Big Data for Macroeconomists and Policy Makers

Gianluca Violante
Princeton University, CEBI, CEPR, IZA and NBER

CEPR & TFI Event – Household Welfare: Do We Need Big Data?
A total of five exabytes of information was created between the dawn of civilization through 2003. That much information is now created every two days.

Eric Schmidt, former CEO of Google, 2010
Big data in household finance

1. Why are big data useful to the policy maker?

- **Ex-ante**: to *monitor* household financial fragility
- **Throughout**: to *target* policy interventions
- **Ex-post**: to *evaluate* effectiveness of intervention
Big data in household finance

1. Why are big data useful to the policy maker?
   - **Ex-ante:** to *monitor* household financial fragility
   - **Throughout:** to *target* policy interventions
   - **Ex-post:** to *evaluate* effectiveness of intervention

2. Are big data always *superior* to survey data?
   - Some *thoughts*
Ex-ante: monitor household balance sheets

- Lesson from GR: interconnection btw hh and bank balance sheets
  - Policy response: stress tests on banks’ side
  - Feasible because banks collect big data at their own cost
Ex-ante: monitor household balance sheets

- Lesson from GR: interconnection btw hh and bank balance sheets
  - Policy response: stress tests on banks’ side
  - Feasible because banks collect big data at their own cost

- Equally important, but absent: households’ stress tests
  - Need: detailed, high frequency data on hh balance sheets
  - Simulate ‘crisis scenarios’ and estimate impact of shocks
  - Inform intelligent design of policy
  - Ongoing research with economists at the NY Fed
Ex-ante: monitor household balance sheets

- Only one determinant of household shock absorption capacity
- Challenging measurement (with available data), but important
- You can trace distribution by: area, income, age, credit score, etc.

G. Violante, "Household Welfare and Big Data"
Throughout: target policies

- **Social insurance**: credit or transfers to households in need

- **Crisis**: urgency calls for *easy-to-implement* policies
  - Easy to implement: depends on available data you have
  - You end up with universal policies with low impact
  - **CARES Act**: Any family of 4 with $AGI < 150,000 gets $3,400

- Big data allow to tailor intervention: *more bang for the buck*

- You channel financial help to those who *really* need it
Throughout: target policies

- **Lockdown** has heterogeneous effects in the population
  - Key: whether you are in a *flexible* or *rigid* occupation
  - Dingel-Neiman definitions based on O*NET tasks

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Income</th>
<th>Flexible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economist</td>
<td>$100K</td>
<td>Y</td>
</tr>
<tr>
<td>Optometrist</td>
<td>$100K</td>
<td>N</td>
</tr>
<tr>
<td>Secondary school teacher</td>
<td>$50K</td>
<td>Y</td>
</tr>
<tr>
<td>Flight attendant</td>
<td>$50K</td>
<td>N</td>
</tr>
<tr>
<td>Telemarketer</td>
<td>$25K</td>
<td>Y</td>
</tr>
<tr>
<td>Cook</td>
<td>$25K</td>
<td>N</td>
</tr>
</tbody>
</table>

- A transfer *conditional on occupation* could be more generous
Ex-post: policy evaluation

1. Impact evaluation using **diff-in-diff**
   - Big data improve design of the quasi-experiment
   - Big data allow to capture *heterogeneous treatment effects*
   - Cannot capture GE feedbacks (the intercept)
1. Impact evaluation using diff-in-diff
   - Big data improve design of the quasi-experiment
   - Big data allow to capture heterogeneous treatment effects
   - Cannot capture GE feedbacks (the intercept)

2. Impact evaluation using: structural equilibrium models
   - State-of-the-art macro models capture heterogeneity in MPC
   - Key for transmission mechanism of shock and policy response
   - Big data allow to properly calibrate these models
   - Assess distributional impact of policy
Ex-post: policy evaluation

- Example: monetary policy tightening in HANK models

- Do we see this same heterogeneity in (big) data?

G. Violante, "Household Welfare and Big Data"
Are big data always superior to survey data?

1. **Virtues** of big data vis-à-vis survey data

   • Large-scale
   
   • Contain novel variables (e.g., individual stocks, mobility)
   
   • High-frequency
   
   • Real time
   
   • High quality (less measurement error)

   They allow to truly **embrace and measure heterogeneity**

       (household balance sheets, preferences, financial literacy, etc...)
Are big data always superior to survey data?

2. **Challenges** associated with the use of big data

- **Representativeness** of the underlying universe
  - Data that are not meant to be representative
  - Coverage is often time-varying / endogenous

- **Access** to the dataset is often restricted
  - Inequality in access (connections / research budget)
  - Replicability of the findings

- **Privacy issues**
  - It’s about protecting: *not an argument against collecting*