

The Transformative Role of Administrative Data in Evaluating Economic Policies

John N. Friedman
Brown University and NBER

January 2015

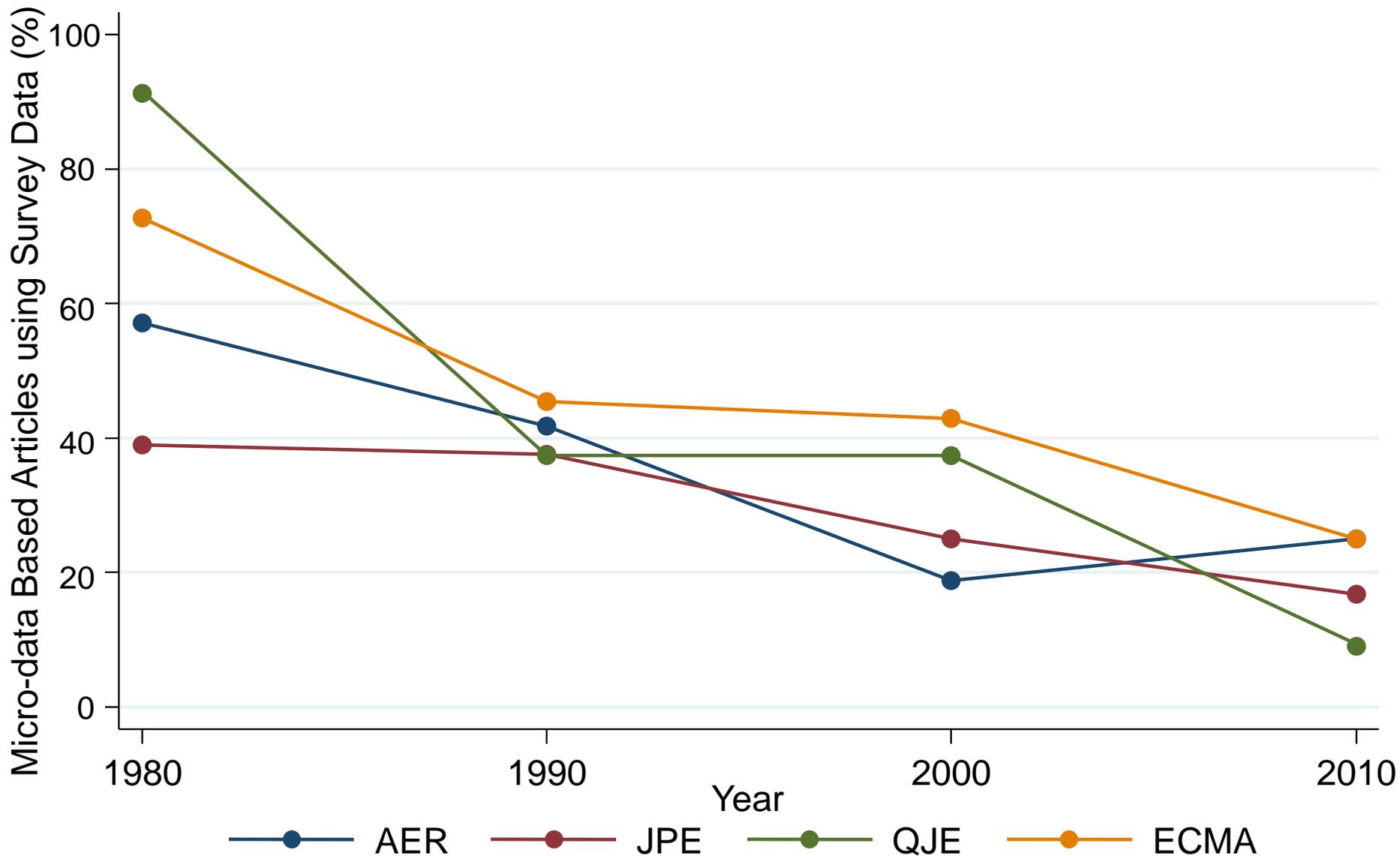
Introduction

- Administrative datasets are transforming the analysis of economic policy
- This talk discusses the key features of this trend using different recent research as examples.
- Conclude with recommendations on the keys to sustainable development of administrative data in the U.S.

Administrative Data in the U.S.

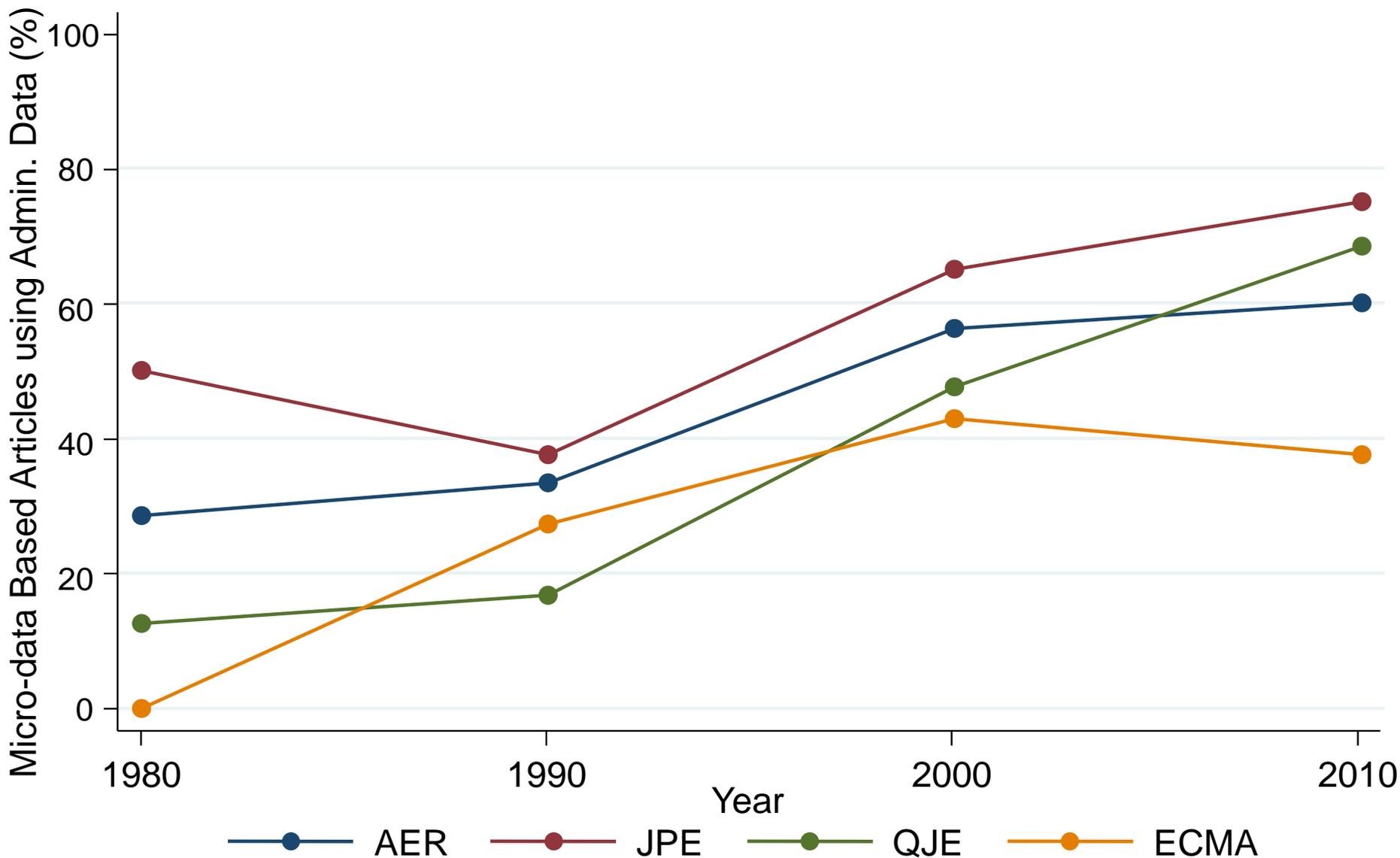
- Papers utilizing admin data ever more dominant in top journals
- Demand for admin. data has led to shift to studying European countries
 - Population registers in Sweden, Austria, Germany, Norway, Denmark
 - Similar data infrastructure starting to be developed in the U.S., but still very fragmented
- Administrative data can help advance research in many different areas:
 - Labor Markets
 - Medicare / Medicaid Health data
 - Education
 - Patents / Innovation
 - Business formation and growth

Use of Pre-Existing Survey Data in Publications in Leading Journals, 1980-2010



Note: "Pre-existing survey" datasets refer to micro surveys such as the CPS or SIPP and do not include surveys designed by researchers for their study. Sample excludes studies whose primary data source is from developing countries.

Use of Administrative Data in Publications in Leading Journals, 1980-2010



Note: "Administrative" datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records). Sample excludes studies whose primary data source is from developing countries.

What are the Benefits of Administrative Data?

1. Excellent Coverage

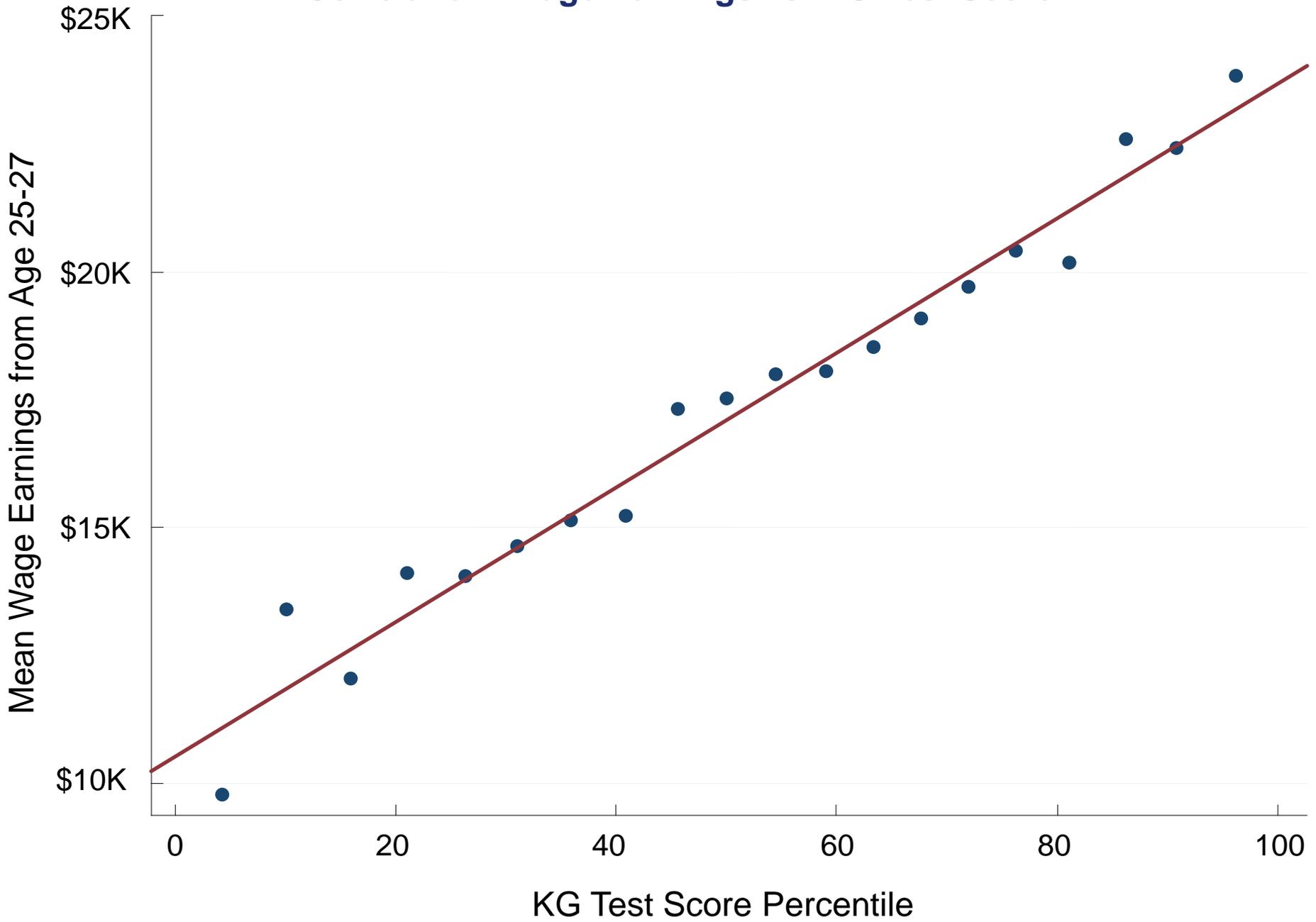
- Longitudinal tracking rates near 100% over long periods of time
- Allows matching of outside data, either retrospectively or in real time

Example: Long-Term Benefits of Kindergarten

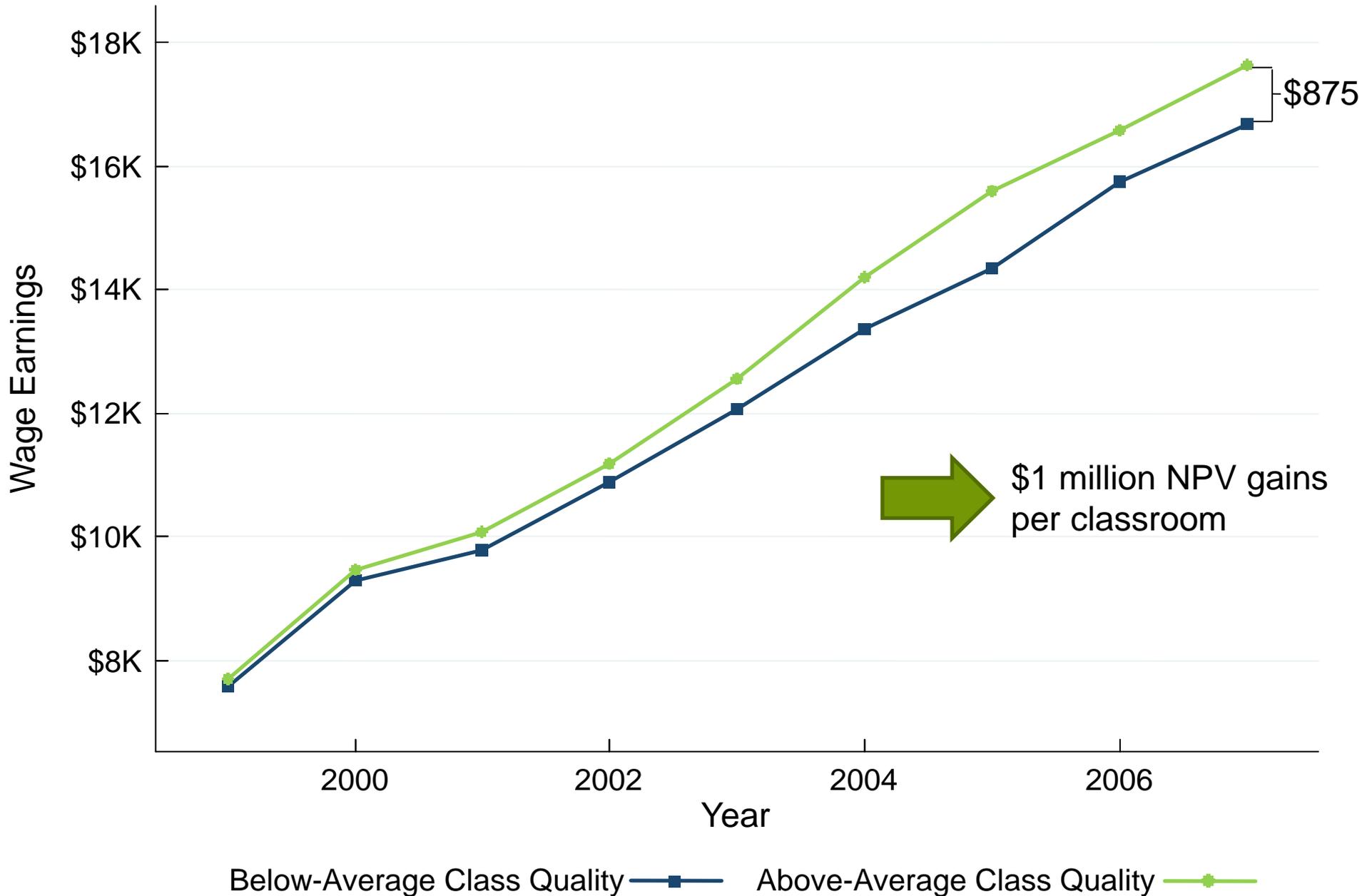
- Question: What are the long-term effects of education policy?
- Project STAR
 - More than 100 papers studying effects on test scores
- We match school records from 12,000 kindergarten students to tax data
 - Match rate of 95%
 - Less attrition after 22 years than at the end of Kindergarten in the original survey data.

Source: Chetty, Friedman, Hilger, Saez, Schanzenbach, Yagan (2011)

Correlation: Wage Earnings vs. KG Test Score



Causal Effect of Kindergarten Class Quality on Earnings by Year



What are the Benefits of Administrative Data?

1. Excellent Coverage

- Longitudinal tracking rates near 100% over long periods of time
- Allows matching of outside data, either retrospectively or in real time

2. **Higher quality information:** virtually no missing data or attrition

- Current Population Survey non-response rate now 31% for income
- Far cheaper to ensure accurate information

Example: Effects of the EITC on Earnings

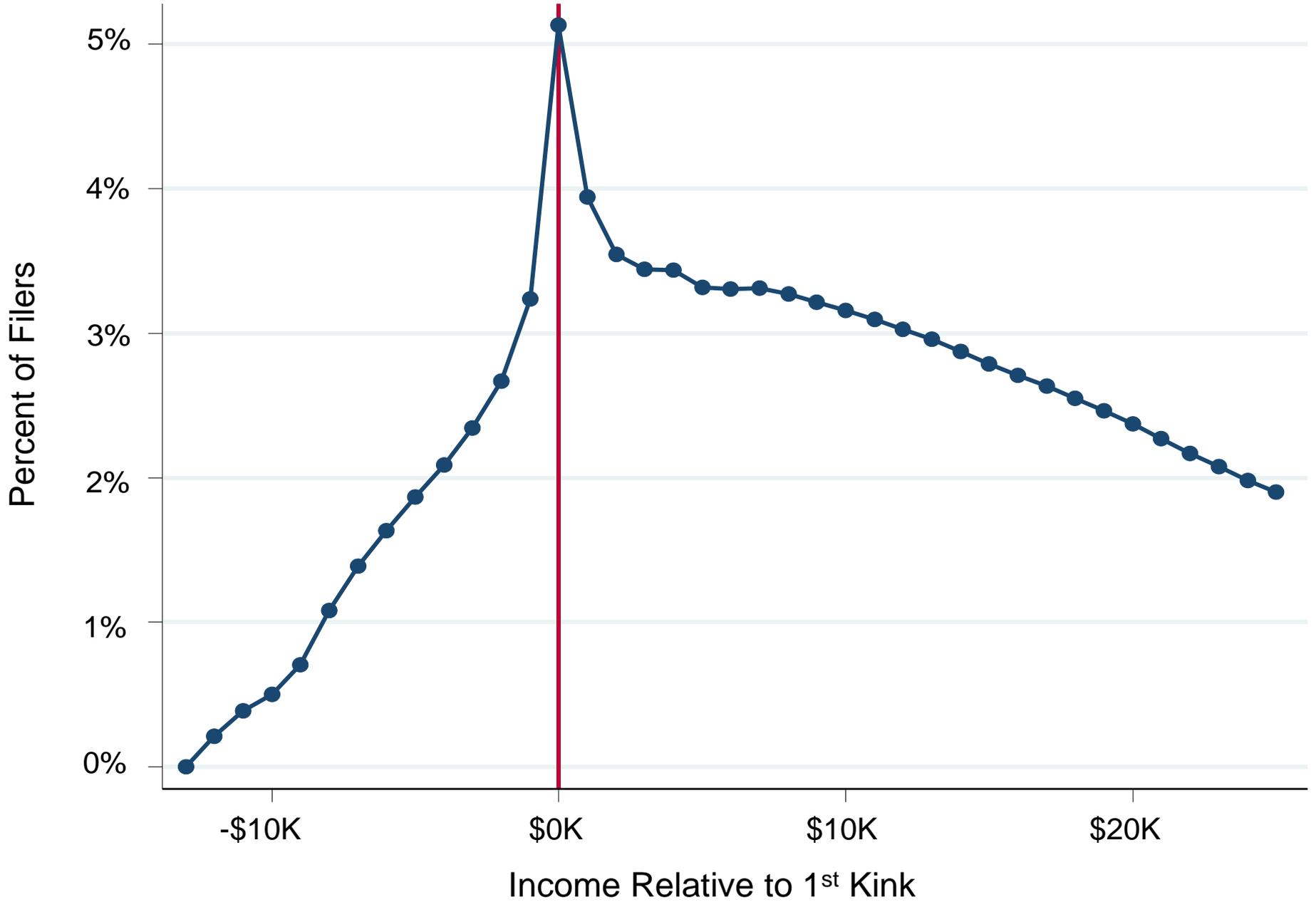
- Question: How does the EITC affect earnings?
 - Largest cash transfer program in the U.S. (\$60 billion)
- Idea: Tax payers who are unaware of the incentives embedded in the EITC will behave as if the program did not exist.
 - We use variation across cities in tax payers' information about the details of the program.
 - Measure information using fraction reporting income exactly to receive the maximum EITC

Source: Chetty, Friedman, Saez (2013)

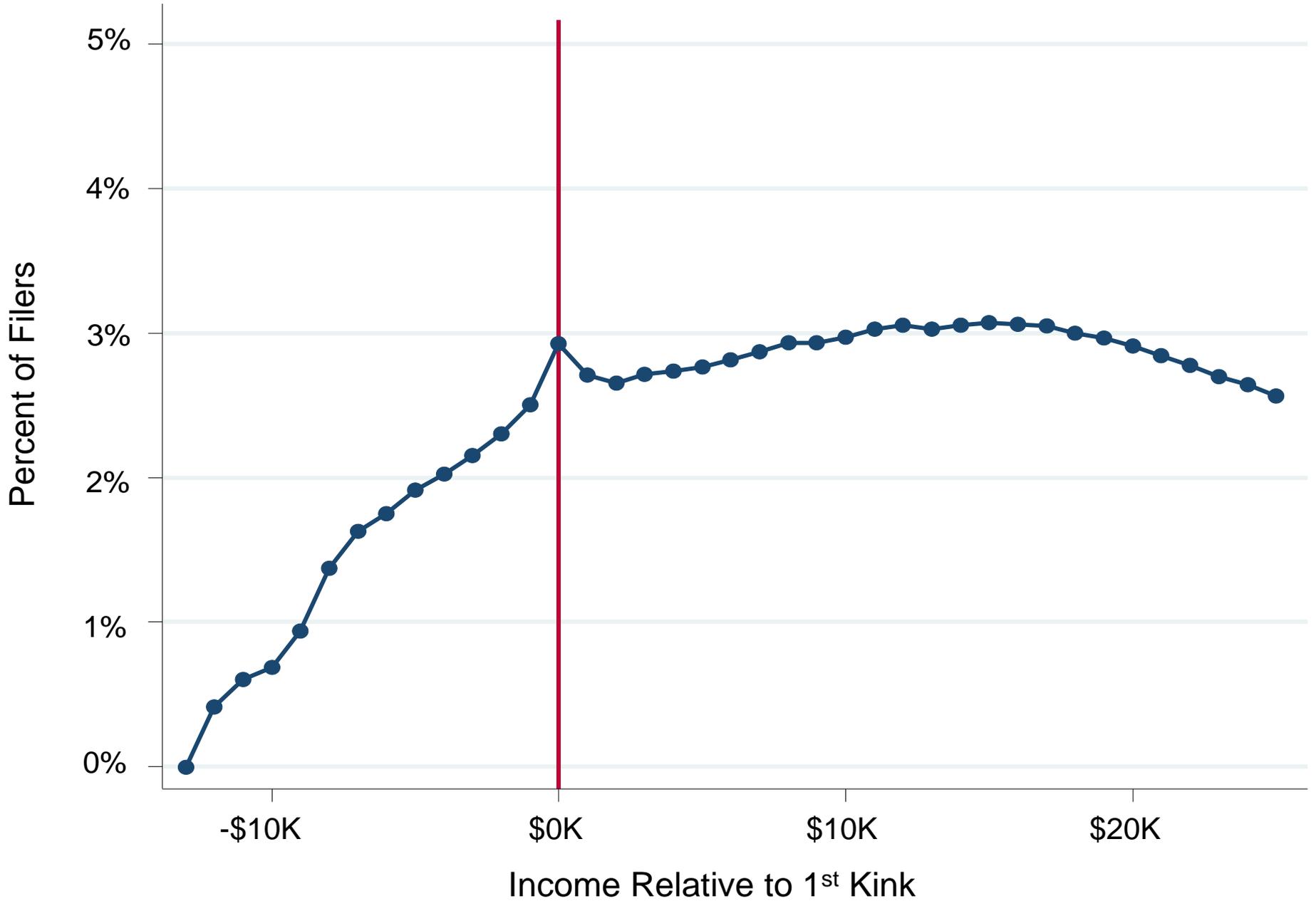
2008 Federal EITC Schedule for a Single Filer with Children



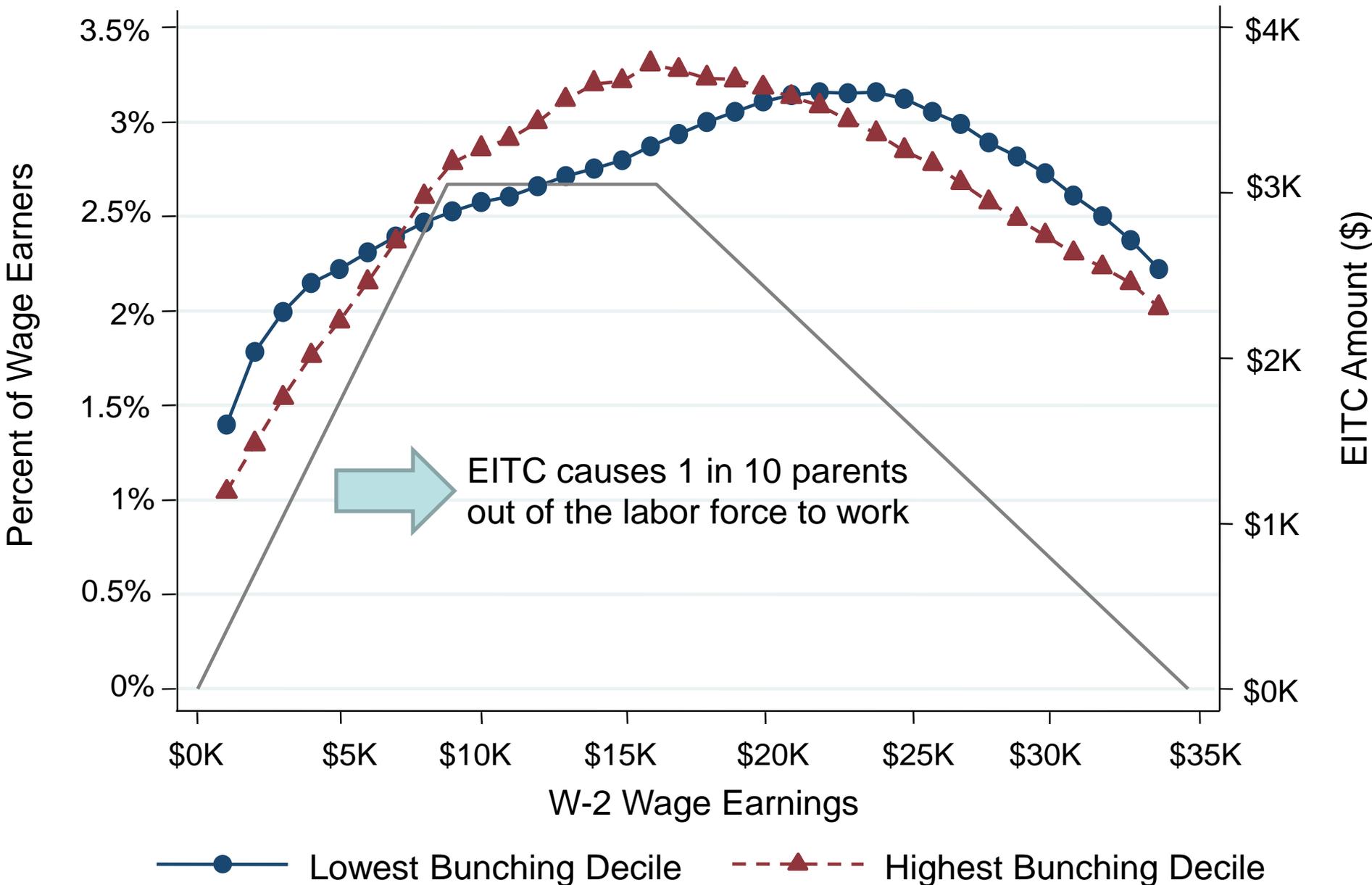
Earnings Distribution in Texas



Earnings Distribution in Kansas



Income Distribution For Single Wage Earners with One Child High vs. Low Bunching Areas



What are the Benefits of Administrative Data?

1. Excellent Coverage

- Longitudinal tracking rates near 100% over long periods of time
- Allows matching of outside data, either retrospectively or in real time

2. Higher quality information: virtually no missing data or attrition

- Current Population Survey non-response rate now 31% for income
- Far cheaper to ensure accurate information

3. Very Large Samples: 2,000 times the size of the CPS

- Can develop new non-parametric, quasi-experimental research designs
- Explore heterogeneity in its own right

Example: Parental Income and College Attendance

- Question: Where do children from different families go to college?
- Administrative data allows analysis of this question with significantly more granularity than previous research
- Work in progress will estimate the distribution of parental income at the college level for release later in 2015

Keys to Sustainable Data Development in the U.S.

1. **Education:** Most law-makers and bureaucrats (and citizens) do not understand the benefits of research that uses administrative data
 - Moving in the right direction (e.g., Murray-Ryan Data Commission Bill)
 - Shared responsibility among academics

Keys to Sustainable Data Development in the U.S.

1. **Education:** Most law-makers and bureaucrats (and citizens) do not understand the benefits of research that uses administrative data.
2. **Resources:** Bandwidth to support projects without burdening host agencies
 - Direct researcher access to micro-level data
 - Fair and open competition for data access based on scientific merit

Keys to Sustainable Data Development in the U.S.

1. **Education:** Most law-makers and bureaucrats (and citizens) do not understand the benefits of research that uses administrative data.
2. **Resources:** Bandwidth to support projects without burdening host agencies
3. **Systematic Electronic Monitoring:** Ensure data integrity and efficient access
 - Careful balance necessary
 - Example from Denmark: Social enforcement

Keys to Sustainable Data Development in the U.S.

1. **Education:** Most law-makers and bureaucrats (and citizens) do not understand the benefits of research that uses administrative data.
2. **Resources:** Bandwidth to support projects without burdening host agencies
3. **Systematic Electronic Monitoring:** Ensure data integrity and efficient access
4. **Data-Sharing Between Agencies:** Current legal requirements prohibit information sharing that would benefit both research and administration, e.g.:
 - FAFSA simplification
 - Social support program eligibility (or not)
 - Sample frame / data verification for Census and other surveys (e.g., National Directory of New Hires)