ECON0024 Economic Policy Analysis

Lecture Notes for Academic Year 22/23

Xiaotian Tian, Kuangjie Ni, Kaicheng Lu, Haruka Shuei

April 27, 2023

Contents

Introduction

These are our unofficial lecture notes for ECON0024 Economic Policy Analysis. Contents are created according to lectures and slides, as well as Matt & Emma's tutorials. Please contact me (xiaotian.tian.20@ucl.ac.uk) should you have any questions or suggestions.

Notations:

- ⋆ indicates chapters that are "model-intensive"
- Red highlights usually indicate important conclusions and results
- Blue highlights usually indicate important concepts

Remarks:

- Some contents are subject to our personal interpretations.
- This is an open-source project. Contributions are welcomed!
  Github repository: https://github.com/Er1kKa-Tian/ECON0024-Notes

Chapter 1

Unemployment Insurance

1.1 Unemployment

1.1.1 Definition of Unemployment

The International Labour Organisation (ILO)'s definition of unemployment: The number of jobless people who want to work, are available to work, and are actively seeking employment

1.1.2 Why we care about unemployment?

1. Unemployment is a sign of inefficiency.
2. Modern view: part of the unemployment arises from search costs (or frictions).
   - Zero unemployment is sub-optimal, because it indicates intensive job searching that induces high search costs.
   - Therefore, some level of unemployment (usually 3% - 6%) are tolerated in reality
3. Being unemployed decreases life satisfaction.

1.2 Introduction to Unemployment Insurance (UI)

1.2.1 Introduction

Many countries have some kind of UI, but it still remains controversial due to its benefit-and-cost tradeoff:

- Main benefit: helps people in a time of need
- Main cost: reduces incentive to search for work while unemployed

The following section will analyse how to design unemployment insurance given this trade-off.

1.2.2 Brief History of Unemployment Insurance

- The first unemployment benefit scheme was introduced in the UK with the *National Insurance Act* in 1911.
- U.S. introduced UI in 1935 in response to the Great Depression.
- Most European countries introduced UI after the WWII during the expansion of welfare state.
- Today, most developed countries have some UI schemes.

1.2.3 Institutional Details

Typically, UI is financed through a payroll tax on employers and/or employees, but the incidence and rates vary a lot across countries. In exchange of the paid in contributions, unemployment benefit are received upon job-loss.

- Usually, the eligibility for UI must be a result of a layoff.
- The length and the level of UI benefits often depend on the amount and time length of contribution.
- Minimum contribution period and minimum contribution amount are often defined.
- The relationship between contribution and benefit-level is often highly non-linear (benefit caps, minimum benefits).

1.2.4 Net Benefit Replacement Rate

We define the Net Benefit Replacement Rate, an important feature of UI, as:

$$\text{Net Benefit Replacement Rate} = \frac{\text{Weekly Benefit}}{\text{Weekly Wage Earnings}}$$

1.3 ⋆ Optimal Unemployment Insurance with No Moral Hazard

1.3.1 Expected Utility Model of Unemployment

A typical objective considered by economists is to maximize agents' welfare given by expected utility.

Specifically, we can calculate the expected utility by:

$$E[U] = (1-p) \times U(c^e) + p \times U(c^u) - \psi(1-p)$$

where

- p is the probability of being *unemployed* ("unemployment rate")
- $1-p$ is the probability of being *employed*
- c^u is the consumption when being *unemployed*
- c^e is the consumption when being *employed*
- $U(.)$ is the utility function: we assume it to be *strictly increasing and concave* i.e. $U'(.) > 0$ and $U''(.) < 0$
- $\psi(.)$ is the job searching cost function: we assume it to be *strictly increasing and convex* i.e. $\psi'(.) > 0$ and $\psi''(.) > 0$ (higher probability of employment $\implies$ higher searching cost)

Assumptions on the individual side

Assume no saving/borrowing and homogeneous agents, we can write the individuals' budget constraints:

- Consumption at employment equals to after-tax wage i.e. $c^e = w - t$
- Consumption at unemployment equals to UI benefit i.e. $c^u = b$

With this assumption, we can express the expected utility as:

$$E[U] = (1-p) \times U(w-t) + p \times U(b) - \psi(1-p) \tag{1.1}$$

Assumptions on the government side:

Assume that the government must have a balanced budget: total tax collected must equal to total benefit given. With this assumption, we can write the government's budget constraint:

$$(1-p) \times t = p \times b$$

Rearranging this equation, it indicates that taxes need to be:

$$t = \frac{p}{1-p} \times b$$

plug these results into the expected utility of individuals (equation 1.1):

$$E[U] = (1-p) \times U\left(w - \frac{p}{1-p} \times b\right) + p \times U(b) - \psi(1-p)$$

Therefore, we can express the government's optimisation problem as:

$$\max_b E[U] = (1-p) \times U\left(w - \frac{p}{1-p} \times b\right) + p \times U(b) - \psi(1-p) \tag{1.2}$$

Further assumption: No moral hazard

Assume that p, the probability of finding a job, does not depend on the benefit level b, i.e. no moral hazard. We can treat p as an exogenous variable here.

Reminder: here, moral hazard refers to the adverse actions taken by insured individuals in response to insurance against adverse outcomes. It exists as long as insurers cannot perfectly monitor insurees.

1.3.2 Optimisation Result: Full Insurance

To find the optimal UI, we solve the optimisation problem (1.2). The FOC is:

$$\frac{\partial E[U]}{\partial b} = -(1-p) \times \frac{p}{1-p} \times U'\left(w - \frac{p}{1-p} \times b\right) + p \times U'(b) = 0$$

$$U'(b) = U'\left(w - \frac{p}{1-p} \times b\right)$$

$$b = w - \frac{p}{1-p} \times b$$

This result implies that

$$c^e = w - t = b = c^u$$

which indicates full insurance – people have the same income when unemployed as they were employed. In practice, this would mean that the net replacement rate = 1. (No moral hazard $\rightsquigarrow$ Full insurance)

1.4 ⋆ Optimal Unemployment Insurance with Moral Hazard

1.4.1 Moral Hazard

The problem with full insurance is that it eliminates incentives to work. To incorporate moral hazard into our framework, we assume that p increases with b: more generous benefits deter job search and hence increase unemployment.

$$p = p(b), \quad \frac{\partial p(b)}{\partial b} > 0$$

1.4.2 Individuals' Response

Now, individuals choose their probability of being unemployed p given the unemployment benefit b:

$$\max_p E[U] = (1-p) \times U(c^e) + p \times U(c^u) - \psi(1-p)$$

The FOC of this problem is:

$$\frac{\partial E[U]}{\partial p} = -U(c^e) + E(c^u) + \psi'(1-p) = 0$$

Manipulate:

$$Pr(employed) = 1 - p = \psi'^{-1}\Big(U(c^e) - U(c^u)\Big)$$

Note that c^e, c^u are also functions of b:

$$Pr(employed) = 1 - p(b) = \psi'^{-1}\Big(U(w - \frac{p}{1-p} \times b) - U(b)\Big) \tag{1.3}$$

Now, we can investigate the effect of changing b on the probability of being employed (using $\frac{\partial f^{-1}(x)}{\partial x} = \frac{1}{f'(x)}$):

$$\frac{\partial Pr(employed)}{\partial b} = \frac{\partial(1-p(b))}{\partial b} = \frac{-\frac{p}{1-p} \times U'\big(w - \frac{p}{1-p} \times b\big) - U'(b)}{\psi''\Big(U\big(w - \frac{p}{1-p} \times b\big) - U(b)\Big)} < 0$$

because $U'(.) > 0$ and $\psi''(.) > 0$ by assumption.

1.4.3 The government's optimisation: Partial Insurance

Our model now becomes:

$$\max_b E[U] = \big(1 - p(b)\big) \times U\left(w - \frac{p(b)}{1-p(b)} \times b\right) + p(b) \times U(b) - \psi\big(1 - p(b)\big)$$

The FOC becomes more complicated:

$$0 = -U'\left(w - \frac{p(b)}{1-p(b)} \times b\right) \times \left(\frac{p'(b) \times b}{1-p(b)} + p(b)\right) + p(b)U'(b) + p'(b) \times \underbrace{\big[-U(c^e) + U(c^u) + \psi'(1-p(b))\big]}_{\text{From the FOC of individuals, we know this is 0}}$$

$$0 = -U'\left(\underbrace{w - \frac{p(b)}{1-p(b)} \times b}_{c^e}\right) \times \left(\frac{p'(b) \times b}{1-p(b)} + p(b)\right) + p(b)U'(\underbrace{b}_{c^u})$$

Rearrange this, we can get the Main Equation of Optimal UI:

$$\underbrace{\frac{U'(c^u) - U'(c^e)}{U'(c^e)}}_{\text{Insurance Value}} = \underbrace{\frac{1}{1-p} \times \epsilon_{p,b}}_{\text{Moral Hazard Cost}} \tag{1.4}$$

where $\epsilon_{p,b} = \frac{b}{p}\frac{dp}{db}$ is the elasticity of unemployment rate with respect to benefits.

Note that the Main Equation of Optimal UI implies: if $\epsilon_{p,b} > 0$, then $0 < c^u < c^e < w$ which means partial insurance will be the optimum. (Moral hazard $\rightsquigarrow$ Partial insurance)

1.4.4 Explaining the Main Equation of Optimal UI

As in a typical optimum, the marginal benefit (insurance value) has to be equal to the marginal cost (moral hazard cost). If the marginal benefit (insurance value) is larger than the marginal cost (moral hazard cost), we need to increase the UI benefit to reach optimum, v.v.

1. **Insurance Value**
   This is the marginal benefit of UI benefit (redistributing one unit from the employed to the unemployed): if we increase the UI benefit by 1 unit, the net benefit will be the insurance value:

$$\text{Insurance Value} = \frac{U'(c^u) - U'(c^e)}{U'(c^e)}$$

2. **Moral Hazard Cost**
   This is the marginal cost of UI benefit (redistributing one unit from the employed to the unemployed): if we want to increase the UI benefit by 1 unit, we need to collect an extra of: $p(b) + p'(b) \times b$. This has to be paid by people who remain employed, so the tax rate has to increase by: $\frac{p(b)+p'(b)\times b}{1-p(b)}$. Rearrange:

$$\text{Moral Hazard Cost} = \frac{p(b)}{1-p(b)} \times \left(1 + p'(b) \times \frac{b}{p(b)}\right) = \frac{p(b)}{1-p(b)} \times (1 + \epsilon_{p,b})$$

   This can be understood as the extra cost imposed on an employed individual in order to redistribute 1 unit to every unemployed individual. The more responsive the unemployment rate is (higher $\epsilon_{p,b}$), the more costly of an increase in UI will be.

1.4.5 Sufficient Statistic Approach

Approximate the Main Equation of Optimal UI

The term $\frac{U'(c^u)-U'(c^e)}{U'(c^e)}$ cannot be estimated because people's utility functions are not observed. Instead, we employ the 1st order Taylor Expansion. Expand $U'(c^u)$ around c^e:

$$U'(c^u) \approx U'(c^e) + U''(c^e)(c^u - c^e)$$

Therefore:

$$\begin{aligned}
\frac{U'(c^u) - U'(c^e)}{U'(c^e)} &\approx \frac{U'(c^e) + U''(c^e)(c^u - c^e) - U'(c^e)}{U'(c^e)} \\
&= \frac{U''(c^e)(c^u - c^e)}{U'(c^e)} \\
&= \underbrace{-\frac{c^e U''(c^e)}{U'(c^e)}}_{\text{Relative Risk Aversion } \gamma} \times \frac{c^e - c^u}{c^e} \\
&= \gamma \times \frac{c^e - c^u}{c^e} \\
&= \gamma \times \frac{\Delta c}{c^e}
\end{aligned}$$

Substituting this into the main equation (1.4):

$$\underbrace{\frac{1}{1-p} \times \epsilon_{p,b}}_{\text{Moral Hazard Cost}} \approx \underbrace{\gamma \times \frac{\Delta c}{c^e}}_{\text{Insurance Value}} \tag{1.5}$$

This equation (1.5) is a transformation of the Main Equation of Optimal UI, which is estimable in practice:

- $\frac{\Delta c}{c^e}$ is the consumption drop at unemployment
- γ is the coefficient of relative risk aversion (higher RRA indicates more risk aversion)
- $\epsilon_{p,b} = \frac{b}{p}\frac{dp}{db}$ is the elasticity of unemployment rate with respect to benefits

We call this the Sufficient Statistic Approach because the optimal UI can be calculated aftering measuring the above three factors.

Pros and Cons of the Sufficient Statistic Approach

The alternative method of determining the optimal UI benefit is known as the structural labour approach:

1. Estimate the key parameters of the utility function, job search function, etc.
2. Simulate counterfactual results with alternative levels of the UI benefit

The first procedure often requires functional assumptions on the utility function and the cost of job search. Compared with this, the Sufficient Statistic Approach focus on some key statistics without identifying the whole model of job search. Thus, it has obvious pros and cons:

- **Key advantage**: easier to implement and less assumptions are needed
- **Key disadvantage**: the derivation of the Main Equation of Optimal UI depends critically on the *envelop theorem* which only holds for small policy deviations (and we used the 1st order Taylor Expansion, which is a local approximation), so it cannot be used to assess the impact of radical changes

1.5 Estimation of Moral Hazard ($\epsilon_{p,b}$)

Moral hazard in UI is thought to manifest itself in the duration of unemployment: economists investigate whether unemployed individuals find jobs more slowly when UI benefits are higher. The key challenge is to use quasi-experiments to identify these effects.

1.5.1 Difference-in-Difference

The idea is to exploit changes in UI laws that affect a *treatment* group and compare that with a *control* group.

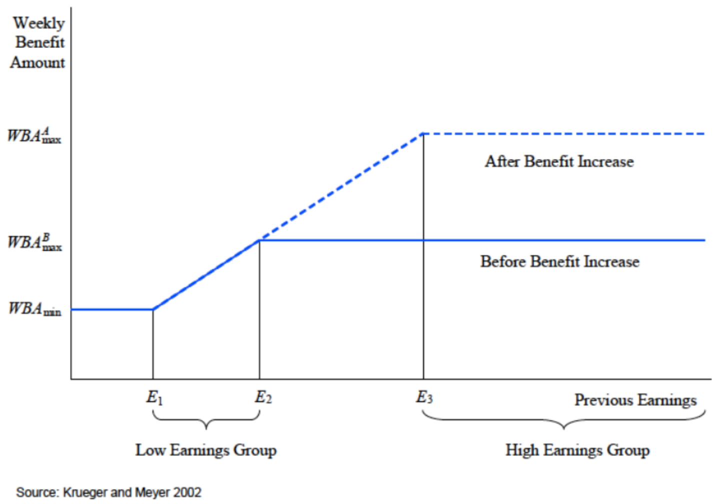


Figure 1.1: Example of DiD: low earnings group as control and high earnings group as treatment

1. Standard DiD Specification with 2 states and 2 time period

   $$log(Unemployment\ Duration)_{ist} = \beta_0 + \beta_1 After_{ist} + \beta_2 Treat_{ist} + \beta_3 Treat_{ist} \times After_{ist} + \epsilon_{ist}$$

   where:

   - $log(Unemployment\ Duration)_{ist}$ is the log unemployment duration of individual i in state s at time t
   - $After_{ist}$ is the time dummy
   - $Treat_{ist}$ is the treatment dummy

   Under the Common/Parallel Trend Assumption (the difference in unemployment duration would have been the same between the treatment/control group in absence of the policy change), β_3 identifies the casual effect (ATT, specifically).

2. Fixed Effects Version

   $$log(Unemployment\ Duration)_{ist} = \alpha_s + \theta_t + \beta log(b_{ist}) + \gamma X_{ist} + \epsilon_{ist}$$

   where:

   - α_s is the state fixed effect
   - θ_t is the time fixed effect
   - X_{ist} includes other control variables

   β identifies the relationship between percentage change in replacement rate and percentage change in unemployment duration under FE/FD assumptions. (Correct specification; Random sampling; Time variations and no perfect multicolinearity in regressors; Zero conditional mean)

Meyer (1990) and many others implemented this method using data on unemployment duration in the U.S. and state-level reforms. The general finding shows a benefit elasticity of 0.4-0.6: 10% rise in unemployment benefits leads to about a 4-6% increase in unemployment duration.

1.5.2 Regression Discontinuity Designs

This empirical strategy exploits jumps in UI benefits at a threshold. ***Main assumption***: Continuity: individuals who are slightly below the threshold and individuals who are slightly above the threshold would have similar unemployment duration in absence of the jump in the UI benefit level at the threshold.

This assumption will not hold if:

- There are *selection/manipulation* on one side of the threshold. For example, the unemployed somehow achieve to lay-off a few days to deliberately reach the threshold.
- There are other policy changes that affect outcomes at the threshold.

Example:
Card-Chetty-Weber (2007) used the fact that in Austria, you get up to 30 weeks of benefits when you have been employed for 36+ months in last 5 years (instead of up to 20 weeks).

Figure 1.2: Jump in UI Benefit

They checked: (1) there is no selection around the threshold (on frequency of layoffs, age, wages, etc.) (2) there is no other policy change at the threshold.

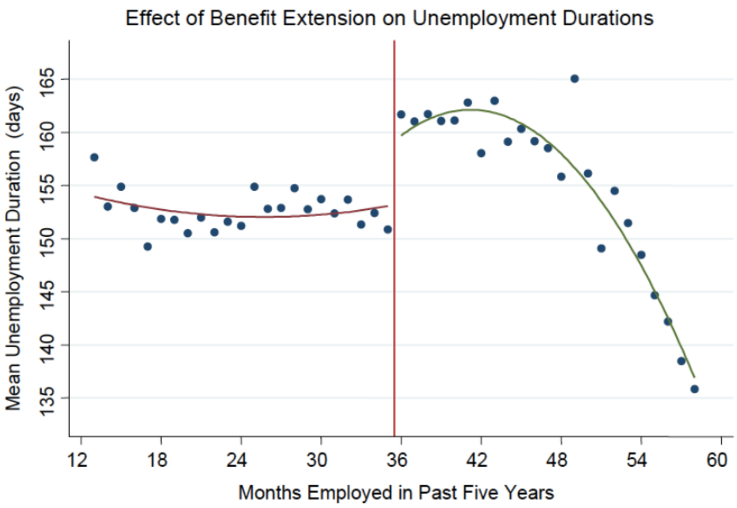


Figure 1.3: RDD Result

They estimated a benefit elasticity $\epsilon_{p,b} \approx 0.3$.

1.5.3 Regression Kink Designs

This strategy is very similar to RDD, but we exploit a kink in UI benefits instead of a jump. ***Main assumption***: In the absence of the kink in the benefit-level, there would be no kink in unemployment duration.

Example:
Kolsrud, Landais, Nilsson and Spinnewijn (2016) studied a kinked UI benefit scheme in Sweden from 1999 to 2000:

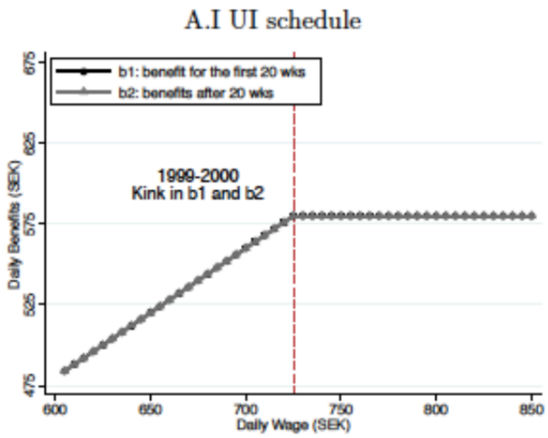


Figure 1.4: Kinked UI Benefits

They found a kink in unemployment duration, indicating the existence of moral hazard:

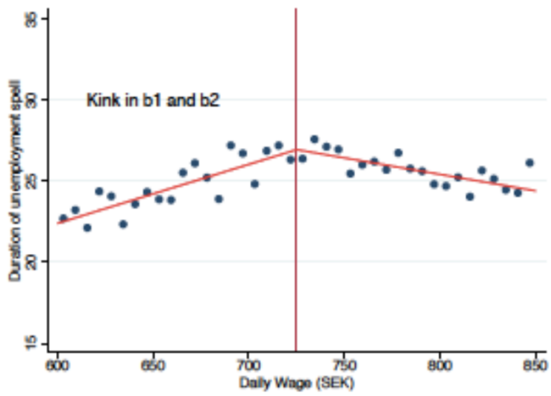


Figure 1.5: Kinked Unemployment Duration

In this case, the assumption can be easily checked: after 2000, the kinked UI benefits was abandoned, and the kink in unemployment duration disappeared:

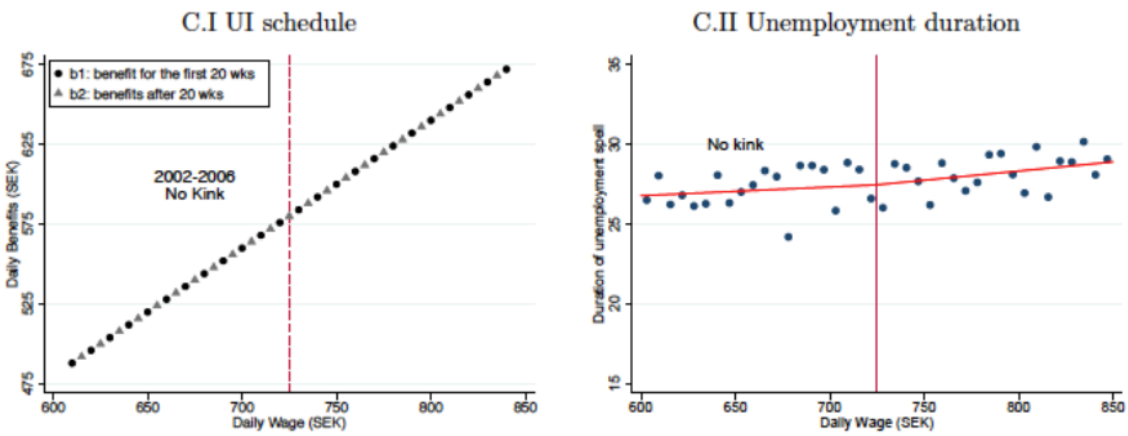


Figure 1.6: Check the Assumption

1.5.4 Extension: Decompose the Effect Using the Slutsky Equation

Decomposition

Similar to the Slutsky Equation, we can decompose the overall effect of changes in UI benefit on job research of the unemployed into a liquidity effect and a moral hazard effect (discussed by Card, Chetty, and Webers (2007)):

$$\underbrace{\frac{\partial s_t}{\partial b_t}}_{\text{Overall Effect}} = \underbrace{\frac{\partial s_t}{\partial A_t}}_{\text{Liquidity Effect}} - \underbrace{\frac{\partial s_t}{\partial w_t}}_{\text{Moral Hazard Effect}}$$

where:

- Liquidity Effect shows how search would change if cash was given to the agent independently of being unemployed or not – giving an extra dollar of UI benefits increases the wealth of the receiver, making them feel "richer."
- Moral Hazard Effect shows the distortion effect of UI benefit – giving an extra dollar of UI benefit distorts the "price" of being unemployed compared with being employed, i.e. the receiver may have lower incentives to look for a job.

Estimation Using Two RDD

Card et. al. (2007), estimated the overall effect and liquidity effect, then calculated the moral hazard effect using the above decomposition.

- Estimating Overall Effect: RDD exploiting the fact that people with less than 36 months of employment in the past 5 years receive 20 rather than 30 weeks of UI. Compare non-employment duration of those just below the 36 months threshold and those just above it.
- Estimating Liquidity Effect: RDD exploiting the fact that firms are forced to pay a lump-sum severance payment equal to 2 months of previous salary to workers who are laid off after 3 years of service. Compare workers who are just below the 3 year cutoff with those just above, and look at the difference in their non-employment duration.

1.6 Evidence on Consumption Smoothing ($\frac{\Delta c}{c^e}$)

The evidence on consumption smoothing is more limited. The key constraint is data: while we have millions of observations of unemployment duration (administrative records), data on consumption is often restricted to surveys, which contain much less samples. (Typically, we need a large sample size to use RDD/RKD.)

1.6.1 Difference-in-Difference

The settings are similar to the DiD strategy mentioned before. Gruber (1997) used food consumption as the dependent variable, and used cross-state and time variations. The dataset used is PSID food consumption.

Gruber identified the following equation:

$$\frac{\Delta c}{c^e} = \beta_0 + \beta_1 \frac{b}{w_i}$$

Their results show that: $\beta_0 \approx 0.24$ and $\beta_1 \approx -0.28$, which means a 10% increase in UI benefits will induce a 2.8% reduction in the consumption drop. Without UI benefits, unemployment will cause a 24% drop in consumption. With the current level of UI benefit (net benefit replacement rate $\frac{b}{w_i} = 0.5$), the consumption is estimated to drop about 10%.

This provides convincing evidence that insurance markets are not perfect and UI does play a consumption smoothing role.

Note that the drop in consumption to the drop in UI benefit is much less than 1-1. This is because UI also affects people's Self Insurance behaviour:

- Saving behaviours

- ◇ Spousal labor supply
- ◇ Borrowing from friends

Provision of UI benefits crowds out such self insurance to some extent.

1.6.2 Using Data of Bank Accounts

Ganong (2016) uses JPMorgan Chase Institute's (JPMCI) data on anonymized data on 210,000 checking accounts that received a direct deposit of unemployment insurance (UI) benefits. This overcomes the small sample problem present in survey data.

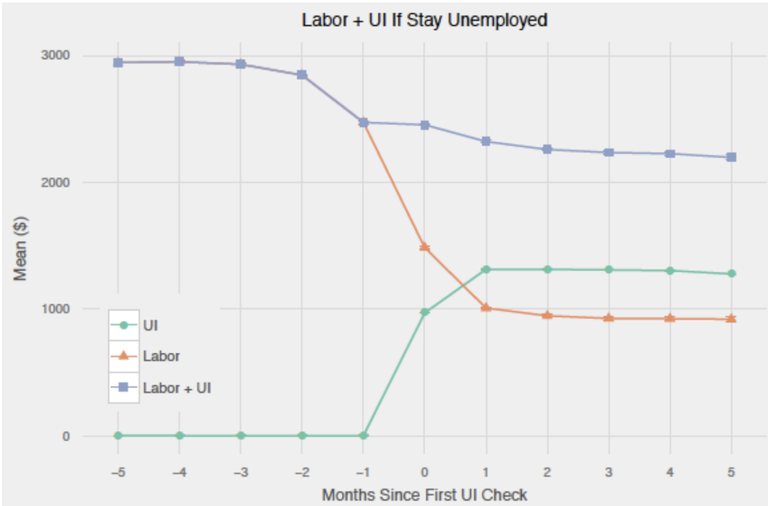


Figure 1.7: Labour Income and UI Benefits

We can see a clear drop in individuals' disposable income after becoming unemployed.

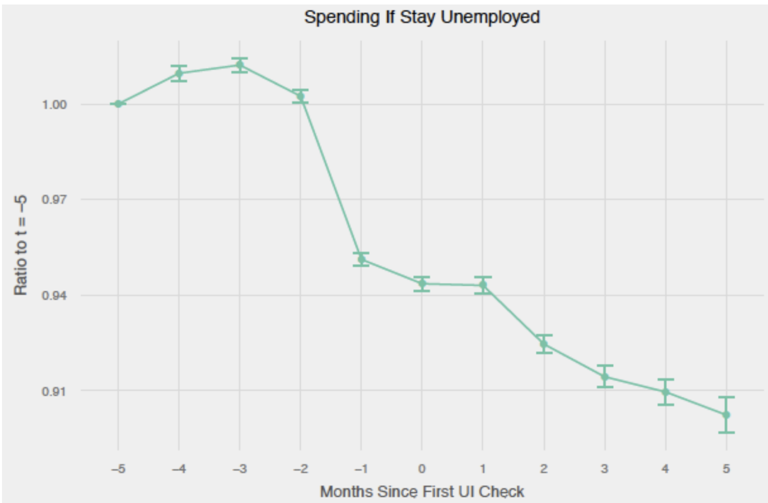


Figure 1.8: Spending if Individuals Stay Unemployed

We can see that, if individuals stay unemployed, their spending will keep decreasing.

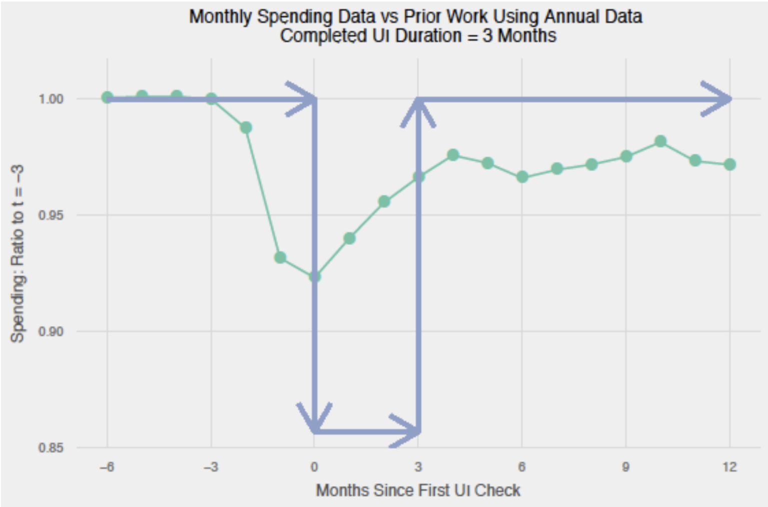


Figure 1.9: Spending if Individuals Find a Job after 3 Months

However, if they find a job after 3 months, their consumption drop will be much less.

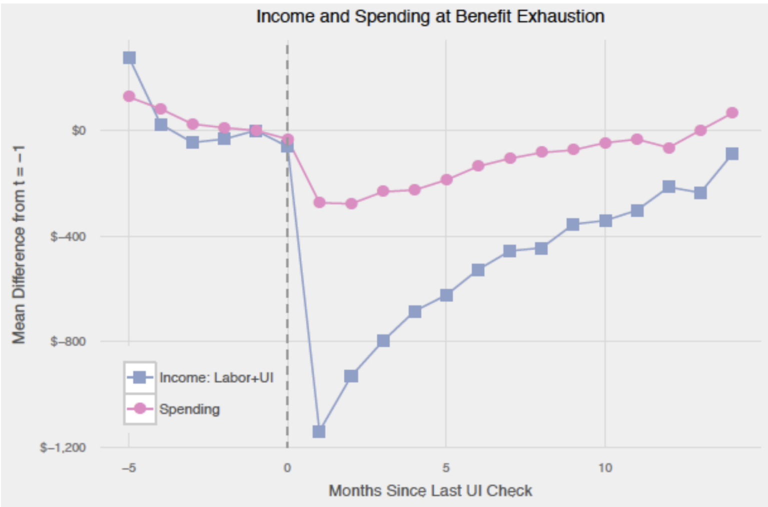


Figure 1.10: Consumption and Income Paths

The figure above clearly indicates that, during unemployment, people are spending more than their income, suggesting that they are smoothing their consumption.

1.6.3 Using Administrative Data

Kolsrud, Landais, Nilsson and Spinnewijn (2016) uses administrative data on unemployment, which is linked to data on income and wealth in Sweden. They use income and wealth changes to create a registry-based consumption measure. Again, this overcomes the small sample problem present in survey data.

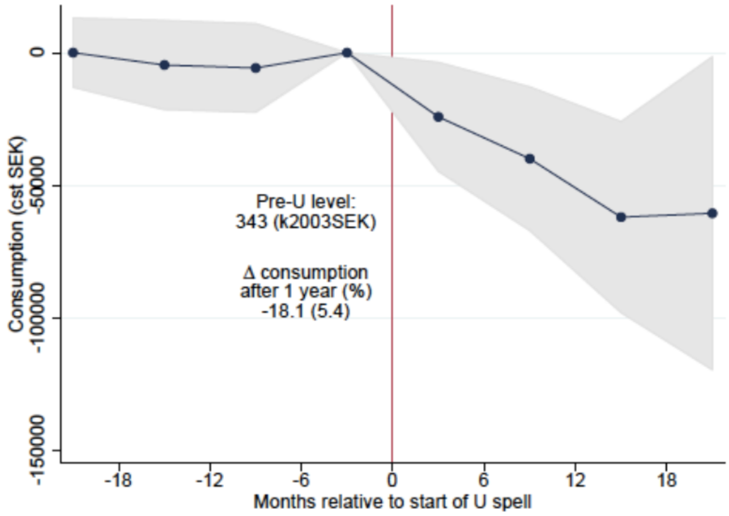


Figure 1.11: Kolsrud's Results

Their results also corroborated the existence of consumption smoothing.

1.7 ⋆ Implication of the Empirical Evidence on the Optimal UI

1.7.1 Is Our Current UI Benefit Optimal?

Recall our "sufficient statistical approach" version (equation 1.5) of the Main Equation of Optimal UI:

$$\underbrace{\frac{1}{1-p} \times \epsilon_{p,b}}_{\text{Moral Hazard Cost}} \approx \underbrace{\gamma \times \frac{\Delta c}{c^e}}_{\text{Insurance Value}}$$

From our empirical researches:

- Consumption drop $\frac{\Delta c}{c^e}$ is around 10%
- Elasticity of probability of unemployment with respect to UI benefit level $\epsilon_{p,b}$ is around 0.3
- The probability of being employed $(1-p)$ is around 90-95%, depending on the state of the economy
- Survey results suggest an average RRA γ around 2

Plug in those values:

$$\underbrace{\frac{1}{0.95} \times 0.3 \approx 0.316}_{\text{Moral Hazard Cost}} > \underbrace{2 \times 0.1 = 0.2}_{\text{Insurance Value}}$$

$$\text{Moral Hazard Cost} > \text{Insurance Value} \;\Rightarrow\; b > b^*$$

Currently, the moral hazard cost is higher than the insurance value, indicating that our UI benefit level is too high (*too generous*). We should cut the benefit level to optimize.

1.7.2 Counter-arguments by Chetty and Szeidl (2007): Consumption Commitment & Risk Aversion

Chetty and Szeidl (2007) pointed out that risk aversion, γ, is poorly identified, and it appears to vary substantially according to specific situations. Specifically, people with consumption commitments tend to have higher risk aversions as explained below.

Consumption Commitment

Chetty and Szeidl extended the standard expected utility model to incorporate goods whose consumption is hard to adjust (e.g. mortgage).

The standard utility model assumes that there is only one consumption good c, and people can cut back on all consumption goods freely anytime. This implies that, when unemployed, consumption of food, housing, cars, furniture... will all drop.

However, in practice, it is hard to adjust many elements of consumption in short run due to adjustment costs.

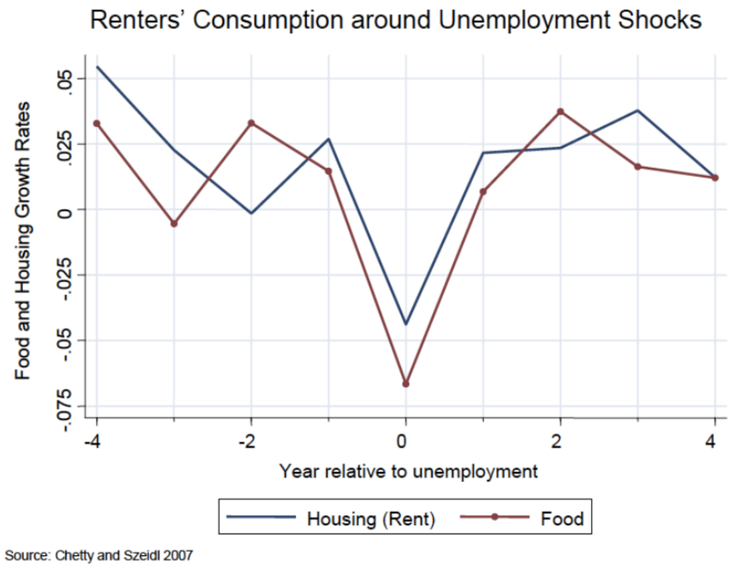


Figure 1.12: Consumption Commitment

Shown above, compared with consumption on foods, rents are much harder to be cut off. Therefore, facing unemployment, individuals may need to reduce more flexible consumption, such as foods. In this case, the actual welfare loss will be amplified.

Higher Risk Aversion

Chetty's model with consumption commitments suggests that risk aversion γ may be greater than 4 when agents are hit by an unemployment shock, because consumption commitment amplifies negative utility effects, making people more dislike negative income shocks.

Plugging in this value:

$$\underbrace{\frac{1}{0.95} \times 0.3 \approx 0.316}_{\text{Moral Hazard Cost}} < \underbrace{4 \times 0.1 = 0.4}_{\text{Insurance Value}}$$

$$\text{Moral Hazard Cost} < \text{Insurance Value} \;\Rightarrow\; b < b^*$$

Therefore, under this more realistic model, our current UI benefit level is lower than the optimal level.

1.8 Should UI Benefits Vary by the Business Cycle?

Our optimization requires equalization of insurance value and moral hazard cost, which contain parameters: $p, \epsilon_{p,b}, \frac{\Delta c}{c^e}, \gamma$. If they vary with the business cycle, our optimal UI benefits should adjust accordingly. Thus, again, this is an empirical question.

1.8.1 Unemployment Duration and the Business Cycle

Schmieder, von Wachter and Bender (2011) conclude that, during recessions:

- the responsiveness of unemployment duration to an additional month of UI benefit is lower. ($\frac{dp}{db}$ is lower)
- the probability of employment is lower ($(1-p)$ is lower or p is higher)

Therefore, the elasticity of unemployment rate with respect to benefits $\epsilon_{p,b} = \frac{b}{p}\frac{dp}{db}$ is lower in recessions. This in turn indicates lower moral hazard costs, and UI benefits should be higher in recessions to optimise.

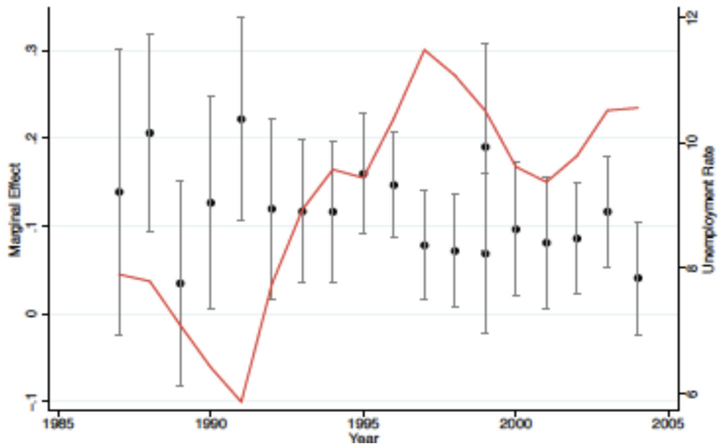


Figure 1.13: Red line: unemployment rate; Marginal effect: shows $\frac{dp}{db}$

1.8.2 Consumption Smoothing and the Business Cycle

Kroft and Notowigdio (2015) replicate and extend works by Gruber (1997). They estimate how the effect of UI on the consumption drop upon unemployment varies with the state unemployment rate in the previous year. They do not find evidence for larger or smaller drop in consumption when unemployment is higher. Thus, we can conclude that the value of consumption smoothing does not vary with the business cycle.

1.8.3 Counter-cyclical UI Benefits

As discussed in the two subsections above: during a recession, moral hazard costs decrease while insurance values do not change. Therefore, it is optimal to implement a counter-cyclical UI benefit scheme: benefits should be larger in recessions and lower in booms.

1.9 Extension: Match Quality and Long-term Effects

We ignored another potential benefit of UI schemes in our simple model above: improvements in match quality. When people become unemployed and face the pressure of purchasing necessities, they may be forced to take worse jobs. For example, an experienced engineer may have no choice but to work in McDonald's. Provision of UI benefits may alleviate such mismatching.

On the other hand, if UI benefits induce people to stay unemployed longer, there may be skill deprecation: an unemployed individual may lose his/her social skills and self-esteem.

Which of these dominates is an empirical question:

- Card, Chetty, and Weber (2007) exploit the same discontinuity in Austria to examine the effect on subsequent wages or on subsequent job tenure, and found no significant improvement
- Schmieder et al. (2015) use RDD to examine the effect of extending UI eligibility from 12 months to 18 months at age 42 in Germany. They found non-employment duration increases by 1 months while re-employment wage decrease by 3%, supporting skill deprecation.
- Nekoei and Weber (2015) use RDD to examine the effect of extending UI eligibility from 7.5 months to 9.5 months at age 42 in Austria. They find that non-employment duration increases

by 3 days, and they also find positive effects on wages, suggesting that UI benefits improve matching quality.

All in all, empirical results are contradictory. In reality, it is very likely that both match quality improvement and skill depreciation exist. The overall effect depends on which one dominates.

Chapter 2

Minimum Wage

2.1 ⋆ Neoclassical Model (Price Theory)

2.1.1 Firm's Problem and Equilibrium

Firms maximise their profits by choosing the level of employment L_j and taking market-level wage w as given:

$$\max_{L_j} pf(L_j) - wL_j$$

The FOC is:

$$\underbrace{pf'(L_j)}_{\text{Marginal Revenue Product}} = \underbrace{w}_{\text{Marginal Cost}}$$

Therefore, the equilibrium employment is:

$$L_j^* = f'^{-1}\left(\frac{w}{p}\right)$$

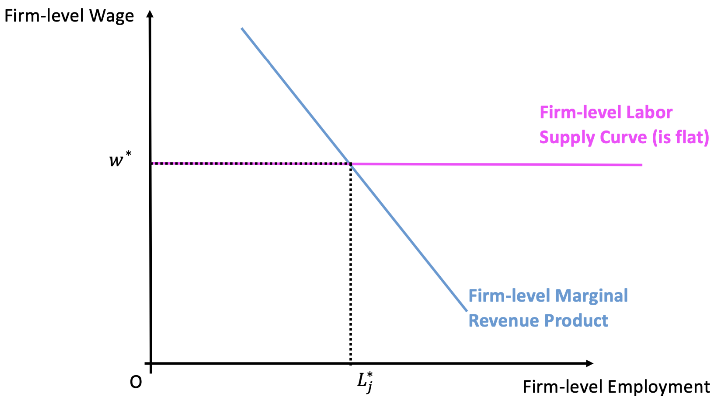


Figure 2.1: Neoclassical Model: Firm's Problem

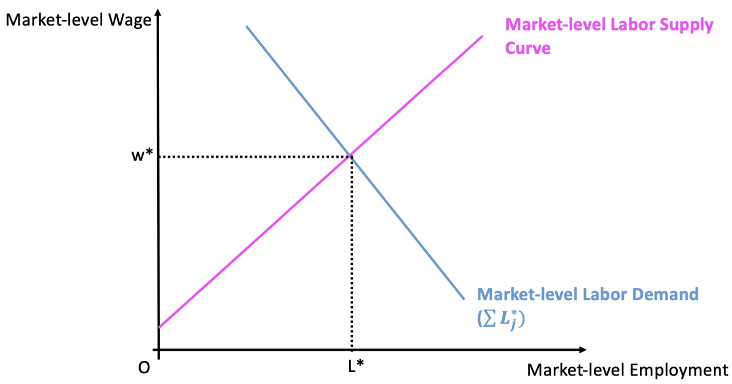


Figure 2.2: Neoclassical Model: Market Equilibrium

2.1.2 Effect of Minimal Wage

At firm's level, a minimal wage above equilibrium induces decrease in employment:

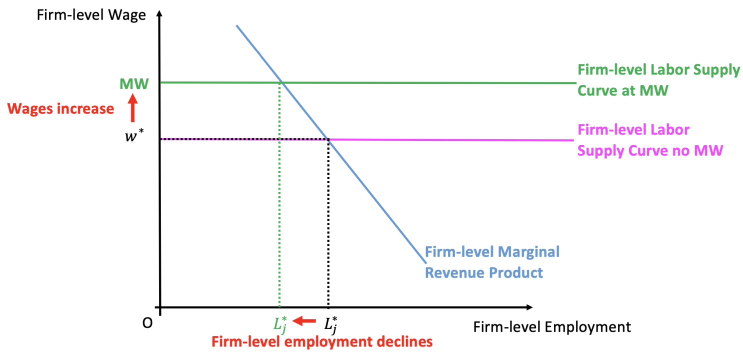


Figure 2.3: Minimal Wage: Firm's Response

At market level, this causes a welfare loss:

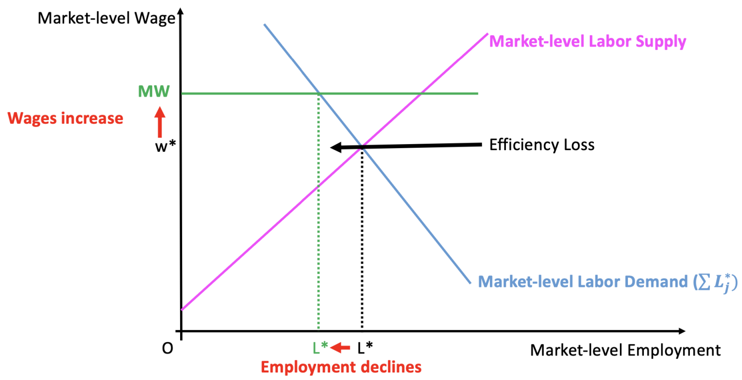


Figure 2.4: Minimal Wage: Market Equilibrium

2.1.3 Debate over the Neoclassical Model and the Consensus

George Stigler (1946): Price Theory

- There is no free lunch
- Minimum wage decreaes employment and leads to efficiency loss
- (This is a testable prediction)

Richard Lester (1947): Common Sense Approach

- Abandoned the model and looked at data instead
- Ask managers about what determines employment
- Most important factor mentioned in responses: demand of output
- Less than 20% reported labour cost as the most important factor
- Conclusion: (small) increase in labour cost will have a limited effect on employment

The Consensus: Milton Friedman (1953): "Essays in Positive Economics"

- Assumptions can be unrealistic and simplifying
- A good model focuses on the most important mechanisms, instead of discussing all irrelevant ingredients
- A good model can explain and predict what happens if circumstances change
- Evidence produced over the 60-80s supports declining employment
- **Consensus**: Chicago-style approach/price theory captures what happens in response to a minimum wage

2.2 Consensus Breaks: Modern Empirical Evidences

2.2.1 Card and Krueger (1994, 1995) and the "Credibility Revolution"

Card and Krueger (1994, 1995)

Card and Krueger (1994, 1995) studied the impact of an increasing minimum wage in New Jersey with a Difference-in-Difference approach. Their control group is Pennsylvania where there was no minimum wage hike. Their results showed that the increase in minimum wage increased both wages and employment in New Jersey. This contradicted the neoclassical model.

- Their DiD design:
$$\Delta E_i = \alpha + \beta' X_i + \gamma NJ_i + \epsilon_i$$
where X_i controls for store characteristics and NJ_i is a dummy takes value 1 for New Jersey.
- Their Exposure design:
$$\Delta E_i = \tilde{\alpha} + \tilde{\beta}' X_i + \tilde{\gamma} GAP_i + \tilde{\epsilon}_i$$
where $GAP_i = NJ_i \times \max\left\{\frac{5.05 - W_{1,i}}{W_{1,i}}, 0\right\}$ accounts for the wage increase due to the policy.

Results:

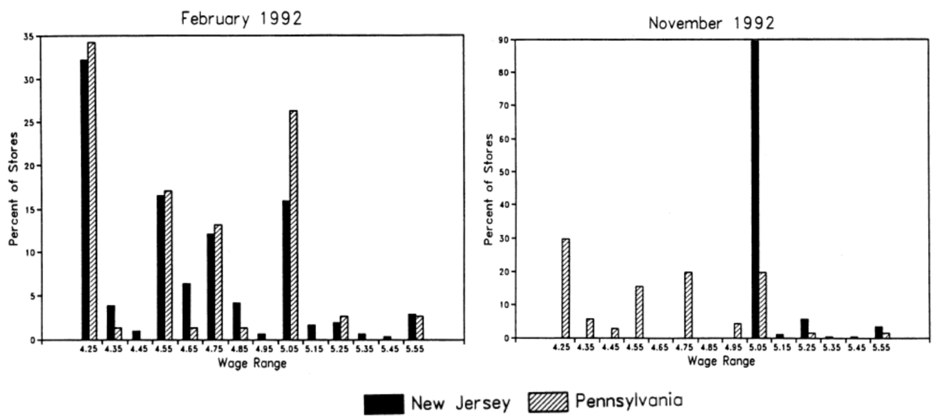


Figure 2.5: Wage Distributions

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

	Model				
Independent variable	(i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap[a]	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership[b]	no	yes	no	yes	yes
4. Controls for region[c]	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls[d]	—	0.34	—	0.44	0.40

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are −0.237 and 8.825, respectively. All models include an unrestricted constant (not reported).

[a]Proportional increase in starting wage necessary to raise starting wage to new minimum rate. For stores in Pennsylvania the wage gap is 0.

[b]Three dummy variables for chain type and whether or not the store is company-owned are included.

[c]Dummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

[d]Probability value of joint *F* test for exclusion of all control variables.

Figure 2.6: 1st Row: DiD; 2nd Row: Exposure Design

Their results provided clear evidence that the increase in minimum wages in New Jersey caused an increase in both wages and employment.

A Lesson Learnt for Economists

- Economic models are important as they can capture the most important mechanisms at play
- Simplicity is value, but we should not be dogmatic about our models
- We need "credible" research designs, that could be used to reject or accept the key predictions of our models
- Alan Krueger (2018): "The idea of turning economics into a true empirical science, where core theoriescan be rejected, is a BIG, revolutionary idea."

2.2.2 New Mimimum Wage Research

New minimum wage research shifted the focus from theory to "reduced form" empirical evidence on employment and wages, investigating the impact of minimum wages directly.

The U.S. has been a fertile ground due to state-level variations that can be exploited:

- Two-way Fixed Effects estimation (e.g. Neumark and Wascher, 1993)
- Usage of administrative data (e.g. Card and Krueger, 2000)
- Border-discontinuity design (e.g. Dube, Lester and Reich, 2010)

On the other hand, most of the evidence focuses on specific demographic groups (e.g. teens) or sectors (e.g. restaurants). It could be possible that minimum wages increase employment in some sectors but decrease it in others.

2.2.3 Frequency Distribution Based Approach

Idea

Cengiz et al. (2019) develop a novel method to assess the overall employment change using changes in distribution:

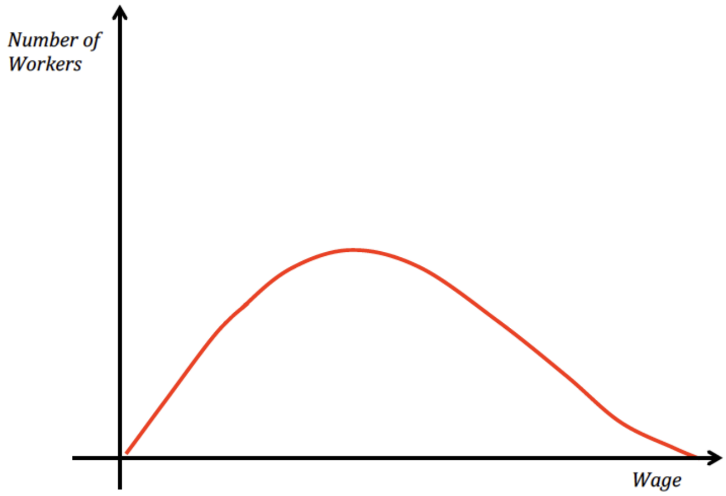


Figure 2.7: Distribution of Employment before Policy Change

After an increase in the minimum wage, jobs with wages below the threshold will disappear and new jobs with wages higher than the threshold will be generated. The net effect on employment is illustrated below:

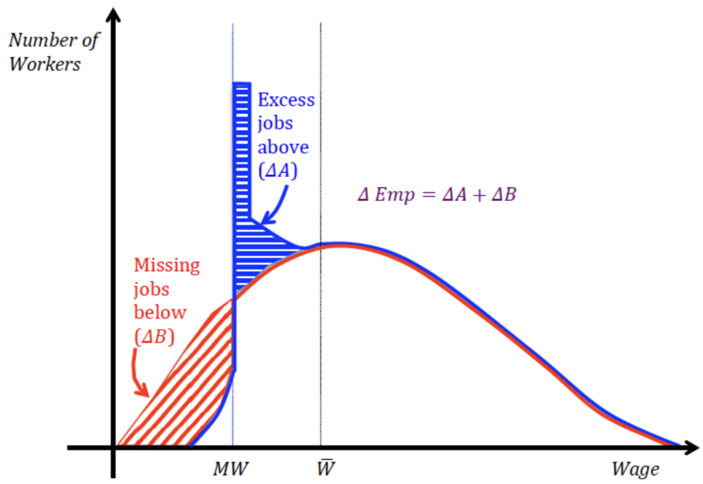


Figure 2.8: Distribution of Employment after Policy Change

The key empirical challenge of this method is to get counterfactual wage distribution. The solution is to use a DiD idea.

During 1999-2000, Washington State raised its minimum wage from \$7.51 to \$9.18, and WA has administrative data on hourly wages. We divide the distribution by wage bins, and calculate the counterfactual by creating a synthetic control group from a number of states.

With this method, Cengiz et al. obtained both actual and counterfactual employment distributions:

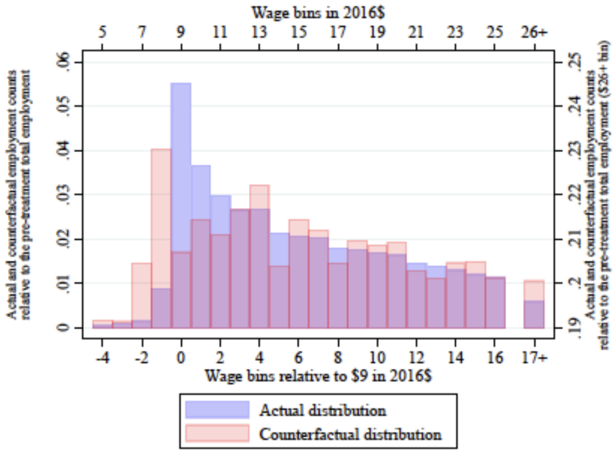


Figure 2.9: Actual and Counterfactual Distributions

Comparing the two distributions, the effect of an increase in minimum wage is retrieved. The patter of change matches Figure 2.8, and, as shown by the red line, there is a redistribution of employment from jobs with wages below the threshold to jobs with wages above the threshold. No reduction in overall employment is observed:

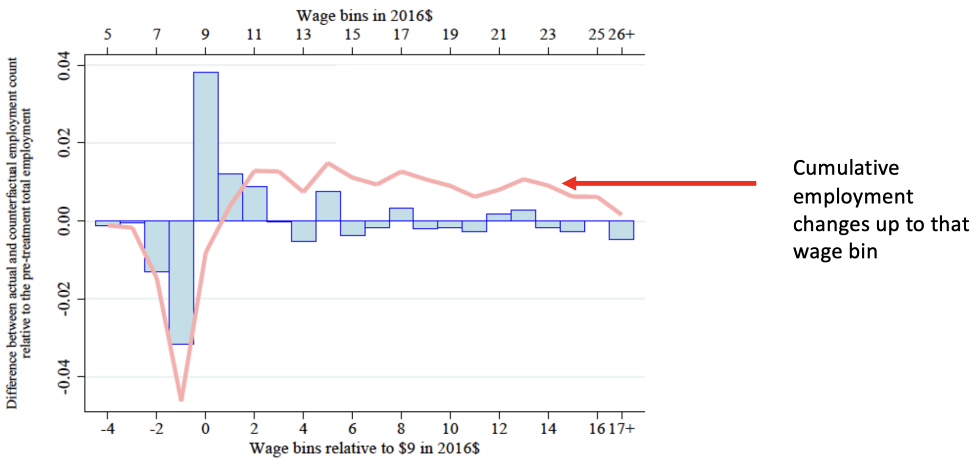


Figure 2.10: Treatment Effect

Taking heterogeneity into account, we cannot find any overall dis-employment either:

◇ No evidence that dis-employment effects are more prominent for larger changes

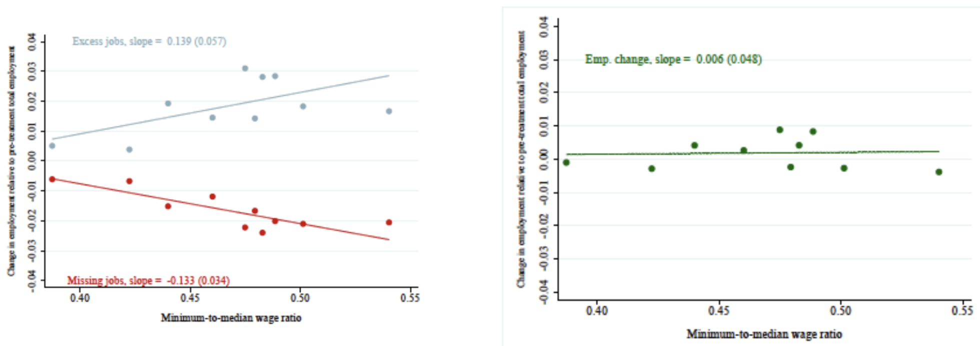


Figure 2.11: Heterogeneity by Magnitude of Increase

◇ No evidence that dis-employment effects are more prominent when unemployment rate is high

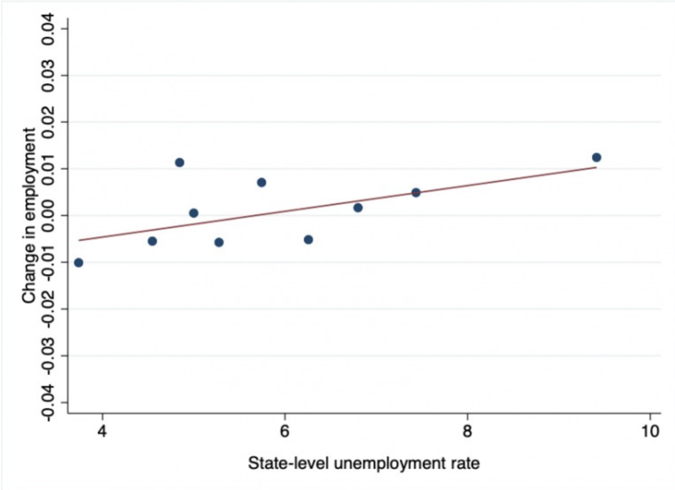


Figure 2.12: Heterogeneity by State Unemployment Level

2.3 ⋆ Monopsonistic Model

2.3.1 Firm's Problem and Equilibrium

Assumptions

Empirical research presented in the previous section show no evidence that an increase in minimum wage will cause dis-employment. This indicates that assumptions we made for the standard neoclassical "price theory" model do not hold. In reality:

◇ Workspaces are differentiated (not homogeneous)

◇ Such differentiation provides each employer with some market power

◇ The labour market is a monopsonistic market

The intuition is that, because some workers want to work at a certain firm, the employer can hire them with lower wages. This creates a wedge between productivity and wages. In equilibrium, fewer workers are hired and wages are lower.

Firm's Optimisation Problem

Firms maximise their profits:

$$\max_{L_j} pf(L_j) - w(L_j)L_j$$

The FOC implies that:

$$\underbrace{pf'(L_j)}_{\text{Marginal Revenue Product}} = \underbrace{w(L_j)\left(1+\frac{1}{\eta}\right)}_{\text{Marginal Cost}}$$

where η is the elasticity of firm-level labour supply defined as $\eta = \frac{\partial L_j}{\partial w}\frac{w}{L_j} > 0$. The equilibrium wage in monopsonistic model is:

$$w^*(L_j) = \underbrace{pf'(L_j)}_{\text{Marginal Revenue Product}} \times \underbrace{\frac{1}{1+\frac{1}{\eta}}}_{\text{Mark-down}(<1)} < w^*_{Neoclassical}$$

The equilibrium wage here is different from the marginal revenue product: firms enjoy a "mark-down." The magnitude of such difference depends on firm-level elasticity of labour supply: if $\eta \to \infty$, we get back to the neoclassical framework.

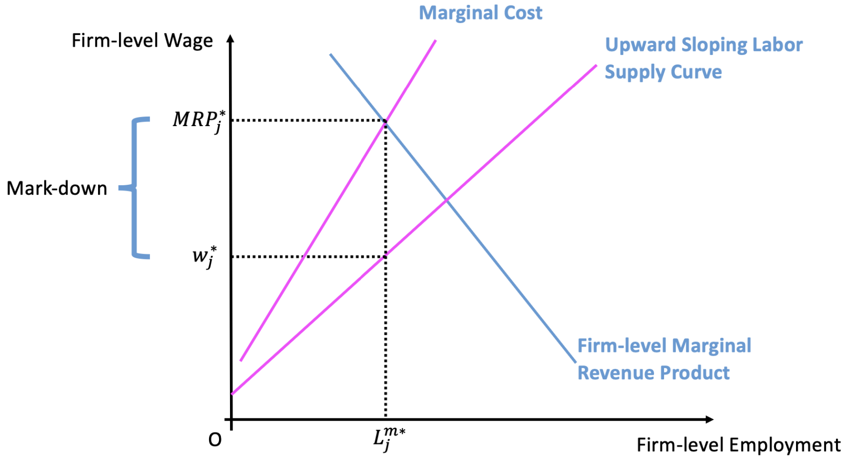


Figure 2.13: Equilibrium in a Monopsonistic Labour Market

Predictions

Several main predictions of this monopsonistic model:

- People feel they are underpaid at their workplace relative to their true productivity – wages are "marked-down"
- People feel they could find a better paying workplace, but those are not ideal (they are too far from home, not nice co-workers)
- Workers with similar skills and ability are paid differently at different firms (there is a firm-level skill premium)
- Efficient firms are larger and pay more to workers with similar skills

2.3.2 Effect of Minimal Wage: Reallocation

In this framework, a minimum wage scheme will cause:

- The least efficient firms will close
- The most efficient firms increase wages, the mark-down (or profit per worker) falls
- Though the wedge between productivity and wage falls, it is still profitable to employ workers
- Firms will hire more with lower mark-down

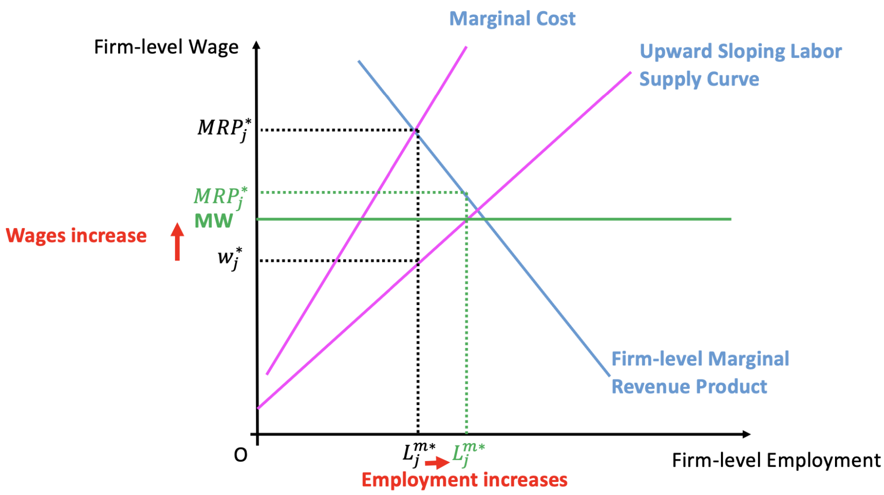


Figure 2.14: Minimum Wages in Monopsonistic Model

The overall effect is featured as a process called reallocation: employees of the least efficient firms are transferred to the more efficient one. This improves efficiency. Meanwhile, though new firms pay more, workers might be worse off in other aspects, so the overall effect on welfare is still ambiguous.

2.3.3 Empirical Research on Reallocation

Historically, the role of reallocation has been featured prominently in the minimum wage debate since the later 19th century. (e.g. advocates to use minimum wage to stop the proliferation of ”sweatshops” in the 1890s; Winston Churchill’s argument for the Trade Boards Act 1909)

Dustmann, Lindner, Schönberg, Umkehrer, and vom Berge (2022)

Dustmann, Lindner, Schönberg, Umkehrer, and vom Berge (2022) provide direct evidence on the impact of the minimum wage on reallocation:

- Reform studied: Introduction of the minimum wage in Germany in 2015
- Data: Administrative employer-employee database covering the life history of all German private sector workers
- Empirical analysis: How wages, employment, and quality of the firm evolved following minimum wage hikes?
- Control groups:
  - Projection based on the observed trend before the introduction of the minimum wage
  - Wage evolution at the higher end of the wage distribution

Results

Positive wage increases for employees around the minimum wage:

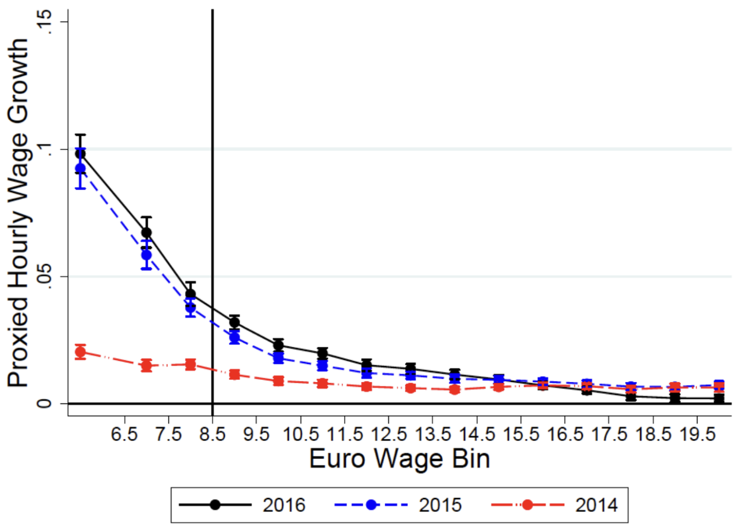


Figure 2.15: Wage Growth as a Function of Wages 2 Years Before

Increase in employment:

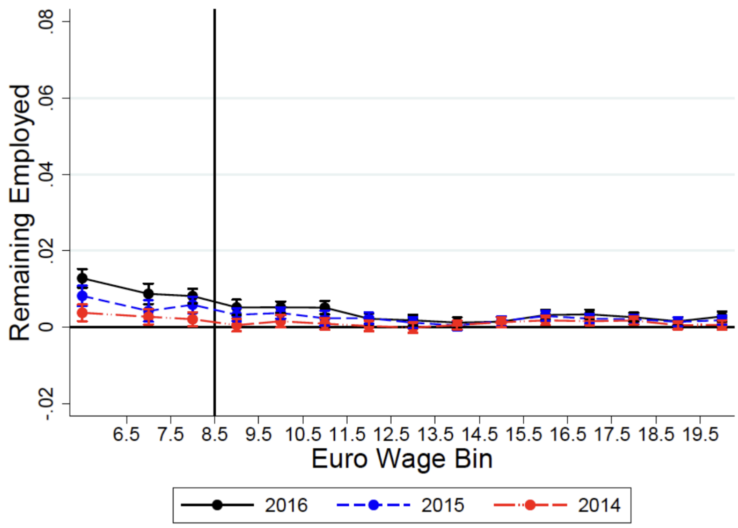


Figure 2.16: Employment Growth as a Function of Employment 2 Years Before

Higher wage premium:

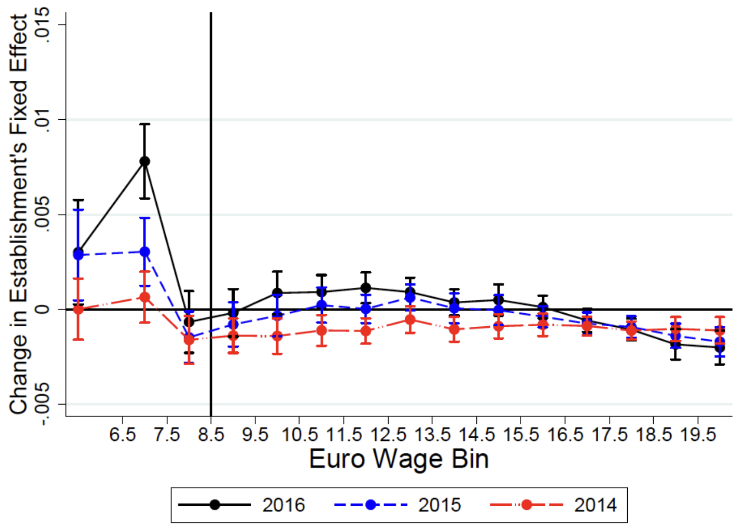


Figure 2.17: Firm Wage Premium as a Function of Wages 2 Years Before

Higher productivity:

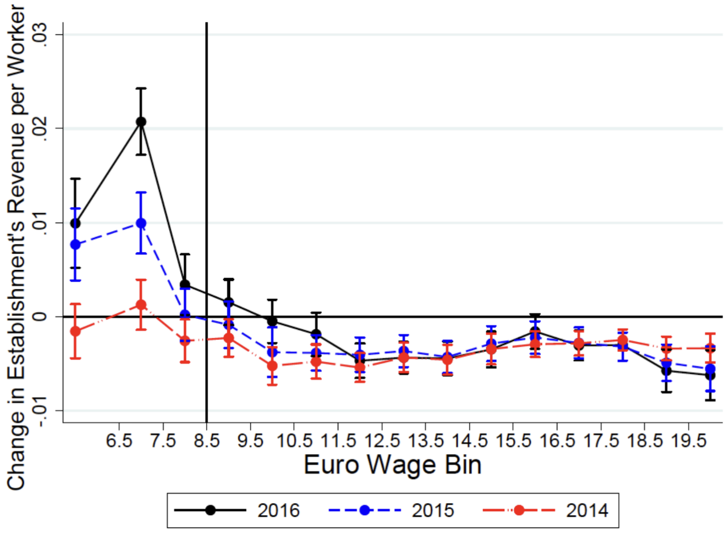


Figure 2.18: Firm's Productivity Growth as a Function of Wages 2 Years Before

Reallocation from small firms to larger firms:

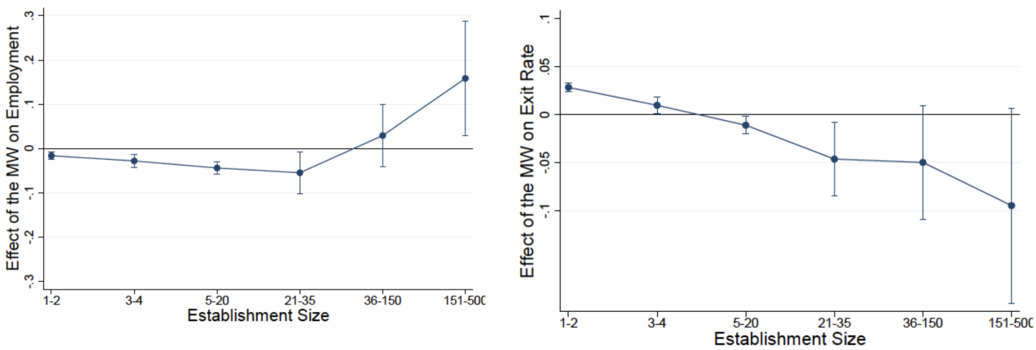


Figure 2.19: The Effect of the Minimum Wage by Firm Size

Key findings:

- Positive and significant effect on wages
- No dis-employment effect
- *Reallocation* of workers to:
  - Firms paying higher wage premium
  - Firms with lower turnover
  - Firms with larger sizes higher productivity

This indicates that minimum wages reallocate workers to more productive firms, contributing productivity improvement. The welfare implication is more ambiguous since there's evidence that workers' commuting distance increased.

Chapter 3

Health as Human Capital[1]

3.1 Intro: Why are some of us less healthy than others?

- Bad genes
- Bad lifestyles
- Poor healthcare
- Poor early environment (e.g. poor parenting, mum drinking during pregnancy)

3.2 Health Spending

3.2.1 Facts about Health Spending: Global

Health spending per capita

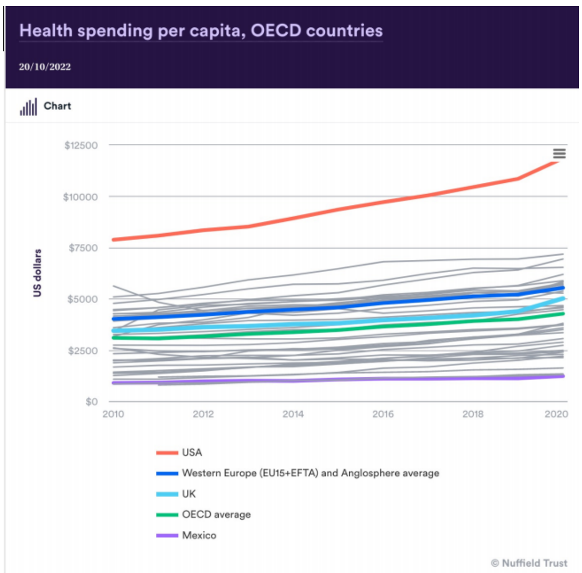


Figure 3.1: Health Spending per Capita, OECD countries

- Health spending per capita is increasing over time. People are living longer than they used to be. The longer you live, the sicker you get and the more money you spend on healthcare.
- US has a higher health spending and a particular healthcare structure.

[1] Written by Kuangjie Ni, Edited by Xiaotian Tian

A Large Fraction Is Government/Compulsory Spending

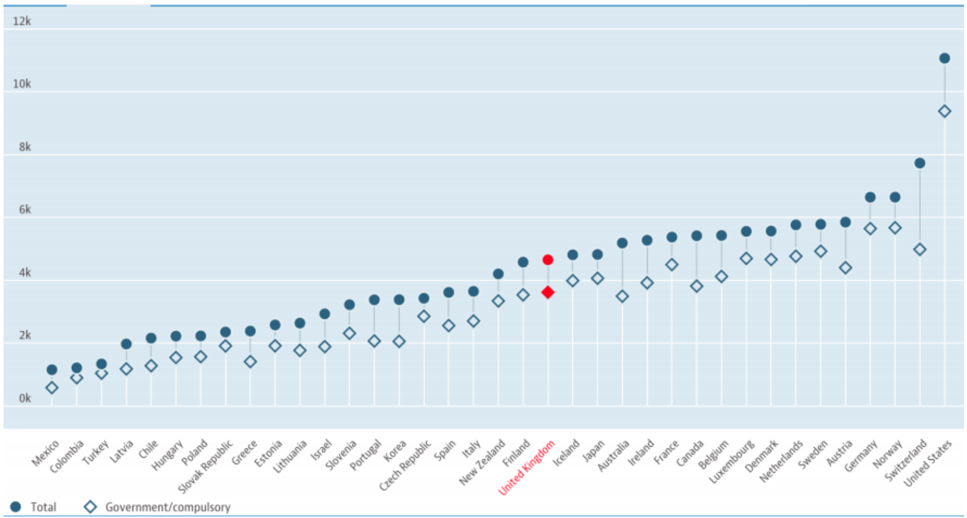


Figure 3.2: Health spending: Total including government/compulsory spending in US dollars/capita, 2019

- Blue dots are the total health spending, and diamonds are the government/compulsory spending.
- US government spends a lot of money on healthcare.

Total Spending as % of GDP

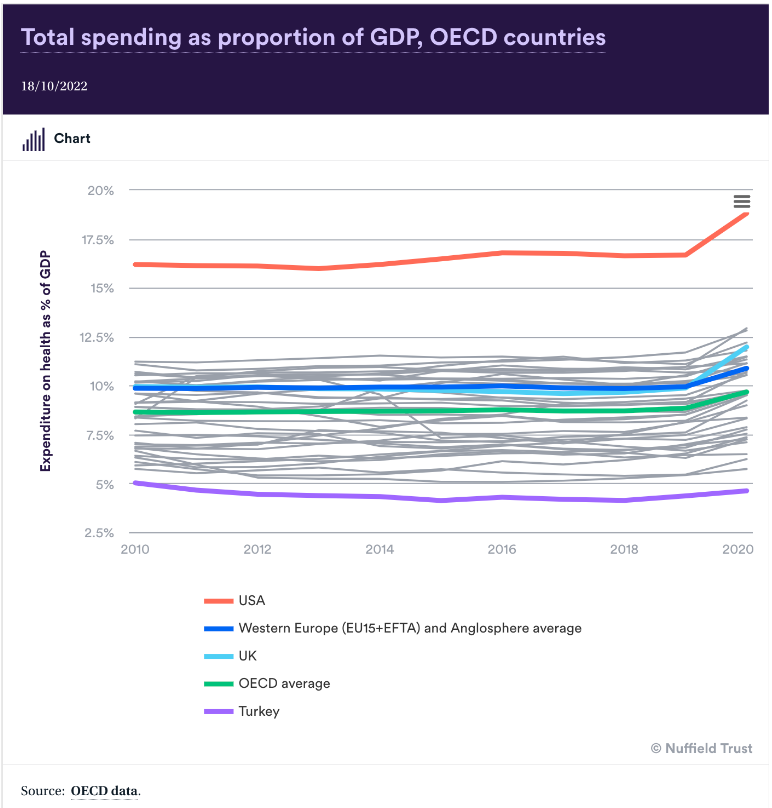


Figure 3.3: Total spending as a proportion of GDP, OECD countries

- US has the highest health spending as a proportion of GDP.
- The total health spending as a proportion of GDP increased during the pandemic.

3.2.2 Facts about Health Spending: the UK

Health as % of UK Total Spending

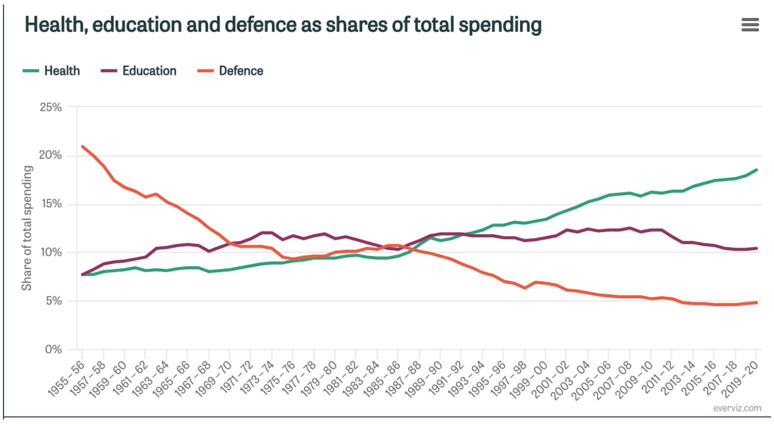


Figure 3.4: Health, education and defence as shares of total spending

- Health spending takes a bigger share of total UK spending over the last 70 years.
- The share of education spending is approximately flat, and the share of defence spending declines significantly.

Components of UK Government Spending

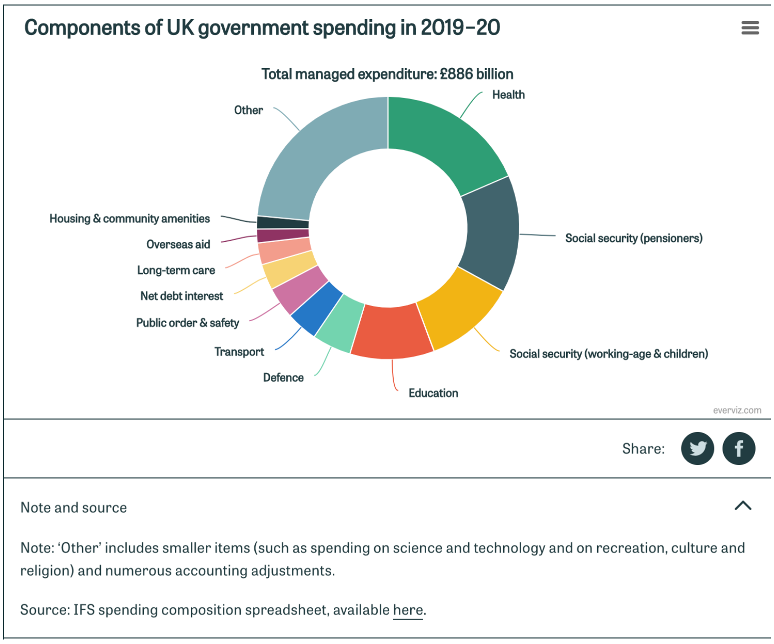


Figure 3.5: Components of UK GOV spending in 2019-20

Total managed UK government expenditure is 886 billion pounds, and health spending is a very important component.

UK Public Spending on Health

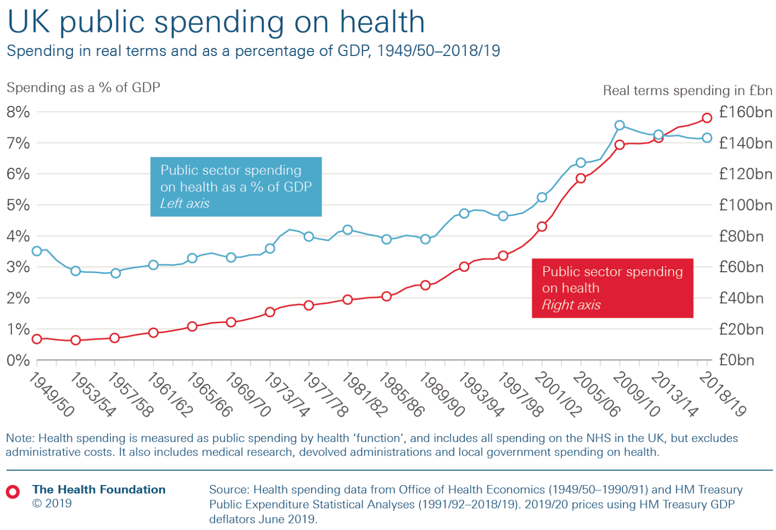


Figure 3.6: UK public spending on health, 1949/50 - 2018/19

- ◇ The blue line shows the UK public sector spending on health as a percentage of GDP (left axis). The red line shows the UK public sector spending on health in real terms (right axis).
- ◇ The UK public sector spending on health has increased both as a percentage of GDP and in real terms since the end of World War II.

Department of Health and Social Care spending

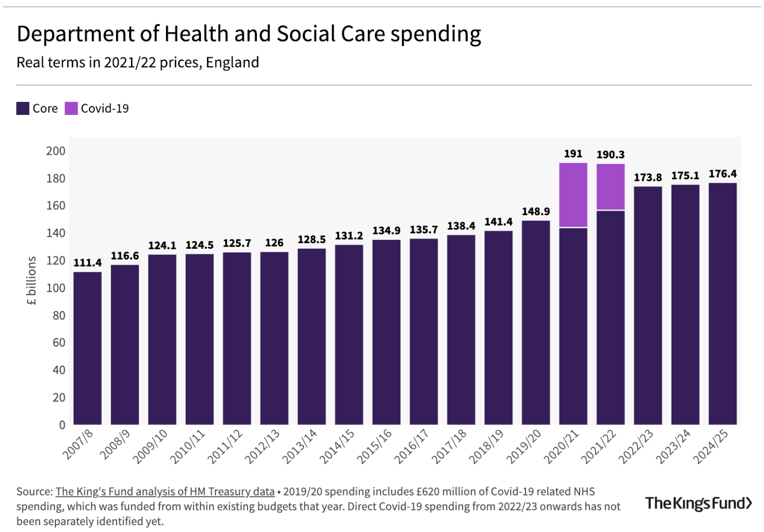


Figure 3.7: Department of Health and Social Care spending, real terms in 2021/22 prices, England

The Department of Health and Social Care spending increased a lot due to Covid-19, and the core spending keeps going up after 2021/22.

UK Health Spending Growth

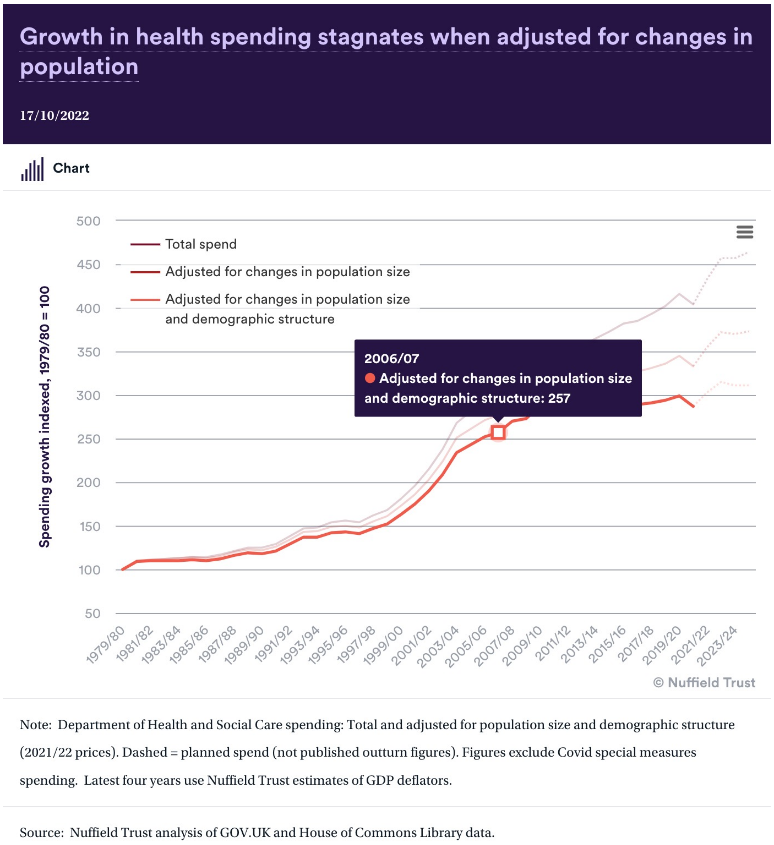


Figure 3.8: Health spending growth indexed, 1979/80 = 100

- ◇ The dark red line shows the total health spending. The red line shows the total health spending adjusted for changes in population size. The orange line shows the total health spending adjusted for changes in population size and demographic structure.
- ◇ Growth in health spending stagnates after adjusted for changes in population (population ageing).

Changes in health spending per capita, adjusted for demographic changes

Time period, political party	Average annual change in per capita health spending (adjusted)
1979–1997, Conservatives	+2.03%
1997–2010, Labour	+5.67%
2010–2015, Con/Lib coalition	-0.07%
2015–2021, Conservatives	-0.03%
2021–2024, Conservatives – committed spend	+2.05%

Figure 3.9: Changes in health spending per capita, adjusted for demographic changes

The Labour government has the highest increase in health spending per capita.

3.2.3 How Is the Health Budget Spent?

How funding flows in the NHS?

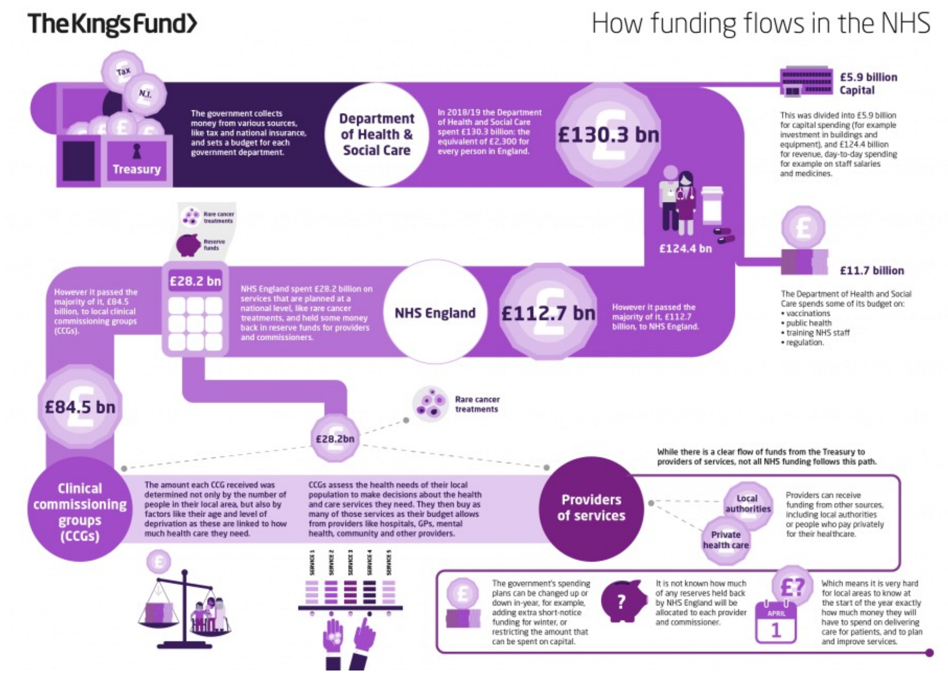


Figure 3.10: How funding flows in the NHS?

- In 2018/19, the Treasury gave the Department of Health and Social Care 130.3 billion pounds. This was divided into 5.9 billion for capital spending and 124.4 billion for day-to-day spending.
- The 124.4 billion was further divided into 11.7 billion for preventive services (vaccinations, public health, training NHS staff, and regulation) and 112.7 billion for NHS England.
- NHS England spent 28.2 billion on national-planned services, such as rare cancer treatments, and passed the remaining 84.5 billion to the clinical commissioning groups (CCGs). The amount each CCG received was determined by the number of people in the local area and by factors such as age (how sick they are).

Example: Bradford CCG

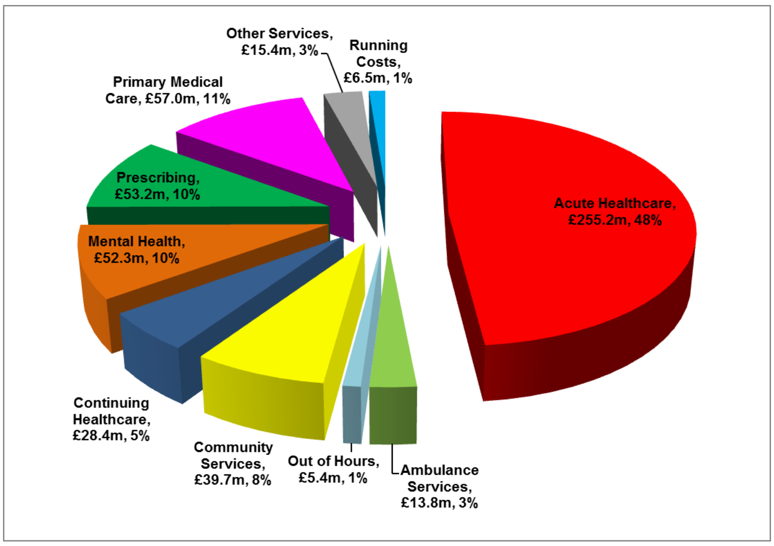


Figure 3.11: Total CCG net expenditure, 2019/20(526.9 million pounds)

Half of the money goes to acute healthcare. A small proportion of the money goes to preventive services, such as community services (health visitors and immunisation).

3.2.4 Preventive Care

3.2.5 Other Factors that Matter to Health other than Healthcare

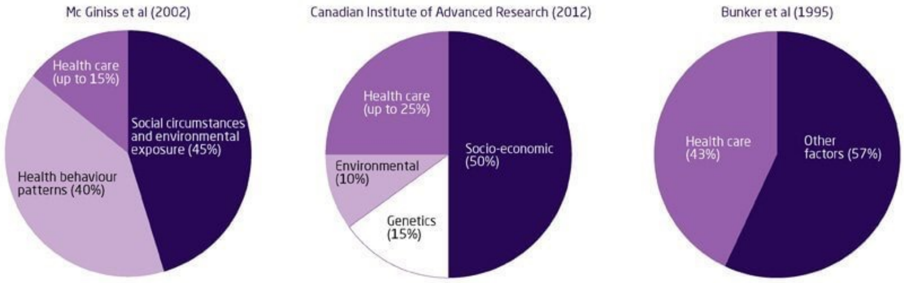


Figure 3.12: There is more to health than healthcare

These three pie charts show the proportion of health that does not depend on the healthcare. Although the proportion varies between different studies, healthcare matters by less than a half in all studies.

Example: The role of lifestyles in the prevention of chronic conditions

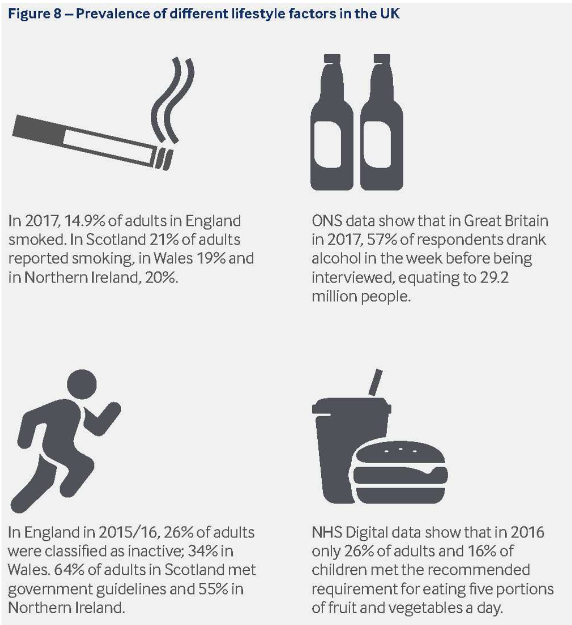


Figure 3.13: Prevalence of different lifestyle factors in the UK

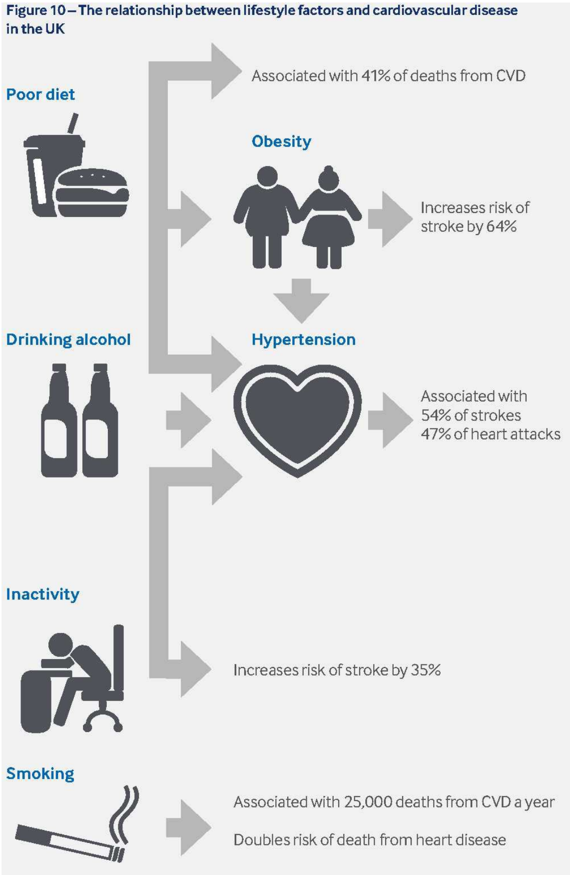


Figure 3.14: The relationship between different lifestyle factors and cardiovascular diseases in the UK

- Poor lifestyles, such as poor diet, drinking alcohol, inactivity and smoking, are associated with cardiovascular diseases.

Rank	England	PAF (%)	Scotland	PAF (%)	Wales	PAF (%)	Northern Ireland	PAF (%)
1	Tobacco	19·26	Tobacco	22·76	Tobacco	20·31	Tobacco	20·01
2	Dietary risks	14·41	Dietary risks	16·12	Dietary risks	16·35	Dietary risks	15·88
3	High blood pressure	13·04	High blood pressure	14·62	High blood pressure	15·53	High blood pressure	14·99
4	High body-mass index	9·57	Alcohol and drug use	12·98	High body-mass index	9·85	Alcohol and drug use	11·50
5	Alcohol and drug use	9·52	High body-mass index	10·70	Alcohol and drug use	9·59	High body-mass index	9·97
6	High total cholesterol	7·44	High total cholesterol	8·49	High total cholesterol	8·07	High total cholesterol	8·35
7	Occupational risks	4·85	High fasting plasma glucose	5·02	High fasting plasma glucose	5·20	High fasting plasma glucose	5·18
8	High fasting plasma glucose	4·84	Occupational risks	4·63	Occupational risks	4·55	Occupational risks	4·30
9	Air pollution	4·04	Air pollution	3·87	Air pollution	3·91	Air pollution	3·58
10	Low physical activity	2·16	Impaired kidney function	2·48	Low physical activity	2·03	Low physical activity	2·52

Behavioural
Environmental and occupational
Metabolic

Figure 2: **PAF for risk factors for all-cause YLLs rate per 100 000 population for England, Scotland, Wales, and Northern Ireland, both sexes, 2016**
PAF=population attributable fraction. YLLs=years of life lost.

Figure 3.15: PAF (Population Attributable Factions) for major risk factors for all-cause YLLs (Years of Life Lost) rate per 100,000 population for England, Scotland, Wales, and Northern Ireland, both sexes, 2016

- The population attributable fraction is the proportional reduction in population disease or mortality that would occur if exposure to a risk factor were reduced to an alternative ideal exposure scenario.
- The ten leading risk factors contributing to YLLs were similar in rank across the four regions of the UK. Although the ranks were similar, the PAF of each risk factor varied in size in different countries, such as a higher PAF from tobacco in Scotland, and from alcohol and drug use in Scotland and Northern Ireland, compared with the other UK regions.
- Poor lifestyles do matter for health, although most money goes to curative health.

3.2.6 How Is the Health Budget Spent?

Worldwide

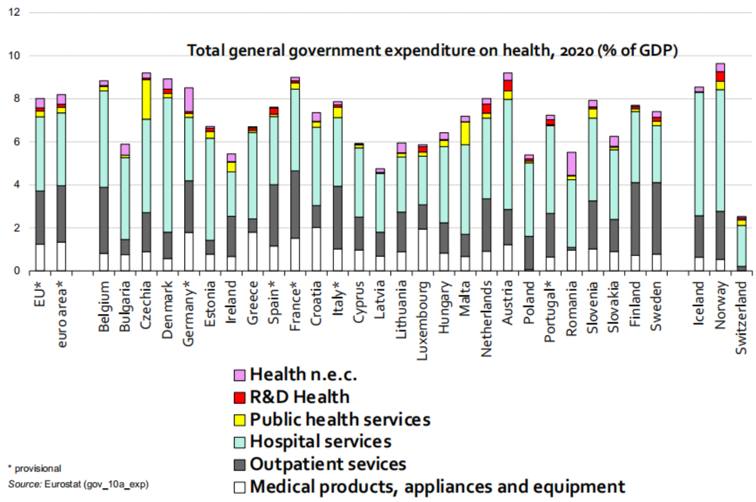


Figure 3.16: Total general government expenditure on health as a percentage of GDP, 2020

- ◇ The biggest share goes to hospital services, followed by outpatient services, so most money is used to treat people in the first place.

UK

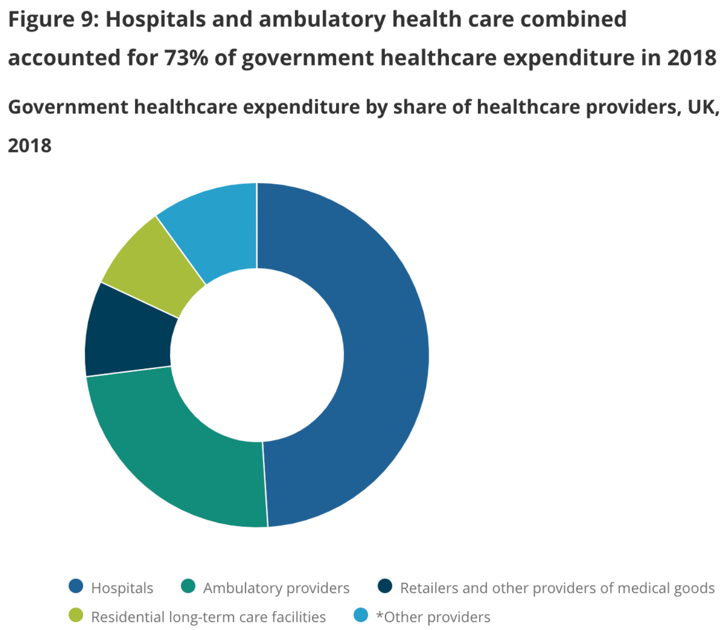


Figure 3.17: Government healthcare expenditure by share of healthcare providers, UK, 2018

- ◇ Hospitals and ambulatory healthcare combined accounted for 73 percent of government healthcare expenditure in 2018

Preventive care expenditure is easy to cut, UK

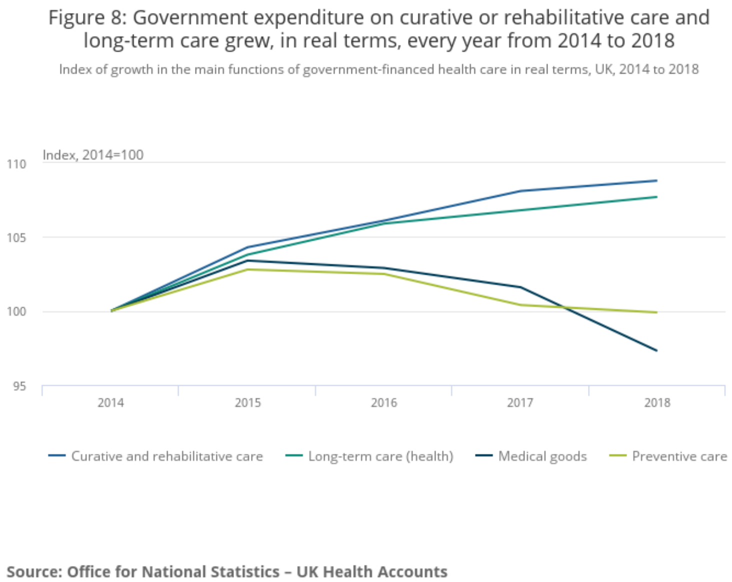


Figure 3.18: Government expenditure on curative or rehabilitative care and long-term care grew, in real terms, every year from 2014 to 2018, index: 2014 = 100

- ◇ Preventive care expenditure has decreased since 2015. This is because it is easy to cut. Curative care is difficult to cut, so the expenditure has increased over time.

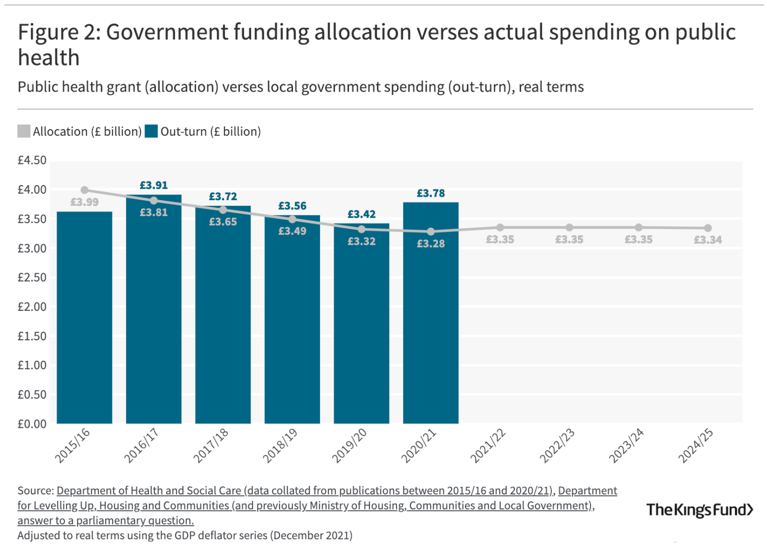


Figure 3.19: Public health grant (allocation) VS local government spending (out-turn), real terms

- ◇ The blue bars show the total local government spending on healthcare, which increased in 2020/21 due to Covid-19.
- ◇ The grey line shows the money allocated to public health prevention. It decreased since 2015/16. This is because it is easy to cut.

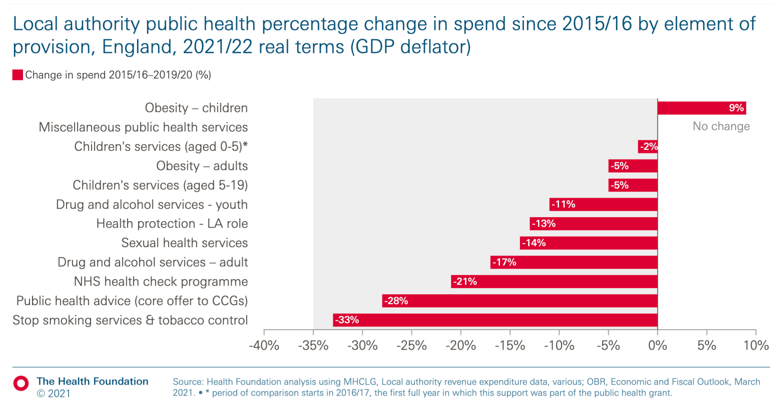


Figure 3.20: Local authority public health percentage change in spending since 2015/16 by element of provision, England, 2021/22 real terms (GDP deflator)

- For preventive healthcare expenditure, only obesity for children increased, since it is a major problem for children in the UK and costs NHS a lot of money. The expenditure is cut across all other categories.

3.2.7 Early Intervention and Prevention

- Early intervention and prevention approaches aim to support health and wellbeing by taking action before health problems worsen, or by preventing health problems for occurring in the first place.
- Public Health Prevention = improving public health through disease prevention a. Clinical interventions such as screening and vaccinations; b. Population-level measures aimed at influencing health behaviours or addressing the social determinants of health (e.g. living conditions, education etc.).
- Early Intervention = strategies aimed to mitigate the effects of problems once they have been identified. – Some specifically targeted at the 'early years', including pregnancy, early parenting and the early parent-child relationship.
- A NICE (National Institute for Health and Care Excellence) review of the cost-effectiveness of 200 interventions found that 30 (15 percent) were cost-saving and 141 (70.5 percent) were cost-effective.

3.2.8 Why government intervention in getting people to become healthier?

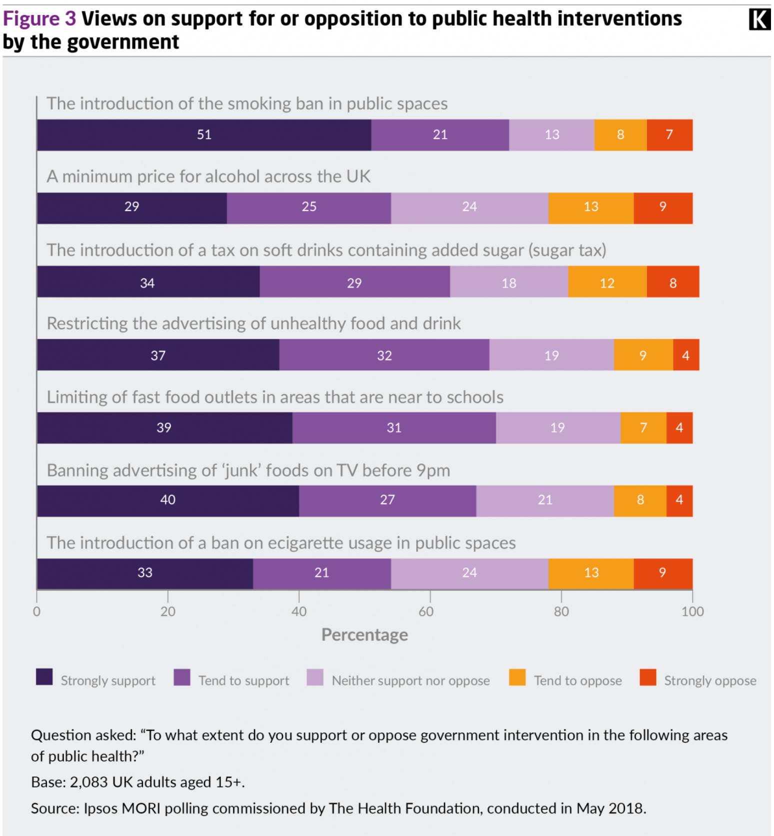


Figure 3.21: Views on support for or opposition to public health interventions by the government

Reasons for government intervention:

- improve human capital, increase productivity, raise tax revenue in the future
- market failure: asymmetric information (e.g. junk food: consumers don't know the content/how much is going to harm them)
- public goods: hospital/public health services, nobody provides unless the government provides, it is expensive
- moral hazards
- negative externality (smoking)
- time-inconsistent preference (procrastination: go to the gym); addictive goods - the government can nudge/remind people
- In order to make progress and develop effective policies, we need first to understand how health is produced.

3.3 ⋆ The Production of Health as Human Capital: The Grossman Model

3.3.1 Introduction

- Grossman (1972) studied how individuals allocate their resources to produce health. He combined the theory of human capital and the theory of the allocation of time to explain the demand for health. Health is a unique good. We use time/resources/money to produce it and enjoy it.

- Four important aspects
  1. Health can be treated both as a consumption and an investment good. As a consumption good, health yields direct utility. As an investment good, health increases the number of healthy days available to participate in the market and non-market activities.
  2. Health lasts for more than one period. It depreciates (e.g. ageing process; if no vaccine immunity, poorer health; if don't go to the gym, poorer health) and can be analyzed like a capital good.
  3. Individuals are not passive consumers of health: they produce it, spending time and money (buy market inputs).
  4. Demand for medical care is derived: people demand medical care not per se, but to produce health.
- As both a consumption and an investment good, optimal health capital accumulation requires the decision-making of its owner – the individual – who is *both the consumer and the producer.*

3.3.2 Preferences, Utility, and Death

Health as a consumption good enters directly into utility. Here, we model an individual's Intertemporal Utility Function as:

$$U = U(\underbrace{\varphi_0 H_0}_{h_0}, ..., \underbrace{\varphi_n H_n}_{h_n}, Z_0, ..., Z_n)$$

where:

- H_0 : inherited stock of health (predetermined)
- H_i : stock of health in the ith time period
- φ_i : service flow per unit stock
- $h_i = \varphi_i H_i$: healthy days
- Z_i : non-health commodity

Length of life is *endogenous*: death occurs when $H = H_{min}$. The lifespan is determined by individuals' choices of working time (TW_i), time spent on producing health investment (TH_i), time spent on producing non-health commodities (T_i), and the purchase of other market inputs (M_i).

Individuals *maximise their utilities* subject to resource (time/budget) constraints (section 3.3.3) and production technologies (section 3.3.4) introduced below.

3.3.3 Two Resource Constraints

Here, we have 2 constraints: the budget constraint and the time constraint.

The Budget Constraint is:

$$\sum_{i=0}^{n} \frac{P_i M_i + V_i X_i}{(1+r)^i} = \sum_{i=0}^{n} \frac{W_i \times TW_i}{(1+r)^i} + A_0$$

$$\text{PV of Health and Non-Health Commodities} = \text{PV of Labour Earnings} + \text{Initial Assets}$$

where:

- P_i : price for health
- V_i : price for input of non-health commodity
- M_i : good inputs in the production of health
- X_i : good inputs in the non-health commodity
- W_i : wage rate
- TW_i : time spent working
- A_0 : initial assets

- r : interest rate

The Time Constraint is:

$$\underbrace{TW_i + TH_i + T_i}_{h_i} + TL_i = \aleph$$

where:

- TW_i : time spent on working
- TH_i : time spent on producing health
- T_i : time spent on producing other goods we enjoy
- TL_i : time loss due to illness
- $\aleph$: overall time endowment ($\aleph = h_i + TL_i$)
- h_i : healthy days ($h_i = TW_i + TH_i + T_i$)

3.3.4 Production Technologies (Production Constraints)

Here, we have 5 production functions in total:

The Healthy Time Production Function is:

$$h_i = f(H_i)$$

which means healthy days can be produced by some technology using health stocks.

The Health Stock Production Function is:

$$H_{i+1} = H_i - \delta_i H_i + I_i$$

where:

- I_i : gross investment
- δ_i : depreciation (exogenous, might vary with age, e.g. older, higher depreciation)

The health stock PF implies that when investment is high, an individual can overcome the high rate of depreciation.

Such specification brings *two problems*:

- Linearity: It does not account for diminishing marginal returns
- It does not account for random shocks

The 2 Household Production Functions are:

Production of health investment (I_i):

$$I_i = I_i(M_i, TH_i; E_i)$$

Production of non-health commodities (Z_i):

$$Z_i = Z_i(X_i, T_i; E_i)$$

where

- M_i = good inputs in the production of health
- X_i = good inputs in the production of non-health commodity
- TH_i : time spent on producing health
- T_i : time spent on producing other goods we enjoy
- E_i : stock of human capital (education) $\Longrightarrow$ better education, better health

We can see that the production of health and non-health commodities needs both time and money. (Cooking and going to the gym also matter.)

The Income Production Function is:

$$Y_i = W_i \times TW_i$$

3.3.5 The Production Possibility Frontier

The Production Possibility Frontier shows the frontier of the feasible set of non-health commodities (Z) health stock (H).

Since an individual dies and cannot consume/produce any non-health commodities, we must have $Z_i = 0$ when $H_i \leq H_{min}$.

This is a WRONG example:

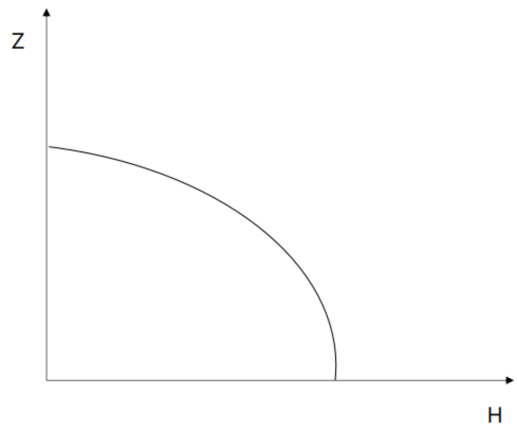


Figure 3.22: Wrong PPF

A CORRECT example and Optimality:

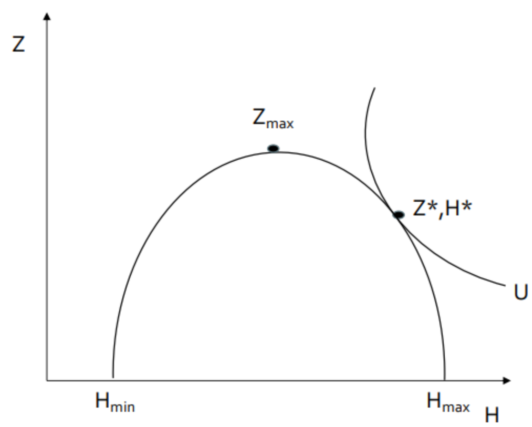


Figure 3.23: Correct PPF and the Optimal Point

We must have H_{min}, Z_{max}, and H_{max}. We reach the optimal point (Z^*, H^*) when the indifference curve is tangent to the PPF. In the following 2 sections, we will explain this is also level of health stock where MEC and COC curves intersect.

3.3.6 MEC and COC Curves

The Marginal Efficiency of Health Capital (MEC) curve shows the value of the additional healthy days gained from a 1-unit increase in the health stock (MPH). MEC is the monetary value of the marginal productivity of health stocks. It is diminishing as an individual's health stock expands.

MEC has the following expression:

$$MEC = MPH \times w$$

where:

- w: wage rate
- MPH: marginal productivity of health stock

The Cost of Capital (COC) curve represents the supply price of capital or the cost of holding an additional unit of health capital. Its position is determined by the depreciation rate of the health stock (δ) and the opportunity cost of holding health capital (r):

$$COC = r + \delta$$

3.3.7 Optimisation, Demand for Health, and Ageing

Optimisationa and Demand for Health

Optimality Conditions for the simplest version of the model:

PV of the Marginal Cost of Gross Health Investment = PV of Marginal Benefits of Health

We have seen this is the tangent point of the FPF and indifference curve. Meanwhile, this is also the point where the Marginal Efficiency of Health Capital (MEC) curve intersects the Cost of Capita (COC) curve:

$$\underbrace{r+\delta}_{MC\ (COC)} = \underbrace{MPH \times w}_{MB\ (MEC)}$$

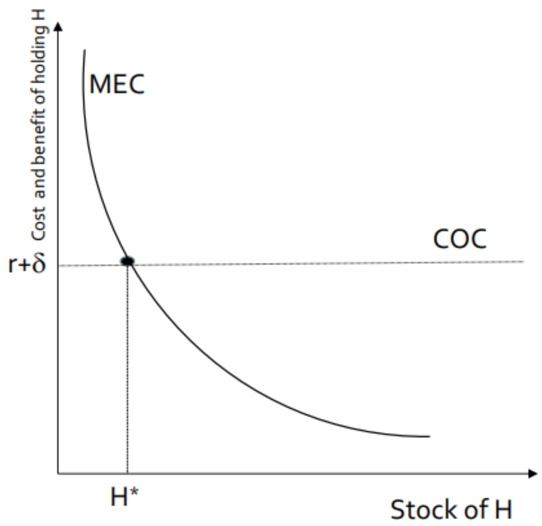


Figure 3.24: MEC and COC of Holding Health Stocks and the Optimal Point

What Happens as We Age?

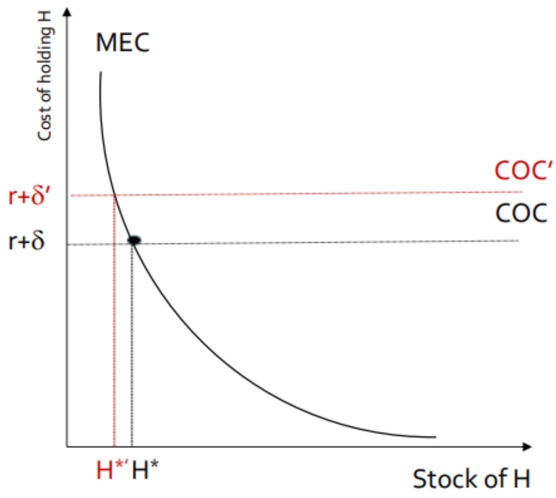


Figure 3.25: Optimality and Aging

The depreciation rate increases ($\delta \to \delta'$) as we age, so the COC shifts upwards to COC', and the optimal level of health stock we can sustain is lower ($H^* \to H^{*'}$).

Health Inequality

See section 3.4.4.

3.3.8 Literature Development and Criticisms

Since 1972, the Grossman model has been the cornerstone for modelling investment in health capital, spurring a great deal of research, extensions, empirical testing and criticism. Grossman himself reviewed the literature many times, e.g. in the 2000 Chapter "The Human Capital Model" in the Handbook of Health Economics.

Some of the main criticisms of the Grossman model:

- It does not preclude an individual choosing to live forever (Ehrlich and Chuma, JPE 1990 "does not determine the length of life")
- It does not provide an adequate conceptual framework for the SES (socioeconomic status)-health gradient (Galama and van Kippersluis, EJ, 2018): They develop a model which incorporates health, longevity, wealth, earnings, education, work, job-related physical and psychosocial health stressors, leisure, health investment (e.g. exercise, medical care) and healthy and unhealthy consumption (including housing, neighbourhood social environment).
- Not faithful to gerontological models of health deficit accumulation (Strulik).
- It does not include an early childhood phase (Heckman, PNAS 2007).

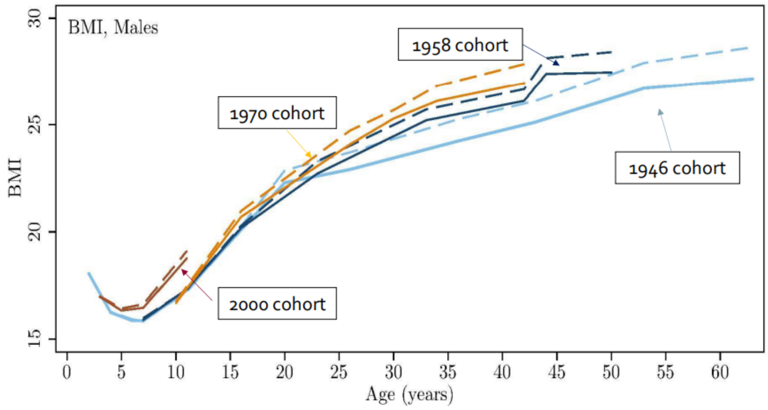
It is a very active area of research, both theoretically and empirically.

3.4 Inequalities in Health

Inequalities in health are not random and are associated with characteristics. Poor people are less healthy, maybe because they eat cheaper food and don't know the benefit of doing exercise (less educated). Education is correlated with health.

More educated → Higher income → More resources and information

3.4.1 SES Inequalities in BMI Widen over the Lifecycle and across British Cohorts



Note: Solid line: high SES; dashed line: low SES. Source: Conti, Mason and Poupakis, 2019.

Figure 3.26: SES Inequalities in BMI Widen over the Lifecycle and across British Cohorts

BMI (Body Mass Index) is not the best measure of body fitness, since it only includes two dimensions (weight and height), it doesn't take into account fat mass, it doesn't include many diseases such as blindness, it is developed in the West and thus the norms may not apply to other ethnicities. However, BMI is a relatively *simple measure*, and it is correlated with obesity and worse body conditions later in life.

Lifecycle:

- The lines are very close in the beginning, the 1946 cohort starts to diverge at the end of adolescence and keeps diverging as they get older.
- Low SES people have higher BMIs than high SES people as they get older.
- For each cohort, BMI generally increases as age increases.

Cohort:

- 2000 cohort has a higher BMI at the young age.
- The health inequality (divergence) presents in every cohort, and we are not doing anything to effectively reduce it.

3.4.2 Health inequalities are everywhere

- Researchers have documented inequalities in the distribution of health by socioeconomic status, gender, and ethnicity.
- Research on socio-economic inequalities in health in the UK has a long history, together with nice data.
- In the early part of the 20th century the British government introduced questions on occupation in the decennial census:
- The 1970-1972 Decennial Supplement of occupational Mortality (OCPS) showed that men in social class V (unskilled) were 2.5 times as likely to die before the age of 65 than those in social class I (professional). Children in social class V families were twice as likely to die as those in social class I.

3.4.3 Landmark Studies in Social Class Inequalities in Health in the UK

Brief Listing

- Black Report (1980)

  - Health inequalities were widening, the problem had little to do with NHS.
  - Four possible explanations: (1) data artefact; (2) social selection (sicker people ⇒ less able to study and work ⇒ poorer); (3) behaviour; (4) material circumstances.
  - Policy recommendation: reduce poverty, spend more on prevention.

- Whitehead Report (1987)
  - Commissioned by the Health Education Council (HEC) and headed by Margaret Whitehead.
  - Health inequalities widened since the Black Report.
  - HEC was scrapped: was campaigning on alcohol, tobacco and diet issues which upset some of the government's financial supporters.
- Acheson Report (1998)
  - Commissioned by the new Labour (Blair) government in 1997, under the chairmanship of a former Chief Medical Officer, Sir Donald Acheson.
  - Similar findings and recommendations as the Black Report: the root cause of inequalities in health was poverty.
- Whitehall I Study
  - Examined over 18,000 male British civil servants over 10 years, starting in 1967.
  - Showed that mortality was higher among those in the lower grade than in the higher grade – for all causes and CHD.
  - Controlling for risk factors (smoking, obesity, exercise, blood pressure) accounted for 40 percent of the gradient.
- Whitehall II Study
  - Examined 10,308 civil servants aged 35-55, 2/3 men and 1/3 women, starting in 1985.
  - Showed social gradient for several diseases.
  - Added job-related stress to the traditional risk factors for low social class.
- Fair Society, Healthy Lives: The Marmot Report (2010)
  - Concluded that reducing health inequalities would require action on six policy objectives: number 1 is "Give every child the best start in life".
- Health Equity in England: The Marmot Review 10 Years On (2020)
  - "This '10 years on' report shows that, in England, health is getting worse for people living in more deprived districts and regions, health inequalities are increasing and, for the population as a whole, health is declining. The data that this report brings together also show that for almost of all the recommendations made in the original Marmot Review, the country has been moving in the wrong direction. In particular, lives for people towards the bottom of the social hierarchy have been made more difficult. Some of these difficulties have been the direct result of government policies, some have resulted from failure to counter adverse trends such as increased economic inequalities or market failures."

The Marmot Review

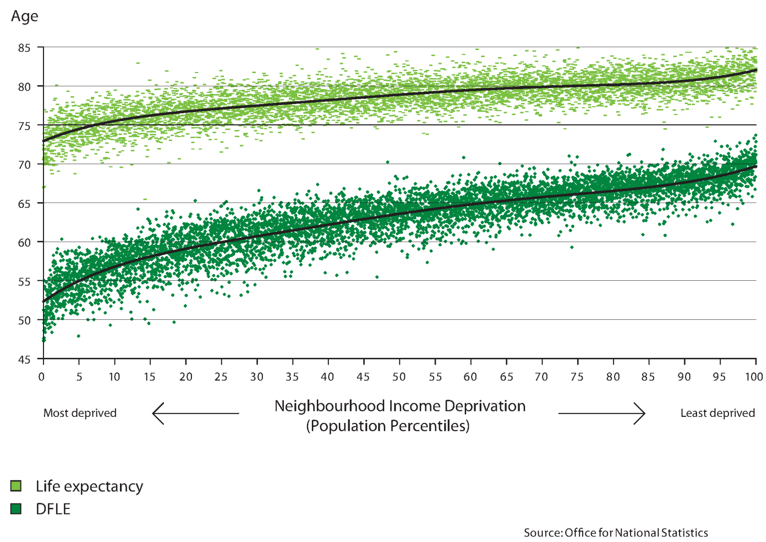


Figure 3.27: Life expectancy and disability-free life expectancy (DFLE) at birth, by neighbourhood income level, England, 1999-2003

- Disability-Free Life Expectancy (DFLE) measures not only how long you live but also how well you live, so there is a gap between life expectancy and DFLE.
- There is a discrepancy in the life expectancy/DFLE between the most deprived and the least deprived people in England.

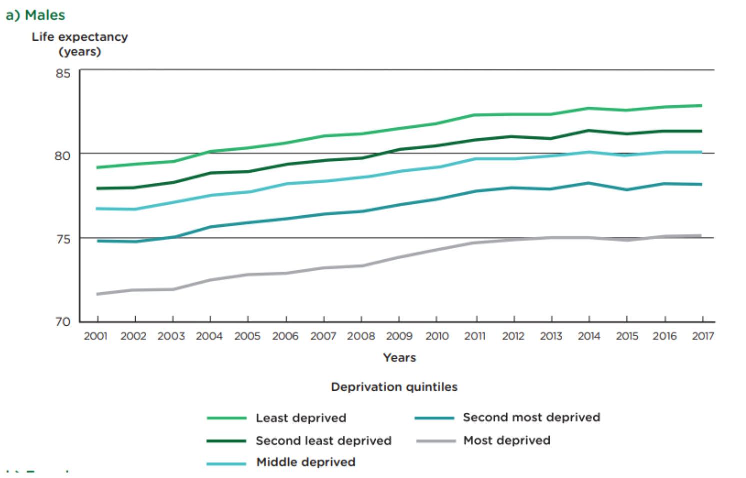


Figure 3.28: Life expectancy at birth by area deprivation quintiles and sex, England, 2003–05 to 2015–17

- Life expectancy is increased for everyone.
- The most deprived has the lowest life expectancy.
- The gaps between groups are almost the same as time passed, so inequalities stay the same.

Public Health England

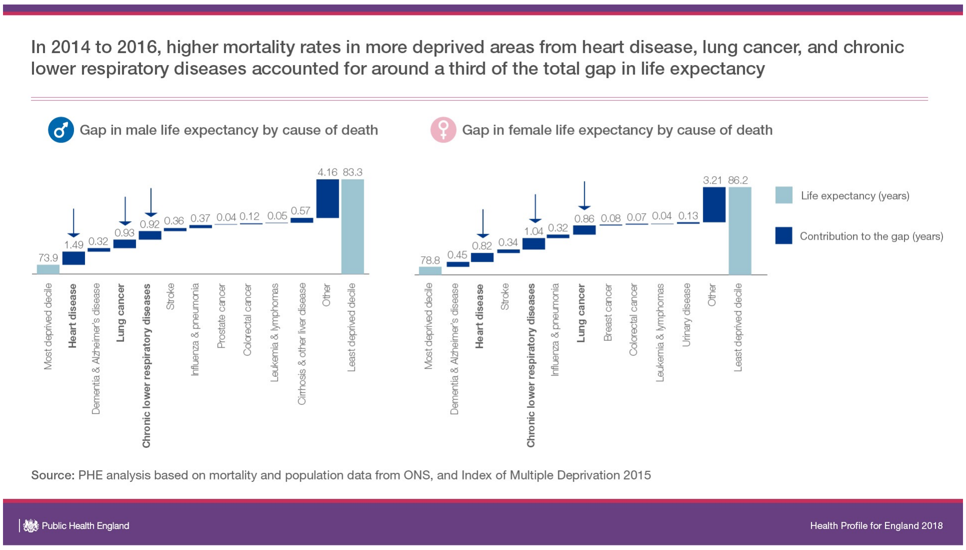


Figure 3.29: What causes of death are driving the gap in life expectancy among the most deprived people and the least deprived people, male/female

In 2014 to 2016, higher mortality rates in more deprived areas from heart disease, lung cancer, and chronic lower respiratory diseases accounted for around a third of the total gap in life expectancy.

The IFS Deaton Review (2019-)

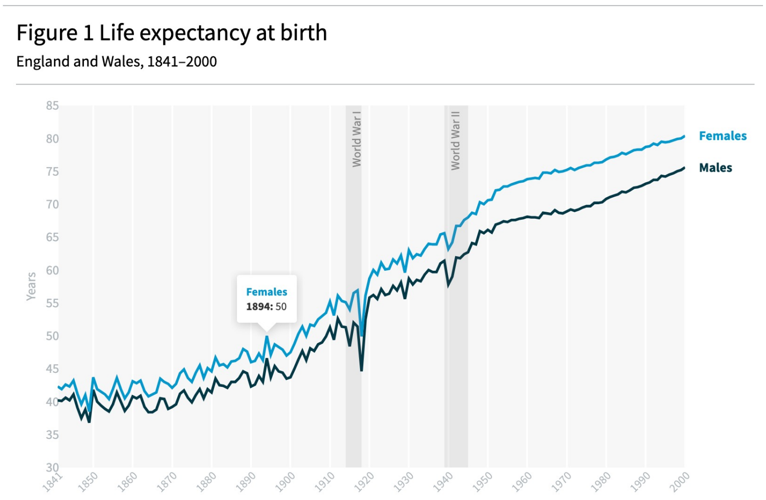


Figure 3.30: Life expectancy at birth, England and Wales, 1841-2000, Males/Females

The life expectancy increased and almost doubled. Females had higher life expectancy. The biggest drop was in World War II.

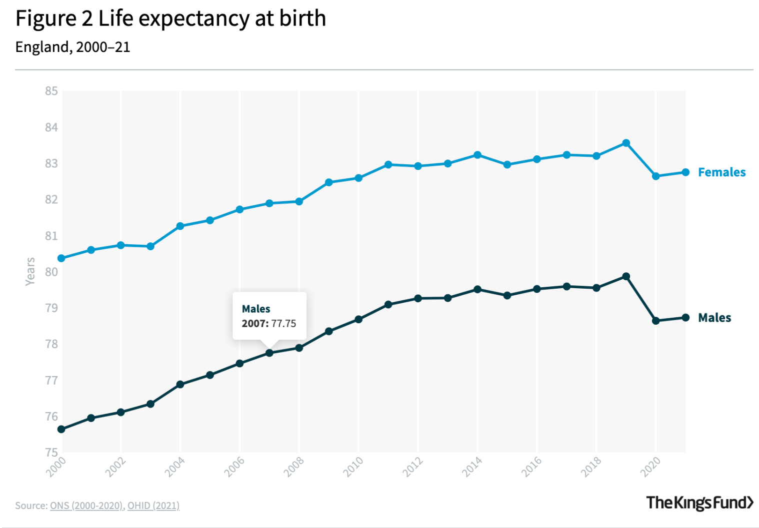


Figure 3.31: Life expectancy at birth, England, 2000-2021, Males/Females

The second biggest drop was during the pandemic (about 1 year).

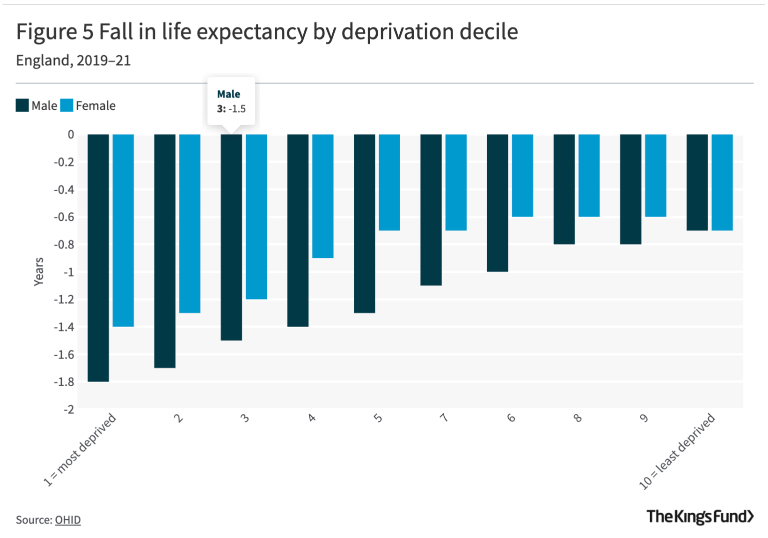


Figure 3.32: Fall in life expectancy by deprivation decile, England, 2019-21

During the Covid-19 pandemic, the most deprived male had a life expectancy drop of 1.8 years, while the least deprived male had a life expectancy drop of 0.7 years. Males tend to have a bigger drop than females.

3.4.4 Why Do Health Disparities Exist?

Understanding the causes of health disparities is of key policy importance for addressing them (health subsidies, health information, nudge, for the poor people).

Potential Causes are:

1. SES can causally affect health. In terms of the Grossman model (Section 3.3.7):
   - In the framework of the Grossman model, SES could actually affect health. Firstly, people with better SES (e.g. higher education) can have higher Marginal Efficiency of Health Capital (MEC), for instance, due to higher wages:

$$MEC \uparrow = MPH \times w \uparrow$$

- Also, they may have a lower rate of depreciation δ due to less stress. This will shift down their COC:
$$COC \downarrow = r + \delta \downarrow$$
- Higher MEC, lower COC $\rightsquigarrow$ Higher equilibrium health stock
- Besides, people with better SES typically have more wealth. Hence, their FPF could be expended compared with disadvantaged people, which also leads to higher health stocks.
- Meanwhile, both health and SES can be affected by third factors. For example: time preferences (Fuchs' hypothesis): people who are willing to delay gratification will invest more in both health and education.

2. Early health can causally affect SES (Health Selection).
For example, unhealthy child $\rightsquigarrow$ less educated $\rightsquigarrow$ lower SES $\rightsquigarrow$ poorer adult health

3.5 Education and Health: Causal Inference

3.5.1 Observed Difference in Health by Education (Compulsory vs. Post-Compulsory)

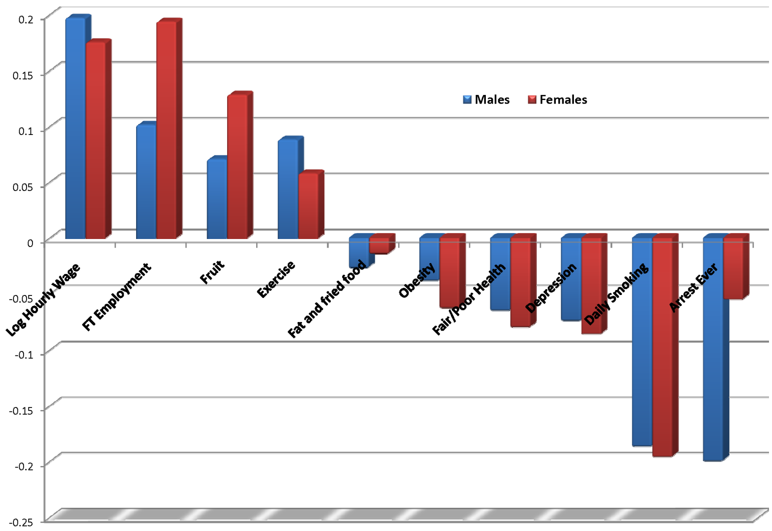


Figure 3.33: Disparities by Education (compulsory vs. post-compulsory)

The graph shows the raw differentials in the outcomes between individuals with post-compulsory and compulsory level of education. Individuals with post-compulsory education tend to have better health.

3.5.2 Causality or Correlation? An Overview of Empirical Strategies

Key question: Is the positive correlation between health and schooling causal?

Literature has used different approaches to establish causality:

- First set of studies: used instrumental variables with questionable instruments for education (e.g. quarter of birth), finding strong effects.
- Second set of studies: exploited features of the educational system:
  - Lleras-Muney (RevStud, 2005) uses compulsory school and child labour laws in thirty states from 1915 to 1939 as instruments for education and finds a significant effect in reducing mortality (using DiD).
  - Clark and Royer (AER, 2013) use regression discontinuity methods (RDD) exploiting two changes to British compulsory schooling laws and find no effects on mortality or other measures of health (discussed in detail in Section 3.5.3). Similar results in a recent paper by Janke, Johnston, Propper et al. (JHE, 2018).

- Third set of studies: twin differences (different between twins, NOT DiD!) (Lundborg, JPopEc 2013). It suggests that completing high school improves health, but additional schooling does not lead to additional health gains. Controlling for certain early life factors that may vary within twin pairs does not alter the main conclusions. Discussed in detail in Section 3.5.4.
- Fourth set of studies: more "structural" models: Conti et al., 2010a,b show that education has a causal component in most outcomes, but with dramatic heterogeneity; Heckman et al., 2018 sets up a dynamic model and show that completing high school is especially effective. See section 3.5.6.
- Recent literature: disagreement still exists. See section 3.5.7.
- Summary: Grossman (NBER wp 21609, 2015) "The Relationship between Health and Schooling: What's New?" concludes "There is enough conflicting evidence in the studies that I have reviewed to warrant more research on the question of whether more schooling does in fact cause better health outcomes."

3.5.3 Clark and Royer: RDD (AER, 2013)

Setup

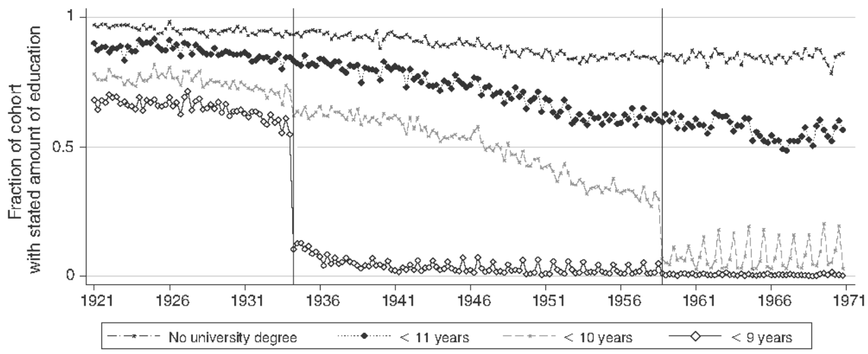

VOL. 103 NO. 6 CLARK AND ROYER: EFFECT OF EDUCATION ON MORTALITY AND HEALTH 2092

Figure 1. Years of Full-Time Education by Quarter of Birth

Notes: Sample includes individuals in the Health Survey for England: 1991–2004. Points represent means among people in each quarter-year of birth cell (all later graphs present data by month-year of birth). The vertical lines are cutoffs corresponding to the first cohorts subject to the new compulsory schooling laws. The first of these took effect on April 1, 1947 and the second took effect on September 1, 1972. Thus, since the two compulsory schooling reforms affected 14 (first reform) and 15 (second reform) year olds, the first cohorts impacted are those born in April 1933 for the first reform and September 1957 for the second reform.

Figure 3.34: Years of Full-Time Education by Quarter of Birth

This figure presents data at the quarter-of-birth level using Health Survey of England data. We study how the introduction of two compulsory schooling laws affected the fraction of cohorts with stated amount of education:

- The first reform took effect on April 1, 1947, which increased the compulsory school leaving age from 14 to 15.
- the second took effect on September 1, 1972, which increased the compulsory school leaving age from 15 to 16.

The vertical lines are cutoffs corresponding to the first cohorts subject to the new compulsory schooling laws. Since the two compulsory schooling reforms affected 14 (first reform) and 15 (second reform) years old, the first cohorts impacted are those born in April 1933 for the first reform and September 1957 for the second reform.

The 1947 change reduced the fraction that completed nine years or less by roughly one half; the 1972 change decreased the fraction that completed ten years or less by roughly one quarter. Hence, the reforms did increase students' years of education and reduce the number of students who dropped out at a particular age.

Nature of Discontinuity

VOL. 103 NO. 6 CLARK AND ROYER: EFFECT OF EDUCATION ON MORTALITY AND HEALTH 2094

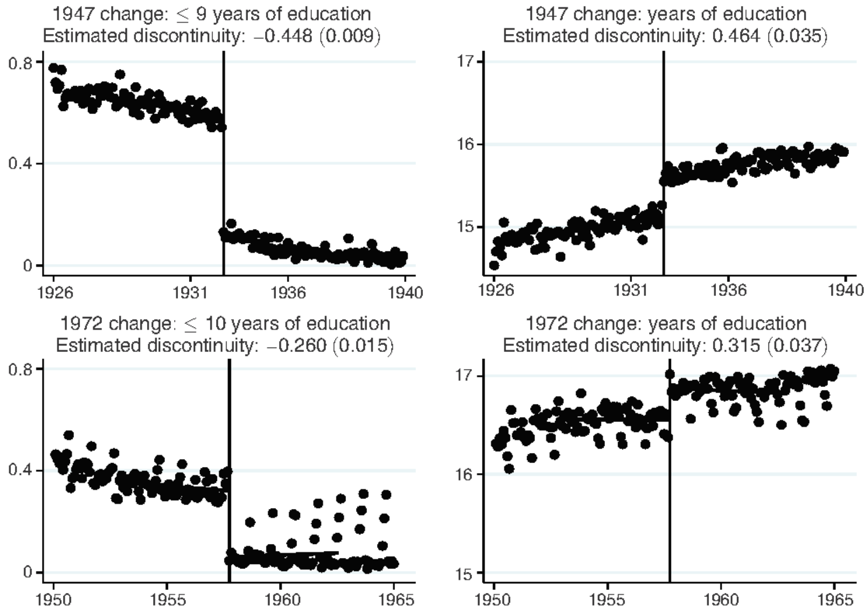


FIGURE 2. THE IMPACT OF THE COMPULSORY SCHOOLING CHANGES ON EDUCATIONAL ATTAINMENT

Notes: Samples are based on pooled General Household Survey and Health Survey for England data. Points represent means among people in each month-year of birth cell. The estimated discontinuities are based on local linear regressions; standard errors are in parentheses. The fitted values of these local linear regressions are also plotted.

Figure 3.35: The Impact of the Compulsory Schooling Changes on Educational Attainment

Here, our running variable is the year of born (cohorts): at certain years of born, there are "jumps" in average years of education.

Outcomes

VOL. 103 NO. 6 CLARK AND ROYER: EFFECT OF EDUCATION ON MORTALITY AND HEALTH 2104

FIGURE 3. THE IMPACT OF THE COMPULSORY SCHOOLING CHANGES ON MORTALITY

Notes: The log odds ratio is defined as the logarithm of the odds of dying for the relevant cohort relative to the January 1926 cohort for the 1947 reform and relative to the March 1950 cohort for the 1972 reform. Points represent the log odds death ratio for each month-year of birth cell. The estimated discontinuities are based on local linear regressions; the standard errors of the estimates are presented in parentheses. The fitted values of these local linear regressions are also plotted.

Figure 3.36: The Impact of the Compulsory Schooling Changes on Mortality

Mortality: The two compulsory schooling laws have no effect on mortality (no discontinuity).

VOL. 103 NO. 6 CLARK AND ROYER: EFFECT OF EDUCATION ON MORTALITY AND HEALTH 2110

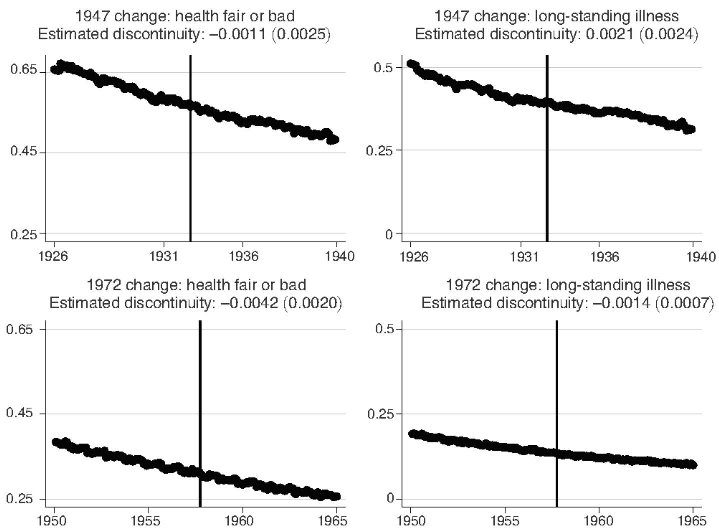


FIGURE 4. THE IMPACT ON REPORTING BEING IN FAIR OR WORSE HEALTH IN THE 2001 CENSUS

Notes: Samples are based on 2001 census data. Points represent means among people in each month-year of birth cell. The estimated discontinuities are based on local linear regressions; standard errors are in parentheses. The fitted values of these local linear regressions are also plotted.

Figure 3.37: The Impact on Reporting Being in Fair or Worse Health in the 2001 Census

Health Fair/Bad or Long-standing Illness: The reforms cause no change in health fair or bad and long-standing illness.

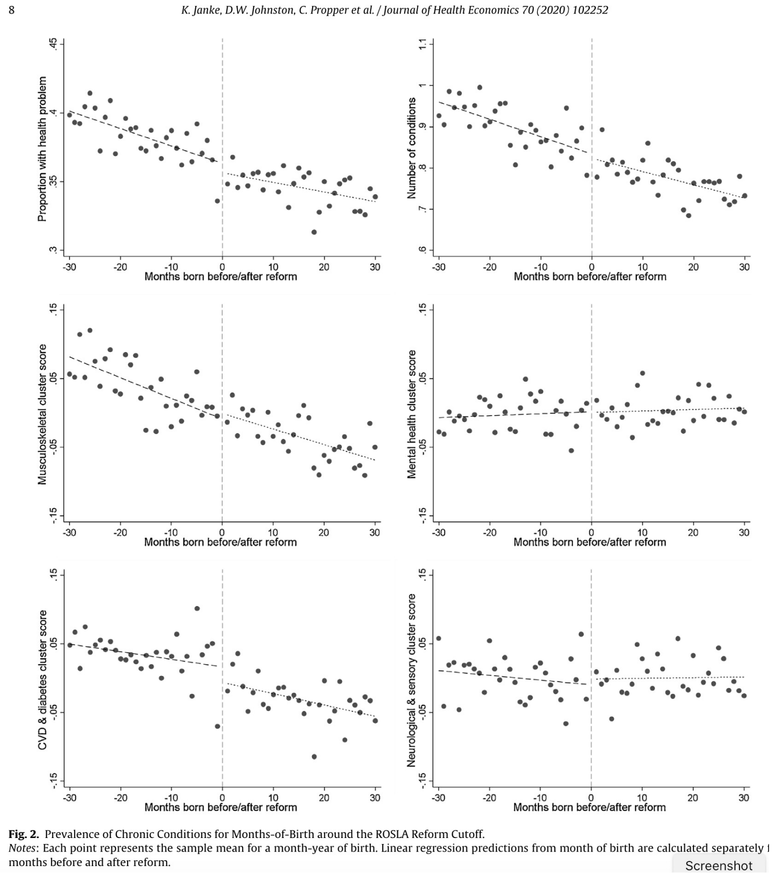


Figure 3.38: Prevalence of Chronic Conditions for Months-of-Birth around the ROSLA Reform Cutoff

Other outcome variables: Janke, Johnston, Propper et al. (Journal of Health Economics, 2020) analysed the same reform and showed that only the prevalence of cardiovascular diseases was affected by the reform.

Conclusion

Higher education attainment shows nearly no effect on health conditions (at least at the cutoff).

Extension: What If We Run an OLS?

K. Janke, D.W. Johnston, C. Propper et al. / Journal of Health Economics 70 (2020) 102252 9

Table 1
ROSLA Reform: Estimates of the Impact of Age Completed Full-Time Education on Chronic Condition Summary Measures.

	OLS Age completed education (1)	OLS Affected by reform (2)	2SLS Age completed education (3)
Summary Measures			
At least one condition	−0.075***	0.002	0.005
	[−0.085,−0.064]	[−0.016,0.019]	[−0.047,0.056]
At least one limiting condition	−0.073***	−0.000	−0.001
	[−0.083,−0.063]	[−0.015,0.015]	[−0.044,0.042]
Number of conditions	−0.347***	0.006	0.019
	[−0.390,−0.304]	[−0.037,0.050]	[−0.108,0.146]
Condition Clusters			
Musculoskeletal conditions	−0.184***	0.034*	0.100
	[−0.212,−0.156]	[−0.005,0.072]	[−0.029,0.228]
Mental health conditions	−0.128***	0.003	0.009
	[−0.159,−0.097]	[−0.041,0.047]	[−0.115,0.132]
CVD & diabetes conditions	−0.060***	−0.050***	−0.147***
	[−0.081,−0.040]	[−0.084,−0.015]	[−0.241,−0.053]
Neurological & sensory conditions	−0.003	−0.004	−0.011
	[−0.032,0.026]	[−0.040,0.033]	[−0.115,0.092]
Sample size	109032	359284	359284

Notes: OLS figures in column (1) are coefficient estimates on age completed full-time education using only the cohorts born prior to the reform (1 September 1957) and individuals who completed full-time education ≤ aged 16 years. OLS figures in column (2) are coefficient estimates on a binary variable indicating whether the individual was born after 1 September 1957 (treatment indicator). 2SLS figures in column (3) are coefficient estimates on age completed full-time education. Control variables include a linear function of month-year of birth, wave number, month of birth (Jan–Dec), and interactions between month-of-birth dummies and the treatment indicator. The condition cluster variables are predicted scores from a principal-components factor analysis with orthogonal rotation of the 17 condition dummy variables. Each condition cluster score has a mean of zero and a standard deviation of one. 95% confidence intervals are presented in brackets; calculated allowing for clustering at the month-year of birth level. *, ** and *** signify *p*-values less than 0.10, 0.05 and 0.01, respectively.

Figure 3.39: ROSLA Reform Estimates

If we directly run an OLS regression of education on health, we can see significant correlations. However, if we use 2SLS (Fuzzy RDD), we see that the coefficients are close to 0.

3.5.4 Lundborg, Journal of Population Economics 2013: Twin Differences

The usual setup in twin studies:

$$Y_i = \alpha_0 + \alpha_1 eduyrs_i + \alpha_2 X_i + \eta_i$$

where Y_i is certain health outcome.

Take the difference between twins' outcomes:

$$\Delta Y = \beta_0 + \beta_1 \Delta eduyrs + \beta_2 \Delta X + \Delta\eta$$

where $\Delta Y = Y_i - Y_j$

This is a fixed-effect regression, which eliminates effects of common characteristics between twins.

However, β_1 cannot not be interpreted as the causal effect of education on health, as education is subject to selection and is likely to be correlated with the error term (endogeneity: ability/effort/parents' investment), so we still need an instrument for $\Delta eduyrs$.

To address this concern, Lundborg exploits the unusually rich MIDUS data and examine to which extent his results are robust to controlling for differences in early life factors, such as parental treatment, within twin pairs.

Lundborg concludes that completing high school improves health, as measured through self-reported health, chronic conditions, and exercise behavior, but that additional schooling does not lead to additional health gains. Controlling for certain early life factors that may vary within twin pairs does not alter the main conclusions.

3.5.5 Explaining/Decomposing the Education-Health Gradient

Cutler and Lleras-Muney (JHE, 2010)

- Cutler and Lleras-Muney (JHE, 2010) examine possible explanations for the relationship between education and health, using several datasets from US and UK. They use OLS regression and see how much of the coefficient goes down when adding additional covariates.
- They are able to explain two-thirds of the gradient.
- "Economic resources" is able to explain 32 percent of the gradient. "Specific knowledge" is able to explain 12 percent of the gradient (US dataset). "Cognitive ability" seems to explain a lot in the UK dataset. "Tastes" and "Personality" explain a little and "social integration" explains somehow.

Table 12
Share of education gradient explainable by different factors.

Factor	Explanatory power					
	NHIS	HRS	NLSY	MIDUS	NCDS	Approximate summary
Economic resources	32%	17%	12%	11%	24%	20%
Additional reduction when add:						
Specific knowledge	12%	NA	NA	NA	NA	12%
Cognitive ability	NA	NA	15%	NA	44%	30%
Tastes	NA	0%	NA	1%	2%	1%
Personality	4%	NA	4%	1%	2%	3%
Social integration	NA	NA	NA	7%	15%	11%

Note: Based on the results in the previous tables. The table reports mortality weighted reductions (see text for explanation).

Figure 3.40: Share of education gradient explainable by different factors

Conti and Hansman (JHE, 2013)

- Conti and Hansman (JHE, 2013) test the robustness of Cutler and Lleras-Muney's (CLM) results by using alternative measures of child personality available in the National Child Development Study:
  - the Rutter Behavior Scale (ages 7, 11 and 16)
  - the British Social Adjustment Guide (BSAG, ages 7 and 11):
- CLM results show that adult personality explains very little of the education-health gradient, but Conti and Hansman (JHE, 2013) show that child personality contributes the education-health gradient to an extent nearly as large as cognition.
- Childhood matters.

Behaviour	Cognition	CLM Personality	CH Rutter+BSAG	CH Motivation
Current smoker	48%	4%	23%	48%
Former smoker	25%	0%	10%	135%
No. cigs smoked	17%	0%	-2%	29%
BMI	22%	7%	23%	8%
Eat fruit daily	12%	5%	16%	21%
No. drinks/w	-19%	10%	13%	82%

Figure 3.41: How much education-health gradient is explained by cognition, personality and motivation

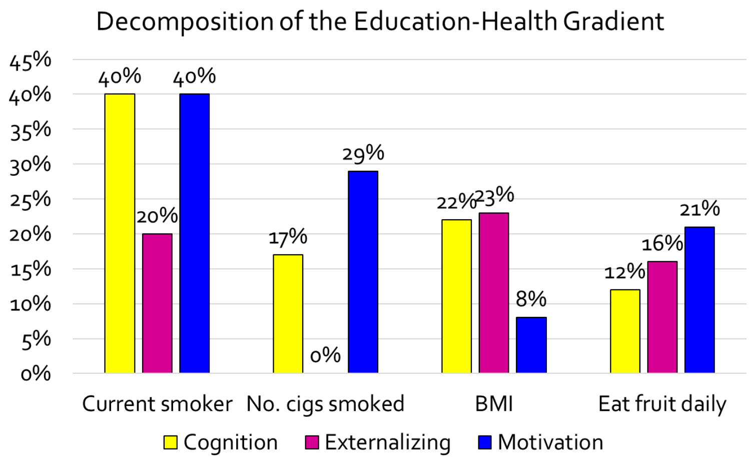


Figure 3.42: Decomposition of the Education-Health Gradient

Non-cognitive factors such as externalising and motivation are at least as important as cognition in explaining the gradient.

3.5.6 Structural Models in Education-Health Gradient Decomposition

The Early Origins of the Gradient: Conti, Heckman et al. (AER, PPS, 2010a,b)

- Conti, Heckman et al. (AER, PPS, 2010a,b) decompose observed educational differentials into causal components of education and components due to selection on early childhood human capital.
- In their framework, early childhood factors affect health both through education and directly. Both channels are quantified.
  - They estimate a semiparametric structural model of the choice of schooling (decision to stay on at 16) and the causal effect of schooling on a variety of outcomes at age 30:
    - labour market (wages and employment)
    - health status (self-reported health, depression and obesity)
    - health behaviours (smoking, exercise)
  - Three childhood (age 10) endowments:
    - cognition (e.g. British Ability Scales)
    - noncognitive traits (e.g. locus of control)
    - health (height, head circumference)
- They look at the mean effect of education on health (like previous literature) and also at heterogeneity in treatment effects for people with different early childhood endowments.
- Data: British Cohort Study (BCS70), a cohort of all individuals born in one week of April 1970 in the United Kingdom.

Outcomes:

- Causal component of education and effects of childhood characteristics: educations and early childhood characteristics both have causal effects on most of the outcomes.

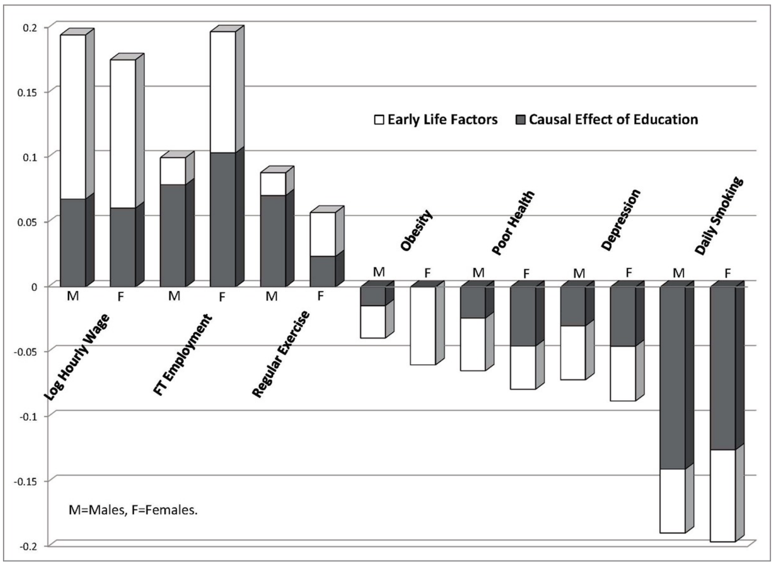


Figure 3.43: How education and early life factors share the causal effects on different health outcomes

- Specific Outcomes
  - Obesity

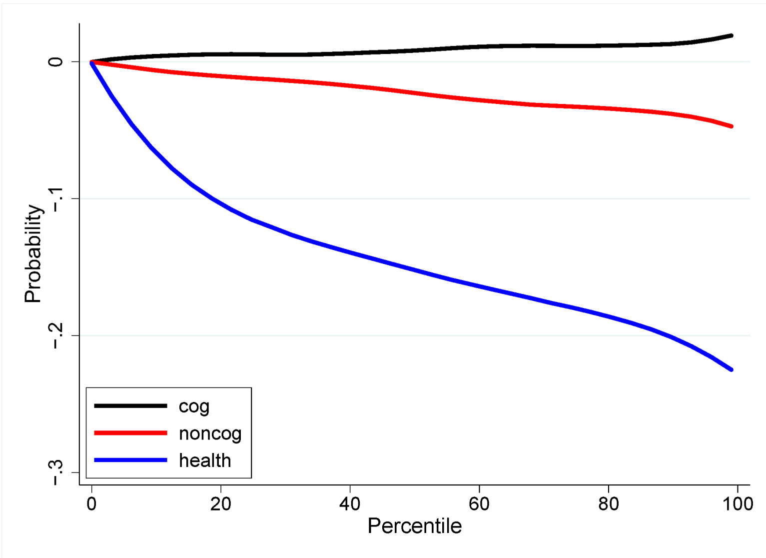


Figure 3.44: The Effects of Childhood Endowments on the Probability of being Obese at Age 30

Cognition matters little. Non-cognitive explains a little bit. Early physical health is the most important determinant of obesity.

  - Smoking

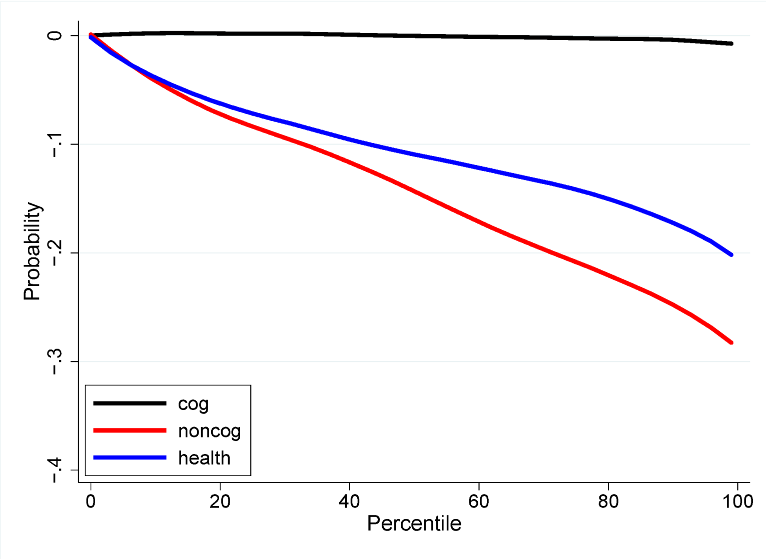


Figure 3.45: The Effects of Childhood Endowments on the Probability of being a Daily Smoker at Age 30

Cognition doesn't affect smoking. Non-cognitive factors (e.g. self-regulation and motivation) and physical health are equally important determinants of smoking.

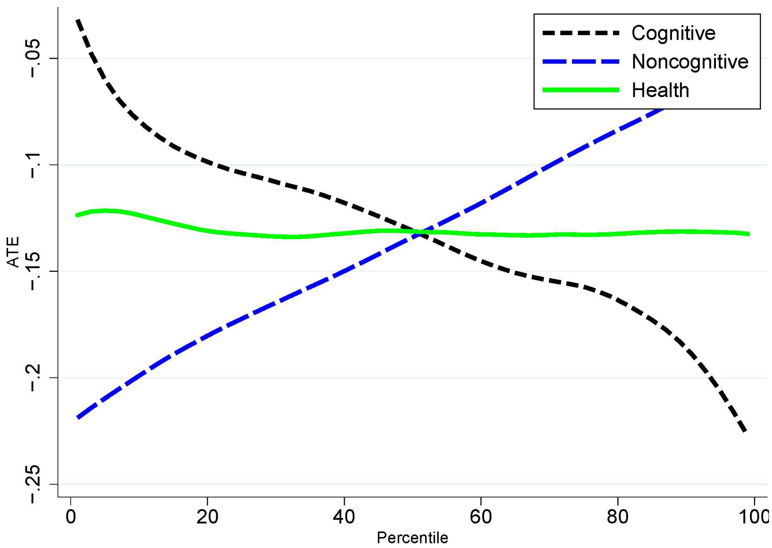


Figure 3.46: Heterogeneity in the Effects of Education on Smoking

Effects of education on smoking is heterogeneous: beneficial effect of education is much bigger at the top of the cognitive ability distribution

Evidence from a Dynamic Model: Heckman, Humphries and Veramendi (JPE 2018)

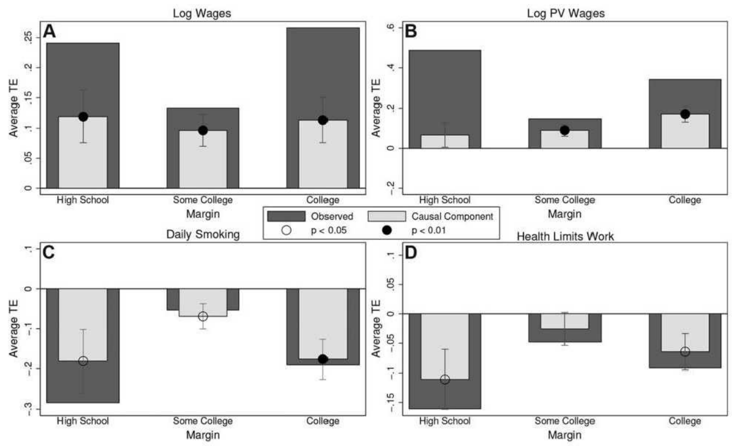


Fig. 3.—Causal versus observed differences by final schooling level (compared to next-lowest level). These figures report pairwise treatment effect, (18), for the indicated schooling nodes. Each bar compares the mean outcomes from a particular schooling level j and the next-lowest level $j-1$ defined for the set of persons who complete schooling at $j-1$ or j. The "Observed" bar displays the observed differences in the data. The "Causal Component" bar displays the estimated average treatment effect to those who get treated (ATE) for the indicated group. The difference between the observed and causal treatment effects is attributed to the effect of selection and ability. Selection includes sorting on gains. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars show one standard deviation and correspond to the 15.87th and 84.13th percentiles of the bootstrapped estimates, allowing for asymmetry. Significance at the 5 percent and 1 percent levels is shown by open and filled circles on the plots, respectively.

Figure 3.47: Evidence from a Dynamic Model

Heckman, Humphries and Veramendi (JPE 2018) extend the structural framework discussed above into a dynamic model of schooling choice and estimate causal effects from multiple levels of schooling. They find substantial continuation value of schooling for high-ability individuals, but not for low ones beyond high school.

- The figure shows how additional level of education affects wages and health compared to previous education level
- The black bars labelled "Observed" are the raw differences found in the data, and the gray bars are estimated causal components.
- The estimated average causal effects are the highest for high schools.

3.5.7 The State of the Literature

- Recent review "The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence" by T. Galama, A. LlerasMuney, H. van Kippersluis, Oxford Research Encyclopedia of Economics and Finance (2018):
  - "There is no convincing evidence of an effect of education on obesity, and the effects on smoking are only apparent when schooling reforms affect individuals' track or their peer group, but not when they simply increase the duration of schooling."
  - "An effect of education on mortality exists in some contexts but not in others, and seem to depend on: (a) gender; (b) the labour market returns to education; (c) the quality of education; (d) whether education affects peers' quality."
- Important step forward would be to understand the quality and content of further education, years beyond the minimum school leaving age, and both short-, medium-, and long-term outcomes.
- Some papers analyse the publication bias towards the positive effects of education on health.

3.5.8 Where Next?

- Grossman (NBER wp 21609, 2015) "The Relationship between Health and Schooling: What's New?" concludes "There is enough conflicting evidence in the studies that I have reviewed to warrant more research on the question of whether more schooling does in fact cause better health outcomes."

- Among promising areas of current and future research:
  - Does schooling quality matter?
  - What are the mechanisms via which schooling influences health and health behaviours?
  - What is the role of genes?
  - What is the role of subjective expectations of returns?

Bibliography

- Barcellos et al. (2018). PNAS. Education can reduce health differences related to genetic risk of obesity.
- Clark, D. and Royer, H. (2013). The Effect of Education on Adult Mortality: Evidence from Britain. American Economic Review, 106(6), 2087-2120.
- Conti, G., and Hansman, C. (2013). Personality and the education–health gradient: A note on "Understanding differences in health behaviours by education". Journal of health economics, 32(2), 480-485.
- Conti, G. and Heckman, J.J. (2010). Understanding the Early Origins of the Education-Health Gradient: A Framework That Can Also Be Applied to Analyze Gene-Environment Interactions. Perspectives on Psychological Science, 5(5), 585-605.
- Conti,G., Heckman,J.J. and Urzua, S. (2010).The Education-Health Gradient. American Economic Review P P, 100(2), 234-238.
- Cutler, D. M., and Lleras-Muney, A. (2010). Understanding differences in health behaviours by education. Journal of health economics, 29(1), 1-28. • [Erlich, I., and H. Chuma (1990) "A model of the demand for longevity and the value of life extensions", Journal of Political Economy 98:761-782.]
- Galama,T. and H. van Kippersluis (2018).A Theory of Socioeconomic Disparities in Health Over the Life Cycle, Economic Journal.
- Galama, Titus J. and Lleras-Muney, Adriana and van Kippersluis, Hans, The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence, Oxford Encyclopedia of Economics and Finance.
- Gilleskie, D. (2008). Health Capital: Theory and Empirical Evidence. Incentives and Choice in Health Care MIT Press: Cambridge, MA.
- Grossman, M. (1972). On the concept of health capital and the demand for health. Journal of Political economy, 80(2), 223-255.
- Grossman, M. (2000). The human capital model. Handbook of health economics, 1, 347-408.
- Grossman, M. (2015).The Relationship between Health and Schooling: What's New? NBER wp 21609.
- Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. Proceedings of the national Academy of Sciences, 104(33), 13250-13255.
- Heckman, J.J., J.E. Humphries and G. Veramendi (2018). Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking. Journal of Political Economy, October S1.
- Janke, K., D.Johnston, C. Propper and M. Shields (2018).The Causal Effect of Education on Chronic Health Conditions. IZA DP 11353.
- Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. The Review of Economic Studies, 72(1), 189-221.
- Lundborg, P. (2013).The health returns to schooling—what can we learn from twins? Journal of population economics, 26(2), 673-701

Chapter 4

The Health Effects of Early Interventions

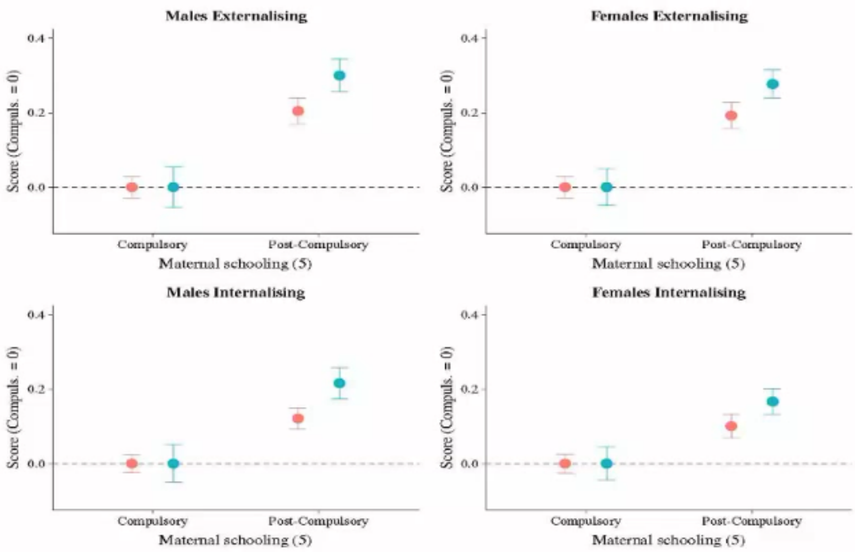
4.1 Questions we are interested in and motivation

◇ We saw the substiantial SES-health gradient last time

◇ Moreover, Attanasio, Blundell. Conti and Mason (2019) showed that this inequality is multidimentional. Inequality in socio-emotional skills (mental health) of 5 year olds has also widened over time.

  * The score on various socio-emotional skills were normalised to 0 for the 1970 cohort, and we can see from figure 4.1 that kids born to more educated mothers are better off in 2000 compared to 1970 (gap became bigger).

  * Interpreting the fig: externalising = social skills; internalising = the ability to focus; higher value is better



Note: orange=1970 cohort; green=2000 cohort. Source: Attanasio, Blundell, Conti, and Mason (2019).

Figure 4.1: inequality in socio-emotional skills

◇ What should we do?

  * Try to look at the forgotten aspect in the Grossman health production model: early childhood development

  * From the data, about a half of the difference in labour market outcomes, health behaviour and health at age 30 can be explained by early life factors happening before $\leq$ 10 years

old, and the rest is explained by the causal effect of edu (Conti and Heckman, 2010)

* Traits of health inequality also begin early in life: high value of C-Reactive Protein (a measure of inflammation) was documented at mid-40s for those in low SES. At the same time, a similar gradient occurred in childhood, where children born to lower SES households are also more likely to have low birth weight (Conti and Heckman, 2010)
* SES-health gradient can occur before birth! The fact that the mother resides in deprived neighbourhoods had a significant negative effect on the size of fetus along different dimensions (Conti et al., 2018)

(a) By Deprivation

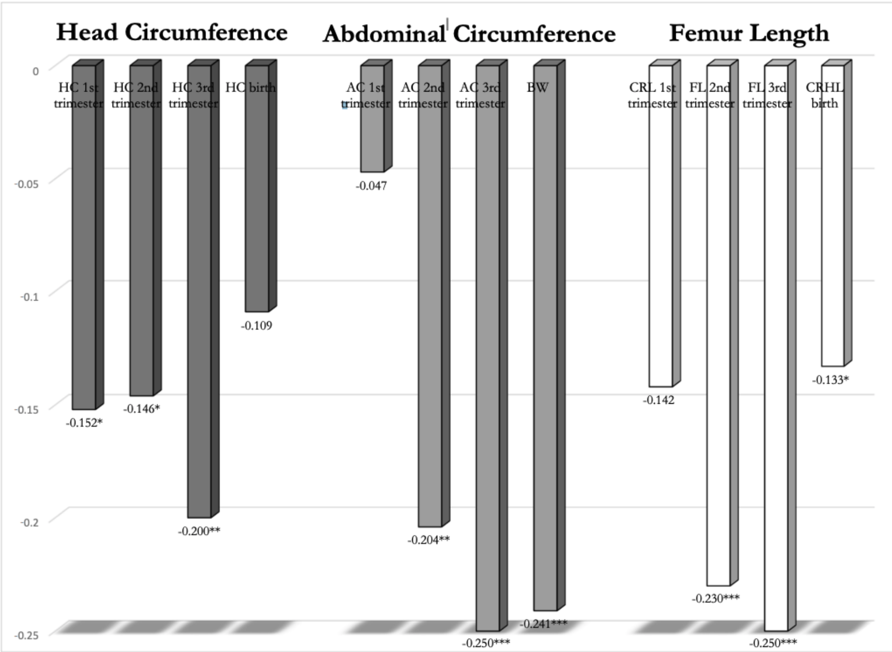


Figure 4.2: SES-health gradient before birth

* This means that the fetuses of poorer moms are smaller, so you will have smaller babies as gestation continues.
* The key takeaway from all these stories is that when the kid is conceived, there are already inequalities that will be carried over throught the lifecycle

OK maybe it makes sense to start early!

4.2 Understanding the causal effect of early life conditions

- The way to look for causal effects involves techniques in labor (quasi-natural experiments from exposure to war, pandemics, policies) and twin-based designs (twin FE, how does the outcome of the high-birth-weight twin look like?)
- Here, we look at effect of early childhood conditions on later life outcomes by exploiting policy shocks and RCTs
- A slight detour before studying effect of various internventions: some major findings in the literature on this topic:
  * The effect of early shocks are heterogeneous, and the heterogeneity is endogenous. Effect of shocks depends on child endowments, resource constraints and production technologies
  * Limited understanding on effect of multiple shocks, and the relative importance of biology and behavioural mechanisms

- Nevertheless, we should intervene becuase there are frictions preventing parental investment in human capital in early years, such as...
  - Information and liquidity constraint
  - Behavioral factors: Time inconsistency, altruism, cognitive constraints, aspirations
- Interventions may help: fetus of smoking moms' are much smaller. A slight caveat here is that higher obesity in the population can lead to too big fetuses (bigger $\neq$ better).

4.3 We should intervene. But intervene how?

4.3.1 Pre-natal Interventions: Home visiting (HV)

Nurse Family Partnerships (NFP)

- A program very popular in the USA
- Nurses going to houses of first-time low-SES mothers at early stage of pregnancy until the child is 2 years old
- Max. 64 visits that provide information and support
- Super high return ($\approx$ 500%)
- Treated kids (those with parents being allocated to NFP group) turned out better in terms of better cognition, higher likelihood to graduate high school with honours, less disabilities, and the mothers spent less money in welfare compared to the control group (those with parents allocated to not-that-good usual care system)
- Details on cost-saving (Olds et al., 2019):

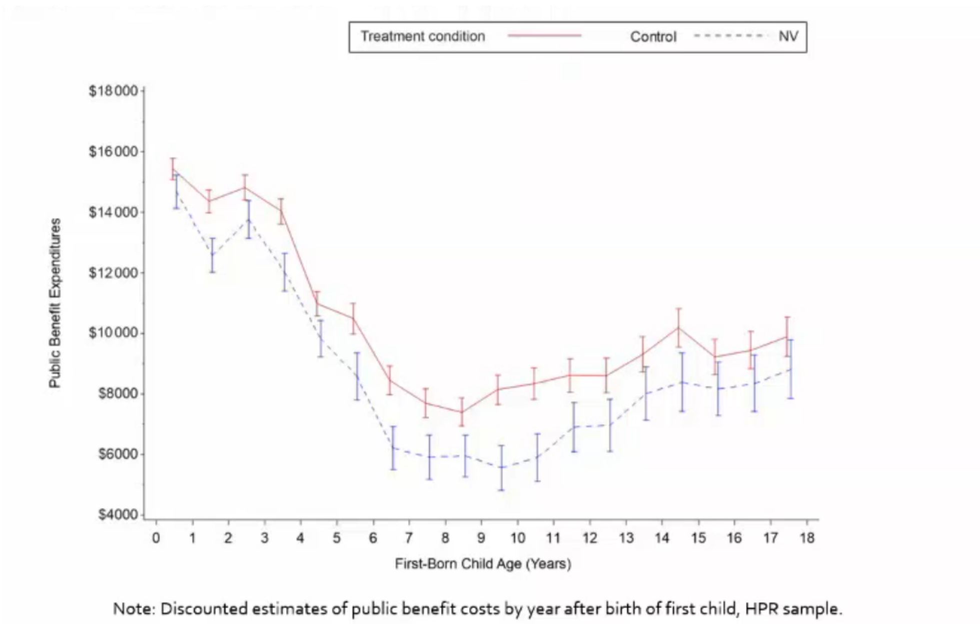


Figure 4.3: Cost saving in NFP (NV) group

So NFP effective and saves money for the government? We cannot say that until we consider what will happen if we scale this up!

Scaling up

- This is a relatively new area of research, and we have relatively good evidence on the effect of HV at scale from Scandinavian countries due to their "welfare state" setup
- Hjort et al. (2017) studied a HV program that was implemented at scale in Denmark from 1937-49

- They found that kids in treated areas saw huge improvements in SES and health compared to the control group (note that this effect is for every child in Denmark!)
- Some examples of improvement:
  - Children in treated areas more likely to survive after 64 years old
  - Less likely to be diagnosed with diseases
  - Fewer nights spent in hospitals

That's Great! But the effect of HV in the UK is unknown due to data problems...

We also have similar programs in the UK called the Healthy Child Programme (HCP) that ...

- provided advice & support (e.g.immunization, examination of the child)
- Identified risks (e.g. abuse)
- Referral to social services if needed

Is it effective? We don't know because there is no data!

However, this program is now facing its own problems

- Massive defunding after Oct 2015 & local authorities taking over this responsibility led to caseload of >500 in some regions
- Health visitors were moved to the frontline during
- The effect of this is by no means trivial. A survey of visitors by Conti and Dow (2020) showed that they are concerned about domestic abuse & violence, parents' mental health, child growth/ development, ...

Does the success of the NFP generalise? Investigation of the Family Nurse Partnership (FNP) programme in the UK

- Note the difference between the UK FNP and the US NFP :)
- Eligibility criteria of FNP: teenage first-time mom (regardless of SES)
- This program was scaled up after being implemented in small scale (10 sites) and received positive feedback
- Result of an RCT that came out in 2014 showed that this program did not impact the primary outcomes. However, it improved other outcomes such as child cognitive & language development, but they were considered secondary. Primary outcomes are health-based, see figure 4.4 for details on the outcomes of interest.

Outcome domain	Four Primary Outcomes	Secondary
Pregnancy & birth	•Changes in prenatal tobacco use •Birth weight	•Intention to breastfeed •Prenatal attachment
Child health & development	•Emergency attendances / admissions within two years of birth	•Injuries & ingestions •Breast feeding (initiation & duration) •Language development
Maternal life course and economic self-sufficiency	•Proportion of women with a second pregnancy within two years of first birth	•Education&Employment •Health status •Social support • Paternal involvement

Figure 4.4: Primary and Secondary outcomes in the UK FNP

- Follow-up studies showed that the effect on cognitive development (reading and writing) persisted, but there was no impacts on child maltreatment

◇ **Overall, the program only improved certain aspects of child development, not others.** "Adding FNP to the usually provided health and social care provided no additional short-term benefits to our primary outcomes"

What could explain the poor generalizability of the evidence?

◇ Difference in culture leading to different behaviour

◇ Difference in trust in public institutions

◇ Difference in counterfactual. The control group gets much worse treatment in the US compared to the UK (remember that control group in the UK gets HV!).

- ∗ In FNP, households get about 39 visits. In control (HV), households get about 16 visits from NHS health visitors
- ∗ FNP and HV can both tell mothers about how to make the kid healthy (breastfeeding, not smoking, ...). However, the unique thing about the FNP was that they told moms how to approach child cognitive development (e.g. play with the child)

Note that the secondary outcomes are still important, and the FNP does have a positive significant impact on child development at 5 & 7 years old.

Also note that recent studies (McConnell et al. 2022) evaluating the effect of the US NFP found that it no longer brings significant improvements on " maternal and newborn health primary or secondary outcomes". The effectiveness of programs can also change over time even for the same country due to different counterfactuals

4.3.2 Post-natal Interventions

The Carolina Abecedarian Project

◇ This is a small-scale randomised experiment in the US in the 1970s

◇ Send poor African-American kids to intensitve daycare from 0-5 years old

◇ What do they have at the daycare centres?

- ∗ Cognitive and behavioural stimulations
- ∗ Healthcare
- ∗ Nutrition (food provision)

◇ It is also very effective!

- ∗ Benefit-to-Cost Ratio = 7.3
- ∗ Group exposed to this treatment had better health in adulthood compared to the control

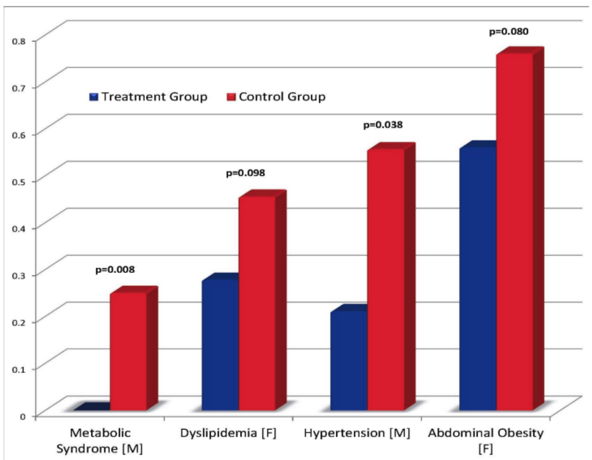


Figure 4.5: Improved health in adulthood (35 yr old)

Effect of implementing this at a larger scale: an examination of the effects of Head Start and Sure Start

Head Start (US)

- A compensatory preschool program in the US providing head start centres (centres providing a variety of services as in the CAP) to low income children
- Carneiro and Ginja (2014) found that "participation in the program reduces the incidence of behavioural problems, health problems, and obesity of male children at ages 12 and 13; lowers depression and obesity among adolescents, and it reduces engagement in criminal activities and idleness for young adults."
- They employed a fuzzy RDD design using discontinuities in the probability of participation induced by program eligibility rules (note that a jump was only observed in males)
- Results in more detail:

Table 6—Reduced-Form Estimates: Ages 12–13 (*Sample: Males*)

	Control mean (1)	Observations (2)	ITT (3)	Marginal effect (4)	RW pv < 0.1 (5)
Behaviors					
Drug use	0.156	1,268	0.118 [0.189]	0.027	No
Overweight	0.198	1,242	−0.379** [0.165]	−0.095	Yes
Grade repetition	0.286	1,285	−0.243 [0.161]	−0.079	No
Alcohol use	0.467	1,289	−0.249 [0.160]	−0.085	No
School damage	0.143	1,209	0.0334 [0.161]	0.008	No
Ever smoke	0.359	1,281	−0.146 [0.160]	−0.048	No
Special education	0.239	1,254	−0.250 [0.169]	−0.075	No
BPI	0.654	1,211	−0.274** [0.125]		Yes
Health					
Health requires use of special equipment	0.0938	1,111	−0.777*** [0.287]	−0.101	Yes
Health requires frequent visits to doctor	0.190	1,273	−0.323* [0.184]	−0.083	No
Health requires use of medicines	0.207	1,251	−0.214 [0.179]	−0.056	No
Health limitations	0.0683	1,115	−0.168 [0.216]	−0.028	No
Cognitive					
PIAT-M	0.047	1,197	0.027 [0.100]		No
PIAT-RR	0.156	1,196	−0.238* [0.133]		No
PIAT-RC	−0.161	1,181	−0.144 [0.113]		No

Notes: Probit (ordinary least squares (OLS) for BPI, PIAT) estimates. Controls excluded from table: cubic in log family income and family size at age four, an interaction between these two variables, cubic in log of average family income and family size for ages zero to two, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age four, race and age dummies, and year and state of residence at age-four effects. *Control mean* is the mean outcome among observations just above the cutoff (at most 25 percent above the cutoff). Marginal effects for discrete outcomes in column 4. Robust standard errors are reported in brackets clustered at the level of each state-year cell, measured at age four.

Figure 4.6: Head Start reduces overweight, behavioural (BPI) and other problems for adolescents

Sure Start (UK)

- This is a program launched in 1999 in England that was initially targeting diadvantaged communities but then scaled up
- The program intends to provide all the help possible to parents in one go
  - Provision of health information

  - * Referral to health services
  - * Stay-and-play, parenting programmes (e.g. FNP)
  - * Parental support, job search assistance

- ◇ The impact of this program is studied under the following framework:

$$D^{ya}_{sql(d)} = \delta^{ya} SS_{dq} + \beta^{ya} X_s + \alpha^{ya} Pop_{al} + \gamma^{ya}_q + \pi^{ya}_{l(d)} + \epsilon^{ya}_{sql(d)}$$

Where

  - * $D^{ya}_{sql(d)}$: an indicator for whether there is any hospitalization of type y at age α for children of sex s born in quarter q and residing in neighborhood l (of LA d)
  - * SS_{dq}: the average number of centers per thousand children aged 0-4 that were open between the child's birth and the age at which the outcome is measured
  - * Gender, size of population in the LAs are controlled for
  - * Also included chort of birth & neighbourhood FEs

- ◇ Results
  - * Sure Start reduces overall hospitalisations. Note that there was an increase in hospitalisation at the age of 1, and this is probably due to hospitalisation for preventative purposes

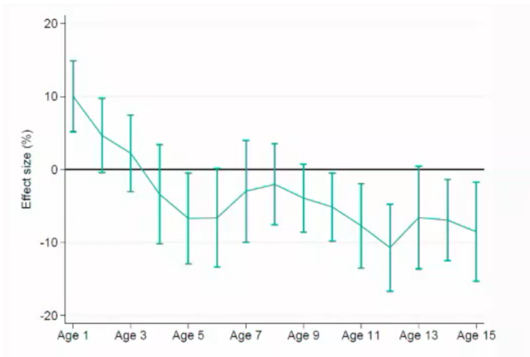


Figure 4.7: Sure Start reducing overall hospitalisation

- ◇ If we decompose by causes of hospitalisation and the extent to which treated areas are deprived

Figure 4.8: Decomposition of the effect of Sure Start

- Greater access to Sure Start substantially increases hospitalisations for infectious illnesses in infancy, which can be because of referral effects or the spreading of the infectious disease in the childcare centres
- However, there are significant and substantial falls in hospitalisations because they build up a stronger immune response. This effect fade out in the longer term as the other kids' immune systems catch up after school
- Less hospitalisation due to external causes at early ages. Kids are able to fight with each other as they grow older
- Big reduction in hospitalisation for mental health at age of 12-13
- The program is more effective in poorer areas (poor = green; rich = light blue; middle = pink)

4.4 Why do these programs work?

This will help us understand the generalisability of programs implemented in one country at one time

How early interventions affect us later?

- A stylised way
  - Health is a component of human capital, and good health will yield multiple returns later in life (e.g. education, earnings,...)
  - Health can be developed throughout the lifecycle by human capital development (think about the factors involved in the production of health) and **investments** by parents and governments. Investments in human capital can be carried out at different stages in life.
  - Investment level can also react to level of human capital
- The shortfall with this very general approach is that there are so many ways to put money (can invest in pre-natal/ post-natal/ interventions at late childhood etc.). How to best allocate our money?
- Attanasio (2015) has developed a model that allows us to understand the impact of early childhood investments
  - This is a model of investment in children's human capital
  - Parents are altruistic (care about child's human capital/ development), and they maximise utility function subject to constraints
  - Human capital change over time because of (a) Investments, which can react to level of HK (b) Environmental variables whose evolution is independent of human capital (c) Past level of HK. Such process is governed by a production function
  - This HK production function is complicated and non-linear
  - The choice variable here is the level of parental investment to children, and parental behaviour plays an important role here - there can be compensation/reinforcement effects.
  - Let's write this down formally. Here, I consider a simple framework where parents fully understand the child's human capital production function of:

$$HK_{i,t+1} = g_t(HK_{i,t}, I_{i,t}, Z_{i,t}, e_{i,t}^{HK})$$

  Where

  - $HK_{i,t/t+1}$: Human capital
  - $I_{i,t}$: Parental Investment
  - $Z_{i,t}$: Other background variables like parents' education
  - $e_{i,t}^{HK}$ Random Shocks, or inputs in the production function that are not directly observed or considered by the researcher

- Notice that the production function varies with time, and there is no saving

We then have the model in full glory

$$\begin{aligned} \underset{C_{i,t},\, I_{i,t}}{\text{maximize}} \quad & U(C_{i,t}, HK_{i,t+1}) \\ \text{subject to} \quad & C_{i,t} + P_t^I I_{i,t} = Y_{i,t}, && \text{(Resource constraint)} \\ & HK_{i,t+1} = g_t(HK_{i,t}, I_{i,t}, Z_{i,t}, e_{i,t}^{HK}) && \text{(Technological constraint)} \end{aligned}$$

- Here, P_t^I is a vector of investment prices and the objective function is for parents
- Do not distinguish between material/time investment here
- Solving this model gives demand functions for investment and consumption
- Investment function is given by:

$$I_{i,t} = f_t(HK_{i,t}; P_t^I; Z_{i,t}; e_{i,t}^{HK}; Y_{i,t}; \pi)$$

  Where π is the vector of parameters characterising the utility and production function
- This is very important for empirical work, as this allows us to estimate the effect of early childhood investment on human capital (i.e. estimate the production function)
- More specifically, P_t^I and $Y_{i,t}$ does not enter the production function, meaning that we can exploit variations in these variables to estimate the production function (Recall the way we estimate SEMs in ECON0019)

4.5 Understanding the mechanisms

We now have a good grasp of the reduced-form relationship between early childhood interventions and later life outcomes. But where does the better health of treated kids come from?

- For example, in the Carolina Abecedarian Project, we see improved health when the treated kids turn 35.
- Is it because of improvements in child development, or is it because of better education/ employment/ income of the treated kids later in life?
- Conti et al. (2016) showed that half of the impacts of the intervention are explained by early development (reduction in the child's BMI and increase in task orientation at ages 1-2)

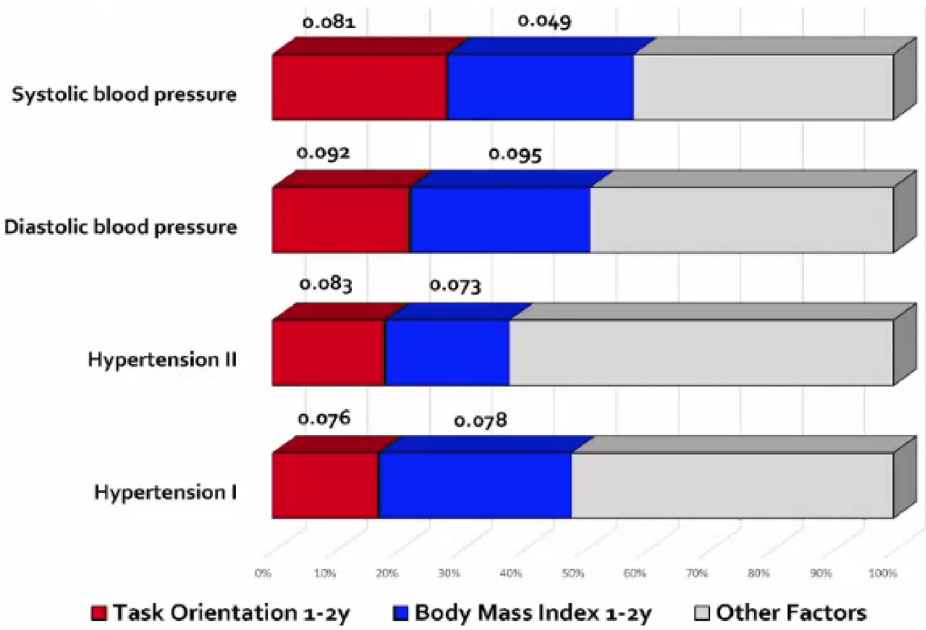


Figure 4.9: Effectiveness of CAP explained by improvements in outcomes in early childhood

Chapter 5

Informality and Development

5.1 Basics and Facts

5.1.1 Definitions

Legalistic Definition

Informal firms and workers are those that operate at the margin of the relevant laws and regulations. This excludes obviously illegal activities, such as drug smuggling, but includes unregistered/tax-evading firms.

We do NOT define informality by sizes.

Two Margins of Informality for Firms

1. Extensive Margin: whether to register the business
2. Intensive Margin: whether registered firms hire informal workers (or not comply with taxes)

For employees, there's only one kind of informality: whether employed informally or not.

5.1.2 Statistical Facts

Informal firms occupy a share around 70% in Brazil (extensive margin); informal workers have a share of 30-80% of the labour force in Latin American countries (either employed by informal firms (extensive margin) or employed informally by formal firms (intensive margin)). Specifically, the intensive margin accounts for 40-44% of informal employment in LAC.

- Informality is pervasive

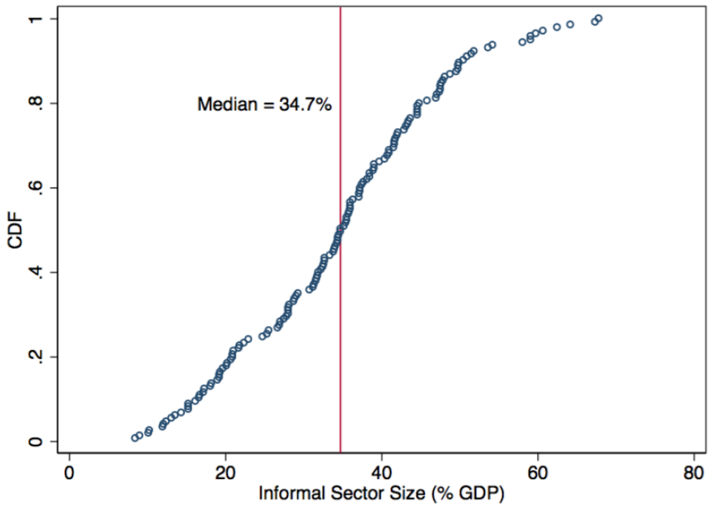


Figure 5.1: Distribution of Informal Sector Size as % of GDP

◇ Informality is negatively correlated with GDP per capita

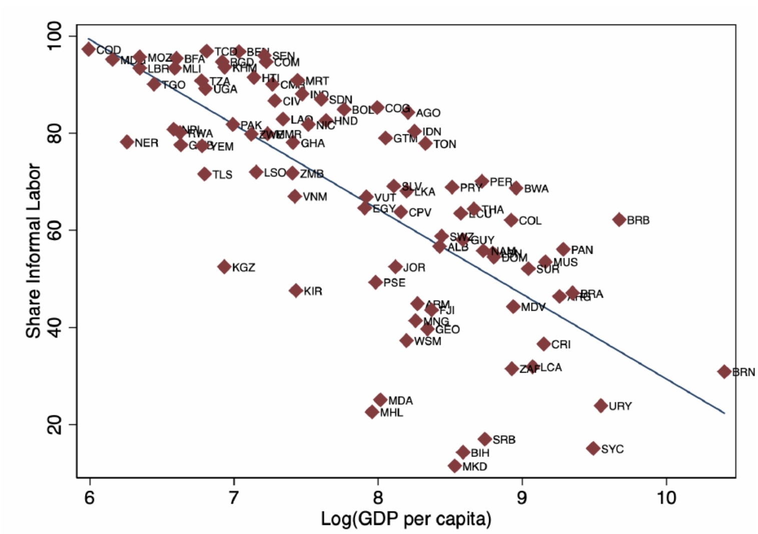


Figure 5.2: Share of Informal Labour against Log GDP per Capita

◇ There are huge variations of informality even within income groups

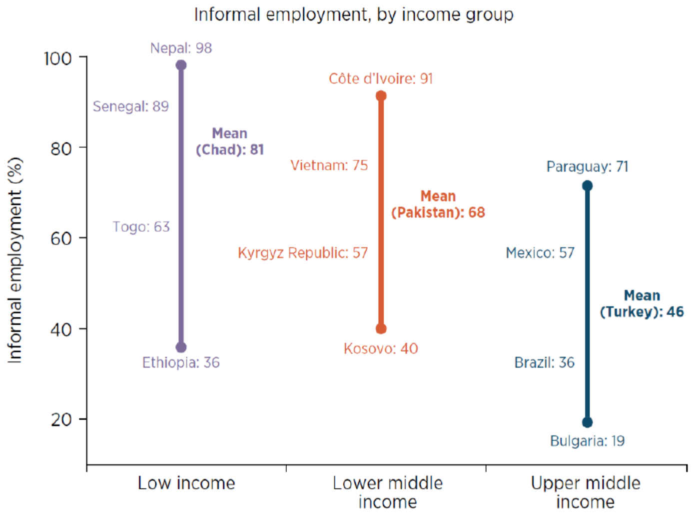


Figure 5.3: Variations within Income Groups (Penny Goldberg, 2022)

◇ Despite the negative correlation found in fig 5.2, economies cannot simply grow out of informality. In the graph below, economies coloured light grey have shrinking informal sectors as they grow. However, on the other hand, economies coloured black have stable/expanding informal sectors as they grow. A representative case is India (coloured red): despite fast growth and development, its informal sector has hardly diminished. (Note that informality reduction itself should not be considered as an economic goal because its welfare implication is ambiguous.)

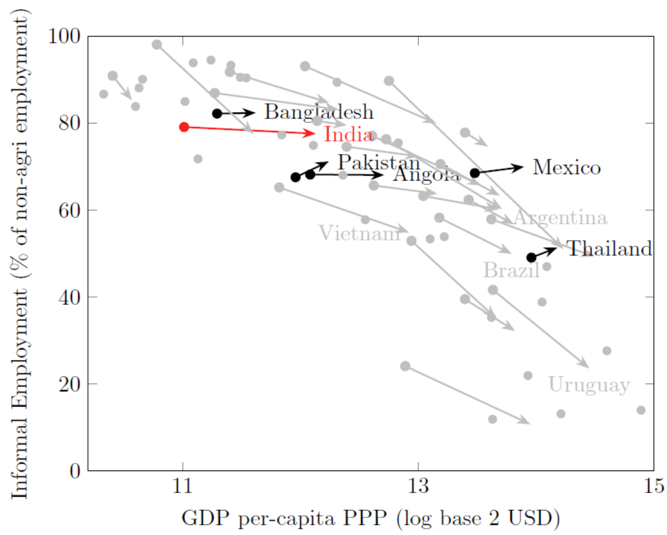


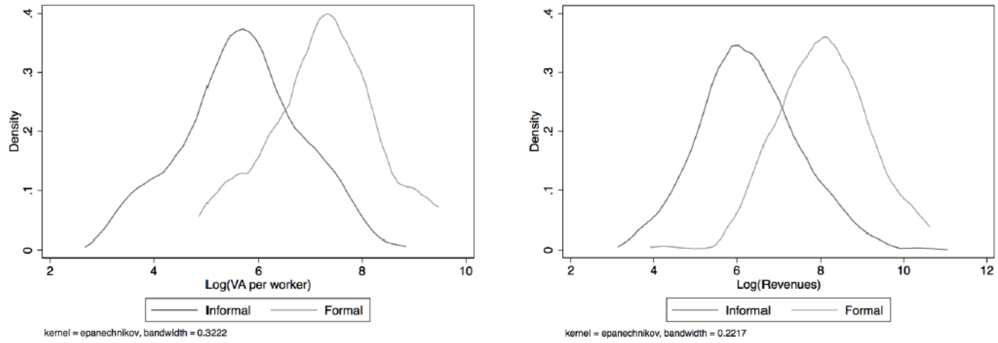
Figure 5.4: Trajectory of Informal Employment (Belavadi, 2021 (PhD Thesis, Penn State))

5.2 Formal vs. Informal Firms: No Duality and Dynamics

Informal firms are on average smaller, pay lower wages, have less educated owners, hire less educated employees, and earn lower profits.

5.2.1 Lack of Duality

Nevertheless, differences above do not necessarily indicate a dualistic view. Formal firms and informal firms coexist within narrowly defined industries, and there is a substantial overlap in the productivity distribution of formal and informal firms, even within industries:



(a) Productivity: Log(VA/Worker)

(b) Size: Log(Revenues)

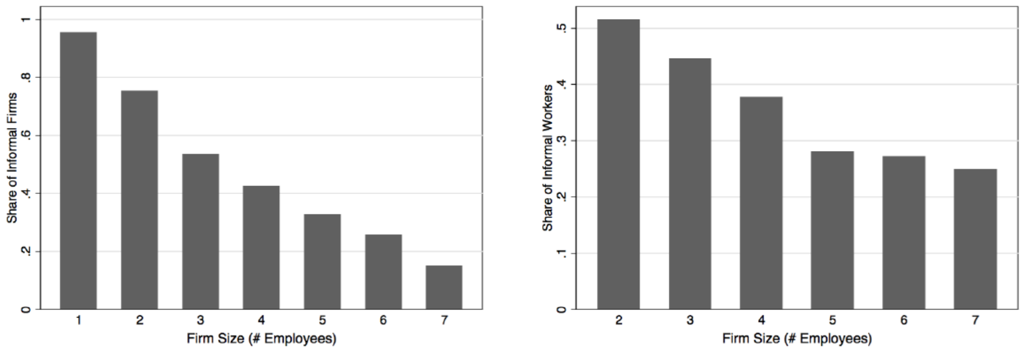
Source: Ulyssea (2018)

Figure 5.5: Large Overlap in Sizes/Productivity Distributions (Ulyssea, 2018)

5.2.2 Missing Middle?

Missing middle refers to situations where there are a large number of firms with either a few employees or many employees, with a thin distribution in between. Missing middle is an indication for economic solidification, which is harmful for growth and development.

The distribution of firms does not shown a "missing middle." Both extensive and intensive margins of informality are negatively correlated with firm sizes:



(a) Panel A: Extensive margin

(b) Panel B: Intensive margin

Source: Ulyssea (2018)

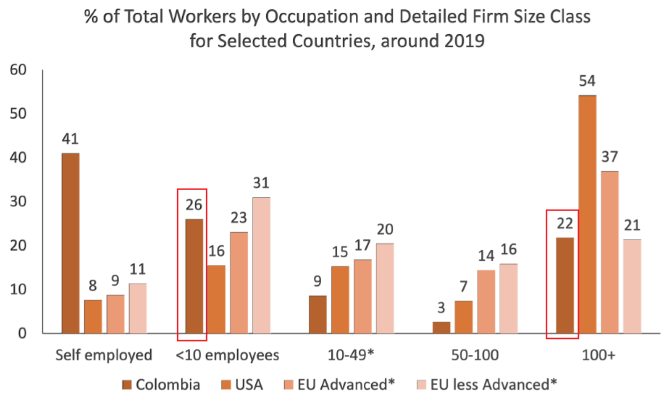
Figure 5.6: Margins of Informality and Firm Sizes - Brazil (Ulyssea, 2018)

However, there is a missing middle in employment distributions:

	Firm size by number of workers	Number of firms	Share	Number of workers	Share
Ecuador	2–10	2,987	99.6	5,764	79.3
	11–100	11	0.03	727	10.0
	101 +	0.1	0.01	773	10.0
	Total	**2,998**	**100.0**	**7,264**	**100.0**
Mexico	2–10	4,535	95.0	9,596	$\boxed{43.5}$
	11–100	211	4.40	4,383	19.9
	101 +	27	0.60	8,044	$\boxed{36.5}$
	Total	**4,773**	**100.0**	**22,024**	**100.0**
Peru	2–10	3,163	98.9	8,586	73.0
	11–100	29	0.08	1,563	13.3
	101 +	4.1	0.02	1,610	13.7
	Total	**3,196**	**100.0**	**11,759**	**100.0**
United States	2–9	4,726	60.1	12,503	9.7
	10–99	1,405	17.9	29,851	23.2
	100 +	1,729	21.9	86,147	67.0
	Total	**7,861**	**100.0**	**128,592**	**100.0**

Source: Cruces and Levy (2019)

Figure 5.7: Missing Middle in Employment - Mexico

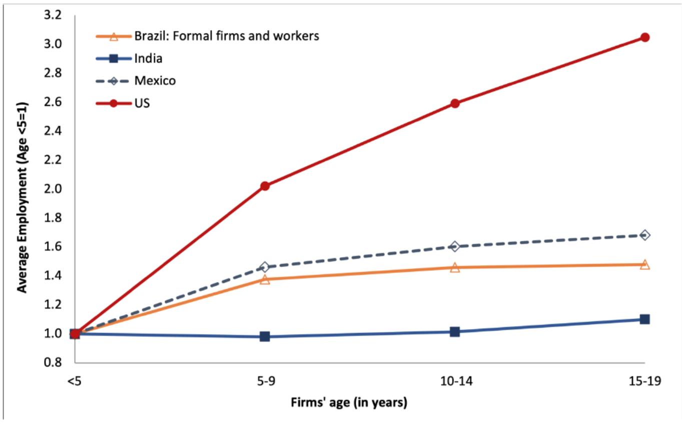


Source: Eslava, (2022)

Figure 5.8: Missing Middle in Employment - Colombia

5.2.3 Informality and Firm Expansions

Overall, evidences show us that firms in developing countries grow less and stagnant firms survive longer:

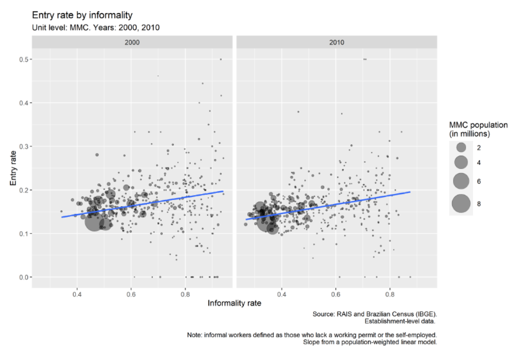


Source: Hsieh & Klenow (2014) + Sedlacek & Ulyssea (2022).

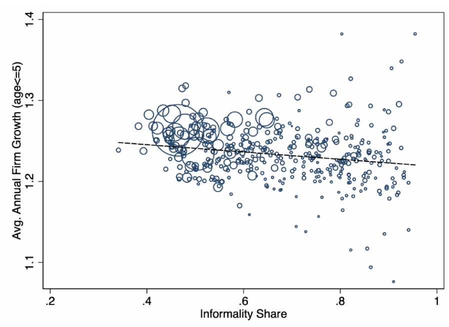
Figure 5.9

- ◇ The red line represents U.S.: its high gradient shows that surviving firms in U.S. expand very fast, and some small firms quickly die out
- ◇ On the other hand, the dash line represents India: expansion of firms is stagnant (maybe due to the family/gender business traditions)
- ◇ The orange line represents the formal sector in Brazil: if the informal sector is included, it will be approximately the same as India

One explanation for such stagnation could be related to informality: as shown in the following figures, there is a positive correlation between entry rate and informality: plenty of firms enter the market as informal ones; on the other hand, those new entrants tend to expand very slow as firms' growth rate is negatively correlated with informality.



Source: Sedlacek & Ulyssea (2022).



Source: Sedlacek & Ulyssea (2022).

Figure 5.10: Left: Entry Rate and Informality; Right: Firm Growth and Informality

In addition, we also have evidences that:

- ◇ Firms' average sizes increase with their ages, in both formal and informal sectors
- ◇ Extensive and intensive margins of informality decrease with firms' age
- ◇ A substantial fractions of informal firms eventually formalise

5.2.4 Employees' Perspective: Wage Gaps & Income

Considering all workers as a whole, there is no significant overall effect of informality on wages after controlling firm fixed effects. (Red boxes)

For unskilled workers, being employed in the formal sector significantly increase their wages. However, for skilled workers, no such effect is observed. The reason behind this difference could be that the minimum wage becomes binding for the unskilled workers as they become formally employed. (Orange boxes)

Dep. Var.: $\log(wage)$

	All Workers (1)	All Workers (2)	Unskilled (3)	Unskilled (4)	Skilled (5)	Skilled (6)
Formal	0.3511***	0.0499	0.4127***	0.2595***	0.2672***	-0.1040
	(0.030)	(0.086)	(0.039)	(0.093)	(0.046)	(0.157)
Skilled	0.2190***	0.0926**				
	(0.035)	(0.046)				
Male	0.1578***	0.1125*	0.1179**	0.0536	0.1628***	0.1445**
	(0.032)	(0.061)	(0.046)	(0.137)	(0.046)	(0.071)
Age	0.0646***	0.0425***	0.0590***	0.0421***	0.0879***	0.0840***
	(0.007)	(0.011)	(0.009)	(0.016)	(0.013)	(0.026)
Age Sqr.	-0.0007***	-0.0005***	-0.0006***	-0.0004**	-0.0010***	-0.0012***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	4,502	4,502	2,596	2,596	1,906	1,906
R-squared	0.200	0.872	0.188	0.902	0.149	0.866
Firm FE	No	Yes	No	Yes	No	Yes

Notes: Regressions use matched employer-employee data for formal and informal firms and their employees from ECINF. Variable *Formal* is a dummy for formal employee; *Skilled* is a dummy for workers with at least high school degree. Robust standard errors in parentheses. Significant at the *** 1 percent, ** 5 percent, and * 10 percent level.

Figure 5.11: Log Wage Regression on Formal Dummies and Controls

Hardly surprisingly, we also have evidence that labour informality is negatively correlated with household income:

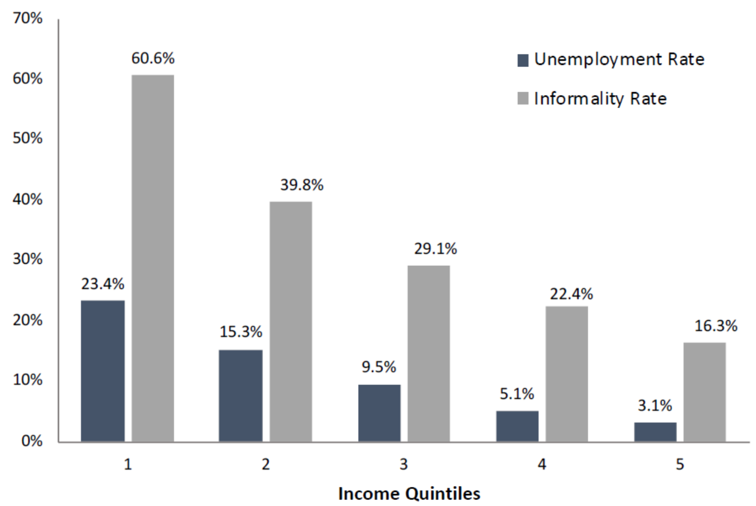
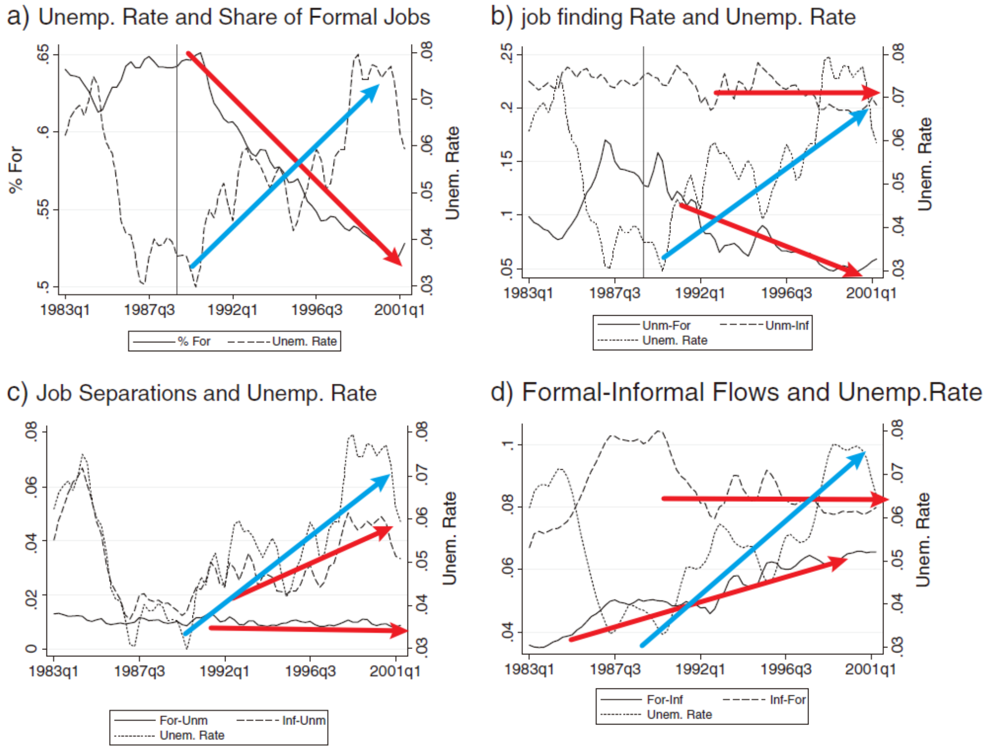


Figure 5.12: Labour Informality and Household Income

5.2.5 Transfer Into and Out of Informality

Hiking unemployment in the 90s sheds light on the patterns of employment in both formal and informal sector:



Source: Bosch and Esteban-Pretel (2012)

Figure 5.13: Transfer Into and Out of Informality (Bosch and Esteban-Pretel, 2012)

- Graph a: informality is counter-cyclical: when there is a recession, the informal sector occupies a higher share in the economy, providing a potential buffer
- Graph b: this shows the probability with which workers find formal/informal jobs from unemployment, and it corroborates graph a: during the recession, it becomes harder to find a job in the formal sector while the job finding rate in the informal sector remains unchanged.
- Graph c: during the recession, job separation (destruction/layoff) increases clearly, indicating worse job security in the informal sector, since there is no firing cost
- Graph d: direct transitions from informality to formality are procyclical (easy to see from graph) as they correlate negatively with unemployment rate; Direct transitions from formality to informality are also procyclical (less easy to see) but less volatile

Overall, we can conclude that the main reason why informality is counter-cyclical is not because of the flow between formal and informal jobs, but the flow between unemployment and formal/informal jobs.

5.3 ⋆ Modelling Informality

5.3.1 Model Setup

For now, we focus on the extensive margin only. Operating *formally*, firms' profit function can be expressed as:

$$\pi_f(\theta) = (1-\tau_y)\underbrace{\theta F(k,l)}_{\equiv y(\theta)} - (1+\tau_w)w_f l - r_f k - \bar{c}_f$$

Operating *informally*, firms' profit function can be expressed as:

$$\pi_i(\theta) = \Big[1 - p\big(y(\theta)\big)\Big]\theta F(k,l) - w_i l - r_i k$$

where:

- τ_y is the revenue tax rate

- θ represents firm's productivity
- τ_w is the payroll tax rate
- $F(.)$ is the production function, increasing and concave in l and k
- w_f, w_i are formal/informal wages (usually assume $w_f > w_i$)
- l is the amount of labour employed
- r_f, r_i are capital rents (usually assume $r_f < r_i$ because formal firms have access to more loans and are considered less risky)
- k represents capital used
- $p(y(\theta))$ represents the "cost of informality," increasing and convex in firm's output y
- $\bar{c}_f$ is the per-period fixed cost that only formal firms must pay

Note that, in addition to those costs, formalisation of a firm must incur in a fixed registration cost, which varies dramatically across countries.

5.3.2 Policy Implications

Our model shows us that policymakers can incentive firms to formalise by:

- Reducing costs of formality: costs of entering the formal sector (registration) and costs of remaining formal (e.g. taxes)
- Increasing the benefits of formality: e.g. more access to capital (lower r_f)
- Increasing the costs of informality: e.g. higher enforcement of existing laws and regulations

5.4 Empirical Evidence and Interpretation

5.4.1 Evaluation of Treatments to Facilitate Formalisation

Regression setting

$$y_{it} = \alpha + \beta \text{Treatment}_{it} + \gamma X_{it} + \epsilon_{it} \tag{5.1}$$

- Unit of analysis (i): typically firm or entrepreneur
- y_{it}: formality dummy or share of formal firms
- Treatment_{it}: formalisation treatment dummy (=1 if treated)
- X_i is a vector of Controls
- ATE is identified ($\beta_{OLS} = E[Y_{1i} - Y_{0i}]$) under the assumption: $(Y_{1i}, Y_{0i}) \perp D_i | X_i$ because there is no selection on levels/gains conditional on X_i.

Results

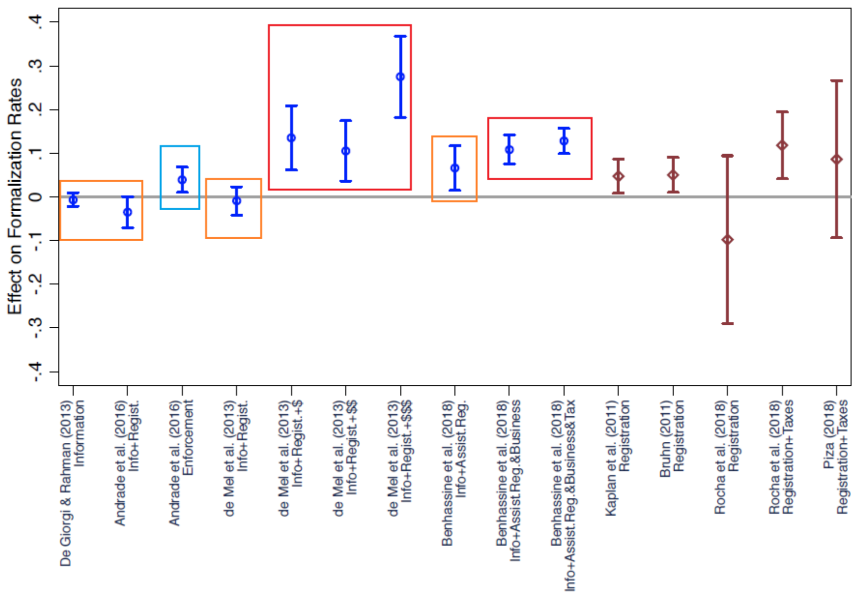


Figure 5.14: Estimated Effects on Formalisation Rates of Different Formalisation Treatments (Ulyssea, 2020)

- Orange squares incorporate treatments that reduce registration costs or provide information, which have no significant effect on formalisation. This reveals that high fixed costs for formalisation may not be the main obstacle.
- Red squares incorporates treatments that reduce ongoing costs, which show significant positive effects on formalisation
- The blue square incorporates the only paper that studied higher enforcement, whose effect is significant. However, in reality, such intervention may have political economy concerns.

5.4.2 Fixed and Ongoing Costs of Formalisation: RCTs in Sri Lanka

This paper written by de Mel et al., 2013 used experiments in Sri Lanka to explore effects of covering registration costs and ongoing costs.

Research Designs and Treatments

- Treatment 1: Provide information + Cover direct monetary costs of firm registrations
- Treatment 2: Treatment 1 + 10,000 rupees (50% of median monthly profits)
- Treatment 3: Treatment 1 + 20,000 rupees
- Treatment 4: Treatment 1 + 40,000 rupees

Treatment 1 only covers the fixed cost of registration the firm, while treatment 2-4 also provide additional money transfers, which can be understood as compensations for discounted future ongoing costs.

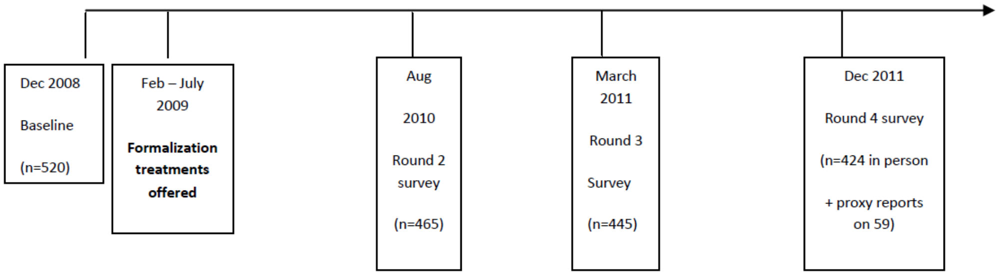


Figure 5.15: Research Timeline

Results: Effects on Formalisation

Dependent Variable: Registered During Intervention Window

	Intention-to-treat		Treatment on the Treated	
	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)
Information and Reimbursement Treatment	-0.00943	-0.0101	-0.0126	-0.0138
	(0.0165)	(0.0254)	(0.0219)	(0.0286)
10,000 Rs Treatment	0.135***	0.134***	0.216***	0.214***
	(0.0380)	(0.0380)	(0.0576)	(0.0515)
20,000 Rs Treatment	0.105***	0.105***	0.167***	0.167***
	(0.0350)	(0.0387)	(0.0534)	(0.0508)
40,000 Rs Treatment	0.275***	0.273***	0.476***	0.471***
	(0.0473)	(0.0453)	(0.0691)	(0.0598)
Strata/Quintuplet dummies	No	Yes	No	Yes
Observations	520	520	520	520
R-squared	0.102	0.284		
P-values for testing:				
10,000 Rs Treatment = 20,000 Treatment	0.5320	0.5264	0.4993	0.4470
10,000 Rs Treatment = 40,000 Treatment	0.0152	0.0086	0.0021	0.0002

Figure 5.16: Effects of Treatments on Formalisation

The figure above shows the Intention-to-Treat (ITT) and Treatment on the Treated (ATT) estimates of corresponding treatment (on the rows). ITT is calculated by directly regressing the formality dummy on the treatment dummy and covariates. ATT is identified as LATE using invitation as an IV for treatment. Put differently, ITT is the treatment effect of assignment while ATT/LATE is the treatment effect on compilers.

Results indicate that only covering the fixed cost of registration is not enough to induce firms to formalise (ITT/ATT of treatment 1 are not significant). Treatment with additional money to cover ongoing costs are more effective, and their effects increase with the amount of payment, as indicated by figures in the red square.

Results: Effects of Formalisation on Firms' Performance

	Monthly profits	Truncated profits (99th)	Truncated profits (95th)	Log profits	Monthly sales	Truncated sales (99th)
Full effect (upper bound)						
Registered with the DS	13,706*	10,834	5,923	0.357*	122,295	99,073
	(8,241)	(6,847)	(4,774)	(0.202)	(85,196)	(72,440)
Effect after controlling for log capital stock (lower bound)						
Registered with the DS	8,996	7,054	2,544	0.168	61,868	44,438
	(8,074)	(6,474)	(4,462)	(0.187)	(85,602)	(71,944)
Observations	2,181	2,181	2,181	2,123	2,139	2,139
Mean for control group in sample	30,537	28,662	25,048	9.87	237,185	211,399

	Log sales	Number of paid workers	Recruited a new worker	Log capital stock
Full effect (upper bound)				
Registered with the DS	0.460	0.525	0.102	0.396
	(0.319)	(0.426)	(0.0902)	(0.258)
Effect after controlling for log capital stock (lower bound)				
Registered with the DS	0.174	0.431	0.106	N/A
	(0.298)	(0.437)	(0.0925)	
Observations	2,088	1,017	1,017	1,009
Mean for control group in sample	11.42	2.35	0.36	12.278

Figure 5.17: Effects of Formalisation on Firms' Performance

Using the treatment as an IV for formalisation, Suresh et al. estimated the effects of formalisation on firms' performance indicators. Overall, there is no significant positive effects. This provides insights that becoming formal may not be beneficial for firms.

	Paid Taxes	Amount of Taxes paid	Formal Accounting	Has Receipt book	Business Bank A/c.	Applied for Business Loan	Applied for Personal Loan
Registered with the D.S.	-0.0643	-8,865	-0.103	$\boxed{0.352^{***}}$	0.0239	-0.00421	-0.0451
	(0.142)	(6,670)	(0.0763)	$\boxed{(0.130)}$	(0.0812)	(0.0770)	(0.0690)
Lag included	No	No	Yes	No	Yes	No	No
Observations	1,036	1,036	1,016	1,049	1,059	1,059	724
Survey rounds question asked	R2, R3, R4	R2, R3, R4	R1-R4	R2, R3, R4	R1-R4	R2, R3, R4	R2, R3
Mean for control group in sample	0.66	6800	0.141	0.31	0.14	0.10	0.056

	Share of Sales made to Govt. (%)	Electric Connection in Bus. Name	Applied for Govt. Contract	Participate in Govt. SME program	Advertised in Last six Months	Business has clear and visible sign	Changed Location
Registered with the D.S.	3.543	-0.152	0.000453	0.0535	$\boxed{0.261^{***}}$	-0.0895	-0.0504
	(2.285)	(0.116)	(0.0540)	(0.0449)	$\boxed{(0.0892)}$	(0.130)	(0.0870)
Lag included	No	No	No	No	Yes	Yes	No
Observations	1,020	724	724	724	1,036	1,030	1,016
Survey rounds question asked	R2, R3, R4	R2, R3, R4	R2, R3	R2, R3	R1-R4	R1-R4	R2, R3, R4
Mean for control group in sample	0.96	0.40	0.022	0.033	0.16	0.56	0.18

Figure 5.18: Effects of Formalisation through Different Channels

Taking a closer look at potential channels through which formalisation could affect firms' performance: we can see that nothing was significantly changed except that now formal firms have receipt books and advertise more.

5.4.3 Other Literatures

Apart from the literature discussed in the lecture, de Andrade et al., 2014 also argues that waiving the lump-sum registration fee is not useful in promoting formalisation. They conducted a RCT in Belo Horizonte, Brazil. In one of the treatments, researchers eliminated all registration costs as well as sanitary tax, municipal inspection fee, and accounting service charge for the first year. Results are summarised in figure 5.19 (column named "Free Cost Difference"): results are at best insignificant, at worst negative.

	Communication vs Control Blocks				Inspector vs Control Blocks			
	Sample Size	Control Mean	Free Cost Difference	Communication Difference	Sample Size	Control Mean	Inspector Assigned Difference	Indirectly Inspected Difference
Administrative data measures of formalizing after interventions began								
Definite match for SIMPLES	1346	0.007	0.00618	−0.00300	5186	0.006	0.00390	−0.00144
			(0.00662)	(0.00446)			(0.00443)	(0.00229)
Definite or probable match for SIMPLES	1346	0.015	0.00177	−0.00706	5186	0.014	0.00421	0.00102
			(0.00773)	(0.00761)			(0.00586)	(0.00384)
Definite match for MEI	1346	0.060	−0.0349***	−0.0155	5186	0.026	0.00313	−0.00557
			(0.0131)	(0.0139)			(0.00826)	(0.00469)
Definite or probable match for MEI	1346	0.067	−0.0370***	−0.0143	5186	0.033	0.0138	−0.00339
			(0.0142)	(0.0161)			(0.0106)	(0.00561)
Definite match for ALF	1346	0.030	0.00311	0.000227	5186	0.032	0.0218**	−0.00540
			(0.0113)	(0.0117)			(0.0110)	(0.00535)
Definite or probable match for ALF	1346	0.041	−0.000335	−0.00597	5186	0.041	0.0327***	0.00206
			(0.0125)	(0.0125)			(0.0124)	(0.00653)
Definitely obtained any type of formal status	1346	0.083	−0.0281*	−0.0120	5186	0.056	0.0245*	−0.0100
			(0.0159)	(0.0174)			(0.130)	(0.0068)
Definitely or most likely obtained any type of formal status	1346	0.104	−0.0350*	−0.0182	5186	0.075	0.0392***	−0.0034
			(0.0183)	(0.0209)			(0.0149)	(0.0083)

Notes: Standard errors in parentheses, clustered at the block level. *, **, and *** indicate significantly different from control mean at the 10, 5, and 1 percent levels, respectively, after controlling for randomization strata. Sampling weights are used for the Inspector vs Control blocks comparisons.
Source: Authors' analysis based on data described in text.

Figure 5.19: Impacts on Formality (de Andrade et al., 2014)

5.4.4 Caveats When Evaluating Those Empirical Studies

RCTs, if correctly carried out, have no concerns about endogeneity by design, but this does not mean that RCT results can be apllied anywhere. Common problems of RCTs are:

- RCTs are costly, so typically researchers can only obtain a *small sample*. This casts doubts on the external validity and general equilibrium effects of RCT results
- There could be *multiple equilibria* in an economy, depending on the level of informality, trust on government, etc.
- A *pre-analysis plan* is ideal to prevent data snooping, but it also confines the potential findings of researches.

Bibliography

Bosch, M., & Esteban-Pretel, J. (2012). Job creation and job destruction in the presence of informal markets. *Journal of Development Economics*, *98*(2), 270–286. https://doi.org/10.1016/j.jdeveco.2011.08.004

de Andrade, G. H., Bruhn, M., & McKenzie, D. (2014). A helping hand or the long arm of the law? experimental evidence on what governments can do to formalize firms. *The World Bank Economic Review*, lhu008. https://doi.org/10.1093/wber/lhu008

de Mel, S., McKenzie, D., & Woodruff, C. (2013). The demand for, and consequences of, formalization among informal firms in sri lanka. *American Economic Journal: Applied Economics*, *5*(2), 122–150. https://doi.org/10.1257/app.5.2.122

Penny Goldberg. (2022). *Penny goldberg's gorman lecture*.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, *108*(8), 2015–2047. https://doi.org/10.1257/aer.20141745

Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics*, *12*(1), 525–546. https://doi.org/10.1146/annurev-economics-082119-121914

Chapter 6

Informality and Labour Markets in Developing Countries

6.1 Basics and Facts: Unemployment, Informality, and Frictions in Labour Markets of Developing Countries

6.1.1 Unemployment and Informality

Unemployment is Not a "Rich Countries' Problem"

Evidences show that unemployment is even a bigger issue for poor countries:

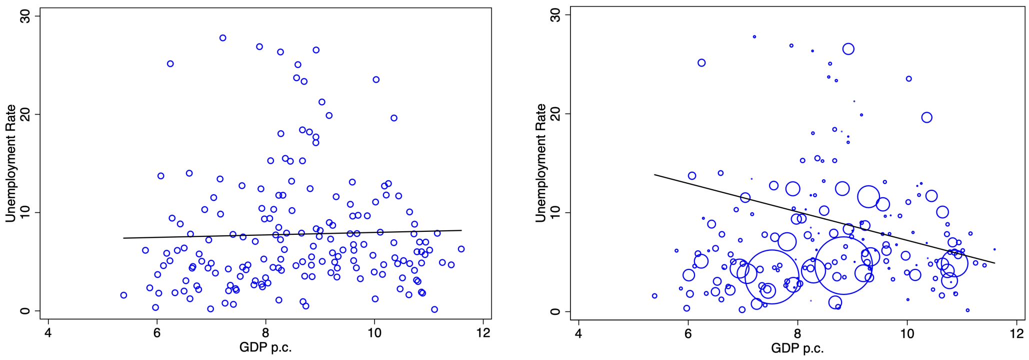


Figure 6.1: Unemployment Rate and GDP per Capita (L: Equal Weight; R: Weighted by Population) (The World Bank, 2022)

Informality and Employment: Issues & Benefits

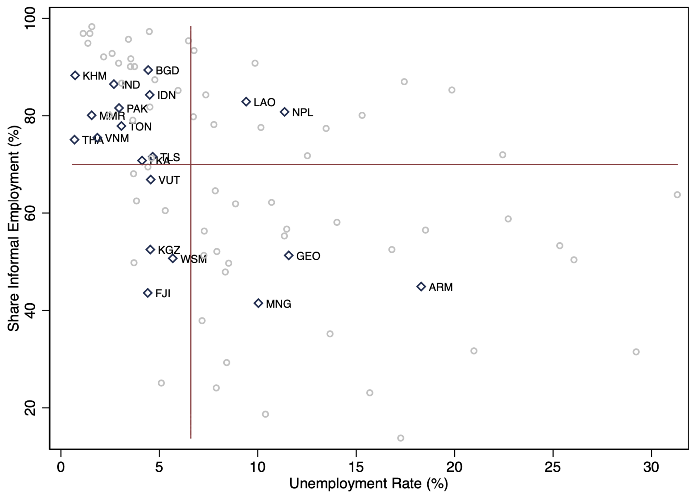


Figure 6.2: Share of Informal Employment and Unemployment Rate

Informality is clearly more common for developing countries, and there is an equilibrium of low unemployment and high informality, indicated by the figure above.

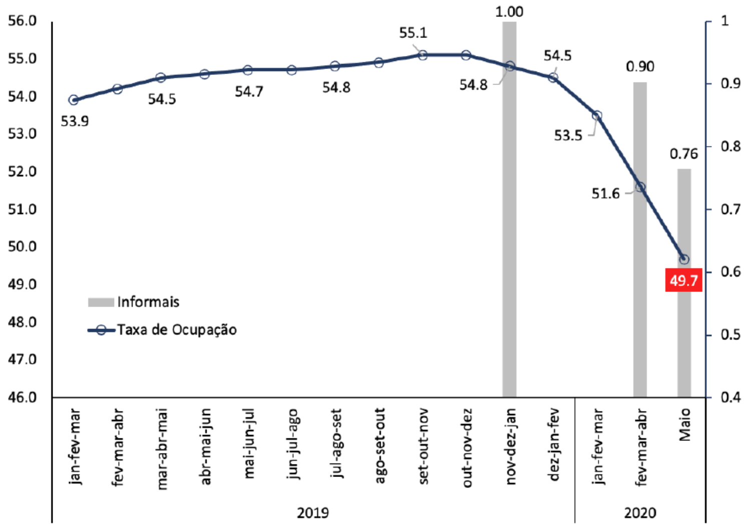


Figure 6.3: Overall Employment and Informal Employment during Crisis

During crisis, high informality could be a problem as there's no firing cost to layoff informal workers. This is observed during Covid in Brazil: while the overall unemployment dropped by 6%, informal employment decreased for around 25%.

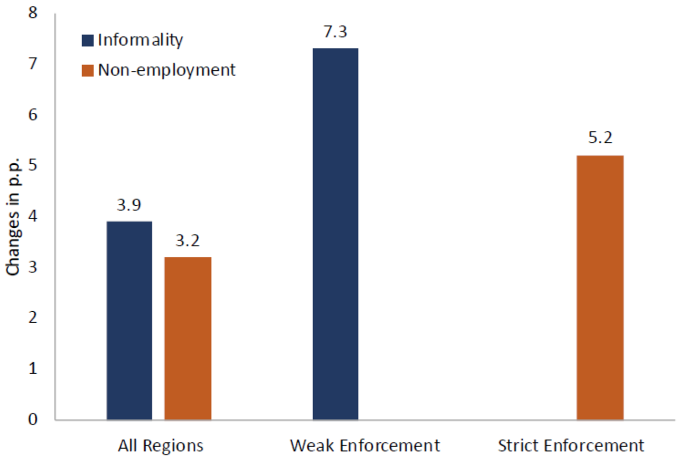


Figure 6.4: Dynamics of Formal/Informal Employment during Brazil's 1990 Trade Liberalisation (Ponczek and Ulyssea, 2022)

However, on the other hand, informal employment could be an employment buffer when the economy is facing some shocks. The figure above shows the dynamics of employment after Brazil's 1990 trade liberalisation, an unilateral trade openness reform which introduced strong competition from imports. In regions with weak enforcement, we witness an upsurge in informal employment while no change in non-employment, indicating a massive transfer from formal to informal sector. In regions with strict enforcement, there is no increase in informality, but tremendous non-employment.

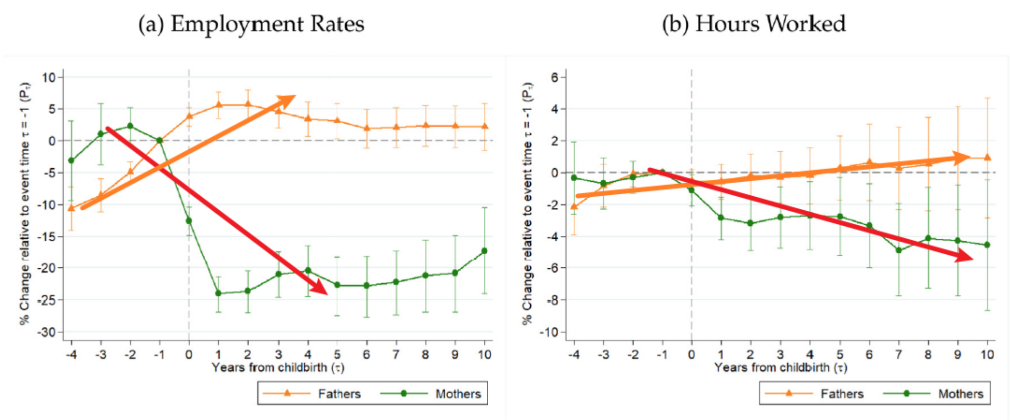


Figure 6.5: Child Cost of Females

Informality may also provides some flexibility to workers. Researchers found significant "child costs" for females: after the birth of the first child, employment indicators of fathers show no sign of deterioration while those of mothers slump. This could be caused by the uneven distribution of caring responsibilities due to social norms. With more time devoted to childcare, it becomes hard for mothers to find formal jobs, and informal employment, with more flexibility, could be an alternative as indicated below:

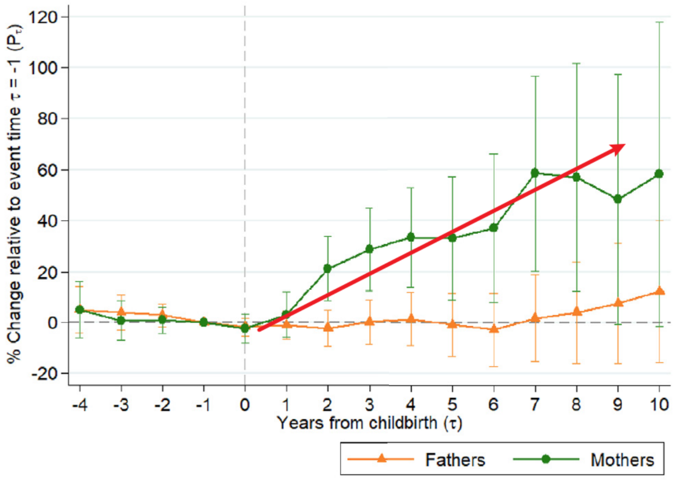


Figure 6.6: Child Birth and Informality

One caveat: the overall welfare effects are ambiguous and always depend on the counterfactual.

6.1.2 Labour Market Frictions

Labour Market Frictions include imperfect information, dispersion of workers and jobs across space (spatial immobility), mobility costs, cultural factors, uncertainty, etc.

Social Network: an Indicator of Information Efficiency

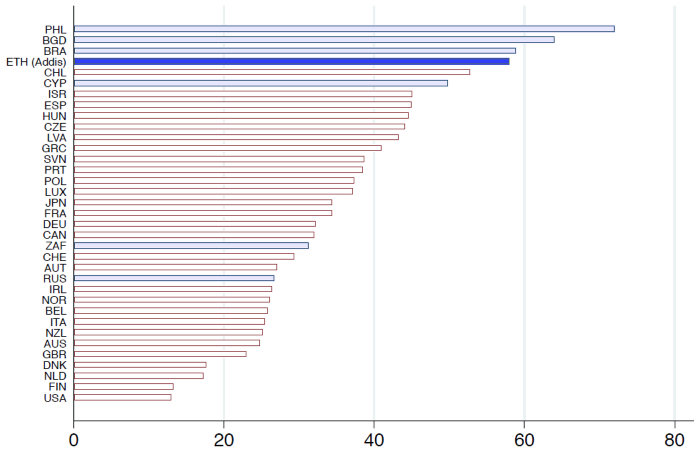
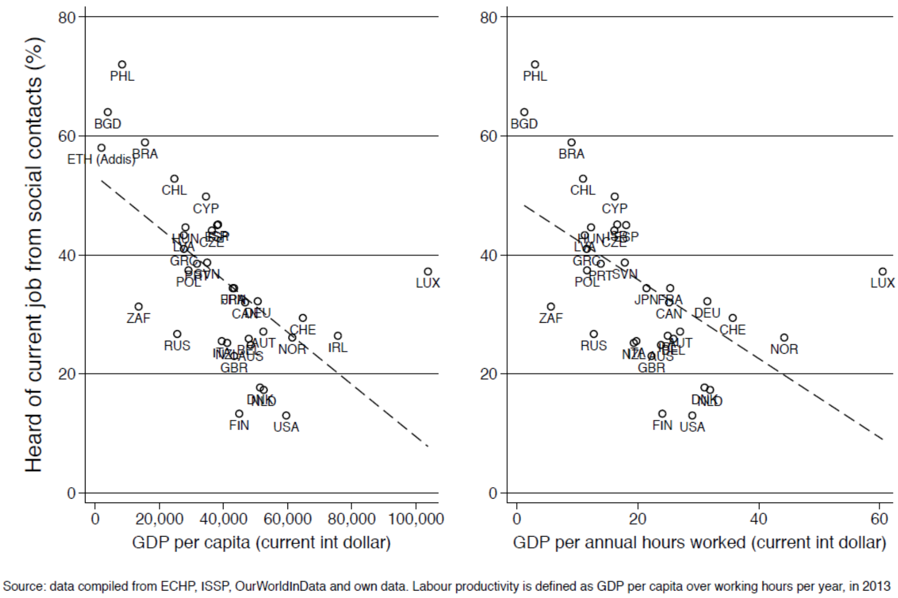


Figure 6.7: Heard of Current Job from Social Networks (%) (Witte, 2018)

The figure above shows the proportion of employed individuals heard of their current jobs from social networks. This measure the importance of social network in an economy's labour market. Typically, higher shares indicate poor alternatives to gather information.

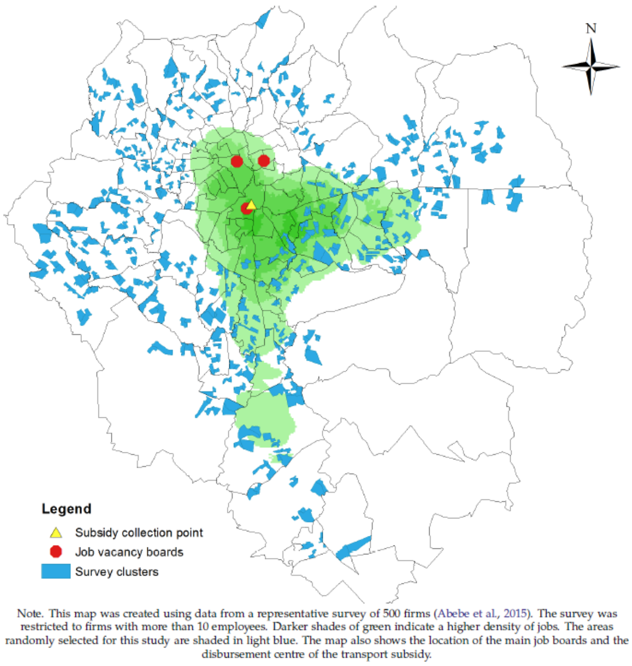


Source: Witte (2018)

Figure 6.8: Heard of Current Job from Social Networks (%) and GDP per Capita (Witte, 2018)

Also, it turns out that there is a negative relationship between shares of employment heard from social networks and GDP per capita. Thus, frictions may be an important problem especially for developing countries.

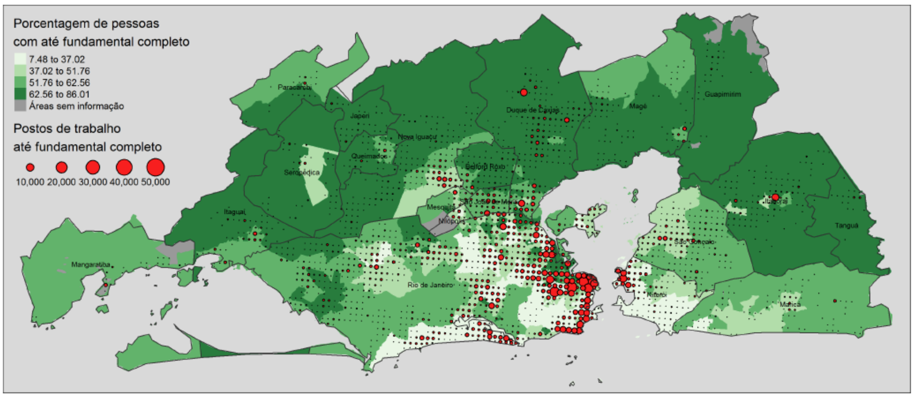
Spatial Frictions



Source: Abebe et al. (2018)

Figure 6.9: Spatial Frictions in Addis Ababa (Abebe et al., 2021)

The blue area represent living places and green area represent job locations. Red dots are job vacancy boards where people find employment information. We can see a clear spatial mismatch between jobs, workers, and information sources.



Source: Campos et al. (2018)

Figure 6.10: Spatial Frictions in Rio de Janeiro

Similar spatial frictions are observed in Rio de Janeiro: areas shaded green are places where low-skilled workers live, and red dots show formal job opportunities for those worker. Red dots are clustered around cities while workers live in rural areas.

6.2 An Empirical Study by Abebe et al., 2021: Transport Subsidy and Job Application Workshop

6.2.1 Motivation and Context

Motivations

The youth typically work less, have lower wages, and face more job insecurities than older workers. This issue is even more severe in developing countries. Besides, lack of access to jobs among young adults is a potentially major source of inefficiency and inequality.

Two main determinants of this lack of jobs:

- High Search Costs due to geographic dispersion, lack of information about vacancies, etc.
- Information Asymmetry between employers and workers

Abebe et al. use two treatments (transport subsidy and job application workshop) to evaluate those two causes and corresponding policies to address them.

Empirical Context: Addis Ababa

- Informal and temporary works are very common
- High turnover: 72% of jobs end within the first 3 months
- Low wage growth even during periods with high economic growth
- The most popular job search method is visiting job vacancy boards
- Visiting vacancy boards is costly due to long travelling distance (as shown by figure 6.9)

6.2.2 Treatments

Treatment 1: Job Application Workshop

Job application workshops are designed to improve job-seekers' ability to present their skills accurately to potential employers (i.e. overcome anonymity). This aims to *tackle with the information asymmetry* between job-seekers and employers.

A workshop incorporates an orientation session and a evaluation session with certifications explaining the nature of the tests and reporting relative grades.

The cost per person is 35 USD including fixed costs related to the development of tests. Excluding fixed costs, it costs only 18.2 USD per person.

Treatment 2: Transport Subsidy

Transport Subsidies are given to individuals with an amount that can cover the cost of a return bus fare from participants' area of residence to the disbursement point (where there are job boards). This aims to reduce the cost of travelling to the city centre to gather job vacancy information and visit firms, hence *reduces search costs*.

The cost per person is 19.8 USD incuding all relevant costs.

6.2.3 Sample and Data Collection

Sampling Strategy

Sampled areas are randomly selected geographic clusters, excluding clusters that are within 2.5km from the city centre. Then researchers use door-to-door sampling to construct a list of all individuals who:

- are between 18-29 years old
- completed high school
- are available to start working in the next 3 months
- don't have a permanent job or enrolled in full-time education

Then individuals are randomly sampled from this list with a higher frequency of from groups with higher educations.

Sample Balancing

Samples are well-balanced: there are no significant differences between the treatment and control groups ex ante.:

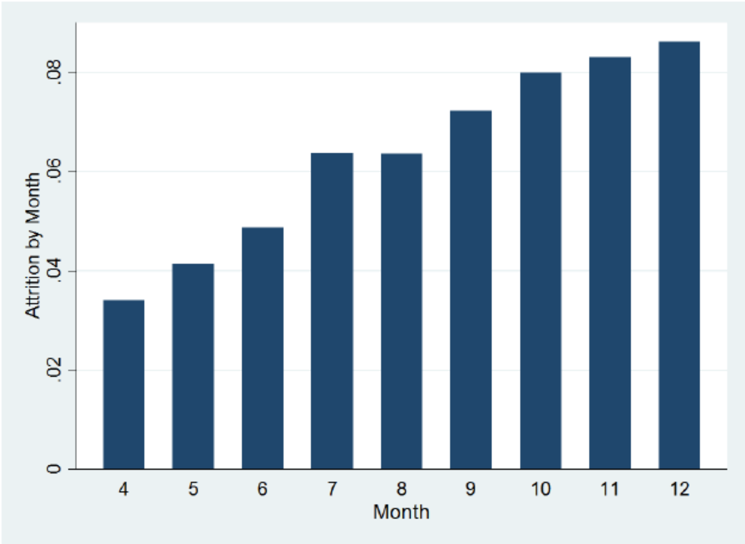
Outcome	Control Mean (1)	SD (2)	Transport (3)	Workshop (4)	N (5)	F-test P (6)
degree	0.18	0.39	0.01	-0.01	3049	0.347
			(0.63)	(0.74)		
vocational	0.43	0.49	0.01	0.01	3049	0.717
			(0.82)	(0.59)		
work	0.31	0.46	-0.01	-0.02	3049	0.881
			(0.61)	(0.56)		
search	0.50	0.50	-0.01	0.00	3049	0.804
			(0.83)	(0.96)		
dipdeg	0.25	0.43	0.00	-0.01	3049	0.557
			(0.94)	(0.68)		
female	0.52	0.50	0.00	0.00	3049	0.968
			(0.98)	(0.96)		
migrant_birth	0.37	0.48	0.01	-0.01	3049	0.530
			(0.72)	(0.84)		
amhara	0.46	0.50	-0.01	-0.06	3049	0.078
			(0.87)	(0.11)		
oromo	0.26	0.44	-0.00	0.02	3049	0.489
			(0.88)	(0.59)		
work_wage_6months	0.46	0.50	-0.00	-0.01	3049	0.659
			(0.99)	(0.67)		
married	0.20	0.40	0.01	-0.03	3049	0.131
			(0.81)	(0.26)		
live_parents	0.52	0.50	-0.01	0.01	3049	0.451
			(0.79)	(0.66)		
experience_perm	0.13	0.34	0.00	-0.01	3049	0.370
			(0.84)	(0.56)		
search_6months	0.75	0.43	-0.01	0.00	3049	0.606

Figure 6.11: Summary and Tests of Balance

If the samples are unbalanced, we should at least control for unbalanced variables later.

Data Collection

A combination of face-to-face interviews and phone follow ups. The *attrition rate is very low*:



Note. Attrition is defined as failure to complete one interview.

Figure 6.12: Attrition Rate from the Phone Survey by Month

6.2.4 Estimation

Pre-Analysis Plan

The authors follow a detailed pre-analysis plan that describes the empirical strategy, outcome variables, subgroup analysis, etc.

For RCTs, pre-analysis plans are typically important: they mitigate the possibility of ex post data snooping. Their effectiveness can be illustrated in this figure:

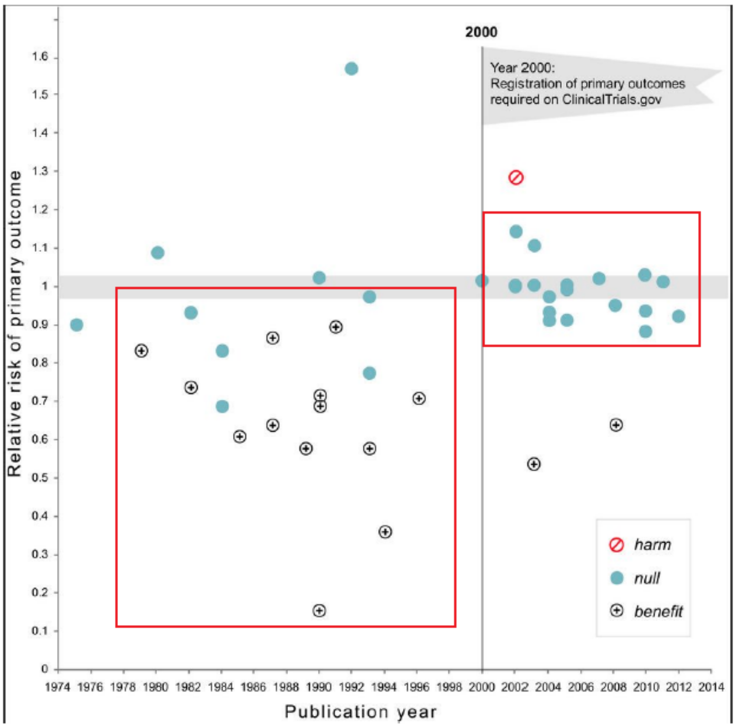


Figure 6.13: Less Significant Results after Requiring Pre-Analysis Plan

The number of researches which claim the effectiveness of a drug decreased dramatically after the requirement of a pre-analysis plan.

Main Regression

$$y_{ic} = \beta_0 + \sum_f \left[\beta_f \text{treatment}_{fic} + \gamma_f \text{spillover}_{fic}\right] + \alpha y_{ic,pre} + \delta x_{ic0} + \mu_{ic}$$

where:

- treatment$_{fic}$: treatment dummy = 1 if an individual was offered treatment f
- spillover$_{fic}$: spillover dummy = 1 if an individual resided in a treated cluster
- x_{ic0}: baseline covariates used for re-randomisation and blocking (and to improve precision)

This regression can capture the *intention-to-treat (ITT)* parameter, and we cluster standard errors at the geographical clusers level.

6.2.5 Results

Overall Effect

	2015				2018			
Outcome	Control mean (1)	Transport (2)	Workshop (3)	Equality (pval) (4)	Control mean (5)	Transport (6)	Workshop (7)	Equality (pval) (8)
Work	0.562	0.037	0.021	0.57	0.693	-0.058*	0.029	0.00
		(0.029)	(0.031)			(0.035)	(0.032)	
		[0.366]	[1.000]			[0.411]	[0.958]	
Hours worked	26.176	0.183	-0.214	0.79	28.250	-2.499*	0.218	0.04
		(1.543)	(1.533)			(1.486)	(1.426)	
		[0.837]	[1.000]			[0.411]	[1.000]	
Monthly wages	857.882	65.879	3.363	0.30	1,531.488	30.916	299.469**	0.02
		(63.864)	(65.667)			(102.352)	(121.383)	
		[0.437]	[1.000]			[0.753]	[0.096]	
Permanent job	0.171	0.033*	0.069***	0.09	0.307	-0.034	-0.010	0.30
		(0.018)	(0.019)			(0.025)	(0.028)	
		[0.215]	[0.004]			[0.411]	[1.000]	
Formal job	0.224	0.054***	0.053***	0.95	0.318	-0.005	-0.007	0.96
		(0.019)	(0.020)			(0.030)	(0.030)	
		[0.032]	[0.021]			[0.753]	[1.000]	
Job satisfaction	0.237	-0.001	0.022	0.45	0.575	-0.025	0.066*	0.01
		(0.027)	(0.027)			(0.037)	(0.036)	
		[0.837]	[1.000]			[0.593]	[0.219]	

Note. In this table we report the *intent-to-treat* estimates of the direct effects of the transport intervention and the job application workshop on primary employment outcomes. These are obtained by OLS estimation of equation (1), weighting each observation by the inverse of the probability of being sampled. Below each coefficient estimate, we report the *s.e.* in parentheses and a q-value in brackets. We correct standard errors to allow for arbitrary correlation at the level of geographical clusters. q-values are obtained using the sharpened procedure of Benjamini et al. (2006). We do this for the data from the first endline in 2015 (Columns 1-4) and then for second endline in 2018 (Columns 5-8). For each endline we report the mean outcome for the control group, the p-value from a F-test of the null hypothesis that transport subsidies and the job application workshop have the same effect, and the number of observations. ***: $p < 0.01$, **: $p<0.05$, *: $p<0.1$.

Figure 6.14: ITT Estimates

Soon after the interventions (2015), significant positive effects are found on participants' rate of entering permanent jobs and formal jobs (orange square). However, 3 years later (2018), significant positive effects are only found for workshops (on monthly wages and job satisfaction). In contrast, transport subsidy is estimated to have some negative effects on employment and hours worked.

Impact Trajectories: Employment and Permanent Employment

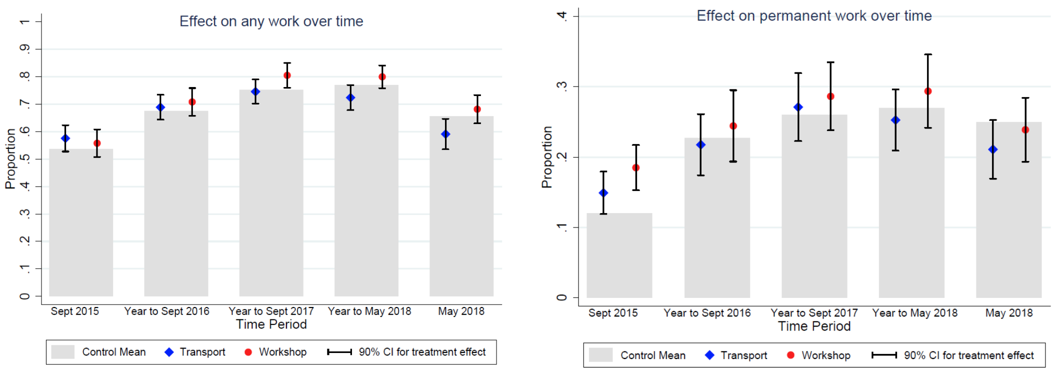


Figure 6.15: Impact Trajectories: Employment and Permanent Employment

Effects of both treatments diminish over time with effects of transport treatment diminishing quicker.

Search Effort

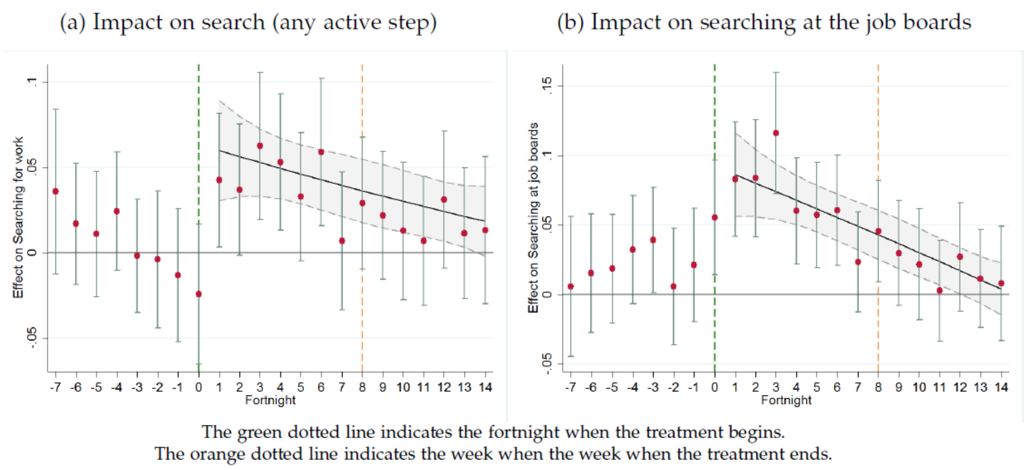


Figure 6.16: Fortnightly Impacts of the Transport Treatment on Job Search

Transport subsidy facilitates job searching immediately after treatment, but such effect vanishes quickly: after 14 days, it becomes insignificant.

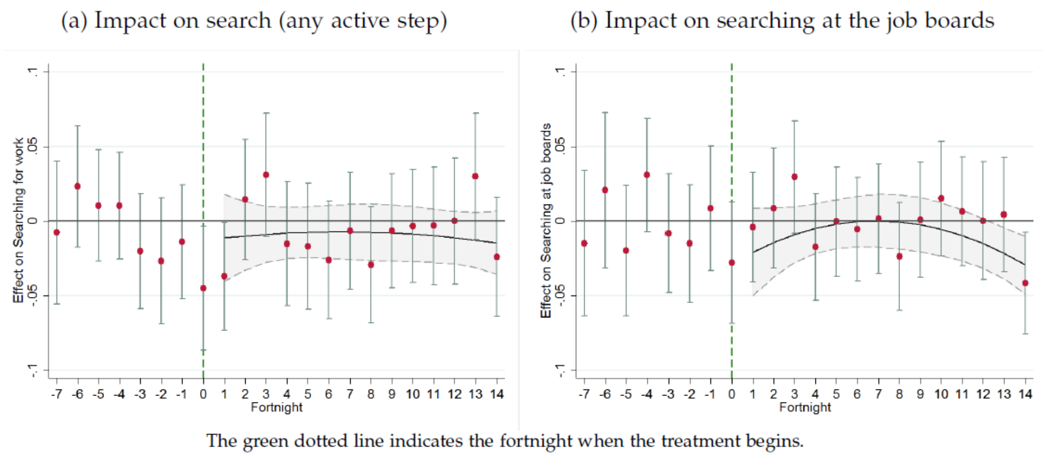


Figure 6.17: Fortnightly Impacts of the Job Application Workshop on Job Search

On the other hand, job application workshop has no effect on job search intensity at all. This indicates that the benefits brought by workshops are likely to be caused by better information efficiency between employers and job-seekers.

Match Quality

Outcome	Control mean (1)	N (2)	ITT Estimates: Transport Coeff (3)	ITT Estimates: Workshop Coeff (4)	Equality pval (5)
Longest tenure (months)	11.845	1,739	0.294	1.197*	0.103
			(0.561)	(0.619)	
Current job tenure (months)	21.326	1,383	0.199	-0.539	0.536
			(1.165)	(0.977)	
Promoted in current job	0.190	1,383	0.022	0.006	0.525
			(0.025)	(0.023)	
Uses skills in current job	0.323	2,016	0.032	0.082**	0.211
			(0.040)	(0.040)	
Earnings conditional on working	2,209.3	1,383	195.0	370.4**	0.283
			(143.1)	(157.6)	

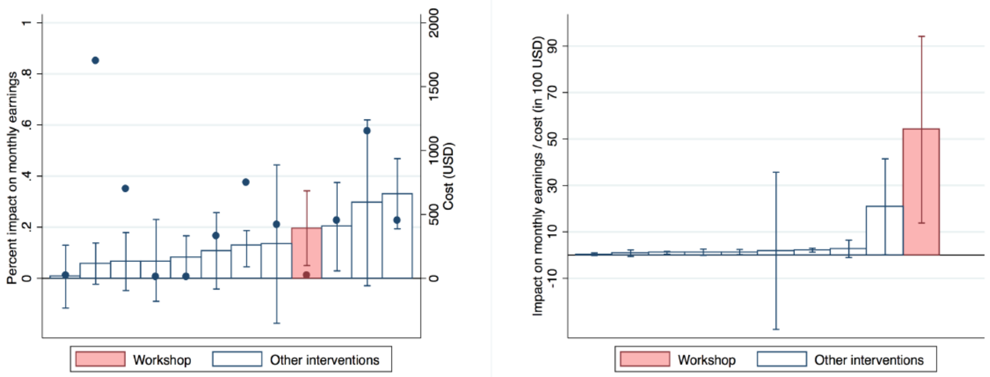
Note. In this table we report the *intent-to-treat* estimates of the impacts of the transport intervention and the job application workshop on several outcomes related to match-quality. These are obtained by OLS estimation of equation (1), weighting each observation by the inverse of the probability of being sampled. Below each coefficient estimate, we report the *s.e.* in parentheses. We correct standard errors to allow for arbitrary correlation at the level of geographical clusters. ***: $p < 0.01$, **: $p<0.05$, *: $p<0.1$.

Figure 6.18: Impacts on Job Tenure and Conditional Earnings

After intervention, workshop treatment shows evidences of improvement in several job matching quality indicators (longest tenure, uses skills in current job, and earnings conditional on working). On the other hand, no significant effects are found for transport subsidy treatment.

Workshop Compared with Other Interventions

(a) Impacts on earnings and cost (b) Impact/cost ratio



Note. We report the estimates of the studies that report earning effects which are included in the review by McKenzie (2017). For some studies, we obtain additional information from the papers (e.g. for Maitra and Mani (2017)).

Figure 6.19: Comparison with Other ALMPs in Developing Countries

Compared with other interventions, job application workshops have significant positive effects on participants' monthly wages with low implementation costs, which endorse it with the highest impact/cost ratio among all policies.

Heterogeneous Effects

	Covariate = 0			Covariate = 1			Transport	Workshop
Baseline covariate	Control mean (1)	Transport (2)	Workshop (3)	Control mean (4)	Transport (5)	Workshop (6)	Equality (pval) (7)	Equality (pval) (8)
Tertiary Education	826.4	15.1 (124.4) [1.000]	470.9** (188.1) [0.034]	1,835.1	54.2 (159.9) [1.000]	37.3 (149.8) [0.993]	0.83	0.07
Male	1,181.9	-40.0 (110.0) [1.000]	132.1 (116.4) [0.087]	1,892.4	104.7 (179.3) [1.000]	475.5* (245.1) [0.363]	0.47	0.21
Active searcher	1,442.2	3.1 (132.7) [1.000]	351.9* (188.9) [0.050]	1,625.8	62.5 (160.0) [1.000]	235.5 (183.1) [0.663]	0.77	0.67
Ever had permanent job	1,465.8	40.2 (104.7) [1.000]	356.5*** (136.7) [0.034]	1,975.7	-42.3 (367.8) [1.000]	-288.7 (350.3) [0.696]	0.82	0.09
Lives close to the centre	1,468.8	41.8 (151.0) [1.000]	406.2** (196.9) [0.042]	1,606.3	52.2 (143.0) [1.000]	141.9 (150.3) [0.696]	0.96	0.29
Predicted endline earnings (above the median)	930.8	123.1 (115.5)	467.1*** (170.3)	2250.4	-226.4 (227.8)	-99.0 (224.1)	0.475	0.0696

Figure 6.20: Heterogeneous Effects on 2018 Wage Earnings by Baseline Characteristics

Treatment effects are heterogeneous within the sample population. In the figure above, the column named "Covariate = 0" indicates the subsample whose baseline covariate (each row) = 0. For example, the orange square on row 1 column 3 indicates the estimated ITT of workshop on an individual without tertiary education.

We can see that while there's no heterogeneity in ITT of transport subsidy (since it is ineffective for everyone...), obvious heterogeneity can be observed in the ITT of job application workshop: disadvantageous participants (with no tertiary education, did not ever have a permanent job, or with predicted endline earnings below median) are more likely to benefit from workshops.

Bibliography

Abebe, G., Caria, A. S., Fafchamps, M., Falco, P., Franklin, S., & Quinn, S. (2021). Anonymity or distance? job search and labour market exclusion in a growing african city. *The Review of Economic Studies*, *88*(3), 1279–1310. https://doi.org/10.1093/restud/rdaa057

Ponczek, V., & Ulyssea, G. (2022). Enforcement of labour regulation and the labour market effects of trade: Evidence from brazil. *The Economic Journal*, *132*(641), 361–390. https://doi.org/10.1093/ej/ueab052

The World Bank. (2022). World development indicators. https://datatopics.worldbank.org/world-development-indicators/

Witte, M. (2018). Job referrals and strategic network formation, 88.

Chapter 7

Monetary Policy Objectives and Monitoring the Economy

7.1 Basic Concepts

7.1.1 What is a Central Bank?

Definition: Central banks are public (or quasi-private) institutions that manage the money supply and sometimes the banking system of a country or monetary union.

Kay Features:

- They act as "Lenders of Last Resort" to commercial banks to prevent bankruns
- A key feature of modern central banks – *Independence*
- Some central banks are responsible for much more than monetary policy:
  - Financial regulation
  - Printing/circulating/retiring physical currency
  - Macroprudential policy (smoothing the credit cycle)

7.1.2 What is Monetary Policy?

Definition: Monetary policy is the management of the money supply to further some economic objective(s).

Idea behind monetary policies: by moderating money available, you can alter behaviour of agents. Monetary policy exploits *trade off* between saving/borrowing and consumption.

Monetary policy stance is usually summarised as "setting an interest rate" (sometime "policy rate"): cutting rates refers to "expansionary" while hiking rates refers to "contractionary".

Central banks may not actually set these interest rates directly, but only *target* for them: in US, the Federal Reserve targets the Federal Fund Rate (FFR) – the average rate charged by one bank to another bank to (uncollateralised) borrow reserves at the Federal Reserve overnight. It's a market rate.

7.1.3 Transmission of Interest Rates and the Yield Curve

Transmission

Monetary policy only sets/targets the overnight rate. To impact the economy, it has to be transmitted through the economy: a rise in the policy rate increases the borrowing cost of banks, so banks will charge other institutions/individuals to borrow in order to maintain their profit margins.

How Does Interest Rates Affects Economy? Consumption Euler Equation

Here, we discuss a simple model for intuition. Assuming log utility and perfect information (no uncertainty), we can derive the Consumption Euler equation:

$$\frac{C_{t+1}}{C_t} = \beta(1 + r_t) = \beta\frac{1 + i_t}{1 + \pi_t}$$

where:

- C_{t+1}, C_t are consumption in the corresponding period
- β represents discount factor
- i represents nominal interest rate
- π represents inflation rate

We can see that, if the nominal interest rate i_t goes up, individuals will be more willing to sacrifice their current consumption C_t in exchange for higher future consumption C_{t+1}.

Yield Curve

Monetary policies changes are reflected by movements of the yield curve.

Overnight lending of reserves is *risk-free*, but other interest rates build in a *risk premium* (due to default risk, inflation risk, interest rate risk, etc.) Long-term interest rates integrate *expected rates* over the duration of the loan plus *risk premium* and *term premium*.

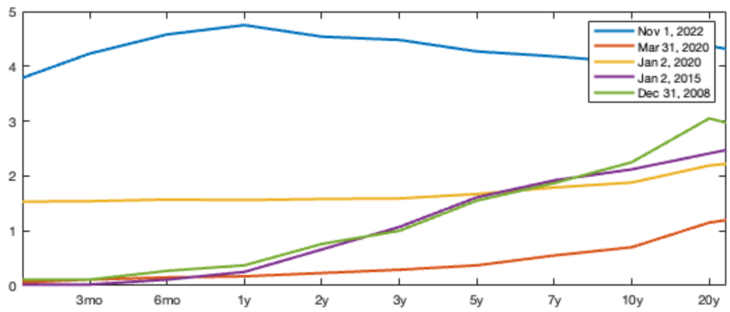


Figure 7.1: Yield Curves

Typically, yield curve if a healthy economy should be upward sloping due to positive risk premium and term premium. An inverted yield curve implies potential economic crisis: it predicts that the central bank will have to cut policy rate dramatically to fight recession in the future. A yield curve hitting the zero lower bound is also undesirable because it means conventional monetary policy instrument is exhausted.

Corporate Bonds?

The yield curves above are built upon government bonds. Corporate bonds will have higher yield curves due to default risks.

7.1.4 Central Bank Mandates

legal Mandates and Dual Mandate Trade-off

Central banks are generally given very specific legal mandate:

- Bank of England: psudo dual mandate
  1. (primiary) Price stability: around 2% inflation
  2. (secondary) Suppot government's growth and employment objectives
- Federal Reserve: dual mandate

  * Maximum employment
  * and Stable prices (2% inflation)
- ECB: only one objective (it is too hard to accommodate business cycles of all Eurozone members)
  * Price stability (2% inflation)

A dual mandate necessitates a trade-off between inflation and unemployment. According to the Neo-Keynesian Phillips Curve:

$$\pi_t = \beta E[\pi_{t+1}] + \kappa u_t, \kappa < 0$$

Any central bank with dual mandates will need to act with two conflicting objectives. Furthermore, there is a hot debate on whether we should set even more objectives for central banks, such as environment and equality.

Flat Philips Curve - Monetary Policy A "Solved Problem"?

From the mid 1990s, inflation has been stabilised around 2% in many developed countries – and even lower after the 2008 GFC. Some central bank started to worry about inflation being too low.

Meanwhile, there seems to be a flat Phillips Curve, indicating that there is no inflation-unemployment trade-off. However, on the flip side, policymakers cannot simply sacrifice employment to bring down inflation – is today's inflation out of control?

7.1.5 Challenges to Central Bank Independence

Independence from political authority is a key tenet of modern central banks. Such insulation from politics allows that:

- Central banks can have long-term thinking, not in a 4-year cycle
- Monetary tools will not be used for political objectives (such as pumping up employment for elections)

When lines are blurred, inflation runs rampant, as in the Argentinean experience. However, threats to central banks' Independence were witnessed recently:

- U.S.: retaliation to perceived "mandate creep:" critics accused the Federal Reserve for going beyond its mandate. This may act as an excuse for political intervention.
- U.K.: the Bank of England directly purchased government bonds from the treasury. Typically, central banks buy government bonds from secondary markets to prevent the government from exploiting the "unqualified purchase" and over-expanding fiscal policies.

7.1.6 Monetary Policy Implementation

Conventional

The Bank of England sets the base rate directly.

Federal Reserve and others target policy rates by conducting Open Market Operations: they purchase and sell government bonds on secondary markets in exchange for reserves held at the central bank to move the rate:

- Sell bonds $\rightsquigarrow$ Less reserves held by commercial banks $\rightsquigarrow$ Banks cut lending or charge higher rates
- Buy bonds $\rightsquigarrow$ More reserves held by commercial banks $\rightsquigarrow$ Banks lend more or charge lower rates

These transactions used to work with the minimum reserve requirement, but now on interest on reserves. Theoretically, changing interest rate paid on reserves have the same effect.

Unconventional

In 2008, interest rates hit the Zero Lower Bound (nominal IR cannot be cut lower than 0, otherwise people hold cash), but the economy was still trapped in recession, so unconventional monetary policy instruments were adopted:

- Forward Guidance: non-binding statements by central banks
  Forward guidance makes a promise about future interest rates, shifting investors' expectation of future interest rates and hence shifting the long-end of yield curve. This works theoretically but empirical evidence is unclear.
- Quantitative Easing: circumvent the yield curve / transmission and purchase bonds directly
  Quantitative easing purchase government bonds, corporate bonds, MBS, etc. directly. Such intervention increases demand and depress supply for those assets, so their price will increase, lowing the yield.
- Central Bank Information Effects?
  The market infer from central banks' actions. Sometime, this causes undesirable consequences: an unexpected low interest rate decision may lead the market to think about "the central bank believes the situation is worse than we evaluated."

7.1.7 Stagflation

Current economic conditions can be characterised as Stagflation: the economy is stagnant with high inflation rate.

Dilemma: for dual mandate central banks, persistently high inflation calls for more aggressive rate hikes, but such actions will push the economy into further recessions.

Further bad news is that today's unemployment is *artificially low*: since 2008, many people have dropped out of the labour force, and they are not counted as unemployed. This has been even worse since the pandemic.

7.1.8 Challenge of the Pandemic – Supply not Demand

The pandemic presents a unique challenge for monetary policy: monetary policy is primarily equipped to *stimulate demand*, but despite an initial drop in demand, most of post-pandemic recession was *supply driven*.

Monetary policy is not likely to have an immediate effect on supply – no amount of interest rate cuts can alleviate lockdown at factories or reopen restaurants. Nevertheless, central banks lowered interest rates to zero immediately. (so "no one can accuse us for doing nothing")

Because the fast pace of the pandemic recession, there is also a challenge for obtaining up-to-date data, discussed at the end of this chapter (section 7.2.4).

7.2 Policy Making and Data Collection

7.2.1 The Policy Cycle

Central banks continuously assess the state of the economy and meet on a regular schedule to decide appropriate action.

In both UK (Monetary Policy Committee) and US (Federal Open Market Committee), there are 8 meetings a year. Research and policy team prepare forecasts, briefings, and make recommendation to committees, and committees make policy decisions.

Occasionally, typically during crises, unscheduled meetings are held for quicker adjustments in policies.

7.2.2 Collecting Data

Assessing the state of the economy is a major challenge due to:

- Release Lag: crucial macroeconomic data is often unavailable. Key releases such as GDP are only available a month after a quarter ends.

- Data Vintages and Revisions: available values of data may be inaccurate. There are at least four vintages of US GDP data, and change from the first to the fourth can be significant (the first one usually has a large error).

An aside: policymakers look at core inflation, which excludes highly volatile food and energy prices, instead of headline inflation data, but food and energy are important in the real life.

7.2.3 Nowcasting

Nowcasting is similar to forecasting, but with an focus on current period's data.

The idea is to infer slow releases such as GDP using quick releases for the current period. We take all data that is available for the current month/quarter, and use historical relationships to predict data that is not yet released.

Factor models and Principal Components Analysis

Nowcasts are typically dynamic factor models using many time series. They are a dimension reduction tool to express a big data set in terms of a few (unknown) underlying time series. Some quickly released data, such as unemployment claims, labour market reports, etc., becomes very important.

Formally, factor models express the whole data set X_t as:

$$X_t = \Lambda F_t + \mu_t, \text{with } dim(X_t) > dim(F_t)$$

F_t here is latent / unobserved. The simplest way to estimate Λ and F_t is to find the uncorrelated factors that explain as much variations in X_t as possible, which is known as Principal Component Analysis:

- Find the time series that explains the variations in X_t the most
- Find a second time series that explain the variations in the residuals in X_t the most, after partialing out the first factor
- ...

The number of factors/principals is determined using *screen plots* or *information criterion*.

One caveat for using such principal component analysis is that estimated $\hat{F}_t$ for a fixed date t depends on the sample used: different vintages of data may yield different estimates. Also, incorporating freshly released data also changes the estimates obtained from previous data.

Financial Data

While macroeconomic data comes with a delay, financial data is produced continuously. However, the challenge is that financial markets are not part of central banks' mandates, so policymakers have to infer from financial data about variables they care about. For example:

- Inferring lending conditions, which are relevant for interest rate decisions
- Derivatives reflect inflation expectations
- Yield curve reflects expectations of future economic conditions

7.2.4 Data during the Pandemic and The Weekly Economic Index

Typically, recessions build over time, so policymakers have some time to collect data, evaluate the situation, and make decisions. However, recessions caused by the pandemic are very fast-pace with abrupt deterioration of economic conditions. In March 2020, policymakers needed data immediately – not in later April when GDP will be released.

Because no traditional data on real activities satisfy the needs, Dr. Daniel Lewis and his colleagues developed the Weekly Economic Index (WEI). They used a panel of 10 non-standard series reflecting consumer activity, industrial production, labour market, etc., that were available at daily or weekly frequency. High frequency data is very noisy, but common factors retrieved are more reliable. WEI measured weekly fluctuations and nowcast GDP well during 2020.

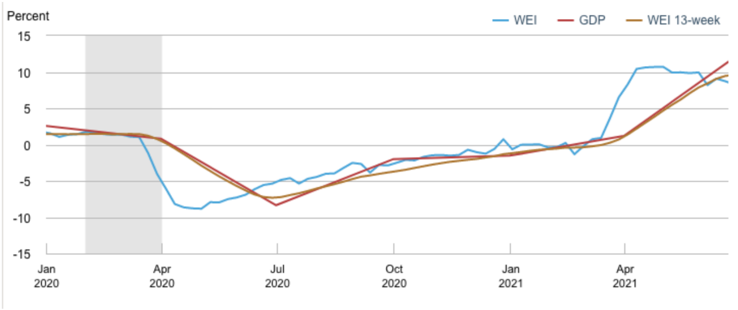


Figure 7.2: Weekly Economic Index (WEI) Nowcasts and Actual GDP

Chapter 8

Estimating the Effects of Monetary Policy[1]

8.1 Introduction

We need to evaluate the effectiveness of monetary policies, but face the following difficulties:

- We cannot conduct RCTs for monetary policies due to feasibility and morality reasons
- We cannot rely on natural experiments because there will be no "control group"

Therefore, we resort to econometric strategies and assumptions. We cannot simply regress outcomes on changes in interest rates due to endogeneity (simultaneity, feedback, etc.), and this will not inform us about the effectiveness of unconventional policies.

8.2 ⋆ Vector Auto-Regression (VAR)

8.2.1 Vector Auto-Regression (VAR) Models

An AR(p) model:

$$y_t = \alpha_y + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + u_{y,t}$$

We can extend this by adding other covariates, and this forms an Auto-regressive Distributed Lag (ADL) Model.

$$\begin{aligned} y_t = \alpha_y &+ \underbrace{\beta_{y,1} y_{t-1} + \beta_{y,2} y_{t-2} + \cdots + \beta_{y,p} y_{t-p}}_{\text{Lags of } y} \\ &+ \underbrace{\beta_{x,1} x_{t-1} + \beta_{x,2} x_{t-2} + \cdots + \beta_{x,p} x_{t-p}}_{\text{Lags of } x} \\ &+ \underbrace{\beta_{z,1} z_{t-1} + \beta_{z,2} z_{t-2} + \cdots + \beta_{z,p} z_{t-p}}_{\text{Lags of } z} + u_{y,t} \end{aligned} \tag{8.1}$$

Express this equation (8.1) in its vector form:

$$y_t = \alpha_y + \sum_{l=1}^{p} \begin{pmatrix} \beta_{l,y}, & \beta_{l,x}, & \beta_{l,z} \end{pmatrix} \begin{pmatrix} y_{t-l} \\ x_{t-l} \\ z_{t-l} \end{pmatrix} + u_{y,t} \tag{8.2}$$

Apply this vector form of ADL model (8.2) to inflation π_t, unemployment rate U_t, and interest rate r_t:

[1] All equations are important in this chapter, so I did not highlight them one-by-one.

$$\begin{cases} \pi_t = \alpha_\pi + \sum_{l=1}^{p} \begin{pmatrix} \beta_{l,\pi}, & \beta_{l,U}, & \beta_{l,r} \end{pmatrix} & \begin{pmatrix} \pi_{t-l} \\ U_{t-l} \\ r_{t-l} \end{pmatrix} + u_{\pi,t} \\ U_t = \alpha_U + \sum_{l=1}^{p} \begin{pmatrix} \rho_{l,\pi}, & \rho_{l,U}, & \rho_{l,r} \end{pmatrix} & \begin{pmatrix} \pi_{t-l} \\ U_{t-l} \\ r_{t-l} \end{pmatrix} + u_{U,t} \\ r_t = \alpha_r + \sum_{l=1}^{p} \begin{pmatrix} \gamma_{l,\pi}, & \gamma_{l,U}, & \gamma_{l,r} \end{pmatrix} & \begin{pmatrix} \pi_{t-l} \\ U_{t-l} \\ r_{t-l} \end{pmatrix} + u_{r,t} \end{cases} \tag{8.3}$$

Again, stack the three equations in 8.3 vertically into a Vector Auto-Regression (VAR) model:

$$\underbrace{\begin{pmatrix} \pi_t \\ U_t \\ r_t \end{pmatrix}}_{Y_{t,3\times 1}} = \underbrace{\begin{pmatrix} \alpha_\pi \\ \alpha_U \\ \alpha_r \end{pmatrix}}_{\alpha_{3\times 1}} + \sum_{l=1}^{p} \left\{ \underbrace{\begin{pmatrix} \beta_{l,\pi}, & \beta_{l,U}, & \beta_{l,r} \\ \rho_{l,\pi}, & \rho_{l,U}, & \rho_{l,r} \\ \gamma_{l,\pi}, & \gamma_{l,U}, & \gamma_{l,r} \end{pmatrix}}_{B_{l,3\times 3}} \underbrace{\begin{pmatrix} \pi_{t-l} \\ U_{t-l} \\ r_{t-l} \end{pmatrix}}_{Y_{t-l,3\times 1}} \right\} + \underbrace{\begin{pmatrix} u_{\pi,t} \\ u_{U,t} \\ u_{r,t} \end{pmatrix}}_{u_{t,3\times 1}} \tag{VAR}$$

$$Y_{t,3\times 1} = \alpha_{t,3\times 1} + \sum_{l=1}^{p} \left\{ B_{l,3\times 3} Y_{t-l,3\times 1} \right\} + u_{t,3\times 1}$$

Simple VARs like this are useful for forecasting.

8.2.2 Impulse Response Functions

An Impulse Response Function (IRF) is the change in one variable associated with a change in another at some lag, which is the cornerstone of empirical macroeconomics.

Conceptually, we think of these as a response of Y_{t+h} to a change in $u_{j,t}$, because changing $u_{j,t}$ does affect variables in previous periods.

- A contemporary ($h = 0$) IRF of a change in $u_{j,t}$ is:

$$\Phi_j^0 = \frac{\partial Y_t}{\partial u_{j,t}} = \iota_j$$

  where ι_j is a vector with 1 in the jth entry and 0 otherwise.

- IRF (Φ) at horizon h can be computed recursively from knowledge of $B_1, \ldots, B_p$:

$$\Phi_j^h = \sum_{k=1}^{min(p,h)} B_k \Phi_j^{h-k} = \sum_{k=1}^{min(p,h)} B_k \frac{\partial Y_t}{\partial u_{j,t-h+k}}$$

8.3 ⋆ Causal Analysis and Structural VAR (SVAR)

8.3.1 A Conceptual Framework for Causal Analysis

We want to study responses to changes in policy that are unpredictable.

If it is predictable:

- The estimated effects are caused by whatever predicts the policy
- Rational gents have already responded to the anticipated policy change

Formally, a change is endogenous if it is predicted or caused by another change. A change is exogenous if it is unpredictable and not a causal result of something else. A movement in a macroeconomic time series that is exogenous is a structural shock. We want to study those structural shocks as as-if random variations in policy, or surprises.

8.3.2 Structural VAR (SVAR) Models

Setup

The vector of VAR residuals u_i is unpredictable but they are *not shocks* because they are cross-correlated, and they are *endogenous*. Exogenous shock should be uncorrelated.

The Structural VAR (SVAR) model assumes the residuals u_t in VAR as linear combinations of n unobserved uncorrelated shocks ϵ_t (SVAR assumption 1). With this, we rewrite the VAR equation:

$$\underbrace{\begin{pmatrix}\pi_t\\ U_t\\ r_t\end{pmatrix}}_{Y_{t,3\times1}} = \underbrace{\begin{pmatrix}\alpha_\pi\\ \alpha_U\\ \alpha_r\end{pmatrix}}_{\alpha_{3\times1}} + \sum_{l=1}^{p}\left\{\underbrace{\begin{pmatrix}\beta_{l,\pi}, & \beta_{l,U}, & \beta_{l,r}\\ \rho_{l,\pi}, & \rho_{l,U}, & \rho_{l,r}\\ \gamma_{l,\pi}, & \gamma_{l,U}, & \gamma_{l,r}\end{pmatrix}}_{B_{l,3\times3}}\underbrace{\begin{pmatrix}\pi_{t-l}\\ U_{t-l}\\ r_{t-l}\end{pmatrix}}_{Y_{t-l,3\times1}}\right\} + \underbrace{\begin{pmatrix}A_{11}, & A_{12}, & A_{13}\\ A_{21}, & A_{22}, & A_{23}\\ A_{31}, & A_{32}, & A_{33}\end{pmatrix}}_{A_{3\times3}}\underbrace{\begin{pmatrix}\epsilon_1\\ \epsilon_2\\ \epsilon_3\end{pmatrix}}_{\epsilon_{t,3\times1}} \tag{SVAR}$$

$$Y_t = \alpha + \sum_{l=1}^{p} B_l Y_{t-l} + \underbrace{A\epsilon_t}_{u_t}$$

and the shocks are:

- $E[\epsilon_t] = 0$ (has zero mean)
- $E[\epsilon_t \epsilon_t'] = I_n$ (has unit variance and zero covariance)
- $E[\epsilon_t \epsilon_s] = 0$ (no serially correlated)

The full-rank matrix A is the Contemporaneous Response Matrix. A_{ij} is the effect of a unit shock j on variable i instantly, which is interpreted causally.

Estimate the Effects of Shocks with Assumptions

$E[u_t] = AE[\epsilon_t] = A \times 0 = 0$ informs us nothing.

$E[u_t u_t'] = AE[\epsilon_t \epsilon_t']A' = AA'$ There are $n(n+1)/2$ equations with n^2 unknowns, so we need additional $n(n-1)/2$ restrictions to find a unique solution.

The further assumptions restraining the unknowns (SVAR assumption 2) take the simplest form of recursive/timing restrictions: we restrict which variables can respond to which shocks contemporaneously. Technically, we impose that:

$$A = \begin{pmatrix}A_{11} & 0 & 0\\ A_{21} & A_{22} & 0\\ A_{31} & A_{32} & A_{33}\end{pmatrix} \tag{SVAR Assumption 2}$$

With ordering $Y_t = \begin{pmatrix}\pi_t\\ U_t\\ r_t\end{pmatrix}$, these restrictions means:

- Inflation only responds to price shocks contemporaneously
- Unemployment responds to both price and labour market (unemployment) shocks contemporaneously
- Interest rate responds to all price, labour market (unemployment), and monetary policy (interest rate) shocks contemporaneously

Under these restrictions , A is the unique lower triangular Cholesky factor of $Cov(u_t) = E[u_t u_t'] = AA'$, and all its entries can be estimated.

Structural Impulse Responses

After estimated the corresponding coefficients in SVAR, we can compute the structural impulse responses – the *dynamic causal effects* of monetary policy.

The structural impulse response of a structural shock ϵ_j as time t on variable Y at time $t+h$ is:

$$\begin{aligned}\Theta_j^h &= \frac{\partial Y_{t+h}}{\partial \epsilon_{j,t}} \\ &= \frac{\partial Y_{t+h}}{\partial u_t'}\frac{\partial u_t}{\partial \epsilon_{j,t}} &&\text{(Chain Rule)} \\ &= \Phi^h A_{.j}\end{aligned}$$

where Φ^h is the h period IRF, and $A_{.j}$ is the jth row of A.

8.4 ⋆ Alternatives to SVAR

8.4.1 Alternative Restrictions

The timing assumptions we used (SVAR Assumption 2) are very strong, so we may want to use others:

- Long-run Restrictions: assume effects of some shocks die out at long horizons
- Sign Restrictions: assume some elements in A have certain signs (e.g. interest hike lowers inflation). This rules out some solutions for A and gives us *a range of possible values*, but not unique point estimates.

8.4.2 Alternative Method: Instrumental Variables

Still using the structure of SVAR, we can use instrumental variables (e.g. an instrument that is correlated with monetary policy shock but not price or labour market shocks) to estimate the corresponding column of A.

To implement this, we regress each series in u_t on instrument z_t by OLS, then normalise the vector of coefficients by the coefficient of $u_{r,t}$ to get A_{mp}.

8.4.3 Alternative Method: Local Projections (LP)

Local Projections

Local Projection (LP) is an alternative method to estimate the effects of monetary policy (or other macro policies). Generally, this is more flexible than VARs. LPs are direct OLS regressions to forecast some outcome:

$$y_{t+h} - y_{t-1} = a + \lambda_h r_t + \sum_{l=1}^{p} \kappa_l X_{t-l} + v_t \qquad \text{(Local Projections)}$$

where X_{t-l} is a set of controls (inflation, unemployment, etc.)

The IRF at horizon h is simply λ_h. However, note that, because r_t is likely to be endogenous, λ_h does not have a causal interpretation yet. It is just the response to the part of r_t that is unpredictable by $X_{t-1}, \ldots, X_{t-p}$.

Local Projections with Instrumental Variables

To deal with endogeneity and obtain causal responses, LPs are always implemented with instrumental variables.

The implementation is just like 2SLS:

- First stage: regress r_t on IV z_t and controls (lags of X); predict $\hat{r}_t$
- Second stage: regress $(y_{t+h} - y_{t-1})$ on $\hat{r}_t$ and controls (lags of X)

λ_h estimated in the second stage is the dynamic causal effect at horizon h.

When choosing instruments, as usual, we need to ensure exogeneity and relevance. Adding controls can help with their validity. Some examples of IV: high frequency financial data and "narrative" rate changes regressed on internal CB forecasts.

In general, we should not use measured shocks directly as independent variables (regressors), because there are measurement errors and scaling issues.

8.5 Empirical Findings

8.5.1 Literatures

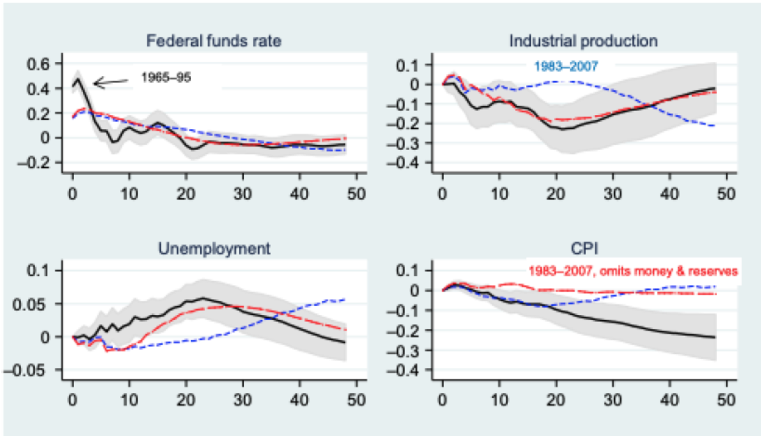


Figure 8.1: Christiano, Eichenbaum, & Evans (1999): Recursive SVAR (Lower Triangle)

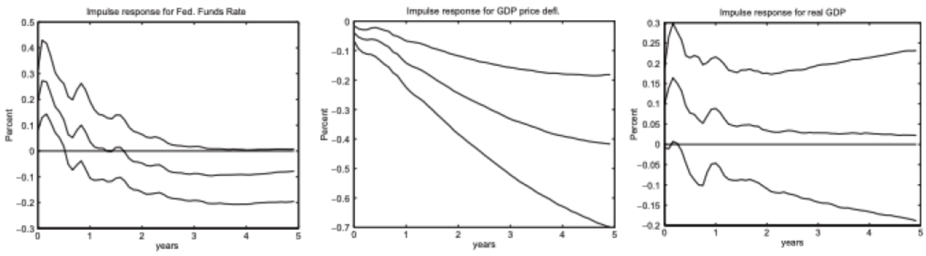


Figure 8.2: Uhlig (2005): SVAR with Sign Restrictions

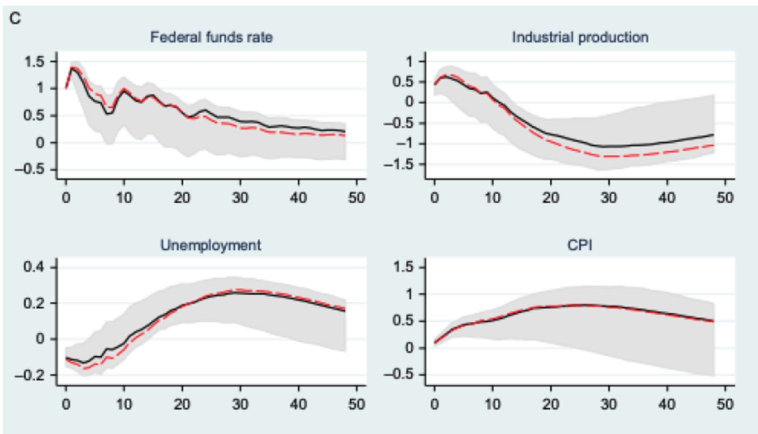


Figure 8.3: Romer & Romer (2004): SVAR with Instrumental Variables

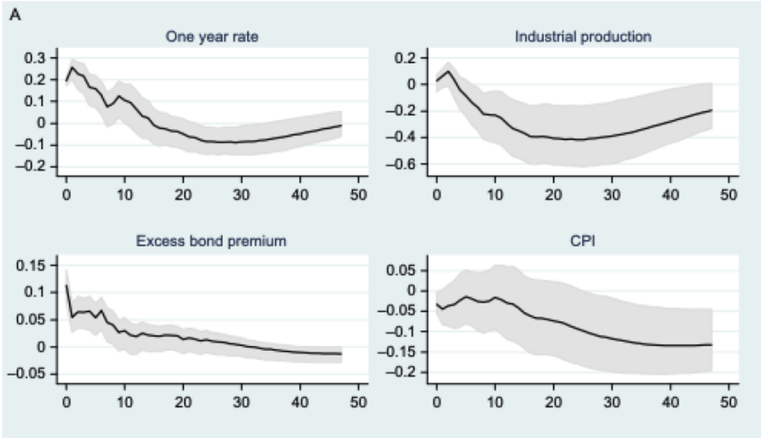


Figure 8.4: Gertler & Karadi (2015): SVAR with Instrumental Variables

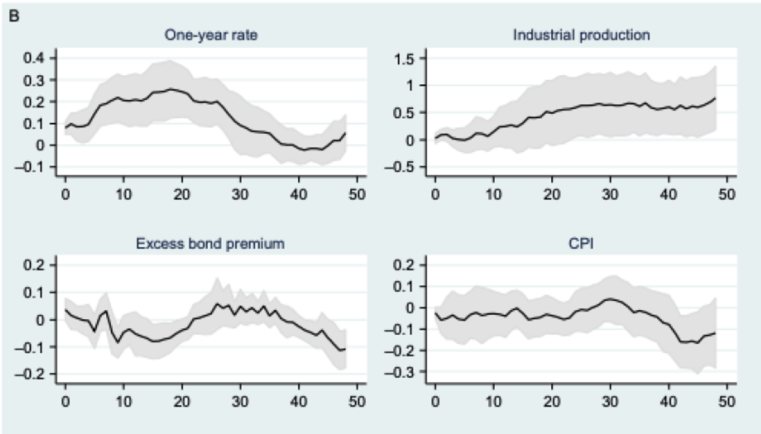


Figure 8.5: Gertler & Karadi (2015): Local Projections (LP) with Instrumental Variables

8.5.2 Evaluating Macroeconometric Models

When estimating causal effects in macroeconomic settings, there is no "truth":

- There is no single way to estimate effects even given the same dataset
- Results are sensitive to restrictions/assumptions imposed
- Also, typically, assumptions cannot be tested, because we cannot estimate the model at all without assumptions

Therefore, it is important to think about robustness: are there findings that are relatively insensitive to alternative assumptions? (i.e. looking for common findings)

8.5.3 Effectiveness of Monetary Policy

Conventional Monetary Policy

Paper	Method, sample	Trough effect of 100 basis point funds peak	% of output explained by shock	Price puzzle?
Christiano et al. (1999)–FFR identification	SVAR, 1965q3–1995q3	−0.7% at 8 quarters	4.4% at 2 years	Yes, but very small
Faust et al. (2004)	HFI, 1991m2–2001m7	−0.6% at 10 months		
Romer and Romer (2004)	Narrative/Greenbook 1970m1–1996m12	−4.3% at 24 months	Major part	No, but prices do not change for 22 months
Uhlig (2005)	Sign restrictions, 1965m1–1996m12	Positive, but not statistically different from 0	5–10% at all horizons.	No (by construction)
Bernanke et al. (2005)	FAVAR, 1959m1–2001m7	−0.6% at 18 months	5% at 5 years	Yes
Smets–Wouters (2007)	Estimated DSGE model, 1966Q1–2004Q4	−1.8 at 4 quarter trough	10% at 1 year (trough)	No
Boivin et al. (2010)	FAVAR, 1962m1–1979m9, 1984m1–2008m12	−1.6% at 8 months in early period −0.7% at 24 months in later period		Only in the early period
Coibion (2012)	"Robust" Romer–Romer methods, 1970m1–1996m12	−2% at 18 months	"Medium" part	Yes, sometimes
Barakchian–Crowe (2013)	HFI, Romer hybrid VAR, 1988m12–2008m6	−5% at 23 months	50% at 3 years	Yes
Gertler–Karadi (2015)	HFI-proxy SVAR, 1979m7–2012m6 (1991m1–2012m6 for instruments)	−2.2% at 18 months	?	No
Amir Ahmadi and Uhlig (2015)	Bayesian FAVAR with sign restrictions, 1960m2–2010m6	−1.3% at 9 months	7% at 24 months	No (by construction)

Figure 8.6: Summary of Leading Results (Ramey (2016))

The answers are surprisingly tentative. In many empirical macroeconomics literatures (half of the examples shown above), we saw the price puzzle: positive interest rate shocks induce higher inflation. Also, response of real activity remains ambiguous with sign restrictions.

Unconventional Monetary Policy

The literature on unconventional policy is still very limited because:

- For the most part, unconventional policies began in 2008
- Most unconventional policies are short-lived and time-varying, so the sample size is very small
- Sampling period contains a massive recession period, which might not be representative

Early evidence shows strong effects of some form of forward guidance and asset purchases *on financial markets*. On the other hand, there is less evidence on their effects on most important *macroeconomic variables*: existing evidence suggests that quantitative easing was effective, but forward guidance is less so.

"QE works in practice, but not in theory."

Interpretability of Estimates

We made some conceptual leaps to recover causal estimates:

- Structural Shocks: central banks don't set interest rates at random and most of their policies are anticipated by the market. Does central banks have to do something surprising to have an effect, or anticipated changes can also be effective?
- Unrealistic assumptions made

Chapter 9

Measuring Teacher Quality

9.1 Empirical Facts: Increasing Educational Spending and Stagnant Outcome

9.1.1 Increasing Educational Spending

Educational Spending in U.S.

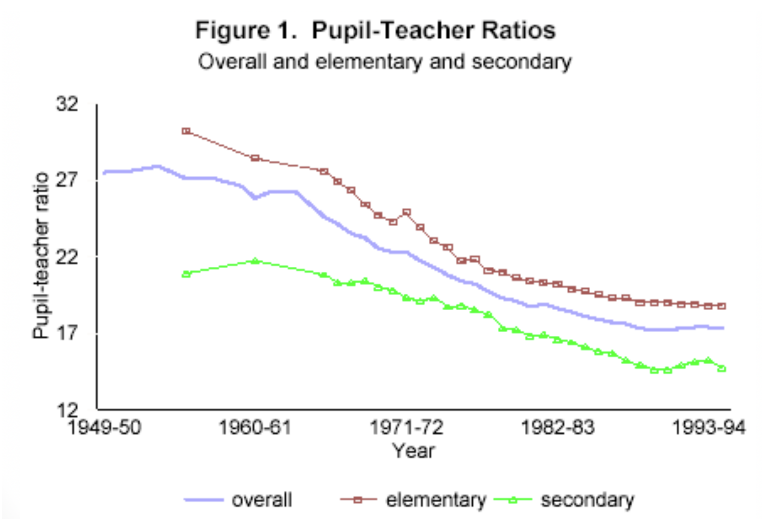


Figure 9.1: U.S. Pupil-Teacher Ratios (Hanushek et al., 2016)

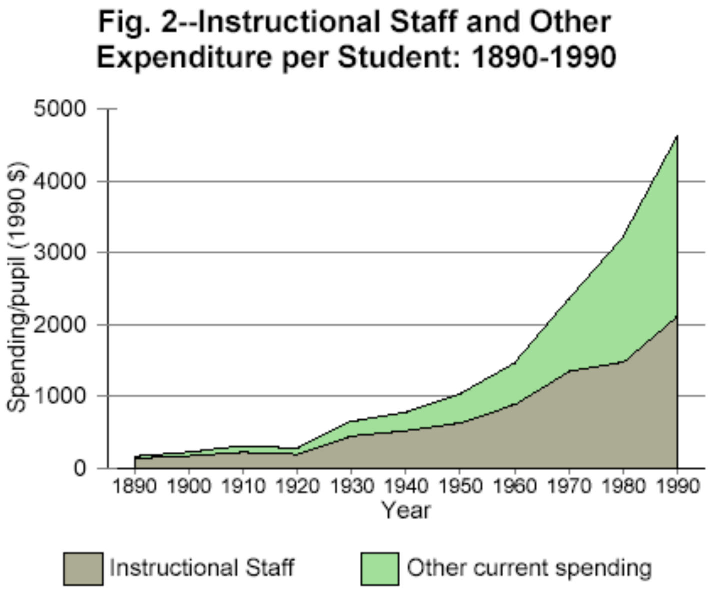


Figure 9.2: U.S. Instructional Staff and Other Expenditure per Student: 1890-1990 (Hanushek et al., 2016)

In U.S., more resources have been allocated to education: the pupil-teacher ratios declined, and the expenditure per student increased.

Educational Spending in the UK

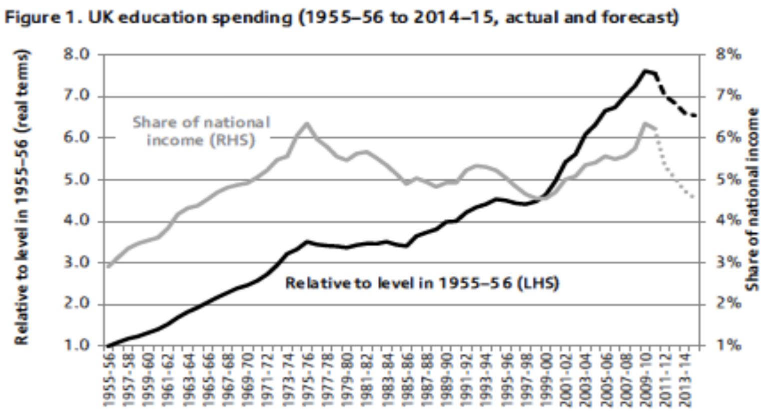


Figure 9.3: UK Education Spending (1955/6-2014/5)

The absolute level of education spending in the UK also kept increasing, though its share of national income levelled off.

9.1.2 Stagnant Outcome

School Results in U.S.

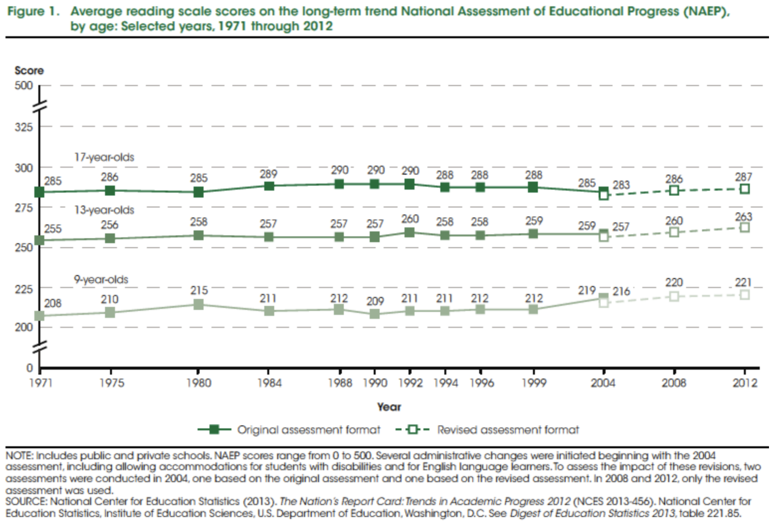


Figure 9.4: Average Reading Scale Scores on NAEP by age 1973-2012

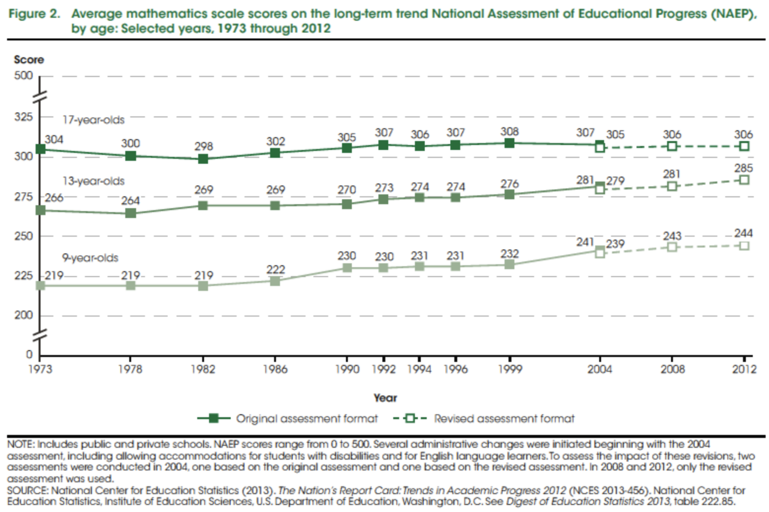


Figure 9.5: Average Mathematics Scale Scores on NAEP by age 1973-2012

However, these was no obvious improvement in education outcomes in U.S.

School Results in the UK

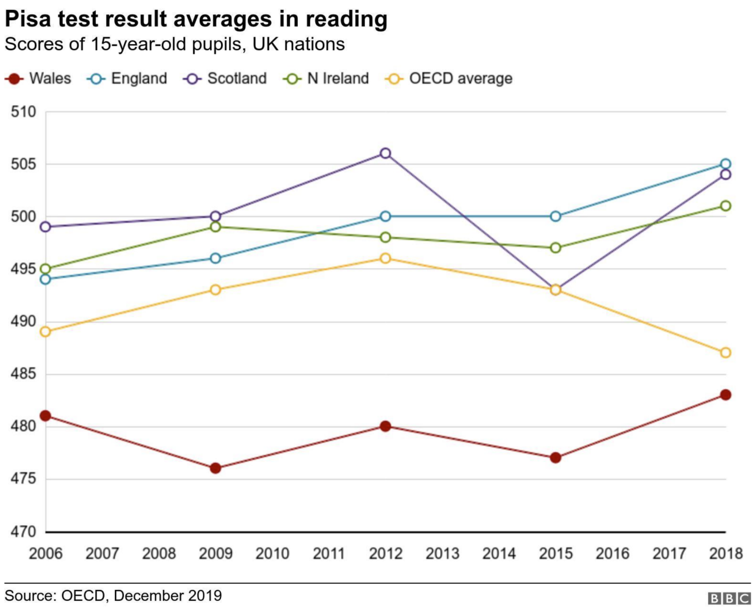


Figure 9.6: PISA Test Result Averages in Reading

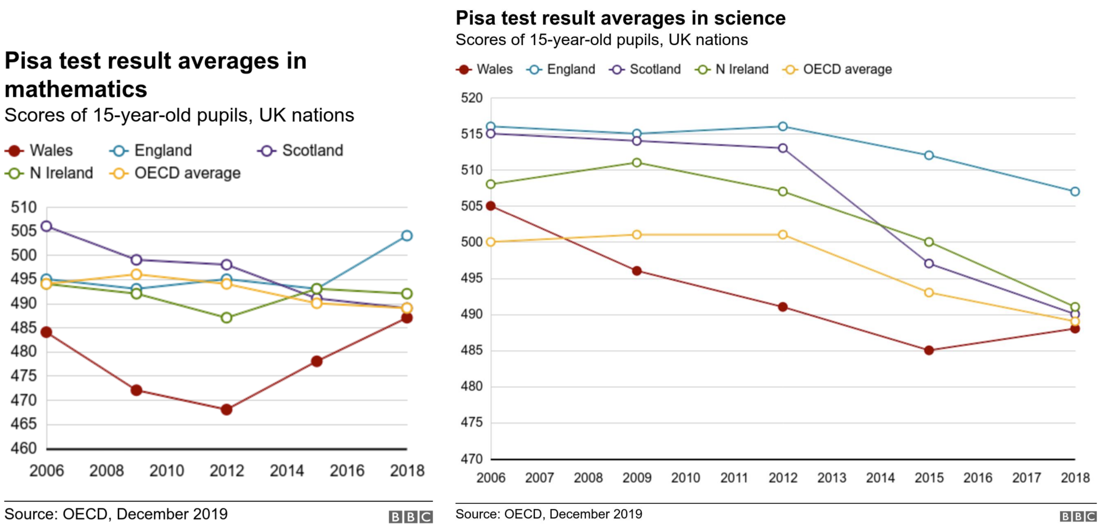


Figure 9.7: PISA Test Result Averages in Math and Science

Similarly, the education outcome in the UK is also stagnant.

9.2 ⋆ Measuring Teacher Quality: Standard Value-Added Model

9.2.1 Motivation

As seen in the previous section, standard measures of school resources do not seem to predict learning, at least when we look at trends. When we examine more micro studies on specific aspects of school quality, such as class size we seem to find some but small impacts of resources on learning.

Meanwhile, the most important school input is likely to be the teacher. Thus, in the following sections, we will develop measures of teacher quality and examine whether estimated teacher quality predicts student learning.

9.2.2 Standard Value-Added Model

Introduction

There are observable measures of teacher quality, such as experience, education, IQ, grades, etc., but they do not predict student learning. Thus, at least at a practical level, important teacher attributes seem to be essentially unobserved.

Therefore, we define teacher quality as the amount of learning observed in students taught by a particular teacher.

The basic idea is to measure the performance of students in two tests across a time interval, take average of all students of a given teacher, and call that teacher quality. This is similar to the league tables in the UK.

Since this measure does not have an exact scale, we calculate the variance / standard deviation and use that as a scale.

Setup

$$A_{ijt} = \theta A_{ij,t-1} + \gamma X_{ijt} + \alpha_{jt} + \epsilon_{ijt} \quad \text{(Standard Value-Added Model)}$$

where:

- i: student; j: teacher; t: grade
- θ: scaling factor (to make tests comparable)
- A_{ijt}: student achievement
- X_{ijt}: student characteristics
- α_{jt}: teacher j's fixed effect in period t (teacher quality)

α measures the teacher quality: it is the average learning of students of teacher j.

This could be written in another way:

$$\underbrace{A_{ijt} - \theta A_{ij,t-1}}_{\text{Value Added}} = \gamma X_{ijt} + \alpha_{jt} + \epsilon_{ijt}$$

9.2.3 Interpretation and Scale

Interpretation and Distinguish Classroom FE

α is really a fixed effect of a regression – it measures the average performance of students taught by a particular teacher, and we call it teacher quality. If only observe one group per teacher, α combines both classroom/peer and teacher FE.

Therefore, we will need to observe at least 2 groups of students taught by the same teacher in order to distinguish classroom/peer fixed effects and teacher fixed effects. With such data, we can run a FE regression with both classroom and teacher FE:

$$\underbrace{A_{ijt} - \theta A_{ij,t-1}}_{\text{Value Added}} = \gamma X_{ijt} + \alpha_{jt} + class_{ijt} + \epsilon_{ijt}$$

After estimating this model, we should also estimate the correct standard error of α using empirical-based error correction or using the covariance between the VA of the same teacher but different groups/years of students. Finally, we standardise VA with the correct standard error and get the result.

Scale

The magnitude of α does not have a direct interpretation. Thus, we compute the standard deviation of α, noted as σ_α, and use it as the unit for α.

For instance, $\alpha = 1$ means that an 1 σ_α increase in teacher quality will induce an 1 σ_A increase in students' value added. Put it differently, α is the change in student achievement measured in σ_A corresponding to an 1 σ_α increase in the teacher quality.

9.2.4 Example

Table 1 The distribution of teacher effectiveness (standard deviations of student achievement)

Study	Location	Teacher effectiveness (σ_τ) Reading	Math
Rockoff (2004)	New Jersey	0.10	0.11
Nye et al. (2004)	Tennessee	0.07	0.13
Rivkin et al. (2005)	Texas	0.15	0.11
Aaronson et al. (2007)	Chicago		0.13
Kane et al. (2008)	New York City	0.08	0.11
Jacob & Lefgren (2008)	Midwest city?	0.12	0.26
Kane & Staiger (2008)	Los Angeles	0.18	0.22
Koedel & Betts (2011)	San Diego		0.23
Rothstein (2009)	North Carolina	0.11	
Hanushek & Rivkin (2010a)	Texas city?		0.11
Average		0.13	0.17

All estimates indicate the standard deviation of teacher effectiveness in terms of student achievement standardized to mean zero and variance one. All are corrected for test measurement error. All except Kane & Staiger (2008) use within-school estimators. Table taken from Hanushek & Rivkin (2010b).

Figure 9.8: Distribution of Teacher Effectiveness

9.3 ⋆ Non-Random Sorting / Selection

9.3.1 Student Sorting / Selection

Our standard value-added model introduced above depends on the critical assumption of conditional random assignment: conditional on A and X, the assignment of students to teachers is random:

$$(A_{0,ijt}, A_{1,ijt}) \perp \mathbb{1}[j = J] | A_{ij,t-1}, X_{ijt}$$

This requires us to include all characteristics that jointly affect selection and outcome. In other words, we cannot have selection on unobservables.

9.3.2 Rothstein 2010: Evidence on Non-random Sorting

Rothstein, 2010 questioned the conditional random assignment assumption. He checked whether future teacher assignments predict current student learning, which is similar to a placebo test.

The intuition is simple: if the 5th grade teacher predicts 4th grade learning, this could only be caused by non-random sorting of students to teachers, because teachers cannot influence students retrospectively.

Specifically, instead of estimating the Standard Value-Added Model, he estimates the following model:

$$A_{ikt-1} = \theta A_{ik,t-2} + \gamma X_{ikt-1} + \alpha_{jt} + \epsilon_{ikt-1}$$

He rejects that α_{jt}'s are jointly equal to zero, and the standard error of α_{jt} is about 0.08 instead of around 0.1 estimated with standard models. This suggests that there is non-random sorting.

9.3.3 Chetty, Friedman, Rockoff 2014: Bias is Small

Setup

Chetty et al., 2014a confirms that Rothstein, 2010 is correct, but the resulting bias is small.

Chetty et al. look at quasi-experiments – changes in teachers across schools.

Specifically:

- 2 schools: school A and school B
- 2 teachers: teacher a works in school A, and teacher b works in school B
- α_{aA} is the estimated value added for teacher a in school A
- α_{bB} is the estimated value added for teacher b in school B
- Teacher a moves to school B

Then, after the movement of teacher a, the change in students' learning should be the same as the change in mean teacher value-added ($\alpha_a - \alpha_b$), assuming teacher movements are random. Otherwise, there is likely to be bias in value added estimates due to sorting.

Results

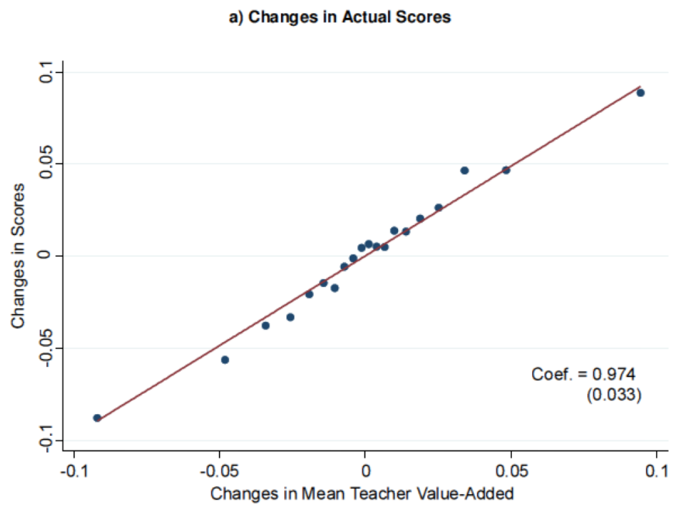


Figure 9.9: Change in Actual Scores and Changes in Mean Teacher Value-Added (Chetty et al., 2014b)

The movement is nearly one-to-one, so bias due to non-random sorting is very small.

But Still Not Settled

There are follow-up discussions from Rothstein and Chetty, Friedman and Rockoff regarding whether changes in teachers is correlated with prior achievement gains of students (or can we say that they are essentially random)?

9.3.4 Kane et al. 2013: Additional Evidence from Random Assignment

Kane et al., 2013 randomly assigned students to teachers in 2011, ensuring there is no systematic sorting of students to teachers. Then, they check how will estimate of value-added in 2009-10 predicts student learning in 2011. The result show that the prediction is well good, which indicates there's no sorting.

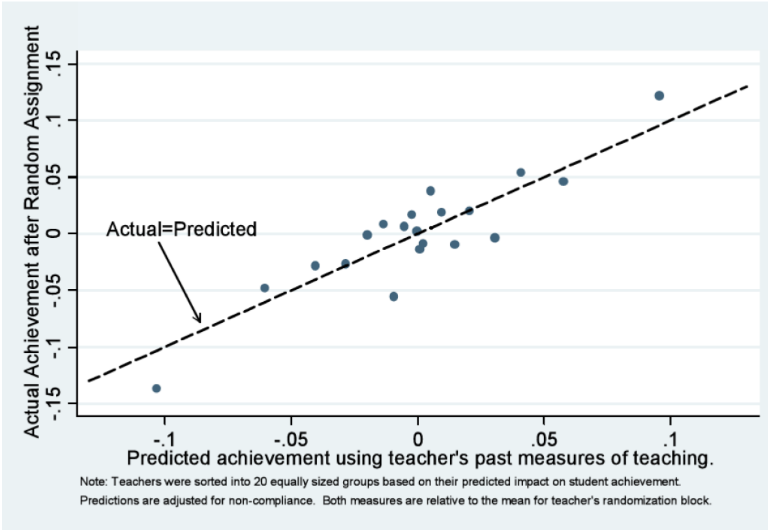


Figure 9.10: Actual and Predicted Achievement in Maths (Kane et al., 2013)

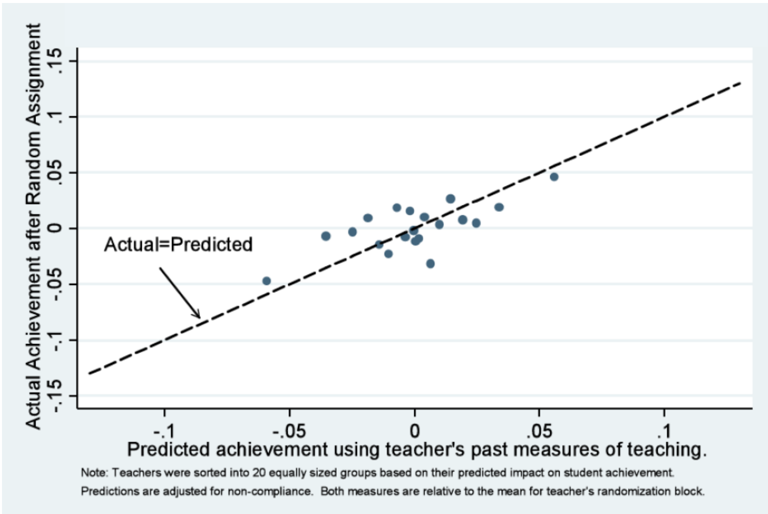


Figure 9.11: Actual and Predicted Achievement in Arts (Kane et al., 2013)

9.4 ⋆ Measurement Issues

9.4.1 Measurement Error

In the Standard Value-Added Model, suppose $\alpha_{jt} = \alpha_j$ (time-unvarying), we cannot observe α_j directly. Instead, we need to estimate it. This comes with a measurement error v_j due to sampling uncertainty (with data from more years, this could be smaller):

$$\widehat{\alpha_j} = \alpha_j + v_j$$

Therefore, assume the measurement error v_j is independent of α_j:

$$Var(\widehat{\alpha_j}) = Var(\alpha_j) + Var(v_j)$$

Hence, $Var(\widehat{\alpha_j})$ overestimates $Var(\alpha_j)$. If the variance of measurement error is large, it will be hard/unreliable to rank teachers based on $\widehat{\alpha_j}$.

9.4.2 Dealing with Measurement Error: Empirical Based Error Correction

We can use the law of total variance to get an estimate of $Var(v_j)$ and use it to retrieve $Var(\alpha_j)$:

$$Var(\widehat{\alpha_j}) = \underbrace{Var(E[\widehat{\alpha_j}|j])}_{\approx Var(\alpha_j)} + \underbrace{E[Var(\widehat{\alpha_j}|j)]}_{\approx Var(v_j)}$$

$$Var(\alpha_j) \approx Var(\widehat{\alpha_j}) - E[Var(\widehat{\alpha_j})|j]$$

Here, $E[Var(\widehat{\alpha_j})|j]$ is just the variance of the FE, which could be obtained from the regression. Using it to proxy $Var(v_j)$, we can retrieve $Var(\alpha_j)$.

However, this method cannot filter out variance from classroom FE shocks. With richer data, we may consider the next method.

9.4.3 Dealing with Measurement Error and Distinguish Teacher / Classroom Effects

If we have data of multiple years, this alternative method can deal with measurement error while distinguish teacher and classroom effects at the same time.

Firstly, we estimate the Standard Value-Added Model for multiple years ($t \neq s$) and get:

- $\widehat{\alpha_{jt}} = \alpha_j + w_{jt} + v_{jt}$
- $\widehat{\alpha_{js}} = \alpha_j + w_{jt} + v_{js}$

where:

- α is teacher's value added (teacher effect)
- w is classroom effect
- v is sampling error

Assume that:

- Classroom shocks are serially uncorrelated and independent of α_j: $Cov(w_{jt}, w_{js}) = 0, (w_{jt}, w_{js}) \perp \alpha_j$
- Sampling errors are serially uncorrelated and independent of α_j: $Cov(v_{jt}, v_{js}) = 0, (v_{jt}, v_{js}) \perp \alpha_j$

Then, we can calculate $Var(\alpha_j)$ by:

$$Cov(\widehat{\alpha_{jt}}, \widehat{\alpha_{js}}) = Cov(\alpha_j, \alpha_j) + \underbrace{Cov(\alpha_j, w_{jt} + v_{jt}) + Cov(\alpha_j, w_{js} + v_{js}) + Cov(w_{jt} + v_{jt}, w_{js} + v_{js})}_{0} = Var(\alpha_j)$$

The previous method can only eliminate $Var(v_{jt})$ and retrieve $Var(\alpha_j + w_{jt})$, which is a mixture of teacher/classroom FE, but this alternative method can also eliminate the variance of classroom FE.

9.4.4 Fluctuations in Teacher Quality

Chetty et al., 2014b extended the model to allow for α_j to vary over time. Specifically, they use $\{\alpha_{jt-T}, \ldots, \alpha_{jt-1}\}$ to predict $\alpha_j t$.

9.5 Predictors of Teacher Quality

9.5.1 Rockoff et al. 2008: Observed Characteristics?

Summary

Question: can you recognise an effective teacher whe you recruit one?

Rockoff et al., 2011 collect several teacher characteristics and correlate the, with α_j. The characteristics they collect include:

◇ Traditional: education, training, college major, SAT scores, ranking of college attended

◇ Non-Traditional: cognitive ability, math knowledge, personal traits, Haberman pre-screener

The outcome variables are: student test scores in math, teacher absences, subjective evaluations of teachers, whether teacher returns to DOE in the following year, whether teacher returns to school in the following year.

Results: most characteristics are not significantly correlated with outcomes at all. Even for those with significant correlations, the magnitudes are tiny.

Specific Results

Table 3: Traditional Predictors of Teacher and Student Outcomes

	Math Achievement	Subjective Evaluation	Teacher Absences	Returned to NYC	Returned to School | NYC
Credentials					
Has a Graduate Degree	-0.014	0.133	0.019	-0.050	-0.016
	(0.024)	(0.138)	(0.412)	(0.033)	(0.035)
	[0.557]	[0.338]	[0.962]	[0.130]	[0.649]
Passed LAST Certification Exam on 1st Attempt *(1=yes)*	0.035	0.123	0.013	-0.053	0.001
	(0.039)	(0.196)	(0.688)	(0.040)	(0.059)
	[0.369]	[0.529]	[0.984]	[0.185]	[0.982]
Teaching Fellow	-0.046	-0.184	1.006	0.118	-0.016
(Relative to Traditionally Certified)	(0.027)*	(0.138)	(0.514)*	(0.031)**	(0.040)
	[0.085]	[0.185]	[0.050]	[0.000]	[0.695]
TFA Corps Member	0.044	-0.052	-0.501	0.128	0.090
(Relative to Traditionally Certified)	(0.030)	(0.140)	(0.422)	(0.035)**	(0.035)**
	[0.151]	[0.710]	[0.235]	[0.000]	[0.011]
Math or Science Major	0.040	-0.048	-1.212	-0.063	-0.007
(Relative to Those Other Than Math, Science, or Education)	(0.031)	(0.183)	(0.529)**	(0.049)	(0.049)
	[0.2]	[0.795]	[0.022]	[0.201]	[0.879]
Education Major	-0.009	-0.117	-0.485	-0.097	0.041
(Relative to Those Other Than Math, Science, or Education)	(0.033)	(0.144)	(0.489)	(0.042)**	(0.038)
	[0.789]	[0.417]	[0.321]	[0.022]	[0.279]
Self-Reported SAT Math Score *(s.d.=1)*	0.012	0.004	-0.119	0.008	0.005
	(0.015)	(0.075)	(0.207)	(0.020)	(0.015)
	[0.41]	[0.960]	[0.564]	[0.686]	[0.715]
Self-Reported SAT Verbal Score *(s.d.=1)*	-0.003	0.035	0.145	0.026	0.004
	(0.014)	(0.081)	(0.228)	(0.020)	(0.022)
	[0.829]	[0.666]	[0.524]	[0.188]	[0.853]
Barrons Rank of College *(s.d.=1)*	0.022	-0.212	0.059	0.027	-0.014
	(0.012)*	(0.087)**	(0.217)	(0.018)	(0.022)
	[0.076]	[0.015]	[0.786]	[0.118]	[0.015]
Control for Student/School Characteristics and Zip Code FE	√	√	√	√	√
Observations	247,903	3,030	4,858	4,877	4,516

Figure 9.12: Traditional Predictors of Teacher and Student Outcomes

Table 4: Non-Traditional Predictors of Teacher and Student Outcomes

	Math Achievement	Subjective Evaluation	Teacher Absences	Returned to NYC	to School | NYC
Cognitive Ability *(Percentile, s.d.=1)*	0.016	0.066	-0.422	0.016	0.021
	(0.012)	(0.058)	(0.227)*	(0.016)	(0.019)
	[0.174]	[0.254]	[0.063]	[0.315]	[0.270]
Math Knowledge for Teaching *(Percent Correct, s.d.=1)*	0.028	0.014	-0.407	0.006	-0.011
	(0.012)**	(0.065)	(0.208)*	(0.014)	(0.017)
	[0.024]	[0.828]	[0.051]	[0.659]	[0.504]
Conscientiousness *(s.d.=1)*	0.011	0.188	0.185	-0.000	0.010
	(0.011)	(0.059)**	(0.169)	(0.013)	(0.020)
	[0.319]	[0.001]	[0.273]	[0.982]	[0.624]
Extraversion *(s.d.=1)*	0.007	0.216	0.086	0.000	0.022
	(0.011)	(0.062)**	(0.189)	(0.015)	(0.017)
	[0.519]	[0.001]	[0.650]	[0.986]	[0.201]
General Efficacy *(s.d.=1)*	0.017	0.019	-0.028	0.037	0.009
	(0.012)	(0.057)	(0.192)	(0.016)**	(0.016)
	[0.149]	[0.736]	[0.885]	[0.024]	[0.591]
Personal Efficacy *(s.d.=1)*	0.012	0.192	0.148	0.015	0.014
	(0.011)	(0.060)**	(0.204)	(0.013)	(0.015)
	[0.271]	[0.001]	[0.470]	[0.280]	[0.372]
Control for Student/School Characteristics and Zip Code FE	√	√	√	√	√
Observations	247,903	3,030	4,858	4,877	4,516

Figure 9.13: Non-Traditional Predictors of Teacher and Student Outcomes

Table 6: Haberman PreScreener Performance and Teacher and Student Outcomes

	Math Achievement	Subjective Evaluations	Teacher Absences	Returned to NYC	Returned to School | NYC
Haberman Top Group	0.033	0.243	0.928	0.009	-0.064
	(0.031)	(0.175)	(0.564)	(0.035)	(0.050)
	[0.297]	[0.167]	[0.100]	[0.793]	[0.206]
Haberman Total Score *(s.d.=1)*	0.021	0.141	0.135	0.027	-0.040
	(0.013)	(0.065)**	(0.230)	(0.018)	(0.020)**
	[0.110]	[0.029]	[0.556]	[0.125]	[0.043]
Controls for School Charateristics and School Zip Code	√	√	√	√	√
Observations	244,235	2,970	4,754	4,773	4,421

Figure 9.14: Haberman PreScreener Performance and Teacher and Student Outcomes

Table 8: Using Factors as Predictors of Teacher and Student Outcomes

	Math Achievement	Subjective Evaluation	Teacher Absences	Returned to NYC	Returned to School | NYC
Factor 1: Cognitive Skills (s.d.=1)	0.033	0.025	-0.227	0.043	0.005
	(0.011)**	(0.065)	(0.195)	(0.016)**	(0.015)
Factor 2: Non-Cognitive Skills (s.d.=1)	0.033	0.272	-0.026	0.009	0.031
	(0.015)**	(0.068)**	(0.243)	(0.017)	(0.020)
F-Test: All Factors Equal Zero (p-value)	0.0023	0.00	0.51	0.03	0.30
Observations	247,903	3,030	4,858	4,877	4,516
Control for Student/School Characteristics and Zip Code FE	√	√	√	√	√

Figure 9.15: Using Factors as Predictors of Teacher and Student Outcomes

Figure 1: Recruitment Information and the Distribution of Predicted Value-Added

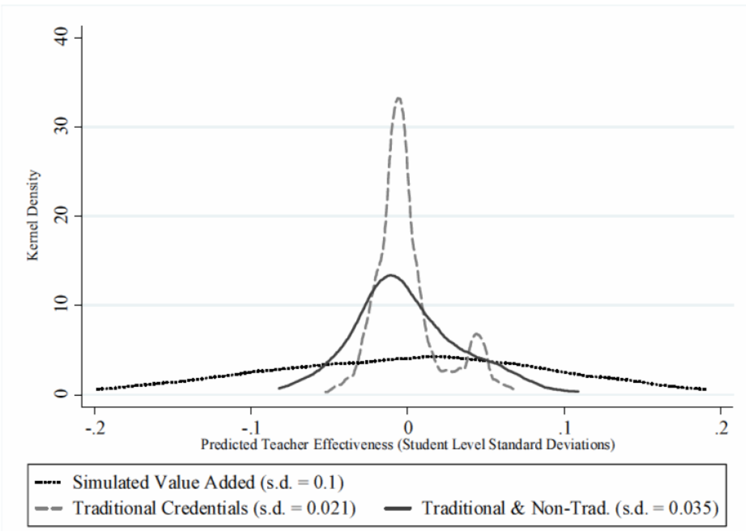


Figure 9.16: Recruitment Information and the Distribution of Predicted Value-Added

9.5.2 Jacob and Lefgren 2005: Principals' Perceptions?

Jacob and Lefgren, 2005 alayse whether pricipals can identify good teachers.

They conclude that subjective principal assessments predict teacher value-added much better than teacher experience, education, and earnings. Principals are good at identifying the best and the worst teachers, but have trouble distinguishing between teachers in the middle of the distribution.

9.6 Long-Term Impacts of Teachers

9.6.1 Low Persistence of Teacher Effects and "Teaching to the Test"

Jacob et al., 2008 conclude that teacher value-added in reading and math *quickly erodes*. One year persistence is only around 20%.

Carrell and West, 2008 explore the random assignment of students to teacher at USAF academy. They find that some teachers get high value-added by "teaching to the test," in order to get better teaching evaluations. The problem is that, by "teaching to the test," they do not teach deep concepts, resulting in low future value-added. Hence, teachers with low current value-added could have high future value-added.

9.6.2 2 Studies on Long-Term Impacts of Teachers

Chetty et al., 2011

Chetty. Friedman, Hilger, Saez, Schanzenbach, Yagan 2011 investigated the STAR Experiment in which 11571 students in Tennessee were randomly assigned to classrooms within their schools from K-3. Originally, STAR was designed as a class size experiment, but it could be used more generally as a classroom experiment that randomly assign students into good and bad classes (measured by end of class test scores of students in that class).

Then, Chetty et al. linked STAR records with tax records.

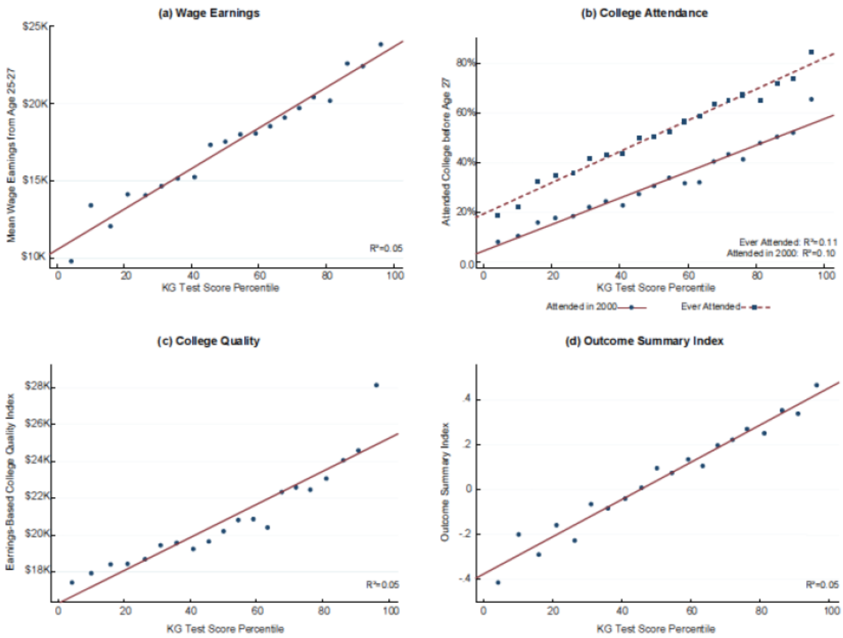


Figure 9.17: Scores and Adult Outcomes

Individual test scores are correlated with adult outcomes.

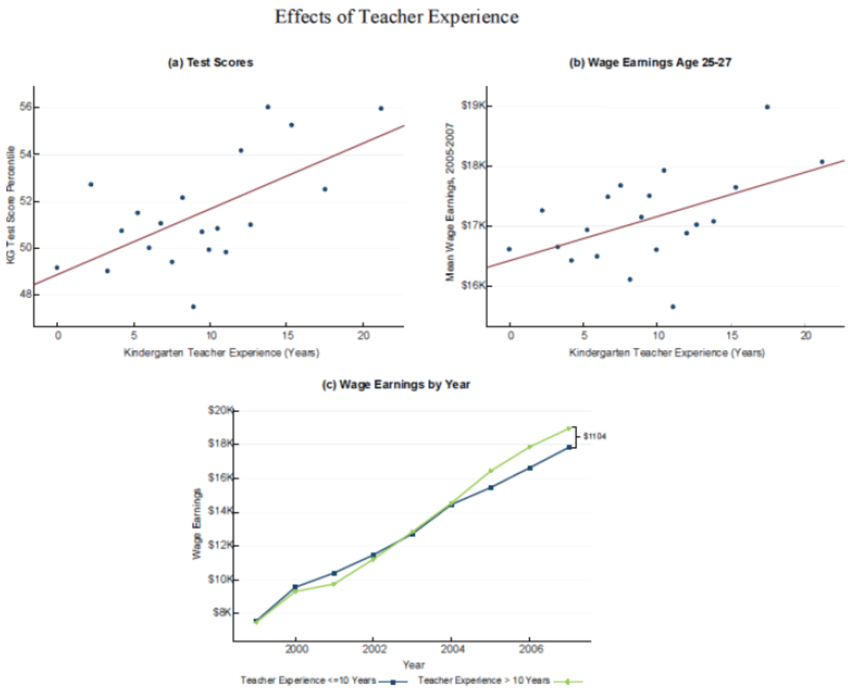


Figure 9.18: Effects of Teacher Experience

Teachers' experience is also correlated with adult outcomes.

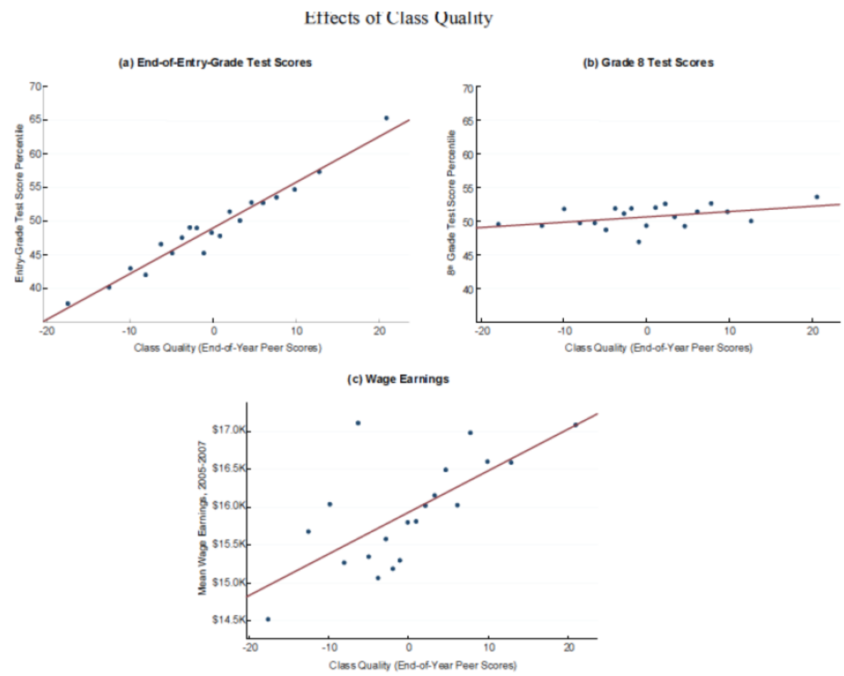


Figure 9.19: Effects of Class Quality

Class quality is also correlated with adult outcomes.

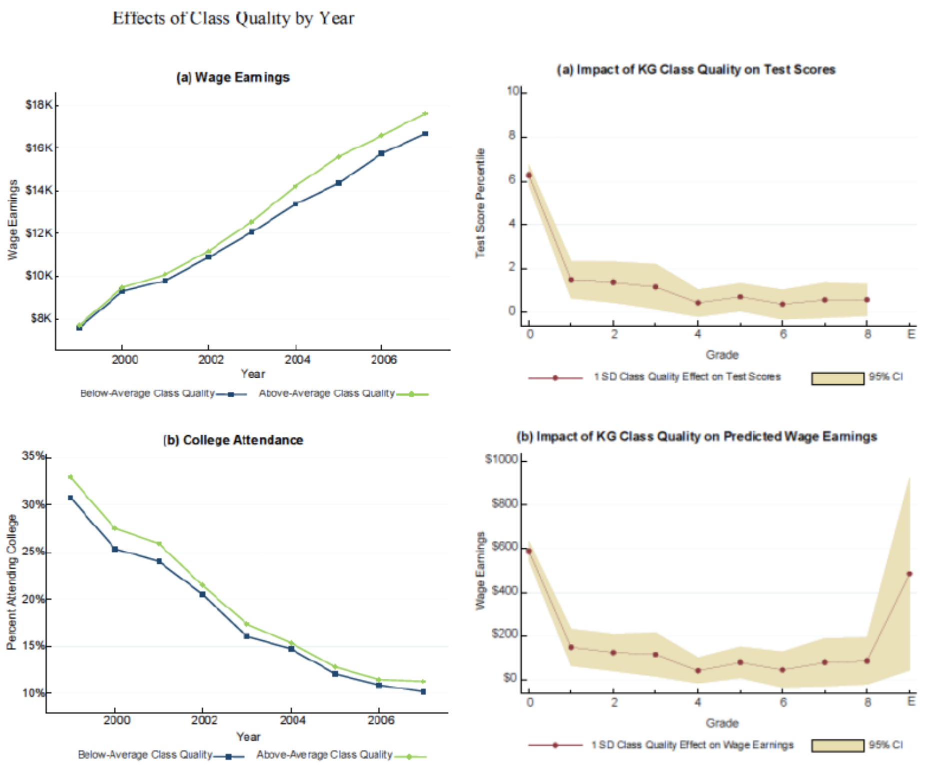


Figure 9.20: Effect of Class Quality by Year

Effect of class quality on test scores diminish as time goes by, but effects on earnings suddenly appear after some years.

Chetty et al., 2014b

Chetty, Friedman, and Rockoff 2014 estimate teacher VA from administrative school records. Specifically, they link VA of teachers assigned to each student to students' tax records. It turns out that distinguishing by grade is not very important.

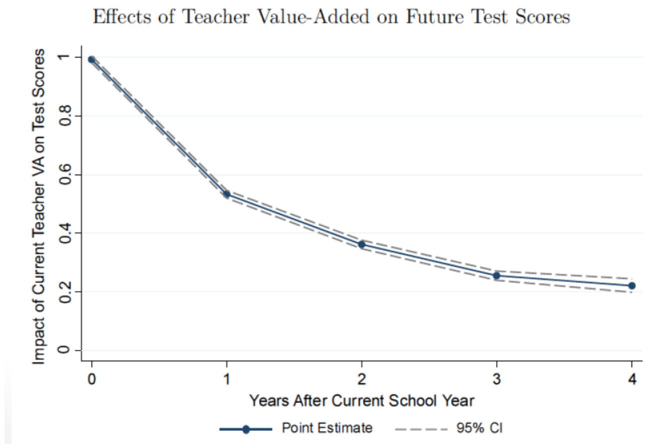


Figure 9.21: Teacher VA and Test Scores

The effect of teacher VA on test score quickly diminishes as age increases.

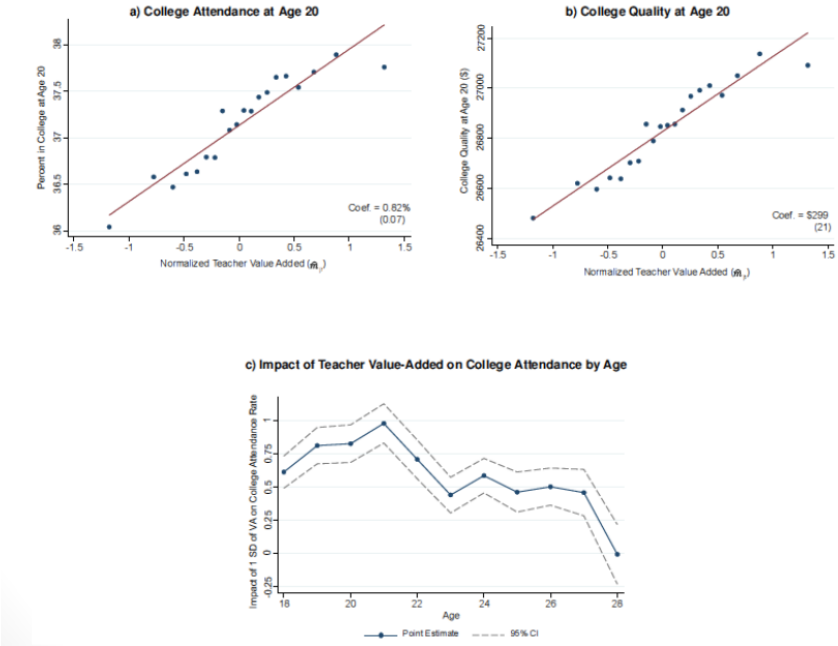


Figure 9.22: Teacher VA and College Outcomes by Age

The effect of teacher VA on college attendance also diminishes.

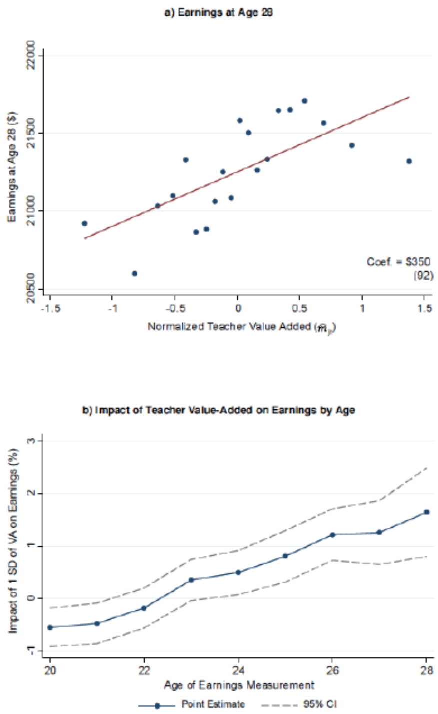


Figure 9.23: Teacher VA and Earnings by Age

However, its effects on earnings increase by time.

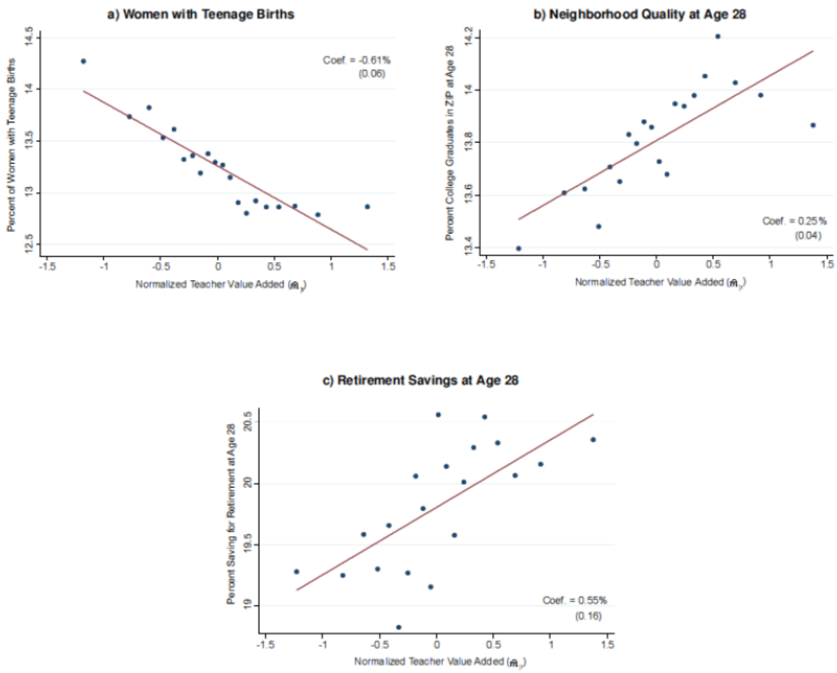


Figure 9.24: Teacher VA and Other Adult Outcomes

Teacher VA is also related to many other adult outcomes.

Conclusions

A lot of literature, including the above two, found that effects of teachers on test scores diminish quickly, but their effects on adult outcomes, such as wages, appear and increase after certain time periods.

9.7 In Developing Countries

Araujo et al., 2016 (Araujo, Carneiro, Cruz, Schady 2016) did a similar study in Ecuador.

Specifically, they randomly assign kindergarden students to classrooms in schools in Ecuador. They have data from multiple years, so it is possible to distinguish teacher and classroom effects.

They test students on multiple dimensions, including Maths, language, executive function, etc. Also, they measure multiple attributes of teachers, including IQ, personality, education, experience, and classroom observations (quality of interaction between teacher and students).

Table 2: Within-school differences in learning outcomes

	Classroom effects								Teacher effects
	2012 cohort (12 tests)				2012 cohort (4 tests)		2013 cohort (4 tests)		Covariance, 2012 and 2013 cohorts (4 tests)
Sample Restriction	Whole sample		Classes with the same teacher throughout the year		Classes for which teachers are the same in both cohorts of children				
	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected	
Language	0.14	0.11	0.15	0.11	0.15	0.10	0.15	0.10	0.09
Math	0.16	0.11	0.16	0.11	0.16	0.11	0.16	0.11	0.09
Executive function	0.14	0.07	0.14	0.07	--	--	--	--	--
TOTAL	0.15	0.11	0.15	0.11	0.16	0.12	0.15	0.10	0.10
Students	13,565	13,565	9,962	9,962	5,904	5,904	6,023	6,023	11,927
Teachers	451	451	334	334	196	196	196	196	196
Schools	204	204	150	150	87	87	87	87	87

Note: In the specifications that include all 12 tests, controls include all baseline child and household characteristics; in the specifications that include only 4 tests, controls include only child age, gender, and the baseline TVIP score.

Figure 9.25: With-in School Differences in Learning Outcomes

Table 5: Teacher characteristics and behaviors, and child learning outcomes

	2012/2013 Cohort					2013/2014 Cohort
	Language	Math	Executive Function	All	Common tests	Common tests
Experience	-0.130	-0.153	-0.115*	-0.159*	-0.157	-
	(0.086)	(0.089)	(0.058)	(0.077)	(0.089)	
Tenured	0.051	0.101*	0.013	0.062	0.086	0.075
	(0.041)	(0.043)	(0.041)	(0.044)	(0.044)	(0.039)
IQ	0.043*	0.038*	0.023	0.041*	0.032	0.016
	(0.021)	(0.018)	(0.016)	(0.018)	(0.018)	(0.018)
Neuroticism	-0.003	-0.003	0.011	0.001	-0.006	0.007
	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.021)
Extraversion	0.026	0.031	0.024	0.032	0.032	0.003
	(0.019)	(0.019)	(0.013)	(0.016)	(0.017)	(0.015)
Openness	0.007	0.018	0.008	0.013	0.022	0.009
	(0.021)	(0.021)	(0.017)	(0.020)	(0.021)	(0.018)
Agreeableness	-0.001	0.000	-0.016	-0.006	-0.009	0.002
	(0.022)	(0.022)	(0.017)	(0.021)	(0.020)	(0.021)
Conscientiousness	-0.025	-0.035	-0.023	-0.032	-0.033	-0.030
	(0.018)	(0.019)	(0.016)	(0.018)	(0.019)	(0.020)
Stroop	0.024	0.034	0.032*	0.035	0.024	0.008
	(0.029)	(0.017)	(0.014)	(0.020)	(0.018)	(0.022)
Parents' Education	0.006	0.007	0.003	0.006	0.004	-0.004
(Average years)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Teacher Attendance	0.028	-0.042	-0.059	-0.027	-0.013	-0.089*
(Days of absence ≥ 4)	(0.050)	(0.034)	(0.039)	(0.037)	(0.034)	(0.040)
CLASS	0.182*	0.192*	0.121*	0.194*	0.184*	0.136**
(lagged)	(0.082)	(0.078)	(0.060)	(0.075)	(0.081)	(0.045)

Figure 9.26: Teacher Characteristics & Behaviours and Child Learning Outcomes

Still, there's no large and significant characteristics that can be used to predict VA of a teacher.

Table 9: Can parents recognize better teachers?

	OLS	OLS Dummy	Ordered Probit
Average score (-*i*)	0.201**	0.162**	0.444**
	(0.065)	(0.054)	(0.142)
Experience	-0.158**	-0.135**	-0.33**
	(0.043)	(0.041)	(0.098)
IQ	0.003	0.002	0.006
	(0.002)	(0.001)	(0.003)
CLASS 2011/2012	0.136**	0.303**	0.116**
	(0.044)	(0.092)	(0.035)

Figure 9.27: Teacher VA and Earnings by Age

However, parents are able to distinguish good teachers.

	N	Mean, SD	Average score (-*i*)	Years of experience	Tenured	IQ	CLASS 2011/2012
Read books, watch pictures or drawings in a book with child	8,359	1.24	0.03	-0.07	0.03	0.0002	0.05
		(1.91)	(0.032)	(0.04)	(0.018)	(0.001)	(0.026)
Tell short stories or tales to child	8,362	1.14	-0.01	-0.08	-0.02	-0.0002	0.03
		(1.81)	(0.037)	(0.04)	(0.021)	(0.001)	(0.029)
Sing to child or sing with, even lullabies	8,362	2.63	-0.057	-0.02	-0.01	-0.001	-0.02
		(2.66)	(0.035)	(0.035)	(0.014)	(0.001)	(0.024)
Take child outside. Go to the park or go for a walk	8,368	1.74	-0.004	-0.043	-0.001	0.0003	0.01
		(1.74)	(0.033)	(0.029)	(0.016)	(0.001)	(0.023)
Play with child with his toys	8,358	1.83	0.02	-0.025	0.02	0.0005	0.03
		(2.41)	(0.041)	(0.046)	(0.021)	(0.001)	(0.028)
Drawing or painting with child	8,363	2.62	0	-0.039	-0.014	-0.0011	0.04
		(2.56)	(0.033)	(0.035)	(0.02)	(0.001)	(0.028)
Play with child to name or count objects or colors	8,364	3.23	-0.01	-0.02	0	-0.0009	0.08**
		(2.59)	(0.036)	(0.04)	(0.018)	(0.001)	(0.027)

Figure 9.28: Teacher VA and Earnings by Age

Meanwhile, parents did not seem to respond to teachers' quality.

Bibliography

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten *. *The Quarterly Journal of Economics*, *131*(3), 1415–1453. https://doi.org/10.1093/qje/qjw016

Carrell, S. E., & West, J. E. (2008, June). Does professor quality matter? evidence from random assignment of students to professors. https://doi.org/10.3386/w14081

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly Journal of Economics*, *126*(4), 1593–1660. https://doi.org/10.1093/qje/qjr041

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632. https://doi.org/10.1257/aer.104.9.2593

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Hanushek, E. A., Machin, S. J., & Woessmann, L. (2016, April 27). *Handbook of the economics of education* [Google-Books-ID: 4_MQCgAAQBAJ]. Elsevier.

Jacob, B. A., & Lefgren, L. (2005, July). Principals as agents: Subjective performance measurement in education. https://doi.org/10.3386/w11463

Jacob, B. A., Lefgren, L., & Sims, D. (2008, June). The persistence of teacher-induced learning gains. https://doi.org/10.3386/w14065

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. (2013). Have we identified effective teachers?: Validating measures of effective teaching using random assignment [Publisher: MET project]. *MET Project Research Paper: Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment (MET Project*. Retrieved January 21, 2023, from https://www.rand.org/pubs/external_publications/EP50156.html

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, *6*(1), 43–74. https://doi.org/10.1162/EDFP_a_00022

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement*. *The Quarterly Journal of Economics*, *125*(1), 175–214. https://doi.org/10.1162/qjec.2010.125.1.175

Chapter 10

Performance Pay in Education

10.1 Motivation

School resources may have limited effectiveness because schools do not face market incentives. Thus, maybe we can introduce some market incentives:

- School choice (allowing parents to choose schools)
- Incentives in schools (for teachers/students)

We will focus on the second one in this lecture.

In reality, designing incentives for teachers is complicated due to the multitasking problem introduced below.

10.2 ⋆ Model of Multitasking (Neal, 2011)

We want to elicit effort from teachers. Specifically, we need to address two main issues:

- Multitasking – there are several activities a teacher undertakes
- Unobservable outputs – It is hard to measure performance

10.2.1 Setup

Basics

The education authority hires 1 teacher to teach 1 student.

The teacher allocates time/effort to two different tasks: t_1 and t_2 are time spent on each task.

Human Capital Production

$$h = f_1 t_1 + f_2 t_2 + e$$

where:

- h: human capital measured in £
- t_1 and t_2: teacher's time spent on each task
- f_1, f_2: constants
- e: random shock with mean zero and independent of t_1, t_2
- $h - e$ is the result of teacher efforts, over which the teacher has control

Observable Performance

The education authority cannot observe h, t_1, t_2.

Only a statistical measure of teacher performance p is observed (e.g. a test score):

$$p = g_1 t_1 + g_2 t_2 + v \tag{10.1}$$

where:

- t_1 and t_2: teacher's time spent on each task
- g_1, g_2: constants
- v: random shock with mean zero and independent of e, t_1, t_2

Teacher's Wage and Utility

Wage:

$$w = s + bp$$

where:

- s: base salary
- b: bonus for performance
- p: student's performance (equation 10.1)

Utility:

$$U = \underbrace{E[s + bp]}_{=E[w]=X} - \underbrace{\left[\frac{1}{2}(t_1 - \overline{t_1})^2 + \frac{1}{2}t_2^2\right]}_{C(t_1,t_2)}$$

where:

- $X = E[w] = E[s + bp]$: teacher's expected income
- $C(t_1, t_2) = \frac{1}{2}(t_1 - \overline{t_1})^2 + \frac{1}{2}t_2^2$: teacher's cost of effort
  - $\overline{t_1}$: the minimal amount of effort spent on task 1 (e.g. showing up)
  - $C(t_1, t_2)$ is increasing in both t_1, t_2

We also have an outside option for the teacher with utility U_0.

Education Authority's Target Function

Suppose the education authority is a benevolent social planner who maximises:

$$\begin{aligned} \max_{t_1, t_2} \quad & E(h) - C(t_1, t_2) \\ \text{s.t.} \quad & h = f_1 t_1 + f_2 t_2 + e, \\ & C(t_1, t_2) = \frac{1}{2}(t_1 - \overline{t_1})^2 + \frac{1}{2}t_2^2 \end{aligned} \tag{10.2}$$

10.2.2 Ideal Optimisation (If All Observable)

If all variables are observable, the education authority can solve the optimisation 10.2 directly:

$$\Longrightarrow \max_{t_1,t_2} f_1 t_1 + f_2 t_2 - \left[\frac{1}{2}(t_1 - \overline{t_1})^2 + \frac{1}{2}t_2^2\right]$$

First order conditions:

$$\begin{cases} f_1 - \frac{1}{2}2(t_1 - \overline{t_1}) & = 0 \\ f_2 - \frac{1}{2}2t_2 & = 0 \end{cases}$$

Optimum with perfect information:

$$\begin{cases} t_1^* & = f_1 + \overline{t_1} \\ t_2^* & = f_2 \end{cases}$$

However, this is not feasible in reality because t_1, t_2 are unobservable.

10.2.3 Optimisation If Only Observe Scores

In reality, the educational authority only observe a performance score p (equatin 10.1). Similar to any typical mechanism design problem with moral hazard, we work backwards by solving teacher's optimisation first.

Step 1: Teacher's Optimisation

Given a level of incentive (b), the teacher's optimal choice of t_1, t_2:

$$\begin{aligned}\max_{t_1,t_2} E[U] &= \underbrace{E[s+bp]}_{=E[w]} - \underbrace{\left[\frac{1}{2}(t_1-\overline{t_1})^2 + \frac{1}{2}t_2^2\right]}_{C(t_1,t_2)} \\ &= s + b(g_1t_1 + g_2t_2) - \left[\frac{1}{2}(t_1-\overline{t_1})^2 + \frac{1}{2}t_2^2\right]\end{aligned}$$

First order conditions for teachers:

$$\begin{cases} t_1 &= bg_1 + \overline{t_1} \\ t_2 &= bg_2 \end{cases} \tag{10.3}$$

We can see that $b \uparrow \implies t_1 \uparrow, t_2 \uparrow$.

Step 2: Educational Authority's Optimisation

Given the teacher's response (equation 10.3), the educational authority's optimisation becomes:

$$\begin{aligned} \max_{b} \quad & E(h) - C(t_1,t_2) = f_1t_1 + f_2t_2 - \left[\frac{1}{2}(t_1-\overline{t_1})^2 + \frac{1}{2}t_2^2\right] \\ \text{s.t.} \quad & t_1 = bg_1 + \overline{t_1}, \\ & t_2 = bg_2 \end{aligned} \tag{10.4}$$

Solving the constrained optimisation above, we can get the optimal choice of incentive:

$$b^* = \frac{f_1g_1 + f_2g_2}{g_1^2 + g_2^2}$$

- $b > 0$ as long as $f_1g_1 + f_2g_2 > 0$
- $b = 0$ if $f_2 = 0$ and $g_1 = 0$
  i.e. *no incentive* if t_2 is not productive and t_1 is not reflected in the test
- $b = 1$ if $f_1 = g_1, f_2 = g_2$
  i.e. *full incentive* if the sensitivity of test scores is the same as the sensitivity of human capital
- f_1g_2, f_2g_2 are substitutes (they both lead to higher bonus)
  - f_1 and g_1 raise the effect of each other on b, so they are complements
  - f_2 and g_2 raise the effect of each other on b, so they are complements

Then, the overall optimal choice of teachers is:

$$\begin{cases} t_1^* &= \frac{f_1g_1^2 + f_2g_1g_2}{g_1^2+g_2^2} + \overline{t_1} \\ t_2^* &= \frac{f_1g_1g_2 + f_2g_2^2}{g_1^2+g_2^2} \end{cases}$$

This is not the most efficient solution where we have perfect information, unless $f_1 = g_1, f_2 = g_2$. However, this is the best outcome achievable.

The educational authority should design the test to make p as close to h as possible.

10.2.4 Use Value-Added (VA) instead of Raw Score

In its simplest form, VA is measured as:

$$VA = T_1 - T_0$$

where:

- T_0: beginning of year test score
- T_1: end of year test score

VA measures the learning during the year.

It is better (more accurate) than using T_1 along, because it accounts for the different starting point of students.

However, simply using $T_1 - T_0$ may not be a good measure of learning. We could also use more complex measures of VA, such as $\frac{T_1}{T_0}$ or controlling for other student characteristics.

10.2.5 Common Problems of Cardinal Performance Pay

Teaching to the Test

Teachers respond to incentives. This issues is especially serious when we are far away from $f_1 = g_1$ and $f_2 = g_2$ (poor test design).

Meanwhile, we cannot simply change the test items frequently to avoid the problem, because we need assessments to be comparable over time.

Corruption/Cheating

Teachers may manipulate which students show up for testing. Jacob and Levitt, 2003 shows that around 5% of the classrooms in Chicago's public school system have this issue.

10.3 Relative Performance Pay

10.3.1 Overview

One of the most important problem of the cardinal performance pay system we discussed above is that *test scores do not have a scale.*

One alternative is to pay teachers based on an ordinal measure of performance, such tournaments between teachers. This is especially useful when observing absolute performance is difficult, but ranking is easy.

Specifically, the educational authority offers a fixed pot of money, which will be paid to workers based on their relative performance. In this case, some will win while some will lose.

The advantages of relative performance pay are:

- It's harder to corrupt.
  For example, even if all teachers collude to exercise low effort, there is a strong incentive to deviate and get the prize.
- No longer need to have similar assessments over time, limiting teaching to the test

10.3.2 Pay for Percentile (Barlevy and Neal, 2012) – Best in Theory

Barlevy and Neal, 2012 designs a "pay for percentile" system, and demonstrates that such a system can elicit efficient effort from all teachers in all classrooms.

Specific procedures:

- For each student is a school system, first form a comparison set of students against which the student will be compared (the general idea is to form a set that contains all other students in the system who begin at the same level of baseline achievement in a comparable classroom setting)

- ◇ At the end of the year, give an assessment to all students.
- ◇ Assign each student to a percentile score based on his end-of-year ranking among the students in his comparison set
- ◇ For each teacher, sum these within-peer percentile scores over all the students she teaches and denote this sum as a percentile performance index
- ◇ Pay each teacher a common base salary plus a bonus that is proportional to her percentile performance index

10.4 Evidences on Performance Pay

Overview

Most studies find performance pay useful.

Program	Place/Time	Description	Study	Results
Career Ladder Evaluation System	Tennessee, 1985–1989	5-stage career ladder for teachers that awarded bonuses after reaching the third stage. Bonuses ranged from $1000 for the third certification level to $7000 for the fifth certification level.	Dee and Keys (2004)	Math scores increased by 3%, reading scores by 2%, but only increases in math were statistically significant. Teachers on the lower 3 rungs were more effective at promoting math achievement, and teachers at the higher rungs were more effective at promoting reading achievement.
CIS	Kenya, 1997–1998	School based program that awarded bonuses to schools for either being the top-scoring school or for showing the most improvement. Bonuses were divided equally among all teachers in a school, who were working with grades 4–8.	Glewwe, Ilias, and Kremer (2010)	The program increased government exam participation. It did not increase scores in the first year, but treatment scores rose by .14 SDs relative to controls in the second year. However, this improvement did not persist after the completion of the program, and there were no improvements on parallel low stakes NGO exams.
ABC	North Carolina 1996–Present	School based program that awards bonuses to all teachers if school-wide scores meet statistical target. $1500 maximum bonus. Part of the state accountability system.	Vigdor (2009)	Large Gains in Math and Reading Proficiency on the State Test. NAEP trends suggest that reading gains are suspect, but math gains may reflect real improvement.
DSAIP	Dallas, 1991–1995	Schools were ranked based on gains in student learning. Approximately the top 20% of schools received awards for each member of their staff. Principals and teachers received $1000 bonuses, and other staff received $500.	Ladd (1999)	Pass rates on standardized tests of Reading and Math increased significantly, but only for white and Hispanic students. Black students did not exhibit significant gains relative to other cities. The dropout rate decreased more in Dallas relative to other cities from 1991 to 1994.
KIRIS	Kentucky, 1992–1996	Schools could earn bonus money if they achieved growth targets for school-wide performance on assessments as well as other objectives.	Koretz and Barron (1998), Koretz (2002)	Scores on KIRIS assessments rose dramatically in all subjects, but Kentucky students showed modest gains or no improvement on many comparable NAEP or ACT tests.
Teachers' Incentive Intervention	Israel, 1995–1997	Schools were ranked based on their relative performance adjusted for student background characteristics. Credits hours, matriculation exam pass rates, and dropout rates served as performance criteria. The top 1/3 of schools received awards. 75% of the award went to bonuses for teachers, 25% of the award went to facilities improvements.	Lavy (2002)	Clear evidence of improved outcomes on most dimensions with larger impacts observed in religious schools. Matriculation certificates did not increase in secular schools, but average test scores increased in both secular and religious schools.
PRP	England, 1999–Present	Teachers submit applications for bonus pay and provide documentation of better than average performance in promoting student achievement. Teachers who are promoted become eligible for future raises if they meet documented criteria.	Atkinson et al. (2009)	No clear evidence of improvement. Given one strategy that sought to adjust for experience differences between treatment and controls, English and Science teachers showed modest improvement. Math teachers did not show improvement.

Program	Place/Time	Description	Study	Results
TAP 1999-	17 states, 227 schools	Statistical VAM method produces teacher performance indices of 1 to 5. Teachers with scores of 3 or greater earn a bonus that increases with their score.	Hudson (2010)	Introduction of TAP raises math achievement relative to samples in a synthetic control group by .15 SDs. Reading impacts positive but smaller and imprecisely estimated.
Israel (experiment)	Israel, 2000–2001	A rank-order tournament among teachers of each subject, with fixed rewards of several levels. Teachers were ranked based on how many students passed the matriculation exam, as well as the average scores of their students.	Lavy (2009)	There were overall improvements in pass rates in Math and English due to an overall change in teaching methods, increased after school teaching, and increased responsiveness to student needs among teachers. Increased exam participation rates also played a role in test score gains.
Andhra Pradesh (Randomized Evaluation Study)	India, 2005–2007	100 schools got group bonuses based on school performance, and 100 got individual bonuses based on teacher performance. Bonuses were awarded based on how much the percentage gain in average test scores exceeded 5%.	Muralidharan and Sundararaman (2010)	After 2 years, students in incentive schools scored better than the control group by .28 SDs in math, and .17 SDs in language. These students also tended to do better on questions of all difficulty. Students at incentive schools also did better in non-incentive subjects.
Achievement Challenge Pilot Project (ACPP)	Little Rock, Arkansas, 2004–2007	Individual teachers were awarded bonuses based on their students' improvement on the Iowa Test of Basic Skills. Awards were determined by the level of growth and number of students a teacher had.	Winters (2008)	There was statistically significant improvement in all three subjects (math, reading, language) tested. Students increased 3.5 Normal Curve Equivalent (NCE) points in math (.16 SDs), 3.3 NCE points in reading (.15 SDs), and 4.6 NCE points in language (.22 SDs).
POINT	Nashville, TN, 2006–2009	Teachers volunteered to participate in a performance pay experiment. Bonuses of 5 K, 10 K, and 15 K were awarded for surpassing the 80%, 90%, and 95% threshold in the historic distribution of value-added.	Springer et al. (2010)	Program involved 5th- through 8th-grade math teachers. Some evidence of achievement gains in 5th-grade math in years two and three, but these gains did not persist over the next school year. No evidence of positive program impacts in other grades. Attrition rates from the study were high years two and three. Attrition is concentrated among inexperienced teachers.
NYC School-Wide Bonus Program	New York City, 2007–2011	Random sample of "high-need" schools participated in a bonus pay scheme. The scheme involved team incentive pay at the school level linked to growth targets, but school compensation committees distributed the bonus money among teachers. The two bonus levels were $3000 per teacher and $1500 per teacher. The program was added on top of an accountability program that already put performance pressure on schools.	Goodman and Turner (2010)	Performance scores were weighted averages of improvements in test score performance and inspections of school environment. Target scores required lower-performing schools to make greater improvements. 2008–2009 was the only full year of implementation. 89% of eligible schools won the maximum bonus. There is no clear evidence that the program improved student achievement.
Portugal's Scale Reform	Portugal, 2007–Present	Abandoned single pay scale in favor of two scale system. Promotion to higher pay scale involved a level jump of about 25% of monthly salary. Teachers in the same school who already worked on the higher pay scale performed the performance assessments for junior teachers.	Martins (2009)	Using schools in the Azores and Madeira as well as private schools as controls, there is no evidence of achievement gains induced by the program and consistent evidence that the program harmed achievement on national exams.
MAP	Florida	Districts choose their own method for measuring teacher contribution to achievement.	No Independent Study	
Procomp	Denver, 2006–Present	Teachers and principals negotiate achievement targets for individual students. Teachers can also earn bonuses for meeting state growth expectations for their students based on statistical targets. Finally, teachers may earn bonuses for serving in a "distinguished" school. School ratings are determined by test scores, parent surveys, and attendance. Maximum performance bonus is 5%.	No Independent Study	
Qcomp	Minnesota, 2007–Present	Much of performance pay linked to evaluations of lesson plans and their implementation. Schools or districts develop their own plans for measuring teacher contributions to measured students achievement.	No Independent Study	

10.4.1 2 Specific Papers on Cardinal Performance Pay

Muralidharan and Sundararaman, 2011

Muralidharan and Sundararaman, 2011 run RCTs in elementary schools in India with individual and group incentives (3% of average pay). Results show significant improvements.

Table 2: Impact of Incentives on Student Test Scores

	[1]	[2]	[3]	[4]	[5]	[6]
	Panel A: Combined (Math and Language)					
	Dependent Variable = Normalized End of Year Test Score					
	Year 1 on Year 0		**Year 2 on Year 1**		**Year 2 on Year 0**	
	[1]	[2]	[3]	[4]	[5]	[6]
Normalized Lagged Test Score	0.499	0.497	0.559	0.568	0.45	0.447
	(0.013)***	(0.013)***	(0.018)***	(0.019)***	(0.015)***	(0.015)***
Incentive School	**0.153**	0.170	0.140	0.130	**0.217**	0.225
	(0.042)* **	(0.042)***	(0.041)***	(0.042)***	**(0.047)*** **	(0.048)***
School and Household Controls	No	Yes	No	Yes	No	Yes
Observations	68678	62614	63004	53032	49498	44213
R-squared	0.29	0.32	0.30	0.32	0.23	0.25
	Panel B: Math					
	Dependent Variable = Normalized End of Year Test Score					
	Year 1 on Year 0		**Year 2 on Year 1**		**Year 2 on Year 0**	
	[1]	[2]	[3]	[4]	[5]	[6]
Normalized Lagged Test Score	0.49	0.492	0.505	0.512	0.418	0.416
	(0.017)***	(0.017)***	(0.025)***	(0.025)***	(0.022)***	(0.023)***
Incentive School	**0.188**	0.205	0.184	0.176	**0.276**	0.286
	(0.049)* **	(0.050)***	(0.050)***	(0.050)***	**(0.055)*** **	(0.056)***
School and Household Controls	No	Yes	No	Yes	No	Yes
Observations	34109	31105	31443	26473	24584	21953
R-squared	0.28	0.30	0.28	0.30	0.23	0.24
	Panel C: Telugu (Language)					
	Dependent Variable = Normalized End of Year Test Score					
	Year 1 on Year 0		**Year 2 on Year 1**		**Year 2 on Year 0**	
	[1]	[2]	[3]	[4]	[5]	[6]
Normalized Lagged Test Score	0.516	0.508	0.617	0.627	0.483	0.476
	(0.014)***	(0.015)***	(0.014)***	(0.014)***	(0.014)***	(0.014)***
Incentive School	**0.119**	0.136	0.098	0.086	**0.158**	0.164
	(0.038)* **	(0.038)***	(0.037)***	(0.038)**	**(0.043)*** **	(0.044)***
School and Household Controls	No	Yes	No	Yes	No	Yes
R-squared	0.319	0.341	0.341	0.366	0.246	0.269

Notes:
1. All regressions include mandal (sub-district) fixed effects and standard errors clustered at the school level.
2. Constants are insignificant in all specifications and are not shown.
3. School controls include an infrastructure and proximity index (as defined in Table 1)
4. Household controls include student caste, parental education, and affluence (as defined in Table 3A)

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 10.1: Outcomes (Muralidharan and Sundararaman, 2011)

Table 8: Teacher Behavior (Observation and Interviews)

	Incentive versus Control Schools (All figures in %)			
<u>Teacher Behavior</u>	Incentive Schools	Control Schools	p-Value of Difference	Correlation with student test scores
	[1]	[2]	[3]	[4]
Teacher Absence (%)	0.26	0.25	0.191	-0.069
Actively Teaching at Point of Observation (%)	0.42	0.43	0.769	0.134***
Did you do any special preparation for the end of year tests? (% Yes)	0.64	0.32	0.000***	0.094**
What kind of preparation did you do? (UNPROMPTED) (% Mentioning)				
Extra Homework	0.42	0.20	0.000***	0.068
Extra Classwork	0.47	0.23	0.000***	0.079**
Extra Classes/Teaching Beyond School Hours	0.16	0.05	0.000***	0.183***
Gave Practice Tests	0.30	0.14	0.000***	0.104***
Paid Special Attention to Weaker Children	0.20	0.07	0.000***	0.000

Notes:
1. Each teacher is treated as one observation with t-tests clustered at the school level.
2. Teacher absence and active teaching in column 4 are coded as means over the year
3. All teacher response variables from the teacher interviews are binary and column 4 reports the correlation between a teacher's stated response and the test scores of students taught by that teacher (controlling for lagged test scores as in the default specifications throughout the paper)

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 10.2: Channels (Muralidharan and Sundararaman, 2011)

Behrman et al., 2015

Behrman et al., 2015 conduct RCTs in secondary schools in Mexico. There are three treatment groups: a group where students get bonus, a group where teacher get bonus, and a group where everyone get bonus.

Average Treatment Effects (ATE) with and without Adjustments for Copiers: All Program Years[a,b,c]

Grade	Year One AY: 2008/2009 T1	T2	T3	Year Two AY: 2009/2010 T1	T2	T3	Year Three AY: 2010/2011 T1	T2	T3
With Copying Adjustment									
Tenth Grade									
ATE	16.9	1.27	31.4	29.1	0.11	46.6	32.3	13.5	63.4
(s.e.)	(4.90)	(5.74)	(5.79)	(4.57)	(5.34)	(7.61)	(4.77)	(5.54)	(10.4)
P-value: TJ = T3	.010	<.001	-	.040	<.001	-	.002	<.001	-
Eleventh Grade									
ATE	13.6	-4.84	18.6	29.7	2.11	43.7	25.2	-2.00	42.1
(s.e.)	(5.40)	(5.50)	(7.39)	(4.89)	(6.05)	(8.33)	(4.24)	(4.31)	(5.64)
P-value: TJ = T3	.545	.004	-	.098	<.001	-	.011	<.001	-
Twelfth Grade									
ATE	9.63	4.71	28.8	21.9	-4.46	34.8	22.7	3.99	56.7
(s.e.)	(6.85)	(6.58)	(6.36)	(5.04)	(6.10)	(6.46)	(7.49)	(7.54)	(15.1)
P-value: TJ = T3	.010	<.001	-	.078	<.001	-	.015	<.001	-

Figure 10.3: Outcomes (Behrman et al., 2015)

Their results are adjusted for potential cheating. Giving everyone the bonus turns out to be the most effective, and the effect increases by time.

Student and Teacher Effort Measures by for Controls and Treatment/Control Difference: Year 3

Grade	C 10	C 11	C 12	T1 - C 10	T1 - C 11	T1 - C 12	T2 - C 10	T2 - C 11	T2 - C 12	T3 - C 10	T3 - C 11	T3 - C 12
Student:												
Avg, hrs/wk study math	4.68	4.45	4.53	.199	.408	.385	-.138	-.070	-.097	.397	.301	.370
				(.094)	(.135)	(.124)	(.091)	(.182)	(.165)	(.112)	(.135)	(.127)
Avg, hrs/wk study non-math subjects	5.56	5.48	5.32	.109	.189	.250	-.161	-.134	-.040	.152	.074	.168
				(.122)	(.122)	(.156)	(.129)	(.168)	(.152)	(.122)	(.134)	(.127)
Frac. pay attention >75% of time	.473	.479		.070	.048	.042	.015	.007	-.006	.101	.070	.050
				(.022)	(.021)	(.024)	(.028)	(.030)	(.026)	(.028)	(.023)	(.032)
Frac. never or almost never text while doing homework	.423	.429	.415	.109	.093	.056	.023	.004	-.007	.126	.097	.061
				(.023)	(.028)	(.027)	(..026)	(.028)	(.280)	(.024)	(.022)	(.225)
Frac. never or almost never watch TV while doing homework	.493	.517	.498	.077	.075	.066	-.021	-.010	-.010	.088	.093	.060
				(.028)	(.018)	(.024)	(.025)	(.022)	(.020)	(.026)	(.022)	(.027)
Frac. Gave Help to Classmates	.599	.608	.643	.056	.058	.026	-.017	-.014	-.041	.086	.087	.026
				(.020)	(.022)	(.023)	(.020)	(.019)	(.028)	(.020)	(.022)	(.028)
Frac. Report Putting Much Effort	.466	.489	.486	.077	.090	.087	-.039	-.029	-.017	.114	.093	.092
				(.022)	(.026)	(.028)	(.021)	(.030)	(.025)	(.022)	(.021)	(.037)
Teacher:												
Gave only multiple choice exams	.351	.115	.155	-.171	.120	.027	-.072	.031	-.047	-.246	-.019	.039
				(.083)	(.106)	(.085)	(.096)	(.066)	(.073)	(.080)	(.075)	(.096)
Frac. prepared students for ALI test	.168	.260	241	.202	.181	.211	.182	.155	.111	.412	.256	.176
				(.103)	(.121)	(.107)	(.091)	(.106)	(.114)	(.106)	(.110)	(.098)
Frac. helped students outside of class to prepare for ALI test	.241	.220	.203	.338	.339	.453	.341	.390	.391	.435	.554	.482
				(.104)	(.126)	(.102)	(.103)	(.111)	(.122)	(.098)	(.092)	(.103)

Figure 10.4: Channels (Behrman et al., 2015)

10.4.2 1 Paper on Relative Performance Pay

Lavy, 2009 analyses the effect of a tournament which is based on performance relative to other teachers in the same subject. (Also a RCT.)

TABLE 4—DID ESTIMATES OF THE EFFECT OF TEACHER BONUSES ON MATH AND ENGLISH OUTCOMES BASED ON THE RANDOMIZED TREATMENT SAMPLE

	Math					
	All quartiles		Estimates by quartile			
	Limited control (1)	Full control (2)	1st (3)	2nd (4)	3rd (5)	4th (6)
Testing rate						
Control group mean	0.802		0.419	0.815	0.903	0.971
Treatment effect	0.046	0.041	0.133	0.055	0.037	−0.021
	(0.027)	(0.021)	(0.051)	(0.035)	(0.021)	(0.029)
	[0.038]	[0.029]	[0.068]	[0.047]	[0.030]	[0.039]
Pass rate						
Control group mean	0.637		0.258	0.503	0.726	0.928
Treatment effect	0.110	0.087	0.146	0.209	0.106	−0.026
	(0.036)	(0.028)	(0.048)	(0.063)	(0.035)	(0.029)
	[0.051]	[0.040]	[0.065]	[0.087]	[0.047]	[0.041]
Average score						
Control group mean	55.046		21.232	46.917	63.946	77.710
Treatment effect	5.469	5.307	9.798	10.920	6.352	−0.861
	(2.292)	(1.950)	(3.497)	(4.104)	(2.122)	(2.493)
	[3.249]	[2.739]	[4.768]	[5.686]	[2.927]	[3.443]
Conditional treatment effect						
Passing rate		0.052	0.051	0.161	0.073	−0.007
Proportion of unconditional effect		59%	35%	77%	69%	–
Average score		2.323	2.465	7.006	3.541	0.839
Proportion of unconditional effect		44%	25%	64%	56%	–
Observations	9,857		2,421	2,365	2,424	2,647

Figure 10.5: Outcomes (Lavy, 2009)

The outcome improves the most in the lower half.

TABLE 8— THE EFFECT OF PAY FOR PERFORMANCE ON TEACHING METHODS AND TEACHER EFFORT

	Math teachers					
	All interviewed teachers		Eligible schools' teachers		RT schools' teachers	
	Sample mean (1)	Treatment-control difference (2)	Sample mean (3)	Treatment-control difference (4)	Sample mean (5)	Treatment-control difference (6)
Teaching methods						
Teaching in small groups	0.661	0.007 (0.051)	0.557	0.111 (0.068)	0.525	0.193 (0.078)
Individualized instruction	0.614	−0.028 (0.060)	0.600	−0.014 (0.087)	0.600	−0.008 (0.125)
Tracking by ability	0.397	0.130 (0.059)	0.471	0.055 (0.073)	0.500	−0.035 (0.102)
Adapting teaching methods to student's ability	0.942	0.011 0.023	0.914	0.038 0.037	0.900	0.030 0.055
Teacher effort						
Added instruction time during the whole year, or before Bagrut exam	0.831	0.015 (0.036)	0.871	−0.025 (0.045)	0.825	−0.022 (0.072)
Added instruction time only before Bagrut exam	0.296	0.071 (0.048)	0.300	0.067 (0.069)	0.150	0.160 (0.082)
Number of additional weekly instruction hours	2.038	1.987 (0.600)	2.959	1.066 (0.809)	2.382	1.246 (0.965)
The teacher initiated the addition of instruction hours	0.709	−0.017 (0.051)	0.714	−0.022 (0.074)	0.625	0.093 (0.105)
Teacher's additional effort was targeted at						
All students	0.587	−0.025 (0.059)	0.614	−0.052 (0.079)	0.575	−0.012 (0.106)
Weak students	0.212	−0.058 (0.042)	0.214	−0.060 (0.059)	0.200	−0.045 (0.066)
Average students	0.011	0.043 (0.018)	0.029	0.025 (0.025)	0.025	0.017 (0.033)
Strong students	0.000	0.006 (0.006)	0.000	0.006 (0.006)	0.000	0.000 (0.000)

Figure 10.6: Channels (Lavy, 2009)

Interestingly, there's no obvious channels through which the improvement happens.

Bibliography

Barlevy, G., & Neal, D. (2012). Pay for percentile. *American Economic Review*, *102*(5), 1805–1831. https://doi.org/10.1257/aer.102.5.1805

Behrman, J. R., Parker, S. W., Todd, P. E., & Wolpin, K. I. (2015). Aligning learning incentives of students and teachers: Results from a social experiment in mexican high schools [Publisher: The University of Chicago Press]. *Journal of Political Economy*, *123*(2), 325–364. https://doi.org/10.1086/675910

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. *The Quarterly Journal of Economics*, *118*(3), 843–877. https://doi.org/10.1162/00335530360698441

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, *99*(5), 1979–2011. https://doi.org/10.1257/aer.99.5.1979

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india [Publisher: The University of Chicago Press]. *Journal of Political Economy*, *119*(1), 39–77. https://doi.org/10.1086/659655

Neal, D. (2011, January 1). Chapter 6 - the design of performance pay in education. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 495–550). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00006-7

Chapter 11

The Economics of Migration: The Migrant[1]

11.1 Introduction and Motivations

11.1.1 Motivations

- Definition of Migration: movement of individuals from one region to another
  - International migration: movement of individuals across international borders
- Causes for Migration:
  - Economic reasons (e.g. education, climate, amenities)
  - Persecution, displacement as a result of war or ethnic cleansing (very dominant in past decades)
  - Preference for the host country
- Consequences of Migration:
  - Host country: direct and immediate effects in economy and society along various dimensions (wages, fiscal effects, innovation, food, habits, etc.)
  - Source country: Immediate effects through withdrawing people (brain drain / brain gain), return and remittances. Also, return migration may bring back new skills and technology learned from the host country to the home country.
- Economically motivated migration always generates a surplus, but has distributional effects at the same time
  - The main beneficiaries of migration are the migrants themselves
  - Key questions: Who else gains in source and sending country, and who may lose?

11.1.2 Focus of Economic Research on Migration

- Migrants:
  - Migration and re-migration decisions
  - Immigrant's performance in the receiving country
  - The selection of immigrants
  - The children of immigrants
- Non-migrants in Host and Source Country:
  - Impact immigration may have on receiving country: wages, employment, prices, fiscal effects, innovations, crime, etc.

[1] Written by Kuangjie Ni, Edited by Xiaotian Tian

  - Impact emigration may have on the sending country: employment, wages, income, children's education, etc.
  - Analysis of remittances
  - Cultural consequences
- The Interaction of Immigrants and Natives
  - Social cohesion, attitudes to immigration, social integration, political outcomes (e.g. voting outcomes)

11.1.3 Emprical Facts: Migration in the International Context

Stock of Migrants

Immigration: used to be a phenomenon of the "New World," but most developed Western nations today have large immigrant populations, and many Asian countries have large internal migration from rural areas to cities.

- World stock of international migrants (UN, Dept of Econ and Soc Affairs, 2019) :
  - 1995: about 161 million people, 2.8 percent of the world population
  - 2019: almost 272 million, 3.5 percent of the world population
- Stock in Europe, Northern America, Australia/New Zealand and Japan :
  - 1995: 92 million, 7.9 percent of the population
  - 2019: more than 150 million, 12 percent of the population

Number of Migrants

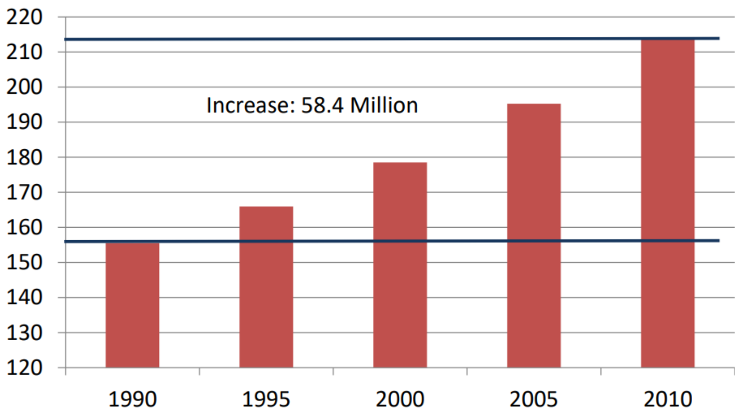


Figure 11.1: Estimated number of international migrants in millions

For the whole world, there is a 58.4 million increase in the number of international migrants from 1990 to 2010, and this figure continues to increase.

Share of Migrants

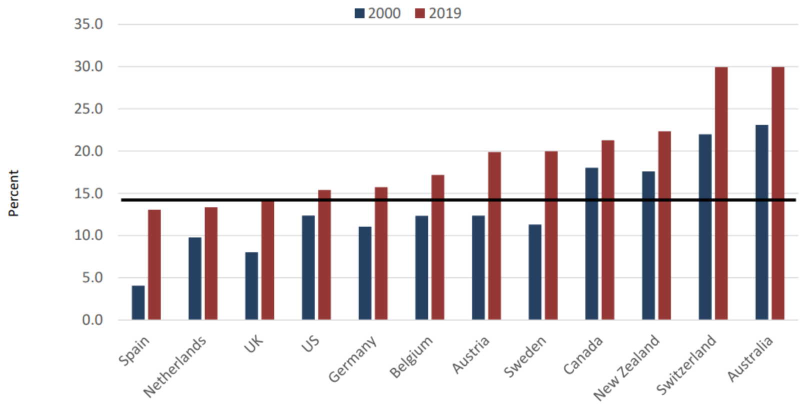


Figure 11.2: Share of foreign-born population, selected countries

- ◇ Blue bars indicate the shares of the foreign-born population in 2000, and red bars indicate the share of the foreign-born population in 2019.
- ◇ The figure in Spain has tripled.
- ◇ In 2019, one in three people in Switzerland and Australia are foreign-born.
- ◇ The horizontal line is the 2019 average for the countries shown on the figure weighted by their total population.

Education of Migrants

Table 4: Immigration and education, by area of origin

	% with lower secondary education	*% with tertiary education*
Natives	31.74	25.83
EU15	35.08	**29.35**
NMS12	23.40	21.03
Other Europe	**49.01**	14.74
North Africa and near Middle East	**50.98**	20.52
Other Africa	39.01	27.84
South and East Asia	40.04	26.26
North America and Oceania	14.10	**49.55**
Latin America	37.19	22.79
All immigrants	*38.05*	*23.51*

The table reports the percentage of natives and immigrants from each area of origin with low (column 1) and high (column 2) education, pooling all destination countries. The sample is restricted to working age population older than 25, not in full-time education and not in military service.
We define immigrants as "foreign born" in all countries, except for Germany where they are defined as foreign nationals.
Source: EU-LFS, years 2007, 2008 and 2009

Figure 11.3: Education of immigrants

- ◇ The table reports the percentage of natives and immigrants from each area of origin with low (column 1) and high (column 2) education, pooling all destination countries. The sample is restricted to the working-age population older than 25, not in full-time education and not in military service. We define immigrants as "foreign born" in all countries, except for Germany where they are defined as foreign nationals.
- ◇ The share of immigrants with lower secondary education is high (around 50 percent) in Other Europe, North Africa and near Middle East. The share of immigrants with tertiary education is high (around 50 percent) in North America and Oceania.

11.1.4 Empirical Facts: Migration in the European context

Brief History of 5 Waves of Migration to Europe

- Europe experienced five major migration waves after WW2.
- 1945 - 1960: Migrations caused by the war. About 20 million people displaced, mainly Germans.
- Second migration movement: economically motivated, started in the early 1950's - 1973 (labour migrations).
- Third wave of migration after 1973: family immigration and reunification of former labour migrants, and asylum migration.
- Fourth big movement: East-West migration, initiated in the late 1980's by a liberalisation of Soviet policy and accelerated by the fall of the Berlin wall in 1989.
- Early 1990's: Movements as consequence of independence and democratisation, of wars inflicted on many areas within and around Europe.
- 2000's: Globalisation, EU enlargement, and labour market rigidities let to large immigration flows into, and across European countries.
- Question: Next decade? Immigration is likely to be increasingly fueled by war prosecution and inequality. Also, social media make people easier to get information about the host countries and then move. Moreover, the large increase in population in Africa and, thus, the immigration imposes pressure on Europe and the US.

Immigrants across Europe: Percentage and Origin

	% Immigrants in total population	*Composition of immigrant population by area of origin*							
		EU15	NMS12	Other Europe	North Africa & Middle East	Other Africa	South and East Asia	North America and Oceania	Latin America
Austria	15.68	17.55	18.7	51.18	3.58	1.2	5.44	1.07	1.29
Belgium	11.76	41.53	6.45	13.83	18.09	10.96	5.48	1.16	2.5
Germany	**14.5**	**25.36**	**8.38**	**46.9**	**7.16**	**2.33**	**6.14**	**2.14**	**1.6**
Denmark	7.98	20.05	5.39	26.27	16.12	4.76	16.75	8.04	2.63
Spain	13.09	13.83	13.76	3.89	15.13	2.86	3.28	0.65	46.6
Finland	2.71	29.86	10.51	33.75	7.16	5.08	8.89	2.73	2.02
France	10.66	27.57	2.99	6.11	40.23	12.08	6.79	1.56	2.67
Greece	7.79	5.85	12.89	61.34	11.98	1.02	4.36	2.21	0.35
Ireland*	15.59	40.16	32.66	3.21	1.54	5.71	9.59	5.6	1.53
Italy	7.41	11.37	18.11	26.72	14.03	5.48	11.27	1.81	11.2
Netherlands	10.66	17.39	3.57	16.64	17.22	5.86	17.45	2.51	19.38
Norway	8.69	30.4	5.54	14.16	11.22	7.58	20.99	4.62	5.49
Portugal	6.48	18.51	3.06	8.31	0.23	45.04	1.73	2	21.12
Sweden	15.16	26.33	8.2	21.56	20.45	4.37	10.8	1.55	6.73
UK	**11.34**	**18.08**	**13.47**	**3.56**	**4.62**	**16.93**	**29.05**	**7.67**	**6.61**
Total	11.27	20.61	10.63	18.91	15.39	8.34	11.25	2.83	12.03
USA	**12.50**	**7.44**	**3.23**	**2.57**	**2.82**	**3.04**	**24.75**	**2.79**	**53.37**

Figure 11.4: Immigrants as a percentage of the total population, years 2007-2009

There is a large difference in immigrant shares across different countries, and the composition by area is different.

Education

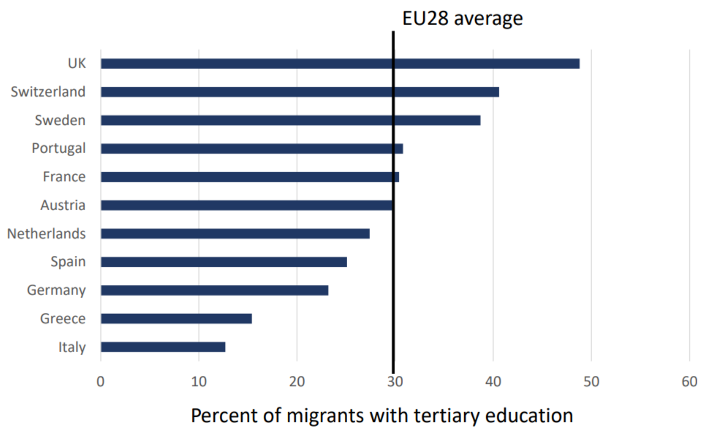


Figure 11.5: Share of immigrants with tertiary education in selected EU countries, 2018

- Immigration is not a homogeneous phenomenon across countries. It is different in terms of education and origins across countries.
- On average in EU28, one in three immigrants has tertiary education.
- The UK attracts the best-educated immigrants. Germany has a relatively low proportion of immigrants with tertiary education.

Composition of Foreign-Born Population

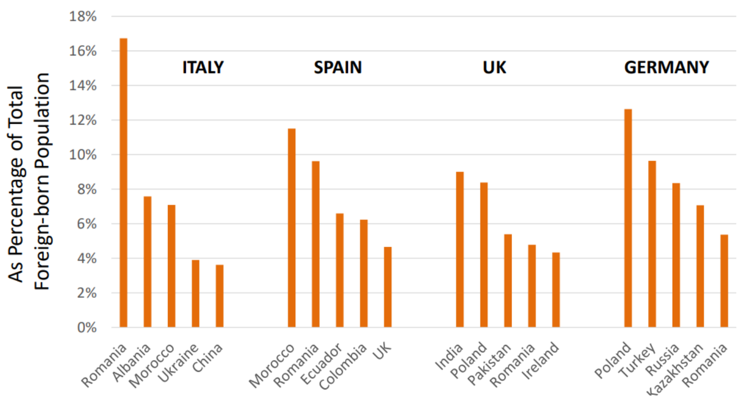


Figure 11.6: Composition of Foreign-Born Population - Main Origin Countries, 2018

There is little overlap in the main origin countries of immigrants across host countries.

Employment Rate

We want to integrate and assimilate immigrants to realise their economic potential, as this is a win-win situation. When immigrants get higher wages in the host country, they pay more taxes.

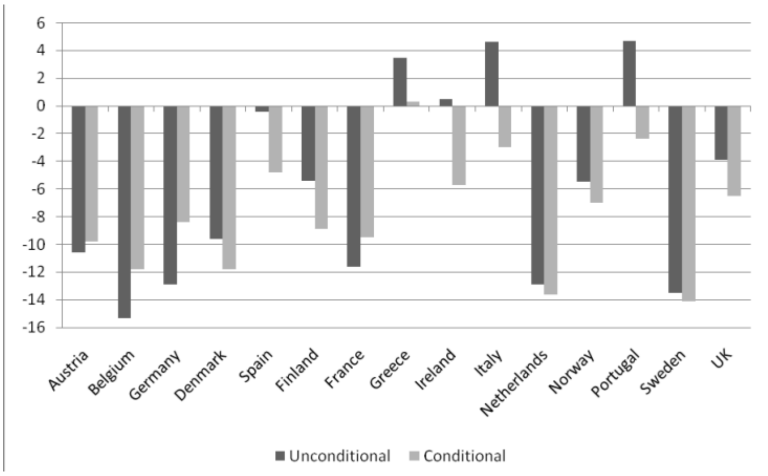


Figure 11.7: Immigrant-native employment differential

Negative bars indicate that there is a higher employment probability of natives than immigrants. Positive bars indicate that there is a higher employment probability of immigrants than natives (e.g. Greece). Southern European countries tend to have positive bars unconditionally. Their shares of immigrants are small, and immigrants come only for joining the labour markets.

Dark grey bars: unconditional; Light grey bars: conditional (controlling for observed characteristics such as education, labour market experience).

Position in Earnings Distribution

Immigrants:

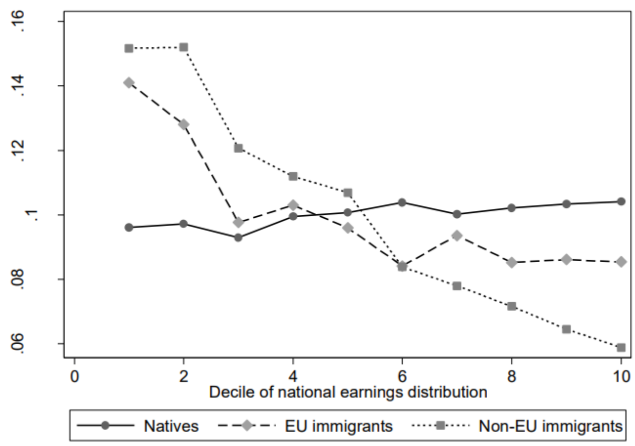
	Decile of national earnings distribution									
	1	2	3	4	5	6	7	8	9	10
Natives	9.6	9.7	9.3	10.0	10.1	10.4	10.0	10.2	10.4	10.5
EU15	11.8	10.7	8.2	9.0	8.9	9.2	10.3	10.0	10.8	11.3
NMS12	18.9	17.3	13.0	13.0	11.1	6.9	7.5	5.5	4.2	2.8
Other Europe	16.1	15.2	10.9	10.7	9.7	9.2	9.4	8.3	6.7	3.8
N.Africa & Middle East	12.8	12.7	12.7	11.4	11.9	8.2	7.5	7.2	7.1	8.6
Other Africa	13.7	15.2	15.0	11.4	13.2	8.0	6.6	6.3	5.5	5.1
South and East Asia	17.0	19.7	12.0	13.7	9.0	7.6	6.5	5.3	4.8	4.5
N.America & Oceania	7.9	6.9	11.6	10.3	10.6	9.2	6.0	9.9	8.6	19.0
Latin America	20.8	19.8	11.8	9.4	10.4	6.2	5.4	4.7	6.1	5.4

The table reports the percentage of natives and immigrants in each decile of the national earnings distribution in Belgium, Germany, Finland, France and Italy pooled. We define immigrants as "foreign born" in all countries except for Germany, where they are defined as foreign nationals.
Source: EULFS, 2009

Figure 11.8: Positions in National Earnings Distribution

- The table reports the percentage of natives and immigrants in each decile of the national earnings distribution in Belgium, Germany, Finland, France and Italy pooled. We define immigrants as "foreign born" in all countries except for Germany, where they are defined as foreign nationals.
- EU15 and Oceania have a relatively larger share of immigrants in the top 10 percent decile of the earnings distribution, while other continents have relatively more immigrants in the lowest 10 percent decile.
- Let's show the idea of this table in a graph below (only looking at EU and non-EU immigrants).

Immigrants and Natives:



The figure reports the share of natives (circles), EU immigrants (rhomb) and non-EU immigrants (squares) in each decile of the national earnings distribution in Belgium, Germany, Finland, France and Italy pooled. Source: EULFS 2009.

Figure 11.9: Immigrant and native earnings distribution

- The figure reports the share of natives (circles), EU immigrants (rhomb) and non-EU immigrants (squares) in each decile of the national earnings distribution in Belgium, Germany, Finland, France and Italy pooled.
- We can see the natives' line is nearly a straight line. However, EU immigrants and non-EU immigrants have a larger proportion in the lowest 10 percent decile of the earnings distribution (e.g. 15 percent of non-EU immigrants is in this decile), and they have a smaller proportion in the top 10 percent decile (e.g. only 6 percent of non-EU immigrants is in this decile).

Children of Immigrants

Table 10: Children in Immigrant households

	Percentage of children (<15) who live in an immigrant household					Percentage of immigrants in adult population	
	EU	*Non-EU*	EU/Non-EU	*Mixed* EU/ Native	Non-EU/ Native	*EU*	*Non-EU*
Austria	3.16	17.47	0.66	4.47	4.32	5.21	8.36
Belgium	4.09	10.69	0.69	3.78	5.11	5.08	5.49
Germany	**1.68**	**7.97**	**0.38**	**2.89**	**6.05**	**2.11**	**3.8**
Spain	1.8	8.04	0.21	2.92	3.43	3.39	8.51
France	1.68	10.08	0.28	2.94	6.52	2.89	6.99
Greece	0.93	9.68	0.08	2.16	2.4	1.18	5.4
Ireland	7.73	4.94	0.61	9.86	2.41	8.96	3.3
Italy	1.66	7.81	0.17	2.94	3.91	1.72	4.6
Netherlands	0.84	12.9	0.35	3.11	6.18	1.5	8.14
Portugal	0.68	5.89	0.32	3.24	6.59	0.54	4.02
UK	**2.12**	**11.03**	**0.48**	**2.37**	**5.06**	**3.03**	**7.44**
Total	**1.86**	**9.43**	0.34	2.95	5.16	**2.58**	**5.96**

The left panel of the table reports the share of children under the age of 15 who live in an immigrant or a mixed household. The right panel reports the share of immigrants in the total population above the age of 15.
EU (Non-EU) households are defined as households where the reference person and her or his spouse – if there is a spouse - is an EU(Non-EU) immigrant. Mixed households are households where the reference person and her or his partner have a different immigrant status. We define immigrants as "foreign born" in all countries except for Germany, where they are defined as foreign nationals.
Source: EULFS, 2007-2009.

Figure 11.10: Children of Immigrants

- The left panel of the table reports the share of children under the age of 15 who live in an immigrant or a mixed household. The right panel reports the share of immigrants in the total population above the age of 15.

Poverty

Table 11: Households with both spouses in bottom decile of earnings distribution

	Percentage of households with both spouses in bottom decile of earnings distribution	Percentage of children (<15) in households with both parents in bottom decile of earnings distribution	Percentage of children (<15) in immigrant household out of all children in households with both parents in bottom decile of the earnings distribution				
					Mixed		
			EU	*Non-EU*	EU/Non-EU	EU/ Native	Non-EU/ Native
Belgium	4.88	4.60	6.50	**23.0**	0.22	2.86	3.39
Germany	1.15	0.80	0	**19.2**	0	0	5.25
France	4.22	3.35	2.54	**19.1**	0.08	0.69	5.70
Italy	4.05	3.30	5.55	**20.0**	0.11	2.01	3.03
Total	**2.98**	**2.53**	**3.62**	**19.8**	**0.10**	**1.26**	**4.57**

Figure 11.11: Households with both spouses in the bottom decile of earnings distribution

- The left panel shows the percentage of households with both spouses in the bottom decile of earnings distribution. The middle panel shows the percentage of children (<15) in households with both parents in the bottom decile of earnings distribution, which is quite in line with the left panel. The right panel shows the percentage of children (<15) in immigrant households out of all children in households with both parents in the bottom decile of the earnings distribution.
- One in five (around 20 percent) of children (<15) in non-EU immigrant households live in poverty.

11.2 ⋆ The Migration Decision

11.2.1 2 Main Reasons for Migration & Refugee Protections

2 Reasons for Migration

Two main reasons for migrations:

- Forced movement due to natural disaster or persecution
- Economic motivated migration for better economic prospects in other areas

This is still a distinction we draw today: between regulations for "asylum" immigrants and "economic" immigrants (1951 Geneva Refugee Convention).

"Asylum" immigrants are not allowed to work in many countries, while economic immigrants can get access to the labour market.

We will only discuss economically motivated migrations here.

Refugees and Their Protection: 1951 Geneva Convention for Refugees (GCR)

Who is a Refugee? 1951 Geneva Convention for Refugees (GCR) defines as: "[a person who] owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country;...

As of April 2015, 145 states have signed the 1951 Convention and 142 have signed both the Convention and the 1967 Protocol which endows the GCR with universal coverage.

This does not address civilians fleeing wars and conflicts (covered in different forms of temporary/subsidiary humanitarian protection).

11.2.2 Classification of Migrations

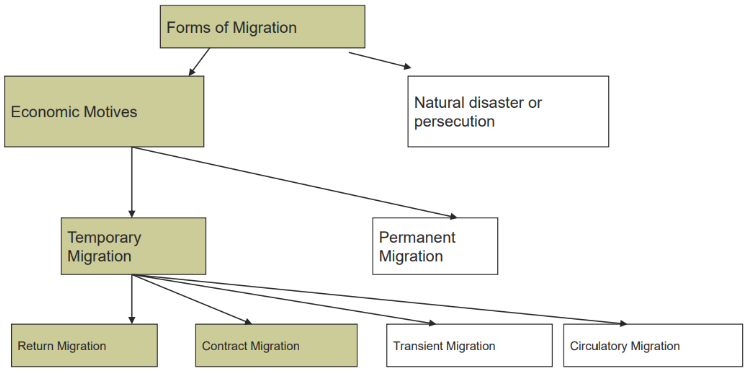


Figure 11.12: Classification of Migrations

11.2.3 ⋆ The Migration Decision Model

Migration decision of individual i from O to D depends on the comparison of the discounted flow of future earnings in O and D (y_{it}^O and y_{it}^D) net of migration cost C_i:

$$K_i = \sum_{t=0}^{T} y_{it}^D \frac{1}{(1+r)^t} - \sum_{t=0}^{T} y_{it}^O \frac{1}{(1+r)^t} - C_i$$

Migration takes place whenever $K_i > 0$.

11.3 ⋆ Assimilation and performance of immigrants in the host countries

11.3.1 Early Literature: Chiswick, 1978

Overview

Large literature in the economics of migration that is concerned with the estimation of immigrants' earnings profiles.

First papers (e.g. Chiswick, 1978) are based on cross-section data (ignoring cohort effects) and assume migrations to be permanent (ignoring selective out-migration). Both assumptions are problematic, and we will discuss them in turn.

Empirical Strategy

Log wage equations, for immigrants (I) and natives (N) :

$$\begin{cases} \ln w_i^I & = b_0^I + b_1^I EX_i + b_2^I YSM_i + e_i^I \\ \ln w_i^N & = b_0^N + b_1^N EX_i + e_i^N \end{cases}$$

where:

- EX_i: potential experience (computed as Age-years of schooling - 6)
- YSM_i: Years since migration
- b_1^I: measures the return to one year of total working experience (experience in host & home country)
- b_2^I: additional increase in earnings each year spent in host country
- $b_1^I + b_2^I$: earnings increase per year in the host country

- b_1^N: earnings increase of natives per year of labour market experience

If $b_1^I + b_2^I > b_1^N$, then the earnings of immigrants grow faster in the host country than earnings of natives (assimilation).

$\delta = b_1^I + b_2^I - b_1^N$ measures the rate of convergence of immigrant/native earnings, which represents the degree of economic assimilation.

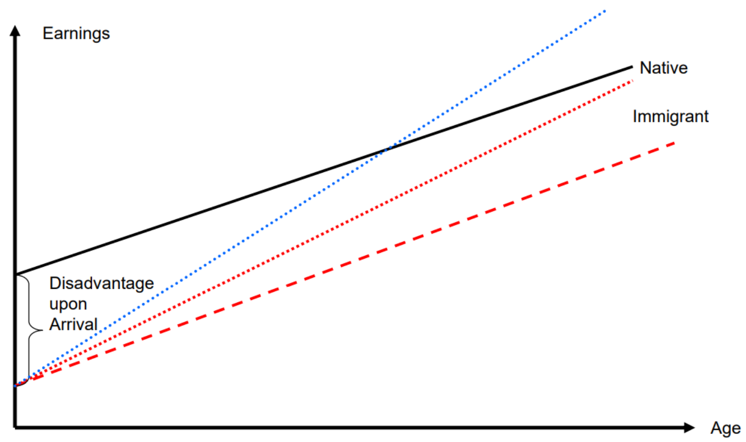


Figure 11.13: Economic Assimilation

Empirical Findings

Chiswick, 1978 uses 1970 US census data to estimate a regression similar to the one above.

Main Findings:

- Immigrants have upon arrival an earnings disadvantage of about 17 percent
- After about 10-15 years in the US labour market, earnings of immigrants overtake those of native workers.

He explains this finding with immigrants, having "more innate ability, are more highly motivated towards labour market success, or self-finance larger investments in post-school training."

11.3.2 Cohort Effects: Borjas, 1985 and Later Literatures

Problem of Early Literatures

Borjas, 1985 argues that estimation based on cross-sectional data may lead to misleading conclusions, because immigrants who differ in their years of residence in the host country have also arrived at different points in time. (Cohort Effects)

If entry wages of immigrants change over time, then this may be picked up by the coefficient on the years since migration variable, confounding differences in immigrant cohort quality with assimilation.

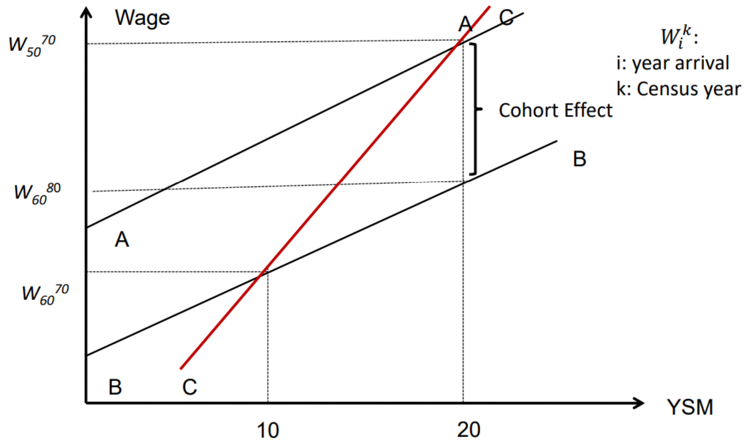


Figure 11.14: Immigration and Assimilation: Cohort Effects

Formally, with 1970 census, we can observe wages for those with 20 years of residence (w_{50}^{70}) and for those with 10 years of residence (w_{60}^{70}).

Estimated wage growth in host country will be:

$$\frac{\Delta w}{\Delta YSM} = \frac{w_{50}^{70} - w_{60}^{70}}{10}$$

This is slope of line CC, which overestimates of immigrant wage growth if cohort effect is negative (later cohorts are worse). In other words, lower initial wages of subsequent cohorts may lead to an overestimate of cross-sectional wage profiles, while improvement in cohort quality leads to an underestimate.

Problem of Chiswick (Algebraically) + Deal with Cohort Effects

We can solve this identification problem by adding another wave of cross-section. Suppose now we also have access to 1980 census data. We can decompose the raw wage growth:

$$\frac{w_{50}^{70} - w_{60}^{70}}{10} = \underbrace{\frac{w_{60}^{80} - w_{60}^{70}}{10}}_{\text{Wage Growth for the 1960 Cohort}} - \underbrace{\frac{w_{50}^{70} - w_{60}^{80}}{10}}_{\text{Cohort Effect}}$$

Again, we can see that:

$$\begin{cases} \text{Quality improves over time} & \Longrightarrow \text{Cohor Effect} < 0 \Longrightarrow \text{Underestimatin of convergence} \\ \text{Quality worsens over time} & \Longrightarrow \text{Cohor Effect} > 0 \Longrightarrow \text{Overestimation of convergence} \end{cases}$$

To deal with this, since we have 2 cross-sections (1970, 1980) now, we can estimate:

$$\Delta w = \frac{w_{50}^{70} - w_{50}^{80}}{30 - 20}$$

However, this FD estimator can only eliminate time-unvarying fixed effect (cohort effect), but it cannot control for differences in macroeconomic conditions (macro effects) in 1970 and 1980

Deal with Cohort Effects + YSM Effects

To discuss this, we need to extend our estimation equations:

$$\begin{cases} \ln w_{it}^I &= b_0^I + b_1^I EX_{it} + b_2^I YSM_{it} + b_m^I C_{im} + \sum_{t=1}^K \gamma_t^I T_{it} + e_{it}^I \\ \ln w_{it}^N &= b_0^N + b_1^N EX_{it} + \sum_{t=1}^K \gamma_t^N T_{it} + e_{it} \end{cases}$$

where:

- t is an index indicating the year in which individual i is observed
- EX_{it} is potential work experience extrapolated from data
- $T_{it} = \mathbb{1}[\text{Year of observation} = t]$ is a sampling time indicating dummy which equals to 1 if individual i is drawn from the cross-section in year t
- γ_t^I, γ_t^N measure the macro effects on log wages of immigrants and natives
- C_{im} is an indicator function, which equals to the calendar year m in which immigrant i arrived

Problem: there is perfect multicolinearity: $YSM_{it} = T_{it}(t - C_{im})$. This means macro effect and cohort effect cannot be separately identified with repeated cross-sections $\Longrightarrow$ we cannot estimate return to YSM without further assumptions.

Potential Solutions:

- Solution 1: Same Cohort Effect within Group
  - Firstly, we can assume the cohort effect b_m^I to be the same for certain cohorts, such as immigrants who arrived within a decade. Then, we can estimated γ_t^I by exploiting variations within a decade. Conversely, we can also assume the macro effect to γ_t^I to be the same for certain cohorts. In the extreme case where there is no macro nor cohort effect, we are estimating the same model as Chiswick, 1978.
  - Assumptions like this need to be carefully justified by data: if one has strong reason to believe that the inflow of immigrants over a particular period is of roughly the same quality (for instance because immigrants all arrived from one particular source country) then this may be a plausible assumption
- Solution 2: Same Macro Effect for Immigrants and Natives
  - Borjas (1985) assumes that the macro effect is the same for immigrants and natives: $\gamma_t^I = \gamma_t^N$
  - Then, we can first estimate the macro effects γ_t^N using data of natives, and use this to identify cohort effects in the immigrant equation
  - Nevertheless, there are evidences suggesting that change in macroeconomic conditions is likely to have different effects on wages of natives and immigrants (Dustmann, Glitz and Vogel, EER 2010)
- Solution 2+: Parameterising Macro Effects
  - Bratsberg et al. (2005) show that the common macro effect assumption leads to serious bias in assimilation profiles for the US
  - The provide an alternative method of estimating γ by parameterising the macro effect at regional level, allowing for different variations for immigrants and natives depending on local unemployment rates

11.3.3 Selective Out-migration

Out-migration is Relevant

We have not yet considered return migrations which is indeed empirically relevant:

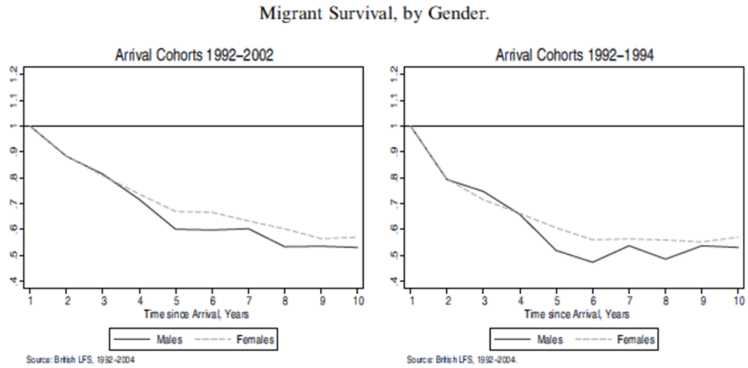


Figure 11.15: Immigrant Survival Rates (Dustmann and Weiss, 2007)

And out-migration is heterogeneous:

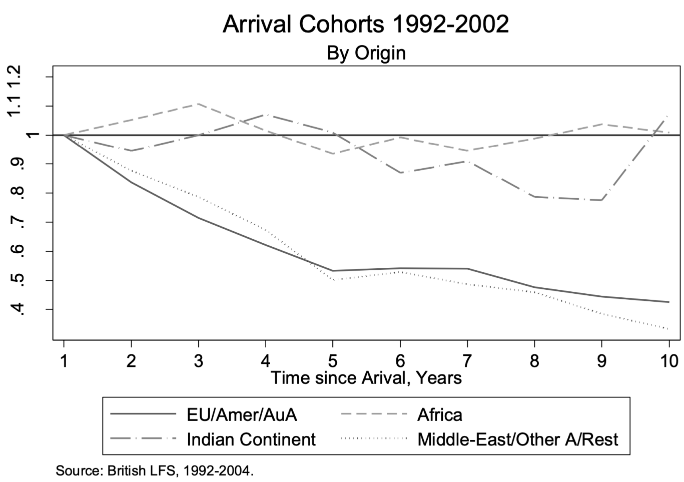


Figure 11.16: Immigrant Survival Rates (Dustmann and Weiss, 2007)

Out-migration Causes Problems

Return migrations could lead to two problems:

- Return migration leads to a possibly selective out-migration, so that earnings profile estimates are biased
- Return migration may lead to heterogeneity in the earnings paths of immigrants, due to differences in behaviour between immigrants induced by differences in their return intentions

Selective Out-migration

There is evidence of selective out-migration on education, age, and earnings; and literatures report both positively and negatively out-selection on observed characteristics. This may cause problems on the estimation of immigrants' earnings profiles:

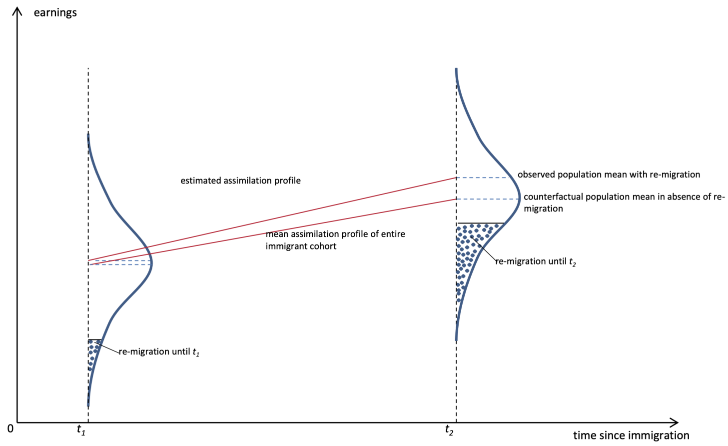


Figure 11.17: Assimilation Profiles, Wages (LFS, 1992-2002)

Ideally, we want to retrieve the lower red line while we will get the higher red line with selections. Panel data will be very useful in tackling with this issue.

An example for this:

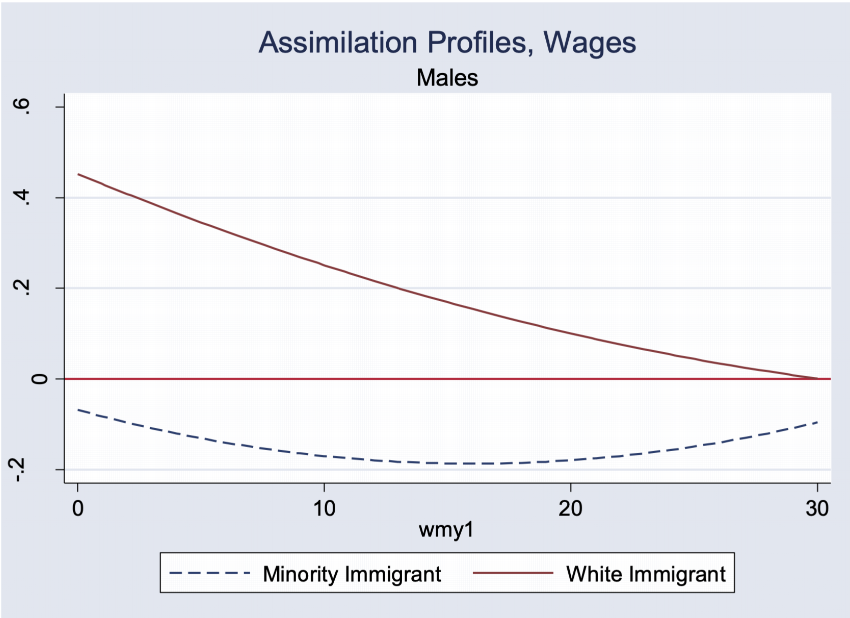


Figure 11.18: Assimilation Profiles, Wages (LFS, 1992-2002)

As shown above, what seems to be "negative assimilation" of white immigrants is actually selective out-migration where those with highest earnings leave the UK (e.g. financial workers).

Bibliography

Borjas, G. J. (1985). Assimilation, changes in cohort quality, and the earnings of immigrants [Publisher: The University of Chicago Press]. *Journal of Labor Economics*, *3*(4), 463–489. https://doi.org/10.1086/298065

Chiswick, B. R. (1978). The effect of americanization on the earnings of foreign-born men [Publisher: University of Chicago Press]. *Journal of Political Economy*, *86*(5), 897–921. Retrieved March 11, 2023, from https://www.jstor.org/stable/1828415

Chapter 12

Immigration and Wages

12.1 Motivation

Chapter 13

The Taxation of Earnings

13.1 Motivation and Facts

13.1.1 Overview of Motivations

Why re-examine earnings taxation?

- New empirical findings on labour supply elasticities
- New insights from optimal tax design theory
- A need to look at the whole income tax/benefit system
- Changes in the economic environment
    - Growth in earnings and wealth inequalities
        - Growth in the top income and wealth share
        - Fall in the relative earnings of the lower educated, especially the relative earnings of low skilled men
    - Changes in employment patterns
        - Growth in single person & single parent households
        - Growth in migration
    - Changes in employment patterns
        - Growth of female labour supply
        - Changes in early retirement behaviour

13.1.2 Facts

Income inequality has risen dramatically since 1970s:

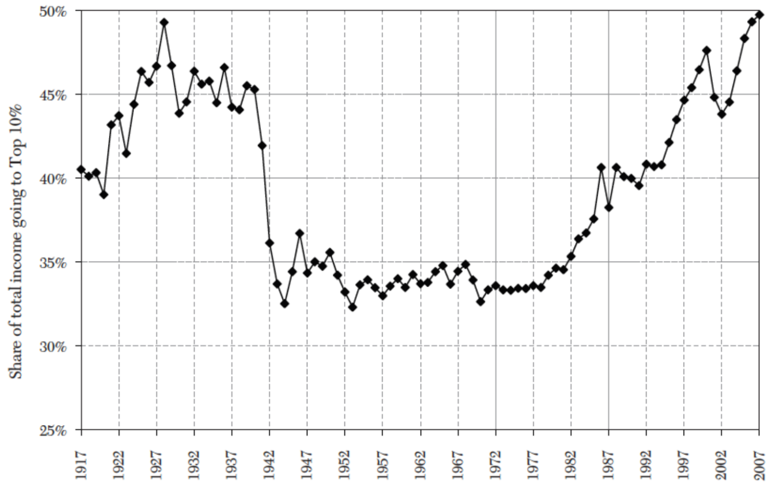


Figure 13.1: The Top Decile Income Share in the United States, 1917–2007 (Atkinson et al., 2011)

The income share increased the most for the top percentile:

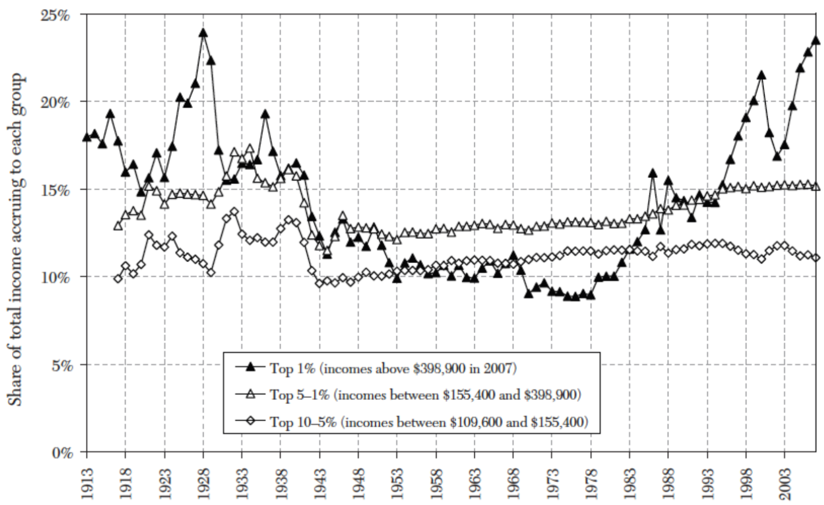


Figure 13.2: Decomposing the Top Decile US Income Share into three Groups, 1913–2007 (Atkinson et al., 2011)

Meanwhile, the income gap between different educational groups also increased:

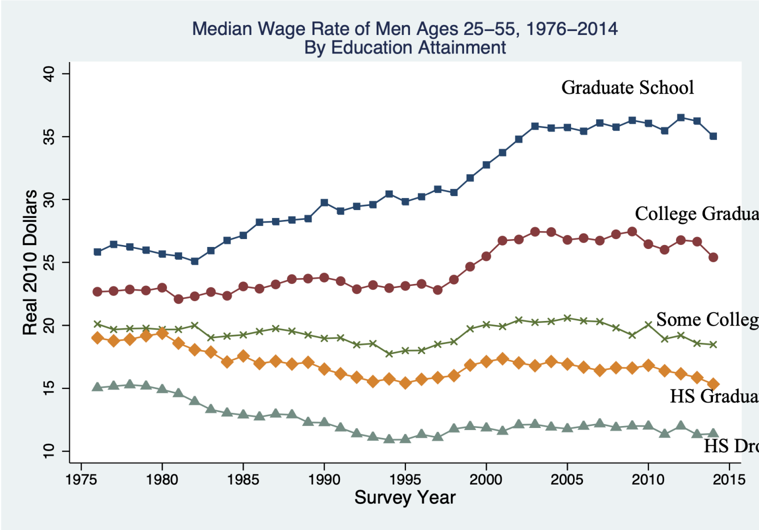


Figure 13.3: Median Wage Rate of Men Ages 25-55, 1976-2014 (Blundell et al., 2018)

13.2 Introduction to Tax System

13.2.1 Example Tax and Benefit Systems

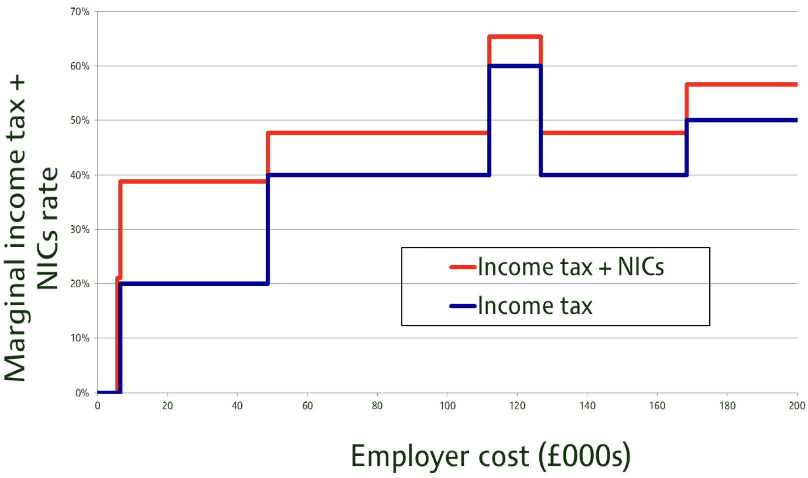


Figure 13.4: Income Tax Schedule for Those Aged Under 65, UK 2010-11 (Mirrlees Review, 2011)

And the tax rate could be different for different sources of income:

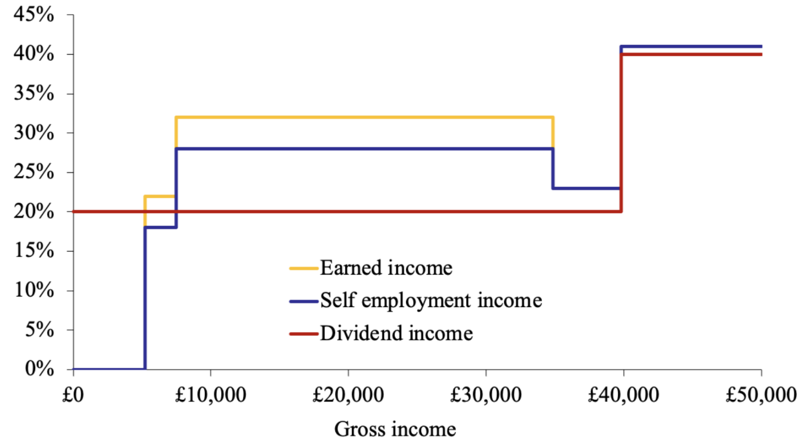


Figure 13.5: Income Tax Schedule by Sources of Income for Those Aged Under 65, UK 2010-11 (Mirrlees Review, 2011)

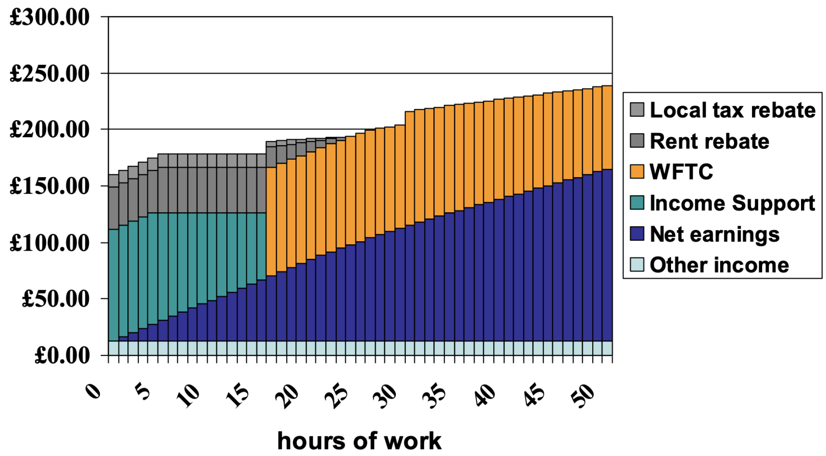


Figure 13.6: Budget Constraint for Single Parent in UK

13.2.2 Measuring Effective Tax Rates

The measurement of effective tax rates should:

- incorporate complete tax and transfer system
- consider both intensive and extensive margins
- account for earned income tax credits

Effective Marginal Tax Rate (MTR)

The Effective Marginal Tax Rate (MTR) is the proportion of an £1 of extra earning retained in the tax and benefit system. This includes all employer taxes and contributions as well as the full set of taxes and benefits.

Participation Tax Rates (PTR)

The Participation Tax Rates (PTR) is the net loss, through taxes and benefits, of earnings in work relative to being out of work. Typically, we assume a person works full-time when calculating PTR.

13.2.3 General Form of Earned Income Tax Credits (EITC)

Overview of EITC

Definition of EITC: A subsidy given to low-income households in return for labour force participation.

Goals of EITC:

- To increase worker welfare
- To offset high MTR and PTR created by income support, rent, and housing benefit schemes

Some features of EITC:

- Kinked benefits: credit depends on earnings and number of children, and there are 3 sections:
  1. Phase-in: credit is flat percentage of earned income or jumps at minimum hours threshold ($MTR < 0$)
  2. Flat range: receive maximum credit ($MTR = 0$)
  3. Phase-out: credit is phased out at a flat rate ($MTR > 0$); in the UK, the phase-out region in non-linear to encourage full-time work
- Credit is based on family earnings and is typically increasing in the number of children
- An individual must be participating the labour force to receive EITC (and subject to a minimal working hours eligibility constraint in the UK, which varies with family strucutres)

EITC in U.S.

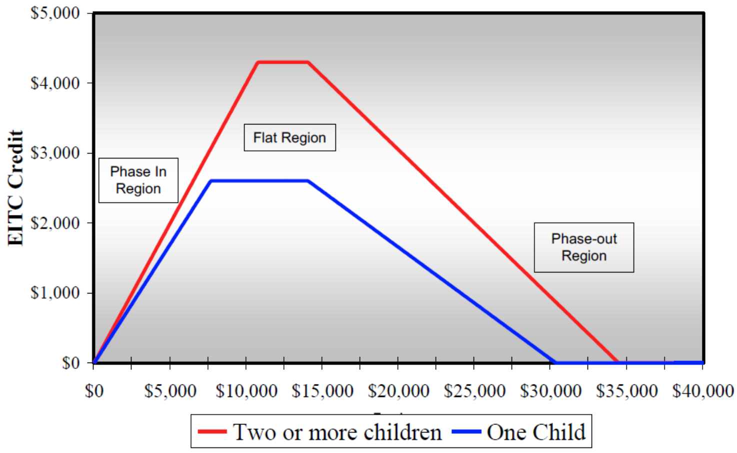
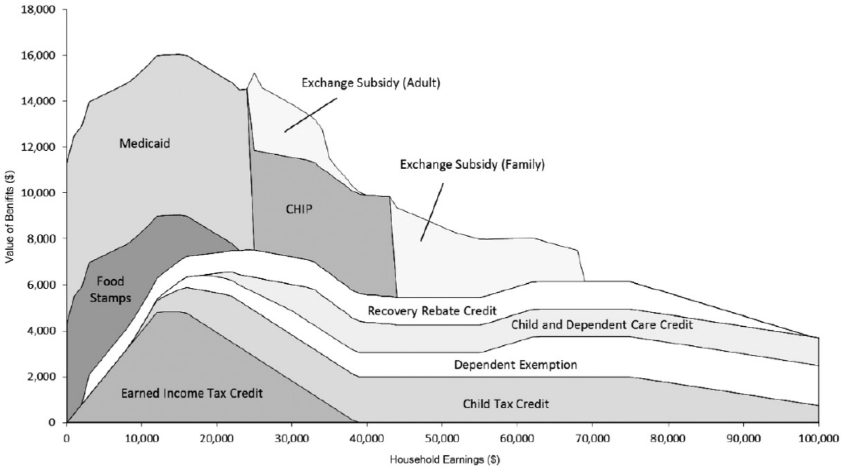


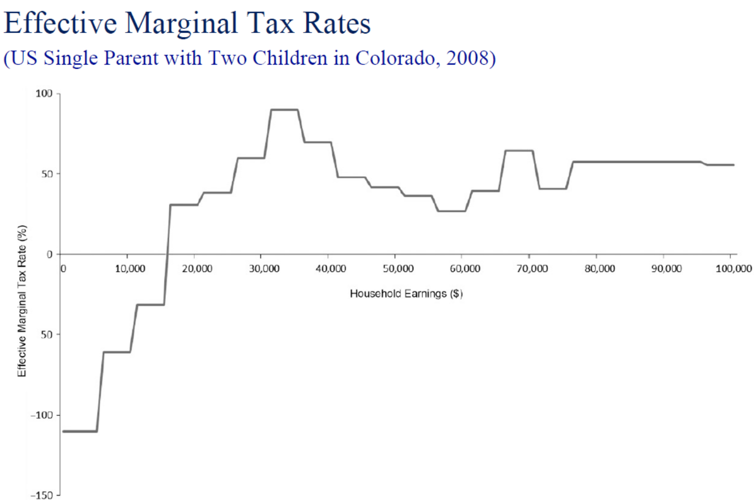
Figure 13.7: The EITC Schedule in US

The US EITC schedule provides larger credit and covers higher earners for families with two or more children.



Source: Urban Institute (NTJ, Dec 2012).
Notes: Value of tax and value transfer benefits for a single parent with two children.

Figure 13.8: Universally Available Tax and Transfer Benefits in US



Source: Urban Institute (NTJ, Dec 2012).
Notes: Value of tax and value transfer benefits for a single parent with two children.

Figure 13.9: US Effective Marginal Tax Rates (MTR)

When income is low, MTR is largely negative, providing big subsidies. As income increases, MTR rises and stabilises at around 50%.

Working Families Tax Credit (WFTC) in the UK

Particular features of UK Working Families Tax Credit:

- Hours of work condition:
  - There is a minimum hours rule – 16 hours per week
  - An additional hours-contingent payment at 30 hours
- Family eligibility – adult credit plus amounts for each child in full time education or younger
- Income eligibility – family net income has to be below a certain threshold

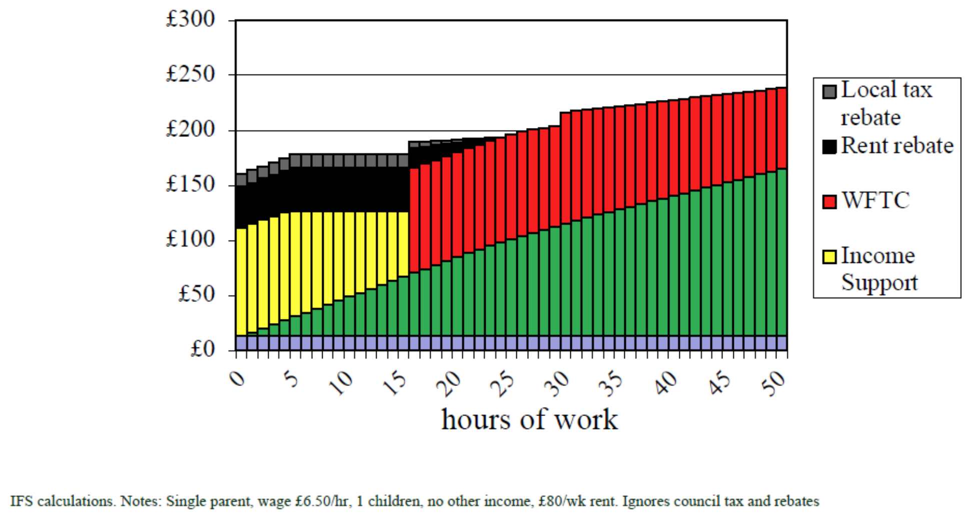


Figure 13.10: The Interaction Between Taxes, Tax Credits and Benefits in UK

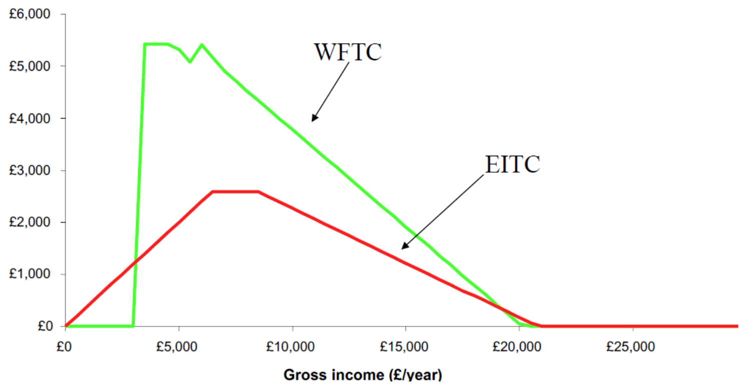


Figure 13.11: The UK and US Tax Credit System Compared

Why are EITCs Typically Directed to Working Age Adults with Children?

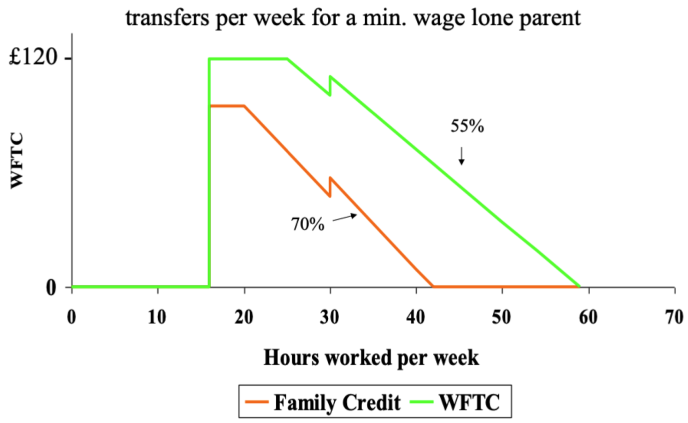
- To promote child welfare
  Supporting a child in the early life could lead to higher human capital in the long run, promoting inter-generational welfare
- To encourage parents with low skills/income to re-enter the labour market
  Income tax credits can be interpreted as a boost of wages for those who are in low-wage professions, making parents less likely to stay at home (especially effective for working mothers due to their high labour supply elasticity)
- To reduce the cost of childcare
  With more disposable income, parents can spend more money instead of time on childcare, and hence increase their working time

How EITCs Attempt to Balance the Dual Objectives of Work Incentives with Redistribution?

In short, EITC schemes are designed to transfer incomes to households in the lower end of the income distribution while provide them with incentives to stay in the labour force.

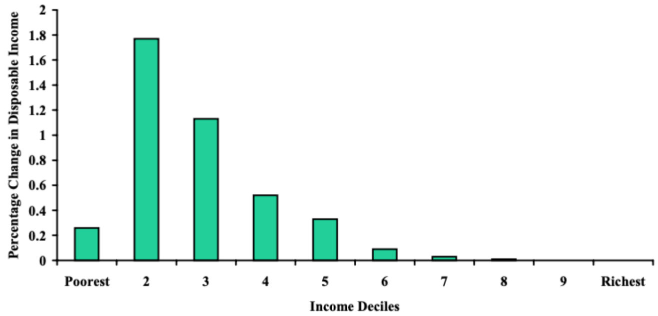
An Case Study of the UK:

- Study the policy shift: family credits → working family tax credits (WFTC)



◇ Distributional effect (Dilnot and McCrae, 1999):

Figure 8: Distributional impact of WFTC

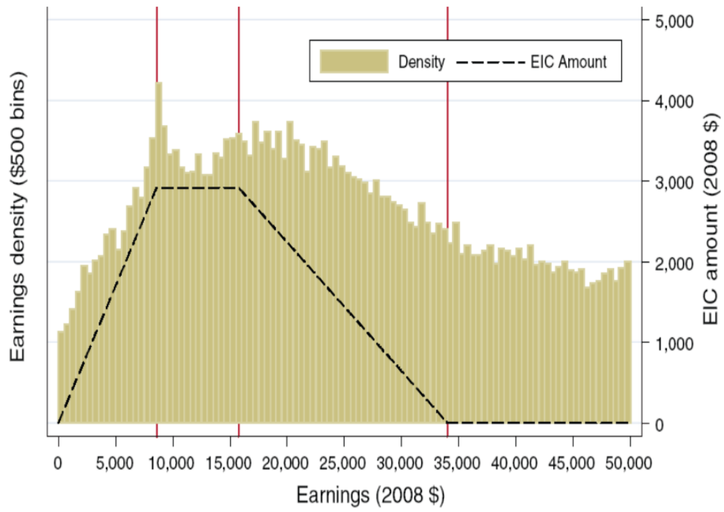


◇ Incentive effect: EITCs shifts up both the initial wage and the budget constraint and hence induces people to both participate in the labour market and work more (if substitution effects dominate). Blundell et al., 2000 used simulation to show that both LFPR and hours worked increase for single parents earning median hourly wage:

Summary

	Mean	*Standard deviation*
Change in participation	+2.20%	[0.42%]
Average change in hours (all)	+0.75	[0.16]
Average change in hours (workers only)	+0.22	[0.04]
Average hours before reform (all)	10.20	
Average hours before reform (workers only)	25.70	

Bunching at Tax Kinks and the EITC



Source: Saez (2010)

Figure 13.12: Bunching at Tax Kinks and the EITC – One Child Families, U.S.

If people are optimising, they should be clustered at kinks (bunching) and avoid gaps. Here, we observe some extents of bunching at the first kink, but not at the second one.

13.3 ⋆ Optimal Tax Design on the Top Earners

13.3.1 Introduction

Mirrlees, 1971 views taxation as an information problem. The government only observes whether people work and how much they earn. The government cannot observe effort – it cannot distinguish a high ability person working few hours from a low ability person working a lot.

In this framework, a tax schedule is chosen to maximise social welfare and raise a required amount of revenue. It needs to reach a balance between redistributive aims and effort incentives. If it taxes high-ability people too much, they may choose to supply less effort. Thus, we need good estimates of labour supply elasticises.

13.3.2 Three Impacts on Social Welfare

Setup

- ◇ Consider a reform that changes the top tax rate τ by a small amount $d\tau$.
- ◇ Let z be the earned income being considered for taxation
- ◇ The top income bracket begins at income z^*
- ◇ There are N taxpayers in the top bracket

Mechanical Effect on Tax Revenue (dM)

With no behaviour response, an increase of the top rate will increase government revenue. This is beneficial to society, as the revenue can be used for government spending or higher transfers.

Mathematically:

$$dM = N \times (z - z^*) \times d\tau > 0 \tag{13.1}$$

Behaviour Response on Tax Revenue (dB)

Higher top rate can also induce top bracket taxpayers to reduce their hours of working (substitution effect) and decrease their earnings. This is a cost to society as tax revenue will fall.

Note that there will be no change for taxpayers in other tax bracket because nothing has changed for them.

Mathematically, this behavioural effect depends on the elasticity of earnings with respect to the net of tax rate $(1-\tau)$

$$e = \frac{dz}{d(1-\tau)}\frac{1-\tau}{z} = -\frac{dz}{d\tau}\frac{1-\tau}{z}$$

$$dz = -e \times z \times \frac{d\tau}{1-\tau}$$

Hence, tax revenue will be reduced by:

$$dB = -N \times e \times z \times d\tau \times \frac{\tau}{1-\tau} \tag{13.2}$$

Welfare Effect (dW)

Any increase in the top rate will reduce the welfare of top bracket taxpayers. This is a loss to society.

If the government values redistribution, then the marginal "value" of income will be small after the income goes above certain level. In the limit, when a taxpayer's income is extremely high, the welfare effect will be negligible relative the the mechanical effect on tax revenue.

The government gives a value of g to an extra £1 earned by top tax bracket taxpayer. This will be strictly less than 1 because the weighted sum of welfare weights in unity.

Thus, the welfare effect of higher marginal tax rate on incomes above z^* is:

$$dW = -g \times N \times (z - z^*) \times d\tau < 0 \tag{13.3}$$

13.3.3 The Choice of the Top Tax Rate

Optimum

Summing up the three effects (equation 13.1, 13.2, and 13.3). The overall effect of a small change in top rate ($d\tau$) is:

$$dM + dB + dW = Nd\tau(z - z^*)\left(1 - g - ea\frac{\tau}{1-\tau}\right)$$

where $a = \frac{z}{z-z^*}$ is the Pareto parameter: higher a indicates a thinner tail.

At optimum, this has to be zero:

$$dM + dB + dW = Nd\tau(z - z^*)\left(1 - g - ea\frac{\tau}{1-\tau}\right) = 0$$

This can be simplified to:

$$\tau^* = \frac{1-g}{1-g+ae} \tag{13.4}$$

Interpretation

There are some nice interpretations of this simple formula:

- Note that a is a parameter of the upper tail of the Pareto distribution with pdf: $f(z) = \frac{C}{z^{1+a}}$
- If g is approximately zero (low weight placed on high-earners / purely revenue maximising), then:
$$\tau^* = \frac{1}{1+ae} \tag{13.5}$$
This is very simply to calculate if we know the Pareto parameter a and taxable income elasticity e
- Estimation of the Pareto parameter a and taxable income elasticity e is important

The next section explores the estimation of those two parameters. We will see that with estimated $a = 1.67$ and $e = 0.46$ (using DiD method), the revenue maximising tax rate for the top 1% is aroun 56%.

13.3.4 Extension: The Choice of the Top Tax Rate with Migration Responses

In addition, individuals may respond to higher taxes by migrating to other countries. This is similar to an extensive margin.

Denote the migration elasticity as m, and ignore distributional effects (g is approximately zero: low weight placed on high-earners / purely revenue maximising).

The optimal top tax rate will be:
$$\tau^* = \frac{1}{1+ae+m} \tag{13.6}$$

By nature, the migration elasticity m is hard to estimated, and there are discussions on whether it changes with the business cycle.

Note that if the migration elasticity m is high, governments need certain accordance to aviod ”tax competitions.”

13.4 Estimating the Pareto Parameter a and Taxable Income Elasticity e

13.4.1 Estimating the Pareto Parameter a

The Pareto parameter a is easy to estimate – we only need to fit the pdf curve with real world data.

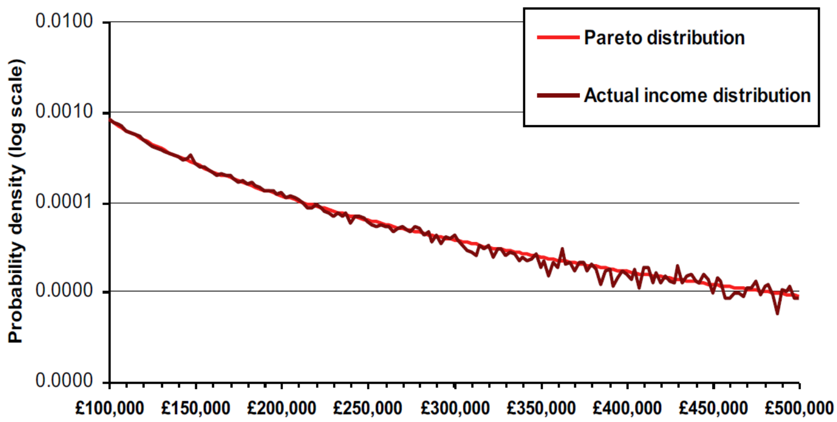


Figure 13.13: Top Income Distribution in the UK

The Pareto parameter for the UK is quite accurately estimated to be 1.67.

Note that Pareto distribution has no moments because its upper tail is not capped.

13.4.2 Measuring Response Elasticity: Overview

Measuring people's responsiveness with respect to a change in MTR is hard. Typically, researchers adopt 4 broad approaches:

- Randomised Control Trials (RCT)
- Quasi-experinments (DiD)
- Bunching Approaches
- Structual Approaches

13.4.3 RCT: Canadian Self Sufficiency Program (SSP)

Setup

Caveat: this RCT assesses workers at the lower end of the income distribution.

The Canadian Self Sufficiency Program (SSP) is designed to answer: do financial incentives encourage work among low skilled lone parents?

To encourage employment among welfare recipients, especially single parents on welfare, randomly selected participants are treated as following:

- A tax credit equals to 50% of earnings is provided (on earnings up to \$ 36000)
- Work at least 30 hours per week to be eligible

The budget constraints of treated/control groups are:

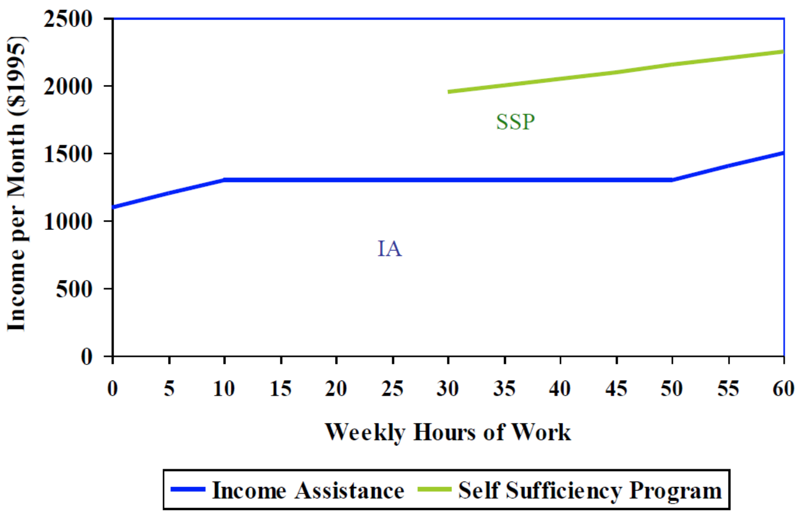


Figure 13.14: Budget Constraint for a Single Parent on Minimum Wage

Results

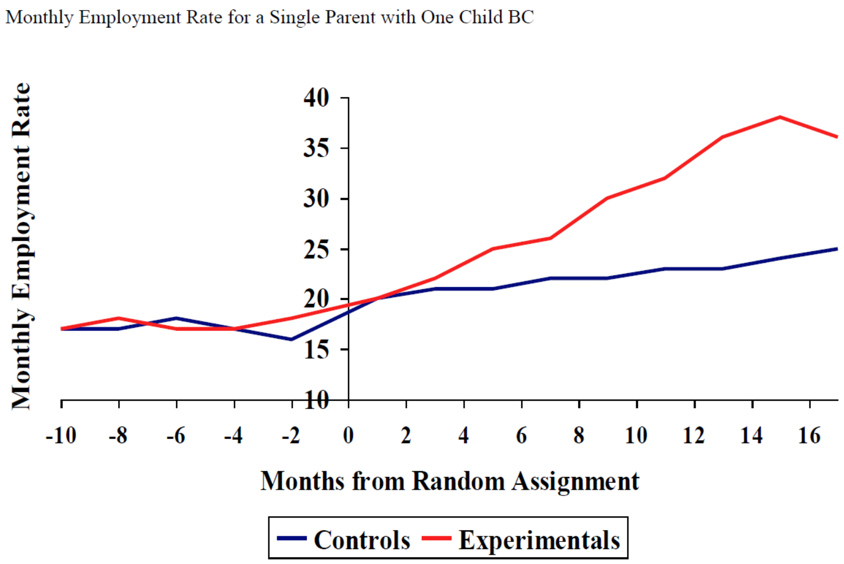


Figure 13.15: Monthly Employment Rate for a Single Parent with One Child

The treatment/control groups are well balanced before the treatment. After treatment, we observe that the employment rate of participants with tax credits becomes significantly higher than the control group.

Caveat for RCTs in Social Science

Unlike in Medicine or Physics, RCTs in social science can never be perfect because we cannot generate a perfect placebo. There will always be some placebo or spillover effects.

13.4.4 Quasi-experiments (Diff-in-Diff)

History of Top Tax Rates in the UK (Brewer et al., 2010)

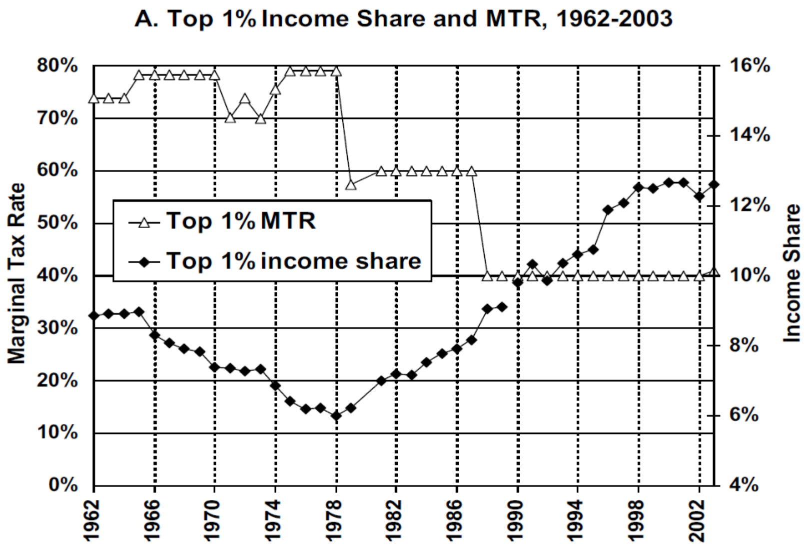


Figure 13.16: Top 1% Income Share and MTR, 1962-2003

From 1978 to 1990, the MTR of the top 1% income earners decreases dramatically.

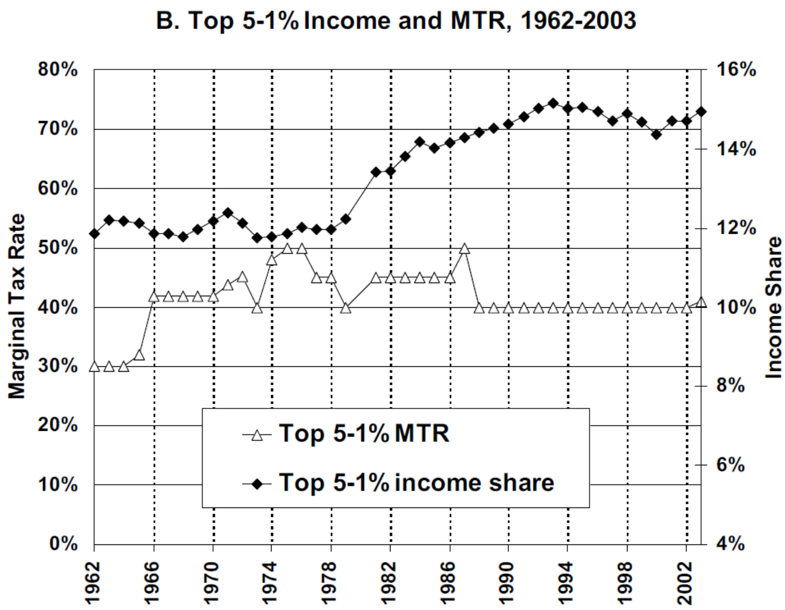


Figure 13.17: Top 1-5% Income Share and MTR, 1962-2003

On the other hand, the MTR of the top 1-5% income earners stays relatively stable, which provides a control group.

A Review of DiD Method

(I use the notation from ECON0021 Microeconometrics here.)

There are two key assumptions of the DiD method:

- ◇ Common Trend: in absense of treatment, the control and treatment group will have the same change in outcome:

$$E[Y_{0,i2} - Y_{0,i1}|D_{i2} = 1] = E[Y_{0,i2} - Y_{0,i1}|D_{i2} = 0]$$

- ◇ Invariant Composition: the composition of control and treatment group cannot be altered

Under those two assumptions, the DiD estimator identifies the average treatment effect on the treated (ATT):

$$\beta_{DiD} = \underbrace{E[Y_{i2} - Y_{i1}|D_{i2} = 1]}_{\text{Diff. in Treatment Group}} - \underbrace{E[Y_{i2} - Y_{i1}|D_{i2} = 0]}_{\text{Diff. in Control Group}} = \underbrace{E[Y_{1,i2} - Y_{0,i2}|D_{i2} = 1]}_{ATT}$$

Compared with ATE, ATT is more relevant here because we are mostly interested in the effect on those who are actually treated.

Result: Taxable Income Elasticities of Top Earners in the UK

	Simple Difference (top 1%)	DiD estimates using top 5-1% as control
1978 vs 1981	0.32	0.08
1986 vs 1989	0.38	0.41
1978 vs 1962	0.63	0.86
2003 vs 1978	0.89	0.64
Full time series	0.69 (0.12)	0.46 (0.13)

Figure 13.18: Estimated Taxable Income Elasticity at the Top in the UK

With updated data, the estimate remains in the 0.35-0.55 range with a centre at 0.46, but the standard error is high.

Meanwhile, be aware that DiD method may have various problems here. For example, there is a key relationship between the size of elasticity and the tax base (Slemrod and Kopczuk, 2002), and there may be adjustment frictions.

13.4.5 Bunching Approaches

Introduction

There is one important source of identification "neglected" in early studies: when there are kinks in the budget constraint, people bunch due to the discontinuous changes in the return to work.

With certain assumptions, we can use non-parametric methods to retrieve a "local" elasticity from the mass at the bunching point.

Throughout the bunching section, we denote the taxable income elasticity as ϵ.

Bunching at Kink Points

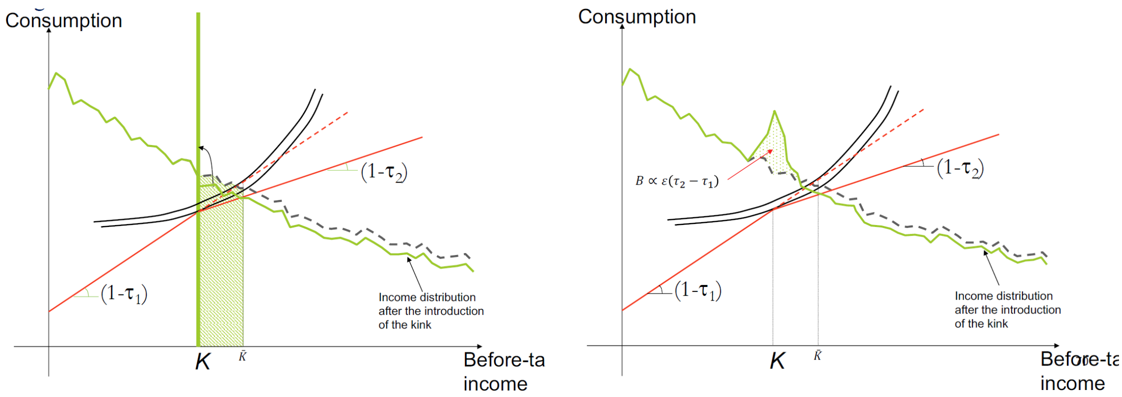


Figure 13.19: Bunching at Kink Points

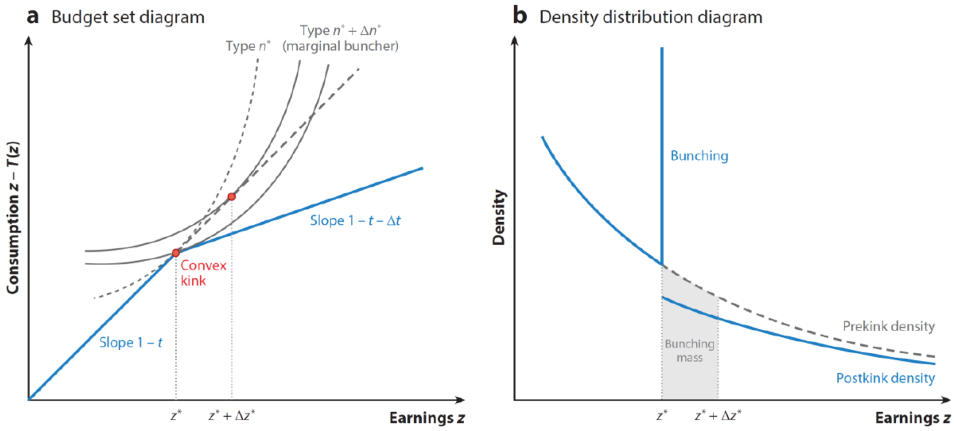


Figure 13.20: Bunching at Kink Points - Another Illustration

Bunching Calculations: Start from Simple Single Upper Tax Bracket

The individual's optimisation problem is:

$$\max c - \frac{\alpha}{1+\frac{1}{\epsilon}} \cdot \left(\frac{e}{\alpha}\right)^{1+\frac{1}{\epsilon}} \quad \text{s.t. } \; c \leq (1-\tau_1)e$$

where $e = wh$ represents earnings and ϵ represents elasticity.

Solving this optimisation, we can obtain the expression for optimal earnings:

$$e^* = \alpha(1-\tau_1)^{\epsilon}$$

Bunching Calculations: Generalise to Several Brackets

We can then generalise our calculation to multiple tax brackets.

Solving the optimisations, we can derive the range of earnings (if there were no kink) where individuals will move to the kink point:

$$k \leq e^* \leq k\left(\frac{1-\tau_1}{1-\tau_2}\right)^{\epsilon} = \overline{k} \tag{13.7}$$

where k is the level of earning where MTR changes (the kink point).

Bunching Calculations: Mass of Bunches

From equation 13.7, we can calculate the mass at the bunching point:

$$\begin{aligned} B &= \int_k^{\overline{k}} f(e^*)de^* \\ &\approx \frac{f(k)+f(\overline{k})}{2} \cdot (\overline{k} - k) \\ &= \frac{f(k)+f(\overline{k})}{2} \cdot k \cdot \left(\left(\frac{1-\tau_1}{1-\tau_2}\right)^{\epsilon} - 1\right) \end{aligned}$$

Then, using the approximation $f(k) \approx f(\overline{k})$ around k and the log approximation, we get the final bunch mass equation for our estimation:

$$B \approx k \cdot f(k) \cdot \epsilon \cdot (\tau_2 - \tau_1) \tag{13.8}$$

We can retrieve an estimate of ϵ by combining estimated mass B, estimated/extrapolated distribution $f(k)$, and institutional parameters τ_2, τ_1, k.

Bunching: Issues and Extensions

In equation 13.8, the estimates of $B, f(k)$ are sensitive to our assumptions. In reality, motivating our assumptions could be hard because MTRs vary by sources of taxable income as well as income level, and they have different elasticities.

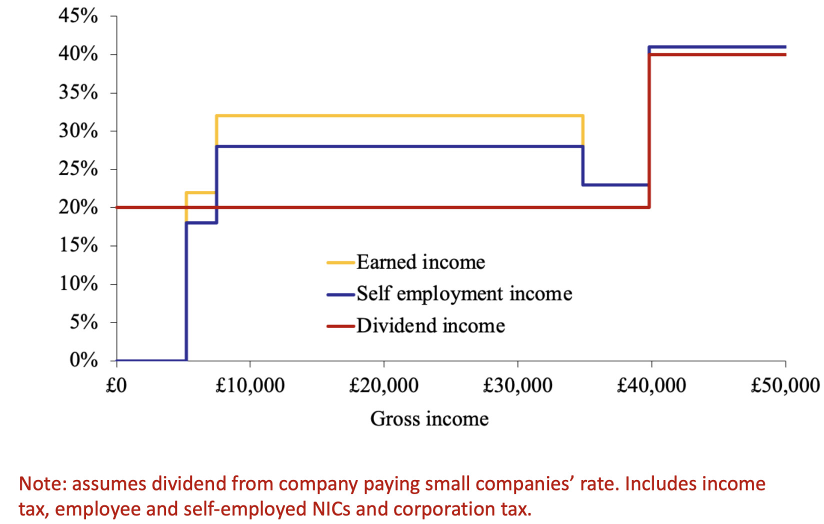


Figure 13.21: MTR of Different Sources of Incomes

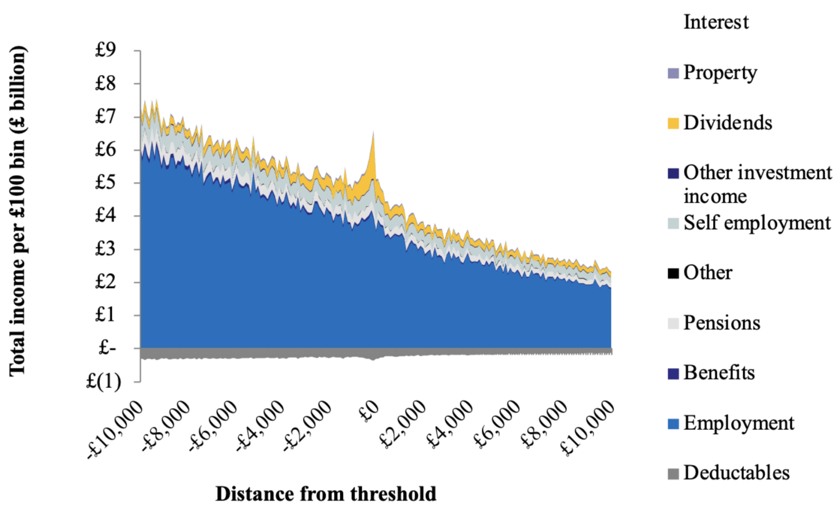


Figure 13.22: Different Sources of Incomes Have Different Elasticities

13.4.6 Discussions and Extensions

Here are some possible topics for discussion:

- Has the taxable income elasticity e changed over time?
- Is the method for estimating e reliable?
- Is Pareto distribution assumption a good one? (Probably yes)
- How would a bargaining model change the arguments? With higher top tax rate τ, the top income earners (usually business owners) may gain additional bargaining power over low-earning workers (Piketty et al., 2014).

13.5 ⋆ The General Tax Schedule

13.5.1 Introduction

The framework of optimal MTR discussed in section 13.3 can be generalised to other income earners. At optimum,the costs and benefits of a small change in MTR are always balanced.

13.5.2 Three Impacts on Social Welfare

Setup

Notations:

- For income z:
  - $H(z)$ is the cumulative distribution (CDF) of individuals
  - $h(z)$ is the density (PDF) of individuals
- Total tax function $T(z)$ depends on income:
  - There is a lumpsum grant $T(0)$
  - Combined with a schedule of MTR $T'(z) = \frac{dT(z)}{dz}$
- Consider a reform that changes the MTR $T'(z)$ by $d\tau$ in a small band of income $(z, z + dz)$

As in subsection 13.3.2, there will always be three impacts of a small change in MTR.

Mechanical Effect on Tax Revenue (dM)

With no behaviour response, an increase of the top rate will increase government revenue. This is beneficial to society, as the revenue can be used for government spending or higher transfers.

Mathematically:

$$dM = (1 - H(z)) \cdot dz \cdot d\tau \tag{13.9}$$

Behaviour Response on Tax Revenue (dB)

Again, higher MTR acts disincentive to working – people may work less (substitution effect) and their taxable incomes will drop. This is a cost to society since tax revenue will decrease.

This behavioural effect depends on the elasticity of earnings with respect to the net of tax rate $(1 - \tau)$. For each individual, the change in MTR ($d\tau$) changes the earning by:

$$dz = -e \cdot z \cdot \frac{d\tau}{1 - T'(z)}$$

Since there are $h(z)dz$ such taxpayers, tax revenue will be reduced by:

$$dB = -h(z) \cdot dz \cdot e \cdot z \cdot d\tau \cdot \frac{T'(z)}{1 - T'(z)} \tag{13.10}$$

Welfare Effect (dW)

Extra taxes also generate a welfare cost.

Let $G(z)$ be the average social value of distributing £1 uniformly from taxpayers with income above z (marginal value placed on income for individuals whose incomes are above z). The welfare effect is:

$$dW = -dM \cdot G(z) \tag{13.11}$$

Note that the "welfare cost" will have the opposite sign.

13.5.3 The Choice of the Top Tax Rate

Optimum

Summing up the three effects (equation 13.9, 13.10, and 13.11). The overall effect of a small change in top rate ($d\tau$) is:

$$dM + dB + dW = \Big\{(1 - G(z)) \cdot (1 - H(z)) \cdot dz \cdot d\tau\Big\} - \Big\{h(z) \cdot dz \cdot e \cdot z \cdot d\tau \cdot \frac{T'(z)}{1 - T'(z)}\Big\}$$

At optimum, this has to be zero:

$$dM + dB + dW = \Big\{(1 - G(z)) \cdot (1 - H(z)) \cdot dz \cdot d\tau\Big\} - \Big\{h(z) \cdot dz \cdot e \cdot z \cdot d\tau \cdot \frac{T'(z)}{1 - T'(z)}\Big\} = 0$$

This can be simplified to:

$$\frac{T'(z)}{1 - T'(z)} = \frac{(1 - G(z)) \cdot (1 - H(z))}{e \cdot z \cdot h(z)} \tag{13.12}$$

$$= \frac{1}{e} \cdot \frac{1 - H(z)}{h(z) \cdot z} \cdot (1 - G(z)) \tag{13.13}$$

Interpretation

We can see that:

- The optimal tax rate decreases with taxable income elasticity e
- The optimal tax rate decreases with $G(z)$ which measures the marginal value placed on income for individuals whose incomes are above z
- The optimal tax rate decreases with the hazard ratio $\frac{1-H(z)}{h(z) \cdot z}$, which measures the thinness of the distribution

13.5.4 Extensions: Negative MTR and Extensive Margins

Negative MTR?

In this framework with only intensive margins, negative MTRs are never optimal: if the MTR were negative in some range, then increasing it a bit in that range would raise the tax revenue and decrease the net earnings of taxpayers in that range ($dB > 0$). However, the behavioural response (working less) would also raise tax revenue as $MTR < 0$ in that range. Therefore, a small tax increase will increase social welfare unambiguously, so optimum cannot be achieved as long as negative MTR exists.

However, if we introduce the extensive margin / participation decision of labour supply, then this argument would change. (next)

Extensive Margins: Setup

Some additional setups:

- If an individual decides to work, he/she gets $z - T(z)$
- If he/she decides not to work, he/she gets $-T(0)$
- Suppose the utility function is linear: $u = c - q$
  - c is disposable income
  - q is the cost of work, distributed with CDF: $P(q|z)$

- Define the elasticity of participation (extensive margin elasticity) η as

$$\eta = \frac{z - T(z) + T(0)}{P} \cdot \frac{\partial P}{\partial q}$$

- Allow MTR to be different across I earning groups indicated by i

Extensive Margins: Optimum and Interpretations

With extensive margins, the formula for optimal taxes will be:

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{1}{e_i h_i} \sum_{j \geq 1}^{I} \left[1 - g_j - \eta_j \cdot \frac{T_j - T_0}{c_j - c_0} \right] \tag{13.14}$$

Literature on labour supply suggests that extensive margin is more responsive than intensive margin. At the bottom of the income distribution, high MTRs are not necessarily desirable and negative participation tax rates could be optimal.

Bibliography

Atkinson, A. B., Piketty, T., & Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, *49*(1), 3–71. https://doi.org/10.1257/jel.49.1.3

Blundell, R., Duncan, A., McCRAE, J., & Meghir, C. (2000). The labour market impact of the working families' tax credit [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-5890.2000.tb00581.x]. *Fiscal Studies*, *21*(1), 75–104. https://doi.org/10.1111/j.1475-5890.2000.tb00581.x

Blundell, R., Joyce, R., Norris Keiller, A., & Ziliak, J. P. (2018). Income inequality and the labour market in britain and the US. *Journal of Public Economics*, *162*, 48–62. https://doi.org/10.1016/j.jpubeco.2018.04.001

Brewer, M., Saez, E., & Shephard, A. (2010). Means-testing and tax rates on earnings. *Mireless Review: Dimensions of Tax Design.*

Dilnot, A., & McCrae, J. (1999, October 3). *Family credit and the working families tax credit.* https://doi.org/10.1920/bn.ifs.1999.0003

Mirrlees, J. (1971). The theory of optimal income taxation. *Review of Economic Studies*, *38*, 175–208. https://doi.org/10.2307/2296779

Piketty, T., Saez, E., & Stantcheva, S. (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, *6*(1), 230–271. https://doi.org/10.1257/pol.6.1.230

Slemrod, J., & Kopczuk, W. (2002). The optimal elasticity of taxable income. *Journal of Public Economics.*

Chapter 14

The Taxation of Capital and Savings

14.1 Guiding Principles for Taxation of Savings

14.1.1 Do Not Distort Optimal Individual Decisions

Economic efficiency argument suggests that the trade-off between consuming today and consuming in the future that individuals face should reflect the return to investment in productive capacity in the economy. At equilibrium, the marginal rate of transformation of consumption goods should be the real interest rate.

If we define the normal return to be the return on a safe investment, then the normal return should reflect this trade-off. Individuals optimise accordingly, maximising their utility. Any tax/subsidy should not distort this trade-off, so taxing normal return on savings cannot be optimal.

We will see this argument is subject to assumptions, and they will be discussed in the end (section 14.2.8).

14.1.2 Atkinson-Stiglitz Theorem (Atkinson and Stiglitz, 1976)

Atkinson and Stiglitz, 1976 show that under two key assumptions that:

- All consumers have preferences separable between consumption and labour
  This indeed states that the marginal benefit of consumption at anytime should not depend on labour supply
- All consumers have the same preference over consumption (i.e. same sub-utility function of consumption)
  This indeed states that all consumers are similar in their desire to smooth consumption across their life-cycle and across potentially uncertain states of the world

Under those assumptions, taxing consumption differently in the first- and the second- period is not optimal (taxing all savings is equivalent). The fundamental idea is we should not create a wedge between the intertemporal marginal rate of substitution (IMRS) and intertemporal marginal rate of transformation (IMRT) between consumer goods in different periods.

14.2 ⋆ Optimal and Non-Optimal Taxation in An Inter-temporal Context

Based on the guiding principles above, we use a simple model to show some insights on which kinds of taxes could be optimal.

14.2.1 Setup: Simple Two-Period Model without Uncertainty

Consider a 2 periods model where:

- An individual receives a fixed income/endowment Y_1 in period 1, and allocates it between consumption in period 1 C_1 and consumption in period 2 C_2
- Savings $Y_1 - C_1$ earn a risk-free rate of return r, and all the payout is consumed in period 2
- All individuals can borrow or lend at this exogenous risk-free interest rate r
- The individual only cares about consumption with a discount rate ρ

14.2.2 Optimisation with No Tax

In absence of taxation, the individual solves this maximisation:

$$\begin{aligned} \max_{C_1, C_2} \quad & V = U(C_1) + \frac{1}{1+\rho} \cdot U(C_2) \\ \text{s.t.} \quad & C_2 = (1+r) \cdot (Y_1 - C_1) \end{aligned} \tag{14.1}$$

This can be written as:

$$\max_{C_1} V = U(C_1) + \frac{1}{1+\rho} \cdot U\big((1+r) \cdot (Y_1 - C_1)\big)$$

Solving this, we obtain the first order condition and the familiar Euler equation for intertemporal allocation of consumption:

$$\frac{\partial V}{\partial C_1} = \frac{\partial U}{\partial C_1} - \frac{1+r}{1+\rho} \cdot \frac{\partial U}{\partial C_2} = 0 \iff \text{IMRS} = \frac{\frac{\partial U}{\partial C_1}}{\frac{\partial U}{\partial C_2}} = \frac{1+r}{1+\rho} \tag{14.2}$$

14.2.3 Pure Income Tax (Not Optimal)

Suppose there is a tax on income at a constant rate t.

Then, the individual has to pay a tax of tY_1 on the endowment received in period 1 and a tax of $t \cdot r \cdot [(1-t)Y_1 - C_1]$ on the interest income received in period 2 (the principal is not taxed, only interest is taxed).

The optimisation problem becomes:

$$\begin{aligned} \max_{C_1, C_2} \quad & V = U(C_1) + \frac{1}{1+\rho} \cdot U(C_2) \\ \text{s.t.} \quad & C_2 = [1 + (1-t)r] \cdot [(1-t)Y_1 - C_1] \end{aligned} \tag{14.3}$$

Solving this, we will get a different FOC and a different IMRS:

$$\frac{\partial V}{\partial C_1} = \frac{\partial U}{\partial C_1} - \frac{1+(1-t)r}{1+\rho} \cdot \frac{\partial U}{\partial C_2} = 0 \iff \text{IMRS} = \frac{\frac{\partial U}{\partial C_1}}{\frac{\partial U}{\partial C_2}} = \frac{1+(1-t)r}{1+\rho} \tag{14.4}$$

Different IMRS in equation 14.2 and 14.4 indicates that, as a pure income tax incorporates tax on capital earnings, the rate of return deceases, distorting the inter-temporal allocation of consumption.

14.2.4 Pure Consumption Tax / VAT (Could Be Optimal)

Suppose that consumption in each period is taxed at a constant rate τ.

Then, a consumption C_i requires an outlay/expenditure of $O_i = (1 + C_i)\tau$. After period 1, savings will be $Y_1 - O_1$, which generates an outlay/expenditure in period 2 of $O_2 = (1 + r) \cdot (Y_1 - O_1)$. Consumption in period 2 will be $C_2 = \frac{O_2}{1-\tau}$.

The optimisation problem becomes:

$$\begin{aligned} \max_{C_1, C_2} \quad & V = U(C_1) + \frac{1}{1+\rho} \cdot U(C_2) \\ \text{s.t.} \quad & C_2 = \frac{(1+r) \cdot [Y_1 - (1+\tau)C_1]}{1+\tau} \end{aligned} \tag{14.5}$$

Solving this, we will get the same IMRS as the case where there's no tax:

$$\frac{\partial V}{\partial O_1} = \frac{1}{1+\tau} \cdot \left(\frac{\partial U}{\partial C_1} - \frac{1+r}{1+\rho} \cdot \frac{\partial U}{\partial C_2} \right) = 0 \iff \text{IMRS} = \frac{\frac{\partial U}{\partial C_1}}{\frac{\partial U}{\partial C_2}} = \frac{1+r}{1+\rho} \tag{14.6}$$

IMRS are the same in equation 14.2 and 14.6. Therefore, a tax on consumption / VAT levied at a constant rate does not distort the intertemporal allocation.

14.2.5 Income Tax with Interest Exemption (Could Be Optimal)

Suppose an income tax is collected, but it exempts interest income. In other words, only the endowment income in period 1 is taxed at a rate τ. This is equivalent to a lump sum tax equals to τY_1.

The optimisation problem becomes:

$$\begin{aligned} \max_{C_1, C_2} \quad & V = U(C_1) + \frac{1}{1+\rho} \cdot U(C_2) \\ \text{s.t.} \quad & C_2 = (1+r) \cdot [(1-t)Y_1 - C_1] \end{aligned} \tag{14.7}$$

Again, solving this, we will get the same IMRS as the case where there's no tax:

$$\frac{\partial V}{\partial C_1} = \frac{\partial U}{\partial C_1} - \frac{1+r}{1+\rho} \cdot \frac{\partial U}{\partial C_2} = 0 \iff \text{IMRS} = \frac{\frac{\partial U}{\partial C_1}}{\frac{\partial U}{\partial C_2}} = \frac{1+r}{1+\rho} \tag{14.8}$$

IMRS are the same in equation 14.2 and 14.8. Therefore, an income tax with interest exemption does not distort the intertemporal allocation.

14.2.6 Tax on Uncertain Returns / Tax on Capital Returns with Exemption on Risk-free Return (Could Be Optimal)

In this setup, we have an capital income tax with exemption for risk-free rate of return on assets – any supernormal/excess capital income will be taxed in period 2.

Specifically, now there are two kinds of assets: risky assets with uncertain return r^R a safe asset with fixed return r^f. Income from the safe asset will not be taxed. Meanwhile, if risky assets actually provide a return higher than the risk-free rate $r^R = r^H > r^f$, then there will be a tax charge of $\tau(r^H - r^f)$ on each unit held.

Symmetrically, if the risky assets turn out to have a low return $r^R = r^L < r^f$, then there will be a tax rebate of $\tau(r^L - r^f)$ on each unit held.

Denote the amount of money invested in risky assets as q^R, the budget constraint will be:

$$C_2 = (1+r^f) \cdot (Y_1 - C_1 - q^R) + [1 - \tau(r^R - r^f)] \cdot q^R$$

Therefore, the optimisation becomes:

$$\begin{aligned} \max_{C_1, C_2} \quad & V = U(C_1) + \frac{1}{1+\rho} \cdot U(C_2) \\ \text{s.t.} \quad & C_2 = (1+r^f) \cdot (Y_1 - C_1 - q^R) + [1 - \tau(r^R - r^f)] \cdot q^R \end{aligned} \tag{14.9}$$

Solving this, we get:

$$\frac{\partial V}{\partial C_1} = \frac{\partial U}{\partial C_1} - \frac{1+r}{1+\rho} \cdot \frac{\partial U}{\partial C_2} = 0 \iff \text{IMRS} = \frac{\frac{\partial U}{\partial C_1}}{\frac{\partial U}{\partial C_2}} = \frac{1+r}{1+\rho} \tag{14.10}$$

The MRTS in equation 14.10 is the same as in equation 14.2, which means such taxation on "supernormal/excess" returns does not distort consumption allocations. Indeed, this is equivalent to a consumption tax.

14.2.7 Arguments for Taxing Excess Return to Savings

As discussed in the previously, taxation on excess returns to savings does not distort consumption decisions. Indeed, if excess returns on savings reflect pure rents (e.g. ownership of land, monopoly firm, or resources), then excess income from savings should be taxed. The real question is "how much should be taxed?"

14.2.8 Arguments for Taxing Normal Return to Savings

In our simple framework discussed above, taxing normal return to savings cannot be optimal. However, there may be some practical deviations from our stylised model.

"Tagging" High-ability Individuals – Heterogeneity in Impatience and Cognitive Ability

In experimental psychology there seems to be wide acceptance that higher-ability individuals are more patient (have lower discount rate). This deviates from the Atkinson-Stiglitz world because we no longer have the homogenous utility assumption. This implies people with high abilities, hence high earnings, also save more to consumer in period 2.

Then, if the rate of discount varies in such a predictable way, tax on normal return to savings is implicitly tax on high-ability individuals, which could be justified by redistributional purposes. This "tagging" idea aligns with Mirrless' view.

Uncertain Earnings (due to ability/productivity uncertainties)

If an individual is uncertain about his/her ability/productivity in the next period, he/she tend to have additional precautionary savings if he/she is risk-averse. In this case, taxing the return to capital/saving could be optimal, which is similar to taxing excess returns discussed before.

Non-separable Preferences between Consumption and Labour Supply

This is a deviation from assumptions in the Atkinson-Stiglitz framework. If individuals have non-separable preferences of consumption and labour supply, then it might make sense to tax less the goods that are complementary with labour supply. If second-period leisure is more complementary with consumption, taxing capital income could be justified.

Bibliography

Atkinson, A. B., & Stiglitz, J. E. (1976). The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, *6*(1), 55–75. https://doi.org/10.1016/0047-2727(76)90041-4

Epilogue

其实这是我用LaTeX写的第一本笔记。在此之前，我几乎没有任何写TeX的经验，心血来潮看了个半小时的youtubde视频就开始写了——真正的从零开始。

在这半年的时间里，我学会了越来越多经济学知识和LaTeX技巧，也得到了很多朋友们的支持和鼓励。对于我来说，相较于对备考的意义，这本笔记更像是大三的纪念册——很多年后，再看到这些文字，我希望大家都可以笑着回想起在UCL一同度过的日子。

感谢本项目的所有参与者！

田晓天
2023年4月3日