

# 1 Model and Estimation

This section introduces the model whose uncertainty is quantified and emphasizes the main economic, mathematical and computational aspects. It is the partial equilibrium, dynamic model of occupational choice developed in Keane and Wolpin (1994) (henceforth KW94). In their survey of dynamic discrete choice structural models, Aguirregabiria and Mira (2010) assign this model to the more general class of Eckstein-Keane-Wolpin models. I largely follow their notation to ease comparisons with other models and, most importantly, to ease the explanation of the estimation method. Besides applications to labour economics, Eckstein-Keane-Wolpin models are used to explain educational and occupational choices at the individual level.

The class of Eckstein-Keane-Wolpin models is structural. This means that, from the perspective of an econometrician, the model structure allows for the estimation of relationships between observable and unobservable variables. This requires the model to be solved for the agents' policy. The policy is the set of rules which describe the agents' optimal behaviour. The relationships between observables and unobservables are governed by exogenous parameters. These parameters may, for example, be utility parameters or distributional parameters which describe the processes of unobserved shocks. Therefore, the exogenous parameters can be estimated given a dataset of observable endogenous variables. Besides the observable states, the observable endogenous variables may also comprise of other parameters like, for instance, payoffs. Estimates for the exogenous parameters allow to use simulations (of states) in order to analyse counterfactual policy scenarios. These policies are represented by changes in some exogenous parameters. For instance, Keane and Wolpin (1997) obtain the following two results based on data from the NLSY79: First, unobserved heterogeneity in the endowment at age sixteen accounts for almost 90% of the variance in lifetime utility whereas shocks to productivity explain 10%. Second, a college tuition subsidy of 2,000 USD increases high school and college graduation by 3.5% and 8.4%, respectively.

As the research code for Keane and Wolpin (1997) is currently in alpha-version, this thesis studies the predecessor model in KW94. The main differences are that the model in KW94 does not contain unobserved permanent agent heterogeneity in endowment and that its choice-specific utility functions feature fewer covariates. This difference in complexity implies a decrease of the computational burden for the UQ but also a worse fit to the data. In fact, this thesis does not use estimates from real data but estimates from simulated data based on arbitrary parameters from KW94.

The section proceeds as follows: First, I introduce the KW94 model specification embedded in the more general Eckstein-Keane-Wolpin framework. This embedding provides additional context to the reader. In the next step, the estimation method simulated

maximum likelihood is presented. This approach is used for the structural estimation of the exogenous model parameters. After remarks on the numerical implementation, I show the estimation results. They include the estimates, the standard errors, and the correlations for all parameters. These results constitute the mean vector and the covariance matrix, which are used to characterize the joint input distribution for the UQ in the next section. The section ends by describing the QoI choice.

## 1.1 Keane and Wolpin (1994)

Aguirregabiria and Mira (2010) define Eckstein-Keane-Wolpin models by four characteristics. The first characteristic is that these models allow for permanent unobserved heterogeneity between agents. The simpler model by KW94 considered here does not use this option in contrast to Keane and Wolpin (1997). The other three characteristics are as follows:

1. Unobservable shocks  $\varepsilon_t$  do not have to be additively separable from the remainder of the utility functions.
2. Shocks  $\varepsilon_t$  can be correlated across choices  $a_t$ .
3. Observable payoffs, or wages,  $W_{a,t}^-$  are not conditionally independent from the unobservable shocks  $\varepsilon_t$  given the observable choices  $a_t$  and the observable part of the state vector  $\mathbf{s}_t^-$ . The reason is that wage shocks enter the wage function directly. If the agent decides against choices with observable payoffs that depend on  $\varepsilon_t$ , these payoffs can not be observed. Therefore, positive shocks and the observation of payoffs for the same choice are positively correlated.

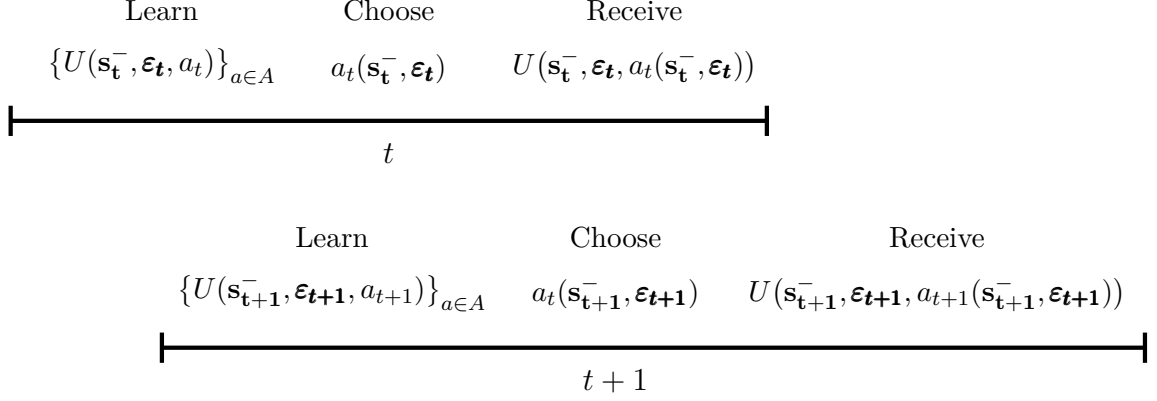
This paragraph describes the Eckstein-Keane-Wolpin model framework without permanent agent heterogeneity in the context of occupational choices as in KW94. In this setting, agents only differ in their draws of unobserved shocks  $\varepsilon_t$ .

In each period, a representative agent receives utility  $U$ . This utility depends on the state space and on choice  $a$  in period  $t \in \{0, 1, 2, \dots, T\}$ . Choices are mutually exclusive. The state space is the set of information in each period which is relevant for present and future utilities. It is split into an observable part  $\mathbf{s}_t^-$  and an unobservable part  $\boldsymbol{\varepsilon}_t$ . Choice  $a_t$  itself is also a function of the state space. This function is the decision rule, or policy, under which the rational agent chooses his utility for period  $t$ . For convenience, or to view utility and choices from different angles,  $U$  is denoted as function of only the states, as a function of  $a_t$  and the states, or as a function of function  $a_t(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t)$  and the states.

For some occupation alternatives, utility and prior decisions may be intertemporally connected: Agents receive a higher utility if they accumulated skills in past occupations that are useful for these alternatives. Other occupations may not reward experience. The observable part of the state space comprises the period, the work experience and the

choice in the previous period. The unobservable part of the state space consists of the alternative-specific shocks  $\{\varepsilon_{a,t}\}_{a \in A}$ . Figure 1 depicts the series of events.

**Figure 1.** Series of events



At the beginning of each period  $t$ , the agent recognizes the reward shocks  $\{\varepsilon_{a,t}\}_{a \in A}$  (as opposed to the observer), and the shocks become part of the unobserved state space  $\boldsymbol{\varepsilon}_t$ . Thus, the alternative-specific utilities  $\{U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t)\}_{a \in A}$  are known to the agent in period  $t$ . However, he can only form expectations about rewards in the future as the alternative-specific shocks  $\{\varepsilon_{a,t}\}_{a \in A}$  are stochastic. The specification in KW94 assumes the rewards shocks  $\{\varepsilon_{a,t}\}_{a \in A}$  to be serially uncorrelated. Therefore, prior shocks do not enter the state space. Next, the agent chooses his occupation  $a_t$  based on the state space information and according to his policy rule  $a_t(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t)$ . Then he receives the occupation-specific reward. The reward can be written as a function composition of utility and policy. This flow repeats for each  $t < T$ . The computation of the optimal policy is sketched in the next paragraph.

Agents are rational and forward-looking. Future utilities are subject to time discount factor  $\delta \in [0, 1]$ . Hence, they choose their optimal sequence of occupations by maximizing the remaining expected, discounted life-time utility. This maximal value is given by value function  $V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t)$  in (1). Like utility  $U$ , I also write value function  $V$  with different emphasis on occupation choice  $a_t$ .

$$V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t) = \max_{\{a\}_{t=0}^T} \left\{ \sum_{t=0}^T \delta^t \int_{\boldsymbol{\varepsilon}_t} U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t) f(\boldsymbol{\varepsilon}_t) d^{|\mathcal{A}|} \boldsymbol{\varepsilon}_t \right\} \quad (1)$$

Value  $V$  depends directly on time  $t$  because  $T$  is finite. Together with the discount factor  $\delta$ , this typically induces life-cycle behaviour. For example, agents invest more in the earlier time periods and work (and consume) more in the following periods. As  $\{\varepsilon_{a,t}\}_{a \in A}$  are the only random parameters and serially independent, the expectation of  $U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t)$  is given by the  $|\mathcal{A}|$ -dimensional integral of  $U$  multiplied by the joint probability density function  $f(\boldsymbol{\varepsilon}_t)$  with respect to  $\boldsymbol{\varepsilon}_t$ .  $|\mathcal{A}|$  denotes the number of occupation choices.

Coursely sketched, the approach to solve the above maximization problem is given by the dynamic programming problem characterized by the Bellman equation (Bellman (1957))<sup>1</sup> that breaks up the problem in (1) into more tractable sub-problems along the time dimension.

$$V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t) = \max_{a_t} \left\{ U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t) + \delta \int_{\boldsymbol{\varepsilon}_{t+1}} \max_{a_{t+1}} V(\mathbf{s}_{t+1}^-, \boldsymbol{\varepsilon}_{t+1}, a_{t+1}) f(\boldsymbol{\varepsilon}_{t+1}) d^{|\mathcal{A}|} \boldsymbol{\varepsilon}_{t+1} \right\} \quad (2)$$

The Bellman equation shows, that solving for the whole sequence of policy functions  $\{a\}_{t=0}^T$  is equivalent to solving iteratively for each optimal, period-specific policy function  $a_t(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t)$ . For this purpose, choose  $a_t$  for each period such that the current period utility and the discounted expected future lifetime utility (given the optimal choice of  $a_{t+1}$ ) are maximized. The finite time horizon eases the problem as the value function for the last period  $T$  simplifies to  $V(\mathbf{s}_T^-, \boldsymbol{\varepsilon}_T) = \max_{a_T} U(\mathbf{s}_T^-, \boldsymbol{\varepsilon}_T, a_T)$ .<sup>2</sup> With this condition, the problem can be solved for all states by iterating backwards: First, one solves for the final period policy  $a_T(\mathbf{s}_T^-, \boldsymbol{\varepsilon}_T)$ . Then this sub-result is plugged in the future value function on the right hand side of (2) to solve for  $a_{T-1}(\mathbf{s}_{T-1}^-, \boldsymbol{\varepsilon}_{T-1})$ , and so forth until  $t = 1$ . By emphasizing time as a dimension, the policies can also be summarized as one function  $a(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t)$ . Given random draws for the unobservable shocks  $\boldsymbol{\varepsilon}_t$  for each period, this policy is used to simulate the occupational paths for a number of agents.

This paragraph addresses the alternative-specific utility functions  $\{U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t)\}_{a \in \mathcal{A}}$  that finally pin down the model in KW94.

There are four different occupations,  $b$ ,  $w$ ,  $e$  and  $h$ , of which occupations  $b$  and  $w$  are defined by the same type of utility function. In the following, I will roughly explain how the first two utility functions model characteristics for working in the blue- and in the white-collar sector and how the latter two equations sketch receiving institutional education and staying at home. The parametrization that distinguishes the blue from the white-collar sector and additional intuition is given later in subsection Estimation Results and in Table 1.

It is assumed that there is a direct mapping from USD to utility. Based on this, the utility functions for occupation  $b$  and  $w$ ,  $U_b$  and  $U_w$ , equal the occupation-specific wage,  $W_{b,t}$  and  $W_{w,t}$ , in USD. The wage equations are given by the Mincer equation for earnings (Mincer (1958)):

$$\begin{aligned} U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, b) &= W_{b,t}^- = \exp \left\{ \beta^b + \beta_e^b x_{e,t} + \beta_b^b x_{b,t} + \beta_{bb}^b x_{b,t}^2 + \beta_w^b x_{w,t} + \beta_{bw}^b x_{w,t}^2 + \varepsilon_{b,t} \right\} \\ U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, w) &= W_{w,t}^- = \exp \left\{ \beta^w + \beta_e^w x_{e,t} + \beta_w^w x_{w,t} + \beta_{ww}^w x_{w,t}^2 + \beta_b^w x_{b,t} + \beta_{bw}^w x_{b,t}^2 + \varepsilon_{w,t} \right\} \end{aligned} \quad (3)$$

<sup>1</sup>For more details, see Raabe (2019), p. 9-19.

<sup>2</sup>More precisely, a finite time horizon in contrast to an infinite time horizon implies that the solution does not require, first, to guess the future value function for an arbitrary last time period, and, second, to iterate backwards in time to obtain a converged value function. On the other hand, the finite time horizon complicates the solution, because it requires one policy function for each time period vice versa solely one policy function for the converged value function of the infinite horizon problem.

Both equations comprise of a constant term, years of schooling  $x_{e,t}$ , linear and quadratic terms of occupation experience, and cross-occupational experience and the respective shocks in  $\boldsymbol{\varepsilon}_t$ .  $\boldsymbol{\beta}$  is the vector of coefficients that multiply the previously defined terms.<sup>3</sup> These coefficients are called covariates by many structural economists.

The utilities for education, or schooling, and staying at home are given by the functions in (4). These functions are also called non-pecuniary rewards.

$$\begin{aligned} U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, e) &= \beta^e + \beta_{he}^e \mathbf{1}(x_{e,t} \geq 12) + \beta_{re}^e (1 - \mathbf{1}(a_{t-1} = e)) + \varepsilon_{e,t} \\ U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, h) &= \beta^h + \varepsilon_{h,t} \end{aligned} \quad (4)$$

$\beta^e$  is the consumption reward of schooling. Function  $\mathbf{1}(x_{e,t} \geq 12)$  indicates whether an agent has completed high school.  $\beta_{he}^e$  is the tuition fee on higher or post-secondary education and  $\beta_{re}^e$  is an adjustment cost for returning to school when the agent chose another occupation the previous period ( $a_{t-1} \neq e$ ).  $\beta^h$  is the mean reward for staying at home.

Write  $\{\varepsilon_{a,t}\}_{a \in A}$  as vector  $\boldsymbol{\varepsilon}_{a,t}$ . It is assumed that  $\boldsymbol{\varepsilon}_{a,t}$  follows a joint normal distribution, such that  $\boldsymbol{\varepsilon}_{a,t} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\varepsilon)$ .  $\boldsymbol{\Sigma}_\varepsilon$  denotes the covariance matrix for shocks  $\boldsymbol{\varepsilon}_{a,t}$ .  $\sigma_a^2$  and  $\sigma_{a(j),a(k \neq j)}^2$  denote the alternative-specific variances and covariances in  $\boldsymbol{\Sigma}_\varepsilon$ . Shocks are serially uncorrelated. Indices  $\{j, k\} \in \mathbb{N}$  are used to denote subsets of  $a$ .

Finally, there is a bijective mapping from periods  $t$  to age 16 to 65. The next subsection describes the estimation method.

## 1.2 Simulated Maximum Likelihood Estimation

To estimate the exogenous model parameters, the approach that this thesis and also KW94 use is the simulated maximum likelihood method (Albright et al. (1977))<sup>4</sup>.

This method can be applied to a set of longitudinal data on occupational choices  $a_t$  and, if available, wages  $W_{a,t}^-$  of a sample of  $i \in I$  individuals starting from age 16. To distinguish from its functional form, let  $\mathcal{W}_{a(k),t}^-$  henceforth denote the measured wages. For each period  $t$ , the recorded choices  $a_0, \dots, a_{t-1}$  imply the occupation-specific experiences  $x_{a,t}$ . Together with  $t$ , they constitute the observable state vector  $\mathbf{s}_t^-$ . Consequently, the measured, observable endogenous variables are  $\mathcal{D} \stackrel{\text{def}}{=} (\mathbf{s}_t^-, \mathcal{W}_{a,t}^-)$ . Given this setup, the goal is to estimate the exogenous model parameters  $\boldsymbol{\theta} \stackrel{\text{def}}{=} (\delta, \boldsymbol{\beta}, \boldsymbol{\Sigma}_\varepsilon)$ .<sup>5</sup> Thus, in the following, every probability is a function of the exogenous model parameters  $\boldsymbol{\theta}$ . The approach to compute the likelihood function  $L_{\mathcal{D}}(\boldsymbol{\theta})$  of the observables in the data begins with the

<sup>3</sup>The notation for  $\boldsymbol{\beta}$  includes two references. The superscript indicates the occupation-specific utility that contains the coefficients. The subscript indicates the occupation-specific experience or abbreviates the condition that regulates the coefficients. Thus, coefficients for constant terms do not have a subscript. Twice the respective subscript mark coefficients for quadratic terms.

<sup>4</sup>See Aguirregabiria and Mira (2010), p. 42-44 and Raabe (2019), p. 21-26 for more details.

<sup>5</sup>Improvements in this thesis' estimation over KW94 are that, first, it is not assumed that the standard errors of the parameters estimates are uncorrelated, and, second, that  $\boldsymbol{\beta}$  is not left out of the estimation.

individual latent variable representation in period  $t$ .

$$a_t = \underset{a}{\operatorname{argmax}} V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t) \quad (5)$$

As  $a_t$  and  $\mathbf{s}_t^-$  are known, the next step is to derive the unobservable shocks  $\boldsymbol{\varepsilon}_t$  in terms of  $a_t$  and  $\mathbf{s}_t^-$ . Therefore, write the set of shocks for which the alternative-specific value function  $V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t(j))$  is higher than the other value functions  $V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t(k \neq j))$  as

$$\boldsymbol{\varepsilon}_t(a_t(j), \mathbf{s}_t^-) \stackrel{\text{def}}{=} \{\boldsymbol{\varepsilon}_t | V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t(j)) = \max_{a_t} V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t)\}. \quad (6)$$

Note that the set condition is a function of the unobservable model parameters  $\boldsymbol{\theta}$ .

Consider first the case of non-working alternatives  $a_t(j) \in [e, h]$ . The probability of choosing  $a_t(j)$  is the probability of set  $\boldsymbol{\varepsilon}_t(a_t(j), \mathbf{s}_t^-)$ . This probability equals the integral of the probability distribution function  $f(\boldsymbol{\varepsilon}_t)$  over all elements of set  $\boldsymbol{\varepsilon}_t(a_t(j), \mathbf{s}_t^-)$  with respect to  $\boldsymbol{\varepsilon}_t$ . Formally,

$$p(a_t(j) | \mathbf{s}_t^-) = \int_{\boldsymbol{\varepsilon}_t(a_t(j), \mathbf{s}_t^-)} f(\boldsymbol{\varepsilon}_t) d^{|\mathcal{A}|} \boldsymbol{\varepsilon}_t. \quad (7)$$

The second case is  $a_t(k) \in [b, w]$ . Assuming the dataset contains wages for the working alternatives  $a_t(k)$ , the probabilities of choosing  $a_t(k)$  take a few steps more to compute. First, note from the wage equations that the alternative-specific shocks  $\boldsymbol{\varepsilon}_{\mathbf{a}, t}$  are log normally distributed. Second, in contrary to the non-working alternatives, using (3), the shocks can directly be expressed as a function of the alternative-specific model parameters  $\beta_{\mathbf{a}(k)}$  by inserting the inferred alternative-specific experiences  $\mathbf{x}_{\mathbf{a}(k), t}$  into  $W_{a(k), t}$  and subtracting the expression from the observed wage  $\mathcal{W}_{a(k), t}^-$  for each individual. Both wages are logarithmized. Thus,

$$\varepsilon_{a(k), t} = \ln(\mathcal{W}_{a(k), t}^-) - \ln(W_{a(k), t}^-). \quad (8)$$

Third, the alternative-specific shocks  $\boldsymbol{\varepsilon}_{\mathbf{a}, t}$  are not distributed independently. Since  $\varepsilon_{a(k), t}$  can be inferred from the observed wage  $\mathcal{W}_{a(k), t}^-$ , this information can be used to form the expectation about the whole error distribution. Therefore, using the conditional probability density function  $f(\boldsymbol{\varepsilon}_t | \varepsilon_{a(k), t})$ , the probability of choosing occupation  $a_t(k)$  conditional on observed states and wages writes

$$p(a_t(k) | \mathbf{s}_t^-, W_{a(k), t}^-) = \int_{\boldsymbol{\varepsilon}_t(a_t(k), \mathbf{s}_t^-)} f(\boldsymbol{\varepsilon}_t | \varepsilon_{a(k), t}) d^{|\mathcal{A}|} \boldsymbol{\varepsilon}_t. \quad (9)$$

Applying integration by substitution yields the following expression for the probability of

the observed wage:<sup>6</sup>

$$p(\mathcal{W}_{a(k),t}^- | \mathbf{s}_t^-) = \omega_t^{-1} \frac{1}{\sigma_{a(k)}} \phi\left(\frac{\varepsilon_{a(k),t}}{\sigma_{a(k)}}\right). \quad (10)$$

Here,  $\omega_t^{-1}$  is the Jacobian of the transformation from observed wage  $\mathcal{W}_{a(k),t}^-$  to error  $\varepsilon_{a(k),t}$  in (8) and  $\phi$  is the standard normal probability density function. Finally, the joint probability of observing choice  $a_t(k)$  and wage  $\mathcal{W}_{a(k),t}^-$  conditional on the observed states is given by the product of the two probabilities in (9) and (10):

$$p(a_t(k), \mathcal{W}_{a(k),t}^- | \mathbf{s}_t^-) = p(a_t(k) | \mathbf{s}_t^-, \mathcal{W}_{a(k),t}^-) p(\mathcal{W}_{a(k),t}^- | \mathbf{s}_t^-) \quad (11)$$

Based on these results, the likelihood contribution of one individual  $i$  can be written as the product of the probability to observe the measured endogenous variables for one individual and for one period over all time periods:

$$L_{\mathcal{D}}^i(\boldsymbol{\theta}) = P(\{a_t^i, \mathcal{W}_{a,t}^{-,i}\}_{t=0}^T) = \prod_{t=0}^T p(a_t^i, \mathcal{W}_{a,t}^{-,i} | \mathbf{s}_t^{-,i}) \quad (12)$$

Therefore, the sample likelihood is given by the product of the individual likelihoods over the whole sample of individuals:

$$L_{\mathcal{D}}(\boldsymbol{\theta}) = P(\{\{a_t^i, \mathcal{W}_{a,t}^{-,i}\}_{t=0}^T\}_{i \in I}) = \prod_{i \in I} \prod_{t=0}^T p(a_t^i, \mathcal{W}_{a,t}^{-,i} | \mathbf{s}_t^{-,i}) \quad (13)$$

Since the probabilities are functions of the exogenous parameters  $\boldsymbol{\theta}$ , the simulated maximum likelihood estimator  $\hat{\boldsymbol{\theta}}$  is the vector of exogenous parameters that maximizes (13). As maximum likelihoods estimates are asymptotically normal<sup>7</sup>, these results are taken as the mean vector for the input parameters in the uncertainty quantification.

The procedure to estimate the parameter vector  $\boldsymbol{\theta}$  using the expressions for the likelihood is as follows: First, The optimization algorithm of choice proposes a parameter vector. Second, the model is solved via backward induction. Third, using the policy functions, the likelihood is computed. These steps are repeated until the optimizer has found the parameter vector that yields the maximal likelihood.

Finally, the calculation of the estimator's covariance is described.<sup>8</sup> The result is used as the covariance matrix for the input parameters in the UQ.

The asymptotic covariance of a maximum likelihood estimator equals the inverse of the Fisher information matrix:  $\text{Var}(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})^{-1}$ . In this thesis, the information matrix

<sup>6</sup>See Raabe (2019), p. 29, 39-40 for the complete derivation.

<sup>7</sup>This property is an advantage of this thesis' estimation approach. It facilitates the uncertainty quantification via Monte Carlo sampling because there is a simple closed form for the (marginal) probability density available. This eases the construction of the desired samples.

<sup>8</sup>See Verbeek (2012), p. 184-186.

$\mathcal{I}(\boldsymbol{\theta})$  is given by the variance of the scores of the parameters.<sup>9</sup> The scores  $\mathbf{s}(\boldsymbol{\theta})$  are the first derivatives of the likelihood function. This can be written in terms of sample and individual likelihoods. Formally, the relationships are given by

$$\mathbf{s}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\partial L_{\mathcal{D}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i \in I} \frac{L_{\mathcal{D}}^i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \sum_{i \in I} \mathbf{s}_i(\boldsymbol{\theta}). \quad (14)$$

Having multiple individual likelihood contributions, the scores are in the form of the Jacobian matrix. Using the property that the expected values of scores,  $\mathbb{E}[\mathbf{s}(\boldsymbol{\theta})]$ , are zero at the maximum likelihood estimator, the variance of the scores is given by (15). It is equal to the inverse of the Fisher information matrix.

$$\mathcal{I}^{-1}(\boldsymbol{\theta}) = \text{Var}(\mathbf{s}(\boldsymbol{\theta})) = \mathbb{E}[\mathbf{s}(\boldsymbol{\theta})\mathbf{s}(\boldsymbol{\theta})']. \quad (15)$$

Hence, the estimator for the asymptotic covariance of the maximum likelihood estimator is given by

$$\hat{\text{Cov}}_J(\hat{\boldsymbol{\theta}}) = \left( \frac{1}{|I|} \sum_{i \in I} \mathbf{s}_i(\hat{\boldsymbol{\theta}})\mathbf{s}_i(\hat{\boldsymbol{\theta}})' \right)^{-1}. \quad (16)$$

$|I|$  is the number of individuals in the data set. The intuition behind the above expression is the following: Estimator  $\hat{\boldsymbol{\theta}}$  maximizes the sample likelihood. This is equivalent to  $\hat{\boldsymbol{\theta}}$  setting the sample scores to zero. However, the individual likelihood may not be zero at the optimal parameter vector for the sample likelihood. This variation is captured by the variance of the individual scores evaluated at  $\hat{\boldsymbol{\theta}}$ . The relations in (14) and (15) then imply that the inverse of the variance of the individual scores is equivalent to the variance of the maximum likelihood estimator.

### 1.3 Numerical Implementation

Besides the standard python libraries, the thesis uses the packages *respy* and *estimagic* to compute the QoI and to estimate the distribution of the input parameters. All other programs can be found in the *Master's Thesis Replication Repository*.

As standard deviations  $\sigma_a$  are restricted to positive numbers, drawing them from the unrestricted estimated joint normal distribution can lead to false results because it is possible to draw parameter values outside of their domain. Therefore, covariance matrix  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$  is written in terms of the lower triangular matrix  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{\mathbf{c}}$  obtained from the Cholesky decomposition of  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ . The contained Cholesky factors are unrestricted and denoted by  $c_i$  and  $c_{i,j}$ .  $i$  and  $j$  are positional indices. Hence, estimates for the Cholesky parameters

---

<sup>9</sup>The computation of  $\text{Cov}(\boldsymbol{\theta})$  by using the Jacobian of the individual likelihood contributions is chosen over other approaches because, first, it yields no error in the inversion step of  $\mathcal{I}(\boldsymbol{\theta})$  and, second, the results are reasonably close to the similar specification in KW94.



and their variation replace the respective estimates for  $\Sigma_{\epsilon}$  in parameter vector  $\theta$  and its covariance matrix  $\text{Cov}(\theta)$  that are presented in the previous subsection.

## 1.4 Estimation Results

This subsection presents estimates  $\hat{\theta}$  for the exogenous parameters and the standard errors  $\text{SE}(\hat{\theta})$ . It also shows the correlations between important estimates.

The second column in Table 1 contains the estimates for the exogenous model parameters  $\theta$ . They are obtained from a simulated dataset of 1000 individuals based on the arbitrary parametrization that is used in Data Set One in KW94.<sup>10</sup> This parametrization has the following economic implications: Occupation in the white-collar sector is more skill-intensive or, more technically, has higher returns to education and occupational experience than occupation in the blue-collar sector. Moreover, experience in the blue-collar sector is rewarded in the white-collar sector but not vice versa. Under this specific parametrization, the diagonal elements of the lower triangular matrix  $c_i$  coincide with the standard deviations of the utility shocks  $\epsilon_{a,t}$  and the non-diagonal elements  $c_{i,j}$  equal the correlations between different alternative-specific shocks  $\epsilon_{a,t}$ . The parameter estimates  $\hat{\theta}$  are precise. This means they equal the parameters with which the model is simulated.

The third column shows this thesis' estimates of the standard errors  $\text{SE}(\hat{\theta})$ . The fourth column shows the standard errors computed in KW94. Given the differences between both estimation specifications, namely the inclusion of  $\beta$  and correlations between standard errors in this thesis, the estimates are reasonably similar. However, the one exception that stands out is the results for the non-diagonal Cholesky factors  $c_{i,j}$ .

I argue that this thesis' estimates are more precise than the estimates in KW94, and therefore, it is correct to use them in the subsequent uncertainty quantification. This claim is based on two reasons. These indicate that KW94, in fact, do not estimate the Cholesky factors, but the standard errors and correlations of shocks  $\epsilon_{a,t}$ . Both expressions are equal for the parametrization in Table 1. Nonetheless, they are conceptually different. Therefore, measures for their variation, which, by construction, also consider parameter values other than the mean estimates, have to differ. The arguments are: First, own estimates of  $\text{SE}(\hat{\theta})$  for the model in terms of standard deviations and correlations of shocks  $\epsilon_{a,t}$  are close to those in KW94. Second, the estimates in KW94 for the non-diagonal elements  $c_{i,j}$ , except of the estimate for  $c_{1,2}$ , would be unlikely corner solutions. For instance, fix the variance for the shocks in  $U_e$ ,  $\sigma_e^2$ , and write this variance in terms of the Cholesky factors<sup>11</sup> such that  $\sigma_e^2 = c_{3,1}^2 + c_{3,2}^2 + c_3^2$ . There are no restrictions that can force the maximum likelihood estimator to attribute  $\sigma_e^2$  mainly to  $c_3$ . In fact and in line with the first argument, the

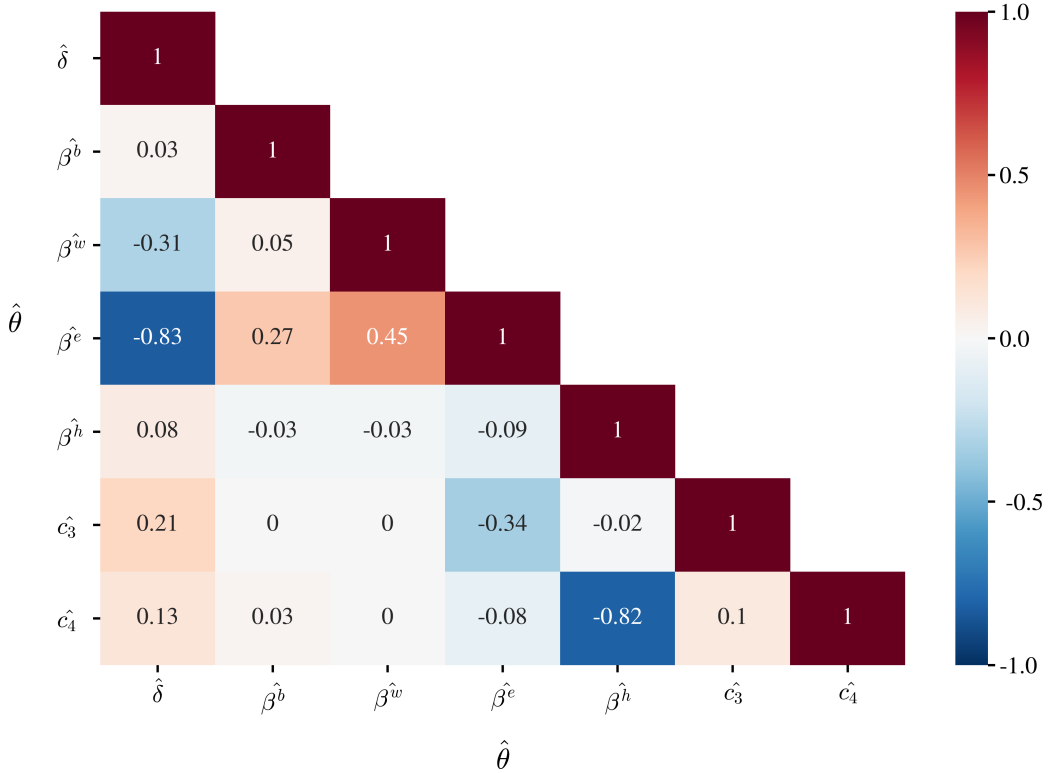
<sup>10</sup>See table 1, p. 658 in Keane and Wolpin (1994); In contrary to the computation of variation measures for  $\hat{\theta}$ , it is sufficient for obtaining  $\hat{\theta}$  to find the individual likelihood of the average agent instead of the sample likelihood in (13).

<sup>11</sup>In general,  $\sigma_{i,j}^2 = \sum_{n=1}^j c_{j,n} c_{i,n}$ .

size of the estimates for the standard errors of  $c_{i,j}$  in KW94 correspond to the size that one would expect for standard deviations of correlation coefficients that range from -1 to 1 and not necessarily for cholesky coefficients. These coefficients are unrestricted, and therefore their standard errors have not to be in this specific size. The previous arguments undermine the credibility of the estimation results in KW94, and as a consequence, the thesis proceeds with its own estimates.

Figure 2 depicts the correlations between the estimates of important parameters in  $\theta$ . In general, the share of high correlations is considerable. Thus, to allow for covariation between the standard errors of the parameter estimates is an important improvement over KW94. The coefficients that stand out are  $\text{corr}(\hat{\delta}, \hat{\beta}^e)$ ,  $\text{corr}(\hat{\delta}, \hat{\beta}^h)$ ,  $\text{corr}(\hat{\beta}^e, \hat{\beta}^w)$ ,  $\text{corr}(\hat{c}_3, \hat{\beta}^e)$  and  $\text{corr}(\hat{c}_4, \hat{\beta}^h)$  with  $-0.83$ ,  $-0.31$ ,  $0.45$ ,  $-0.34$  and  $-0.82$ , respectively.

**Figure 2.** Correlations between estimates for important input parameters



The intuition behind these results can be obtained from the following insight: Negative correlations imply similar effects, and positive correlations imply opposing effects on the likelihood of observed endogenous variables  $\mathcal{D}$ . For instance, consider an individual that decides for a long occupation in the education sector in the first years and then continues to work in the white-collar sector for the rest of his life. The likelihood to observe this individual increases when  $\delta$  rises because all individuals get more patient, and therefore, ceteris paribus, they invest more in education. However, the same likelihood also increases if the educational utility constant  $\beta^e$  rises. Hence, because they can compensate each other, the likelihood around the optimal parameter  $\hat{\theta}$  decreases less for changes of both

parameters in opposing directions than for changes in the same direction. Therefore, parameters  $\delta$  and  $\beta^e$  are negatively correlated in terms of the score function in (14) around  $\hat{\theta}$ . It follows from (16) that their standard errors are negatively correlated.

The above example provides intuition for  $\text{corr}(\hat{\delta}, \hat{\beta}^e) = -0.83$ . An analogous reasoning for the same example can explain  $\text{corr}(\hat{\delta}, \hat{\beta}^w) = -0.31$ . Yet, this correlation is smaller because  $U_{w,t}$  has less covariates than  $U_{e,t}$ .  $\text{corr}(\hat{c}_3, \hat{\beta}^e) = -0.34$  and  $\text{corr}(\hat{c}_4, \hat{\beta}^h) = -0.82$  can be explained by a similar argument: Individuals decide for occupation in education or home sector if the respective utilities are high. This can be achieved by high constant terms or by high positive shocks. The latter can only happen to some individuals if the Cholesky factors are large because these factors are components of the respective shock variance. Shocks  $\varepsilon_{a,t}$  are known to the agents prior to their decision  $a_t$ . Negative shocks have a smaller impact on choosing occupation  $e$  or  $h$  because individuals tend to decide against these alternatives anyway. Thus,  $c_3$  and  $\beta^e$ , and  $c_4$  and  $\beta^h$  can impact the likelihood in the same direction. Therefore, their standard errors are negatively correlated. Moreover, the latter relationship is stronger because  $U_{h,t}$  has a lower level and less covariates.

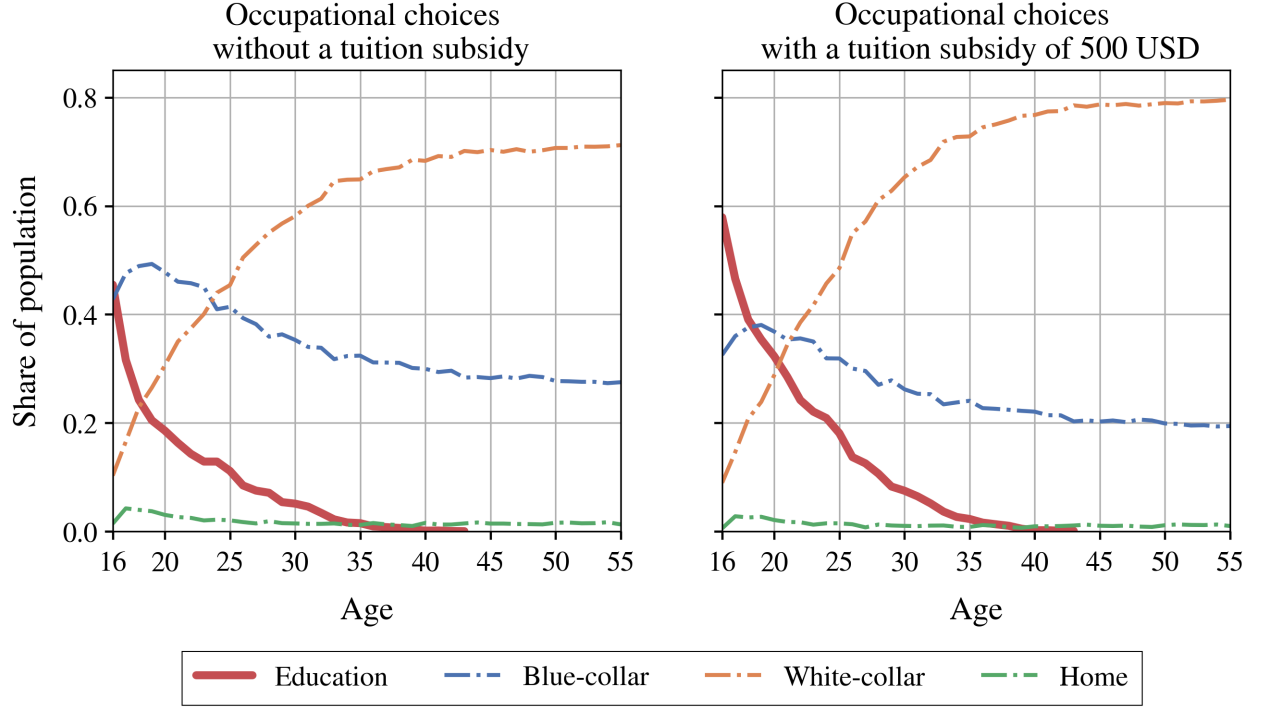
## 1.5 Quantity of Interest

The QoI is the effect of a 500 USD subsidy on annual tuition costs for higher education on the average years of education. Formally,  $\beta_{he}^{e,pol} = \beta_{he}^e - 500$ , where  $\beta_{he}^{e,pol}$  represents the subsidised tuition costs. In KW94, the effect is an increase of 1.44 years.<sup>12</sup> The same figure computed with *respy* is 1.5.

Figure 3 depicts a comparison between the shares of occupations in the different sectors for a sample of 1000 individuals over their relevant lifetime between two different scenarios. The left graph shows the occupation paths under baseline parametrization  $\hat{\theta}$  and the right graph the paths for the same model with subsidised tuition costs.

The red, blue, white, and green lines mark the shares of individuals occupied in the education, blue-collar, white-collar, and home sector, respectively. Both graphs show the typical life-cycle behaviour. Many agents tend to invest in their education early and continue in the white-collar sector as this sector rewards education. Another large group works in the blue-collar sector, and some of them switch to the white-collar sector, as well. This switch from white- to blue-collar is caused by an accumulated blue-collar experience that is also rewarded in the white-collar sector and by positive shocks. The home sector is relatively irrelevant because the participation therein is comparably low for all ages.

<sup>12</sup>See table 4, p. 668 in Keane and Wolpin (1994).

**Figure 3.** Comparison of occupation paths between scenarios

The QoI is the sum of the differences between the education shares at each age in the right and in the left graph as depicted by the red lines.<sup>13</sup> This is because the vertical axis can also be interpreted as the share of one year that the average agent is occupied in the education sector. Comparing both graphs, we can see that the tuition subsidy incentivises younger individuals to stay in the education sector for a longer time and older individuals to work in the white-collar sector. The latter observation is a consequence of the first because the white-collar sector rewards education.

The QoI, the impact of a 500 USD tuition subsidy for higher education on average schooling years, is chosen because it is relevant to society in many areas, for example, education, inequality, and economic growth. The discussion section expands on this point. The QoI's relevance allows me to illustrate the importance of UQ in economics in the context of political decisions.

The next section shows the results of the first part of the UQ, the uncertainty propagation.

<sup>13</sup>If the red lines would depict continuous functions instead of discrete points in time, the QoI would be the difference between the integrals of the education shares as a function of time in the policy and the base scenario.

**Table 1.** Estimates for the distribution of input parameters

Parameter	Mean	Standard error (SE)	SE in KW94
<i>General</i>			
$\delta$	0.95	0.000 84	-
<i>Blue-collar</i>			
$\beta^b$	9.21	0.013	0.014
$\beta_e^b$	0.038	0.0011	0.0015
$\beta_b^b$	0.033	0.000 44	0.000 79
$\beta_{bb}^b$	-0.0005	0.000 013	0.000 019
$\beta_w^b$	0.0	0.000 67	0.0024
$\beta_{ww}^b$	0.0	0.000 029	0.000 096
<i>White-collar</i>			
$\beta^w$	8.48	0.0076	0.0123
$\beta_e^w$	0.07	0.000 47	0.000 96
$\beta_w^w$	0.067	0.000 55	0.000 90
$\beta_{ww}^w$	-0.001	0.000 017	0.000 070
$\beta_b^w$	0.022	0.000 33	0.0010
$\beta_{bb}^w$	-0.0005	0.000 021	0.000 030
<i>Education</i>			
$\beta^e$	0.0	330	459
$\beta_{he}^e$	0.0	155	410
$\beta_{re}^e$	-4000	202	660
<i>Home</i>			
$\beta^h$	17 750	390	1442
<i>Lower Triangular Cholesky Matrix</i>			
$c_1$	0.2	0.0015	0.0056
$c_2$	0.25	0.0013	0.0046
$c_3$	1500	108	350
$c_4$	1500	173	786
$c_{1,2}$	0.0	0.0064	0.023
$c_{1,3}$	0.0	143	0.412
$c_{2,3}$	0.0	116	0.379
$c_{1,4}$	0.0	232	0.911
$c_{2,4}$	0.0	130	0.624
$c_{3,4}$	0.0	177	0.870

## References

- Aguirregabiria, V. and P. Mira (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics* 156(1), 38–67.
- Albright, R. S., S. Lerman, and C. F. Manski (1977). *Report on the Development of an Estimation Program for the Multinomial Probit Model*. Cambridge Systematics.
- Bellman, R. E. (1957). *Dynamic Programming*. Princeton, NJ: Princeton University Press.
- Gabler, J. (2019). *A Python Tool for the Estimation of (Structural) Econometric Models*. URL: <https://github.com/OpenSourceEconomics/estimagic>.
- Keane, M. P. and K. I. Wolpin (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence. *Review of Economics and Statistics* 76(4), 648–672.
- Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. *Journal of Political Economy* 105(3), 473–522.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy* 66(4), 281–302.
- NLSY79 (1990). *National Longitudinal Survey of Youth 1979 Cohort, 1979–1990 (Rounds 1–11)*. URL: <https://www.nlsinfo.org/content/cohorts/nlsy79>.
- Raabe, T. (2019). A unified estimation framework for some discrete choice dynamic programming models. Master’s thesis, Bonn Graduate School of Economics.
- respy (2019). *A Python package for the simulation and estimation of a prototypical finite-horizon dynamic discrete choice model based on Keane & Wolpin (1997)*. URL: <https://github.com/OpenSourceEconomics/respy>.
- Stenzel, T. (2020). *Master’s Thesis Replication Repository*. URL: <https://github.com/HumanCapitalAnalysis/thesis-projects-tostenzel>.
- Verbeek, M. (2012). *A Guide to Modern Econometrics* (4 ed.). Wiley.