Labour Economics

Empirical Issues

Endogeneity Example

Consider the regression equation

$$\log(w_i) = X_i\beta_1 + Z_i\beta_2 + u_i$$

- X_i are worker characteristics that affect productivity
- Z_i are non-wage job characteristics that reduce the desirability of the job

The vector of coefficients β_2 should be positive – less desirable jobs should induce higher wages. However, if any non-included job characteristics N_i (which would have positive coefficient β_O) are positively correlated with elements of Z_i, this will result in an upward bias in the estimates of β_2.

$$\text{plim}(\hat{\beta}_2) \rightarrow \frac{\text{cov}(Z,W)}{\text{var}(Z)} = \beta_2 + \beta_O \frac{\text{cov}(N,Z)}{\text{var}(Z)} = \beta_2 + [\text{POS BIAS}]$$

Consider another omitted variable, a missing skill characteristic P_i that is positively correlated with X_i. If more highly productive workers 'buy' better work conditions (meaning that there is a negative correlation between P_i and Z_i), then

$$\text{plim}(\hat{\beta}_2) \rightarrow \frac{\text{cov}(Z,W)}{\text{var}(Z)} = \beta_2 + \beta_O \frac{\text{cov}(P,Z)}{\text{var}(Z)} = \beta_2 + [\text{NEG BIAS}]$$

Could overcome these worker characteristic measurement problems by running a fixed-effects regression, however it would be necessary to ensure that job changes were independent of skills and job conditions, otherwise endogeneity would still be a problem.

Ability bias is a particularly common example of endogeneity. It occurs when:

$$\text{Cov}(\epsilon, \text{EDUC}) \neq 0$$

Consider the following Mincer equation

$$\log(w_i) = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \epsilon$$

If we consider ability A_i to be an omitted variable that should have positive coefficient β_O then, because we expect ability and education to be positively correlated, this yields positive bias:

$$\text{plim}(\hat{\beta}_1) \rightarrow \frac{\text{cov}(S,W)}{\text{var}(S)} = \beta_1 + \beta_O \frac{\text{cov}(A,S)}{\text{var}(S)} = \beta_2 + [\text{POS BIAS}]$$

Classical Measurement Error

When dependent variable x is measured with error, the regression suffers from attenuation bias, meaning that the estimated regressors are biased toward zero.

$$x = x^* + u, \quad u \sim N(0, \sigma_u^2)$$

$$y = \beta x^* + e, \quad e \sim N(0, \sigma_e^2)$$
$$y = \beta(x - u) + e$$

$$y = \beta x - \beta u + e$$
$$y = \beta x + w$$

$$\text{plim}(\hat{\beta}) \to \frac{\text{cov}(x,y)}{\text{var}(x)} = \frac{\beta\text{var}(x^*)}{\text{var}(x^*) + \text{var}(u)} = \beta\left[\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_u^2}\right]$$
$$\lim_{\sigma_u^2 \to \infty}\left(\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_u^2}\right) = 0$$

Difference in Difference Estimates

The technique often employs some particular policy change that affects one group of individuals (the treatment group) but not a second group (a control group). The estimate of the effect of the policy change on outcomes is then constructed by comparing the change in outcomes for the treated group to the change for the control group.

Consider some outcome y_{it} over i individuals and time span t. Let δ_{it} be an indicator variable equal to 1 if the individual was part of the treatment group and 0 otherwise. Let η_i be an individual specific effect and m_t be a time-varying term common to all individuals. This can be written as:

$$y_{it} = \gamma\delta_{it} + \eta_i + m_t + \epsilon_{it}$$

Taking first differences will eliminate the individual fixed effects:

$$\Delta y_{it} = \gamma\Delta\delta_{it} + \Delta m_t + \Delta\epsilon_{it}$$

The OLS for $\hat{\gamma}$ will thus be:

$$\gamma = \Delta\bar{y}_e - \Delta\bar{y}_c$$

Thus the estimated effect of the policy change is simply the change in mean outcomes for the treated group minus the change for the control group. The very important underlying assumptions of this simple model are:

- The time effects m_t must be the same for treatments and controls (can try to analyse this by looking at trends prior and after change)
- The composition of the treatment and control groups must remain the same before and after the policy change (shouldn't be a problem in balanced panels where the same individuals are tracked over time)

If we think that different individuals may respond differently to the policy change (heterogeneous responses), then the above estimator can be interpreted as estimating what is known as the "average effect of treatment on the treated". This may mean that the effect we estimate is specific to the group affected by the change, and is potentially not going to be the same for other groups.

Heterogeneity

Heterogeneity refers to differences across the units being studied (e.g. firms or individuals). This could lead to problems of heteroskedasticity, omitted variable bias, and inaccurate parameter values for particular groups (e.g. men and women, young and old workers, city or rural).

- Ignore the issue and then interpret the estimated β's as weighted averages of the heterogeneous population parameters
- Estimate separate regressions for particular demographic groups

Functional Form

- Do we allow for a backward bending labour supply curve?
- Interaction terms?
- Non-linear relationships: e.g. overtime wages, progressive taxes
- In the Mincer equation: cubics quartics on x and /or S?

Selection Bias

Sampling bias is systematic error due to a non-random sample of a population, causing some members of the population to be less likely to be included than others. If there are any systematic differences at all between the two population segments, this will mean that:

$$E(e_i|x_i) \neq E(e_i|x_i, y_i > 0)$$

For example, the coefficient of wages on labour supply will actually include two components: a potentially causal effect of the wage upon labour supply, and a compositional effect whereby changes in the wage alters the mix of people who choose to work. We may wish to disambiguate these effects.

Selection Models

Consider the labour supply equation where h_i is the number of hours worked and Q_i is a vector of control variables:

$$\log(h_i) = \alpha \log(w_i) + \beta Q_i + \epsilon_i$$

If we know the wage rates for both working and non-working individuals, we can use a probit model to estimate parameters α and β with MLE methods.

$$\begin{aligned}
\Pr(h_i > 0) &= \Pr(\epsilon_i > -\alpha\, log(w_i) - \beta Q_i) \\
&= \Pr\left(\frac{\epsilon_i}{\sigma_\epsilon} > -\frac{[\alpha\, log(w_i) + \beta Q_i]}{\sigma_\epsilon}\right) \\
&= \Pr\left(z_i > -\frac{[\alpha\, log(w_i) + \beta Q_i]}{\sigma_\epsilon}\right) \\
&= 1 - \Phi\left(-\frac{[\alpha\, log(w_i) + \beta Q_i]}{\sigma_\epsilon}\right) \\
\Pr(h_i > 0) &= \Phi\left(\frac{\alpha\, log(w_i) + \beta Q_i}{\sigma_\epsilon}\right)
\end{aligned}$$

This yields an likelihood function of:

$$L = \prod \Phi\left(\frac{\alpha\, log(w_i) + \beta Q_i}{\sigma_\epsilon}\right)^{I_i} \prod 1 - \Phi\left(\frac{\alpha\, log(w_i) + \beta Q_i}{\sigma_\epsilon}\right)^{1-I_i}$$

$$\log L = \sum I_i \log\left[\Phi\left(\frac{\alpha\, log(w_i) + \beta Q_i}{\sigma_\epsilon}\right)\right] + (1 - I_i)\log\left[1 - \Phi\left(\frac{\alpha\, log(w_i) + \beta Q_i}{\sigma_\epsilon}\right)\right]$$

We can use a very similar technique to solve for the actual probability that individuals work positive hours:

$$P(I_i = 1) = P(h_i > 0) = \Phi\left(\frac{\alpha\, log(w_i) + \beta Q_i}{\sigma_\epsilon}\right)$$

Using this we can calculate the inverse Mills ratio, which is simply the ratio of the pdf to the cdf:

$$\lambda_i = \frac{\phi\left(\frac{\alpha\, log(w_i) + \beta Q_i}{\sigma_\epsilon}\right)}{\Phi\left(\frac{\alpha\, log(w_i) + \beta Q_i}{\sigma_\epsilon}\right)}$$

We can then include these as an independent variable in the regression which includes wages and the vector Z_i of taste and other variables that effect hours worked:

$$h_i = \phi_0 + \phi_1 w_i + \phi_2 Z_i + \epsilon_i$$
$$h_i = \phi_0 + \phi_1 w_i + \phi_2 Z_i + \theta\lambda_i + \psi_i$$

It turns out that including the inverse mills ratio in the structural regression this way allows us to obtain unbiased estimators where $E(\psi_i) = 0$. Note that this requires knowledge of all workers wages w_i, which is seldom possible as no data will be available for non-participant wages.

Imputed Wages

We could just use the structural equation of observable characteristics of non-workers to impute a wage for non-workers using the worker coefficients (where X are worker productivity variables):

$$w_i = \beta_0 + \beta_1 X_i + \nu_i$$

However, OLS estimates of β_1 would be biased, as it only includes individuals whose wage exceeds the reservation wage. This would only produce unbiased estimates if unobserved variables like motivation and intelligence are the same for participants and non-participants, which is unlikely.

We can instead apply a similar technique to estimate an alternate version of the wage equation that adjusts for this selection bias, where ψ_i is an altered error term:

$$w_i = \beta_0 + \beta_1 X_i + \theta\widehat{\lambda_i} + \psi_i$$

The $\widehat{\lambda_i}$ term is calculated per the usual form, but this time using the results from a reduced form equation of labour supply. Substituting the structural hours equation into the structural wage equation we have:

$$\begin{aligned}
h_i &= \phi_0 + \phi_1 w_i + \phi_2 Z_i + \epsilon_i \\
&= \phi_0 + \phi_1(\beta_0 + \beta_1 X_i + \nu_i) + \phi_2 Z_i + \epsilon_i \\
&= \phi_0 + \phi_1\beta_0 + \phi_1\beta_1 X_i + \phi_1\nu_i + \phi_2 Z_i + \epsilon_i \\
&= (\phi_0 + \phi_1\beta_0) + \phi_1\beta_1 X_i + \phi_2 Z_i + (\phi_1\nu_i + \epsilon_i) \\
h_i &= \alpha_0 + \alpha_1 X_i + \alpha_2 Z_i + \eta_i
\end{aligned}$$

Hence for the inverse Mills ratio we have:

$$\widehat{\lambda_i} = \frac{\phi(-\alpha_0 + \alpha_1 X_i + \alpha_2 Z_i)}{\Phi(-\alpha_0 + \alpha_1 X_i + \alpha_2 Z_i)}$$

- Use probit regression to find unbiased estimates of structural parameters $\alpha_0, \alpha_1, \alpha_2$
- Use these results to calculate an estimate of the inverse Mills ratio $\widehat{\lambda_i}$ for each worker
- These values in the wage equation to find unbiased estimates of β_0, β_1
- Use these parameters to calculate imputed wages for non-workers
- Run the regression again using both actual and imputed data

Instrumental Variables

Suppose we have the following equations:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$
$$x_i = a + z_i + v_i$$

If $Cov(v, \epsilon) \neq 0$ then OLS estimates of y will be biased. Instead we can use z as an instrument:

$$y_i = \beta_0 + \beta_1(a_i + z_i + v_i) + \epsilon_i$$
$$y_i = \beta_0 + \beta_1 a + \beta_1 z_i + \beta_1 v_i + \epsilon_i$$
$$y_i = \gamma_0 + \gamma_1 z_i + \phi_i$$

The new error term ϕ should now be uncorrelated with z. Typically we don't actually substitute in like this; we first regress x on z to get the predicted values of x, then sub these into the original equation and run the regression as normal (2SLS). Either way we now have:

$$\text{plim } \beta_{OLS} = \beta + \frac{\text{Cov}(x, \epsilon)}{\text{Var}(x)}$$
$$\text{plim } \beta_{IV} = \beta + \frac{\text{Cov}(z, \epsilon)}{\text{Cov}(z, x)}$$

If $\text{Cov}(z, \epsilon) = 0$, then the IV estimator will be unbiased. Unfortunately we cannot test for this (valid instruments); we can only test for the size of $\text{Cov}(z, x)$ (weak instruments). But if there is some remaining correlation between our instrument z and ε, combined with only a weak relationship between z and x, then IV could yield more biased estimates than OLS.

IV eliminates measurement error attenuation bias, so long as the IV itself is measured accurately.

Interpretation of IV estimates is slightly different, as the coefficients now represent the marginal returns to individuals affected by the particular instrument employed. Depending on the instrument used, these may not be representative of the sample or population as a whole. This is called the Local Average Treatment Effect. For example, if we used weather conditions as an instrument for a regression of demand for wheat against price (weather should correlate with price through supply but should otherwise be independent of demand), what we actually estimate will be the effect on demand of a price increase caused by bad weather. This may or may not be similar to the price changes caused by other factors.

As another example, if the schooling reform leads to a bigger change in schooling for those with the highest returns to schooling, then IV will provide an upward biased estimate of the average return to schooling. Formally, consider a case where the returns to education are heterogeneous β_i, and an intervention program Z is used as an instrument.

$$\Delta W_i = \beta_i \cdot \Delta S_i$$

$$\begin{aligned}\text{plim}(b_{IV}) &= \frac{\text{Cov}(W, Z)}{\text{Cov}(S, Z)} \\ &= \frac{E(\beta_i \cdot \Delta S_i)}{E(\Delta S_i)} \\ &= \frac{E(\beta_i) \cdot E(\Delta S_i | \beta_i)}{E(\Delta S_i)}\end{aligned}$$

If $E(\Delta S_i | \beta_i) = E(\Delta S_i)$ then this instrument will be unbiased.

Instrumental variables become even more complicated when we have multiple excluded variables (e.g. ability and return to education), as z must be uncorrelated with either of these variables.

Control Function

Under the control function approach, the second stage involves adding the residuals from the first stage of the 2SLS as an additional regressor to the original equation, perhaps interacted with other variables as well. If only the residuals are added (no interactions), then IV and the Control Function approach yield identical estimates. But the control function approach is more general by allowing the addition of interaction terms. Such additional terms allow the researcher to relax certain required assumptions i.e. full independence is no longer required.

Identification Problems

- It is very hard to identify both demand and supply equations (e.g. for labour demand and supply), as you need variables in the supply equation that are not in the demand equation and similarly for the supply equation
- In other words, two variables are not enough to map out simultaneous shifts in both the demand and the supply curves. We need to be sure one is not moving
- To overcome this problem you could try to find cases where wage changes are exogenous (e.g. minimum wages), or where the supply of workers is fixed
- Cross section variation in wages is also problematic because firms in a competitive market should be paying the same wage – if they aren't they are probably hiring difference
- There is also the problem of extensive versus intensive adjustment; total industry demand could fall as inefficient firms shut down, while remaining firms could hire more
- Estimates of labour supply using cross-sectional data (regressing hours of work on the current wage) confuses response of labour supply to wage changes of three types:
    - Movements along a given lifetime wage profile i.e. wages generally rise with labour market experience
    - Shifts in the wage profile
    - Changes in the slope of the wage profile

Twins and Siblings Studies

- If identical twins have equal abilities, using a within twin difference estimate of the effects of schooling on earnings should yield an unbiased estimate of the average returns to schooling.
- If twins are raised together, family background should also be identical. If not raised together, it is possible to identify the separate effects of ability and costs on schooling
- Measurement error yields more attenuation bias in within-twin difference estimates, as schooling is often highly correlated within twins, and thus taking the difference in schooling will remove some of the schooling signal relative to the noise
- Identical twins studies suggest that OLS estimates are upward-biased by 10%
- It is more difficult to argue that ability is equal among non-identical twin siblings. Thus if ability differences significantly drive schooling differences within families, sibling-based estimates may be more biased than OLS in the wider population
- If tastes drive schooling differences instead, then sibling estimates may be less biased than OLS in the wider population
- Siblings-based studies contain a small positive ability bias, but less than in standard OLS estimates

Labour Supply

Static Model

Initial assumptions:

- Individuals gain utility from consuming goods and leisure
- An individual's time is divided fully between working and leisure
- There is free choice of hours of work (perhaps by changing employers)
- The price of consumption goods p is assumed fixed at one, so changes in the hourly wage rate w represent real wage changes
- Individuals consume all their income each period (no saving over multiple periods)

Solving the following optimization problem:

$$\max_{C,L} U(C,L) \ \ \text{s.t.}\ C = w \cdot (T - L) + Y$$

We find that at the optimum the slope of the indifference curves equals the slope of the budget constraint. This is equivalent to saying that the ratio of marginal utilities is equal to the wage:

$$\frac{dC}{dL} = -w$$
$$\frac{U_L(C,L)}{U_C(C,L)} = w$$

Individuals may optimally choose $H = 0$ (i.e. to exit the labour force) in situations where they:

- have strong preferences for leisure over consumption
- have a large amount of unearned income Y
- are only offered a low wage w such that $w < w_r$

An increase in unearned income Y:

- Will raise the reservation wage if leisure is a normal good
- Can never increase labour force participation when leisure is a normal good

The effect of a wage increase on the labour force participation decision can only be positive or zero, as there is no income effect from a wage increase if you are currently not working, and the substitution effect is always positive. An individual labour supply curve may thus be backward-bending in the wage. Below the reservation wage, labour supply will be zero. As the wage moves above w_r, labour supply increases, but as the wage continues to rise and hours of work are higher, the income effect may come to dominate the substitution effect, resulting in labour supply falling.

Non-Linear Budget Constraints

There are many reasons for budget constraints to be non-linear:

- Overtime wage premiums
- Fixed costs and benefits of working
- Progressive income taxes
- Welfare payments with differing clawback rates (rental assistance, concession cards, medicare, low tax offsets, family tax benefits)
- Earned income tax credits (EITCs)

Calculating optimal solutions then involves optimizing over each linear piece of the budget set, and at each kink point.

Household Production

This can be simply modelled as an additional form of income: $C = C_M + C_D$. Let $f'(H_D)$ be the marginal product of household production. Optimization will occur when:

$$\frac{U_l(C, L)}{U_C(C, L)} = w = f'(H_D)$$

The effect of an increase in the wage rate w on market labour supply is likely to be more positive than in the basic model without home production. There is still an income effect (negative) which offsets the substitution effect (positive), but the substitution possibilities are greater with home production as the individual can substitute away from home production as well as from leisure if the wage rises. If initially working in the market, a wage increase will lead to less time in home production, as it becomes relatively less productive.

Family Decision Making

The simplest way of treating families is to assume that all income is pooled and joint utility is maximised, i.e. $\max U(C, L_1, L_2)$ s.t. $Y_1 + Y_2$. Unfortunately, numerous observations across economics seem to contradict this income pooling assumption. For example, there is evidence in many countries that children have more family resources spent on them when the mother has more control over the family's income.

A more general model is the collective model, in which utility maximization problems are solved separately for each household member, incorporating a 'sharing rule' which determines how much non-labour income each member receives. If all you observe are individual labour supply decisions of the two household members, plus joint household unearned income y (as is often the case in our survey data), it is possible to identify individual preferences and the sharing rule (up to an additive constant). The model does not, however, tell us how the sharing rule is specified.

Dynamic Model

Here we assume that individuals maximise utility over their lifetime. Often an additively separable utility function of the following form is assumed (where ρ is the discount rate, X_t is a vector of factors that affect utility and consumption choices, e.g. children, age):

$$\sum_{t=1}^{T} \frac{1}{(1+\rho)^{t-1}} U(C_t, L_t, X_t)$$

This optimization problem can be solved using standard Lagrangean techniques:

$$\frac{U_L(C_t, L_t, X_t)}{U_C(C_t, L_t, X_t)} = w_t$$

$$U_C(C_t, L_t) = \frac{1 + r_{t+1}}{1+\rho} U_C(C_{t+1}, L_{t+1})$$

If r is constant and equal to ρ, than the Euler equation shown above simplifies, and the individual simply consumes the same amount in each period. The solution to the optimisation problem is then that the individual chooses optimal consumption, leisure and savings so that the marginal utility of wealth in period t is equal to the discounted value of the marginal utility of wealth in period $t + 1$. In other words, the individual cannot be made better off by shifting consumption over periods.

Labour Supply Functions

- In ascending elasticity order:
- Marshallian labour supply functions hold unearned (or full) income constant
- Hicksian labour supply functions hold utility constant
- Frisch labour supply functions hold the marginal utility of wealth constant

Empirical Evidence

- Male labour supply responses to wage changes are that elasticities are small. Some estimates are actually small negatives, suggesting males may be at the backward bending part of the labour supply curve. But the perceived wisdom is that the elasticity is a small positives (+0.2).
- Female labour supply responses to wage changes are generally larger, but elasticities are still in the inelastic range (+0.4 to +0.5). Elasticities for women have also fallen over time
- Female working hours tend to fall with male partner income
- Participation falls with more children and younger children in the home
- Participation rises with education levels. Wages rise with education so opportunity cost of not working rises

Labour Demand

Marshall's Laws of Demand

The elasticity of labour demand will be high if:

- The elasticity of demand for the final product is high (scale effect)
- Other factors of production are easily substituted for labour (substitution effect)
- The supply of these other factors is elastic
- The share of labour's cost in total costs is large

Elasticity of Substitution

- Elasticity of substitution: percentage change in capital/labour ratio divided by the percentage change in the wage/interest rate ratio
- This is infinite if K and L are perfect substitutes, and zero if they are perfect complements
- The marginal productivity of each factor is equal to its real cost multiplied by the markup, at the optimum. If the product market is perfectly competitive, $v = 1$ and there is no markup

Wage Own Price Elasticities

- The reduction in the demand for labour from an increase in wage rates can be broken down into a substitution effect and a scale effect
- The *substitution* effect measures the effect of a change in an input price on the amount of inputs used to produce a *given* output level (say Q_a)
- The *scale* effect measures the effect of a change in output levels (the scale of operation) on the amount of inputs used, holding input prices constant
- Both the substitution and scale effects reduce the quantity of labour demanded when the wage rate increases; the long-run labour demand curve unambiguously slopes down
- Since there is no substitution effect possible in the short run (with a fixed capital stock), the short-run demand for labour will be steeper than the long-run demand for labour

Wage Cross Price Elasticities

- The conditional cross-price elasticity (substitution effect) is positive, as if capital prices rise, firms will substitute labour for capital when holding Y constant
- The scale effect has an ambiguous sign a priori, but can be shown to be negative if the production function is homogeneous; if either input price increases, costs for the firm rise, so the firm will optimally reduce output
- Since the substitution effect in the two-factor model is always positive, we know that capital and labour must be p-substitutes
- In contrast, the unconditional cross-price elasticity will have an ambiguous sign. The positive substitution effect and the generally negative scale effect offset each other
- If the unconditional cross-price elasticity is positive, the substitution effect dominates and capital and labour are gross substitutes

Trade-off Between Workers and Hours

Number of workers N and hours worked per worker H are not perfect substitutes, because:

- Fixed costs Z of employing an extra worker (irrespective of hours worked)
- Payment of overtime premiums x to induce workers to work beyond the standard work T
- The marginal efficiency of hours worked by an individual $e'(H)$ may increase over some range of H, but should eventually decrease as H becomes large due to fatigue

$$C = \begin{cases} [wT + (1+x)w(H-T) + Z]N + rK, & H > T \\ [wH + Z] + rK, & H < T \end{cases}$$

Firms thus solve the new optimization problem:

$$\min_{H,N,K} C \ \text{ s.t. } \ F\big(K, Ne(H)\big) \geq Y$$

Signs of the conditional elasticities:

- Increase in Z: always increase hours and decrease workers
- Increase in w: always reduce hours and increase workers
- Increase in x: reduce hours and increase workers if using overtime, no effect otherwise
- Increase in T: reduce hours and increase workers if using overtime, no effect otherwise

Adding scale effects to these elasticities (i.e. unconditional elasticities) does not change the signs of these effects generally, but does change their size.

Dynamic Labour Demand

Adjustment of labour by the firm is costly:

- Hiring costs: advertising, recruitment, training
- Firing costs: redundancy payouts (depend on employment protection legislation)

The presence of adjustment costs is often used as an explanation of “labour hoarding” during recessions i.e. firms do not lay off as many workers as static model would suggest.

Quadratic and symmetric costs $c(\Delta L) = \frac{b}{2}(\Delta L)^2$ yield the prediction that firms adjust employment slowly towards the desired long-run equilibrium level. Firms avoid large changes in employment in any period, as costs are higher (function of the square of the change in L). These are very tractable, but are unrealistic.

Piece-wise linear costs $c(\Delta L) = a\Delta L$ if $\Delta L > 0, c(\Delta L) = b\Delta L$ if $\Delta L < 0$ lead to firms jumping immediately to one of two desired levels (no need to smooth), or remaining at the existing level.

Stylized Facts of Labour Demand

- The elasticity of conditional (constant output) labour demand (substitution effect only) is negative and less than unity, and most probably in the interval [−0.15, −0.75]. Hamermesh suggests that "best guess" is most likely around −0.30.
- Studies of unconditional elasticities are more rare, but the evidence suggests that the own labour demand elasticity is around -1.0, thus scale effects are large
- Physical capital and skilled labour are p-complements
- Technical progress is complementary to skilled labour
- Labour demand is less elastic for skilled labour than unskilled labour
- Both number of workers N and hours of work H are p-substitutes for capital
- It is likely that workers N and hours of work H are p-substitutes

Compensating Differentials

Theoretical Model

Basic Idea

- Workers have heterogeneous preferences over consumption AND job characteristics.
- Firms differ in technology (costs) of providing "positive" working conditions, but providing them is always costly
- Workers have information about all jobs and are mobile, so can choose the optimal job for
- Perfect competition ensures firms that spend more on better conditions must be only able to afford to pay lower wages

Simple Model

- Let $g(Z)$ be the pdf of the taste for clean jobs
- Let $f(B)$ be the pdf of the costs of providing clean jobs across firms
- Workers will take a dirty job if $\Delta w > Z$
- Firms will supply a dirty job if $B > \Delta w$

First overall wages adjust so that overall supply equals demand. Think of this as normalizing both g and f distributions so that they both integrate to one. Next, Δw will adjust so that the area under the supply and demand distribution curves is equal – so that the supply of dirty jobs equals the demand for dirty jobs. This means that Δw will adjust such that the following condition holds:

$$\int_0^{\Delta w} g(Z)\, dB = \int_{\Delta w}^{\infty} f(B)\, dB$$

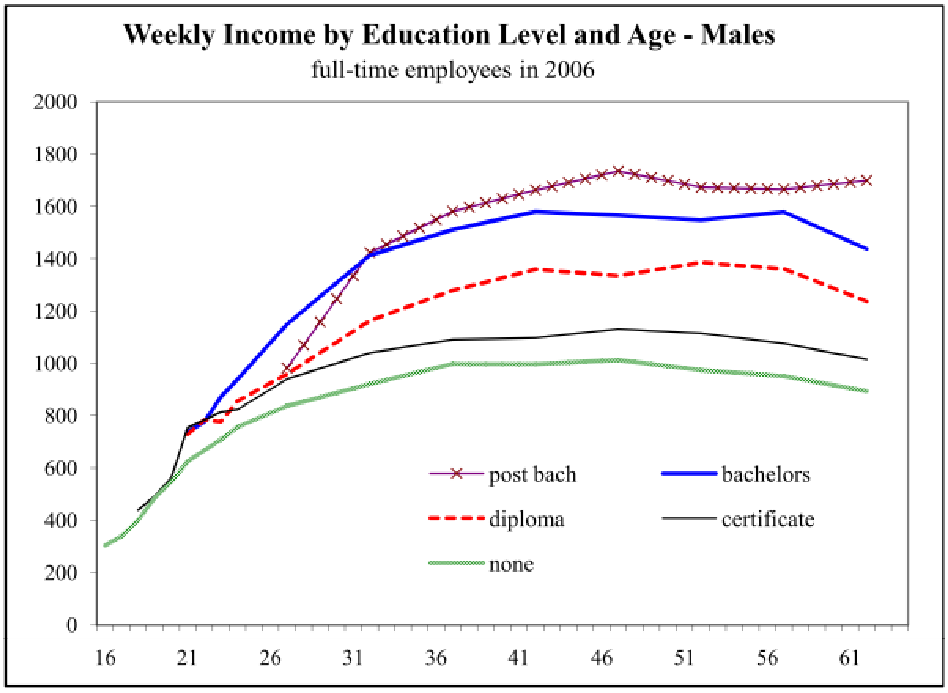
The equilibrium Δw only tells us about the willingness to pay for clean jobs Z for those individuals at the margin of the job choice decision. It does not tell us about the average willingness to pay in the population E(Z). Thus care should be taken in interpreting estimates of compensating wage differentials in what they can say about the distribution of Z in the population.

Empirical Evidence

- One consistent finding is that workers earn premiums for jobs with a higher risk of death
- There are also positive estimates of wage premia for the following:
    - night shifts
    - inflexible work schedules
    - standing a lot
    - noisy environments
    - lower job security
- Rebecca Edwards (2006) finds that having paid maternity leave is related to lower wages, as theory would predict
- Difficult to accurately measure job characteristics as we have to use average industry values

Education and Training

Income-Age Profit in Australia



Internal Rate of Return

This is the subjective discount rate that equilises present value of lifetime earnings under both 'more school' and 'work now' alternatives. For example, with opportunity costs on the left and lifetime income premium on the right:

$$\sum_{(t=0)}^{2} \frac{42{,}000}{(1+r)^t} = \sum_{t=3}^{47} \frac{15{,}000}{(1+r)^t}$$

Solving for r we find $r = 10.61\%$, which represents the internal rate of return. If this rate is above what could be achieved with some other type of investment, after allowing for risk differences (education is a risky investment, as the expected premia is merely an average with wide variation), then the investment is a profitable one.

Mincer Equation

The canonical mincer equation is linear in schooling S_i and quadratic in age/experience X_i:

$$\log(w_i) = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \epsilon_i$$

Models based upon maximising the present value of lifetime earnings predict the following sequence:

- individuals will specialise in investment (full-time schooling) in the beginning, where the stock of human capital is low and the benefits flow over the longest period.
- At some point, it is optimal to combine learning and earning, so the individual enters the labour market. Investment in on-the-job training at the beginning is high, leading to low wages both due to a low stock of human capital, and a high proportion of time devoted to learning
- As time passes, earnings increase rapidly due to accumulating human capital and falling investment in on-the-job training
- Eventually earnings may hit a maximum and decrease to retirement, as investments stop and human capital begins to depreciate

Endogeity and selection bias are likely to both be problems in this model. Also, the assumption of linear returns to years of education is likely unrealistic (e.g. sheepskin effects). Mincer equations should really be estimated only over longitudinal data (following the same individuals), though in practise they are usually estimated with cross-sectional data, which is only valid in the long-run steady state. Technical change (changing returns to human capital) or secular changes in the quality of education would violate this assumption.

Necessary Assumptions

- The return to schooling (ρ) is constant across levels of schooling
- Human capital accumulation (after full-time schooling) declines linearly with labour market experience
- The length of the working life n is the same for all individuals, irrespective of schooling i.e. those with more schooling work to an older age

Heterogeneity

This may emerge from:

- individual differences in the marginal return to schooling
- individual differences in the marginal costs (or tastes) of schooling

Causes of OLS Bias

- If individuals that would have done well without schooling acquire more schooling, then our estimates of the causal effect of education will be biased.
- Variation in b i : People with a higher marginal return from education will tend to acquire more schooling. Estimating the average effect by comparing those with more and less schooling ignores this selection. Implicitly it assumes that those that obtained less schooling would have obtained the same benefit as those that obtained more schooling.
- variation in r i : individuals that face higher marginal costs of education will be less likely to obtain schooling. Comparing those with more or less schooling ignores these differences. Importantly, the OLS estimates might be biased downwards if some individuals that don't go to school have

very high marginal benefits from schooling but don't go to school because they also have very high marginal costs

Returns to Education – the Bottom Line

- Estimates of the return to schooling vary a lot by study, with the range being that a year of schooling raises earnings by 6% to 12%
- Returns in the US for college education rose significantly in the 1980s, and slowly since
- Some evidence of a rising "college premium" in the UK and Canada also; less in Aus
- Returns to college are higher in the US than in the UK, Canada, Australia and Europe

Social Returns to Education

The tradition limited definition of 'social returns' simply meant including the full direct cost of education (private and public), and measuring the full benefit as before-tax income. Australian estimates showed that social returns were still sizeable, though consistently lower by 10-40% than private returns.

The more recent wider definition seeks to include such factors as lower crime, reduce welfare dependency, higher civic participation, better population health, faster adoption of new technologies, etc. There is some evidence in favour of these.

The Screening Hypothesis

Employers cannot observe productivity directly so use education to screen job applicants. Employers should eventually learn the productivity levels of workers but it may take a long time, and hiring costs are large. So employers use signals in their hiring process. The implication of this is that the social returns to education may be much lower than the private returns.

For a separating equilibrium (i.e. only high ability workers choose a higher education level that signals their higher productivity) requires that higher productivity workers have lower costs of obtaining an education or a greater distaste for learning. Otherwise, the signal would be useless and firms would not pay a premium for more highly-educated workers.

A separating equilibrium unambiguously makes low-ability workers worse off (as otherwise they would get the average wage higher than their marginal product), and may make high ability workers better off, so long as the cost of education is sufficiently low for them.

Evidence for and against the screening hypothesis is mixed. Firms could use cheaper signals like IQ tests, though this may be a poorer signals. Self-employed persons don't need to signal, so we would expect self-employed persons of a given income to have lower education; some evidence to support this. There may also be some evidence that people who get a job using the skills from their degree earn no more on average than those who do not use their skills.

On the Job Training

There are two main types of on-the-job training:

- General: learning skills that are useful in many firms, not just the one currently worked in
- Examples: driving vehicles, manual crafts, teaching, working with computers, doing economics research
- Specific: raises the productivity of the worker only in the firm currently working in
- Examples: firm specific equipment or processes, operational procedures of the firm, i.e. who in the firm does what, which forms to fill out, etc

Consider three wage levels before (MRP_1), during (MRP_0), and after training (MRP_2), such that $MRP_0 < MRP_1 < MRP_2$.

If general training, if firm does not pay worker MRP_2 afterwards, the worker will leave and earn that wage elsewhere (skills transferable). So firm cannot afford to pay above MRP_0 during training (as they can never recoup this), and will not as worker can capture all benefits of training. Hence, firms may offer the training, but workers will pay the full cost and gain the full benefits afterwards.

If specific training, then the worker cannot capture the value by leaving, so firm is under no obligation to pay above MRP_1 after training. But worker has bargaining power by merely threatening to quit if not paid something for staying as firm will lose all that investment. So firm will pay a bit above MRP_1. Knowing this, the worker will be willing to bear some of the costs of the training in the short term as is raises his/her bargaining power. Possible wage profile is a slightly upward-sloping line. This may explain rising wage profile over tenure.

The evidence seems to indicate that firms pay for a lot more general training that would be expected. This may be an indication that labour markets are not as flexible – harder for workers to move to another job.

Unemployment

Sources of Unemployment

Frictional Unemployment

- Frictions in labour markets - it takes time to match workers to jobs.
- The assumptions of the neo-classical model of labour markets (full information held by workers about jobs available and wages paid on all jobs) are relaxed.

Structural Unemployment

- 1. where worker skills do not match job requirements, and
- 2. where workers choose to search longer given other income (unemployment benefits).
- Basic search theory begins with the decision to take a particular job on offer, or to keep searching. This will be a function of what the individual receives if they continue searching.

Labour Market Disequilibrium

- Wages held above market-clearing levels
- Minimum wages, unions, implicit contracts, efficiency wages, insider-outsider theory
- Related to the Keynesian view, Phillips curves, etc, as these require wage stickiness - otherwise, wages would fall in recessions and unemployment would not exist

Theories of Sticky Wages

Implicit Contracts

- One version assumes that workers prefer wage stability over job stability. This assumption is required for the model to explain the existence of both downward rigid wages and temporary lay-offs of workers (common in the US)
- 2. In another version (we will discuss this one), it is not assumed that workers prefer wage stability over job stability. The assumption is not necessary if we are not interested in explaining the existence of layoffs

- Instead, we assume that workers are risk averse and cannot borrow, while firms are less risk averse and can borrow. Assume also that firm promises are credible to workers
- In this situation, workers and firms enter an implicit agreement: firms agree not to sack workers or reduce wages in recessions, while workers agree to accept below MRP wages over the business cycle
- This is beneficial for both, as it allows firms to pay lower wages and individuals to gain higher expected utility
- Workers can still leave if wages are below those offered elsewhere, so wages are downward rigid but not upward rigid, consistent with the data

Efficiency Wages

Firms offer wages above individual's alternative best wage

- 1. Attract better workers: Bigger pool of applicants, so can choose most dependable, motivated, experienced
- 2. Keep best workers, as loyal given fair treatment: Firms then more willing to pay for on-the-job training to raise worker productivity
- 3. Elicits more effort, as workers fear being sacked: Less shirking i.e. way of motivating effort, as wage in this job is higher than alternative
- 4. Raises worker morale: Workers feel they are treated more fairly; important in firms where teamwork is involved
- 5. Better nutrition in developing countries: Firms like Nike may do this. Lowers absenteeism through less sickness and raises energy levels

Insider-Outsider Theory

- Theory based on existence of fixed costs of employment and firm-specific skills
- Not perfect competition between those currently in jobs and the pool of unemployed
- The unemployed are the outsiders. New entrants to the labour market may also be outsiders
- Insiders can threaten to quit if not given wage payments, so firms lose sunk costs of hiring and training costs. Also, insiders can threaten to not co-operate with new hires for fear of replacement, and even harass new hires if they replace other insiders
- Outsiders cannot underbid insiders even in recessions. They cannot merely offer to work for lower wages and get the jobs due to the fixed costs and "rent-related" issues
- There has been several studies of this model in Australia over the years, as it has been quite influential, with some supporting evidence

Basic Job Search

Background

- Take the wage offer distribution as exogenous (partial-partial), unchanging over time and known to the individual
- Information regarding the location of vacant jobs and the compensation that they offer is imperfect
- This information must be acquired and evaluated before a worker can or is willing to become employed
- Individuals search or "shop" for the best job, but such shopping requires time and resources, and the benefits of continued search (a better paying job) are uncertain and in the future

Model Setup

- Infinitely lived unemployed job seeker, risk neutral and not liquidity constrained
- Infinitely lived jobs characterized by a wage w
- Cost of search c is interpreted as a flow per unit of time that includes the value of time spent searching and the direct out-of-pocket expenses
- Let time be represented by a sequence of discrete periods of length h
- Let b represent the benefits associated with looking for work – perhaps increased leisure time, unemployment benefits, etc
- Let $\beta(h)$ denote the discount rate applied to future benefits and costs per period of length h
- Let $q(n,h)$ represent the probability of receiving n offers in a period of length h spent searching
- Let the cdf $F(w)$ represent the distribution of wage offers
- let V denote the value of searching during the next period
- Let W denote the present value of stopping; namely of accepting the best offer received w during any period h and working forever at that wage

Hence the value of searching in the next period is given by:

$$V_h = (b-c)h + \beta(h)E\left[\max\left(V_{h+1}, W(x)\right)\right]$$

$$V = (b-c)h + \beta(h)\sum_{n=1}^{\infty}\left[q(n,h)\int_0^{\infty}\max\left(V, W(x)\right) dG(x,n) + q(0,h)V\right]$$

Reservation Wage

The worker will continue to search if $V > W(w)$. It turns out that there is some unique critical wage w^* that will satisfy this expression with equality. We call this the reservation wage. It is usually assumed that this is constant over time, though this may not be true owing to liquidity constraints and finite working life.

- Increases in the value of leisure b (and reductions in direct cost of search c) raise the reservation wage
- Increases in the interest rate (rate of discounting the future) reduce the reservation wage, as individuals become more impatient
- Increases in the job offer arrival rate raise the reservation wage, as individuals are more likely to get a high wage offer in a given time period

Hazard Rate

The hazard rate is the probability of leaving unemployment in any period, given that the individual is unemployed. If the reservation wage is constant, this hazard rate will be constant. If the reservation wage falls with time (for reasons cited above) then the hazard rate will rise over time – positive duration dependence.

Job Offer Arrival Rate

Increases in the job offer arrival rate raise the reservation wage, as individuals are more likely to get a high wage offer in a given time period. This means that increases in the job offer arrival rate λ have a direct effect of raising the hazard rate out of unemployment. However, increases in λ also lead to an increase in the reservation wage, which then reduces the hazard rate (an indirect effect). It has been shown that a sufficient condition for the direct effect to dominate is that the distribution of wage offers has a “log-concave” probability density function.

Wage Offer Distribution

How do changes in the wage offer distribution F(w) affect the reservation wage and the hazard?

- Increases in the mean of F(w) result in increases in the reservation wage w∗, but w∗ increases by less than the increase in the mean of F(w)
- Increases in the mean of F(w) result in increases in the hazard i.e. the direct effect on the hazard dominate the indirect effect via the less than one-for-one increase in w∗
- Increases in the variance of F(w) also result in increases in the reservation wage w∗ (workers wait for a wage from the upper tail of the wage distribution).

On the Job Search

1. If $\lambda_e = 0$ so there is no on-the-job search for the moment:

$$w^* = b + \frac{\lambda_u}{r+\delta}\int_{w^*}^{\infty}[1 - F(x)]\,dx$$

   Comparing this to Equation 3, the only difference is that including job destruction changes the "discount rate" to $r+\delta$ rather than just r. This reduces the reservation wage, and thus increases the unemployment hazard, as noted above.

2. If $\lambda_e > 0$, the job-seeker takes account of the possibility of future income associated with continuing to search while employed, so this lowers the reservation wage. She does not have to stay unemployed to search for a better paying job.

3. If $\lambda_e = \lambda_u$ then $w^* = b$. Why? The individual is just as likely to receive an acceptable offer when employed as when unemployed.

4. If $\lambda_e > \lambda_u$ then $w^* < b$, i.e. there is an incentive to accept 'bad' jobs. This might occur if employers are more likely to hire someone who is already employed than someone who is unemployed (perhaps a signal that they are more "employment ready").

On the Job Search provides a rationale for wages being positively related to both work experience (finding better matches over time) and with job tenure (jobs with higher wages last longer, as the individual is less likely to leave for a better paying job).

It also provides a rationale for the observation that a large proportion of wage increases of individuals occurs when individuals change jobs, rather than within job wage growth.

Equilibrium Search

These models generalize the partial-partial model by incorporating an endogenous determination of wages and the job offer rate, as firms compete against each other to hire workers and maximise profits. The trouble with these models was that initially they predicted a degenerate distribution at the reservation wage; as firms would have no incentive to offer more than this and no reason to offer less. It is possible to avoid this by assuming heterogenous preference for leisure b, but this being the sole basis for a non-degenerate distribution is unsatisfactory. A more recent model was able to generate a non-degenerate distribution without heterogeneity in worker preferences or firm productivity on the basis that some firms may pay more than others to keep quit rates down, as having more workers leads to greater aggregate profits, even if the profit per worker is less.

Inequality

Problem

There has been a substantial rise in wage inequality in much of the western world, especially Australia and the US, since the 1970s. The college premium in the US has risen dramatically since the 1980s despite continual expansion in higher education. This has not occurred in Australia, however it has not really fallen either despite much greater expansion in university education in this period compared to the US (e.g. uni reforms of the 1990s). How can we explain these phenomena?

Commonly Proposed Factors

- Skill-biased technological change: The Canonical Model focuses on understanding the rising college education wage premium in the US in particular, and looks at the race between the increase in supply of educated workers and increased demand via technological change
- Decline of unions: Unions tend to compress wage structures within firms, thus de-unionisation trends may help explain rising inequality. Borland estimates that 30% of the increase in variance for men and 15% for women is due to these factors
- Globalisation and Offshoring: Increasing international trade and offshoring may explain the loss of some middle skilled jobs in developed countries. Technological change (in communications in particular but also reductions in freight costs) may speed this process
- Minimum wage reductions: Falling real minimum wages in the US over the 1979-1988 period explain up to 25% (males) and 30% (females) of the change in the standard deviation of earnings over this period. Not really an issue in Australia

Canonical Skill-Biased Technical Change Model

Motivation

If there were no changes in the relative demand for different groups of workers, there should be a negative relationship between changes in relative wages and changes in relative supplies of these different groups (i.e. standard downward sloping demand for different types of workers). Looking at US data for recent decades, we do not find this expected negative correlation. This seems to indicate significant demand changes. Factoring in industry changes (e.g. decline in manufacturing) explained some but not all of this effect.

Assumptions

- Labour force is split into two skill levels: high skill H and low skill L
- The production function for the economy is of constant elasticity of substitution form (CES)
- Technological change is factor-augmenting, increasing A_L and A_H
- Labour markets are competitive for both types of labour
- The elasticity of substitution between L and H is given by σ

The Model

Aggregate production is given by:

$$Y = \left[(A_L L)^{\frac{\sigma-1}{\sigma}} + (A_H H)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

Note that if $\sigma = 1$, this simplifies to the Cobb-Douglas case. Wages are shown below.

$$w_L = \frac{\partial Y}{\partial L} = (A_L)^{\frac{\sigma-1}{\sigma}} \left[(A_L)^{\frac{\sigma-1}{\sigma}} + \left(A_H \frac{H}{L} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\sigma-1}}$$

$$w_H = \frac{\partial Y}{\partial H} = (A_H)^{\frac{\sigma-1}{\sigma}} \left[(A_H)^{\frac{\sigma-1}{\sigma}} + \left(A_L \frac{H}{L} \right)^{-\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\sigma-1}}$$

Taking the ratio of these wages and then logging the result gives:

$$\frac{w_H}{w_L} = \left(\frac{A_H}{A_L} \right)^{\frac{\sigma}{\sigma-1}} \left(\frac{H}{L} \right)^{-\frac{1}{\sigma}}$$

$$\log\left(\frac{w_H}{w_L} \right) = \frac{\sigma}{\sigma-1} \log\left(\frac{A_H}{A_L} \right) - \frac{1}{\sigma} \log\left(\frac{H}{L} \right)$$

This equation clearly shows that the wage ratio increases with the relative technological factor augmentation of H compared to L, and decreases with the H/L ratio in the workforce.

$$\frac{\partial \log(w_H/w_L)}{\partial \log(A_H/A_L)} = \frac{\sigma-1}{\sigma}$$

This shows us that skill biased technical change in favour of H workers only increase the wage premium if $\sigma > 1$, that is if H and L are gross substitutes. Most studies estimate σ to be between 1.4 and 2.

Problems with the Canonical Model

- It does not provide an explanation for falling real wages for the low skilled, unless there has been technical regress (however globalization, decline of unions, minimum wage falling)
- The model implies a falling trend in the rate of skill-biased technological change since the early 1990s, but there seems to be no reason for this (i.e. technology keeps improving at a rapid rate)
- The model treats technological change as exogenous, but skill-based technologies may be adopted more quickly when supplies of educated workers rise
- It is not readily able to explain different trends in residual earnings inequality for different groups (increasing for more educated, declining for less educated)

Job Polarisation

- There has been job polarisation over this period also i.e. simultaneous growth in the share of employment in high skill, high wage occupations and in low skill, low wage occupations
- Job polarisation has been at least as pronounced in Europe as in the US
- Growth in high skill occupation groups, and in low skill services, but less growth in middle skill occupations such as office administration, production workers and operators
- Routinization hypothesis: Progress in communication and information technology has made doing certain "routine" tasks using computers and machinery very cheap. Thus administrative and clerical work and routine production work can now be done by computers and overseas
- At the same time, the relative demand for workers who perform complementary non-routine tasks has increased
- Abstract tasks /non-routine cognitive: problem solving, intuition, creativity (high end jobs)
- Non-routine manual tasks: situational adaptability, visual and language recognition, in person interactions (low end service jobs)
- Changes in industrial structure do not explain much of the job polarisation in the US