Economic Review* 101.6, pp. 2754-2781. DOI: 10.1257/aer.101.6.2754.

Cunha, F. and J. Heckman (2007). "The Technology of Skill Formation". In: *American Economic Review* 97.2, pp. 31-47. DOI: 10.1257/aer.97.2.31.

Geary, R. C. (1942). "Inherent Relations between Random Variables". In: *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences* 47, pp. 63-76. URL: http://www.jstor.org/stable/20488436.

Heckman, J. J. and J. A. Smith (1993). "Assessing the Case for Randomized Evaluation of Social Programs". In: *Measuring Labor Market Measures: Evaluating the Effects of Active Labour Market Policies*. Ed. by K. Jensen and P. K. Madsen. Copenhagen: Danish Ministry of Labor, pp. 35-96.

Heckman, J. J, J. Smith, and N. Clements (1997). "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts". In: *Review of Economic Studies* 64.4, pp. 487-535. URL: http://www.jstor.org/stable/2971729.

Heckman, J. J. and E. Vytlacil (2005). "Structural Equations, Treatment Effects, and Econometric Policy Evaluation1". In: *Econometrica* 73.3, pp. 669-738. DOI: 10.1111/j.1468-0262.2005.00594.x.

Imbens, G. W. and J. D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica* 62.2, pp. 467-475. DOI: 10.2307/2951620.

Kotlarski, I. (1967). "On Characterizing the Gamma and the Normal Distribution". In: *Pacific Journal of Mathematics* 20, pp. 69-76.

Mardia, K. V. (1970). "Measures of Multivariate Skewness and Kurtosis with Applications". In: *Biometrika* 57.3, pp. 519-530. URL: http://www.jstor.org/stable/2334770.

Reiersol, O. (1950). "Identifiability of a Linear Relation between Variables Which Are Subject to Error". In: *Econometrica* 18.4, pp. 375-389. URL: http://www.jstor.org/stable/1907835.

Robins, J. M. (1997). "Causal Inference from Complex Longitudinal Data". In: *Latent Variable Modeling and Applications to Causality*. Ed. by M. Berkane. New York: Springer, pp. 69-117.

Robinson, P. M. (1988). "Root-N-Consistent Semiparametric Regression". In: *Econometrica* 56.4, pp. 931-954. URL: http://www.jstor.org/stable/1912705.

Rueschendorf, L. (1981). "Sharpness of Frechet-bounds". In: *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57.2, pp. 293-302. DOI: 10.1007/BF00535495.

Vytlacil, E. (2002). "Independence, Monotonicity, and Latent Index Models: An Equivalence Result". In: *Econometrica* 70.1, pp. 331-341. DOI: 10.1111/1468-0262.00277.