Chapter 2: Applications of discrete and dynamic choice models

Joan Llull

Structural Econometrics for Labor Economics
and Industrial Organization

IDEA PhD Program

I. Introduction

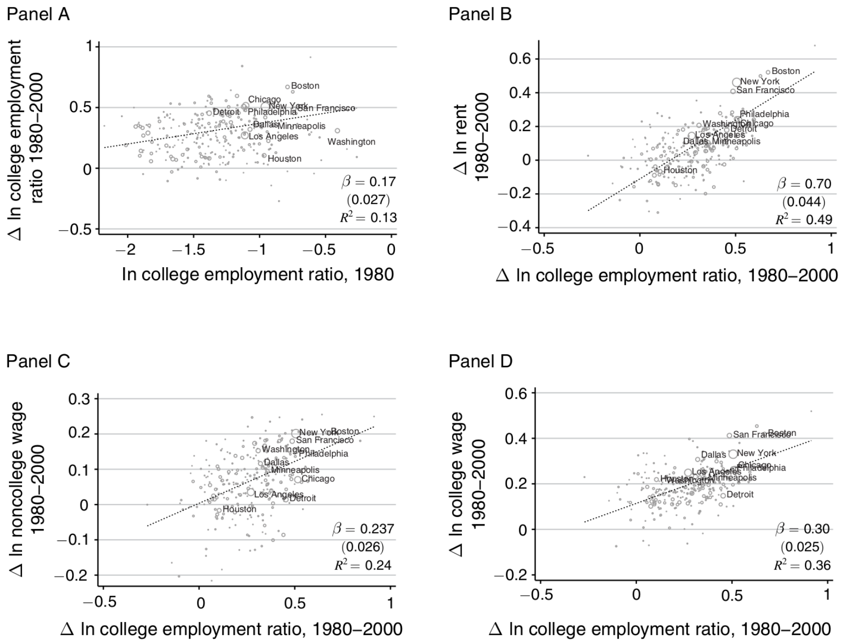
In this chapter we review three pieces of empirical research that serve as an illustration of the usage of (static) discrete choice models, selection models, and dynamic programming in empirical micoeconometrics. The first application is a paper by Rebecca Diamond, published in the American Economic Review in 2016, in which she estimates a location choice model to explain the diverging location choices by skill in the 1980-2000 period. This paper is an application of static discrete choice models, and it is also representative of an important literature in Urban Economics, the one estimating spatial equilibrium models. As an application for selection models in structural estimation, we present a review of the standard static female labor supply model based on the Handbook of Labor Economics chapter by Michael Keane, Kenneth Wolpin, and Petra Todd on structural estimation, published in 2011. The static female labor supply model is often replaced by a dynamic model of labor supply, especially when human capital considerations are in place, but the static model is a useful stepping stone in that direction. Finally, to review the estimation of (equilibrium) continuous choice dynamic models, we review a classic and very well cited paper by James Heckman, Lance Lochner, and Christopher Taber, published in the first issue of the Review of Economic Dynamics, which presents a dynamic model of human capital accumulation to revisit the sources of wage inequality. This model is close to the dynamic models often specified in Macroeconomics, but it is estimated with techniques that are more commonly used in empirical microeconomics.

II. Spatial equilibrium: Diamond (2016)

Spatial equilibrium models date back from Rosen (1979) and Roback (1982). The Rosen-Roback model is a static model that features one-time migration decisions. Many different versions of this type of spatial equilibrium models have been estimated in the literature. These models typically feature (some of) the following: i) perfectly mobile labor markets, potentially with moving costs, ii) fixed land and endogenous housing markets, iii) local amenities, iv) productivity differentials across cities, and v) local price differences. They are somewhat at the intersection between Urban and Labor Economics (cities vs individuals as the subject of interest). The paper by Diamond (2016) is a paper that fits very well in this category, featuring most of these items, and allowing for endogenous amenities as the main novelty.

This section provides a stylized version of the model. Diamond's research question is motivated by the observation that cities that had the largest proportion of college graduates among their workers in 1980 also experienced the largest increase in the college to high school ratio in the 1980-2000 period:

Figure 1. Changes in Wages, Rents, and College Employment Ratios, 1980-2000



Note: Source: Diamond (2016).

This pattern implies an increasing sorting across cities by education over these two decades. This increasing sorting is also associated to larger increases in rent prices and wages for both college and non-college workers. Diamond's paper aims at understanding the reasons behind this increasing sorting observed in the data.

Let Y_j denote output in city j. Also let U_j denote the supply of unskilled (non-college) labor in the city, S_j denote the supply of skilled (college) labor, and K_j denote capital. The aggregate demands of the two labor inputs are given by the first order conditions on the following production function:

$$Y_j = K_j^{1-\alpha}(\theta_{Uj}U_j^{\rho} + \theta_{Sj}S_j^{\rho})^{\frac{\alpha}{\rho}}. \tag{1}$$

The terms θ_{Uj} and θ_{Sj} denote city-education group-specific productivity shocks, and are the key exogenous force driving the endogenous changes produced by the model for the 1980-2000 period. Workers are paid their marginal product, denoted by W_{kt} for $k \in \{U, S\}$, which is given by the first order conditions on the aggregate firm's problem in each city.

Workers of skill k choose consumption of housing services and goods, as well as location, to maximize their utility. Let c denote their consumption of goods, and h denote their consumption of housing services. Also let $G(A_j(U_j, S_j), \boldsymbol{X}_j, \boldsymbol{z}) + \varepsilon_j$ denote an amenity function, which depends on endogenous amenities $A_j(U_j, S_j)$ and on exogenous city-specific amenities $\boldsymbol{X}_j$, as well as observable and unobservable idiosyncratic variables $\boldsymbol{z}$ and ε_j. Let P denote (U.S. level) current price index, and let R_j denote rent prices. The worker's decision problem is thus given by:

$$\begin{aligned} \max_{\{j,c,h\}} & \zeta \ln c + (1-\zeta) \ln h + G(A_j(U_j, S_j), \boldsymbol{X}_j, \boldsymbol{z}) + \varepsilon_j \\ \text{s.t.} \qquad & Pc + R_j h \leq w_{kj}. \end{aligned} \tag{2}$$

This problem determines labor supply in each location and housing demand.The housing supply is given by:

$$R_j = F(C_j, L_j), \tag{3}$$

where C_j are the construction costs at city j and L_j denotes land availability.

Most of the elements of the model are city-education-specific (capital letters), and individual-specific variables only appear in the individual decision problem, as a result of observable and unobservable amenities. The estimation is, thus, a two-stage procedure. In the first step, the parameters associated to individual-specific regressors $\boldsymbol{z}$ are estimated by maximum likelihood, as a conditional logit model, collapsing city-education group-specific parts into city-education dummies. In the second step, the remaining parameters of the model are jointly estimated by two-step Generalized Method of Moments (GMM) in first differences, using city level variation. In the specification of the moment conditions for the GMM problem, the instruments used for estimation include the exogenous variables of the model plus a combination of education group-specific Bartik shocks and their interactions with two variables that affect the housing supply elasticity: a regulatory index and an index of land availability.

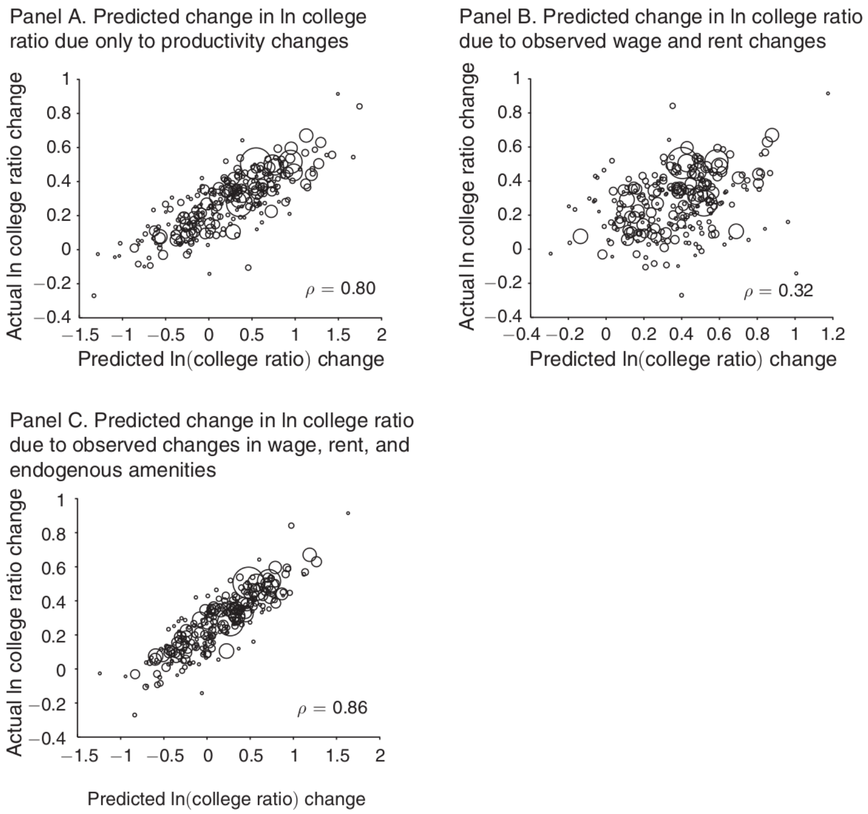
The GMM equations include the two first order conditions on the firm's problem in order to estimate productivity, a linear expression that links the estimated city-education level utility dummies δ_{jk} to its primitives, an equilibrium expression for the housing market that links rents with interest rates, regulation and land availability indices, and endogenous housing demand variables, and an expression for endogenous amenities. The latter is given by the following expression:

$$\Delta A_j = \gamma \Delta \ln \frac{S_j}{U_j} + \epsilon_j, \tag{4}$$

where A_j denotes an amenity index computed from the combination of many different variables at the city level, using a factor model.

Using the estimated parameters, the author uses the model to simulate a set of counterfactual scenarios that allow her to decompose the observed patterns described in Figure 1 into the different factors that may have generated them. In particular, all graphs in the figure plot the predicted change in the college to non-college employment ratio between 1980 and 2000 in different counterfactuals (horizontal axis) against the actual change observed in the data (vertical axis):

Figure 2. Predicted Changes in the log of College Employment Ratio, 1980-2000



Note: Source: Diamond (2016).

The graphs also include the correlation between actual and predicted values. Each graph represents a different counterfactual exercise. In Figure 2.A, the author simulates an economy in which rents and amenity values are as in 1980, but wages are determined by productivity levels of year 2000 (but labor supplies of

1980). In this context, productivity differences drive an important part of the observed changes (the correlation between predicted and actual values is 0.8). Figure 2.B shows an alternative exercise that allows wages and rents to adjust to the new productivity levels, but keeps endogenous amenities fixed to 1980 levels. In this counterfactual, the correlation between predicted and actual values is substantially reduced, which indicates that these factors, if anything, go in the opposite direction to the changes described in Figure 1. This is so because even though some cities became relatively more productive for high skilled workers than others, which increased wages substantially, further college workers move there, which mitigates the increase in wages predicted in the first place, and increases housing rents, which prevents some individuals to move there. Given this result then, where is the missing variation coming from? Figure 2.C show that endogenous amenities play an important role. In particular, once amenities are also allowed to endogenously react, the correlation between data and predictions rises to 0.86. The interpretation of this result is that, despite the increases in housing rents and the mitigation of the initial increases in wages, cities with further college individuals became more desirable cities to live in, which makes more people willing to live there, even though the increasing housing rents make college workers more willing to pay for the extra amenities than non-college workers.

III. Married woman's labor force participation

In this section, we review the basic unitary household model to describe female labor force participation decisions. In this model, the household is well represented by a single decision unit that takes into account the utilities of the two members of the couple in the decision process. Let c denote consumption, n denote the number of children in the household, $\boldsymbol{x}$ denote a vector of observable characteristics, ε denote some unobservable factor that affects couple's valuation of wife's leisure/home production, and $d = 1$ if the wife works and zero otherwise. The couple's utility is given by $U(c, d, n, \boldsymbol{x}, \varepsilon(1-d))$, with $\partial U/\partial c > 0$, $\partial^2 U/\partial c^2 < 0$, and $U(c, 1, n, \boldsymbol{x}, \varepsilon) > U(c, 0, n, \boldsymbol{x}, 0)$ for some values of ε, meaning positive and decreasing marginal utility of consumption, and some preference for leisure/home production over labor, at least for some realizations of the taste shock. Typically, $U(c, 1, n, \boldsymbol{x}, \varepsilon) > U(c, 1, n', \boldsymbol{x}, \varepsilon)$ for $n > n'$ as well, indicating stronger preference for not working in the presence of children.

The husband is assumed to work, generating income y. The wife receives a wage offer $\omega(\boldsymbol{x}, \upsilon)$ and decides whether to work or not accordingly. If the wife works, the household incurs in child care cost of π per child. The budget constraint is:

$$c = y + [\omega(\boldsymbol{x}, \upsilon) - \pi n]d. \tag{5}$$

Unobservable variables that influence the couple's utility ε and those entering in the woman's wage offer υ are serially uncorrelated and are jointly distributed as $F(\varepsilon, \upsilon | y, \boldsymbol{x}, n)$. Let $\mathbb{1}\{\cdot\}$ denote the indicator function, which equals one if the condition in the argument is satisfied. The probability that the wife participates is:

$$\begin{aligned}
&\Pr(d = 1 | \boldsymbol{x}, n, y) \\
&\quad = \int \mathbb{1}\{U(y + \omega(\boldsymbol{x}, \upsilon) - \pi n, 1, n, \boldsymbol{x}, \varepsilon) - U(y, 1, n, \boldsymbol{x}, 0) > 0\} dF(\varepsilon, \upsilon | y, \boldsymbol{x}, n) \\
&\quad \equiv G(y, \boldsymbol{x}, n).
\end{aligned} \tag{6}$$

In the context of this model, we can proceed with structural or non-structural estimations. The goal of structural model is to estimate all or some of the primitive elements of the model, $U(\cdot)$, $\omega(\cdot)$, and $F(\cdot)$. Non-structural estimation is concerned about the estimation of $G(\cdot)$. On top of the structural/non-structural dichotomy, we can also decide to make additional assumptions about the functional forms of the primitives and estimate their parameters (parametric approach) or try to recover these functions non-parametrically.

In order to analyze the convenience of these four approaches, suppose you want to test to what extent the following elements affect participation: i) wages, ii) husband's earnings, and iii) childcare costs. Which of the four approaches you choose crucially determines your ability to address each of the three goals.

In a non-structural non-parametric approach, we do not need to make further assumptions: we simply estimate $G(\cdot)$ non-parametrically. However, in order to achieve the first goal, we would need further assumptions. In particular, we need that some element of $\boldsymbol{x}$ affects wages and not the participation directly. This variable (exclusion restriction) would allow us to infer how participation probabilities vary when we vary wages. Let $\boldsymbol{x}_1$ denote the partition of the vector $\boldsymbol{x}$ that affects wages but does not enter the utility function directly. The effect of wage changes on participation could be inferred from $\partial G / \partial \boldsymbol{z}'$. The second goal is clearly feasible (within sample) without further assumptions, since $\partial G / \partial y$ is identified. The third goal, on the contrary, is unfeasible without further assumptions, because G and π cannot be separately identified.

Parametrically specifying $G(\cdot)$ leads to the parametric non-structural approach. For example, you can specify it as a probit or logit. This approach takes us to a similar situation as before, except that now, $\partial G / \partial y$ is also identified out of sample.

The non-parametric structural approach requires identifying $U(\cdot)$, $\omega(\cdot)$, and $F(\cdot)$ separately without imposing additional assumptions about functional forms. This is infeasible provided that wages are only observed for the individuals who work. With further non-parametric assumptions and data on wages for the women who work, one could go a bit further. For example, if $\omega(\cdot)$ is assumed to be additively separable, one should be able to identify the deterministic part of the wage function, given that there are exclusion restrictions that affect the participation decision but not wages (y, n, and potentially some elements in $\boldsymbol{x}$ not included in $\boldsymbol{x}_1$, denoted by $\boldsymbol{x}_2$). Further assumptions on $F(\cdot)$ could also lead to partial identification of $U(\cdot)$.

Finally, consider the structural parametric approach with the following very standard parametric assumptions:

$$U(c, d, n, \boldsymbol{x}, n, \varepsilon(1-d)) \equiv c + (1-d)[\boldsymbol{x}_2'\boldsymbol{\beta} + \gamma n + \varepsilon], \tag{7}$$

$$\omega(\boldsymbol{x}, \upsilon) = \boldsymbol{x}'\boldsymbol{\delta} + \upsilon, \tag{8}$$

and:

$$(\varepsilon, \upsilon)'|y, \boldsymbol{x}, n \sim \mathcal{N}(\mathbf{0}, \Sigma). \tag{9}$$

Given this parametrization, the difference in utilities in Equation (6) is:

$$\begin{aligned} &U(y + \omega(\boldsymbol{x}, \upsilon) - \pi n, 1, n, \boldsymbol{x}, \varepsilon) - U(y, 1, n, \boldsymbol{x}, 0) > 0\} dF(\varepsilon, \upsilon | y, \boldsymbol{x}, n) \\ &\quad = \boldsymbol{x}'\boldsymbol{\delta} - [\pi + \gamma]n - \boldsymbol{x}_2'\boldsymbol{\beta} + \upsilon - \varepsilon. \end{aligned} \tag{10}$$

With only data on choices, one can identify $\pi + \gamma$, $\boldsymbol{\delta}_1$, and $\boldsymbol{\delta}_2 - \boldsymbol{\beta}$, where $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are the partitions of δ associated, respectively, to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. With further data on wages for women who work, $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are separately identified using the standard Heckman selection approach. With wage data and exclusion restrictions, the parameters of Σ could also be identified (the correlation, from the coefficient of the Heckman selection model, the variance of ν from the variance in wages, and the variance of ε as the inverse of a coefficient that multiplies $\boldsymbol{x}_1\boldsymbol{\delta}_1$ in the discrete choice problem).

As for the identification goals, the elasticity of labor supply with respect to wages is only identified, again, if there are exclusion restrictions. The effect of husband's income is identified in this case (it is zero by assumption, indeed; other utility functions would lead to different effects). Finally, the effect of changing the cost of child care can also be identified in this context, even though only

$\gamma + \pi$ is identified. The reason is that childcare cost only enters the utility as an effect on consumption. In this context, therefore, the elasticity to childcare cost is equivalent to that of wages.

In sum, in all approaches we need an exclusion restriction to identify the extent to which participation reacts to wage changes. The parametric approach allows making predictions of the effect of exogenous variables on choices out of sample, whereas in the non-parametric approach, these predictions can only be made in sample. Finally, the structural approach allows the researcher to make counterfactual policy simulations, identifying the effect of childcare costs on participation.

IV. Human capital accumulation: Heckman, Lochner, and Taber (1998)

In the long tradition in the literature on wage inequality, partially discussed in Chapter 1, the canonical model typically takes a partial equilibrium approach. Heckman, Lochner, and Taber (1998) gave a new look to this issue providing a general equilibrium overlapping generations model of labor earnings, skill formation, and physical capital accumulation with heterogeneous human capital. These authors develop new methods for estimating the demand of unobserved human capital and to determine the substitution relationships between skills and capital consistent with general equilibrium. Using their estimated model, the authors find that a model of skill-biased technical change with a trend estimated from the aggregate technology is consistent with the observed increase in the college-high school wage gap as well as overall inequality, whereas immigration contributes little to rising inequality. In this sense, it provides a tool to correct for endogenous labor supply adjustments to skill-biased technical change, ignored in the aggregate production function applications reviewed in the previous chapter.

Let h_a denote the stock of human capital of a given individual at age a, b_a denote her assets, and $e \in \{S, U\}$ denote her educational level. Also let i_t denote the interest rate at calendar time t, and r_{et} the prices of skills in education group e at time t. Consider the following life-cycle maximization problem:

$$\begin{aligned} V(h_a, b_a, e, i_t, r_{et}) \equiv \max_{c,g} & \left\{ \frac{c^{1-\gamma}}{1-\gamma} + \beta V(h_{a+1}, b_{a+1}, e, i_{t+1}, r_{et+1}) \right\}, \\ \text{s.t. } & b_{a+1} \leq b_a[1 + (1-\tau)i_t] + (1-\tau) r_{et} h_a (1-g) - c, \end{aligned} \tag{11}$$

where γ is the coefficient of relative risk aversion, β is the discount factor, and τ is the proportional tax rate. Let ω denote individual ability (unobserved by the econometrician), and δ denote the depreciation rate of skills. On-the-job human capital of a person with education e and ability ω accumulates as:

$$h_{a+t}(\omega, e) = \omega g^{\eta_e} h_a(\omega, e)^{\psi_e} + (1-\delta)h_a(\omega, e), \tag{12}$$

with $0 < \eta_e < 1$ and $0 < \psi_e < 1$ for $e \in \{S, U\}$. The authors assume a discrete distribution for ω, with eight points of support (four observable types, denoted by k, and based on quartiles of AFQT test, two education groups). Individuals are assumed to have perfect foresight of future prices and interest rates in equilibrium (because there are no aggregate shocks in the economy). Individuals work until age a_R, when are forced to retire, and afterwards live until age $\bar{a}$ without perceiving labor income.

The decision of education is taken at the front end, comparing lifetime discounted utility and discounted direct cost of education π_e plus some non-pecuniary benefits expressed in present terms, ε_e, with $\varepsilon_S - \varepsilon_U \sim \mathcal{N}(\mu_k, \sigma)$. Therefore, the education decision is the outcome of the following discrete choice problem:

$$\max_e \left[V^E(\omega, e, t) - \pi_e + \varepsilon_e\right], \tag{13}$$

where $V^E(\omega, e, t)$ is the lifetime discounted present value of education e when the individual is of ability ω entering in the labor market at cohort t.

Aggregate output Y_t is determined by the following nested CES technology:

$$Y_t = \left\{\alpha K_t^{\phi} + (1-\alpha)[\theta_t L_{St}^{\rho} + (1-\theta_t)L_{Ut}^{\rho}]^{\frac{\phi}{\rho}}\right\}^{\frac{1}{\phi}}. \tag{14}$$

Skill-biased technical change, determined by the evolution of θ_t is given by a time trend, similar to the seminal work by Katz and Murphy (1992), as we discussed in Chapter 1. In particular, the parameter θ_t is allowed to vary over time so that:

$$\ln\left(\frac{\theta_t}{1-\theta_t}\right) = \ln\left(\frac{\theta_0}{1-\theta_0}\right) + \varphi t. \tag{15}$$

The equilibrium in this economy is given by the sequence of interest rates $\{i_t\}_{t=0}^{\infty}$ and skill prices $\{r_{Ut}, r_{St}\}_{t=0}^{\infty}$ that clear the market subject that the aggregate firm maximizes profits, and workers maximize lifetime discounted utility.

The parameters to be estimated include the following: the risk aversion γ; the discount factor β; the tax rate τ; the points of support for the ability distribution, $\{\omega_{ek}\}_{k\in\{1,2,3,4\}}^{e\in\{S,U\}}$; the parameters of the human capital accumulation function $\{\{h_0(k,e)\}_{k\in\{1,2,3,4\}}, \eta_e, \psi_e\}_{e\in\{S,U\}}$ and δ; the tuition cost $\{\pi_e\}_{e\in\{S,U\}}$; the parameters of the distribution variance of the non-pecuniary costs $\{\{\mu_k\}_{k\in\{1,2,3,4\}}, \sigma\}$; and the parameters of the production function $\{\alpha, \theta_0, \varphi, \phi, \rho\}$. Given that the authors do not observe consumption in their data-set, they assume that β and γ take common values in the literature, assume that $\delta = 0$, and fix the value of τ to fit within the range of estimates in the literature. The tuition costs π_e are estimated as average values from the data. The estimation of the remaining parameters is carried in a step-wise fashion.

In the first step, the parameters of the aggregate production function are estimated. At old ages, say $a > a^*$ for some a^*, individuals no longer invest in human capital (that is, $g \approx 0$). Therefore:

$$w(a^* + 1, t + 1, h_{a^*+1}) \equiv r_{et+1} h_{a^*+1} = r_{et+1} h_{a^*}(1 - \delta), \tag{16}$$

which implies:

$$\frac{w(a^* + \ell, t + \ell, h_{a^*+\ell})}{w(a^*, t, h_{a^*})} = \frac{r_{et+\ell}(1 - \delta)^\ell}{r_{et}}. \tag{17}$$

Normalizing $r_{e0} = 1$, the sequence of skill prices is identified up to a scale $(1-\delta)^t$. Given these skill prices, the aggregate stocks of skill units can be recovered from the skill prices:

$$\frac{wage\ bill_{et}}{r_{et}(1 - \delta)^t} = \frac{L_{et}}{(1 - \delta)^t}. \tag{18}$$

As in Chapter 1, the relative demands of the two labor inputs give an expression for the relative skill prices:

$$\ln \frac{r_{St}}{r_{Ut}} = \frac{\theta_t}{1 - \theta_t} + (\phi - 1) \ln \frac{L_{St}}{L_{Ut}} = \ln \left(\frac{\theta_0}{1 - \theta_0} \right) + \varphi t + (\phi - 1) \ln \frac{L_{St}}{L_{Ut}}. \tag{19}$$

Given (18), the last term is identified, and θ_0, φ, and ϕ can be recovered from the coefficients of a linear regression. The remaining parameters of the aggregate production function are estimated in a similar manner.

In the second step, the parameters of the human capital production function are estimated by non-linear least squares. Note that the investment decision g is not observed, and, therefore, it has to be replaced by the solution of the dynamic problem. This solution is computed by backwards induction. In order to estimate $h_0(k, e)$, they parameterize $h_{a_R}(k, e)$ and recover the former from the estimates of the latter and the other parameters of the human capital production function.

Finally, in the third step estimation the authors estimate an auxiliary probit model in which they recover a non-parametric estimate of $\{(1 - \tau)[V^E(\omega, S, t) - V^E(\omega, U, t)] + \mu_k\}/\sigma$ and a coefficient associated to their estimated tuition costs π_e. Then, they recover the structural parameters from these estimates. The variance σ is recovered as the coefficient associated to π_e, and μ_k is recovered comparing the non-parametric estimates to the values predicted by the model (given the parameters estimated in steps one and two).

After estimation, the authors use the estimated model to decompose the sources of increasing wage inequality into different elements. To do so, they simulate the economy in transition from one steady state that represents mid-1970s to another in which the technology is permanently (and unexpectedly) shifted, progressively over 30 years, at a speed consistent with their estimate of φ. This shift makes the prices of skilled labor to increase and that of unskilled labor to decrease, leading to a college-high school wage gap increase, depressed initially by differential investments by high and low educated. At the beginning, high skill workers invest more because of the higher return, but later on, the opportunity cost of training becomes larger, and they slow down. This is especially true for young workers that enter in the labor market in the onset of the change. For them, the opportunity cost is too large, and they invest much less. This evidence in investment is linked to Katz and Murphy (1992), and to the later results in Card and Lemieux (2001) and Jeong, Kim, and Manovskii (2015), which we discussed in Chapter 1. In the second phase of the transition, a change in the extensive margin of investment also takes place, and more individuals go to college, even though they invest less on the job afterwards (as it is also the case for high school workers).

After this primary analysis, they present a battery of results that show the importance of different channels in explaining increasing wage inequality. In particular, they explore the effect of changes in demographics (a baby boom) and immigration of unskilled workers, finding evidence in favor of the former explaining most of the increase in wage inequality, finding a little role for the latter.