

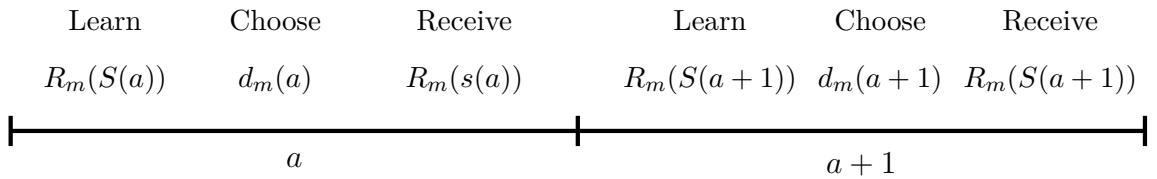
1 Model and Parametrization

This section introduces the economic model to which the UQ is applied. It is the dynamic model of occupational choice developed in Keane and Wolpin (1994). [Add further remarks on its relevance and Keane and Wolpin (1997).] First, I outline the general choice-theoretic framework and the concrete specification of the model. This includes the set of uncertain model parameters. Then, I present the approach that Keane and Wolpin use to estimate these model parameters given data on occupational choices. After that, their results for the means and variances of the respective estimates are shown. These constitute the parametrization for the input distribution in the UQ. The section continues with remarks on the numerical implementation. It concludes with the presentation of the QoI.

1.1 Economic Model

Keane and Wolpin (1994) develop a dynamic economic model to explain occupation decisions between compulsory schooling and pension age. In the general choice-theoretic framework, a representative agent decides for occupation m between M alternatives at each age $a \in A$. These alternatives are mutually exclusive. The decision for occupation m' at age a is denoted by $d_{m'}(a) = 1$. This equation is equivalent to $d_{m \in M \setminus m'}(a) = 0$. Each occupation decision is immediately rewarded by $R_m(S(a))$. Partly, rewards and prior decisions are intertemporally connected: Agents receive higher rewards if they accumulated skills in past occupations that are useful for the present one. $S(a)$ denotes the state space. The state space is the set of information at age a relevant for the present and future rewards for each occupation choice. R_m is subject to random shocks. The flow of events is depicted in Figure X.

FIGURE 1: Flow of events



At the beginning of each age a , the agent recognizes the reward shocks, and the shocks become part of the state space. Thus, $R_m(S(a))$ is known to the agent at age a . However, he can only form expectations about rewards in the future. Given Keane and Wolpin's concrete specification, it is assumed that shocks to the reward function are not serially correlated. Therefore, prior shocks do not enter the state space. Thus, the state space comprises the history of choices or, in other words, the occupation experience and the realizations of the contemporaneous shocks. Next, the agent chooses his occupation based

on state space information. Then he receives the occupation-specific reward. This flow repeats for each $a < A$.

Agents are rational and forward-looking. Future rewards 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 rewards *at each age*. This maximal value is given by the value function $V(S(a), a)$.

$$V(S(a), a) = \max_{(d_m(a))_{m \in M}} \mathbb{E} \left[\sum_{\tau=a}^T \delta^{\tau-a} \sum_{m=1}^M R_m(a) d_m(a) | S(a) \right] \quad (1)$$

Value V depends directly on age a because A is finite. To re-write the maximization problem as Bellman equation (Bellman (1957); equation 3), it is noted that the value function equals the maximal choice-specific value function (equation 2).

$$V(S(a), a) = \max_{m \in M} V_m(S(a), a) \quad (2)$$

$$V_m(S(a), a) = \max_{d_m(a)} \left\{ R_m(S(a), a) + \delta \mathbb{E} [V(S(a+1), a+1)] \right\} \quad (3)$$

Taken together, the above two equations constitute the dynamic programming problem for the general choice-theoretic framework. The agent computes his potential return for every possible scenario order to make the optimal decision. Therefore, this problem is solved for every point in the state space.

The economic model of occupational choice in Keane and Wolpin (1994) is specified by the four return functions R_m in equation block 3 for the following m choices: occupation one, occupation two, school and home. s_a is the number of years of schooling, x_{1a} is the number of years that the agent worked in occupation one. Analogously, x_{2a} is the experience in occupation two. α_1 and α_2 are vectors that determine the effect of experience and cross-experience in the return functions for occupation one and occupation two. β_0 is the consumption value of schooling. Indicator function 1 indicates whether an agent has completed high school, and β_1 is the post-secondary tuition cost of schooling. β_2 is an adjustment cost for returning to school when the agent had another occupation the year before ($d_{a-1} = 0$). γ_0 is the mean value for staying at home. An additional assumption is $\delta = 0.9$.

$$\begin{aligned}
R_1(a) &= w_{1a} = \exp(\alpha_{10} + \alpha_{11}s_a + \alpha_{12}x_{1a} - \alpha_{13}x_{1a}^2 + \alpha_{14}x_{2a} - \alpha_{15}x_{2a}^2 + \epsilon_{1a}) \\
R_2(a) &= w_{2a} = \exp(\alpha_{20} + \alpha_{21}s_a + \alpha_{22}x_{1a} - \alpha_{23}x_{1a}^2 + \alpha_{24}x_{2a} - \alpha_{25}x_{2a}^2 + \epsilon_{2a}) \\
R_3(a) &= \beta_0 - \beta_1 1(s_a \geq 12)(s_a) - \beta_2(1 - d_3(a - 1)) + \epsilon_{3a} \\
R_4(a) &= \gamma_0 + \epsilon_{4a}
\end{aligned} \tag{4}$$

1.2 Parameter Estimation

Given longitudinal data on occupational choices m of a sample with size I , the model parameters θ in table 2 can be estimated. The approach that Keane and Wolpin (1994) use is the simulated maximum likelihood method (Albright et al. (1977); see Keane et al. (2011) for a detailed application to a similar model). Denote the first and the last period observed for agent i by a_{a_0i} and a_{Ai} . The sample likelihood function for θ is given by

$$L(\theta; \text{data}) = \prod_{i=1}^{i=I} \prod_{a=a_0}^{a=A} Pr(d_{ia} = 1, S_{ia})^{d_{ia}} Pr(d_{ia} = 0, S_{ia})^{1-d_{ia}}. \tag{5}$$

Assuming normality of ϵ , equation 5 is analytic. The estimation alternates the solution of the dynamic programming problem on the complete parameter grid with the maximum likelihood estimation of the parameters for each period in the data. Thus in each iteration, the agent's problem is solved for each state. This gives the population's joint distribution of choices and parameters. The distribution is represented by the age-specific likelihood. Then, the state vector that maximizes the probability of the distribution of occupational choices in the data is obtained by maximizing the likelihood function.

Keane and Wolpin (1994) simulate data on fictional parameter values test the estimation method with good results. Table 2 presents their estimates for mean and standard deviation of model parameters θ . [Substitute true params with params + bias]

The mean estimates for θ have the following economic implications: Occupation two is more skill-intensive or, more formally, has higher returns to education and occupational experience than occupation one. Additionally, experience in occupation one is rewarded in occupation two but not vice versa. The c parameters are the Cholesky factors of the lower triangular matrix C obtained from a Cholesky decomposition of the covariance matrix of the error vector. In this specific parametrization, matrix C coincides with shocks covariance matrix Σ . Therefore, the error terms are uncorrelated, as well.

The estimates in table 1 are the moments used to describe the input parameter distribution for the uncertainty quantification. However, estimates for parameters based on fictional simulations are an important drawback. The estimates imply uncorrelated input parameters. This does not correspond to the respective economic literature. [Include resources.] A methodological consequence is that the UQ, precisely the computation of Sobol' indices, is simplified. The reason for this parametrization approach is that the

numerical implementation, in particular, the part for the likelihood function, is erroneous. Therefore, this thesis foregoes to compute its own parameter estimates.

TABLE 1: Model Parametrization^a

Parameter	Mean ^b	Standard Deviation	Label
α_{10}	9.21	0.0016	
α_{11}	0.038	0.000 16	
α_{12}	0.033	0.000 14	
α_{13}	0.0005	0.000 011	
α_{14}	0.0	0.0011	
α_{15}	0.0	0.000 058	
α_{20}	8.48	0.0019	
α_{21}	0.07	0.000 095	
α_{22}	0.067	0.000 14	
α_{23}	0.001	0.000 008	
α_{24}	0.022	0.000 40	
α_{25}	0.0005	0.000 041	
β_0	0.0	100	
β_1	0.0	145	
β_2	4000	317	
γ_0	17 750	159	
c_{11} ^c	0.2	0.0030	
c_{21}	0.0	0.016	
c_{22}	0.25	0.0044	
c_{31}	0.0	0.261	
c_{32}	0.0	0.224	
c_{33}	1500	218	
c_{41}	0.0	0.123	
c_{42}	0.0	0.406	
c_{43}	0.0	-0.588	
c_{44}	1500	360	

^a The numbers are taken from table 4.1 in Keane and Wolpin (1994), p.665.

^b [Maybe small note on Bayesian perspective of true parameter values.]

^c $\epsilon_{1t} = a_{11}\eta_{1t}$
 $\epsilon_{2t} = a_{21}\eta_{2t} + a_{22}\eta_{2t}$
 $\epsilon_{3t} = a_{31}\eta_{3t} + a_{32}\eta_{3t} + a_{33}\eta_{3t}$
 $\epsilon_{4t} = a_{41}\eta_{4t} + a_{42}\eta_{4t} + a_{43}\eta_{4t} + a_{44}\eta_{4t}$
 $\eta_k \sim N(0, 1), \quad k = 1, \dots, 4$

1.3 Numerical Implementation

The thesis uses the python package **(alias?)** to compute the QoI. It is important to transform the input parameters from the format in Keane and Wolpin (1994) to the respy convention. This includes, in particular, the transformation of the Cholesky parameters c into standard deviations and correlations. All details can be found in the **(alias?)**.

1.4 Quantity of Interest

The QoI for the uncertainty quantification is the effect of a \$500 college tuition subsidy on the average years of schooling. Formally, $\beta_2 1(13 \leq s_a \leq 16) = \beta_2 1(13 \leq s_a \leq 16) - 500$. In Keane and Wolpin (1994), the effect is an increase of 1.44 years (see table 4, p.668). The same figure computed with **(alias?)** is 1.449. Given other model outputs, the general ability of **(alias?)** to replicate the model solution in Keane and Wolpin (1994) is good. Therefore, the QoI computations can be regarded as sufficiently precise. I choose this quantity because it is relevant to society in many respects, for example, education and inequality. Section XX expands this point. The QoI's relevance allows illustrating the importance of UQ in economics in the context of political decisions.

References

- 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.
- Keane, M. P., P. E. Todd, and K. I. Wolpin (2011). The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4A, pp. 331–461. Amsterdam, Netherlands: Elsevier Science.
- 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.
- OpenSourceEconomics (2019). respy. Python package for the simulation and estimation of a prototypical finite-horizon dynamic discrete choice model based on Keane & Wolpin (1997). Version: develop@XX. <https://github.com/OpenSourceEconomics/respy/commit/XX>.
- Stenzel, T. (2020). *Master’s Thesis Replication Repository*. Version: master. <https://github.com/HumanCapitalAnalysis/thesis-projects-tostenzel>.