Homework 2

ECON 24450 Spring, 2021

Jake Underland

2021-04-27

Contents

1. Labor Supply Theory.

Using the lecture notes as a guide, show that 1) an increase in the wage has an ambiguous effect on labor supply and 2) a traditional welfare program unambiguously reduces labor supply.

Solution.

1) The uncompensated utility maximization problem:

$$\max_{x,l} u(x,l) \quad s.t. \quad px \leq y + w(T-l)$$

where x is amount of goods consumed, l is time spent on leisure, T is total time, and y is non-labor income. Solving the above,

$$\max_{x,l} \mathcal{L} = u(x,l) + \lambda[y + w(T-l) - px]$$
$$[x] : \; u_x(x^*, l^*) - \lambda p = 0$$
$$[l] : \; u_l(x^*, l^*) - \lambda w = 0$$
$$[y] : \; y + w(T-l^*) - px^* = 0$$

In the optimal state, define

$$\begin{aligned}
x^* &\equiv x^u(p,w,y) \\
l^* &\equiv l^u(p,w,y) \\
\lambda^* &\equiv \lambda(p,w,y) \\
h^u &\equiv T - l^* \ldots \textit{(The uncompensated labor demand/supply)} \\
\epsilon_{h^u,w} &\equiv \frac{\partial h^u/h^u}{\partial w/w} = \frac{\partial h^u}{\partial w} \cdot \frac{w}{h^u} \textit{(The uncompensated labor demand/supply elasticity)}
\end{aligned}$$

Now, we look at the dual expenditure minimization problem.

$$\min_{x,l} px + wl \quad s.t. \quad u(x,l) \geq \bar{u}$$

$$\min_{x,l} \mathcal{L} = px + wl + \lambda[\bar{u} - u(x,l)]$$

$$\begin{aligned}
[x]: &\ p - \lambda u_x(x^*,l^*) = 0 \\
[l]: &\ w - \lambda u_l(x^*,l^*) = 0 \\
[y]: &\ \bar{u} - u(x^*,l^*) = 0
\end{aligned}$$

In the optimal state, define

$$\begin{aligned}
\bar{u} &\equiv u(x^*,l^*) \\
x^* &\equiv u_x(p,w,\bar{u}) \\
l^* &\equiv u_l(p,w,\bar{u}) \\
h^c(p,w,\bar{u}) &\equiv T - l^c(p,w,\bar{u}) \ldots \textit{(The compensated labor demand/supply)} \\
\epsilon_{h^c,w} &\equiv \frac{\partial h^c/h^c}{\partial w/w} = \frac{\partial h^c}{\partial w} \cdot \frac{w}{h^c} \textit{(The uncompensated labor demand/supply elasticity)}
\end{aligned}$$

Now, we define the expenditure function $E(\cdot)$ and discuss its properties:

$$\begin{aligned}
E(p,w,\bar{u}) &\equiv p \cdot x^c(p,w,\bar{u}) + w \cdot l^c(p,w,\bar{u}) \\
&= p \cdot x^c(p,w,\bar{u}) + w \cdot (T - h^c(p,w,\bar{u})) \\
&= \underbrace{p \cdot x^c(p,w,\bar{u}) - w \cdot h^c(p,w,\bar{u})}_{E^*(p,w,\bar{u}) \textit{(minimum needed value of non-labor income)}} + wT
\end{aligned}$$

Now, differentiate $E^*(\cdot)$ w.r.t w to yield

$$\frac{\partial E^*}{\partial w} = \frac{\partial E(p,w,\bar{u})}{\partial w} - T$$

By the envelope theorem,

$$= l^c(p,w,\bar{u}) - T = T - h^c - T = -h^c]$$

We can also derive concavity:

$$\begin{aligned}
\frac{\partial^2}{\partial w^2} E^* &= \frac{\partial}{\partial w} - h^c < 0 \\
&\implies \frac{\partial}{\partial w} h^c > 0
\end{aligned}$$

The Slutsky Decomposition follows as:

$$h^c(p,w,\bar{u}) = h^u(p,w,y)$$

where $y = E^*(p, w, \bar{u})$. Totally differentiate both sides of this equation w.r.t w to get:

$$\begin{aligned}
\frac{\partial}{\partial w} h^c(p, w, \bar{u}) &= \frac{\partial}{\partial w} h^u(p, w, y) + \frac{\partial}{\partial y} h^u(p, w, y) \cdot \frac{\partial y}{\partial w} \\
&= \frac{\partial}{\partial w} h^u(p, w, E^*) + \frac{\partial}{\partial y} h^u(p, w, E^*) \cdot \frac{\partial E^*}{\partial w} \\
\implies \frac{\partial h^c}{\partial w} &= \frac{\partial h^u}{\partial w} + \frac{\partial h^u}{\partial y} \cdot \underbrace{(-h^c)}_{\text{From Envelope theorem}} \\
\implies \frac{\partial h^u}{\partial w} &= \frac{\partial h^c}{\partial w} + \frac{\partial h^u}{\partial y} \cdot h^c \qquad (\star)
\end{aligned}$$

Where

$$\begin{aligned}
&\frac{\partial h^u}{\partial y} = \frac{\partial (T - l^u(p, w, y))}{\partial y} = -\frac{\partial l^u}{\partial y} < 0 \ldots \text{(Income Effect)} \\
&\frac{\partial h^c}{\partial w} > 0 \ldots \text{(Substitution Effect)} \\
&h^c > 0
\end{aligned}$$

Thus, the substitution and income effect each work in the opposite direction in the final equation we have derived, rendering the effect of an increase in wage on labor supply ambiguous.

2) Formulating the welfare program model:

$$\begin{aligned}
G &= \text{max benefit} \\
S &= G - t(wh) \ldots \text{actual budget where } t \text{ is tax rate (benefit reduction)} \\
dw &= -tw \\
dy &= G
\end{aligned}$$

Totally differentiate uncompensated labor demand w.r.t w:

$$\begin{aligned}
\frac{dh^u}{dw} &= \frac{\partial h^u}{\partial w} + \frac{\partial h^u}{\partial y} \cdot \frac{dy}{dw} \\
\implies dh^u &= \frac{\partial h^u}{\partial w} \cdot dw + \frac{\partial h^u}{\partial y} \cdot dy
\end{aligned}$$

Plugging in $(\star)$,

$$\begin{aligned}
dh^u &= (\frac{\partial h^c}{\partial w} + \frac{\partial h^u}{\partial y} \cdot h^c)(-tw) + \frac{\partial h^u}{\partial y} \cdot dy \\
&= -tw\frac{\partial h^c}{\partial w} + \frac{\partial h^u}{\partial y}(-twh^c + dG) \\
\implies \frac{dh^u}{h} &= (-t) \cdot \frac{w}{h}\frac{\partial h^c}{\partial w} + \frac{\partial h^u}{\partial y}\frac{y}{h}(\frac{G - twh}{y})
\end{aligned}$$

Where $h = h^u = h^c$.

$$\implies \frac{dh^u}{h} = \epsilon_{h^c,w}(-t) + \epsilon_{h^u,y} \cdot \frac{S}{y}$$

Where $\epsilon_{h^c,w} > 0$, $-t < 0$, $\epsilon_{h^u,y} < 0$, $\frac{S}{y} > 0$. Thus, the substitution effect $\epsilon_{h^c,w}(-t)$ of welfare and income affect $\epsilon_{h^u,y} \cdot \frac{S}{y}$ of welfare are both negative, and unambiguously reduce labor supply.

2.

2.

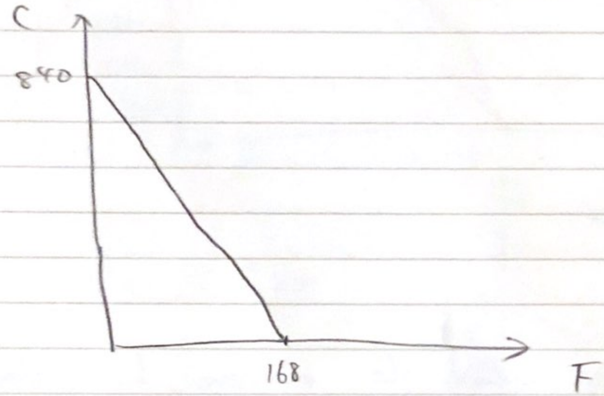
a).

The budget constraint is

$$C = w(168 - F)$$

with $w = \$5$, it is

$$C = 840 - 5F$$



Plugging this expression into her utility,

$$U(C(F), F) = \frac{1}{4} \ln(840 - 5F) + \frac{3}{4} \ln F$$

FOC [F] $$-\frac{5}{3360 - 20F} + \frac{3}{4F} = 0$$

$$\Rightarrow F^* = 126$$

SOC :

$$-\frac{3}{4F^2} - \frac{100}{(3360 - 20x)^2} < 0.$$

Thus, Ann chooses to work $168 - 126 = \underline{42 \text{ hours}}$.

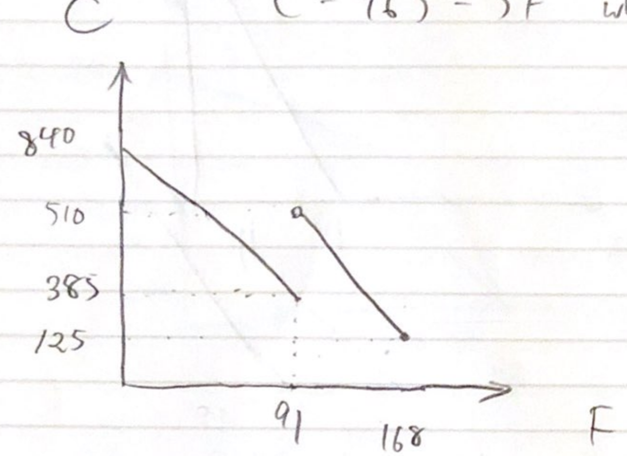
The household's total weekly income is $42 \times \$5 = \underline{\$210}$

2.

b) The new budget constraint is

$$C = 840 - 5F \quad \text{when } C \geq 385, \text{ or } F \leq 91$$

$$C = 965 - 5F \quad \text{when } C < 385 \text{ or } F > 91$$



Ann will join as even without the program, her optimal choice for leisure was 126 hours, but under the program, this would improve her well being without any trade-off.

Her maximization problem is.

$$\max_F \mathcal{L} = \frac{1}{4}\ln(965-5F) + \frac{3}{4}\ln F + \mu[F-91]$$

[F]

$$\frac{-5}{3860-20F} + \frac{3}{4F} + \mu F = 0$$

$$\mu(F-91) = 0$$

Kuhn-Tucker : $F > 91, \ \mu = 0$

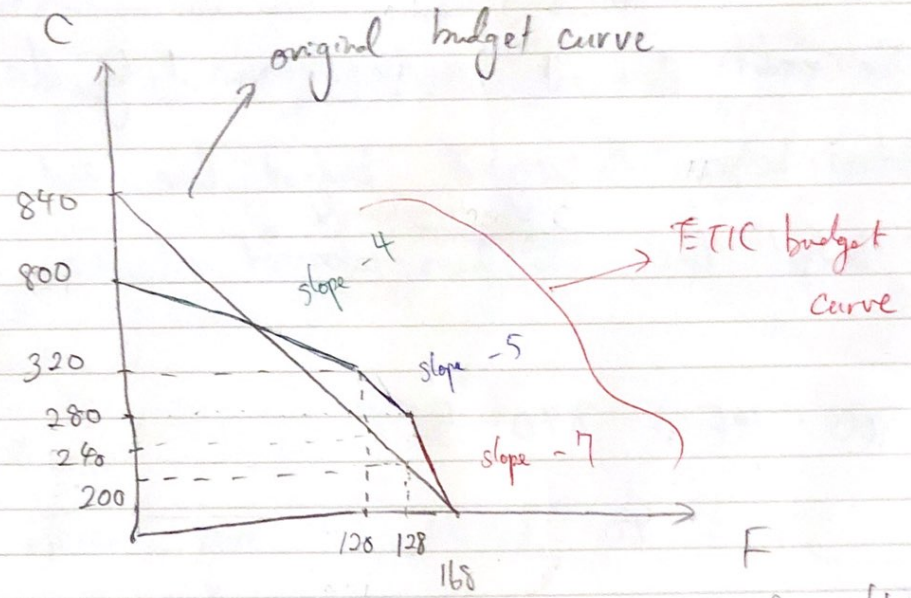
or $F = 91, \ \mu > 0$

In this case, the optimal solution is when $\mu = 0$.

$F^* = 144.75$ she will work $168 - 144.75$ $= 23.25$ hours and and her weekly income will be \$241.25.

2.

c)



$S = G + w(1-t)h$

When $C < 200$:

$$C = 5 \times (1+0.4) \times (168 - F)$$
$$= 1176 - 7F$$

$= G + 5(1-0.2)(168-F)$

When $200 \le C < 240$:

$$C = 5(128 - F) + 280$$

$$= 920 - 5F$$

When $C \geq 240$:

$$C = 5 \times (1 - 0.2) \times (120 - F) + 320$$

$$= 800 - 4F$$

The maximum credit is the maximum distance between the old budget constraint without subsidies and the new one with the EITC. Since credit only accumulates until $C = 200$, and increases with earned income, max credit is the level of tax credit at $C = 200$, or

$200 \times 0.4 = 80$.

2.

c) The credit phase out can be determined by the intersection between the original budget line and the EITC budget line. The lines intersect at

$$800 - 4F = 840 - 5F$$

$$F = 40.$$

At this point, the phase out tax completely erases the tax credit accumulated prior, leaving Ann in the same state she would be without any subsidy program.

d) We solve the optimization problem Ann faces for each portion of the curve and choose the one that maximizes utility.

$$\max_F \mathcal{L} = \frac{1}{4}\ln(1176 - 7F) + \frac{3}{4}\ln F + \mu[F - 128]$$

$$[F] \quad \frac{-7}{4704 - 28F} + \frac{3}{4F} + \mu F = 0$$

$$\mu(F - 128) = 0.$$

Assuming $\mu = 0$, $F = 126$. However, this violates the Kuhn-Tucker inequality, so $\mu > 0$ and $\underline{F = 128}$.

d)

$$\max_F \quad \frac{1}{4}\ln(920-5F) + \frac{3}{4}\ln F \quad \text{s.t.}$$

$$120 \leq F \leq 128.$$

$$[F] \quad \frac{-5}{3680-20F} + \frac{3}{4F} = 0$$

$$\Rightarrow F = 150$$

However, since $120 \leq F \leq 128$,

$$\underline{F^* = 128}$$

$$\max_F \mathcal{L} = \frac{1}{4}\ln(800-4F) + \frac{3}{4}\ln F + \mu[F-120]$$

$$\text{FOC} \quad [F] \quad \frac{-1}{800-4F} + \frac{3}{4F} + \mu F = 0$$

$$\mu(F-120) = 0.$$

Assuming $\mu = 0$, $F = 144$

However, this violates the inequality constraint so the constraint binds and $F^* = 120$.

Thus we have the two solutions $F = 120, 128$.

Utility is higher when $F = 128$.

So, Ann will work $168 - 128 = 40$ hours

and earn income $C = 920 - 5 \times 128 = \underline{\$280}$

e). In summary:

Baseline:	Work	42 hours
	Earnings	$\times 5$ \$210
	Transfers	\$0
	Income	\$210
Cash Welfare:	Work	23.25 hrs.
	Earnings	\$116.25
	Transfers	\$125
	Incomes	\$241.25
EITC:	Work	40 hrs
	Earnings	\$200
	Transfers	\$80
	Incomes	\$280

We see that Cash welfare discourages Ann from working and leads to the shortest number of hours worked, which is consistent with theory. We also know that EITC is most effective in incentivizing work while providing benefits, and compared to the base case, it hardly lowers hours worked. It is also cheaper compared to cash welfare.

3.

Pset 2

(3).

a. $U_{1/2} = \ln X_{1/2} + \frac{\ln(1-L_{1/2})}{2}$

$U_1 = \ln X_1 + \ln(1-L_1)$

$U_2 = \ln X_2 + 2\ln(1-L_2)$

$p = 1 \quad w = 1, \quad y = 1/4$

The budget constraint of the workers is given as:

$$X_i = \frac{1}{4} + L_i$$

Substituting the above expression into each worker's utility function and taking the FOC yields

$i = 1/2 \quad \ln\left(\frac{1}{4} + L_{1/2}\right) + \frac{\ln(1-L_{1/2})}{2}$

$$\text{FOC} \quad [L_{1/2}] \quad \frac{1}{\frac{1}{4}+L_{1/2}} - \frac{1}{2(1-L)} = 0$$

$$2(1-L) = \frac{1}{4} + L$$

$$3L = \frac{7}{4}$$

$$L^*_{1/2} = \frac{7}{12}$$

$i = 1$

$\ln\left(\frac{1}{4}+L_1\right) + \ln(1-L_1)$

$$\text{FOC} \quad [L_1] \quad \frac{1}{\frac{1}{4}+L} - \frac{1}{1-L} = 0$$

$$(1-L) = \frac{1}{4}+L$$

$$\Rightarrow \quad L^*_1 = \frac{3}{8}$$

(3)

a. $i = 2$

$$\ln\left(\frac{1}{4} + L_2\right) + 2\ln(1 - L_2)$$

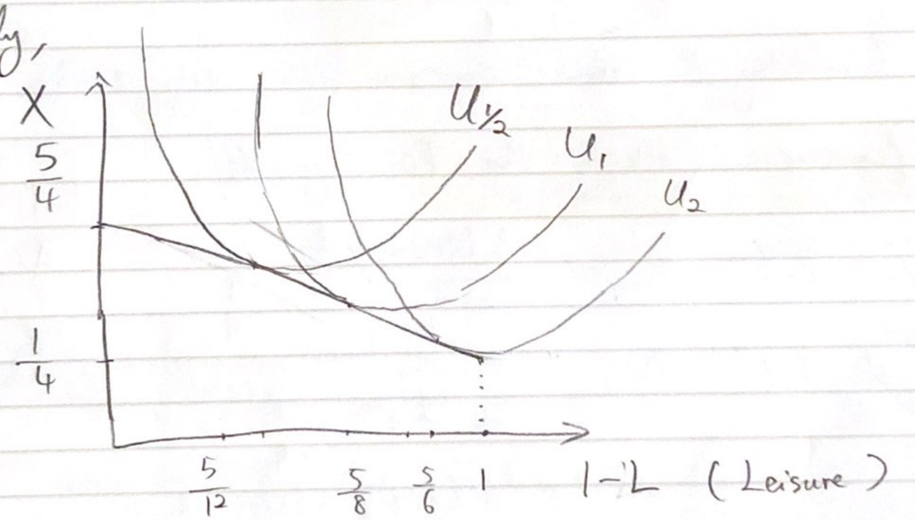
$$\text{FOC } [L_2] \quad \frac{1}{\frac{1}{4} + L_2} - \frac{2}{1 - L_2} = 0$$

$$\Rightarrow 1 - L_2 = \frac{1}{2} + 2L_2$$

$$3L_2 = \frac{1}{2}$$

$$L_2^* = \frac{1}{6}$$

Graphically,



Because U_2 values leisure time the most, their indifference curve is shifted to the right and they choose to work less. The value of i indicates each worker's subjective appreciation for leisure.

(3)

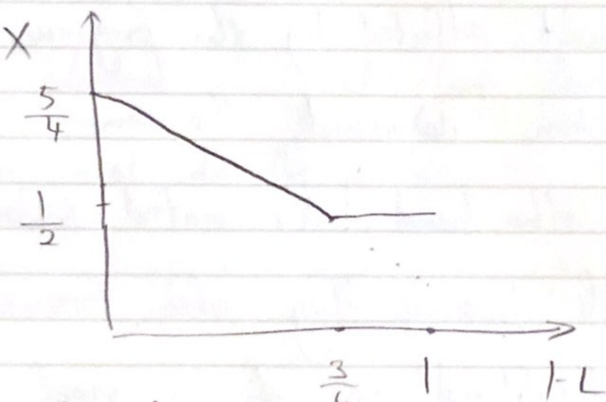
b. The new budget constraint is

$$X_i = \frac{1}{4} + L_i + \overbrace{\left[\frac{1}{4} - L_i\right]}^{S}$$

$$= \frac{1}{2} \quad \text{while } L_i \le \frac{1}{4}$$

$$X_i = \frac{1}{4} + L_i \quad \text{when } L_i \ge \frac{1}{4}.$$

This new constraint is mapped as



Notice that the budget constraint for worker i when $L_i \ge \frac{1}{4}$ is exactly the same as before.

Thus, the optimal response for each worker is

$$\max\left\{ U_i\left(\tfrac{1}{2}, 0\right),\ U_i\left(\tfrac{1}{4} + L_i^*, L_i^*\right) \right\}$$

where $U_i(X_i, L_i^*) = \ln X_i + i \ln(1 - L_i^*)$

for $i = \frac{1}{2}$: $U_{1/2}(\frac{1}{2}, 0) = -0.69 < -0.62 = U_{1/2}\left(\underset{\frac14 + L_{1/2}^*}{\frac{5}{6}}, \underset{L_{1/2}^*}{\frac{7}{12}}\right)$

So, under this new system, $i = \frac{1}{2}$ will choose $L_{1/2} = 7/12$

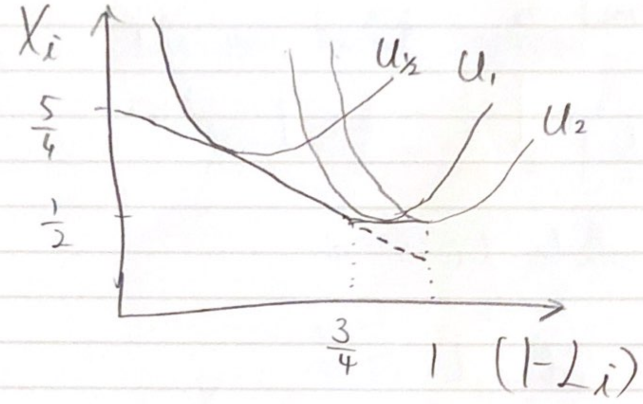
for $i = 1$: $U_1(\frac{1}{2}, 0) = -0.69 > -0.94 = U_1\left(\frac{5}{8}, \frac{3}{8}\right)$.

So $L_1 = 0$.

for $i = 2$: $U_2(\frac{1}{2}, 0) = -0.69 > -1.24 = U_2\left(\frac{5}{12}, \frac{1}{6}\right)$

So, $L_2 = 0$

b) Graphically, this result is explained as follows.



$U_{1/2}$ was already located towards the left, away from the region of the budget constraint affected by the program. Thus, its tangency condition remained the same. U_1 and U_2, on the other hand, are located towards the right of the graph because they value leisure more. As such, the heightened area of the graph where $L_i \leq \frac{1}{4}$ leads to tangency/intersection with a a higher utility indifference curve.

(3)

(c).

The new budget constraint now is:

$$X_i = \frac{1}{4} + L_i + \left[\frac{1}{4} - \frac{1}{2}L_i\right]$$

$$= \frac{1}{2} + \frac{1}{2}L_i \qquad \text{when } L_i \leq \frac{1}{2}$$

$$X_i = \frac{1}{4} + L_i \qquad \text{when } L_i > \frac{1}{2}$$

When $L_i > 1/2$, the optimization problem that the workers face is identical to what they faced in (a). When $L_i \leq \frac{1}{2}$, they must solve the problem

$$\max_{L_i} \ \ln\left(\frac{1}{2} + \frac{1}{2}L_i\right) + i \ln(1 - L_i) \quad \text{s.t. } L_i \leq \frac{1}{2}$$

For $i = 1/2$

$$\text{FOC } [L_{1/2}] = \frac{1}{2\left(\frac{1}{2} + \frac{1}{2}L_{1/2}\right)} - \frac{1}{2}\frac{1}{(1-L_i)} = 0$$

Solving this we get $L_{1/2} = \frac{1}{3} < 1/2$

$$L_{1/2}^{w} = 1/3$$

$$X_{1/2}^{w} = \frac{1}{2} + \frac{1}{2} \times \frac{1}{3} = \frac{2}{3}$$

$$U_{1/2}\left(\frac{2}{3}, \frac{1}{3}\right) = -0.61 > -0.62 = U_{1/2}\left(\frac{5}{6}, \frac{7}{12}\right)$$

Thus, $L_{1/2} = \frac{1}{3}$

For $i=1$,

$$FOC\ [L_1] = \frac{1}{2(\frac{1}{2}+\frac{1}{2}L)} - \frac{1}{(1-L)} = 0$$

$$\Rightarrow L_1^w = 0 \leq \frac{1}{2}.$$

$$X_1^w = \frac{1}{2}$$

$$U_1(\tfrac{1}{2}, 0) = -0.69 > -0.94 = U_1(\tfrac{5}{8}, \tfrac{3}{8})$$

Thus, $L_1 = 0$

For $i=2$,

$$FOC\ [L_2] = \frac{1}{2(\frac{1}{2}+\frac{1}{2}L)} - \frac{2}{(1-L)} = 0$$

$$\Rightarrow L_2 = -\frac{1}{3} < 0$$

$$\Rightarrow L_2 = 0.$$

$$U(\tfrac{1}{2}, 0) = -0.69 > -1.24 = U_2(\tfrac{5}{12}, \tfrac{1}{6}).$$

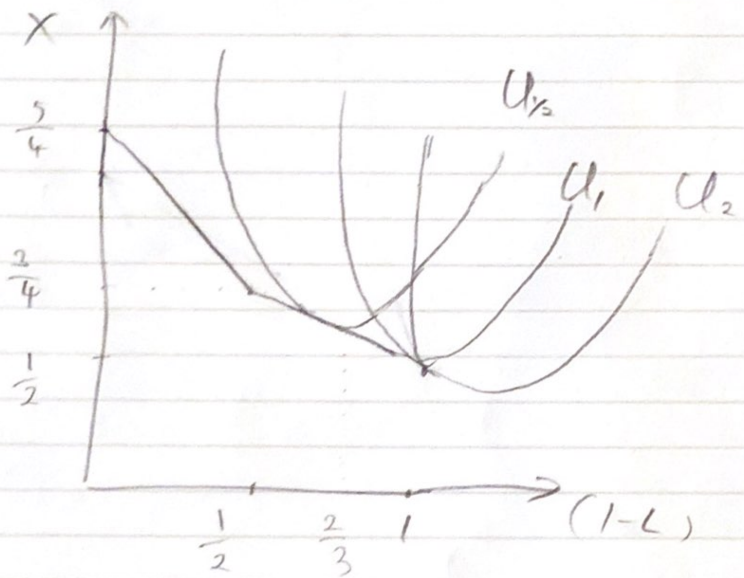
Thus, $L_2 = 0$

Therefore, this policy change did not change the labor supply of $i=1,2$, only $i=\frac{1}{2}$. Total government outlays actually increase, since in (b). $i=\frac{1}{2}$ was not on welfare, but now they are.

(B).

(C).

Graphically, we can map the budget constraint.



The altered area of the budget curve did not change the end point, at $X = \frac{1}{2}$. Since U_1 and U_2 are located to the right, their extremums are also towards the right, and so unless this end point shifts up, their optimal point, given by their highest utility indifference curve that intersects with the budget, remains the same. This reflects how highly workers 1 and 2 value leisure. Worker ½, on the other hand, has $U_{1/2}$ on the region of the graph that shifted up, and so their optimal point changed. Practically speaking, this shows how the welfare program with a ½ tax on earnings encourages participation but discourages labor.

4.

4.

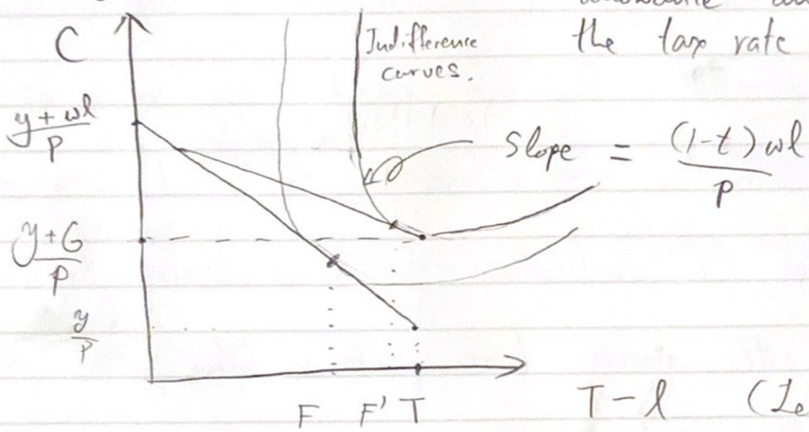
(a) The effects of a decrease in the wage rate towards labor supply are inconclusive. The income effect of a decreased wage rate would induce more labor, as the mother tries to make up for lost income. However, the substitution effect of a decrease in the wage rate makes time spent on labor less valuable and thus reduces the number of hours worked. For someone who is on the margin of working or not working it can be assumed that they are almost indifferent to the amount of labor-income they are making right now and the amount of income they will have once they quit work. Thus, if the wage rate decreases, the substitution effect of labor depreciating in value relative to leisure will most likely trump the effects of an income effect inducing more labor.

Anderson and Levine (2000) show empirical support for this.

4.

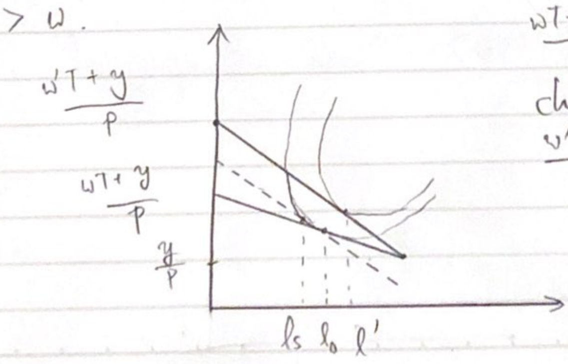
(b).

Budget Constraint. Let G be the lump sum allowance and t be the tax rate of the phase out.



As we have seen in problem 1, this form of subsidy which is analogous to a traditional welfare program unambiguously reduces labor supply, as can be seen by comparing $F < F'$.

c). Publicly provided child care for all yields the opposite effect of a), namely, an increase in the wage rate. Thus, the income effect demotivates labor while the substitution effect induces more labor. The effects can be decomposed graphically: $w' > w$.



$\frac{wT+y}{P}$ is the budget curve without child care for all.
$\frac{w'T+y}{P}$ is the budget curve with child care for all.
l' is labor under child care for all, and $l_0 - l_s$ is the magnitude of the substitution effect.

4.

(c). The effect of child care for all on labor participation & supply is ambiguous and depends on the preferences of each individual. Thus, when compared with phased out lump sum allowance, which almost definitely reduces labor supply, it is less likely to reduce labor supply.

5. Data exercise

In a few weeks, we will discuss disability insurance in detail. This data exercise is based on the paper The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust (Black, Daniel, and Sanders 2002), which you should read before starting this question.

```
library(tidyr)
library(dplyr)
library(data.table)
library(Hmisc)
library(ivpack)
library(stargazer)
library(ggplot2)
```

a.

What is the causal relationship of interest in this paper? Write down the structural equation. Note: this equation does not appear in the paper.

The paper estimates the marginal returns to medical care by investigating the relationship between costs of medical inputs and child mortality rates.

$$\Delta y = \alpha_0 + x\alpha_1 + \alpha_2 \Delta(medical\, inputs) + \epsilon$$

Where y is child mortality rates.

b.

What is the problem with estimating this equation in cross-sectional data? In what direction is the estimate likely to be biased?

Cross-sectional studies estimate returns to additional, incremental spending that occur in some areas but not others, and on average find similar health outcomes across section. However, they fail to account for the endogenous correlation of the health risk of a population and medical expenditure for that population. Thus, given the similarity in health outcomes across sections, this type of study finds trouble identifying a significant correlation between medical spending and improvement of health outcomes. This likely biases the estimates downward, towards the conclusion that the effect of medical spending on mortality is null.

c.

Fuzzy RD is just a form of IV. What is the instrument in this context? Explain how the instrument allows you to get an unbiased estimate of the causal relationship of interest.

The instrument is the VLBW indicator. Assuming continuity of other relevant factors contributing to an infant's mortality, infants that are just below the threshold of $1500g$ receive higher medical inputs (summarized in medical expenditure) than infants just above due to medical customs and guidelines. Thus, this instrument allows us to isolate medical spending and overall medical inputs from the individual infant's risk level, and show the effect that medical spending has on infant mortality while controlling for health.

d.

Write down the "first stage" regression discontinuity equation and the "reduced form" regression discontinuity equation. (These do appear in the paper.) Explain each term of the equations and the coefficients of interest.

$$Y_i = \alpha_0 + \alpha_1 VLBW_i + \alpha_2 VLBW_i \times (g_i - 1500) + \alpha_3(1 - VLBW_i) \times (g_i - 1500) + \alpha_t + \alpha_s + \delta X_i' + \epsilon_i$$

The first stage is when Y_i denotes costs and the reduced form is when Y_i denotes the infant one-year mortality rate. $VLBW_i$ is a dummy variable that indicates whether an infant is classified as $VLBW$ ($< 1500g$), and its coefficient α_1 indicates the degree to which being $VLBW$ affects mortality. The second and third terms are gram trend terms below and above the threshold, parameterized so that their coefficients are equal when the trends are the same. α_t and α_s are indicators for the year of birth t and state of birth s. X_i' are the controls used that deal with other newborn characteristics not specified in the rest of the model, and ϵ_i is the error term.

e.

Name one or more key threats to the validity of the empirical strategy used by Almond et al. (2010). Be specific to their context. Describe what tests you would perform to assess these threats.

One threat is that the summary measures utilized for medical inputs may not be exhaustive, and that unobserved factors that reduce mortality may also contribute to the discontinuity in mortality around the 1500 g threshold. For example, since the 1500 g. cut-off is well known in the medical community, infants born under 1500 g. may be paid especially close attention by physicians and medical staff, and treated with more care than those who clear the 1500 g. line. This is outlined in Angert and Adam (2009). Many of these behavioral changes in the medical staff may not be picked up by the summary measure (medical costs), and result in a slight upward bias. This threat can be analyzed by obtaining records of infants just below and above the 1500 g. threshold with identical treatment/medical input records and analyze for a notable difference in their mortality.
Moreover, using the summary statistic of medical cost may run the risk of overlooking price changes corresponding with weight, but as the authors suggest in the paper, this threat can be tested by analyzing for discontinuity in pricing across the VLBW threshold, among other methods.

f.

Download the data set from: http://data.nber.org/lbid/adkw.dta. Import the data set to your preferred statistical analysis program (e.g., Stata, R). By downloading the data set, you agree to the following NCHS data rules: 1) Use the data in this dataset for statistical reporting and analysis only. 2) Make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery. 3) Not link this dataset with individually identifiable data from other NCHS or non- NCHS datasets.

i.

i. Replicate Figure 2-A, which gives the main finding of the paper.

A. *Group births into 1-ounce (28.3495-gram) bins, radiating outward from the 1500 gram threshold.*

```
# loading in data
dat <- read.csv("DataExercise_RD/adkw.csv")
```

```
# creating bin categories
ub <- max(dat$dbirwt)
lb <- min(dat$dbirwt)
s1 <- seq(from = 1500, to = ub, by = 28.3495)
s2 <- rev(seq(from = 1500, to = lb, by = -28.3945))
s2 <- head(s2, -1)
bins <- c(s2, s1)
bins <- c(-Inf, bins, Inf)
bin_groups <- seq(1:(length(bins) - 1))
bins; bin_groups
```

```
## [1]     -Inf 1443.211 1471.605 1500.000 1528.350 1556.699      Inf
```

```
## [1] 1 2 3 4 5 6
```

```r
# cutting data into bins
dat$bwt_bins <- cut(dat$dbirwt, breaks = bins, labels = bin_groups)
```

B. *For each bin, plot the bin's mean one year mortality against the bin's median birthweight.*

```r
gdat <- dat %>%
  group_by(bwt_bins) %>%
  mutate(bin_medianwt = median(dbirwt, na.rm = TRUE)) %>%  # median bwt
  mutate(bin_meandeath = mean(death1year, na.rm = TRUE)) %>%  # mean mortality
  mutate(bin_meanage = mean(gestat, na.rm = TRUE)) # mean gestational age
```

```r
ggplot(gdat, aes(x = bin_medianwt, y = bin_meandeath)) +
  geom_point(pch = 16) +
  xlim(1400, 1600) +
  geom_vline(xintercept = 1500) +
  labs(x = "Birth weight (g)", y = "Mean 1 Year Mortality") +
  ggtitle("One Year Mortality")
```

C. *Interpret the graph you created, which is the main finding of the paper.*

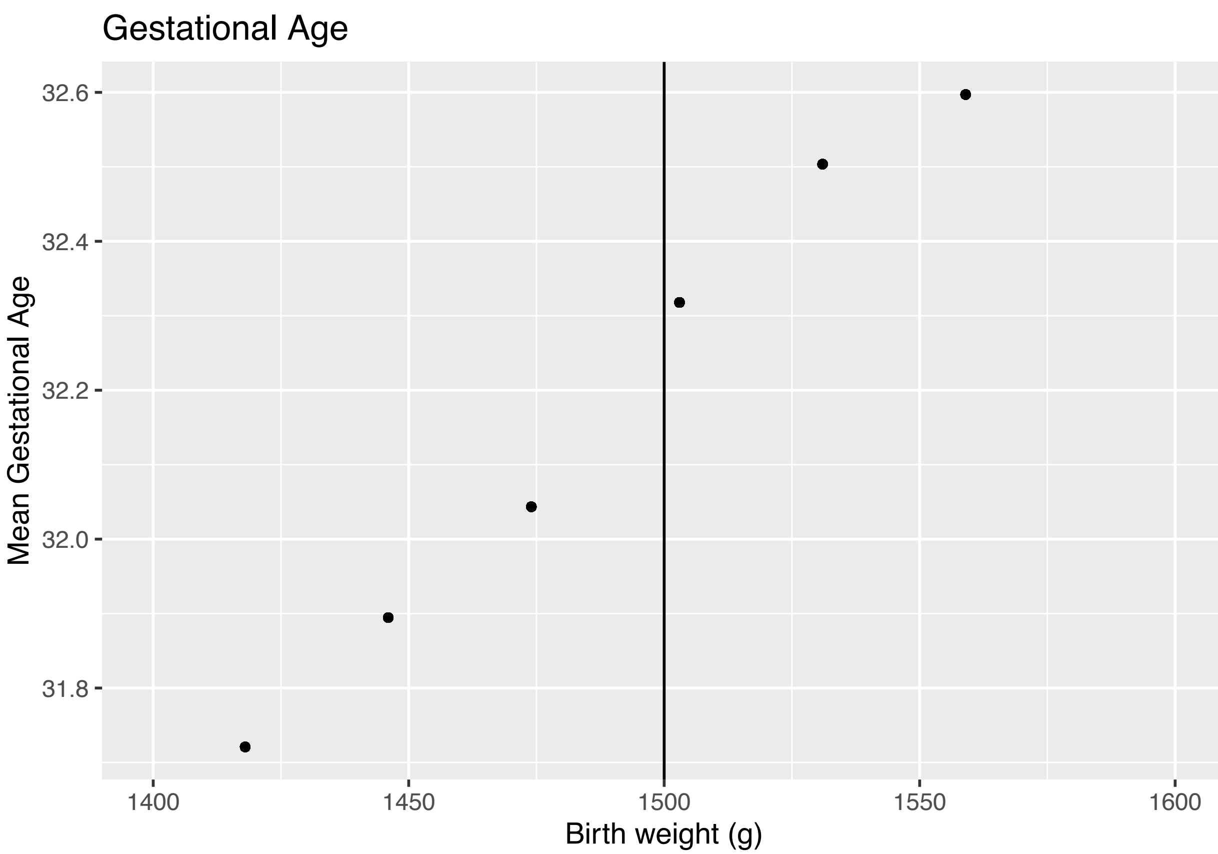
Due to the limitations of our data, we are unable to observe the same discontinuity that was observed in the paper. Mean mortality seems to be continuous around the 1500g. line, and if there is a discontinuity it seems to be between the bins with 1443.211-1471.605 g. and 1471.605-1500.000 g. Thus, replicating the analyses in the paper on this data set may not be appropriate.

ii.

Test for differences in observable covariates across the 1500 gram threshold (replicate Figure 5-A from the paper).

A. *Adapt the method used in (i) to replicate Figure 5-A (the covariate of interest is gestational age).*

```
ggplot(gdat, aes(x = bin_medianwt, y = bin_meanage)) +
  geom_point(pch = 16) +
  geom_vline(xintercept = 1500) +
  xlim(1400, 1600) +
  labs(x = "Birth weight (g)", y = "Mean Gestational Age") +
  ggtitle("Gestational Age")
```



B.

Which RD identifying assumption does this figure support? Explain.
This graph shows continuity of other covariates that likely contribute to child mortality around the 1500 g. line. Continuity in the covariates allows us to assume that the only factor contributing to the discontinuity observed in the original data is the binary indicator of whether a child crosses the 1500 g. threshold or not, which acts as a proxy for quantity of medical inputs received by the infant. Without this assumption, there may be underlying factors contributing to the discontinuity and our research design will not be able to isolate and pick up the effects of medical inputs on child mortality.

iii.

Above, you wrote down the reduced form equation that appears in the paper.

A. *Construct a dummy variable $VLBW_i$ that equals one if the infant is classified as very low birthweight (i.e., birthweight strictly less than 1500 grams), and 0 otherwise.*

```
# creating dummy
dat <- dat %>%
  mutate(VLBW = ifelse(dbirwt < 1500, 1, 0),
         trend.down = VLBW * (dbirwt - 1500),
         trend.up = (1 - VLBW) * (dbirwt - 1500))
```

B. *Ignore the terms α_t, α_s and δX_i in the reduced form equation. Estimate the reduced form equation by OLS. Obtain heteroskedastic-robust standard errors.*

```
# OLS estimate
reduced.ols <- lm(death1year ~ VLBW + trend.down + trend.up, data = dat)
```

```
# Heteroskedastic robust standard errors
hkse <- function(reg){robust.se(reg)[,2]}
ols_error <- hkse(reduced.ols)
```

```
## [1] "Robust Standard Errors"
```

C. *Construct a table that contains the estimated coefficients $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ from the reduced form equation. Report standard errors in parentheses below each estimate. In your table, indicate whether each estimated coefficient is significant at a 1%, 5% or 10% level.*

```
stargazer(reduced.ols, type="latex", title = "OLS Results",
          se = list(ols_error), header = FALSE, no.space=TRUE)
```

Table 1: OLS Results

	Dependent variable:
	death1year
VLBW	-0.010^{***}
	(0.002)
trend.down	-0.0001^{***}
	(0.00003)
trend.up	-0.0002^{***}
	(0.00003)
Constant	0.063^{***}
	(0.001)
Observations	202,071
R^2	0.001
Adjusted R^2	0.0005
Residual Std. Error	0.233 (df = 202067)
F Statistic	33.752^{***} (df = 3; 202067)
Note:	*p<0.1; **p<0.05; ***p<0.01