

# The Global Income Dynamics (GID) Project: An Overview and Some New Insights

---

Fatih Guvenen

U of Minnesota, FRB Minneapolis, and NBER

*OECD's LinkEED and Book Launch Event*

*December 9, 2021*

# The Global Income Dynamics (GID) Project: A Bird's Eye View

**Joint initiative:** Princeton (Gianluca Violante), Stanford (Luigi Pistaferri), U of Minnesota & MEBDI (Fatih Guvenen)

# The Global Income Dynamics (GID) Project: A Bird's Eye View

**Joint initiative:** Princeton (Gianluca Violante), Stanford (Luigi Pistaferri), U of Minnesota & MEBDI (Fatih Guvenen)

**Goal:** To build an open-access, cross-country

# The Global Income Dynamics (GID) Project: A Bird's Eye View

**Joint initiative:** Princeton (Gianluca Violante), Stanford (Luigi Pistaferri), U of Minnesota & MEBDI (Fatih Guvenen)

**Goal:** To build an open-access, cross-country

- 1 **harmonized** database
- 2 of detailed **micro statistics**
- 3 on **income inequality & income dynamics**
- 4 based on **panel** data
- 5 from **Administrative records**

# The Global Income Dynamics (GID) Project: A Bird's Eye View

**Joint initiative:** Princeton (Gianluca Violante), Stanford (Luigi Pistaferri), U of Minnesota & MEBDI (Fatih Guvenen)

**Goal:** To build an open-access, cross-country

- 1 harmonized database
  - 2 of detailed micro statistics
  - 3 on income inequality & income dynamics
  - 4 based on panel data
  - 5 from Administrative records
- ▶ Project has been underway for the past 3+ years.

## Current State: GID Version 1.0

- ▶ **13 countries:** US, UK, CAN, FRA, ITA, SPA, GER, NOR, SWE, DEN, MEX, BRA, ARG.
- ▶ **54 economists** in 13 country teams.
- ▶ More than **1 million statistics** per country (work completed)

## Current State: GID Version 1.0

- ▶ **13 countries:** US, UK, CAN, FRA, ITA, SPA, GER, NOR, SWE, DEN, MEX, BRA, ARG.
- ▶ **54 economists** in 13 country teams.
- ▶ More than **1 million statistics** per country (work completed)
- ▶ Database infrastructure/website currently being built. Expected to go live in **first half of 2022**.

## Current State: GID Version 1.0

- ▶ **13 countries:** US, UK, CAN, FRA, ITA, SPA, GER, NOR, SWE, DEN, MEX, BRA, ARG.
- ▶ **54 economists** in 13 country teams.
- ▶ More than **1 million statistics** per country (work completed)
- ▶ Database infrastructure/website currently being built. Expected to go live in **first half of 2022**.
- ▶ A Special Issue of **Quantitative Economics** with 13 papers written by country teams will be published in 2022.



# Current State: GID Version 1.0

- ▶ **13 countries:** US, UK, CAN, FRA, ITA, SPA, GER, NOR, SWE, DEN, MEX, BRA, ARG.
- ▶ **54 economists** in 13 country teams.
- ▶ More than **1 million statistics** per country (work completed)
- ▶ Database infrastructure/website currently being built. Expected to go live in **first half of 2022**.
- ▶ A Special Issue of **Quantitative Economics** with 13 papers written by country teams will be published in 2022.

## GID Version 2.0:

- ▶ Plan is to gradually **open to more countries** starting in 2022.
  - About 10 country teams ready to join. Final number likely to be higher.

- **Argentina:** "The Evolution of the Earnings Distribution in a Volatile Economy: Evidence from Argentina," Blanco, Diaz de Astarloa, Drenik, Moser, and Trupkin.
- **Brazil:** "Earnings Inequality and Dynamics in the Presence of Informality: The Case of Brazil" by Engbom, Gonzaga, Moser, and Olivieri.
- **Canada:** "Four Decades of Canadian Earnings Inequality and Dynamics Across Workers and Firms", by Bowlus, Gouin-Bonenfant, Liu,Lochner, and Park.
- **Denmark:** "Trends in Income Risk in Denmark 1987-2016" by Leth-Petersen and Saeverud
- **France:** "Inequality and Earnings Dynamics in France: National Policies and Local Consequences", by Kramarz, Nimier-David, and Delemotte.
- **Germany:** "Inequality and Income Dynamics in Germany", by Drechsel-Grau, Peichl, Schmieder, Schmid, Walz, and Wolter.
- **Italy:** "Labor Reforms and Earnings Dynamics: The Italian Case", by Hoffman, Malacrino, and Pistaferri.
- **Mexico:** "Income Dynamics and Inequality: The Case of Mexico", by Puggioni, Calderón, Cebreros Zurita, Fernandez Bujanda, Gonzalez, and Jaume.
- **Norway:** "Earnings Dynamics and Its Intergenerational Transmission: Evidence from Norway", by Halvorsen, Ozkan, and Salgado.
- **United States:** "U.S. Long-Term Earnings Outcomes by Sex, Race, Ethnicity, and Place of Birth", by Abowd, McKinney, and Janicki.
- **United Kingdom:** "Income Dynamics in the United Kingdom 1975-2020", by Bell, Bloom, and Blundell.
- **Spain:** "Income Risk Inequality: Evidence from Spanish Administrative Record", by Arellano, Bonhomme, De Vera, Hospido, and Wei.
- **Sweden:** "Income Dynamics in Sweden 1985-2016," by Friedrich, Laun, and Meghir.

# Background: Data on Income Inequality

- ▶ Several harmonized cross-country databases (of statistics) on income inequality are available:
  - [World Inequality Database](#) (WID.world) spearheaded by the work of T. Atkinson, T. Piketty and E. Saez
  - [World Income Inequality Database](#) (WIID2) maintained at the United Nations University
  - [OECD Income Distribution Database](#) (IDD)
  - [Luxembourg Income Study](#) (LIS)

# Background: Data on Income Inequality

- ▶ Several harmonized cross-country databases (of statistics) on income inequality are available:
  - [World Inequality Database](#) (WID.world) spearheaded by the work of T. Atkinson, T. Piketty and E. Saez
  - [World Income Inequality Database](#) (WIID2) maintained at the United Nations University
  - [OECD Income Distribution Database](#) (IDD)
  - [Luxembourg Income Study](#) (LIS)
  
- ▶ There are also journal special issues on income inequality across countries
  - [2010 RED Special Issue](#) on “Cross-Sectional Facts for Macroeconomists” (income dynamics as well)
  - The forthcoming “[Deaton Review](#)” (IFS)

# Why This Project?

- ▶ Existing cross-country databases are typically:
  - based on [survey data](#)
  - focused on [cross-sectional](#) statistics (e.g. inequality)
  - reported for broad demographic groups (age, gender, etc.)
  
- ▶ OECD's LinkEED launching today is most similar to GID. Unfortunately, no database component for now.

# Why This Project?

- ▶ Existing cross-country databases are typically:
  - based on [survey data](#)
  - focused on [cross-sectional](#) statistics (e.g. inequality)
  - reported for broad demographic groups (age, gender, etc.)
  
- ▶ The GID Project aims to extend the focus to:
  - [Administrative](#) data, which is now available from many countries
  - To study [income dynamics](#) and other long-run phenomena
  - [Finely-defined subpopulations](#) (examples later)
  
- ▶ OECD's LinkEED launching today is most similar to GID. Unfortunately, no database component for now.

# I. Survey Data

- ▶ Household surveys have been tremendously valuable in micro and macro research for many decades.
  - For example, 4000+ papers written using the US Panel Study of Income Dynamics alone.

# I. Survey Data

- ▶ Household surveys have been tremendously valuable in micro and macro research for many decades.
  - For example, 4000+ papers written using the US Panel Study of Income Dynamics alone.
- ▶ **But survey data** also faces well-known challenges:
  - **Measurement error**; and sample **attrition** in panels
  - Lack of representativeness in the **tails of distribution**
  - Small sample size/**Statistical power**
  - **Increasing non-response**, declining response quality (National Academies of Sciences (2013), Meyer et al (2016))



# I. Survey Data

- ▶ Household surveys have been tremendously valuable in micro and macro research for many decades.
  - For example, 4000+ papers written using the US Panel Study of Income Dynamics alone.
- ▶ **But survey data** also faces well-known challenges:
  - **Measurement error**; and sample **attrition** in panels
  - Lack of representativeness in the **tails of distribution**
  - Small sample size/**Statistical power**
  - **Increasing non-response**, declining response quality (National Academies of Sciences (2013), Meyer et al (2016))
- ▶ These drawbacks matter for cross-sectional analyses *but* are much more problematic for studying dynamics.
  - Will discuss specific examples later.

# I. Administrative Data

- ▶ **Minimal** measurement error and **little or no** attrition in most countries.

# I. Administrative Data

- ▶ **Minimal** measurement error and **little or no** attrition in most countries.
- ▶ **Size:** granular analyses of sub-populations and **high-order moments**

# I. Administrative Data

- ▶ **Minimal** measurement error and **little or no** attrition in most countries.
- ▶ **Size:** granular analyses of sub-populations and **high-order moments**
  - Income changes have heavy tails: Probability of 3+ sigma shock:
    - ▶ **0.27%** under a Normal
    - ▶ **2.4%** or **9X** more likely in the data.
    - ▶ Effect on welfare or risk premium substantial.

# I. Administrative Data

- ▶ **Minimal** measurement error and **little or no** attrition in most countries.
- ▶ **Size:** granular analyses of sub-populations and **high-order moments**
  - Income changes have heavy tails: Probability of 3+ sigma shock:
    - ▶ **0.27%** under a Normal
    - ▶ **2.4%** or **9X** more likely in the data.
    - ▶ Effect on welfare or risk premium substantial.
- ▶ Ability to **link** data across various sources in some countries (income, expenditures, health, education)

# I. Administrative Data

- ▶ **Minimal** measurement error and **little or no** attrition in most countries.
- ▶ **Size:** granular analyses of sub-populations and **high-order moments**
  - Income changes have heavy tails: Probability of 3+ sigma shock:
    - ▶ **0.27%** under a Normal
    - ▶ **2.4%** or **9X** more likely in the data.
    - ▶ Effect on welfare or risk premium substantial.
- ▶ Ability to **link** data across various sources in some countries (income, expenditures, health, education)
- ▶ **Caveat:** In countries with large informal sectors, coverage limited to formal sector. More on this later.

How GID Can Help Shed Light on  
Important Substantive Questions: An Example

# Trends in Individual Income Uncertainty

Rising income uncertainty in the US became conventional wisdom.



# Trends in Individual Income Uncertainty

- ▶ **Question:** What happened to individual income uncertainty in the US since the 1980s?
- ▶ Answer crucial for many questions and policy design.

Rising income uncertainty in the US became conventional wisdom.

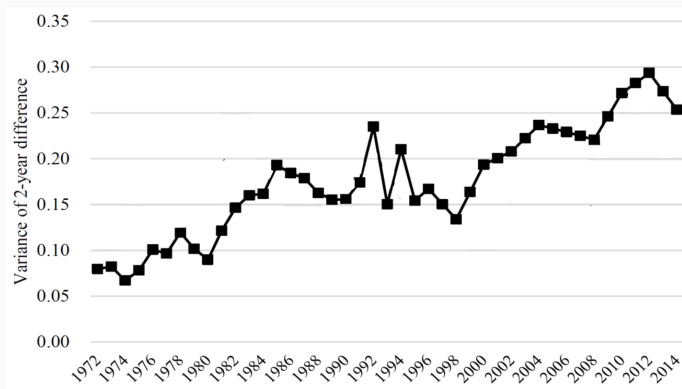
# Trends in Individual Income Uncertainty

- ▶ **Question:** What happened to individual income uncertainty in the US since the 1980s?
- ▶ Answer crucial for many questions and policy design.
- ▶ Seminal paper by Moffitt and Gottschalk (1994):
  - Income volatility **increased substantially** in PSID from 1968 to 1988.
  - **30+ papers confirmed the result** and extended to 2010s.

Rising income uncertainty in the US became conventional wisdom.

# US Income Volatility in Survey Data **UP**

**Figure 1:** Variance of Income Growth (2-year)



Moffitt and Zhang (AEA P&P, 2018)

- ▶ Variance of income growth in PSID nearly tripled from 1970s to 2010s.

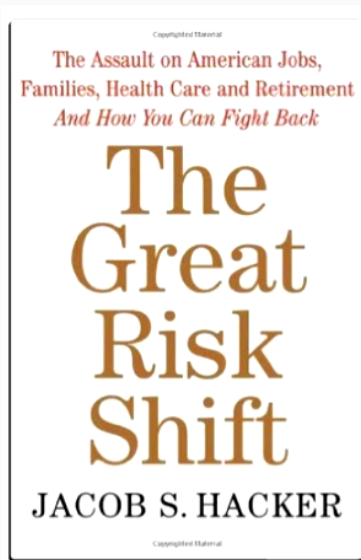
# Trends in Individual Income Uncertainty

- ▶ Question: What happened to individual income uncertainty in the US since the 1980s?
- ▶ Answer crucial for many questions and policy design.
- ▶ Seminal paper by Moffitt and Gottschalk (1994):
  - Income volatility increased substantially in PSID from 1968 to 1988.
  - 30+ papers confirmed the result and extended to 2010s.
- ▶ That income uncertainty is higher today became conventional wisdom.

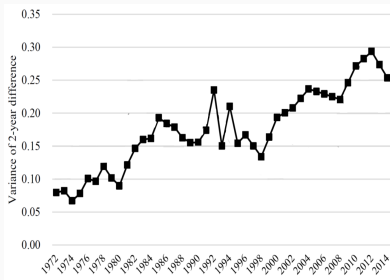
## Evidence Has Caused Some Alarm

*The volatility of family incomes has gone up — way, way up... In fact over the past generation the economic instability of American families has risen much faster than income inequality*

(Hacker 2006)



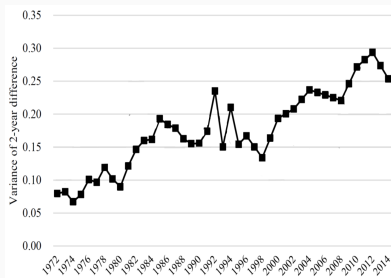
# US Income Volatility Trending UP



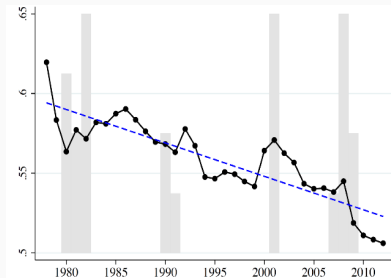
(a) Survey (PSID)

Left: Moffitt and Zhang (AEA P&P, 2018);

# US Income Volatility Trending UP? or **DOWN?**



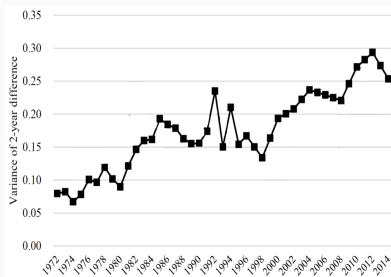
(c) Survey (PSID)



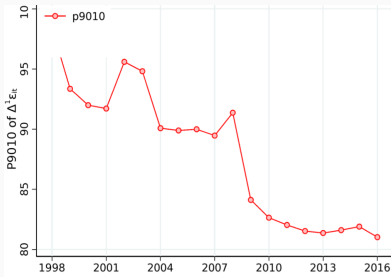
(d) Administrative (SSA Data)

Left: Moffitt and Zhang (AEA P&P, 2018); Right: Guvenen, Ozkan, Song (JPE, 2014)

# US Income Volatility Trending UP? or **DOWN?**



(e) Survey (PSID)



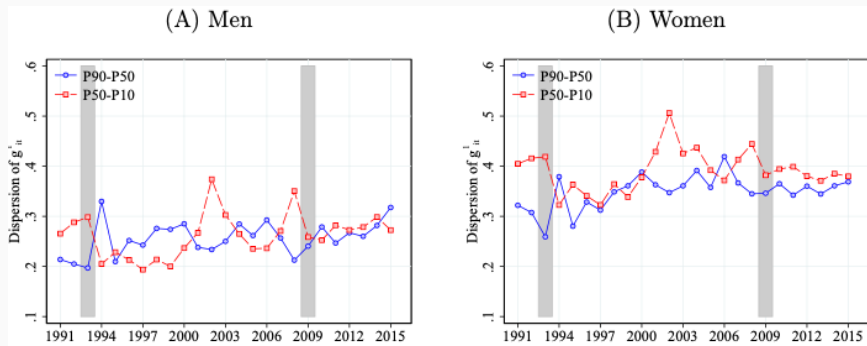
(f) Administrative (LEHD Data for GID project)

Left: Moffitt and Zhang (AEA P&P, 2018); Right: **Abowd and McKinney (2019)**



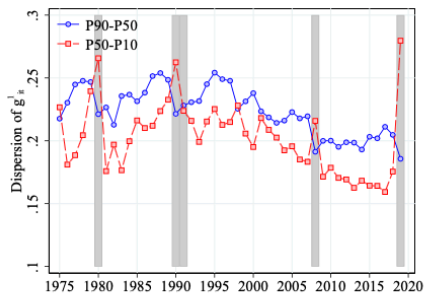
# GID Database: How About Other Countries?

Figure 2: France

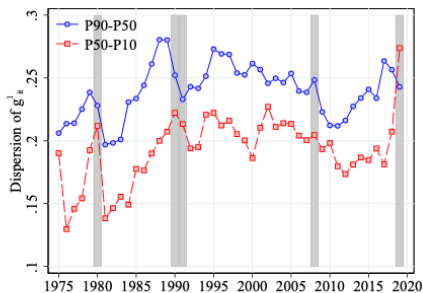


- ▶ NB: The sources for the following figures are the GID country team working papers listed above.

Figure 3: UK

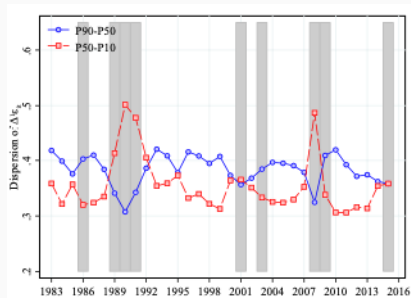


(a) Male

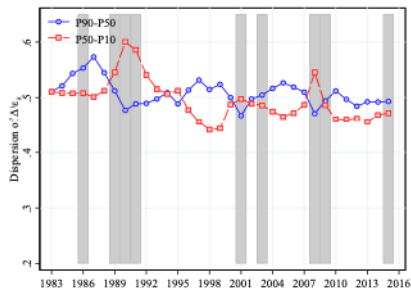


(b) Female

**Figure 4:** Canada



**(a) Men**



**(b) Women**

Figure 5: Norway

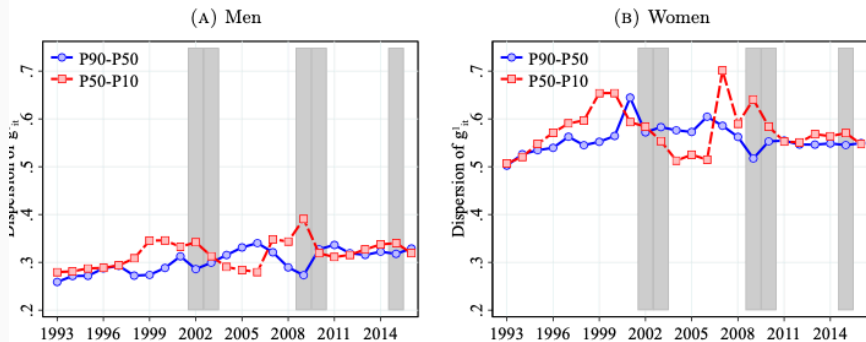
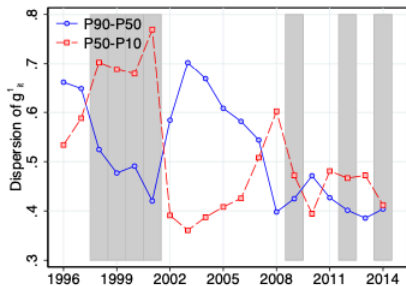
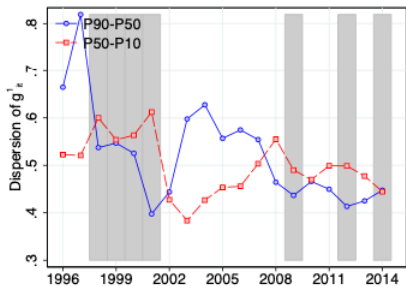


Figure 6: Argentina



(a) Men



(b) Women

Figure 7: Brazil

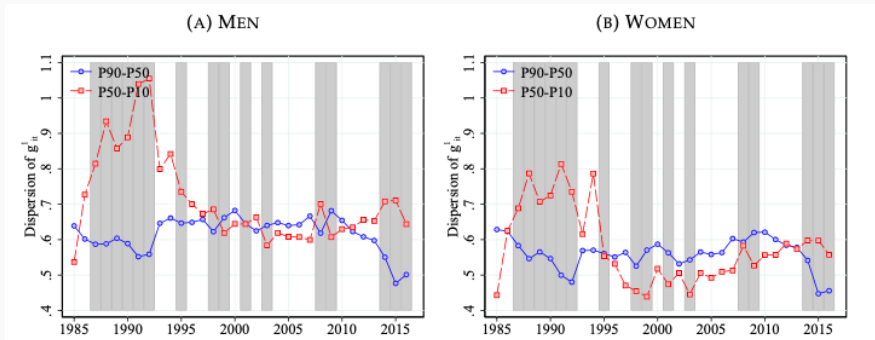
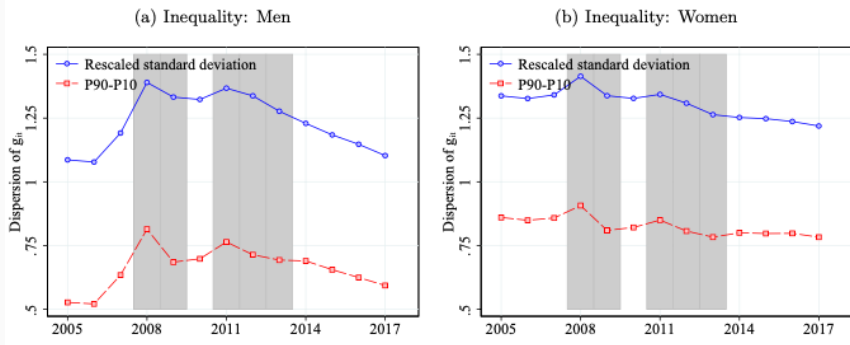


Figure 8: Spain



- ▶ Overall, **income volatility** has been
  - **declining** in several countries (Brazil, Argentina, UK after 1990)
  - **relatively flat** in some others (Canada, Norway, France), and
  - **increasing in a few** (men in Spain and women in UK pre-1990s) although it **appears more cyclical than a trend**, especially Spain.
  
- ▶ Overall, GID database contains very granular statistics on income volatility that will allow digging deeper into this question.
  - e.g., P90-P50, P50-P10, and other pctlile differences, of one- and five-year income change, conditional on permanent income, age, etc.



# Taking Stock

- ▶ Overall, **income volatility** has been
  - **declining** in several countries (Brazil, Argentina, UK after 1990)
  - **relatively flat** in some others (Canada, Norway, France), and
  - **increasing in a few** (men in Spain and women in UK pre-1990s) although it **appears more cyclical than a trend**, especially Spain.
- ▶ This first look does **not reveal any broad-based increase in income uncertainty** as suggested by US survey data.
- ▶ Overall, GID database contains very granular statistics on income volatility that will allow digging deeper into this question.
  - e.g., P90-P50, P50-P10, and other pctlile differences, of one- and five-year income change, conditional on permanent income, age, etc.

# Back to the GID Database: Longitudinal Dimension

- ▶ Cross-sectional data (inequality):
  - Provides **static snapshots** of a distribution, but **misses** longitudinal **dynamics**

# Back to the GID Database: Longitudinal Dimension

- ▶ Cross-sectional data (inequality):
  - Provides **static snapshots** of a distribution, but **misses** longitudinal **dynamics**
- ▶ Panel data:
  - Allows study of several dimensions important for welfare:
    - ▶ **Income Dynamics** (volatility, skewness, persistence, etc.)
    - ▶ **Mobility** (intra- and inter-generationally)
    - ▶ **Tail behavior** (probability of extreme wage cuts vs. wage hikes)
    - ▶ Inequality in long-run (“permanent”) income

### III. Heterogeneity Dimension

- ▶ Income risk, inequality, and mobility patterns vary significantly:
  - by cohort, gender, education, employer, permanent income, geographical location, etc.

### III. Heterogeneity Dimension

- ▶ Income risk, inequality, and mobility patterns vary significantly:
  - by cohort, gender, education, employer, permanent income, geographical location, etc.
- ▶ In GID Version 1.0, a large number of statistics are reported for groups defined by
  - the interaction of **gender** x **cohort** x **age** group x **“permanent” income**.

### III. Heterogeneity Dimension

- ▶ Income risk, inequality, and mobility patterns vary significantly:
  - by cohort, gender, education, employer, permanent income, geographical location, etc.
- ▶ In GID Version 1.0, a large number of statistics are reported for groups defined by
  - the interaction of **gender** x **cohort** x **age** group x **“permanent” income**.
  - **Permanent income** is based on past 3-year average income. **40 quantile bins** plus more at the top.
    - ▶ This variable **captures substantial heterogeneity** conditional on all other observables.

### III. Heterogeneity Dimension

- ▶ Income risk, inequality, and mobility patterns vary significantly:
  - by cohort, gender, education, employer, permanent income, geographical location, etc.
- ▶ In GID Version 1.0, a large number of statistics are reported for groups defined by
  - the interaction of **gender** x **cohort** x **age** group x **“permanent” income**.
  - **Permanent income** is based on past 3-year average income. **40 quantile bins** plus more at the top.
    - ▶ This variable **captures substantial heterogeneity** conditional on all other observables.
  - For a typical country, **4,000+ distinct moments** for a single statistic (e.g., volatility of 1-year log income change).

# Detailed Statistics

- ▶ [List of statistics](#) drawn from a broad set of recent empirical and theoretical papers.
- ▶ List has been revised and expanded several times since we started.



# Detailed Statistics

- ▶ **List of statistics** drawn from a broad set of recent empirical and theoretical papers.
- ▶ List has been revised and expanded several times since we started.
- ▶ Emphasis on providing flexible quantile-based statistics to allow computation of alternative measures:
  - Volatility: Standard deviation, P90-P10, but also P90-P50, P99-P90, P95-P5, etc.
  - Skewness: 3rd centralized moment, Kelley's skewness, etc.

# Detailed Statistics

- ▶ **List of statistics** drawn from a broad set of recent empirical and theoretical papers.
- ▶ List has been revised and expanded several times since we started.
- ▶ Emphasis on providing flexible quantile-based statistics to allow computation of alternative measures:
  - Volatility: Standard deviation, P90-P10, but also P90-P50, P99-P90, P95-P5, etc.
  - Skewness: 3rd centralized moment, Kelley's skewness, etc.
- ▶ Statistics calculated for 3 different income measures: **raw income**, **residualized income**, **permanent income**.

# Overview of Available Statistics

## ▶ Income levels (static):

- **Time-series statistics** (up to 40 years for some countries)
- Measures of inequality for annual and “permanent” income
- Empirical densities
- and others

# Overview of Available Statistics

## ▶ Income levels (static):

- Time-series statistics (up to 40 years for some countries)
- Measures of inequality for annual and “permanent” income
- Empirical densities
- and others

## ▶ Income changes (dynamics):

- 1-year and 5-year income changes (transitory and persistent changes)
- Log change and arc-percent change (to allow extensive margin—zero income)
- Moments on dispersion, skewness, and kurtosis (tails)
- Short- and Long-run mobility statistics
- For very finely defined groups

# Overview of Available Statistics

## ▶ Income levels (static):

- Time-series statistics (up to 40 years for some countries)
- Measures of inequality for annual and “permanent” income
- Empirical densities
- and others

## ▶ Income changes (dynamics):

- 1-year and 5-year income changes (transitory and persistent changes)
- Log change and arc-percent change (to allow extensive margin—zero income)
- Moments on dispersion, skewness, and kurtosis (tails)
- Short- and Long-run mobility statistics
- For very finely defined groups

## ▶ We have additional conditioning variables (education, occupation)

- Will be added after initial launch (when we can afford it)

# Harmonization

- ▶ Harmonization has been a key focus from the beginning.

# Harmonization

- ▶ Harmonization has been a key focus from the beginning.
- ▶ On the data side:

# Harmonization

- ▶ Harmonization has been a key focus from the beginning.
- ▶ On the data side:
  - Comparable **income definitions**: before-tax, annual, wage/salary income (exclude self employment income), including overtime, bonuses, vacation pay, and other items.
    - ▶ In a few countries (UK and Argentina) income data is recorded monthly and converted to annual



# Harmonization

- ▶ Harmonization has been a key focus from the beginning.
- ▶ On the data side:
  - Comparable **income definitions**: before-tax, annual, wage/salary income (exclude self employment income), including overtime, bonuses, vacation pay, and other items.
    - ▶ In a few countries (UK and Argentina) income data is recorded monthly and converted to annual
  - **Same sample selection** criteria
  - **Similar time periods**: For heterogeneity statistics, all countries use 20 years of data, going back from last available year.

# Harmonization

- ▶ Harmonization has been a key focus from the beginning.
- ▶ On the data side:
  - Comparable **income definitions**: before-tax, annual, wage/salary income (exclude self employment income), including overtime, bonuses, vacation pay, and other items.
    - ▶ In a few countries (UK and Argentina) income data is recorded monthly and converted to annual
  - **Same sample selection** criteria
  - **Similar time periods**: For heterogeneity statistics, all countries use 20 years of data, going back from last available year.
- ▶ But most important:
  - **Single master code to generate all statistics** from very first step of cleaning raw data to the very end results.
  - Master code was written by two team members, **Serdar Ozkan** and **Sergio Salgado**, so a big shout-out to both!

# Informal Sector

- ▶ In countries with large informal sectors, administrative data misses many workers.
- ▶ Relevant for Mexico, Argentina, and Brazil. But as GID coverage broadens, it will matter for more countries.

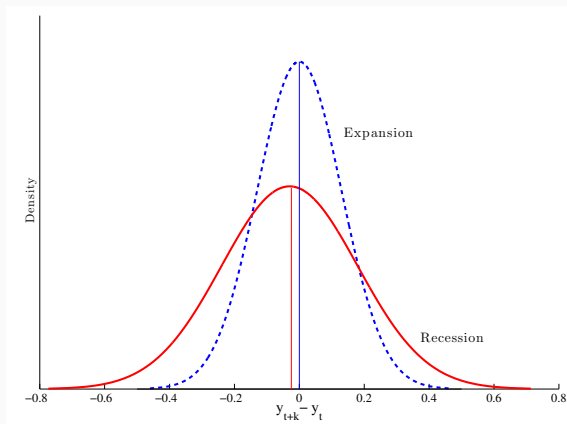
# Informal Sector

- ▶ In countries with large informal sectors, administrative data misses many workers.
- ▶ Relevant for Mexico, Argentina, and Brazil. But as GID coverage broadens, it will matter for more countries.
- ▶ Teams for these 3 countries were asked to do a validation study with other datasets available for each country.
- ▶ These issues are discussed in the QE papers providing guidance about how to approach the use of reported statistics.
- ▶ This is a challenging issue that we will continue to work on in future iterations.

Insights from GID: Another Example  
What Happens to Individual Income Risk in  
Recessions?

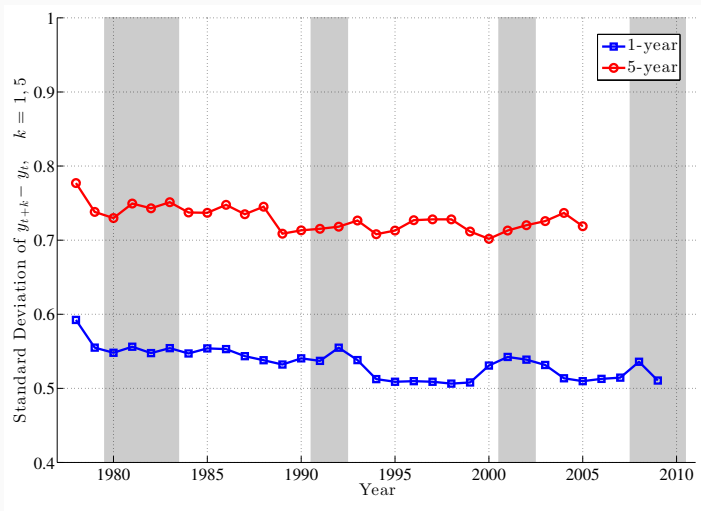
# How Does Income Risk Change over the Business Cycle?

## Hypothesis #1: Countercyclical Variance of Income Shocks

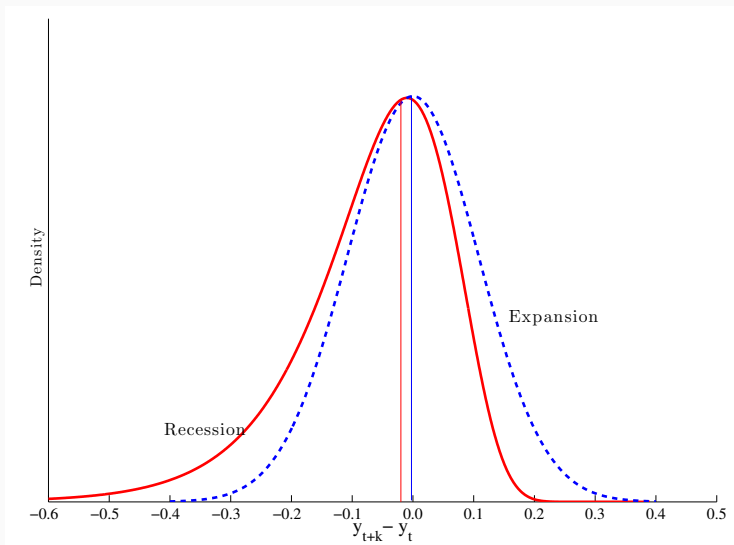


- ▶ Parametric estimates from US survey data found a large rise in variance in recessions.

# Plotting Standard Deviation of Income Change Shows No Rise in Recessions

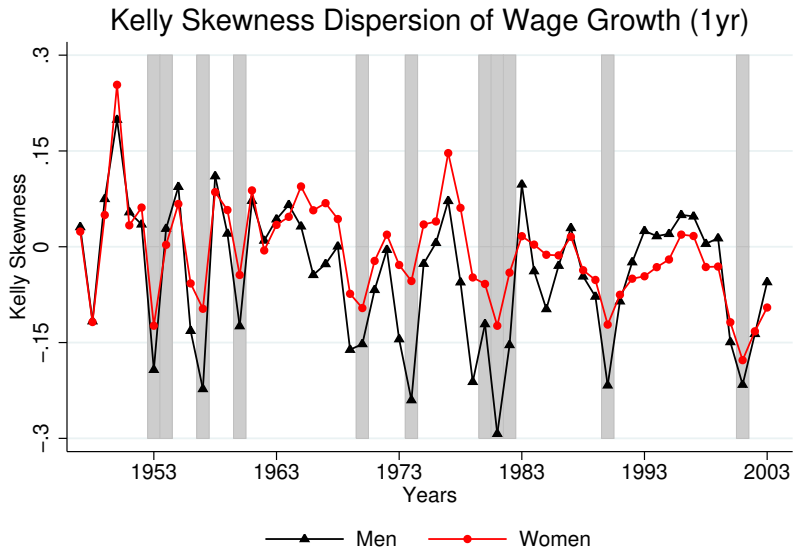


## Hypothesis #2: Procyclical Skewness

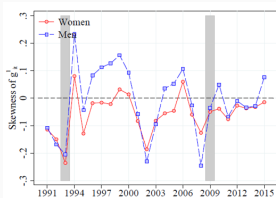




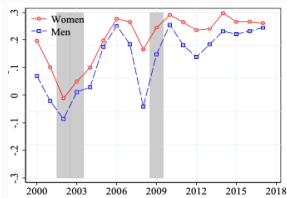
# US SSA Data: Skewness of income changes procyclical



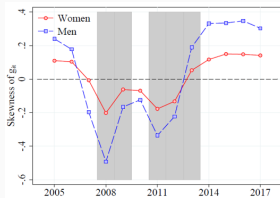
# Is Skewness Procyclical in other Countries Too? –Yes



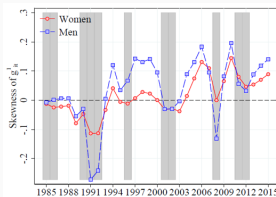
(a) France



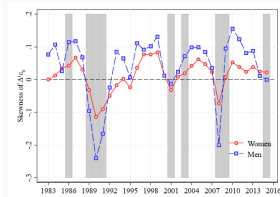
(b) Germany



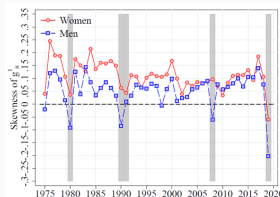
(c) Spain



(d) Sweden

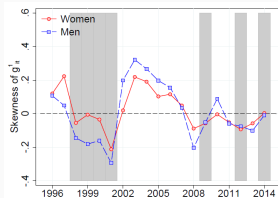


(e) Canada

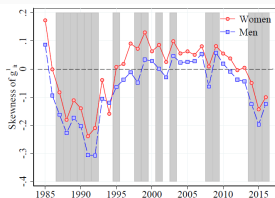


(f) UK

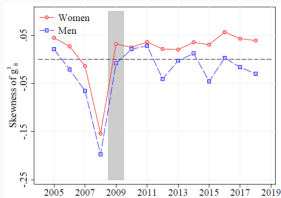
# Is Skewness Procyclical in other Countries Too? –Yes



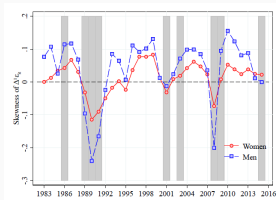
(g) Argentina



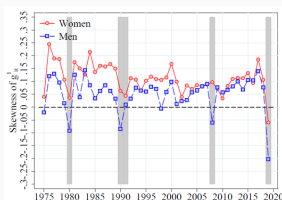
(h) Brazil



(i) Mexico



(j) Canada



(k) UK

# Conclusions

- ▶ The GID project is a database of statistics on income inequality and income dynamics.
- ▶ We plan (hope?) to expand it to other variables, including those related to firm-worker connection.

# Conclusions

- ▶ The GID project is a database of statistics on income inequality and income dynamics.
- ▶ We plan (hope?) to expand it to other variables, including those related to firm-worker connection.
- ▶ The OECD volume released today is a seminal contribution on the firm-worker links and sets an example for us and others that follow.

# Conclusions

- ▶ The GID project is a database of statistics on income inequality and income dynamics.
- ▶ We plan (hope?) to expand it to other variables, including those related to firm-worker connection.
- ▶ The OECD volume released today is a seminal contribution on the firm-worker links and sets an example for us and others that follow.
- ▶ Using administrative data in countries with high informal employment is still a challenge and our future efforts will focus on the best way to address it.

# Conclusions

- ▶ The GID project is a database of statistics on income inequality and income dynamics.
- ▶ We plan (hope?) to expand it to other variables, including those related to firm-worker connection.
- ▶ The OECD volume released today is a seminal contribution on the firm-worker links and sets an example for us and others that follow.
- ▶ Using administrative data in countries with high informal employment is still a challenge and our future efforts will focus on the best way to address it.
- ▶ We welcome all feedback and suggestions, especially when the database goes live.

All GID Project papers are posted online here:

<https://mebdi.org/global-income-dynamics-project/>

Thanks!