Part VII

Accounting for the Endogeneity of Schooling
Accounting for the endogeneity of schooling

- Much of the CPS-Census literature on the returns to schooling ignores the choice of schooling and its consequences for estimating “the rate of return”.
- It ignores uncertainty.
- It is static and ignores the dynamics of schooling choices and the sequential revelation of uncertainty.
- It also ignores ability bias.
Accounting for the endogeneity of schooling

- Economists since C. Reinhold Noyes (1945) in his comment on Friedman and Kuznets (1945) have raised the specter of ability bias, noting that the estimated return to schooling may largely be a return to ability that would arise independently of schooling.

- Griliches (1977) and Willis (1986) summarize estimates from the conventional literature on ability bias.

- For the past 30 years, labor economists have been in pursuit of good instruments to estimate “the rate of return” to schooling, usually interpreted as a Mincer coefficient.
Accounting for the endogeneity of schooling

- However, the previous sections show that, for many reasons, the Mincer coefficient is not informative on the true rate of return to schooling, and therefore is not the appropriate theoretical construct to gauge educational policy.

- Card (1999) is a useful reference for empirical estimates from instrumental variable models.
Accounting for the endogeneity of schooling

- Even abstracting from the issues raised by the sequential updating of information, and the distinction between *ex ante* and *ex post* returns to schooling, which we discuss further below, there is the additional issue that returns, however defined, vary among persons.

- A random coefficients model of the economic return to schooling has been an integral part of the human capital literature since the papers by Becker and Chiswick (1966), Chiswick (1974), Chiswick and Mincer (1972) and Mincer (1974).
Accounting for the endogeneity of schooling

- In its most stripped-down form and ignoring work experience terms, the Mincer model writes log earnings for person $i$ with schooling level $S_i$ as

$$\ln y_i = \alpha_i + \rho_i S_i,$$

where the “rate of return” $\rho_i$ varies among persons as does the intercept, $\alpha_i$.

- For the purposes of this discussion think of $y_i$ as an annualized flow of lifetime earnings.

- Unless the only costs of schooling are earnings foregone, and markets are perfect, $\rho_i$ is a percentage growth rate in earnings with schooling and not a rate of return to schooling.
Accounting for the endogeneity of schooling

Let \( \alpha_i = \bar{\alpha} + \varepsilon_{\alpha_i} \) and \( \rho_i = \bar{\rho} + \varepsilon_{\rho_i} \) where \( \bar{\alpha} \) and \( \bar{\rho} \) are the means of \( \alpha_i \) and \( \rho_i \).

Thus the means of \( \varepsilon_{\alpha_i} \) and \( \varepsilon_{\rho_i} \) are zero.

Earnings equation (14) can be written as

\[
\ln y_i = \bar{\alpha} + \bar{\rho}S_i + \{\varepsilon_{\alpha_i} + \varepsilon_{\rho_i}S_i\}. \tag{15}
\]

Equations (14) and (15) are the basis for a human capital analysis of wage inequality in which the variance of log earnings is decomposed into components due to the variance in \( S_i \) and components due to the variation in the growth rate of earnings with schooling (the variance in \( \bar{\rho} \)), the mean growth rate across regions or time (\( \bar{\rho} \)), and mean schooling levels (\( \bar{S} \)).

See, e.g. Mincer, 1974, and Willis, 1986.
Accounting for the endogeneity of schooling

- Given that the growth rate $\rho_i$ is a random variable, it has a distribution that can be studied using the methods surveyed below.

- Following the representative agent tradition in economics, it has become conventional to summarize the distribution of growth rates by the mean, although many other summary measures of the distribution are possible.

- For the prototypical distribution of $\rho_i$, the conventional measure is the “average growth rate” $E(\rho_i)$ or $E(\rho_i|X)$, where the latter conditions on $X$, the observed characteristics of individuals.

- Other means are possible such as the mean growth rates for persons who attain a given level of schooling.
Accounting for the endogeneity of schooling

- The original Mincer model assumed that the growth rate of earnings with schooling, $\rho_i$, is uncorrelated with or is independent of $S_i$.

- This assumption is convenient but is not implied by economic theory.

- It is plausible that the growth rate of earnings with schooling declines with the level of schooling.
Accounting for the endogeneity of schooling

- It is also plausible that there are unmeasured ability or motivational factors that affect the growth rate of earnings with schooling and are also correlated with the level of schooling.

- Rosen (1977) discusses this problem in some detail within the context of hedonic models of schooling and earnings.

- A similar problem arises in analyses of the impact of unionism on relative wages and is discussed in Lewis (1963).
Accounting for the endogeneity of schooling

- Allowing for correlated random coefficients (so $S_i$ is correlated with $\varepsilon_{\rho_i}$) raises substantial problems that are just beginning to be addressed in a systematic fashion in the recent literature.

- Here, we discuss recent developments starting with Card’s (1999) random coefficient model of the growth rate of earnings with schooling, a model that is derived from economic theory and is based on the analysis of Becker’s model by Rosen (1977).

- We consider conditions under which it is possible to estimate the mean effect of schooling and the distribution of returns in his model.

- The next section considers the more general and recent analysis of Carneiro, Heckman, and Vytlacil (2005).
Accounting for the endogeneity of schooling

- In Card’s (1999, 2001) model, the preferences of a person over income \( y \) and schooling \( S \) are
  \[ U(y, S) = \ln y(S) - \varphi(S) \quad \varphi'(S) > 0 \]
  and
  \[ \varphi''(S) > 0. \]

- The schooling-earnings relationship is \( y = g(S) \).

- This is a hedonic model of schooling, where \( g(S) \) reveals how schooling is priced out in the labor market.
Accounting for the endogeneity of schooling

- This specification is written in terms of annualized earnings and abstracts from work experience.

- It assumes perfect certainty and abstracts from the sequential resolution of uncertainty that is central to the modern literature.

- In this formulation, discounting of future earnings is kept implicit.

- The first order condition for optimal determination of schooling is

\[
\frac{g'(S)}{g(S)} = \varphi'(S). \tag{16}
\]
Accounting for the endogeneity of schooling

- The term \( \frac{g'(s)}{g(s)} \) is the percentage change of earnings with schooling or the “growth rate” at level \( s \).

- Card’s model reproduces Rosen’s (1977) model if \( r \) is the common interest rate at which agents can freely lend or borrow and if the only costs are \( S \) years of foregone earnings.

- In Rosen’s setup, an agent with an infinite lifetime maximizes \( \frac{1}{r} e^{-rS} g(S) \) so \( \varphi(S) = rS + \ln r \), and \( \frac{g'(S)}{g(S)} = r \).
Accounting for the endogeneity of schooling

- Linearizing the model, we obtain

\[
\frac{g'(S_i)}{g(S_i)} = \beta_i(S_i) = \rho_i - k_1 S_i, \quad k_1 \geq 0,
\]

\[
\varphi'(S_i) = \delta_i(S_i) = r_i + k_2 S_i, \quad k_2 \geq 0.
\]

- Substituting these expressions into the first order condition (16), we obtain that the optimal level of schooling is

\[
S_i = \frac{(\rho_i - r_i)}{k}, \text{ where } k = k_1 + k_2.
\]

- Observe that if both the growth rate and the returns are independent of \( S_i \), \((k_1 = 0, k_2 = 0)\), then \( k = 0 \) and if \( \rho_i = r_i \), there is no determinate level of schooling at the individual level.

- This is the original Mincer (1958) model.
Accounting for the endogeneity of schooling

- One source of heterogeneity among persons in the model is $\rho_i$, the way $S_i$ is transformed into earnings.

- School quality may operate through the $\rho_i$ for example, as in Behrman and Birdsall (1983), and $\rho_i$ may also differ due to inherent ability differences.

- A second source of heterogeneity is $r_i$, the “opportunity cost” (cost of schooling) or “cost of funds.”

- Higher ability leads to higher levels of schooling.

- Higher costs of schooling results in lower levels of schooling.
Accounting for the endogeneity of schooling

- We integrate the first order condition (16) to obtain the following hedonic model of earnings,

\[ \ln y_i = \alpha_i + \rho_i S_i - \frac{1}{2} k_1 S_i^2. \]  (17)

- To achieve the familiar looking Mincer equation, assume \( k_1 = 0 \).

- This assumption rules out diminishing “returns” to schooling in terms of years of schooling.
Accounting for the endogeneity of schooling

- Even under this assumption, \( \rho_i \) is the percentage growth rate in earnings with schooling, but is not in general an internal rate of return to schooling.

- It would be a rate of return if there were no direct costs of schooling and everyone faces a constant borrowing rate.

- This is a version of the Mincer (1958) model, where \( k_2 = 0 \), and \( r_i \) is constant for everyone but not necessarily the same constant.
Accounting for the endogeneity of schooling

- If $\rho_i > r_i$, person $i$ takes the maximum amount of schooling.
- If $\rho_i < r_i$, person $i$ takes no schooling and if $\rho_i = r_i$, schooling is indeterminate.
- In the Card model, $\rho_i$ is the person-specific growth rate of earnings and overstates the true rate of return if there are direct and psychic costs of schooling.
Accounting for the endogeneity of schooling

- This simple model is useful in showing the sources of endogeneity in the schooling earnings model.

- Since schooling depends on $\rho_i$ and $r_i$, any covariance between $\rho_i - r_i$ (in the schooling equation) and $\rho_i$ (in the earnings function) produces a random coefficient model.

- Least squares will not estimate the mean growth rate of earnings with schooling unless, $\text{Cov}(\rho_i, \rho_i - r_i) = 0$. 
Accounting for the endogeneity of schooling

- Dropping the \( i \) subscripts, the conditional expectation of log earnings given \( s \) is

\[
E (\ln y \mid S = s) = E (\alpha \mid S = s) + E (\rho \mid S = s) s.
\]

- The first term produces the conventional ability bias if there is any dependence between \( s \) and raw ability \( \alpha \).

- Raw ability is the contribution to earnings independent of the schooling level attained.

- The second term arises from sorting on returns to schooling that occurs when people make schooling decisions on the basis of growth rates of earnings with schooling.

- It is an effect that depends on the level of schooling attained.
Accounting for the endogeneity of schooling

- In his Woytinsky Lecture (1967), Becker points out the possibility that many able people may not attend school if ability \( (\rho_i) \) is positively correlated with the cost of funds \( (r_i) \).
- A meritocratic society would eliminate this positive correlation and might aim to make it negative.
- Schooling is positively correlated with the growth rate \( (\rho_i) \) if \( \text{Cov}(\rho_i, \rho_i - r_i) > 0 \).
- If the costs of schooling are sufficiently positively correlated with the growth rate, then schooling is negatively correlated with the growth rate.
Accounting for the endogeneity of schooling

- Observe that $S_i$ does not directly depend on the random intercept $\alpha_i$.

- Of course, $\alpha_i$ may be statistically dependent on $(\rho_i, r_i)$.

- In the context of Card’s model, we consider conditions under which one can identify $\bar{\rho}$, the mean growth rate of earnings in the population as well as the full distribution of $\rho$.

- First we consider the case where the marginal cost of funds, $r_i$, is observed and consider other cases in the following subsections.
Estimating the mean growth rate of earnings when $r_i$ is observed

- A huge industry surveyed in Card (1999) seeks to estimate the mean growth rate in earnings, $E(\rho_i)$, calling it the “causal effect” of schooling.

- For reasons discussed earlier in this chapter, in general, it is not an internal rate of return.

- However, it is one of the ingredients used in calculating the rate of return as we develop further below.

- The “causal effect” may also be of interest in its own right if the goal is to estimate pricing equations for labor market characteristics.

- We discuss some simple approaches for identifying causal effects before turning to a more systematic analysis below.
Estimating the mean growth rate of earnings when $r_i$ is observed

- Suppose that the cost of schooling, $r_i$, is measured by the economist.
- Use the notation “⊥” to denote statistical independence.
- Assume
  \[ r_i \perp (\rho_i, \alpha_i). \]
- This assumption rules out any relationship between the cost of funds ($r_i$) and raw ability ($\alpha_i$) with the growth rate of earnings with schooling.
Estimating the mean growth rate of earnings when $r_i$ is observed

- For example, it rules out fellowships based on ability.
- We make this assumption to illustrate some ideas and not because of its realism.
- Observing $r_i$ implies that we observe $\rho_i$ up to an additive constant.
- Recall that $S_i = \frac{(\rho_i - r_i)}{k}$, so that $\rho_i = r_i + kS_i$ and $\bar{\rho} = E(\rho_i) = \bar{r} + kE(S_i)$. 
Estimating the mean growth rate of earnings when $r_i$ is observed

- $r_i$ is a valid instrument for $S_i$ under the assumption that $k_1 = 0$.

- It is independent of $\alpha_i, \rho_i$ (and hence $\varepsilon_{\alpha i}, \varepsilon_{\rho i}$) and is correlated with $S_i$ because $S_i$ depends on $r_i$. 
Estimating the mean growth rate of earnings when \( r_i \) is observed

**Form**

\[
\frac{\text{Cov}(\ln y_i, r_i)}{\text{Cov}(S_i, r_i)} = E \left\{ (r_i - \bar{r}) \left[ (\alpha_i - \bar{\alpha}) + (\rho_i - \bar{\rho})(S_i - \bar{S}) + \bar{\rho}S_i + \rho_i \bar{S} - \bar{\rho} \bar{S} \right] \right\}
\]

\[
E \left\{ \left[ \frac{\rho_i - r_i}{k} \right] [r_i - \bar{r}] \right\}
\]

\[
\frac{1}{k} E[(\Delta r)(\Delta \rho)(\Delta \rho - \Delta r)] - \frac{\bar{\rho} \sigma_r^2}{k},
\]

\[
- \frac{\sigma_r^2}{k}
\]

where \( \Delta X = X - E(X) \).
Estimating the mean growth rate of earnings when $r_i$ is observed

- As a consequence of the assumed independence between $r_i$ and $(\rho_i, \alpha_i)$, $E[(\Delta r)(\Delta \rho)^2] = 0$ and $E[(\Delta r)^2 \Delta \rho] = 0$, so

$$\frac{\text{Cov}(\ln y_i, r_i)}{\text{Cov}(S_i, r_i)} = \bar{\rho}.$$
Estimating the mean growth rate of earnings when \( r_i \) is observed

- Observe that \( \bar{\rho} \) is not identified by this argument if \( \rho_i \not\perp r_i \) (so the mean growth rate of earnings depends on the cost of schooling).

- In that case, \( E[(\Delta r)(\Delta \rho)^2] \neq 0 \) and \( E[(\Delta r)^2(\Delta \rho)] \neq 0 \).

- If \( r_i \) is known and \( r_i = L_i \gamma + M_i \), where the \( L_i \) are observed variables that explain \( r_i \) and \( E(M_i | L_i) = 0 \), then \( \gamma \) is identified, provided a rank condition for instrumental variables is satisfied.

- We require that \( L_i \) be at least mean independent of \( (M_i, \rho_i, \alpha_i) \).

- From the schooling equation we can write \( S_i = (\rho_i - L_i \gamma - M_i)/k \) and \( k \) is identified since we know \( \gamma \).
Estimating the mean growth rate of earnings when $r_i$ is observed

- Observe that we can estimate the distribution of $\rho_i$ since $\rho_i = r_i + kS_i$, $k$ is identified and $(r_i, S_i)$ are known.

- This is true even if there are no instruments $L$, ($\gamma = 0$), provided that $r_i \perp \rho_i, \alpha_i$.

- With the instruments that satisfy at least the mean independence condition, we can allow $r_i \not\perp \rho_i$ and all parameters and distributions are still identified.

- The model is fully identified provided $r_i$ is observed and $L_i \perp (M_i, \rho_i, \alpha_i)$.

- Thus, we can identify the mean return to schooling.
Estimating the mean growth rate when \( r_i \) is not observed

- If \( r_i \) is not observed and so cannot be used as an instrument, but we know that \( r_i \) depends on observed factors \( L_i \) and \( M_i \),
  \[ r_i = L_i \gamma + M_i \] and \( L_i \perp \perp (M_i, \alpha_i, \rho_i) \), then our analysis carries over and the mean growth rate \( \bar{\rho} \) is identified.

- Recall that \( \ln y_i = \alpha_i + \bar{\rho} S_i + (\rho_i - \bar{\rho})S_i \).

- Substitute for \( S_i \) to get an expression of \( y_i \) in terms of \( L_i \),
  \[ \ln y_i = \alpha_i + \rho_i(\rho_i - L_i \gamma - M_i)/k. \]
Estimating the mean growth rate when $r_i$ is not observed

- We obtain the vector moment equations:

$$\text{Cov}(\ln y_i, L_i) = \bar{\rho} \text{ Cov}(S_i, L_i),$$

so $\bar{\rho}$ is identified from the population moments because the covariances on both sides are available.

- Partition $\gamma = (\gamma_0, \gamma_1)$, where $\gamma_0$ is the intercept and $\gamma_1$ is the vector of slope coefficients.

- From the schooling equation, we obtain

$$S_i = \frac{\rho_i - L_i \gamma_1 - M_i}{k} - \frac{\gamma_0}{k}$$

$$= -\frac{L_i \gamma_1}{k} + \frac{\rho_i - M_i}{k} - \frac{\gamma_0}{k}.$$
Estimating the mean growth rate when $r_i$ is not observed

- We can identify $\gamma_1/k$ from the schooling equation, as well as the mean growth rate $\bar{\rho}$.

- However, we cannot identify the distribution of $\rho_i$ or $r_i$ unless further assumptions are invoked.

- We also cannot separately identify $\gamma_0$, $\gamma_1$ or $k$.

Adding selection bias

- Selection bias can arise in two distinct ways in the Becker-Card-Rosen model: through dependence between $\alpha_i$ and $\rho_i$ and through dependence between $\alpha_i$ and $r_i$.

- Allowing for selection bias,

$$
E(\ln y_i \mid S_i) = E(\alpha_i \mid S_i) + E(\rho_i S_i \mid S_i)
$$

$$
= E(\alpha_i \mid S_i) + E(\rho_i \mid S_i)S_i.
$$

- If there is an $L_i$ that affects $r_i$ but not $\rho_i$ and is independent of $(\alpha_i, M_i)$, i.e., $L_i \perp \perp (\alpha_i, \rho_i, M_i)$, and $E(r_i \mid L_i)$ is a nontrivial function of $L_i$, in the special case of a linear schooling model,

$$
E(\ln y_i \mid L_i) = E(\alpha_i \mid L_i) + E(\rho_i S_i \mid L_i)
$$

$$
= \eta + \bar{\rho}E(S_i \mid L_i).
$$
Adding selection bias

- Since we can identify $E(S_i \mid L_i)$ we can identify $\bar{\rho}$.

- Thus, under the stated conditions, the instrumental variable (IV) method identifies $\bar{\rho}$ when there is selection bias.

- In a more general nonparametric case for the schooling equation, which we develop in the next section of this chapter, this argument breaks down and $\bar{\rho}$ is not identified when $\rho_i$ determines $S_i$ in a general way.
Adding selection bias

- The sensitivity of the IV method to assumptions about special features of Card’s model is a simple demonstration of the fragility of the method.

- We return to this model later and use it to motivate recent developments in the literature on identifying information available to agents when they make their schooling decisions.
Summary

- Card’s version of the Becker (1967)-Rosen (1977) model is a useful introduction to the modern literature on heterogeneous “returns to schooling.

- \( \rho_i \) is, in general, a person-specific growth rate of log earnings with schooling and not a rate of return.

- There is a distribution of \( \rho_i \) and no scalar measure is an adequate summary of this distribution.

- Recent developments in this literature, to which we now turn, demonstrate that standard instrumental variable methods are blunt tools for recovering economically interpretable parameters.