SOCIAL INTERACTION
AND SICKNESS ABSENCE

Assar Lindbeck, Mårten Palme
and Mats Persson
Puzzle in sickness absence:

The number of persons absent from work because of sickness, Sweden 1974-2003
Background

Social interaction

Social norms

Consequences for moral hazard in social insurance

Social multiplier
Background

Difficult to measure norms directly

Instead: indicators of the operations of norms
Background

We study effects of group behavior on individual behavior

Problem: Difficult to separate group effects from effects of selection and unobserved heterogeneity
Manski (1999, 2000)

- Correlated effect.
- Contextual effect.
- Social interaction.
Case study: Sickness absence in Sweden

Variation across geographical areas

Data set = LOUISE + National Social Insurance Board
Case study: Sickness absence in Sweden

Variation across geographical areas

Data set = LOUISE + National Social Insurance Board

How to define geographical areas?
SAMS – Small Area Market Statistics

- Geographical definition constructed by Statistics Sweden
- Idea is to create an area with the same type of housing within each church parish
- Around 1,000 individuals within each SAMS
- About 9,299 SAMS districts
\[ \bar{S}_n = \sum_{i=1}^{N_n} S_{in} / N_n, \quad n = 1, \ldots, 9003 \]

Large standard deviation \((\sigma = 5.4)\), large range \((330)\).

Can socio-economic factors explain the variation?

Three types of variables.
Variables in the $X$ vector

<table>
<thead>
<tr>
<th>for the individual</th>
<th>age (all ages from 18 to 64, one dummy for each age, i.e., 46 dummies)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>education (seven levels, one dummy for each, i.e., six dummies)</td>
</tr>
<tr>
<td></td>
<td>gender (one dummy)</td>
</tr>
<tr>
<td></td>
<td>marital status (two dummies)</td>
</tr>
<tr>
<td></td>
<td>having children aged 3 or younger (one dummy)</td>
</tr>
</tbody>
</table>
Variables in the $X$ vector (cont’d)

<table>
<thead>
<tr>
<th><strong>for the workplace:</strong></th>
<th><strong>industry</strong> (60 industries, i.e., 59 dummies)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>sector</strong></td>
<td><em>(central government authority, state-owned enterprise, local government authority, etc.), 10 dummies)</em></td>
</tr>
<tr>
<td><strong>size of workplace</strong></td>
<td><em>(21 dummies: 1 employee, 2-10, 11-20, 21-30, …, 91-100, etc.)</em></td>
</tr>
</tbody>
</table>
Variables in the $X$ vector (cont’d)

| for the geographical area: | town or countryside (one dummy)  
|                            | life expectancy in the region  
|                            | local unemployment  
|                            | (expressed by the incidence of unemployment, 19 dummy variables, one for each 5-percent interval) |
\[ S_{in} = \alpha + X_{in}' \beta + \varepsilon_{in} \]
\[ S_{in} = \alpha + X_{in} \beta + \varepsilon_{in} \]

\[ \bar{\varepsilon}_n = \sum_{i}^{N_n} \varepsilon_{in} / N_n, \quad n = 1, ..., 9003 \]

Large remaining variation.

\[ \sigma \quad 5.4 \rightarrow 4.6 \quad \text{R. about the same.} \]

Is this a random phenomenon?
\[ S_{in} = \alpha_n + X_{in} \beta + \varepsilon_{in} \]

Neighborhood-specific intercepts.

\( F = 5.880. \)

Significant neighborhood effects at 1 percent level.
What does this reflect?
What does this reflect?

Unobserved permanent factors?
What does this reflect?
Unobserved permanent factors?
Should not influence changes.
What does this reflect?
Unobserved permanent factors?
Should not influence changes.

\[ \Delta S_{in} = \alpha + X_{in} \beta + \varepsilon_{in} \]
What does this reflect?

Unobserved permanent factors?

Should not influence changes.

\[ \Delta S_{in} = \alpha + X_{in} \beta + \varepsilon_{in} \]

Alternatively: specific \( \alpha_n \). Still, significant neighborhood effects.

\( F \) test : 1.559
Conclusion: There is something else.
Conclusion: There is something else.

Could that be social norms?
Conclusion: There is something else.

Could that be social norms?

Is individual work absence causally related to average work absence in the local neighborhood?
One way of modelling social norms is

\[ S_{in} = \alpha + X_{in}' \beta + \gamma \bar{S}_n + \varepsilon_{in} \]
One way of modelling social norms is

\[ S_{in} = \alpha + X_{in} \beta + \gamma \bar{S}_n + \varepsilon_{in} \]

However: the reflection problem.
One way of modelling social norms is

\[ S_{in} = \alpha + X_{in}' \beta + \gamma \bar{S}_n + \varepsilon_{in} \]

However: the reflection problem.

Four different approaches to deal with this problem.
Four approaches:
Four approaches:

1. Public sector employees have higher work absence rate. We use the share of public sector employees as instrumental variable.
Four approaches:

2. We analyze the interaction between neighborhoods and workplaces. Are you more likely to be influenced by the absence behavior of your neighbors if they are also your workmates?
Four approaches:
3. We ask whether movers within the country tend to adjust their sick-absence to the area where they are moving to.
Four approaches:

4. We ask whether immigrants tend to adjust their sick-absence behavior to the behavior of natives in their neighborhood.
Four approaches:
Four approaches:

Within each of these four approaches, we use alternative specifications and estimation techniques.
Four approaches:

Within each of these four approaches, we use alternative specifications and estimation techniques.

Each approach has specific weaknesses, but all approaches point in the same direction.
Four approaches:

Within each of these four approaches, we use alternative specifications and estimation techniques.

Each approach has specific weaknesses, but all approaches point in the same direction.
Public-sector vs. private-sector employees
Public-sector vs. private-sector employees

\[ Z_n \equiv \text{proportion of population in neighborhood } n \text{ that works in the public sector.} \]

IV approach
\[
\begin{align*}
\bar{S}_n & = a + X_{in}' b + cZ_n + e_n \\
S_{in}^{\text{priv}} & = \alpha + X_{in}' \beta + \gamma \hat{S}_n + \varepsilon_{in}
\end{align*}
\]
\[
\begin{align*}
\bar{S}_n &= a + X'_{in} b + cZ_n + e_n \\
S^\text{priv}_{in} &= \alpha + X'_{in} \beta + \gamma \hat{S}_n + \varepsilon_{in}
\end{align*}
\]

This gives us an estimate of $\gamma$ in the equation

\[
S^\text{priv}_{in} = \alpha + X'_{in} \beta + \gamma \bar{S}_n + \varepsilon_{in}
\]
Let’s the do it the other way around:

\[
\begin{align*}
\bar{S}_n &= \alpha + X_{in}' b + c(1 - Z_n) + e_n \\
S_{in}^{\text{publ}} &= \alpha + X_{in}' \beta + \gamma \hat{S}_n + \varepsilon_{in}
\end{align*}
\]
<table>
<thead>
<tr>
<th>Population</th>
<th>Number of individuals and observations</th>
<th>Regressor</th>
<th>Reduced form ($\mu$ in eq. (4))</th>
<th>First step in IV regression ($c$ in eq. (3))</th>
<th>IV estimate ($\gamma$ in eq. (3))</th>
</tr>
</thead>
<tbody>
<tr>
<td>All those who work in private sector</td>
<td>2,839,410 ind. 14,556,753 obs.</td>
<td>Share of population in neighborhood $n$ that work in public sector ($Z_{nt}$)</td>
<td>0.0387*** (0.0013) $R^2 = 0.020$</td>
<td>6.670*** (0.0116) $R^2 = 0.499$</td>
<td>0.581*** (0.0199) $R^2 = 0.0215$</td>
</tr>
<tr>
<td>All those who work in public sector</td>
<td>1,956,740 ind. 10,502,405 obs.</td>
<td>Share of population in neighborhood $n$ that work in private sector ($1 - Z_{nt}$)</td>
<td>-0.0438*** (0.0017) $R^2 = 0.026$</td>
<td>-5.752*** (0.0123) $R^2 = 0.512$</td>
<td>0.762*** (0.0302) $R^2 = 0.274$</td>
</tr>
<tr>
<td>All employees</td>
<td>4,796,150 ind. 25,059,158 obs.</td>
<td>Share of population in neighborhood $n$ that work in public sector ($Z_{nt}$)</td>
<td>0.0418*** (0.0011) $R^2 = 0.024$</td>
<td>6.222*** (0.0084) $R^2 = 0.503$</td>
<td>0.672*** (0.0173) $R^2 = 0.0252$</td>
</tr>
</tbody>
</table>
Problems with the IV approach?

Selection problem? Some individuals choose to live in specific neighborhoods?
Problems with the IV approach?

Selection problem? Some individuals choose to live in specific neighborhoods?

Control for gender, age, education, etc.
Problems with the IV approach?

Selection problem? Some individuals choose to live in specific neighborhoods?

Control for gender, age, education, etc. Unobservables that are uncorrelated with these?
Problems with the IV approach?

Selection problem? Some individuals choose to live in specific neighborhoods?

Control for gender, age, education, etc. Unobservables that are uncorrelated with these?

Workplace fixed effects?
Refutability test

\[ EDU_{\text{int}}^{\text{priv}} = a + k_t + X_{\text{int}}' \beta + \mu_1 Z_{nt} + e_{nt} \]

\[ EDU_{\text{int}}^{\text{publ}} = a + k_t + X_{\text{int}}' \beta + \mu_2 Z_{nt} + e_{nt} \]

\[ \mu_1 = 0.113^{***} \]

\[ \mu_2 = -0.198^{***} \]
Interaction Model

\[ S_{ijk} = \delta + X_{ijk}' \beta + \gamma(CA_{jk} \cdot \bar{S}_k) + \lambda_j + \mu_k + \varphi CA_{jk} + \varepsilon_{ijk} \]

\( j \) – sub-index for work place.
\( k \) - sub-index for neighborhood.
CA – concentration of workers from the same neighborhood working at the same work place.

**Identification:** through the interaction between the average work absence level in the neighborhood and the concentration of workers from the same neighborhood working at the same workplace.

**Note:** controls for fixed workplace and neighborhood effects as well as the concentration of workers from the same workplace.
Interaction Model

\[ S_{ijk} = \delta + X_{ijk}' \beta + \gamma (CA_{jk} \cdot \bar{S}_k) + \lambda_j + \mu_k + \varphi CA_{jk} + \varepsilon_{ijk} \]

**Interpretation:** if there is an above average work absence in the neighborhood, the expected work absence increases with factor \( \gamma \) with the concentration of workers from the same neighborhood.
Interaction Model

• Corresponding model for the effect of the social interaction at the workplace.

\[ S_{ijk} = \delta + X_{ijk} \beta + \gamma (CA_{jk} \ast \overline{S}_j) + \lambda_j + \mu_k + \varphi CA_{jk} + \varepsilon_{ijk} \]
Results from the Interaction Model

Table 4: The strength-of-network effect

<table>
<thead>
<tr>
<th>Number of obs. and ind.</th>
<th>$\hat{\nu}$</th>
<th>$\frac{\partial S_{wn}}{\partial \bar{S}_n}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>24,449,603 obs.</td>
<td>2.146***</td>
<td>0.0502</td>
<td>0.012</td>
</tr>
<tr>
<td>4,693,560 ind.</td>
<td>(0.0273)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The numbers of observations and individuals in this table are somewhat smaller than the corresponding numbers in Table 1. The reason is that for each individual, we have deleted the individual himself from the data when computing the averages $\bar{S}_n$. For some neighborhoods, there is only one individual who works in each workplace; these cases therefore do not appear in the regression.
Approach No. 3: Movers within the country.

• If an individual moves from a high-absence neighborhood to a low-absence one does he/she change work absence behavior?
Approach No. 3: Movers within the country.

\[
S_{\text{mover}}^{\text{int}} - S_{\text{mover}}^{\text{in},t-1} = \alpha + \lambda_t + (X_{\text{int}} - X_{\text{in},t-1}) \beta + \eta \cdot (\bar{S}_{\text{m},t-1}^{\text{non-mover}} - \bar{S}_{\text{n},t-1}^{\text{non-mover}}) + \varepsilon_{\text{imnt}}
\]

\[
S_{\text{mover}}^{\text{int}} - S_{\text{in},t-1} = \alpha + \lambda_t + (X_{\text{int}} - X_{\text{in},t-1}) \beta + \eta \cdot (\bar{S}_{\text{m},t-1}^{\text{non-mover}} - \bar{S}_{\text{n},t-1}^{\text{non-mover}}) +
\]

\[+ \delta \cdot D_{it} \cdot (\bar{S}_{\text{m},t-1}^{\text{non-mover}} - \bar{S}_{\text{n},t-1}^{\text{non-mover}}) + \varepsilon_{\text{imnt}} \]

\[
D_{it} = \begin{cases} 
1 & \text{if individual } i \text{ has moved at time } t \text{ to an area with } \\
& \text{higher absensce, i.e., if } \bar{S}_{\text{m},t-1}^{\text{non-mover}} > \bar{S}_{\text{n},t-1}^{\text{non-mover}} \\
0 & \text{otherwise.}
\end{cases}
\]
**Approach No. 3: Movers within the country.**

Table 2: Movers within Sweden

<table>
<thead>
<tr>
<th>Specification</th>
<th>No. of individuals and observations</th>
<th>$R^2$</th>
<th>$\hat{\eta}$</th>
<th>$\hat{\delta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetric specification (5)</td>
<td>1,551,059 ind. 2,202,466 obs.</td>
<td>0.0055</td>
<td>0.032*** (0.00678)</td>
<td>—</td>
</tr>
<tr>
<td>Asymmetric specification (5')</td>
<td>1,551,059 ind. 2,202,466 obs.</td>
<td>0.0055</td>
<td>0.028*** (0.0094)</td>
<td>0.052 (0.0777)</td>
</tr>
</tbody>
</table>
Approach No. 4: immigrants vs. natives

\[ S_{in}^m = \alpha + \beta X_{in}^m + \gamma \bar{S}_n^s + \varepsilon_{in} \]
<table>
<thead>
<tr>
<th>Region</th>
<th>Number of ind. and obs.</th>
<th>Estimate of $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All regions</td>
<td>720,742 ind. 3,376,753 obs.</td>
<td>0.629*** (0.0063)</td>
</tr>
<tr>
<td>Nordic countries</td>
<td>210,059 ind. 1,088,923 obs.</td>
<td>0.651*** (0.0174)</td>
</tr>
<tr>
<td>EU (except Nordic countries)</td>
<td>77,982 ind. 358,797 obs.</td>
<td>0.461*** (0.0242)</td>
</tr>
<tr>
<td>Europe (except EU)</td>
<td>154,378 ind. 744,440 obs.</td>
<td>0.126*** (0.0176)</td>
</tr>
<tr>
<td>Africa</td>
<td>38,422 ind. 163,554 obs.</td>
<td>0.091*** (0.0308)</td>
</tr>
<tr>
<td>North America</td>
<td>21,655 ind. 92,321 obs.</td>
<td>0.278*** (0.0360)</td>
</tr>
<tr>
<td>Latin America</td>
<td>36,556 ind. 167,644 obs.</td>
<td>0.345*** (0.0347)</td>
</tr>
<tr>
<td>Asia</td>
<td>173,447 ind. 723,644 obs.</td>
<td>0.237*** (0.0157)</td>
</tr>
<tr>
<td>Oceania</td>
<td>3,626 ind. 14,146 obs.</td>
<td>0.222*** (0.0766)</td>
</tr>
<tr>
<td>Former Soviet Union</td>
<td>4,398 ind. 22,566 obs.</td>
<td>0.037 (0.0949)</td>
</tr>
</tbody>
</table>
• We have excluded neighborhoods with less than 20% natives.
• Sensitivity analysis: only recent immigrants.
Conclusions

• Is there evidence of social interaction?
• What is the magnitude of this effect?
• Four different strategies: Each captures a specific form of social interaction.
• Other forms are also possible:
  – National level through the mass media.
  – Social interactions outside the neighborhood: relatives, bowling clubs or unions.
Lower bound estimates:

• Interaction model: effect only through interaction between work and neighborhood networks.

• Movers within Sweden: Only partial effect since the entire adjustment is not likely to happen the first year.

• 0.032 and 0.050, respectively.
Upper bound estimates:

- IV model.
- Immigrants.
- 0.672 and 0.581, respectively.