Nonparametric Inference on State Dependence among Temporary Workers

Inhyuk Choi

Pennsylvania State University

December 2019
Suppose we observe the following data (0 = unemployed, 1 = employed):

<table>
<thead>
<tr>
<th></th>
<th>period 0</th>
<th>period 1</th>
<th>period 2</th>
<th>period 3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>agent 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>agent 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

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):

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

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:

<table>
<thead>
<tr>
<th></th>
<th>Parametric</th>
<th>Nonparametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete outcomes</td>
<td>Magnac (2000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prowse (2012)</td>
<td>my paper</td>
</tr>
</tbody>
</table>

**Contribution**

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

**Observable outcomes** \( Y_{it} \in J := \{0, 1, \ldots, J\} \)
- \( Y_i := (Y_{i0}, Y_{i1}, \ldots, Y_{iT}) \in \mathcal{Y} \)

**Unobservable potential outcomes** \((U_{it}(0), U_{i1}(1), \ldots, U_{it}(J)) \in J^{J+1}\)
- \( U_i(y) := (U_{i1}(y), \ldots, U_{iT}(y)) \)
- \( U_i := (Y_{i0}, U_i(0), U_i(1), \ldots, U_i(J)) \in \mathcal{U} \)

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

\[
Y_{it} = \sum_{y=0}^{J} \mathbb{1} \{ Y_{i(t-1)} = y \} \ U_{it}(y) = U_{it}(Y_{i(t-1)}) \quad \forall \ t \geq 1. \quad (1)
\]
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.
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 $P$.
- $\mathcal{P}^* \subseteq \mathcal{P}^\dagger \subseteq \mathcal{P} \iff$ identified set $\subseteq$ admissible set $\subseteq$ set of all possible $P$. 
Identification (cont.)

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

\[
\mathbb{P}[Y = y, X = x] = \mathbb{P}_P[Y = y, X = x]
\]

Observable pmf of \((Y, X)\) \quad Probability of an event when \((U, X)\) is distributed according to \(P\), and \(Y\) is determined through (1)

\[
= \mathbb{P}_P[Y_0 = y_0, U_t(y_{t-1}) = y_t \ \forall \ t \geq 1, X = x]
\]

\[
= \sum_{u \in \mathcal{U}_{oeq}(y)} P(u, x),
\]

Linear in \(\{P(u, x) | u \in \mathcal{U}, x \in \mathcal{X}\}\) \hfill (3)

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

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

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

Proposition 1 (Torgovitsky, 2019)

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

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

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

State dependence can be measured by the proportion of agents with

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

- $\text{noSD}_t(P) := \mathbb{P}_P[U_t(0) = U_t(1) = \cdots = U_t(J)]$
- $\text{SPSD}_t(P) := \mathbb{P}_P[U_t(0) < U_t(1) < \cdots < U_t(J)]$
- $\text{SNSD}_t(P) := \mathbb{P}_P[U_t(0) > U_t(1) > \cdots > U_t(J)]$
- $\text{PSD}_t(P) := \mathbb{P}_P[U_t(0) \leq U_t(1) \leq \cdots \leq U_t(J)] - \text{noSD}_t(P) - \text{SPSD}_t(P)$
- $\text{NSD}_t(P) := \mathbb{P}_P[U_t(0) \geq U_t(1) \geq \cdots \geq U_t(J)] - \text{noSD}_t(P) - \text{SNSD}_t(P)$
- $\text{MSD}_t(P) := 1 - \text{noSD}_t(P) - \text{SPSD}_t(P) - \text{SNSD}_t(P) - \text{PSD}_t(P) - \text{NSD}_t(P)$
Proposition 2

Suppose that $P^\dagger = P$. Then the sharp identified sets for $SPS_{t}$, $SNS_{t}$, $PSD_{t}$, $NSD_{t}$, and $MSD_{t}$ are given by

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

and $[0, 1]$, respectively.
SPSD\(t(P)\) can be modified to be conditional on realizations of \(Y\).

- SPSD among those with \(Y_t = y\):

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

- SPSD among those with \(Y_t = Y_{t-1} = y\):

\[
\text{SPSD}_t(P | yy) := \mathbb{P}_P[U_t(0) < U_t(1) < \cdots < U_t(J) | Y_t = y, Y_{t-1} = y]
\]

\[
= \frac{\mathbb{P}_P[U_t(j) = j \text{ for all } j \neq y, Y_t = y | Y_{t-1} = y]}{\mathbb{P}[Y_t = y | Y_{t-1} = y]}
\]

\[
= \text{Proportion of the observed persistence in } y \\
\text{that is due to SPSD}
\]
Monotone Treatment Response

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

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

or equivalently,

$$\text{noSD}_t(P) + \text{SPSD}_t(P) + \text{PSD}_t(P) + \text{MSD}_t(P) = 1.$$
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, σ)** 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 $\nabla_P[U_t(y)]$ do not depend on $t$. 
Assumption DSC  Every $P \in \mathcal{P}^\dagger$ is such that for each $y \in J$ 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

- 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 $\geq (J + 1)^{T+1} + 2 \times \dim(P) = 118,179$
- Linear programming solver and symbolic modeling language used (Gurobi and MPL)
## Results

**Table: Estimated identified sets for the BHPS data**

<table>
<thead>
<tr>
<th>Parameters of Interest</th>
<th>Identification</th>
<th>Parameters of Interest</th>
<th>Identification</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MTR</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>WST</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>ST (m = 0, σ = 0.1)</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>ST (m = 0)</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>ST (m = 1)</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>DSC</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>MIV</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>MTS</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPSD_t</th>
<th>.000</th>
<th>.000</th>
<th>.000</th>
<th>.000</th>
<th>.000</th>
<th>.000</th>
<th>.000</th>
<th>.000</th>
<th>.000</th>
<th>.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPSD_t(·</td>
<td>0)</td>
<td>.952</td>
<td>.948</td>
<td>.943</td>
<td>.939</td>
<td>.939</td>
<td>.218</td>
<td>.216</td>
<td>.196</td>
<td>.195</td>
</tr>
<tr>
<td>SPSD_t(·</td>
<td>00)</td>
<td>.347</td>
<td>.351</td>
<td>.351</td>
<td>.351</td>
<td>.351</td>
<td>.240</td>
<td>.240</td>
<td>.240</td>
<td>.240</td>
</tr>
<tr>
<td>SPSD_t(·</td>
<td>1)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.683</td>
<td>.683</td>
<td>.683</td>
<td>.683</td>
</tr>
<tr>
<td>SPSD_t(·</td>
<td>11)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.797</td>
<td>.794</td>
<td>.794</td>
<td>.794</td>
</tr>
<tr>
<td>SPSD_t(·</td>
<td>2)</td>
<td>.981</td>
<td>.979</td>
<td>.979</td>
<td>.979</td>
<td>.979</td>
<td>.215</td>
<td>.213</td>
<td>.200</td>
<td>.199</td>
</tr>
<tr>
<td>SPSD_t(·</td>
<td>22)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.220</td>
<td>.218</td>
<td>.205</td>
<td>.203</td>
</tr>
</tbody>
</table>

1 = imposed explicitly, 2 = imposed implicitly
### Table: 95% confidence regions for the BHPS data

<table>
<thead>
<tr>
<th></th>
<th>MTR</th>
<th>WST</th>
<th>ST ( (m = 0, \sigma = 0.1) )</th>
<th>ST ( (m = 0) )</th>
<th>ST ( (m = 1) )</th>
<th>MTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

|        | SPSD\( t \) | SPSD\( t(\cdot | 0) \) | SPSD\( t(\cdot | 00) \) | SPSD\( t(\cdot | 1) \) | SPSD\( t(\cdot | 11) \) | SPSD\( t(\cdot | 2) \) | SPSD\( t(\cdot | 22) \) |
|--------|-------------|-------------------|-----------------|-------------|----------------|-------------|----------------|
|        | .000        | .000              | .000            | .000        | .000           | .000        | .000           |
|        | .956        | .954              | .954            | .405        | .406           | .420        | .448           |
|        | .000        | .000              | .000            | .000        | .000           | .000        | .000           |
|        | .448        | .448              | .468            | .448        | .448           | .460        | .448           |
|        | 1.00        | 1.00              | 1.00            | 1.00        | 1.00           | 1.00        | 1.00           |
|        | .000        | .000              | .000            | .000        | .000           | .000        | .000           |
|        | .521        | .521              | .552            | .521        | .521           | .554        | .521           |
|        | 1.00        | 1.00              | 1.00            | 1.00        | 1.00           | 1.00        | 1.00           |
|        | .000        | .000              | .000            | .000        | .000           | .000        | .000           |
|        | .999        | .996              | .997            | .416        | .416           | .424        | .425           |
|        | 1.00        | 1.00              | 1.00            | .426        | .425           | .433        | .426           |

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

<table>
<thead>
<tr>
<th>Model</th>
<th>(b_1)</th>
<th>(b_2)</th>
<th>(b_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTR</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WST</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ST ((m = 0, \sigma = 0.1))</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MTS</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(b)</td>
<td>(n)</td>
<td>(b_1)</td>
<td>(b_2)</td>
</tr>
<tr>
<td>(\text{SPSD}_t)</td>
<td>.000</td>
<td>.196</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>.405</td>
<td>.419</td>
</tr>
<tr>
<td>(\text{SPSD}_t(\cdot</td>
<td>0))</td>
<td>.000</td>
<td>.240</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>.448</td>
<td>.457</td>
</tr>
<tr>
<td>(\text{SPSD}_t(\cdot</td>
<td>00))</td>
<td>.000</td>
<td>.683</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>1.00</td>
<td>.700</td>
</tr>
<tr>
<td>(\text{SPSD}_t(\cdot</td>
<td>1))</td>
<td>.000</td>
<td>.313</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>.521</td>
<td>.532</td>
</tr>
<tr>
<td>(\text{SPSD}_t(\cdot</td>
<td>11))</td>
<td>.000</td>
<td>.794</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>1.00</td>
<td>.996</td>
</tr>
<tr>
<td>(\text{SPSD}_t(\cdot</td>
<td>2))</td>
<td>.000</td>
<td>.200</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>.416</td>
<td>.428</td>
</tr>
<tr>
<td>(\text{SPSD}_t(\cdot</td>
<td>22))</td>
<td>.000</td>
<td>.205</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>.426</td>
<td>.436</td>
</tr>
</tbody>
</table>

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

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

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 \( \sigma_A^2 \).

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 \( t \):

\[
\text{ATE}_t \equiv \mathbb{E} \left[ \mathbb{I} \left\{ \gamma + X'_{it} \beta + \lambda Y_{i0} + A_i + V_{it} \geq 0 \right\} - \mathbb{I} \left\{ X'_{it} \beta + \lambda Y_{i0} + A_i + V_{it} \geq 0 \right\} \right]
\]
ATE and SD

\[ \text{ATE}_t(P) := \mathbb{E}_P[U_t(1) - U_t(0)] \]

\[ = (\mathbb{P}_P[U_t(0) = 0, U_t(1) = 1] + \mathbb{P}_P[U_t(0) = 1, U_t(1) = 1]) \]

\[ - (\mathbb{P}_P[U_t(0) = 1, U_t(1) = 0] + \mathbb{P}_P[U_t(0) = 1, U_t(1) = 1]) \]

\[ = \text{SPSD}_t(P) - \text{SNSD}_t(P) \]
Assumption ST and Lower Bounds on $\text{SPSD}_t$

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

$$
P_P[Y_{t} = 0] = P_P[U_{t}(0) = U_{t}(1), Y_{t} = 0] + P_P[U_{t}(0) \neq U_{t}(1), Y_{t} = 0]$$

$$
= P_P[U_{t}(0) = 0, U_{t}(1) = 0] + P_P[U_{t}(0) \neq U_{t}(1), Y_{t} = 0]$$

$$
= P_P[U_{t-1}(0) = 0, U_{t-1}(1) = 0] + P_P[U_{t-1}(0) \neq U_{t-1}(1), Y_{t-1} = 0]$$

$$
= P_P[Y_{t-1} = 0]
$$

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

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

$$
= 1 - [\text{SPSD}_t(P) + \text{PSD}_t(P) + \text{MSD}_t(P)]
$$

$$
\neq 1.
$$
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

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

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

<table>
<thead>
<tr>
<th></th>
<th>Period $t$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$P[Y_t = 0]$</td>
<td>.029</td>
<td>.020</td>
<td>.019</td>
<td>.030</td>
</tr>
<tr>
<td>$P[Y_t = 1]$</td>
<td>.030</td>
<td>.026</td>
<td>.022</td>
<td>.020</td>
</tr>
<tr>
<td>$P[Y_t = 2]$</td>
<td>.941</td>
<td>.954</td>
<td>.959</td>
<td>.949</td>
</tr>
<tr>
<td>$P[Y_t \neq Y_{t-1}]$</td>
<td>–</td>
<td>.061</td>
<td>.046</td>
<td>.052</td>
</tr>
<tr>
<td>$P[Y_t = 0 \mid Y_{t-1} = 0]$</td>
<td>–</td>
<td>.340</td>
<td>.510</td>
<td>.547</td>
</tr>
<tr>
<td>$P[Y_t = 1 \mid Y_{t-1} = 1]$</td>
<td>–</td>
<td>.356</td>
<td>.357</td>
<td>.364</td>
</tr>
<tr>
<td>$P[Y_t = 2 \mid Y_{t-1} = 2]$</td>
<td>–</td>
<td>.976</td>
<td>.980</td>
<td>.969</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Proportion of individuals with ···</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>periods of $Y_t = 0$</td>
<td>.937</td>
<td>.045</td>
<td>.007</td>
<td>.004</td>
<td>.007</td>
</tr>
<tr>
<td>periods of $Y_t = 1$</td>
<td>.935</td>
<td>.043</td>
<td>.013</td>
<td>.005</td>
<td>.003</td>
</tr>
<tr>
<td>periods of $Y_t = 2$</td>
<td>.013</td>
<td>.010</td>
<td>.019</td>
<td>.078</td>
<td>.880</td>
</tr>
<tr>
<td>spells of $Y_t = 0$</td>
<td>.937</td>
<td>.049</td>
<td>.015</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>spells of $Y_t = 1$</td>
<td>.935</td>
<td>.052</td>
<td>.012</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>spells of $Y_t = 2$</td>
<td>.013</td>
<td>.023</td>
<td>.965</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>transitions</td>
<td>.891</td>
<td>.065</td>
<td>.038</td>
<td>.006</td>
<td>–</td>
</tr>
</tbody>
</table>
An inferential approach based on direct sample analogs of $\theta_i^*$ and $\theta_u^*$ untenable

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

**Strategy** (Chernozhukov-Hong-Tamer, 2007)$^1$

1. Transform the characterization of $\Theta^*$ 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

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

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

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

The restriction function \( \rho \) partitioned into two components

- \( \rho_s : \mathcal{P} \rightarrow \mathbb{R}^{d_s} \) (stochastic component)
  - Assumed that \( \exists m_\rho : \mathcal{W} \times \mathcal{P} \rightarrow \mathbb{R}^{d_s} \) for which \( \rho_s(P) = \mathbb{E}[m_\rho(\mathcal{W}, P)] \)
  - \( m_{\rho,s}(\mathcal{W}, P) : \text{the } s^{\text{th}} \text{ component of } m_\rho(\mathcal{W}, P) \)

- \( \rho_d : \mathcal{P} \rightarrow \mathbb{R}^{d_\rho - d_s} \) (deterministic component)
  - Not depending on the distribution of \( \mathcal{W} \)
The Criterion Function (cont.)

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

\[ \mathcal{P}^* = \{ P \in \mathcal{P}^\dagger_d | \mathbb{E}[m_{\text{eq},w}(W, P)] = 0 \ \forall w \in \mathcal{W}, \]
\[ \mathbb{E}[m_{\rho,s}(W, P)] \geq 0 \ \forall s = 1, \ldots, d_s \}, \]

where \( \mathcal{P}^\dagger_d := \{ P \in \mathcal{P} | \rho_d(P) \geq 0 \}. \)

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

\[ \mathcal{R}^* := \{(P, \lambda) \in \mathcal{P}^\dagger_d \times \mathbb{R}_+^{d_s} | \mathbb{E}[m_{\text{eq},w}(W, P)] = 0 \ \forall w \in \mathcal{W}, \]
\[ \mathbb{E}[m_{\rho,s}(W, P)] - \lambda_s = 0 \ \forall s = 1, \ldots, d_s \} \]

so that \( \mathcal{P}^* \) is the projection of the first component of \( \mathcal{R}^* : \)

\[ \mathcal{P}^* = \{ P \in \mathcal{P} | (P, \lambda) \in \mathcal{R}^* \text{ for some } \lambda \in \mathbb{R}_+^{d_s} \}. \]
Write \( \{m_{\rho,s}\}_{s=1}^{d_s} \) and \( \{m_{\text{eq},w}\}_{w \in \mathcal{W}} \) together as \( \{m_j\}_{j=1}^{d_m} \) with \( d_m = d_s + d_{\mathcal{W}} \).

- 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 \((P, \lambda) \in \mathcal{R}^* \text{ iff } Q(P, \lambda) = 0\) and \((P, \lambda) \in \mathcal{P}_d^+ \times \mathbb{R}^{d_s}_+\).

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

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

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 \( Q_n : \)

\[
\overline{Q}_n(t) := \inf_{(P, \lambda) \in P_d^\dagger(t) \times \mathbb{R}_+^{ds}} Q_n(P, \lambda),
\]

where \( P_d^\dagger(t) := \{ P \in P_d^\dagger | \theta(P) = t \} \).

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

- Confidence regions for \( \Theta^* \) 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
     \[ Q_{SS}^S(t) := \inf_{(P, \lambda) \in \mathcal{P}_d^+(t) \times \mathbb{R}^{d_s}_{+}} Q_{SS}^S(P, \lambda), \]
     where $Q_{SS}^S(P, \lambda)$ defined analogously to $Q_n(P, \lambda)$, but constructed using a subsample $\{W_i^*\}_{i=1}^b$ randomly drawn from $\{W_i\}_{i=1}^n$ without replacement.

   - The SS test rejects $H_0 : t \in \Theta^*$ when $\overline{Q}_n(t)$ is larger than the $1 - \alpha$ quantile of $Q_{SS}^S(t)$ based on $B$ random subsamples.

   - A $1 - \alpha$ SS confidence region for $\Theta^*$ 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 $\mathcal{P}_d^+(t) \times \mathbb{R}^{d_s}_{+}$ (computationally infeasible here)
A rejection of the null hypothesis $H_0 : \mathcal{P}^* \neq \emptyset$

$\rightarrow$ $\not\exists$ 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

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

whose distribution can be approximated as before.

A level $\alpha$ misspecification test rejects $H_0 : \mathcal{P}^* \neq \emptyset$ 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

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

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

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

where \(\{W_i^*\}_{i=1}^{n}\) 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

$$\tilde{Q}_n(t) := \min_{(P, \lambda, g, h)} Q^*_n(P, \lambda, g, h)$$

s.t. $(P, \lambda) \in \hat{R}^*(t)$ and $(P, \lambda) + n^{-1/2}(g, h) \in \mathcal{P}^\dagger_d(t) \times \mathbb{R}^{d_s}_+$,

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

- The distribution of $\tilde{Q}_n(t)$ approximated by redrawing $\{W_i^*\}_{i=1}^n$ a large number $(B)$ of times and computing $\tilde{Q}_n(t)$ for each draw

- *The CNS test* rejects $H_0 : t \in \Theta^*$ when $\overline{Q}_n(t)$ is larger than the $1 - \alpha$ quantile of these $B$ values of $\tilde{Q}_n(t)$.

- A $1 - \alpha$ *CNS confidence region for $\Theta^*$ is the set of all $t$ for which the CNS test does not reject.*