1 Model and Estimation

This section introduces the economic model whose uncertainty is quantified. 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 model class is structural. This means that, from the perspective of an econometrician, the model structure allows for the estimation of relationships between observable and unobservable state variables. These relationships 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 using simulations (of states) in order to analyse counterfactual policy scenarios. Changes in some exogenous parameters represent these policies. For example, 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%. And 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 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 data simulated on arbitrary parameters choices that are taken from KW94.

The section proceeds as follows: First, I introduce the KW94 model specification embedded in the more general Eckstein-Keane-Wolpin framework. 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. These 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)

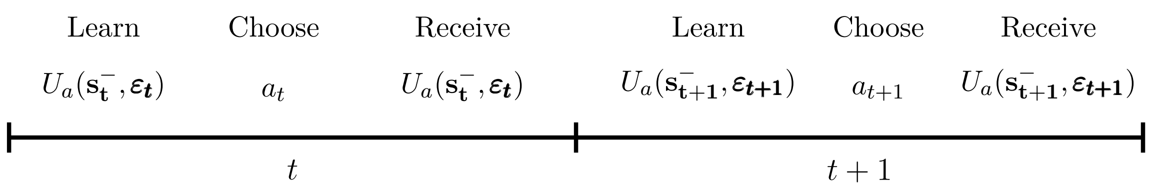
Aguirregabiria and Mira (2010) define Eckstein-Keane-Wolpin models by four characteristics. The first characteristic is that these models allow for permanent unobserved heterogeneity between agents. The simpler model by KW94 considered here does not use this option in contrast to Keane and Wolpin (1997).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

1. Unobservable shocks ε_t do not have to be additively separable from the remainder of the utility functions.
2. Shocks ε_t can be correlated across choices a_t.
3. Observable payoffs, or wages, $W^{-}_{a,t}$ are not conditionally independent on the unobservable shocks ε_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. The agent can observe this prior to his decision. This decision can then lead to the non-observation of alternative-specific wages.

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

A representative agent decides for action, or occupation, a_t from the set of alternatives A in each time period t. These alternatives are mutually exclusive. From each decision, agents obtain the alternative-specific utility U_a. Notation U_a indicates that the utility depends on occupation choice a_t. In each time period, Utilities U_a are subject to random shocks $\varepsilon_{a,t}$ which are also alternative-specific. 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. $\boldsymbol{S_t}$ denotes the state space. The state space is the set of information in each period t relevant for the present and future utilities for each occupation choice a_t. The observable part of the state space comprises the period, the work experience and the choice in the previous period. It is denoted by vector $\boldsymbol{s_t^-}$. The unobservable part of the state space is denoted by vector $\boldsymbol{\varepsilon_t}$ and consists of the alternative-specific shocks $\boldsymbol{\varepsilon_{a,t}}$. The sequence of events is depicted in Figure 1.

Figure 1. Sequence of events



At the beginning of each period t, the agent recognizes the reward shocks $\boldsymbol{\varepsilon_{a,t}}$ (as opposed to the observer), and the shocks become part of the unobserved state space $\boldsymbol{\varepsilon_t}$. Thus, the utilities $U_a(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t})$ are known to the agent in period t. However, he can only form expectations about rewards in the future as the alternative-specific shocks $\boldsymbol{\varepsilon_{a,t}}$ are stochastic. The specification in KW94 assumes the rewards shocks $\boldsymbol{\varepsilon_{a,t}}$ 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. Then he receives the occupation-specific reward U_a. This flow repeats for each $t < T$.

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 $V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t})$.

$$V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) = \max_{\{a\}_{t=0}^{T}} \left\{ \sum_{t=0}^{T} \delta^t \int_{\boldsymbol{\varepsilon_t}} U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t) f(\boldsymbol{\varepsilon_t}) d^{|A|} \boldsymbol{\varepsilon_t} \right\} \tag{1}$$

Value V depends directly on time t because T is finite. Together with the discount factor δ, 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 $\boldsymbol{\varepsilon_{a,t}}$ are the only random parameters and serially independent, the expectation of $U(\boldsymbol{s_t^-}, \boldsymbol{\varepsilon_t}, a_t)$ is given by the $|A|$-dimensional integral of U multiplied by the joint probability density function $f(\boldsymbol{\varepsilon_t})$ with respect to $\boldsymbol{\varepsilon_t}$. $|A|$ 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)).[1]

$$V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) = \max_{a_t} \left\{ U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t) + \delta \int_{\boldsymbol{\varepsilon_t}} \max_{a_{t+1}} V_{a_{t+1}}(\mathbf{s_{t+1}^-}, \boldsymbol{\varepsilon_{t+1}}) f(\boldsymbol{\varepsilon_{t+1}}) d^{|A|} \boldsymbol{\varepsilon_{t+1}} \right\} \tag{2}$$

The Bellman equation states, 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 $V(\mathbf{s_T^-}, \boldsymbol{\varepsilon_T}) = \max_{a_T} U(\mathbf{s_T^-}, \boldsymbol{\varepsilon_T}, a_T)$. With this condition the problem can be solved for all states by iterating backwards. Given initial states and random draws for the unobservable shocks $\boldsymbol{\varepsilon_t}$ for each period, these policy equations are used to simulate the occupational paths for a number of agents.

This paragraph addresses the alternative-specific utility functions $U_a(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t})$ that finally pin down the model structure in KW94. There are four different occupations, b, w, e and

[1] For more details, see Raabe (2019), p. 9-19.

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 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 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)):

$$\begin{aligned} U_b(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) &= W_{b,t}^- = \exp\Big\{\beta^b + \beta_e^b x_{e,t} + \beta_b^b x_{b,t} + \beta_{bb}^b x_{b,t}^2 + \beta_w^b x_{w,t} + \beta_{ww}^b x_{w,t}^2 + \varepsilon_{b,t}\Big\} \\ U_w(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) &= W_{w,t}^- = \exp\Big\{\beta^w + \beta_e^w x_{e,t} + \beta_w^w x_{w,t} + \beta_{ww}^w x_{w,t}^2 + \beta_b^w x_{b,t} + \beta_{bb}^w x_{b,t}^2 + \varepsilon_{w,t}\Big\} \end{aligned} \tag{3}$$

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 $\boldsymbol{\varepsilon_{a,t}}$. $\boldsymbol{\beta}$ is the vector of coefficients that multiply the previously defined terms.[2] These coefficieints are called covariates by many structural economists.

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

$$\begin{aligned} U_e(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) &= \beta^e + \beta_{he}^e \mathbf{1}(x_{e,t} \geq 12) + \beta_{re}^e(1 - \mathbf{1}(a_{t-1} = e)) + \varepsilon_{e,t} \\ U_h(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) &= \beta^h + \varepsilon_{h,t} \end{aligned} \tag{4}$$

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

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 $\boldsymbol{\varepsilon_{a,t}}$. σ_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 and k 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.

[2] 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. Coefficients for quadratic terms are marked by twice the respective subscript.

1.2 Simulated Maximum Likelihood Estimation

To estimate the exogenous model parameters, the approach that this thesis and also KW94 uses is the simulated maximum likelihood method (Albright et al. (1977))[3].

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, ..., 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 $\mathbf{m} \stackrel{\text{def}}{=} (\mathbf{s^-_t}, \mathcal{W}^-_{a,t})$. Given this setup, the goal is to estimate the exogenous model parameters $\boldsymbol{\theta} = (\delta, \boldsymbol{\beta}, \boldsymbol{\Sigma_\varepsilon})$.[4] Thus, in the following, every probability is a function of the exogenous model parameters $\boldsymbol{\theta}$. The approach to compute the likelihood function $L_{\mathbf{m}}(\boldsymbol{\theta})$ of the observables in the data begins with the individual latent variable representation in period t.

$$a_t = \underset{a}{\operatorname{argmax}}\, V_a(\mathbf{s^-_t}, \boldsymbol{\varepsilon_t}) \tag{5}$$

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 $V_{a(j)}$ is higher than the other value functions $V_{a(k\neq j)}$ in time t as

$$\boldsymbol{\varepsilon_t}(a_t(j), \mathbf{s^-_t}) \stackrel{\text{def}}{=} \{\boldsymbol{\varepsilon_t} | V_{a_t(j)}(\mathbf{s^-_t}, \boldsymbol{\varepsilon_t}) = \max_a V_a(\mathbf{s^-_t}, \boldsymbol{\varepsilon_t})\}). \tag{6}$$

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,

$$\mathrm{p}\Big(a_t(j)|\mathbf{s^-_t}\Big) = \int_{\boldsymbol{\varepsilon_t}(a_t(j), \mathbf{s^-_t})} f(\boldsymbol{\varepsilon_t}) d^{|A|}\boldsymbol{\varepsilon_t}. \tag{7}$$

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. In the first step, note from the wage equations that the the alternative-specific shocks $\boldsymbol{\varepsilon_{a,t}}$ 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 $\boldsymbol{\beta_{a(k)}}$ by inserting the inferred alternative-specific experiences $\boldsymbol{x_{a(k),t}}$ into $W_{a(k),t}$ and subtracting the expression from the observed wage $\mathcal{W}^-_{a(k),t}$ for each individual. Both

[3] see Aguirregabiria and Mira (2010), p. 42-44 and Raabe (2019), p. 21-26 for more details.

[4] 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 β is not left out of the estimation.

wages are logarithmized. Thus,

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

Third, the alternative-specific shocks $\boldsymbol{\varepsilon_{a,t}}$ are not distributed independently. Since $\varepsilon_{a(k),t}$ can be inferred from the observed wage $\mathcal{W}^{-}_{a(k),t}$, the 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

$$\mathrm{p}\Big(a_t(k)|\mathbf{s_t^-}, W^{-}_{a(k),t}\Big) = \int_{\boldsymbol{\varepsilon_t}(a_t(k),\mathbf{s_t^-})} f(\boldsymbol{\varepsilon_t}|\varepsilon_{a(k),t}) d^{|A|}\boldsymbol{\varepsilon_t}. \tag{9}$$

Applying integration by substitution yields the following expression for the probability of the observed wage:[5]

$$\mathrm{p}\Big(\mathcal{W}^{-}_{a(k),t}|\mathbf{s_t^-}\Big) = \omega_t^{-1}\frac{1}{\sigma_{a(k)}}\phi\Bigg(\frac{\varepsilon_{a(k),t}}{\sigma_{a(k)}}\Bigg) \tag{10}$$

Here, ω_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 ϕ 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):

$$\mathrm{p}\Big(a_t(k), \mathcal{W}^{-}_{a(k),t}|\mathbf{s_t^-}\Big) = \mathrm{p}\Big(a_t(k)|\mathbf{s_t^-}, \mathcal{W}^{-}_{a(k),t}\Big)\mathrm{p}\Big(\mathcal{W}^{-}_{a(k),t}|\mathbf{s_t^-}\Big) \tag{11}$$

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:

$$L^i_{\mathbf{m}}(\boldsymbol{\theta}) = P\Big(\{a^i_{t,}, \mathcal{W}^{-,i}_{a,t,}\}^T_{t=0}\Big) = \prod_{t=0}^{T} \mathrm{p}\Big(a^i_t, \mathcal{W}^{-,i}_{a,t}|\mathbf{s_t}^{-,\mathbf{i}}\Big) \tag{12}$$

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

$$L_{\mathbf{m}}(\boldsymbol{\theta}) = P\Big(\Big\{\{a^i_{t,}, \mathcal{W}^{-,i}_{a,t,}\}^T_{t=0}\Big\}_{i\in I}\Big) = \prod_{i\in I}\prod_{t=0}^{T} \mathrm{p}(a^i_t, \mathcal{W}^{-,i}_{a,t}|\mathbf{s_t}^{-,\mathbf{i}}) \tag{13}$$

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[6], these results are taken as the

[5]See Raabe (2019), p. 29 and p.39-40 for the complete derivation.

[6]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.

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 maximal likelihood.

Finally, the calculation of the estimator's covariance is described.[7] 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: $\text{Var}(\theta) = \text{I}(\theta)^{-1}$. In this thesis, the information matrix $\text{I}(\boldsymbol{\theta})$ is given by the variance of the scores of the parameters.[8] The scores $\text{s}(\boldsymbol{\theta})$ 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

$$\text{s}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\partial L_{\mathbf{m}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i \in I} \frac{L^i_{\mathbf{m}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \sum_{i \in I} \text{s}_i(\boldsymbol{\theta}). \tag{14}$$

Having multiple individual likelihood contributions, the scores are in the form of the Jacobian matrix. Using the property that the expected values of scores, $\mathbb{E}[\text{s}(\theta)]$, 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.

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

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

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

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.

[7] see Verbeek (2012), p. 184-186.

[8] 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 $I(\boldsymbol{\theta})$ and, second, the results are reasonably close to the similar specification in KW94.

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 σ_a are restricted to positive numbers, drawing them from the unrestricted estimated joint normal distribution can lead to false results. Therefore, covariance matrix $\boldsymbol{\Sigma_\epsilon}$ is written in terms of the lower triangular matrix $\boldsymbol{\Sigma_\epsilon^c}$ obtained from the Cholesky decomposition of $\boldsymbol{\Sigma_\epsilon}$. 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 and their variation replace the respective estimates for $\boldsymbol{\Sigma_\epsilon}$ in the specification presented in the previous subsection.

1.4 Estimation Results

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

The second column in Table 1 contains the estimates for the exogenous model parameters $\boldsymbol{\theta}$. They are obtained from a simulated dataset of 1000 individuals based on the arbitrary parametrization that is used in Data Set One in KW94.[9] 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 $\boldsymbol{\epsilon_{a,t}}$ and the non-diagonal elements $c_{i,j}$ equal the correlations between different alternative-specific shocks $\boldsymbol{\epsilon_{a,t}}$. The parameter estimates $\hat{\boldsymbol{\theta}}$ are precise. This means they equal the parameters with which the model is simulated.

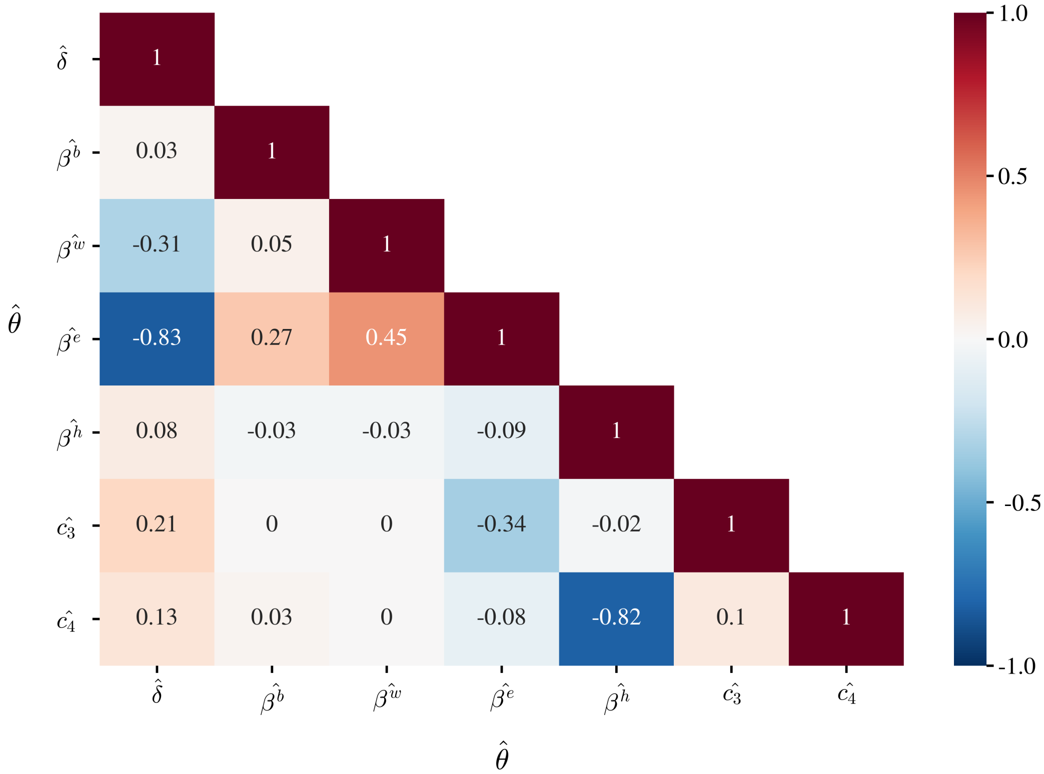
The third column shows this thesis' estimation of the standard errors $\text{SE}(\hat{\boldsymbol{\theta}})$. The fourth column shows the standard errors computed in KW94. Given the differences between both estimation specifications, namely the inclusion of β and correlations between standard errors in this thesis, the estimates are reasonably similar. However, the one exception that stands out is the results for the non-diagonal Choleksy 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 $\boldsymbol{\epsilon_{a,t}}$. 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

[9] see table 1, p. 658; In contrary to the computation of $\text{Corr}(\hat{\boldsymbol{\theta}})$, it is sufficient to find the average likelihood instead of the sample likelihood in (13).

than the mean estimates, have to differ. The arguments are: First, own estimates of SE($\hat{\boldsymbol{\theta}}$) for the model in terms of standard deviations and correlations of shocks $\boldsymbol{\varepsilon_{a,t}}$ 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, $\overline{\sigma_e^2}$, and write this variance in terms of the Cholesky factors such that $\overline{\sigma_e^2} = c_{3,1}^2 + c_{3,2}^2 + c_3^2$[10]. There are no restrictions that can force the maximum likelihood estimator to attribute σ_e^2 mainly to c_3. In fact and in line with the first argument, the 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. Cholesky coefficients, however, are unresticted, and therefore their standard errors have not to be in this specific size. These arguments undermine the credibility of the estimation results in KW94, and as a consequence, the thesis proceeds with the own estimates.

Figure 2 depicts the correlations between the estimates of important parameters in $\boldsymbol{\theta}$. In general, the share of high correlations is considerable. The coefficients that stand out are corr($\hat{\delta}, \hat{\beta^e}$), corr($\hat{\delta}, \hat{\beta^h}$), corr($\hat{\beta^e}, \hat{\beta^w}$), corr($\hat{c_3}, \hat{\beta^e}$) and 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 $\mathbf{m}$. For instance, consider an individual that

[10]In general, $\sigma_{i,j}^2 = \sum_{n=1}^{j} c_{j,n} c_{i,n}$.

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 δ 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 β^e rises. Hence, because they can compensate each other, the likelihood around the optimal parameter $\hat{\boldsymbol{\theta}}$ decreases less for changes of both parameters in opposing directions than for changes in the same direction. Therefore, parameters δ and β^e are negatively correlated in terms of the score function in (14) around $\hat{\boldsymbol{\theta}}$. It follows from (16) that their standard errors are negatively correlated.

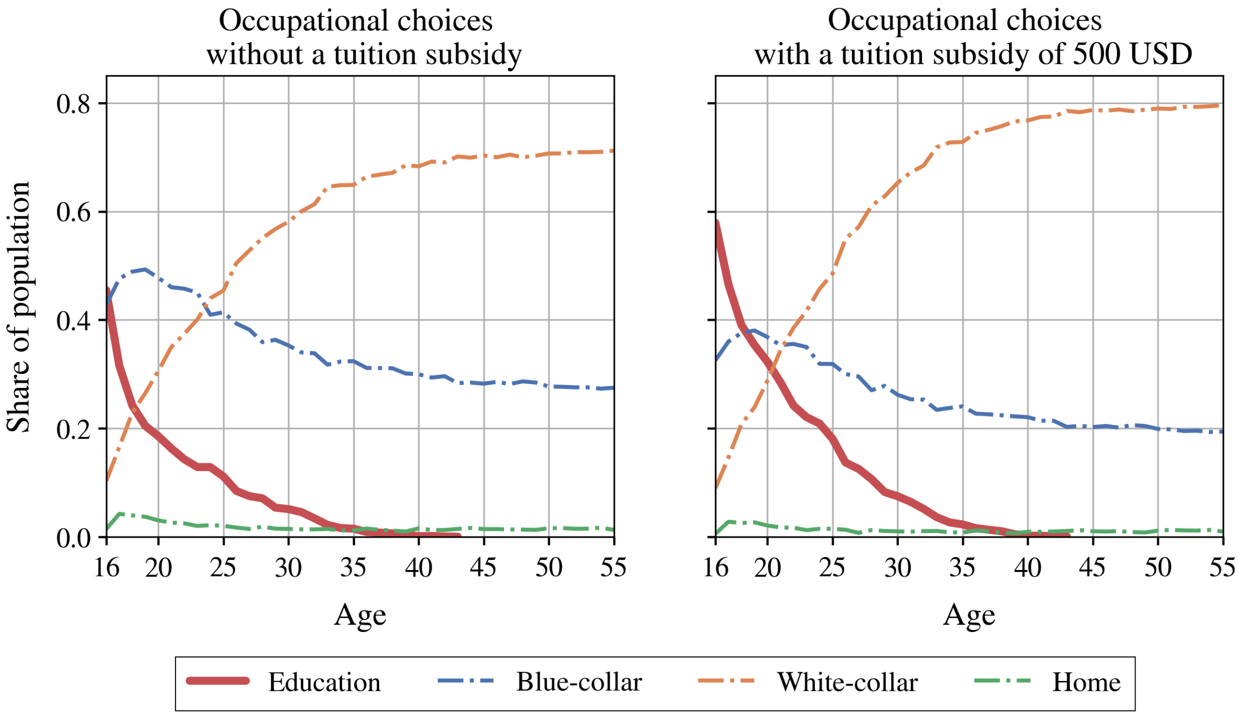
The above example provides intuition for $\text{corr}(\hat{\delta}, \hat{\beta^e}) = -0.83$ An analogous reasoning for the same example can explain $\text{corr}(\hat{\delta}, \hat{\beta^w}) = -0.31$. Yet, this correlation is smaller because $U_{w,t}$ has less covariates than $U_{e,t}$. $\text{corr}(\hat{c_3}, \hat{\beta^e}) = -0.34$ 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 because the factors are components of the respective shock variance. Shocks $\boldsymbol{\varepsilon_{a,t}}$ 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 β^e, and c_4 and β^h can impact the likelihood in the same direction. And 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, $\beta_{he}^{e,pol} = \beta_{he}^{e} - 500$. In KW94, the effect is an increase of 1.44 years (see table 4, p. 668). 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 scenarios. The left graph shows the occupation paths under baseline parametrization $\hat{\theta}$ and the right graph the paths for the same model with tuition subsidy.

Figure 3. Comparison of occupation paths between scenarios



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. Another large group works in the blue-collar, and some of them switch to the white-collar sector, as well. This switch is caused by an accumulated experience that is also rewarded in the white-collar sector and by positive shocks. As the relative participation in the home sector is low, this sector is relatively irrelevant for all ages. The QoI is the difference between the red line in the left graph and in the right graph, averaged over the number of years. The reason for this is that 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 and older individuals to work in the white-collar sector. The latter observation is a consequence of the first because education is rewarded in the white-collar sector.

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.

Table 1. Estimates for the distribution of input parameters

Parameter	Mean	Standard error (SE)	SE in KW94
General			
δ	0.95	0.000 84	-
Blue-collar			
β^b	9.21	0.013	0.014
β^b_e	0.038	0.0011	0.0015
β^b_b	0.033	0.000 44	0.000 79
β^b_{bb}	−0.0005	0.000 013	0.000 019
β^b_w	0.0	0.000 67	0.0024
β^b_{ww}	0.0	0.000 029	0.000 096
White-collar			
β^w	8.48	0.0076	0.0123
β^w_e	0.07	0.000 47	0.000 96
β^w_w	0.067	0.000 55	0.000 90
β^w_{ww}	−0.001	0.000 017	0.000 070
β^w_b	0.022	0.000 33	0.0010
β^w_{bb}	−0.0005	0.000 021	0.000 030
Education			
β^e	0.0	330	459
β^e_{he}	0.0	155	410
β^e_{re}	−4000	202	660
Home			
β^h	17 750	390	1442
Lower Triangular Cholesky Matrix			
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

(1990). National longitudinal survey of youth 1979 cohort, 1979–1990 (rounds 1–11). Bureau of Labor Statistics.

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.

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.

OpenSourceEconomics (2019). respy. Python package for the simulation and estimation of a prototypical finite-horizon dynamic discrete choice model based on Keane & Wolpin (1997). https://github.com/OpenSourceEconomics/respy.

Raabe, T. (2019). A unified estimation framework for some discrete choice dynamic programming models. Master's thesis, Bonn Graduate School of Economics.

Stenzel, T. (2020). *Master's Thesis Replication Repository.* Version: master. https://-github.com/HumanCapitalAnalysis/thesis-projects-tostenzel.

Verbeek, M. (2012). *A Guide to Modern Econometrics.* Wiley.