Nonparametric Inference on State Dependence among Temporary Workers

Inhyuk Choi

Pennsylvania State University

December 2019

Persistence in Employment

Suppose we observe the following data ($0 =$ unemployed, $1 =$ employed):

	period 0	period 1	period 2	period 3	$\cdots$
agent 1	0	0	0	1	$\cdots$
agent 2	1	1	1	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Serial correlation in $(Y_{i0}, \ldots, Y_{iT}) \implies \exists$ a causal effect of $Y_{i(t-1)}$ on Y_{it}?

- State dependence vs. Persistent latent heterogeneity
- Important implications for the design of labor market programs

How to Distinguish SD from Heterogeneity?

Parametric dynamic binary response models (e.g. Heckman, 1981):

$$Y_{it} = \mathbb{1}\left\{\gamma Y_{i(t-1)} + X_{it}'\beta + \lambda Y_{i0} + A_i + V_{it} \geq 0\right\} \quad \forall t \geq 1,$$

where A_i and V_{it} are unobservable.

- Arbitrary functional form restrictions on the distribution of heterogeneity
- Usually motivated by analytic convenience, rather than economic theory

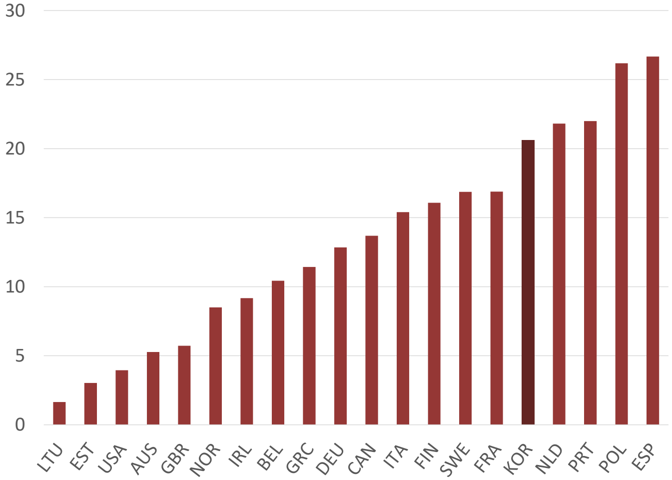
A nonparametric dynamic potential outcomes model (Torgovitsky, 2019):

$$Y_{it} = Y_{i(t-1)} U_{it}(1) + (1 - Y_{i(t-1)}) U_{it}(0),$$

where $U_{it}(0)$ and $U_{it}(1)$ represent the potential outcomes.

Temporary Employment

Figure: Temporary employment, % of salary workers, 2015 (OECD)



The Literature

The papers that examined SD in employment dynamics:

	parametric	nonparametric
binary outcomes	Heckman (1981)	Torgovitsky (2019)
discrete outcomes	Magnac (2000) Prowse (2012)	my paper

Contribution

- The first study that explores whether and to what extent there is SD among temporary workers, based on the nonparametric framework

Outline

1. The Model
2. Identification
3. Parameters of Interest
4. Identifying Assumptions
5. Application
6. Conclusion

The Model

Observable outcomes $Y_{it} \in \mathcal{J} := \{0, 1, \ldots, J\}$

- $Y_i := (Y_{i0}, Y_{i1}, \ldots, Y_{iT}) \in \mathcal{Y}$

Unobservable potential outcomes $(U_{it}(0), U_{i1}(1), \ldots, U_{it}(J)) \in \mathcal{J}^{J+1}$

- $U_i(y) := (U_{i1}(y), \ldots, U_{iT}(y))$
- $U_i := (Y_{i0}, U_i(0), U_i(1), \ldots, U_i(J)) \in \mathcal{U}$

Y_i is related to $(U_i(0), U_i(1), \ldots, U_i(J))$ through

$$Y_{it} = \sum_{y=0}^{J} \mathbb{1}\left\{Y_{i(t-1)} = y\right\} U_{it}(y) = U_{it}(Y_{i(t-1)}) \qquad \forall\ t \geq 1. \tag{1}$$

The Model (cont.)

Observable covariates $X_i := (X_{i0}, X_{i1}, \ldots, X_{iT}) \in \mathcal{X}$ with $|\mathcal{X}| < \infty$

- Observed heterogeneity:

  The dist'n of $(U_i(0), U_i(1), \ldots, U_i(J))|X_i = x$ is different for each $x \in \mathcal{X}$.

- Unobserved heterogeneity:

  The dist'n of $(U_i(0), U_i(1), \ldots, U_i(J))|X_i = x$ need not be degenerate.

Identification

A structure for the model with (1) is a pmf P on $\mathcal{U} \times \mathcal{X}$.

- A function P with domain $\mathcal{U} \times \mathcal{X}$ is a pmf iff P takes values in $[0, 1]$, and

$$\sum_{u \in \mathcal{U}, x \in \mathcal{X}} P(u, x) = 1. \tag{2}$$

- Let $\rho : \mathcal{P} \to \mathbb{R}^{d_\rho}$ be a function representing restrictions on P.
- $\mathcal{P}^\star \subseteq \mathcal{P}^\dagger \subseteq \mathcal{P} \iff$ identified set $\subseteq$ admissible set $\subseteq$ set of all possible P

Identification (cont.)

$P \in \mathcal{P}^\star$ requires that for every $y := (y_0, y_1, \ldots, y_T) \in \mathcal{Y}$ and $x \in \mathcal{X}$,

$$\begin{aligned}
\underbrace{\mathbb{P}[Y = y, X = x]}_{\text{Observable pmf of } (Y, X)} &= \underbrace{\mathbb{P}_P[Y = y, X = x]}_{\substack{\text{Probability of an event when } (U, X) \text{ is distributed} \\ \text{according to } P, \text{ and } Y \text{ is determined through (1)}}} \\
&= \mathbb{P}_P[Y_0 = y_0, U_t(y_{t-1}) = y_t \ \forall \ t \geq 1, X = x] \\
&= \underbrace{\sum_{u \in \mathcal{U}_{\text{oeq}}(y)} P(u, x),}_{\text{Linear in } \{P(u,x) \ | \ u \in \mathcal{U}, x \in \mathcal{X}\}}
\end{aligned} \tag{3}$$

where $\mathcal{U}_{\text{oeq}}(y) := \{u \in \mathcal{U} \mid u_0 = y_0, u_t(y_{t-1}) = y_t \ \forall \ t \geq 1\}$.

Identification (cont.)

Usually interested in a parameter $\theta : \mathcal{P} \to \mathbb{R}$ and its identified set

$$\Theta^\star := \{\theta(P) \mid P \in \mathcal{P}^\star\}.$$

Proposition 1 (Torgovitsky, 2019)

Suppose that $\mathcal{P}^\dagger$ is closed and convex, and that θ is a continuous function of P. Then, as long as $\mathcal{P}^\star$ is nonempty, the identified set $\Theta^\star$ is given by $[\theta_l^\star, \theta_u^\star]$, where

$$\theta_l^\star := \min_{P \in \mathcal{P}^\star} \theta(P) = \min_{\{P(u,x)\in[0,1] \ \mid \ u\in\mathcal{U}, x\in\mathcal{X}\}} \theta(P) \text{ s.t. } \rho(P) \geq 0, (2), \text{ and } (3)\ \forall y, x,$$

$$\theta_u^\star := \max_{P \in \mathcal{P}^\star} \theta(P) = \max_{\{P(u,x)\in[0,1] \ \mid \ u\in\mathcal{U}, x\in\mathcal{X}\}} \theta(P) \text{ s.t. } \rho(P) \geq 0, (2), \text{ and } (3)\ \forall y, x.$$

State Dependence

State dependence can be measured by the proportion of agents with

$$\sum_{j=0}^{J} \mathbb{1}\left\{\sum_{y=0}^{J} \mathbb{1}\left\{U_t(y) = j\right\} = J+1\right\} \neq 1.$$

- $\text{noSD}_t(P) := \mathbb{P}_P[U_t(0) = U_t(1) = \cdots = U_t(J)]$

- $\text{SPSD}_t(P) := \mathbb{P}_P[U_t(0) < U_t(1) < \cdots < U_t(J)]$
- $\text{SNSD}_t(P) := \mathbb{P}_P[U_t(0) > U_t(1) > \cdots > U_t(J)]$

- $\text{PSD}_t(P) := \mathbb{P}_P[U_t(0) \leq U_t(1) \leq \cdots \leq U_t(J)] - \text{noSD}_t(P) - \text{SPSD}_t(P)$
- $\text{NSD}_t(P) := \mathbb{P}_P[U_t(0) \geq U_t(1) \geq \cdots \geq U_t(J)] - \text{noSD}_t(P) - \text{SNSD}_t(P)$

- $\text{MSD}_t(P) := 1 - \text{noSD}_t(P) - \text{SPSD}_t(P) - \text{SNSD}_t(P) - \text{PSD}_t(P) - \text{NSD}_t(P)$

State Dependence (cont.)

Proposition 2

Suppose that $\mathcal{P}^\dagger = \mathcal{P}$. *Then the sharp identified sets for* SPSD_t, SNSD_t, PSD_t, NSD_t, *and* MSD_t *are given by*

$$\left[0, \sum_{y=0}^{J} \mathbb{P}[Y_{t-1} = y, Y_t = y]\right],$$

$$\left[0, \sum_{y=0}^{J} \mathbb{P}[Y_{t-1} = y, Y_t = J - y]\right],$$

$$[0, 1 - \mathbb{P}[Y_{t-1} = 0, Y_t = J] - \mathbb{P}[Y_{t-1} = J, Y_t = 0]],$$

$$[0, 1 - \mathbb{P}[Y_{t-1} = 0, Y_t = 0] - \mathbb{P}[Y_{t-1} = J, Y_t = J]],$$

and $[0, 1]$, *respectively.*

Conditional State Dependence

$\text{SPSD}_t(P)$ can be modified to be conditional on realizations of Y.

- SPSD among those with $Y_t = y$:

$$\text{SPSD}_t(P \mid y) := \mathbb{P}_P[U_t(0) < U_t(1) < \cdots < U_t(J) \mid Y_t = y]$$

- SPSD among those with $Y_t = Y_{t-1} = y$:

$$\begin{aligned}\text{SPSD}_t(P \mid yy) &:= \mathbb{P}_P[U_t(0) < U_t(1) < \cdots < U_t(J) \mid Y_t = y, Y_{t-1} = y] \\ &= \frac{\mathbb{P}_P[U_t(j) = j \text{ for all } j \neq y, Y_t = y \mid Y_{t-1} = y]}{\mathbb{P}[Y_t = y \mid Y_{t-1} = y]} \\ &= \text{Proportion of the observed persistence in } y \\ &\quad\ \text{that is due to SPSD}\end{aligned}$$

Monotone Treatment Response

Assumption MTR *Every $P \in \mathcal{P}^{\dagger}$ satisfies for all $t \geq 1$,*

$$\mathrm{SNSD}_t(P) + \mathrm{NSD}_t(P) = 0,$$

or equivalently,

$$\mathrm{noSD}_t(P) + \mathrm{SPSD}_t(P) + \mathrm{PSD}_t(P) + \mathrm{MSD}_t(P) = 1.$$

Stationarity

Assumption ST ($m = 0$) *For every $P \in \mathcal{P}^\dagger$, the joint distribution of $(U_t(0), U_t(1), \ldots, U_t(J))$ associated with P is invariant across $t \geq 1$.*

Assumption ST ($m = 1$) *For every $P \in \mathcal{P}^\dagger$, the joint distribution of $(U_{t-1}(0), U_t(0), U_{t-1}(1), U_t(1), \ldots, U_{t-1}(J), U_t(J))$ associated with P is invariant across $t \geq 1$.*

Stationarity (cont.)

Assumption ST($m = 0, \sigma$) *Let $\sigma \geq 0$ be a number chosen by the researcher. Define*

$$S_t(u; P) := \mathbb{P}_P[U_t(y) = u(y) \text{ for each } y \in \mathcal{J}]$$

with $u := (u(0), \ldots, u(J))$. Then for every $P \in \mathcal{P}^\dagger$, $u \in \mathcal{J}^{J+1}$, and $t \geq 1$,

$$(1 - \sigma)S_t(u; P) \leq S_{t+1}(u; P) \leq (1 + \sigma)S_t(u; P).$$

Weak Stationarity

Assumption WST *Every* $P \in \mathcal{P}^{\dagger}$ *is such that for all* $y \in \mathcal{J}$, *both* $\mathbb{E}_P[U_t(y)]$ *and* $\mathbb{V}_P[U_t(y)]$ *do not depend on* t.

Diminishing Serial Correlation

Assumption DSC *Every* $P \in \mathcal{P}^\dagger$ *is such that for each* $y \in \mathcal{J}$ *and* $t \geq 1$, $\text{Corr}_P(U_t(y), U_{t+s}(y))$ *is decreasing in* $|s|$ *for* $s \in \{1-t, \ldots, T-t\}$.

If Assumption WST holds, Assumption DSC becomes a linear restriction:

- $\mathbb{E}_P[U_t(y) \cdot U_{t+s}(y)]$ is decreasing in $|s|$ for $s \in \{1-t, \ldots, T-t\}$.

Monotone Instrumental Variables

Assumption MIV *Every* $P \in \mathcal{P}^\dagger$ *is such that for each* $y \in \mathcal{J}$ *and* $t \geq 1$,

(i) $\mathbb{P}_P[U_t(y) = J \mid X = x]$ *is weakly increasing or weakly decreasing in one or more components of* $x \in \mathcal{X}$, *and*

(ii) $\mathbb{P}_P[U_t(y) = 0 \mid X = x]$ *is weakly decreasing or weakly increasing in one or more components of* $x \in \mathcal{X}$.

Monotone Treatment Selection

Assumption MTS *Every* $P \in \mathcal{P}^{\dagger}$ *is such that for all* $y \in \mathcal{J}$, $y_{t-2} \in \mathcal{J}$, *and* $t \geq 2$,

(i) $\mathbb{P}_P[U_t(y) = J \mid Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}]$ *is weakly increasing in* $y_{t-1} \in \mathcal{J}$, *and*

(ii) $\mathbb{P}_P[U_t(y) = 0 \mid Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}]$ *is weakly decreasing in* $y_{t-1} \in \mathcal{J}$.

Data and Computation

Data

- 4,888 obs. from British Household Panel Survey (2005–2009)
- Each worker's employment status is classified into:
  - 0 unemployed
  - 1 temporarily-employed
  - 2 permanently-employed

Computation

- With $J = 2$ and $T = 3$, $\dim(P) = (J+1)^{(J+1)T+1} = 59,049$ (w/o covariates)
- The number of constraints $> (J+1)^{T+1} + 2 \times \dim(P) = 118,179$
- Linear programming solver and symbolic modeling language used (Gurobi and MPL)

Results

Table: Estimated identified sets for the BHPS data

MTR		1	1	1	1		1	1	1	1
WST			1	2	2			1	2	2
ST ($m = 0, \sigma = 0.1$)			1	2	2			1	2	2
ST ($m = 0$)				1	2				1	2
ST ($m = 1$)					1					1
DSC					1					1
MIV					1					1
MTS						1	1	1	1	1
SPSD_t	.000 .952	.000 .948	.000 .943	.000 .939	.000 .939	.000 .218	.000 .216	.000 .196	.000 .195	.000 .195
$\text{SPSD}_t(\cdot \mid 0)$	.000 .347	.000 .351	.000 .351	.000 .351	.000 .351	.000 .240	.000 .240	.000 .240	.000 .240	.000 .240
$\text{SPSD}_t(\cdot \mid 00)$	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 .683	.000 .683	.000 .683	.000 .683	.000 .683
$\text{SPSD}_t(\cdot \mid 1)$	.000 .402	.000 .394	.000 .394	.000 .394	.000 .394	.000 .314	.000 .313	.000 .313	.000 .313	.000 .313
$\text{SPSD}_t(\cdot \mid 11)$	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 .797	.000 .794	.000 .794	.000 .794	.000 .794
$\text{SPSD}_t(\cdot \mid 2)$	.000 .981	.000 .979	.000 .979	.000 .979	.000 .979	.000 .215	.000 .213	.000 .200	.000 .199	.000 .197
$\text{SPSD}_t(\cdot \mid 22)$	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 .220	.000 .218	.000 .205	.000 .203	.000 .201

1 = imposed explicitly, 2 = imposed implicitly

Results (cont.)

Table: 95% confidence regions for the BHPS data

MTR	1	1	1	1	1	1
WST	1	2	2	1	2	2
ST ($m = 0, \sigma = 0.1$)	1	2	2	1	2	2
ST ($m = 0$)		1	2		1	2
ST ($m = 1$)			1			1
MTS				1	1	1
SPSD_t	.000 .956	.000 .954	.000 .954	.000 .405	.000 .406	.000 .420
$\text{SPSD}_t(\cdot \mid 0)$	.000 .448	.000 .448	.000 .468	.000 .448	.000 .448	.000 .460
$\text{SPSD}_t(\cdot \mid 00)$	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 1.00
$\text{SPSD}_t(\cdot \mid 1)$	.000 .521	.000 .521	.000 .552	.000 .521	.000 .521	.000 .554
$\text{SPSD}_t(\cdot \mid 11)$	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 1.00	.000 1.00
$\text{SPSD}_t(\cdot \mid 2)$	.000 .999	.000 .996	.000 .997	.000 .416	.000 .416	.000 .424
$\text{SPSD}_t(\cdot \mid 22)$	.000 1.00	.000 1.00	.000 1.00	.000 .426	.000 .425	.000 .433

1 = imposed explicitly, 2 = imposed implicitly

Results (cont.)

Table: 95% confidence regions for different subsample sizes ($b_1 = n^{2/3}$, $b_2 = n^{3/4}$, $b_3 = n^{4/5}$)

MTR	1	1	1	1
WST	1	1	1	1
ST ($m = 0, \sigma = 0.1$)	1	1	1	1
MTS	1	1	1	1
b	n	b_1	b_2	b_3
SPSD_t	.000 .196	.000 .405	.000 .419	.000 .437
$\text{SPSD}_t(\cdot \mid 0)$	.000 .240	.000 .448	.000 .457	.000 .480
$\text{SPSD}_t(\cdot \mid 00)$	.000 .683	.000 1.00	.000 .700	.000 .702
$\text{SPSD}_t(\cdot \mid 1)$	.000 .313	.000 .521	.000 .532	.000 .561
$\text{SPSD}_t(\cdot \mid 11)$	.000 .794	.000 1.00	.000 .996	.000 .983
$\text{SPSD}_t(\cdot \mid 2)$	.000 .200	.000 .416	.000 .428	.000 .445
$\text{SPSD}_t(\cdot \mid 22)$	.000 .205	.000 .426	.000 .436	.000 .454

1 = imposed explicitly

Conclusion

Summary

- Extended the DPO model to allow for multiple outcomes
- Found little evidence of SD among temp workers in Britain
- Obtained excessively wide confidence regions

Future research

- Measure SD among temp workers in other countries
- Develop or apply a new inferential approach
- Build a structural model to describe the mechanism

Dynamic Binary Response Models

$$Y_{it} = \mathbb{1}\left\{\gamma Y_{i(t-1)} + X_{it}'\beta + \lambda Y_{i0} + A_i + V_{it} \geq 0\right\} \text{ for all } t \geq 1$$

A1 $V_i \equiv (V_{i1}, \ldots, V_{iT}) \sim N(0, I_T)$, *where* I_T *is the* T*-dim identity matrix.*

A2 V_i *is independent of* (Y_{i0}, X_i, A_i), *where* $X_i \equiv (X_{i0}, X_{i1}, \ldots, X_{iT})$.

A3 $A_i \sim N(0, \sigma_A^2)$ *for some unknown* σ_A^2.

A4 A_i *is independent of* (X_i, Y_{i0}).

The MLE of $(\gamma, \beta, \lambda, \sigma_A^2)$ consistent and asymptotically normal if the above is valid

- Enabling the construction of a consistent estimator of the ATE at time t :

  $$\text{ATE}_t \equiv \mathbb{E}\left[\mathbb{1}\left\{\gamma + X_{it}'\beta + \lambda Y_{i0} + A_i + V_{it} \geq 0\right\} - \mathbb{1}\left\{X_{it}'\beta + \lambda Y_{i0} + A_i + V_{it} \geq 0\right\}\right]$$

ATE and SD

$$\begin{aligned}
\mathrm{ATE}_t(P) :&= \mathbb{E}_P[U_t(1) - U_t(0)] \\
&= (\mathbb{P}_P[U_t(0) = 0, U_t(1) = 1] + \mathbb{P}_P[U_t(0) = 1, U_t(1) = 1]) \\
&\quad - (\mathbb{P}_P[U_t(0) = 1, U_t(1) = 0] + \mathbb{P}_P[U_t(0) = 1, U_t(1) = 1]) \\
&= \mathrm{SPSD}_t(P) - \mathrm{SNSD}_t(P)
\end{aligned}$$

Assumption ST and Lower Bounds on SPSD_t

Assumption ST and $\text{noSD}_t(P) = 1 \implies$ stationary distribution of Y

$$\begin{aligned}
\mathbb{P}_P[Y_t = 0] &= \mathbb{P}_P[U_t(0) = U_t(1), Y_t = 0] + \mathbb{P}_P[U_t(0) \neq U_t(1), Y_t = 0] \\
&= \mathbb{P}_P[U_t(0) = 0, U_t(1) = 0] + \mathbb{P}_P[U_t(0) \neq U_t(1), Y_t = 0] \\
&= \mathbb{P}_P[U_{t-1}(0) = 0, U_{t-1}(1) = 0] + \mathbb{P}_P[U_{t-1}(0) \neq U_{t-1}(1), Y_{t-1} = 0] \\
&= \mathbb{P}_P[Y_{t-1} = 0]
\end{aligned}$$

- If Assumption ST holds, non-stationary Y implies $\text{noSD}_t(P) \neq 1$.
- If Assumptions MTR and ST hold, non-stationary Y implies

$$\begin{aligned}
\text{noSD}_t(P) &= 1 - \text{SPSD}_t(P) - \text{SNSD}_t(P) - \text{PSD}_t(P) - \text{NSD}_t(P) - \text{MSD}_t(P) \\
&= 1 - [\text{SPSD}_t(P) + \text{PSD}_t(P) + \text{MSD}_t(P)] \\
&\neq 1.
\end{aligned}$$

Fixed Effects

Assumption FE *Let* $U_t := (U_t(0), U_t(1), \ldots, U_t(J))$. *For every* $P \in \mathcal{P}^\dagger$, *there exists a random variable* A *such that*

$$\mathbb{P}_P[U_t = u \mid Y_{t-1}, \ldots, Y_1, Y_0, A] = \mathbb{P}_P[U_1 = u \mid Y_0, A] \quad \textit{(almost surely)}$$

for all $u \in \mathcal{J}^{J+1}$ *and all* $t \geq 2$.

Proposition 3 (Torgovitsky, 2019)

Let $t > s \geq 1$, *and define* $Y^{0,s} := (Y_0, Y_1, \ldots, Y_s)$. *If Assumption FE holds, then for any* $P \in \mathcal{P}^\dagger$, *every* $u \in \mathcal{J}^{J+1}$ *and every* $y \in \mathcal{J}^s$,

$$\mathbb{P}_P[U_t = u, Y^{0,s-1} = y] = \mathbb{P}_P[U_s = u, Y^{0,s-1} = y].$$

Data Description

Table: Descriptive statistics on (un)employment dynamics in the BHPS

	Period t				
	0	1	2	3	
$\mathbb{P}[Y_t = 0]$	.029	.020	.019	.030	
$\mathbb{P}[Y_t = 1]$	.030	.026	.022	.020	
$\mathbb{P}[Y_t = 2]$	.941	.954	.959	.949	
$\mathbb{P}[Y_t \neq Y_{t-1}]$	–	.061	.046	.052	
$\mathbb{P}[Y_t = 0 \mid Y_{t-1} = 0]$	–	.340	.510	.547	
$\mathbb{P}[Y_t = 1 \mid Y_{t-1} = 1]$	–	.356	.357	.364	
$\mathbb{P}[Y_t = 2 \mid Y_{t-1} = 2]$	–	.976	.980	.969	
	Proportion of individuals with $\cdots$				
	0	1	2	3	4
periods of $Y_t = 0$	.937	.045	.007	.004	.007
periods of $Y_t = 1$	.935	.043	.013	.005	.003
periods of $Y_t = 2$	.013	.010	.019	.078	.880
spells of $Y_t = 0$	.937	.049	.015	–	–
spells of $Y_t = 1$	.935	.052	.012	–	–
spells of $Y_t = 2$	.013	.023	.965	–	–
transitions	.891	.065	.038	.006	–

Statistical Inference

An inferential approach based on direct sample analogs of $\theta_l^\star$ and $\theta_u^\star$ untenable

- Would be consistent, with their asymptotic dist'n highly nonstandard

Strategy (Chernozhukov-Hong-Tamer, 2007)[1]

1. Transform the characterization of $\Theta^\star$ in Prop. 1 into a criterion function
2. Use an appropriate sample analog of this criterion function as the basis for statistical inference

[1] The following discussion is largely taken from Torgovitsky (2019).

The Criterion Function

$\mathcal{W} := \text{supp}(Y, X)$, the joint support of the observable data $W := (Y, X)$

For each $w := (w_y, w_x) \in \mathcal{W} \subset \mathbb{R}^{d_W}$, define

$$m_{\text{oeq},w}(W, P) := \mathbb{1}[Y = w_y, X = w_x] - \sum_{u \in \mathcal{U}_{\text{oeq}}(w_y)} P(u, w_x).$$

The restriction function ρ partitioned into two components

- $\rho_s : \mathcal{P} \to \mathbb{R}^{d_s}$ (stochastic component)
  - Assumed that $\exists m_\rho : \mathcal{W} \times \mathcal{P} \to \mathbb{R}^{d_s}$ for which $\rho_s(P) = \mathbb{E}[m_\rho(W, P)]$
  - $m_{\rho,s}(W, P)$: the s^{th} component of $m_\rho(W, P)$
- $\rho_d : \mathcal{P} \to \mathbb{R}^{d_\rho - d_s}$ (deterministic component)
  - Not depending on the distribution of W

The Criterion Function (cont.)

The DPO model can be viewed as a moment inequality model:

$$\begin{aligned}\mathcal{P}^\star = \{P \in \mathcal{P}_d^\dagger \mid \mathbb{E}[m_{\text{oeq},w}(W,P)] = 0 \ \forall w \in \mathcal{W}, \\ \mathbb{E}[m_{\rho,s}(W,P)] \geq 0 \ \forall s = 1,\dots,d_s\},\end{aligned}$$

where $\mathcal{P}_d^\dagger := \{P \in \mathcal{P} \mid \rho_d(P) \geq 0\}$.

Letting $\lambda \in \mathbb{R}_+^{d_s}$ denote a vector of positive slackness variables,

$$\begin{aligned}\mathcal{R}^\star := \{(P,\lambda) \in \mathcal{P}_d^\dagger \times \mathbb{R}_+^{d_s} \mid \mathbb{E}[m_{\text{oeq},w}(W,P)] = 0 \ \forall w \in \mathcal{W}, \\ \mathbb{E}[m_{\rho,s}(W,P)] - \lambda_s = 0 \ \forall s = 1,\dots,d_s\}\end{aligned}$$

so that $\mathcal{P}^\star$ is the projection of the first component of $\mathcal{R}^\star$:

$$\mathcal{P}^\star = \{P \in \mathcal{P} \mid (P,\lambda) \in \mathcal{R}^\star \text{ for some } \lambda \in \mathbb{R}_+^{d_s}\}.$$

The Criterion Function (cont.)

Write $\{m_{\rho,s}\}_{s=1}^{d_s}$ and $\{m_{\text{oeq},w}\}_{w\in\mathcal{W}}$ together as $\{m_j\}_{j=1}^{d_m}$ with $d_m = d_s + d_W$.

- The first d_s components of $\{m_j\}_{j=1}^{d_m}$ correspond to $\{m_{\rho,s}\}_{s=1}^{d_s}$.

A natural choice of population criterion function is

$$Q(P,\lambda) := \sum_{j=1}^{d_s} (\mathbb{E}m_j(W,P) - \lambda_j)^2 + \sum_{j=d_s+1}^{d_m} (\mathbb{E}m_j(W,P))^2 ,$$

implying $(P,\lambda) \in \mathcal{R}^\star$ iff $Q(P,\lambda) = 0$ and $(P,\lambda) \in \mathcal{P}_d^\dagger \times \mathbb{R}_+^{d_s}$.

Given an i.i.d. sample $\{W_i\}_{i=1}^n$, a sample analog of $Q(\cdot,\cdot)$ would be

$$Q_n(P,\lambda) := \sum_{j=1}^{d_s} n\left(\overline{m}_{n,j}(P) - \lambda_j\right)^2 + \sum_{j=d_s+1}^{d_m} n\overline{m}_{n,j}(P)^2,$$

where $\overline{m}_{n,j}(P) := n^{-1}\sum_{i=1}^n m_j(W_i,P)$ for $j = 1,\ldots,d_m$.

The Criterion Function (cont.)

To define a sample criterion function for a given parameter $\theta(\cdot)$, profile Q_n :

$$\overline{Q}_n(t) := \inf_{(P,\lambda)\in\mathcal{P}_d^\dagger(t)\times\mathbb{R}_+^{d_s}} Q_n(P,\lambda),$$

where $\mathcal{P}_d^\dagger(t) := \{P \in \mathcal{P}_d^\dagger \mid \theta(P) = t\}$.

$Q_n(t)$ serves as a test statistic for a test of $H_0 : t \in \Theta^\star$.

- Confidence regions for $\Theta^\star$ constructed by collecting all $t \in \Theta$ for which H_0 is not rejected

Critical Values

How to approximate the distribution of $\overline{Q}_n(t)$ under the null hypothesis?

1. Subsampling
   - The distribution of $\overline{Q}_n(t)$ under H_0 approximated by that of
   $$\overline{Q}_b^{SS}(t) := \inf_{(P,\lambda)\in\mathcal{P}_d^{\dagger}(t)\times\mathbb{R}_+^{d_s}} Q_b^{SS}(P,\lambda),$$
   where $\overline{Q}_b^{SS}(P,\lambda)$ defined analogously to $Q_n(P,\lambda)$, but constructed using a subsample $\{W_i^{\star}\}_{i=1}^{b}$ randomly drawn from $\{W_i\}_{i=1}^{n}$ without replacement.
   - *The SS test* rejects $H_0 : t \in \Theta^{\star}$ when $\overline{Q}_n(t)$ is larger than the $1-\alpha$ quantile of $\overline{Q}_b^{SS}(t)$ based on B random subsamples.
   - A $1-\alpha$ *SS confidence region for* $\Theta^{\star}$ is the set of all t for which the SS test does not reject.
2. The shape restriction approach of Chernozhukov et al. (2015)
   - Based on a careful approximation of $\overline{Q}_n(t)$ considering the shape of the constraint set $\mathcal{P}_d^{\dagger}(t)\times\mathbb{R}_+^{d_s}$ (computationally infeasible here)

Testing for Misspecification

A rejection of the null hypothesis $H_0 : \mathcal{P}^\star \neq \emptyset$

$\rightarrow$ $\nexists$ an admissible $P \in \mathcal{P}^\dagger$ consistent with the observed data.

$\rightarrow$ Some of the assumptions embodied in $\mathcal{P}^\dagger$ are false.

$\rightarrow$ The model is misspecified.

A natural statistic for such a test is

$$\overline{Q}_n := \inf_{(P,\lambda)\in\mathcal{P}_d^\dagger\times\mathbb{R}_+^{d_s}} Q_n(P,\lambda),$$

whose distribution can be approximated as before.

A level α misspecification test rejects $H_0 : \mathcal{P}^\star \neq \emptyset$ when $\overline{Q}_n$ is larger than the $1-\alpha$ quantile of the simulated distribution.

- Such a test always fails to reject when the estimated identified set is non-empty since $\overline{Q}_n = 0$ in such cases.

The CNS Test

$$Q_n^\star(P,\lambda,g,h) := \sum_{j=1}^{d_s}\left(\nu_{n,j}^\star(P) + n^{-1}\sum_{i=1}^n \nabla m_j(W_i,P)[g] - h_j\right)^2 + \sum_{j=d_s+1}^{d_m}\left(\nu_{n,j}^\star(P) + n^{-1}\sum_{i=1}^n \nabla m_j(W_i,P)[g]\right)^2$$

- (g,h) are parameters that serve as local deviations to (P,λ).
- $\nabla m_j(W_i,P)[g] := \frac{\partial}{\partial\kappa} m_j(W_i,P+\kappa g)|_{\kappa=0}$
- For each $j = 1,\ldots,d_m$,

$$\nu_{n,j}^\star(P) := \frac{1}{\sqrt{n}}\sum_{i=1}^n [m_j(W_i^\star,P) - \overline{m}_{n,j}(P)],$$

where $\{W_i^\star\}_{i=1}^n$ is a bootstrap sample drawn i.i.d. with replacement from $\{W_i\}_{i=1}^n$.

The CNS Test (cont.)

The distribution of $\overline{Q}_n(t)$ approximated by that of

$$\widetilde{Q}_n(t) := \min_{(P,\lambda,g,h)} Q_n^\star(P,\lambda,g,h)$$
$$\text{s.t. } (P,\lambda) \in \widehat{\mathcal{R}}^\star(t) \text{ and } (P,\lambda) + n^{-1/2}(g,h) \in \mathcal{P}_d^\dagger(t) \times \mathbb{R}_+^{d_s},$$

where $\widehat{\mathcal{R}}^\star(t) := \left\{(P,\lambda) \in \mathcal{P}_d^\dagger(t) \times \mathbb{R}_+^{d_s} \mid Q_n(P,\lambda) \leq (1+\tau)\overline{Q}_n(t)\right\}$, with $\tau > 0$ given

- The distribution of $\widetilde{Q}_n(t)$ approximated by redrawing $\{W_i^\star\}_{i=1}^n$ a large number (B) of times and computing $\widetilde{Q}_n(t)$ for each draw
- *The CNS test* rejects $H_0 : t \in \Theta^\star$ when $\overline{Q}_n(t)$ is larger than the $1-\alpha$ quantile of these B values of $\widetilde{Q}_n(t)$.
- A $1-\alpha$ *CNS confidence region for* $\Theta^\star$ is the set of all t for which the CNS test does not reject.