Eckstein–Keane–Wolpin models
An invitation for transdisciplinary collaboration*

Open Source Economics

This version: April 8, 2021

Abstract

We present background material on a class of structural microeconometric models to facilitate transdisciplinary collaboration in their future development. We describe the economic framework, mathematical formulation, and calibration procedures for the so-called Eckstein–Keane–Wolpin (EKW) models. We provide an exemplifying analysis of the seminal model outlined in Keane and Wolpin (1997) and present our group's ensemble of research codes that allow for its specification, simulation, and calibration. We summarize our efforts drawing on research outside economics to address the computational challenges in applying EKW models and improve the reliability and interpretability of their results.

*Corresponding author: Philipp Eisenhauer, p.eisenhauer@uni-bonn.de. We thank Janoś Gabler, Annica Gehlen, Holger Gerhardt, Boryana Ilieva, Fedor Ishakov, Lena Janys, Moritz Mendel, Tim Mensinger, Tobias Raabe, Klara Röhrl, John Rust, and Rafael Suchy for their valuable input. Julian König and Luis Wardenbach provided outstanding research assistance.

Contents

1 Introduction

Economists use structural microeconometric models to study individual decision-making. These models specify the objective of individuals, their economic environment, and the institutional and informational constraints under which they operate. Calibration of the model to observed data on individual decisions and experiences allows quantifying the importance of competing economic mechanisms in determining economic outcomes and forecasting the effects of policy proposals (Wolpin, 2013).

We restrict our exposition to the class of Eckstein–Keane–Wolpin (EKW) models (Aguirregabiria & Mira, 2010)[1]. Labor economists use them to study human capital investment decisions. Human capital comprises the knowledge, skills, competencies, and attributes embodied in individuals facilitating the creation of personal, social, and economic well-being (Becker, 1964). Differences in human capital attainment lead to inequality in various life outcomes such as labor market success and health across and within countries (OECD, 2001).

In Bhuller et al. (2018), for example, we apply an EKW model to analyze the mechanisms determining schooling decisions in Norway. We calibrate the model using Norwegian population panel data with nearly career-long earnings histories. After validating our model using a mandatory schooling reform, we gain insights into the underlying economic mechanisms that generate the effects of the policy and forecast the impacts of several policy alternatives.

We offer this handout to facilitate transdisciplinary collaboration in the future development of EKW models. We first describe their economic framework, mathematical formulation, and calibration procedure. We then turn to the seminal model outlined in Keane and Wolpin (1997) as an example and present our group's ensemble of research codes that allow for its specification, simulation, and calibration. Finally, we summarize our efforts drawing on research outside economics to address the computational challenges in applying EKW models and improve the reliability and interpretability of their results.

Throughout, we only offer a limited number of seminal references and textbooks that invite further study. We introduce acronyms and symbols as needed, and our notation draws on the reviews by Aguirregabiria and Mira (2010), Arcidiacono and Ellickson (2011), and Puterman (1994).

1. See Adda et al. (2017), Blundell et al. (2016) for recent publications.

2 Setup

We now present the basic setup of the EKW models. We first describe the economic framework, then turn to its mathematical formulation, and finally outline the calibration procedure.

2.1 Economic framework

EKW models describe sequential decision-making under uncertainty (Gilboa, 2009; Machina & Viscusi, 2014). At time $t = 1, \ldots, T$ each individual observes the state of the economic environment $s_t \in S$ and chooses an action a_t from the set of admissible actions $\mathcal{A}$. The decision has two consequences: an individual receives an immediate utility $u_t(s_t, a_t)$ and the economy evolves to a new state s_{t+1}. The transition from s_t to s_{t+1} is affected by the action but remains uncertain. Individuals are forward-looking. Thus they do not simply choose the alternative with the highest immediate utility. Instead, they take the future consequences of their current action into account.

A policy $\pi = (a_1^\pi(s_1), \ldots, a_T^\pi(s_T))$ provides the individual with instructions for choosing an action in any possible future state. It is a sequence of decision rules $a_t^\pi(s_t)$ that specify the action at a particular time t for any possible state s_t under π. The implementation of a policy generates a sequence of utilities that depends on the objective transition probability distribution $p_t(s_t, a_t)$ for the evolution of state s_t to s_{t+1} induced by the model. Individuals have rational expectations (Muth, 1961) so their subjective beliefs about the future agree with the objective transition probabilities of the model.

Figure 1 depicts the timing of events in the model for two generic periods. At the beginning of period t, an individual fully learns about the immediate utility of each alternative, chooses one of them, and receives its immediate utility. Then the state evolves from s_t to s_{t+1} and the process is repeated in $t + 1$.

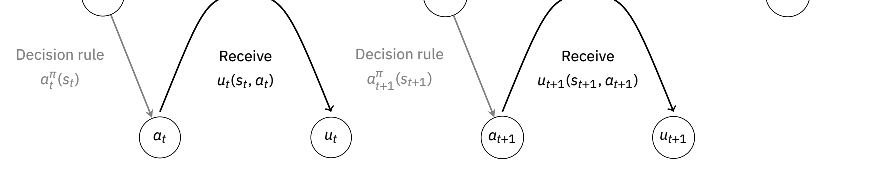


Figure 1. Timing of events

Individuals face uncertainty and they seek to maximize the expected total discounted utilities. An exponential discount factor $0 < \delta < 1$ parameterizes their time preference and captures a taste for immediate over future utilities.

Equation (1) provides the formal representation of the individual's objective. Given an initial state s_1, individuals implement the policy π from the set of all possible policies Π that maximizes the expected total discounted utilities over all T decision periods given the information available in the first-period state, s_1:

$$\max_{\pi \in \Pi} \mathrm{E}_{s_1}^{\pi} \left[\sum_{t=1}^{T} \delta^{t-1} u_t(s_t, a_t^{\pi}(s_t)) \right]. \tag{1}$$

The superscript of the expectation emphasizes that each policy π induces a different probability distribution over the sequences of utilities.

2.2 Mathematical formulation

EKW models are set up as a standard Markov decision process (MDP) (Puterman, 1994; White, 1993). When making sequential decisions under uncertainty, the task is to determine the optimal policy π^* with the largest expected total discounted utilities $v_1^{\pi^*}(s_1)$ as formalized in equation (1). In principle, this requires evaluating the performance of all policies based on all possible sequences of utilities, each weighted by the probability with which they occur. Fortunately, however, the multistage problem can be solved by a sequence of simpler inductively defined single-stage problems.[2]

The value function $v_t^{\pi}(s_t)$ captures the expected total discounted utilities under policy π from period t onwards for an individual experiencing state s_t:

$$v_t^{\pi}(s_t) = \mathrm{E}_{s_t}^{\pi} \left[\sum_{j=0}^{T-t} \delta^j \, u_{t+j}(s_{t+j}, a_{t+j}^{\pi}(s_{t+j})) \right].$$

Then we can determine $v_1^{\pi}(s_1)$ for any policy by recursively evaluating equation (2):

$$v_t^{\pi}(s_t) = u_t(s_t, a_t^{\pi}(s_t)) + \delta \; \mathrm{E}_{s_t}^{\pi} \left[v_{t+1}^{\pi}(s_{t+1}) \right]. \tag{2}$$

Equation (2) expresses the total value $v_t^{\pi}(s_t)$ of adopting policy π going forward as the sum of its immediate utility and all expected discounted future utilities.

The principle of optimality (Bellman, 1954) allows to construct π^* by solving the optimality equations (3) for all s and t recursively:

$$v_t^{\pi^*}(s_t) = \max_{a_t \in A} \left\{ u_t(s_t, a_t) + \delta \; \mathrm{E}_{s_t}^{\pi^*} \left[v_{t+1}^{\pi^*}(s_{t+1}) \right] \right\}. \tag{3}$$

2. Optimal decisions in an MDP are a deterministic function of the current state s only, i.e., an optimal decision rule is always deterministic and Markovian. We restrict our notation to this special case right from the beginning.

Algorithm 1. Backward induction procedure

for $t = T, \ldots, 1$ **do**
    if t = T **then**
        $v_T^{\pi^*}(s_T) = \max_{a_T \in A} \Big\{ u_T(s_T, a_T) \Big\} \qquad \forall\, s_T \in S$
    else
        Compute $v_t^{\pi^*}(s_t)$ for each $s_t \in S$ by
            $v_t^{\pi^*}(s_t) = \max_{a_t \in A} \Big\{ u_t(s_t, a_t) + \delta \ \mathrm{E}_{s_t}^{\pi} \Big[v_{t+1}^{\pi^*}(s_{t+1}) \Big] \Big\}.$
        and set
            $a_t^{\pi^*}(s_t) = \arg\max_{a_t \in A} \Big\{ u_t(s_t, a_t) + \delta \ \mathrm{E}_{s_t}^{\pi} \Big[v_{t+1}^{\pi^*}(s_{t+1}) \Big] \Big\}.$
    end if
end for

The optimal value function $v_t^{\pi^*}$ is the sum of the expected discounted utilities in t over the remaining time horizon assuming the optimal policy is implemented going forward. The optimal action is choosing the alternative with the highest total value:

$$a_t^{\pi^*}(s_t) = \arg\max_{a_t \in A} \left\{ u_t(s_t, a_t) + \delta \ \mathrm{E}_{s_t}^{\pi^*} \left[v_{t+1}^{\pi^*}(s_{t+1}) \right] \right\}.$$

Algorithm 1 allows to solve the MDP by a simple backward induction procedure. In the final period T, there is no future to take into account, and the optimal action is choosing the alternative with the highest immediate utilities in each state. With the decision rule for the final period at hand, the other optimal decisions can be determined recursively following equation (3) as the calculation of their expected future utilities is straightforward given the relevant transition probabilities.

2.3 Calibration procedure

EKW models are calibrated to data on observed individual decisions and experiences under the hypothesis that the individual behaves according to the model. The goal is to back out information on utility functions, preference parameters, and transition probabilities. This requires the full parameterization θ of the model.

Economists have access to information for $i = 1, \ldots, N$ individuals in each time period $t = 1, \ldots, T_i$. For every observation (i, t) in the data, we observe the action a_{it}, some components $\bar{u}_{it}$ of the utility, and a subset $\bar{s}_{it}$ of the state s_{it}. Therefore, from an economist's point of view, we need to distinguish between two types of state variables $s_{it} = (\bar{s}_{it}, \epsilon_{it})$. At time t, the economist and individual both observe $\bar{s}_{it}$ while ϵ_{it} is only observed by the individual. In summary, the data $\mathcal{D}$ has the following structure:

$$\mathcal{D} = \{a_{it}, \bar{s}_{it}, \bar{u}_{it} : i = 1, \ldots, N; t = 1, \ldots, T_i\},$$

where T_i is the number of observations for which we observe individual i.

Numerous calibration procedures for different settings exist (Davidson & MacKinnon, 2003; Gourieroux & Monfort, 1996). We briefly outline likelihood-based and simulation-based calibration. Independent of the calibration criterion, it is necessary to solve for the optimal policy π^* at each candidate parameterization of the model.

Likelihood-based calibration seeks to find the parameterization $\hat{\theta}$ that maximizes the likelihood function $\mathcal{L}(\theta \mid \mathcal{D})$, i.e. the probability of observing the given data as a function of θ. As we only observe a subset $\bar{s}_t$ of the state, we can determine the probability $p_{it}(a_{it}, \bar{u}_{it} \mid \bar{s}_{it}, \theta)$ of individual i at time t in $\bar{s}_{it}$ choosing a_{it} and receiving u_{it} given parametric assumptions about the distribution of ϵ_{it}. The objective function takes the following form:

$$\hat{\theta} \equiv \arg\max_{\theta \in \Theta} \underbrace{\prod_{i=1}^{N} \prod_{t=1}^{T_i} p_{it}(a_{it}, \bar{u}_{it} \mid \bar{s}_{it}, \theta)}_{\mathcal{L}(\theta \mid \mathcal{D})}.$$

In simulation-based calibration, our goal is to find the parameterization $\hat{\theta}$ that yields a simulated data set from the model that closest resembles the observed data. More precisely, the goal is often to minimize the weighted squared distance between a set of moments M_D computed on the observed data and the same set of moments computed on the simulated data $M_S(\theta)$. The objective function takes the following form:

$$\hat{\theta} \equiv \arg\min_{\theta \in \Theta} (M_D - M_S(\theta))' W (M_D - M_S(\theta)).$$

3 Example

We now present an exemplifying analysis of a canonical EKW model on human capital investment. The model was initially studied in Keane and Wolpin (1997) to explore the career decisions of young men about their schooling, work, and occupational choice. We first outline the basic setup of the model, provide some descriptive statistics of the empirical data used for its calibration, and then explore selected economic insights.

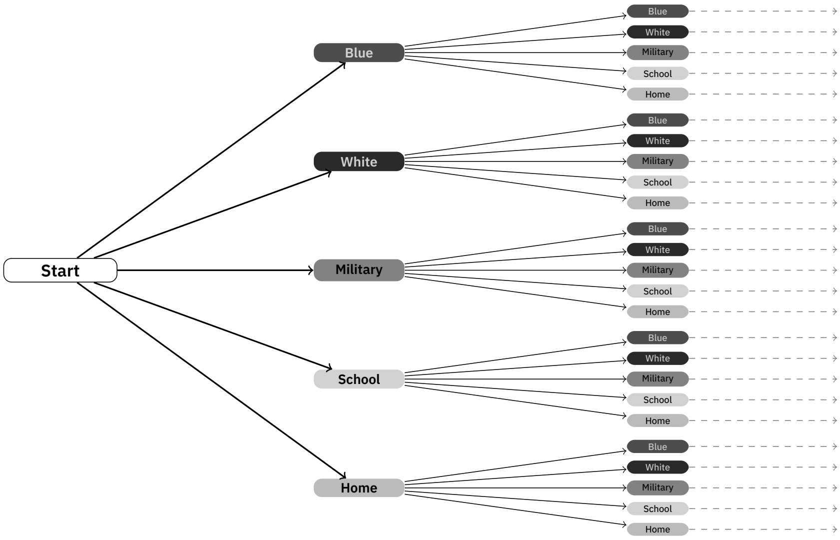


Figure 2. Decision tree

3.1 Basic setup

We follow individuals over their working life from young adulthood at age 16 to retirement at age 65 where the decision period $t = 16, \dots, 65$ is a school year. Figure 2 illustrates the initial decision problem as individuals decide $a \in \mathcal{A}$ whether to work in a blue-collar or white-collar occupation ($a = 1, 2$), to serve in the military ($a = 3$), to attend school ($a = 4$), or to stay at home ($a = 5$).

Individuals are already heterogeneous when entering the model. They differ with respect to their level of completed schooling h_{16} and have one of four different $\mathcal{J} = \{1, \dots, 4\}$ alternative-specific skill endowments $\boldsymbol{e} = \left(e_{j,a}\right)_{\mathcal{J} \times \mathcal{A}}$.

The immediate utility $u(\cdot)$ of each alternative consists of a non-pecuniary utility $\zeta_a(\cdot)$ and, at least for the working alternatives, an additional wage component $w_a(\cdot)$. Both depend on the level of human capital as measured by their occupation-specific work experience $\boldsymbol{k}_t = \left(k_{a,t}\right)_{a \in \{1,2,3\}}$, years of completed schooling h_t, and alternative-specific skill endowment $\boldsymbol{e}$. The immediate utilities are influenced by last-period choices a_{t-1} and alternative-specific productivity shocks $\boldsymbol{\epsilon}_t = \left(\epsilon_{a,t}\right)_{a \in \mathcal{A}}$ as well. Their general form is given by:

$$u(\cdot) = \begin{cases} \zeta_a(\boldsymbol{k}_t, h_t, t, a_{t-1}) + w_a(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) & \text{if} \quad a \in \{1, 2, 3\} \\ \zeta_a(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) & \text{if} \quad a \in \{4, 5\} \end{cases}.$$

Work experience $\boldsymbol{k}_t$ and years of completed schooling h_t evolve deterministically.

$$\begin{aligned} k_{a,t+1} &= k_{a,t} + 1[a_t = a] \quad \text{if} \quad a \in \{1,2,3\} \\ h_{t+1} &= h_t \;\; + 1[a_t = 4]. \end{aligned}$$

The productivity shocks $\boldsymbol{\varepsilon}_t$ are uncorrelated across time and follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Given the structure of the utility functions and the distribution of the shocks, the state at time t is $s_t = \{\boldsymbol{k}_t, h_t, t, a_{t-1}, \boldsymbol{e}, \boldsymbol{\varepsilon}_t\}$.

Theoretical and empirical research from specialized disciplines within economics informs the specification of each $u_a(\cdot)$ and we discuss the exact functional form of the per-period utility in the blue-collar occupation as an example.[3]

Equation (4) shows the parameterization of the non-pecuniary utility from working in a blue-collar occupation:

$$\begin{aligned} \zeta_1(\boldsymbol{k}_t, h_t, a_{t-1}) = \alpha_1 &+ c_{1,1} \cdot 1[a_{t-1} \neq 1] + c_{1,2} \cdot 1[k_{1,t} = 0] \\ &+ \vartheta_1 \cdot 1[h_t \geq 12] + \vartheta_2 \cdot 1[h_t \geq 16] + \vartheta_3 \cdot 1[k_{3,t} = 1]. \end{aligned} \tag{4}$$

It includes job amenities α_1 and mobility and search costs $(c_{1,1}, c_{1,2})$ that capture the extra effort for individuals who only recently started working in a blue-collar occupation. Additional components depend on whether an individual has a high school ϑ_1 or college ϑ_2 degree. There is a detrimental impact of leaving the military after a single year ϑ_3.

The wage component $w_1(\cdot)$ is given by the product of the market-equilibrium rental price r_1 and an occupation-specific skill level $x_1(\cdot)$. The latter is determined by the overall level of human capital. This specification leads to a standard logarithmic wage equation in which the constant term is the skill rental price $\ln(r_1)$ and wages follow a log-normal distribution.

The occupation-specific skill level $x_1(\cdot)$ is determined by a skill production function, which includes a deterministic component $\Gamma_1(\cdot)$ and a multiplicative stochastic productivity shock $\epsilon_{1,t}$:

$$x_1(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,1}, \epsilon_{1,t}) = \exp\big(\Gamma_1(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,1}) \cdot \epsilon_{1,t}\big).$$

Equation (5) shows the parameterization of the deterministic component of the skill production function:

$$\begin{aligned} \Gamma_1(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,1}) = e_{j,1} &+ \beta_{1,1} \cdot h_t + \beta_{1,2} \cdot 1[h_t \geq 12] \\ &+ \gamma_{1,1} \cdot k_{1,t} + \gamma_{1,2} \cdot (k_{1,t})^2 \\ &+ \gamma_{1,3} \cdot 1[k_{1,t} > 0] + \gamma_{1,4} \cdot t + \gamma_{1,5} \cdot 1[t < 18] \\ &+ \gamma_{1,6} \cdot 1[a_{t-1} = 1] + \gamma_{1,7} \cdot k_{2,t} + \gamma_{1,8} \cdot k_{3,t}. \end{aligned} \tag{5}$$

3. All additional details are available in Appendix A.

There are several notable features. Skills increase with schooling $\beta_{1,1}$ and blue-collar work experience ($\gamma_{1,1}, \gamma_{1,2}$). There are so-called sheep-skin effects (Hungerford & Solon, 1987; Jaeger & Page, 1996) associated with completing a high school $\beta_{1,2}$ and graduate $\beta_{1,3}$ education that capture the impact of completing a degree beyond just the associated years of schooling. Also, there is a first-year blue-collar experience effect $\gamma_{1,3}$ while skills depreciate when not employed in a blue-collar occupation in the preceding period $\gamma_{1,6}$. Other work experience ($\gamma_{1,7}, \gamma_{1,8}$) is transferable.

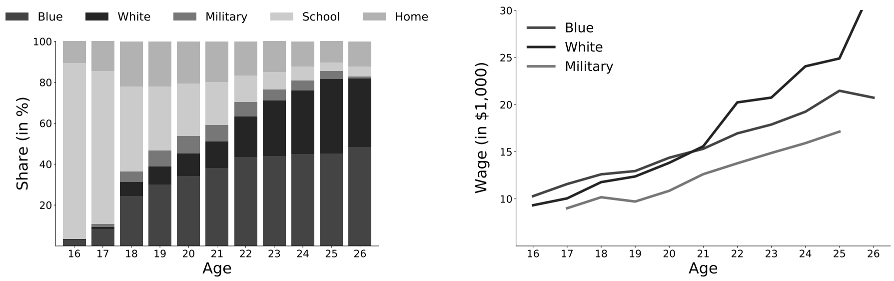
3.2 Empirical data

We analyze the original dataset used by Keane and Wolpin (1997) and thus only provide a brief description here.[4] The authors construct their sample based on the National Longitudinal Survey of Youth 1979 (NLSY79) (Bureau of Labor Statistics, 2019). The NLSY79 is a nationally representative sample of young men and women living in the United States in 1979 and born between 1957 and 1964. Individuals were followed from 1979 onwards and repeatedly interviewed about their schooling decisions and labor market experiences. Based on this information, individuals are assigned to either working in one of the three occupations, attending school, or simply staying at home.

Keane and Wolpin (1997) restrict attention to white males that turn 16 between 1977 and 1981 and exploit the information collected between 1979 and 1987. Thus individuals in the sample are all between 16 and 26 years old. While the sample initially consists of 1,373 individuals at age 16, this number drops to 256 at the age of 26 due to sample attrition, missing data, and the short observation period. Overall, the final sample consists of 12,359 person-period observations.

Figure 3 summarizes our information about choices and wages by age. We show the distribution of choices on the left, and report average wages on the right. Initially, roughly 86% of individuals enroll in school, but this share steadily declines with age. Nevertheless, about 39% obtain more than a high school degree and continue their schooling for more than twelve years. As individuals leave school, most of them initially pursue a blue-collar occupation. But the relative share of the white-collar occupation increases as individuals entering the labor market later have higher levels of schooling. At age 26, about 48% work in a blue-collar occupation and 34% in a white-collar occupation. The share of individuals in the military peaks around age 20 when it amounts to 8%. At its maximum around age 18, approximately 20% of individuals stay at home.

4. We provide additional details in Appendix B.

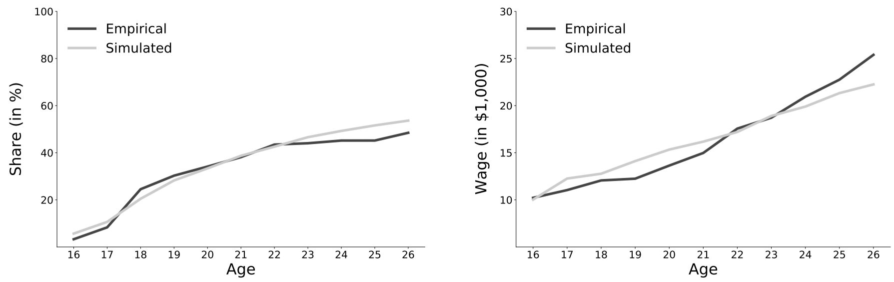


(a) Choices

(b) Average wage

Notes: The wage is a full-time equivalent deflated by the gross national product deflator, with 1987 as the base year. We do not report the wage if less than ten observations are available.

Figure 3. Data overview



(a) Blue-collar

(b) Average wage

Notes: We simulate a sample of 1,000 individuals using the calibrated model.

Figure 4. Model fit

Overall, average wages start at about $10,000 at age 16 but increase considerably up to about $25,000 at age 26. While wages in the blue-collar occupation are initially highest with about $10,286, wages in the white-collar occupation and military start around $9,000. However, wages in the white-collar occupation increase steeper over time and overtake blue-collar wages around age 21. At the end of the observation period, wages in the white-collar occupation are about 50% higher than blue-collar wages with $32,756 as opposed to only $20,739. Military wages remain lowest throughout.

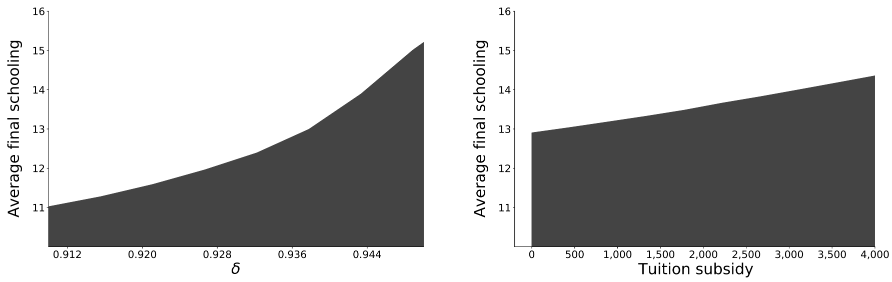
We fit the model to the empirical data using maximum likelihood calibration. Figure 4 shows the overall agreement between the empirical data and a dataset simulated using the calibrated parameters within the support of the data. On the left, we show the choice probability of working in a blue-collar occupation, while we plot the average wage across all occupations on the right.

Overall, the values of the calibrated parameters of the model are in broad agreement with the relevant literature. For example, individuals discount future utilities by 6% per year, and wages increase by about 7% with each additional year of schooling.

3.3 Economic insights

Figure 5 illustrates the ability of the model to quantify the impact of economic mechanisms and to forecast the effect of public policies. On the left, we vary the discount factor capturing time preferences between 0.91 and 0.95 while we introduce a tuition subsidy of up to $4,000 on the right. In both cases, we are interested in the changes to average final schooling.

Increases in the discount factor and the tuition subsidy both result in higher average final schooling. However, they do so for very different reasons. While individuals emphasize the future benefits of their schooling investment in the former, they react to a reduction of its immediate cost in the latter.



(a) Time preference

(b) Tuition subsidy

Notes: We simulate a sample of 1,000 individuals using the calibrated model.

Figure 5. Economic mechanism and policy forecast

4 Pipeline

We are actively developing an ensemble of research codes that provide an analysis pipeline for EKW models. Among them are `respy` and `estimagic`. The former allows for the flexible specification and simulation of EKW models while the latter provides the means for their calibration. We briefly showcase the typical workflow of using both packages in our research.

Figure 6 illustrates a typical workflow. Initially, the user provides the empirical data, the parameterization of the model, and other options to `respy`.

All together define the structure of the model, and we can construct the functionality for the simulation of data and the evaluation of the criterion function. `estimagic` allows calibrating the model to the empirical data. The results from the calibration steps are used to, for example, analyze the economic mechanisms underlying the observed behaviors.

Figure 7 shows the model specification files for Keane and Wolpin (1997). The file on the left sets the parameter values for the utility functions and the distribution of the unobservable state variables. On the right, we provide details on the construction of the observed state variables and numerous tuning parameters for the numerical solution of the model.

Figure 8 depicts the dashboard provided by `estimagic` to monitor the progress and parameter values of the calibration in real-time. This allows us to detect problems during calibration right away and facilitates the debugging process.

We adopt a modern software engineering workflow in the development of both packages and tutorials, source code, testing harness, as well as imple-

```
import respy as rp
from estimagic import maximize

# obtain model input
params, options, df = rp.get_example_model("kw_97_extended_respy")

# process model specification
log_like = rp.get_log_like_func(params, options, df)
simulate = rp.get_simulate_func(params, options)

# perform calibration
results, params_rslt = maximize(log_like, params, "nlopt_bobyqa")

# conduct analysis
df_rslt = simulate(params_rslt)
```

Figure 6. Typical workflow

		value	name
category	**name**		
delta	**delta**	9.370735e-01	delta_delta
wage_white_collar	**constant**	8.741888e+00	wage_white_collar_constant
	exp_school	6.548940e-02	wage_white_collar_exp_school
	exp_white_collar	1.763655e-02	wage_white_collar_exp_white_collar
	exp_white_collar_square	-4.215936e-02	wage_white_collar_exp_white_collar_square
	exp_blue_collar	3.431936e-02	wage_white_collar_exp_blue_collar
	exp_military	1.406945e-02	wage_white_collar_exp_military
	hs_graduate	-3.599855e-03	wage_white_collar_hs_graduate
	co_graduate	2.301313e-03	wage_white_collar_co_graduate
	period	9.577717e-03	wage_white_collar_period
	is_minor	-1.509984e-01	wage_white_collar_is_minor

(a) Parameterization

	value
estimation_draws	200
estimation_seed	500
estimation_tau	500
interpolation_points	-1
n_periods	50
simulation_agents	5000
simulation_seed	132
solution_draws	500
solution_seed	456
monte_carlo_sequence	random
covariates	{'hs_graduate': 'exp_school >= 12', 'co_gradua...

(b) Options

Figure 7. Model specification

mentation details are available in their respective online documentations at https://respy.rtfd.io and https://estimagic.rtfd.io.

5 Improvements

The implementation of EKW models poses several computational challenges. Among them are numerical integration, global optimization, function approximation, and efficient parallelization. We now describe some of our efforts to align `respy` and `estimagic` with the state-of-the-art in computational methods. We have concluded our preparatory work and actively seek input from domain experts for further improvements and joint publication.

5.1 Numerical integration

The solution of EKW models requires the evaluation of millions of integrals to determine the future value of each action in each state. In Eisenhauer, Gabler, and Suchy (2020), we draw on the extensive literature on numerical integration (Davis & Rabinowitz, 2007; Gerstner & Griebel, 1998) to improve the precision and reliability of their solution. The current practice in economics is to implement a random Monte Carlo integration which introduces considerable numerical error and computational instabilities (Judd & Skrainka, 2011).

We consider the optimality equation in a generic time period t to clarify the structure of the integral. Let $v_t^{\pi}(s_t, a_t)$ denote the action-specific value function of choosing action a_t in state s_t while continuing with the optimal policy going forward.

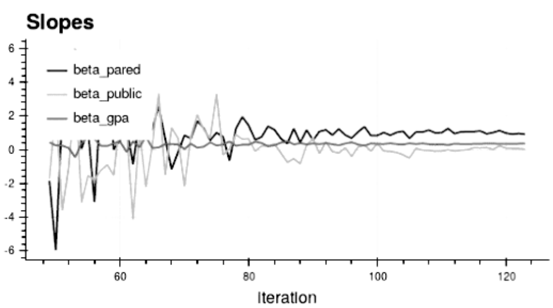


Figure 8. Dashboard

$$
\begin{aligned}
v_t^{\pi}(s_t, a_t) &= u_t(s_t, a_t) + \delta \ \mathrm{E}_{s_t}\left[v_{t+1}^{\pi^*}(s_{t+1}) \right] \\
&= u_t(s_t, a_t) + \delta \int_S v_{t+1}^{\pi^*}(s_{t+1}) \ \mathrm{d}p_t(a_t, s_t) \\
&= u_t(s_t, a_t) + \delta \int_S \max_{a_{t+1} \in A} \left\{ v_{t+1}^{\pi^*}(s_{t+1}, a_{t+1}) \right\} \mathrm{d}p_t(a_t, s_t).
\end{aligned}
$$

Let's consider an atemporal version of the typical integral from Keane and Wolpin (1997) as an example. As outlined earlier, individuals can choose among five alternatives. Each of the alternative-specific utilities is, in part, determined by a stochastic continuous state variable $\boldsymbol{\varepsilon}$. The transition of all other state variables x is deterministic. This results in a five-dimensional integral of the following form:

$$
\int \max_{a \in A} \left\{ v^{\pi^*}(x, \boldsymbol{\varepsilon}, a) \right\} \phi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\varepsilon}) \, \mathrm{d}\boldsymbol{\varepsilon} \quad \forall\, x \in X,
$$

where $\boldsymbol{\varepsilon}$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and probability density function $\phi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$.

5.2 Global optimization

The calibration of EKW models is challenging due to a large number of parameters and multiplicity of local minima. In Eisenhauer, Gabler, and Röhrl (2020), we draw on the literature on global optimization to assess and improve the reliability of the calibrations (Locatelli & Schoen, 2013; Nocedal & Wright, 2006).

We conduct a benchmarking exercise using Keane and Wolpin (1997), Keane and Wolpin (1994) as a well-known and empirically grounded test case. Depending on the calibration procedure, particular challenges arise. For example, while likelihood-based calibration requires smoothing of the choice probabilities, simulation-based calibration involves the optimization of a noisy function. We provide guidelines for selecting the appropriate algorithm in each setting and showcase diagnostics to assess the reliability of the calibration results.

6 Extensions

We are actively pursuing several extensions to the standard analysis of EKW models. For example, we draw on the methodological literature on robust-decision making and uncertainty quantification to account for the uncertainties within and outside the model (Hansen, 2015). We also work with the German Institute for Economic Research and Statistics Norway to improve the available data for the calibration of the models. Again, we have concluded our preparatory work and actively seek input from domain experts for further improvements and joint publication.

6.1 Robust decision making

The uncertainties involved in human capital investments are ubiquitous (Becker, 1964). Individuals usually make investments early in life when they are still uncertain about their abilities and tastes. Their returns also depend on demographic, economic, and technological trends that only start to unfold years from now. However, the treatment of uncertainty in EKW models of human capital investment is very narrow. A model provides individuals with a formalized view of their economic environment and implies unique probabilities for all possible future events. Individuals have no fear of model misspecification.

In Eisenhauer and Suchy (2020), we address this shortcoming by formulating, implementing, and exploring an EKW model of robust human capital investment where individuals face risk within a model and ambiguity about the model (Arrow, 1951). Ambiguity arises as individuals do not know the true model and consider a whole set of models as reasonable descriptions of their economic environment. Individuals fear model misspecification and thus seek robust decisions, i.e., decisions that perform well over the whole range of models.

We incorporate methods from robust optimization (Ben-Tal et al., 2009; Rahimian & Mehrotra, 2019; Wiesemann et al., 2014) and robust Markov decision processes (Iyengar, 2005; Nilim & El Ghaoui, 2005) that allow to construct decision rules that explicitly take potential model misspecification into account.

6.2 Uncertainty quantification

There are numerous sources of uncertainty in the policy forecasts produced by a calibrated EKW model. The model is subject to misspecification, its numerical implementation introduces approximation error, the data is subject to measurement error, and the calibrated parameters remain partly uncertain. However, economists display incredible certitude as they disregard all uncertainty (Manski, 2013) in their forecasts.

In Eisenhauer, Gabler, and Janys (2020), we draw on a rich literature in other disciplines where a proper accounting of the uncertainty in forecasts from complex computational models is mandatory (Saltelli et al., 2008; Saltelli et al., 2004; Smith, 2014). However, uncertainty quantification for EKW models poses several unique challenges. They usually have a large number of uncertain and correlated parameters, and the quantity of interest is time-consuming to compute and a complex function of the model parameters. Using machine learning methods, we set up an emulator that approximates the full model but is fast to evaluate. We revisit the analysis of Keane and Wolpin (1997), Keane and Wolpin (1994) to showcase our approach in a well-known and empirically-motivated setting and characterize the uncertainty in their key findings.

We construct our approximating emulator using recent advances in surrogate modeling (Forrester et al., 2008) and machine learning (Hastie et al., 2008; Murphy, 2012).

6.3 Model validation

The validation of computational models is a prerequisite for their use in other disciplines (Morin et al., 2012; Oberkampf & Roy, 2010). However, it is extremely rare in economics as drastic regime shifts are seldom available in observational data and costly to implement in large-scale experiments.

In Bhuller et al. (2018), we calibrate an EKW model on Norwegian population panel data with nearly career-long earnings histories. Due to the richness of the data, we can validate the model using a mandatory schooling reform. Our data includes substantial geographic variation in compulsory schooling across Norway between 1960 and 1975 as mandatory schooling increased from seven to nine years at different points in time across municipalities. We split our data into a calibration and validation sample. We only use pre-reform data in our calibration, forecast the effect of increasing mandatory schooling by two years, and compare our forecast with the post-reform outcome. Doing so allows us to assess our model's ability to extrapolate individual responses outside the support of our calibration data. We use the validated model to gain insights into the underlying economic mechanisms that generate the effects of the policy and forecast the effects of several policy alternatives.

6.4 Nonstandard expectations

When economists analyze individual decision-making through the lens of an EKW model, they impose rational expectations. The subjective beliefs about the future correspond to the objective transition probabilities induced by the model.

In Eisenhauer, Haan, et al. (2020), we relax this assumption. We analyze and quantify the effect of biased expectations about wage growth in part-time employment on life cycle wage profiles of female workers. We design specific survey questions for the German Socio-Economic Panel (Goebel et al., 2019) and elicit the expected wage trajectories for full-time and part-time employment directly. Thus, we can incorporate the belief elicitation directly in our life cycle model.

References

Adda, J., Dustmann, C., & Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, *125*(2), 293–337. [3]

Aguirregabiria, V., & Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, *156*(1), 38–67. [3]

Altonji, J. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics*, *11*(1), 48–83. [A-2]

Arcidiacono, P., & Ellickson, P. B. (2011). Practical methods for estimation of dynamic discrete choice models. *Annual Review of Economics*, *3*(1), 363–394. [3]

Arrow, K. J. (1951). Alternative approaches to the theory of choice in risk-taking situations. *Econometrica*, *19*(4), 404–437. [16]

Becker, G. S. (1964). *Human capital*. New York City, NY, Columbia University Press. [3, 16]

Bellman, R. E. (1954). The theory of dynamic programming. *Bulletin of the American Mathematical Society*, *60*(6), 503–515. [5]

Ben-Tal, A., El Ghaoui, L., & Nemirowski, A. (2009). *Robust optimization*. Princeton, NJ, Princeton University Press. [16]

Bhuller, M., Eisenhauer, P., & Mendel, M. (2018). The option value of schooling. *Unpublished Working Paper*. [3, 17]

Blundell, R., Costa Dias, M., Meghir, C., & Shaw, J. (2016). Female labor supply, human capital, and welfare reform. *Econometrica*, *84*(5), 1705–1753. [3]

Bureau of Labor Statistics. (2019). *National longitudinal survey of youth 1979 cohort, 1979-2016 (rounds 1-27)*. Columbus, OH, Center for Human Resource Research, Ohio State University. [10, B-1]

Davidson, R., & MacKinnon, J. G. (2003). *Econometric theory and methods*. New York City, NY, Oxford University Press. [7]

Davis, P. J., & Rabinowitz, P. (2007). *Methods of numerical integration*. Cambridge, MA, Dover Books on Mathmatics. [14]

Eisenhauer, P., Gabler, J., & Janys, L. (2020). Uncertainty quantification for structural econometric models. *in preparation for submission*. [16]

Eisenhauer, P., Gabler, J., & Röhrl, K. (2020). Global optmization in structural econometric models. *Unpublished Working Paper*. [15]

Eisenhauer, P., Gabler, J., & Suchy, R. (2020). Advanced numerical integration in structural econometric models. *Unpublished Working Paper*. [14]

Eisenhauer, P., Haan, P., Ilieva, B., Schrenker, A., & Weizsäcker, G. (2020). Biased wage expectations and female labor supply: A structural approach. *Unpublished Working Paper*. [17]

Eisenhauer, P., & Suchy, R. (2020). Robust human capital investment under risk and ambiguity. *Unpublished Working Paper*. [16]

Forrester, D. A. I. J., Sobester, D. A., & Keane, P. A. J. (2008). *Engineering design via surrogate modelling: A practical guide*. Chichester, England, John Wiley & Sons. [17]

Gerstner, T., & Griebel, M. (1998). Numerical integration using sparse grids. *Numerical algorithms*, *18*(3-4), 209. [14]

Gilboa, I. (2009). *Theory of decision under uncertainty*. New York City, NY, Cambridge University Press. [4]

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., & Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik*, *239*(2), 345–360. [17]

Gourieroux, C., & Monfort, A. (1996). *Simulation-based econometrics*. Oxford, United Kingdom, Oxford University Press. [7]

Hansen, L. P. (2015). Uncertainty outside and inside economic models. *Journal of Political Economy*, *122*(5), 945–987. [15]

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning: Data mining, inference, and prediction*. New York City, NY, Springer. [17]

Hungerford, T., & Solon, G. (1987). Sheepskin effects in the returns to education. *Review of Economics and Statistics*, *69*(1), 175–177. [10, A-3]

Iyengar, G. N. (2005). Robust dynamic programming. *Mathematics of Operations Research*, *30*(2), 257–280. [16]

Jaeger, D. A., & Page, M. E. (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *The Review of Economics and Statistics*, *78*(4), 733–740. [10, A-3]

Judd, K. L., & Skrainka, B. S. (2011). High performance quadrature rules: How numerical integration affects a popular model of product differentiation. *CEMMAP Working Paper (CWP03/11)*. [14]

Keane, M. P., & Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, *105*(3), 473–522. [1, 3, 7, 10, 13, 15, 16, A-1, A-4, A-5, B-1, B-2, B-7]

Keane, M. P., & Wolpin, K. I. (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. [15, 16]

Locatelli, M., & Schoen, F. (2013). *Global optimization*. Philadelphia, PA, SIAM. [15]

Machina, M. J., & Viscusi, K. (2014). *Handbook of the economics of risk and uncertainty*. Amsterdam, Netherlands, North-Holland Publishing Company. [4]

Manski, C. F. (2013). *Public policy in an uncertain world: Analysis and decisions*. Cambridge, MA, Harvard University Press. [16]

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*(4), 281–302. [A-3]

Mincer, J. (1974). *Schooling, experience and earnings*. New York City, NY, National Bureau of Economic Research. [A-3]

Morin, A., Urban, J., Adams, P. D., Foster, I., Sali, A., Baker, D., & Sliz1, P. (2012). Shining light into black boxes. *Science*, *336*(6078), 159–160. [17]

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA, MIT Press. [17]

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*(3), 315–335. [4]

Nilim, A., & El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, *53*(5), 780–798. [16]

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. New York City, NY, Springer. [15]

Oberkampf, W. L., & Roy, C. J. (2010). *Verification and validation in scientific computing*. Cambridge, England, Cambridge University Press. [17]

OECD. (2001). *The well-being of nations: The role of human and social capital*. Paris, France, OECD Publication Service. [3]

Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. New York City, NY, John Wiley & Sons. [3, 5]

Rahimian, H., & Mehrotra, S. (2019). Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659 [math.OC]*. [16]

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: The primer*. Chichester, England, John Wiley & Sons. [16]

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. Chichester, England, John Wiley & Sons. [16]

Smith, R. C. (2014). *Uncertainty quantification: Theory, implementation, and applications*. Philadelphia, PA, SIAM. [16]

White, D. J. (1993). *Markov decision processes*. New York City, NY, John Wiley & Sons. [5]

Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, *62*(6), 1358–1376. [16]

Wolpin, K. I. (2013). *The limits to inference without theory*. Cambridge, MA, MIT University Press. [3]

Appendix A Computational implementation

We use the same computational implementation as in Keane and Wolpin (1997). We outline the immediate utility functions for each of the five alternatives. We first focus on their common overall structure and then present their parameterization. Throughout we provide the economic motivation for their specification.

We follow individuals over their working life from young adulthood at age 16 to retirement at age 65. The decision period $t = 16, \ldots, 65$ is a school year, and individuals decide $a \in \mathcal{A}$ whether to work in a blue-collar or white-collar occupation ($a = 1, 2$), to serve in the military ($a = 3$), to attend school ($a = 4$), or to stay at home ($a = 5$).

Individuals are initially heterogeneous. They differ with respect to their initial level of completed schooling h_{16} and have one of four different $\mathcal{J} = \{1, \ldots, 4\}$ alternative-specific skill endowments $\boldsymbol{e} = \left(e_{j,a}\right)_{\mathcal{J}\times\mathcal{A}}$.

The immediate utility $u_a(\cdot)$ of each alternative consists of a non-pecuniary utility $\zeta_a(\cdot)$ and, at least for the working alternatives, an additional wage component $w_a(\cdot)$. Both depend on the level of human capital as measured by their occupation-specific work experience $\boldsymbol{k}_t = \left(k_{a,t}\right)_{a\in\{1,2,3\}}$, years of completed schooling h_t, and alternative-specific skill endowment $\boldsymbol{e}$. The immediate utility functions are influenced by last-period choices a_{t-1} and alternative-specific productivity shocks $\boldsymbol{\epsilon}_t = \left(\epsilon_{a,t}\right)_{a\in\mathcal{A}}$ as well. Their general form is given by:

$$u_a(\cdot) = \begin{cases} \zeta_a(\boldsymbol{k}_t, h_t, t, a_{t-1}) + w_a(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) & \text{if } a \in \{1, 2, 3\} \\ \zeta_a(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) & \text{if } a \in \{4, 5\}. \end{cases}$$

Work experience $\boldsymbol{k}_t$ and years of completed schooling h_t evolve deterministically:

$$\begin{aligned} k_{a,t+1} &= k_{a,t} + 1[a_t = a] \qquad \text{if } a \in \{1, 2, 3\} \\ h_{t+1} &= h_t \;\; + 1[a_t = 4]. \end{aligned}$$

The productivity shocks are uncorrelated across time and follow a multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Given the structure of the utility functions and the distribution of the shocks, the state at time t is $s_t = \{\boldsymbol{k}_t, h_t, t, a_{t-1}, \boldsymbol{e}, \boldsymbol{\epsilon}_t\}$.

Empirical and theoretical research from specialized disciplines within economics informs the exact specification of $u_a(\cdot)$. We now discuss each of its components in detail.

A.1 Non-pecuniary utility

We present the parameterization of the non-pecuniary utility for all five alternatives.

Blue-collar. Equation (A.1) shows the parameterization of the non-pecuniary utility from working in a blue-collar occupation:

$$\begin{aligned}\zeta_1(\boldsymbol{k}_t, h_t, a_{t-1}) = \alpha_1 + c_{1,1} \cdot 1[a_{t-1} \neq 1] + c_{1,2} \cdot 1[k_{1,t} = 0] \\ + \vartheta_1 \cdot 1[h_t \geq 12] + \vartheta_2 \cdot 1[h_t \geq 16] + \vartheta_3 \cdot 1[k_{3,t} = 1].\end{aligned} \tag{A.1}$$

A constant α_1 captures the net monetary-equivalent of on the job amenities. The non-pecuniary utility includes mobility and search costs $c_{1,1}$, which are higher for individuals who never worked in a blue-collar occupation before $c_{1,2}$. The non-pecuniary utilities capture returns from a high school ϑ_1 and a college ϑ_2 degree. Additionally, there is a detrimental effect of leaving the military early after one year ϑ_3.

White-collar. The non-pecuniary utility from working in a white-collar occupation is specified analogously. Equation (A.2) shows its parameterization:

$$\begin{aligned}\zeta_2(\boldsymbol{k}_t, h_t, a_{t-1}) = \alpha_2 + c_{2,1} \cdot 1[a_{t-1} \neq 2] + c_{2,2} \cdot 1[k_{2,t} = 0] \\ + \vartheta_1 \cdot 1[h_t \geq 12] + \vartheta_2 \cdot 1[h_t \geq 16] + \vartheta_3 \cdot 1[k_{3,t} = 1].\end{aligned} \tag{A.2}$$

Military. Equation (A.3) shows the parameterization of the non-pecuniary utility from working in the military:

$$\zeta_3(k_{3.t}, h_t) = c_{3,2} \cdot 1[k_{3,t} = 0] + \vartheta_1 \cdot 1[h_t \geq 12] + \vartheta_2 \cdot 1[h_t \geq 16]. \tag{A.3}$$

Search costs $c_{3,1} = 0$ are absent but there is a mobility cost if an individual has never served in the military before $c_{3,2}$. Individuals still experience a non-pecuniary utility from finishing high-school ϑ_1 and college ϑ_2.

School. Equation (A.4) shows the parameterization of the non-pecuniary utility from schooling:

$$\begin{aligned}\zeta_4(k_{3,t}, h_t, t, a_{t-1}, e_{j,4}, \epsilon_{4,t}) = e_{j,4} + \beta_{tc_1} \cdot 1[h_t \geq 12] + \beta_{tc_2} \cdot 1[h_t \geq 16] \\ + \beta_{rc_1} \cdot 1[a_{t-1} \neq 4, h_t < 12] \\ + \beta_{rc_2} \cdot 1[a_{t-1} \neq 4, h_t \geq 12] + \gamma_{4,4} \cdot t \\ + \gamma_{4,5} \cdot 1[t < 18] + \vartheta_1 \cdot 1[h_t \geq 12] \\ + \vartheta_2 \cdot 1[h_t \geq 16] + \vartheta_3 \cdot 1[k_{3,t} = 1] + \epsilon_{4,t}.\end{aligned} \tag{A.4}$$

There is a direct cost of attending school such as tuition for continuing education after high school β_{tc_1} and college β_{tc_2}. The decision to leave school is reversible, but entails adjustment costs that differ by schooling category $(\beta_{rc_1}, \beta_{rc_2})$. Schooling is defined as time spent in school and not by formal credentials acquired. Once individuals reach a certain amount of schooling, they acquire a degree. There is no uncertainty about grade completion (Altonji, 1993) and no part-time enrollment. Individuals value the completion of high-school and graduate school $(\vartheta_1, \vartheta_2)$.

Home. Equation (A.5) shows the parameterization of the non-pecuniary utility from staying at home:

$$\begin{aligned}\zeta_5(k_{3,t}, h_t, t, e_{j,5}, \epsilon_{5,1}) = e_{j,5} &+ \gamma_{5,4} \cdot 1[18 \leq t \leq 20] + \gamma_{5,5} \cdot 1[t \geq 21] \\ &+ \vartheta_1 \cdot 1[h_t \geq 12] + \vartheta_2 \cdot 1[h_t \geq 16] \\ &+ \vartheta_3 \cdot 1[k_{3,t} = 1] + \epsilon_{5,t}.\end{aligned} \tag{A.5}$$

Staying at home as a young adult $\gamma_{5,4}$ is less stigmatic as doing so while already being an adult $\gamma_{5,5}$. Additionally, possessing a degree $(\vartheta_1, \vartheta_2)$ or leaving the military prematurely ϑ_3 influences the immediate utility.

A.2 Wage component

The wage component $w_a(\cdot)$ for the working alternatives is given by the product of the market-equilibrium rental price r_a and an occupation-specific skill level $x_a(\cdot)$. The latter is determined by the overall level of human capital:

$$w_a(\cdot) = r_a \, x_a(\cdot).$$

This specification leads to a standard logarithmic wage equation in which the constant term is the skill rental price $\ln(r_a)$ and wages follow a log-normal distribution.

The occupation-specific skill level $x_a(\cdot)$ is determined by a skill production function, which includes a deterministic component $\Gamma_a(\cdot)$ and a multiplicative stochastic productivity shock $\epsilon_{a,t}$:

$$x_a(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) = \exp\left(\Gamma_a(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,a}) \cdot \epsilon_{a,t}\right).$$

Blue-collar. Equation (A.6) shows the parameterization of the deterministic component of the skill production function:

$$\begin{aligned}\Gamma_1(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,1}) = e_{j,1} &+ \beta_{1,1} \cdot h_t + \beta_{1,2} \cdot 1[h_t \geq 12] \\ &+ \beta_{1,3} \cdot 1[h_t \geq 16] + \gamma_{1,1} \cdot k_{1,t} + \gamma_{1,2} \cdot (k_{1,t})^2 \\ &+ \gamma_{1,3} \cdot 1[k_{1,t} > 0] + \gamma_{1,4} \cdot t + \gamma_{1,5} \cdot 1[t < 18] \\ &+ \gamma_{1,6} \cdot 1[a_{t-1} = 1] + \gamma_{1,7} \cdot k_{2,t} + \gamma_{1,8} \cdot k_{3,t}.\end{aligned} \tag{A.6}$$

There are several notable features. The first part of the skill production function is motivated by Mincer (1958, 1974) and hence linear in years of completed schooling $\beta_{1,1}$, quadratic in experience $(\gamma_{1,1}, \gamma_{1,2})$, and separable between the two of them. There are so-called sheep-skin effects (Hungerford & Solon, 1987; Jaeger & Page, 1996) associated with completing a high school $\beta_{1,2}$ and graduate $\beta_{1,3}$ education that capture the impact of completing a degree beyond just the associated years of schooling. Also, there is a first-year blue-collar experience effect $\gamma_{1,3}$ while skills depreciate when not employed in a blue-collar occupation in the preceding period $\gamma_{1,6}$. Other work experience ($\gamma_{1,7}, \gamma_{1,8}$) is transferable.

White-collar. The wage component from working in a white-collar occupation is specified analogously. Equation (A.7) shows the parameterization of the deterministic component of the skill production function:

$$\begin{aligned}\Gamma_2(\boldsymbol{k}_t, h_t, t, a_{t-1}, e_{j,2}) = e_{j,2} &+ \beta_{2,1} \cdot h_t + \beta_{2,2} \cdot 1[h_t \geq 12] \\ &+ \beta_{2,3} \cdot 1[h_t \geq 16] + \gamma_{2,1} \cdot k_{2,t} + \gamma_{2,2} \cdot (k_{2,t})^2 \\ &+ \gamma_{2,3} \cdot 1[k_{2,t} > 0] + \gamma_{2,4} \cdot t + \gamma_{2,5} \cdot 1[t < 18] \\ &+ \gamma_{2,6} \cdot 1[a_{t-1} = 2] + \gamma_{2,7} \cdot k_{1,t} + \gamma_{2,8} \cdot k_{3,t}.\end{aligned} \tag{A.7}$$

Military. Equation (A.8) shows the parameterization of the deterministic component of the skill production function:

$$\begin{aligned}\Gamma_3(k_{3,t}, h_t, t, e_{j,3}) = e_{j,3} &+ \beta_{3,1} \cdot h_t \\ &+ \gamma_{3,1} \cdot k_{3,t} + \gamma_{3,2} \cdot (k_{3,t})^2 + \gamma_{3,3} \cdot 1[k_{3,t} > 0] \\ &+ \gamma_{3,4} \cdot t + \gamma_{3,5} \cdot 1[t < 18].\end{aligned} \tag{A.8}$$

Contrary to the civilian sector there are no sheep-skin effects from graduation ($\beta_{3,2} = \beta_{3,3} = 0$). The previous occupational choice has no influence ($\gamma_{3,6} = 0$) and any experience other than military is non-transferable ($\gamma_{3,7} = \gamma_{3,8} = 0$).

Remark 1. Our parameterization for the immediate utility of serving in the military differs from Keane and Wolpin (1997) as we remain unsure about their exact specification. The authors state in Footnote 31 (p. 498) that the constant for the non-pecuniary utility $\alpha_{3,t}$ depends on age. However, we are unable to determine the precise nature of the relationship. Equation (C3) (p. 521) also indicates no productivity shock $\epsilon_{a,t}$ in the wage component. Table 7 (p. 500) reports such estimates.

A.3 Overview of parameters

Table A.1. Overview of parameters in the Keane and Wolpin (1997) extended model.

Parameter	Description
Preference and type-specific parameters	
δ	discount factor
$e_{j,a}$	initial endowment of type j in alternative a specific skills
Common parameters immediate utility	
α_a	return non-wage working conditions
ϑ_1	non-pecuniary premium of finishing high-school
ϑ_2	non-pecuniary premium finishing college
ϑ_3	non-pecuniary premium of leaving military early
Schooling-related parameters	
$\beta_{a,1}$	return additional year of completed schooling
$\beta_{a,2}$	skill premium high-school graduate
$\beta_{a,3}$	skill premium college graduate
β_{tc_1}	tuition cost high-school
β_{tc_2}	tuition cost college
β_{rc_1}	re-entry cost high-school
β_{rc_2}	re-entry cost college
$\beta_{5,2}$	skill premium high-school graduate
$\beta_{5,3}$	skill premium college graduate
Experience-related parameters	
$\gamma_{a,1}$	return same-sector experience
$\gamma_{a,2}$	return squared same-sector experience
$\gamma_{a,3}$	premium having worked in sector before
$\gamma_{a,4}$	return age effect
$\gamma_{a,5}$	return age effect being a minor
$\gamma_{a,6}$	premium remaining in same sector
$\gamma_{a,7}$	return civilian cross-sector experience
$\gamma_{a,8}$	return non-civilian sector experience
$\gamma_{3,1}$	return same-sector experience
$\gamma_{3,2}$	return squared same-sector experience
$\gamma_{3,3}$	premium having worked in sector before
$\gamma_{3,4}$	return age effect
$\gamma_{3,5}$	return age effect being a minor
$\gamma_{4,4}$	return age effect
$\gamma_{4,5}$	return age effect being a minor

continued on next page

continued from previous page

$\gamma_{5,4}$	return age between 17 and 21
$\gamma_{5,5}$	return age older than 20
Mobility and search parameters	
$c_{a,1}$	premium switching to occupation a
$c_{a,2}$	premium for working first time in occupation a
$c_{3,2}$	premium for serving first time in military
Error correlation	
$\sigma_{a,a}$	standard deviation of shock in alternative a
$\sigma_{i,j}$	correlation between shocks of alternative $a = i$ and $a = j$ with $i \neq j$

Note: The listed parameters are represented as an overview. The immediate utilities for the alternatives do not necessarily include all of them.

Appendix B Empirical data

We use the same data as in Keane and Wolpin (1997). They construct their sample based on the National Longitudinal Survey of Youth 1979 (NLSY79) (Bureau of Labor Statistics, 2019). The NLSY79 is a nationally representative sample of young men and women living in the United States in 1979 and born between 1957 and 1964. Individuals were followed from 1979 onwards and repeatedly interviewed about their educational decisions and labor market experiences. Based on this information, individuals are assigned to either working in one of the three occupations, attending school, or simply staying at home. The decision period is the school year.

They restrict attention to white males that turn 16 between 1977 and 1981 and exploit the information collected between 1979 and 1987. Thus individuals in the sample are all between 16 and 26 years old. While the sample initially consists of 1,373 individuals at age 16, this number drops to 256 at the age of 26 due to sample attrition, missing data, and the short observation period. Overall, the final sample consists of 12,359 person-period observations. Within Figure B.1 we depict the sample sizes with respect to the age of the individuals.

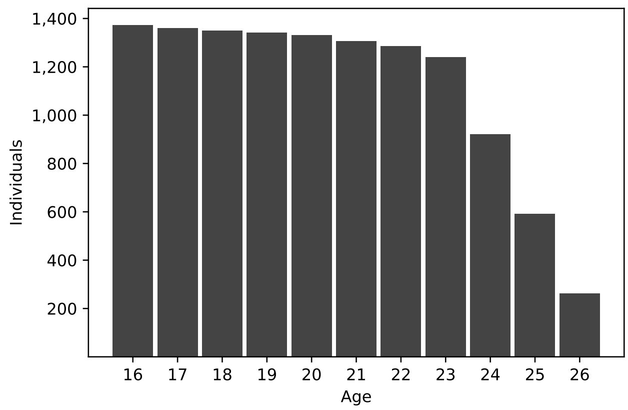


Figure B.1. Sample size with respect to the age.

B.1 Basic structure

First we explore the basic structure of the dataset. All individuals enter the model initially at the same age and are then observed for a varying number of consecutive years. Each year, the individual's decision to work in either a white or blue collar occupation, attend school, enroll in the military, or remain at home is recorded. If working, the dataset potentially also contains that year's wage as a full-time equivalent. We show the data structure for an exemplary individual in Table B.1.

Table B.1. Overview of the dataset structure from Keane and Wolpin (1997).

Period	Age	Experience Schooling	Choice	Wage	Experience Blue-Collar	Experience White-Collar	Experience Military
0	16	11	schooling	-	-	-	-
1	17	12	schooling	-	0	0	0
2	18	13	schooling	-	0	0	0
3	19	14	schooling	-	0	0	0
4	20	15	schooling	-	0	0	0
5	21	16	home	-	0	0	0
6	22	16	white-collar	14,062.7	0	0	0
7	23	16	white-collar	15,921.2	0	1	0
8	24	16	white-collar	18,602.7	0	2	0
9	25	16	white-collar	19,694.0	0	3	0
10	26	16	white-collar	20,611.0	0	4	0

B.2 Basic descriptives

We are able to reproduce the descriptive statistics from Keane and Wolpin (1997) and present some of them below.

Choices. Table B.2 depicts the total and Table B.3 the relative choice frequencies of the individuals clustered by age.

Table B.2. Absolut choice frequencies.

Age	Blue-Collar	White-Collar	Military	Schooling	Home	All
16	45	4	1	1178	145	1373
17	113	15	20	1014	197	1359
18	331	92	70	561	296	1350
19	406	115	107	420	293	1341
20	454	149	113	341	273	1330
21	498	170	106	275	257	1306
22	559	256	90	169	212	1286
23	546	336	68	105	185	1240
24	416	284	44	65	112	921
25	267	215	24	24	61	591
26	127	88	2	13	32	262
All	3762	1724	645	4165	2063	12359

Table B.3. Relative choice frequencies.

Age	Blue-Collar	White-Collar	Military	Schooling	Home
16	3.28	0.29	0.07	85.80	10.56
17	8.31	1.10	1.47	74.61	14.50
18	24.52	6.81	5.19	41.56	21.93
19	30.28	8.58	7.98	31.32	21.85
20	34.14	11.20	8.50	25.64	20.53
21	38.13	13.02	8.12	21.06	19.68
22	43.47	19.91	7.00	13.14	16.49
23	44.03	27.10	5.48	8.47	14.92
24	45.17	30.84	4.78	7.06	12.16
25	45.18	36.38	4.06	4.06	10.32
26	48.47	33.59	0.76	4.96	12.21
All	30.44	13.95	5.22	33.7	16.69

Initially, roughly 86% of the individuals are enrolled in school, but this share steadily declines with age. Nevertheless, about 39% obtain more than a high school degree and continue their education for more than twelve years. As individuals leave school, most of them initially pursue a blue-collar occupation. But the relative share of the white-collar occupation increases as individuals entering the labor market later have higher levels of schooling. At age 26, about 48% work in a white-collar occupation and 34% in a blue-collar occupation. The share of individuals in the military peaks around age 20 with 8%. At its maximum around age 18, approximately 20% of individuals stay at home.

Wages. We reproduce the average real wages by occupation in Table B.4 and plot the standard deviation by age in Figure B.2.

The average wages in the blue-collar occupation are initially highest with about $10,286, while wages in the white-collar occupation and military start around $9,000. However, wages in the white-collar occupation increase steeper over time and overtake blue-collar wages around age 21. At the end of the observation period, wages in the white-collar occupation are about 50% higher compared to blue-collar wages with $32,756 as opposed to only $20,739. Military wages remain lowest throughout.

Table B.4. Average real wages by occupation.

Age	Blue-Collar	White-Collar	Military	All
16	10,286.74	9,320.76	-	10,217.74
17	11,572.89	10,049.76	9,005.36	11,036.60
18	12,603.82	11,775.34	10,171.87	12,060.75
19	12,949.84	12,376.42	9,714.60	12,246.68
20	14,363.66	13,824.01	10,852.51	13,635.87
21	15,313.45	15,578.14	12,619.37	14,977.00
22	16,947.90	20,236.08	13,771.56	17,561.28
23	17,884.95	20,745.56	14,868.65	18,719.84
24	19,245.19	24,066.64	15,910.84	20,942.42
25	21,473.31	24,899.23	17,134.46	22,754.54
26	20,738.91	32,756.04	25,216.83	25,390.90
All	16,436.96	20,295.00	12,255.75	17,115.60

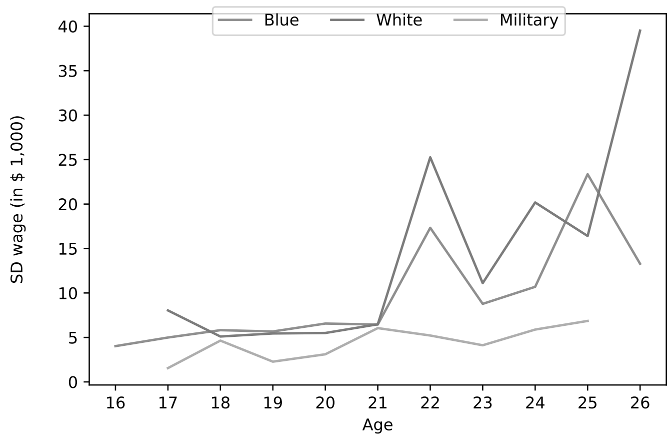


Figure B.2. Wage standard deviation by age.

Note: In the calculation of wage moments we have excluded data points (age) for which only less than 10 observations were available.

Initial schooling. Individuals that enter the model differ with respect to their unobservable type $\{e_{a,t}\}_{a\in\mathcal{A}}$ and the level of initial schooling. The following table and figure illustrate the distribution of initial schooling. Two-thirds of individuals (67.15%) enter with 10 years of schooling, while $20.18 + 4.22 + 0.01 = 24.41\%$ of individuals have less than 10 years of schooling, and 7.5% of individuals were 11 years in school. The average years of initial schooling in the sample amounts to 9.76 years.

Table B.5. Initial years of schooling.

Years	Number	Frequency
7	13	0.01
8	58	0.04
9	277	0.20
10	922	0.67
11	103	0.08

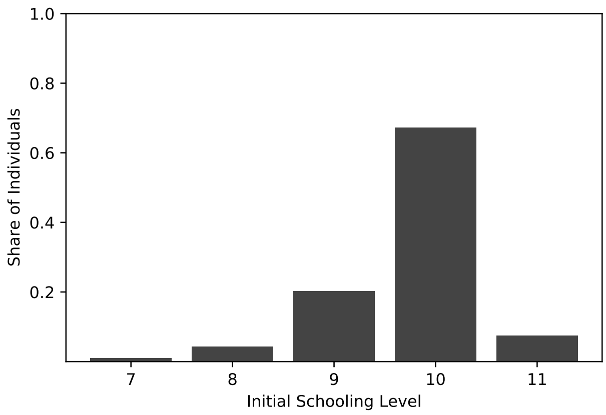


Figure B.3. Distribution of initial schooling.

The amount of schooling an individual obtains is an important determinant for life cycle decisions. The following table illustrates the years spent on each activity by the initial level of schooling.

Table B.6. Average choice frequencies with respect to the initial years of schooling.

	7	8	9	10	11
Blue-Collar	4.15	4.18	4.06	3.15	3.19
White-Collar	0.10	0.27	0.54	1.03	1.21
Military	0.02	0.11	0.35	0.51	0.68
School	0.12	0.29	1.00	2.35	2.56
Home	5.00	4.50	3.21	2.11	1.58
Total	9.39	9.36	9.17	9.16	9.23

Years spent in blue-collar occupations decrease with the initial level of schooling. While individuals with 7 years of initial schooling work on average 4.15 in a blue-collar occupation, individuals that have 11 years of schooling spend only 3.19 years in a blue-collar occupation. Years spent in white-collar occupations increase with the initial level of schooling. In particular, nearly no individual with only 7 years of schooling is occupied in a white-collar job. While individuals at the top distribution spend on average 1.21 years in white-collar ocucpations. The initial level of schooling is a strong predictor for additional schooling. While individuals who enter the model with only 7 years of schooling do not even spend 2 more month in school (on average), those with 11 years of initial schooling add on average 2.5 years. Notably, the descriptives are derived under the instance that the average duration of individuals in the sample differs by the level of initial schooling. Individuals with 7 years of initial schooling are observed for 9.4 periods, while individuals with 10 years of initial schooling are only observed for 9.16 years.

Transition matrix. We illustrate two directions of transition probabilities. Table B.7 depicts the transition probabilities from the choice origin to the choice destination, this is to say, the percentage of individuals who are in an originating alternative in period t (column) and will be in a destination alternative in period $t + 1$ (row).

Table B.7. Transition probabilities from origin to destination.

	Blue-Collar	White-Collar	Military	Schooling	Home
Blue-Collar	0.73	0.10	0.01	0.03	0.12
White-Collar	0.20	0.67	0.01	0.06	0.06
Military	0.10	0.03	0.80	0.01	0.06
Schooling	0.12	0.09	0.02	0.64	0.13
Home	0.31	0.08	0.04	0.10	0.47

Table B.8 depicts the transition probabilities from the choice destination to the choice origin, this is to say, the percentage of individuals who are in a destinating alternative in period t (column) and were in a certain originating alternative in period $t - 1$ (row).

Table B.8. Transition probabilities from destination to origin.

	Blue-Collar	White-Collar	Military	Schooling	Home
Blue -Collar	0.62	0.07	0.02	0.13	0.16
White-Collar	0.18	0.52	0.01	0.20	0.09
Military	0.04	0.01	0.73	0.11	0.11
Schooling	0.04	0.03	0.00	0.87	0.06
Home	0.20	0.04	0.02	0.27	0.46

The diagonal indicates that choices are quite persistent. There is limited mobility between white and blue collar occupations. However, more people transition from a white-collar to a blue-collar occupation than from a white-collar to a blue-collar occupation. Blue-collar workers are twice more likely to transition into home than white-collar workers. Individuals who are not in school at a certain will most likely not return to school.

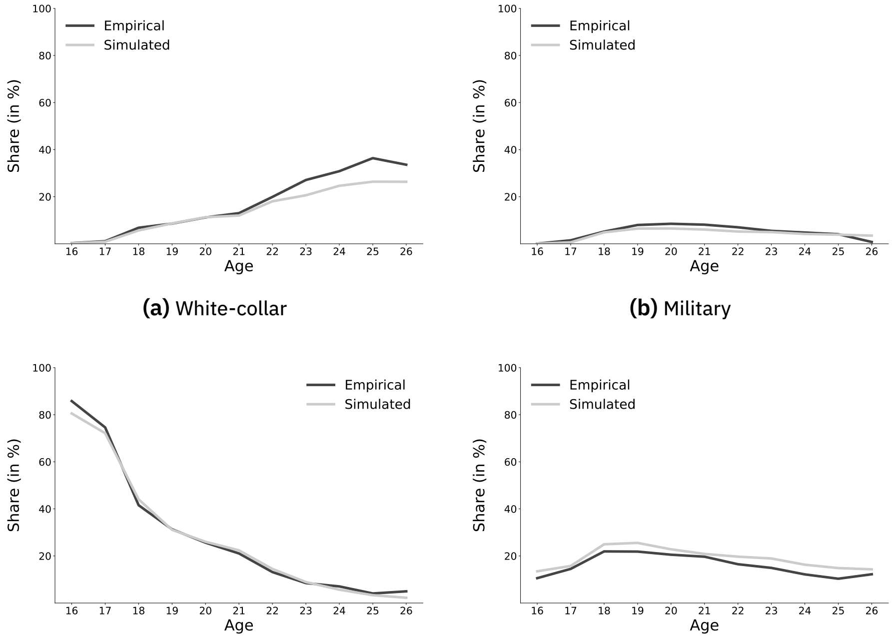
In the original data set choices at age 15 are not available. Consequently, the transition probabilities reported in Table 2 from Keane and Wolpin (1997) cannot be replicated. However, their transition probabilities can be approximated by imputing schooling exprience for individuals at age 15. We use the following rule: If an individual at age 16 has 9 or more years of schooling, then he was in school at age 15. If an individual at age 16 has less than 9 years of schooling, then he was not in school at age 15. The new transition probabilities for schooling are closer to those reported in Keane and Wolpin (1997).

Table B.9. Transition probabilities from origin to destination with imputed schooling exprience for individuals at age 15.

	Blue-Collar	White-Collar	Military	Schooling	Home
Blue-Collar	0.73	0.10	0.01	0.03	0.12
White-Collar	0.20	0.67	0.01	0.06	0.06
Military	0.10	0.03	0.80	0.01	0.06
Schooling	0.10	0.07	0.01	0.70	0.12
Home	0.31	0.08	0.04	0.10	0.47

B.3 Model fit

We provide a full stack of comparisons between simulated and empirical data. We simulate a sample of 1,000 individuals using the calibrated model.



(a) White-collar

(b) Military

(c) School

(d) Home

Figure B.4. Model fit