Uncertainty Quantification for an Eckstein-Keane-Wolpin model with correlated input parameters

Master Thesis Presented to the

Department of Economics at the

Rheinische Friedrichs-Wilhelm-Universität Bonn

in Partial Fulfillment of the Requirements of the Degree of

Master of Science (M.Sc.)

Supervisor: Prof. Dr. Philipp Eisenhauer

Submitted in January 2020 by:
Tobias Stenzel
Matriculation Number: 2971049

Contents

Abbreviations

[see UQ book , center table, two horizontal lines to the top and at the bottom...]

List of Figures

List of Tables

1 Introduction

Forecasts are statements about future events based on past and present data. We use these statements as information to guide our behaviour in order to improve our future. Eventually, the purpose of every science as a whole is to develop forecasts for this very reason. Risk and uncertainty are central to forecasting. The uncertainty that accompanies specific forecast statements is crucial to evaluate how much weight to put on this statement in future decisions. Therefore, to report the accompanying uncertainty should be part of every serious scientific forecast, or, to put it mildly, has to be considered as a good scientific practice.

2 Uncertainty Quantification Framework

2.1 Overview of Uncertainty Quantification

Model-based forecasting includes two main steps: The first step is the calibration. This means, the input parameters of the model are estimated. The second step is the prediction. The prediction contains the evaluation of the estimated parameters with the model to make statements about the future. These statements are made in probabilistic way. Thereby, the uncertainty of these statement is considered.

There are four sources of uncertainty in modern forecasts that are based on complex computational models. The first source, the model uncertainty, is the uncertainty of whether the mathematical model represents the reality appropriately.[1] The second source, the input uncertainty, is the uncertainty about the size of the input parameters of the model. The third one, the numerical uncertainty, comes from potential errors and uncertainties introduced by the conversion from mathematical to computational model. The last source of uncertainty, the measurement uncertainty, is the accuracy of the experimental data that is used to approximate and calibrate the model.

The thesis deals with the second source of uncertainty, the input uncertainty. In my view, this is the source for which uncertainty quantification offers the most instruments and also the strongest instruments. This results from the fact, that the estimation step yields basic measures for the variability or uncertainty in the input parameter estimates. These can then be used to compute a wide variety of measures for the input uncertainty.

The following explains the basic notation for the quantification of the input uncertainty. An essential step is to define the quantity that one wants to predict from the model. It is called Quantity of Interest (henceforth QoI) and denoted by q. The uncertain model parameters are denoted by vector $\boldsymbol{\theta}$. The function that computes QoI q by evaluating a model and, if necessary, post-processing the model output is denoted by $\mathcal{M}$. Thus,

$$q = \mathcal{M}(\boldsymbol{\theta}). \tag{1}$$

Large-scale UQ applications draw from various disciplines like probability, statistics, analysis and numeric. They are used in a combined effort for parameter estimation, surrogate model construction, parameter selection, uncertainty propagation, LSA and GSA, amongst others.

Parameter estimation covers the calibration step. There is a large number of estimation techniques for various types of models. The thesis uses a maximum likelihood approach

[1] However, It seems that there are not many powerful instruments to evaluate and improve the model uncertainty except of comparing statements derived from the model to the data and then improving it where approriate.

detailed in the Model and Estimation section. However, the parameter estimation is is not the main focus.

If the run time of a model is too long to compute some UQ measures, surrogate models are constructed to substitute the original model $\mathcal{M}$. These surrogate models are functions of the model input parameters that are faster to evaluate. They are also called interpolants, because these functions are computed from a random sample of input parameter vectors drawn from the input distribution and evaluated by the model. Therefore, evaluations of the surrogate model interpolate this sample. Some techniques yield functions that have properties which simplify the computation of some UQ measures tremendously.

Another way to reduce computation, not necessarily of the model, but of UQ measures, is to reduce the number of uncertain input parameters. This is the approach that the thesis takes vice versa the construction of a surrogate model. The technique is called Morris sampling and detailed in the following GSA subsection.

Uncertainty propagation is the core of the prediction step. It comprises the construction of the QoI's probability distribution by propagating the input uncertainty through the model. For instance, this can be achieved by repeatedly evaluating a sample of random input parameters by the model (as also required for the construction of a surrogate model). Uncertainty propagation also involves the computation of descriptive statistics like the probabilities for a set of specific events in the QoI range. Conceptually, this is simple. The results are presented in the Uncertainty Propagation section.

LSA

The subfields I pick and do not pick.

QoI explanation.

correlated inputs

2.2 Global Sensitivity Analysis

2.2.1 Morris Screening

2.2.2 Sobol' Indices

$$S_i = \frac{\mathrm{Var}_i[Y|X_i]}{\mathrm{Var}[Y]} \tag{2}$$

$$S_i = \frac{\mathrm{Var}_i[\mathbb{E}_{\sim i}[Y|X_i]]}{\mathrm{Var}[Y]} \tag{3}$$

$$S_{ij} = \frac{\mathrm{Var}_{ij}[\mathbb{E}_{\sim\{i,j\}}[Y|X_i, X_j]]}{\mathrm{Var}[Y]} - S_i - S_j \tag{4}$$

$$S_{\mathrm{u}} = \frac{\mathrm{Var}_{\mathrm{u}}[\mathbb{E}_{\sim\mathrm{u}}[Y|X_{\mathrm{u}}]]}{\mathrm{Var}[Y]} - \sum_{\mathrm{w}\subset\mathrm{u}} S_{\mathrm{w}} \tag{5}$$

$$S_i^{\mathrm{T}} = \sum_{i\in\mathrm{u}} S_{\mathrm{u}} \tag{6}$$

$$\mathrm{Var}[Y] = \mathrm{Var}_i[\mathop{\mathbb{E}}_{\sim i}[Y|X_i]] + \mathop{\mathbb{E}}_{i}[\mathrm{Var}_{\sim i}[Y|X_i]] \tag{7}$$

$$1 = \frac{\mathrm{Var}_i[\mathbb{E}_{\sim i}[Y|X_i]]}{\mathrm{Var}[Y]} + \frac{\mathbb{E}_i[Var_{\sim i}[Y|X_i]]}{\mathrm{Var}[Y]} \tag{8}$$

$$1 = S_i + S_{\sim i}^T \tag{9}$$

$$S_{\sim i}^T = \frac{\mathbb{E}_i[\mathrm{Var}_{\sim i}[Y|X_i]]}{\mathrm{Var}[Y]} \tag{10}$$

$$S_i^T = \frac{\mathbb{E}_{\sim i}[\mathrm{Var}_i[Y|X_{\sim i}]]}{\mathrm{Var}[Y]} \tag{11}$$

$$S_{\mathrm{u}}^{clo} = \frac{\mathrm{Var}_{\mathrm{u}}[\mathbb{E}_{\sim\mathrm{u}}[Y|X_{\mathrm{u}}]]}{\mathrm{Var}[Y]} \tag{12}$$

$$Y = \mathcal{M}(\mathrm{x}) = \mathcal{M}_0 + \sum_{i=1}^{M} \mathcal{M}_i(x_i) + \sum_{1\leq i\leq j\leq M} \mathcal{M}_{ij}(x_i, x_j) + ... + \mathcal{M}_{12..M}(\mathrm{x}) \tag{13}$$

$$S_i = \frac{\mathrm{Cov}[\mathcal{M}_i(x_i), Y]}{\mathrm{Var}[Y]} \tag{14}$$

$$S_i = \frac{\mathrm{Var}[\mathcal{M}_i(x_i)]}{\mathrm{Var}[Y]} + \frac{\mathrm{Cov}[\mathcal{M}_i(x_i)]}{\mathrm{Var}[Y]} \tag{15}$$

$$S_i = \frac{\mathrm{Var}_i[\mathcal{M}_i(x_i)]}{\mathrm{Var}[Y]} \tag{16}$$

$$S_{ij} = \frac{\mathrm{Var}_{ij}[\mathcal{M}_{ij}(x_i, x_j)]}{\mathrm{Var}[Y]} \tag{17}$$

$$\mathrm{Var}[Y] = \sum_{i=1}^{M} \mathrm{Var}[\mathcal{M}_i(x_i)] + \sum_{1 \leq i \leq j \leq M} \mathrm{Var}[\mathcal{M}_{ij}(x_i, x_j)] + ... + \mathrm{Var}[\mathcal{M}_{12..M}(\mathrm{x})] \tag{18}$$

$$\begin{aligned} S_i^T = S_i + \sum_{j \neq i} S_{ij} + \sum_{1 \leq i \leq j \leq M, \{j,k\} \neq i} S_{ijk} + ... = \sum_{i \in \mathrm{w}} S_{\mathrm{w}} = \\ \frac{1}{\mathrm{Var}[Y]} \sum_{i \in \mathrm{w}} \mathrm{Var}_i[\mathcal{M}_{\mathrm{w}}(x_{\mathrm{w}})] \end{aligned} \tag{19}$$

$$S_{\mathrm{u}}^{clo} = \frac{\mathrm{Var}_{\mathrm{u}}[\mathcal{M}_{\mathrm{u}}(\mathrm{x}_{\mathrm{u}})]}{\mathrm{Var}[Y]} + \sum_{\mathrm{w} \subseteq \mathrm{u}} \frac{\mathrm{Var}_{\mathrm{w}}[\mathcal{M}_{\mathrm{w}}(\mathrm{x}_{\mathrm{w}})]}{\mathrm{Var}[Y]} \tag{20}$$

2.3 Surrogate Models and Spectral Expansions

[Scheidegger: Also called Interpolator in the literature]

[Univariate Effects as a measure for comparative statics]

[Philipp: Please add a plot to your thesis (not our notebook) that implements the idea of the uncertainty cone in Figure 1. 2 in our textbook. For example, Figure 1 from KW97 could use such a cone for hte out of support predictions in the occupational shares.]

3 Uncertainty Quantification in the Economic Literature

The need for UQ as an essential part of quantitative economic studies has long been recognized in the economics profession.[2] Also GSA in particular has had strong advocates.[3] However, the demanded evolution of research practice has only been met by a few publications until today. This literature review summarizes these publications with regards to the UQ subfields that are emphasized in the prior section. These are uncertainty propagation and GSA. Table 1 gives an overview of the major topics, analyses, measures and methods in the literature.

Table 1. Overview of UQ literature

Content	Number of articles
Topics	
Climate economics	8
Macroeconomics	4
Analyses	
Uncertainty propagation	8
Globabl sensitivity analysis	7
Local sensitivity analysis	2
Measures	
Sobol' indices	6
Univariate effects	4
Density-based measures	2
Methods	
Monte Carlo sampling	7
Latin hypercube sampling	3
Surrogate model	7
Polynomial chaos expansions	2
Intrusive methods	2
	14

I find 14 contributions that meet the described criteria. Arguably, because UQ is more accomplished in climatology, a large share of research comes from climate economics. Another field where UQ finds some application is macroeconomics. Remarkably, no contribution computes their own estimates for the parametric model uncertainty. The earlier publications tend to use the conceptually simple Monte Carlo uncertainty propagation.

[2]See Hansen and Heckman (1996), Kydland (1992) and Canova (1994), amongst others.
[3]See Canova (1995) and Gregory and Smith (1995).

However, some prefer Latin hypercube sampling. The idea of Latin hypercube sampling is to improve the speed with which the random draws cover the whole variable range. For this purpose, the range is divided into equally probable intervals. Then, one draws only once from each possible interval combination by discarding further draws of the same combinations. The later contributions focus on GSA. Harenberg et al. (2019) gives a well-argued explanation about why GSAs are better than LSAs. GSA measures are Sobol' indices, univariate effects and two density-based measures. The majority of papers use surrogate models to save computation time. The recent works use more sophisticated methods like polynomial chaos expansions (as first applied in Harenberg et al. (2019)) or intrusive approaches (see, for instance, Scheidegger and Bilionis (2019)). This section concludes by explaining the choice of measures and methods made in this thesis and by comparing them to those used in the literature.

Harrison and Vinod (1992) suggest to use uncertainty propagation via Monte Carlo sampling for applied general equilibrium modeling to inspect the uncertainty in model inputs. As a showcase, they propagate the distributions of 48 elasticities through a taxation model by drawing 15,000 input parameter vectors. They analyse their results graphically, using a histogram for their QoI as well as confidence intervals for its mean. For further use, N denotes the size of a Monte Carlo sample.

Canova (1994) proposes to perform a Monte Carlo uncertainty propagation to reflect upon the calibration of dynamic general equilibrium models. The author also addresses challenges and methods for parameter calibration. Canova illustrates his approach by plotting distributions and computing moments and prediction intervals for QoIs in an asset-pricing (N=10,000) and a real business cycle model (N=1,000). Moreover, he analyzes the QoIs' sensitivity towards the uncertainty of individual input parameters by propagating different specifications of input distributions.

More recent examples for Monte Carlo uncertainty propagation investigate climate models, such as Webster et al. (2012). Examples using Latin hypercube sampling are Mattoo et al. (2009) and Hope (2006).

Recently, Harenberg et al. (2019) compare measures from LSA to measures from GSA for multiple QoIs of the canonical, macroeconomic real business cycle model. Thereby, they provide a context for GSA within UQ. The computed sensitivity measures are Sobol' indices and univariate effects. They are obtained by polynomial chaos expansions. For this purpose, Harenberg et al. introduce the leave-one-out error estimator as a measure to select an orthogonal polynomial as the surrogate model.

The concept behind this estimator is the following: Take an arbitrary set A of a large number of n input parameter vectors. From this set, create a set B of n sets that contains every possible permutation of set A but leaving out one parameter vector. Then, for each candidate surrogate model specification, first, compute n surrogate models by evaluating each element of set B. Second, for each specification, compute the mean of the squared errors between actual and surrogate model evaluated at each element of B. This is the leave-one-out error. Finally, one chooses a surrogate model (computed from an arbitrary element of B) for the specification with the lowest error.

The authors come to the following conclusion: On the one hand, a LSA can easily be misleading because its perspective is not broad enough. In particular, they criticise the one-at-a-time approach on which LSAs rely. One-at-a-time methods base on changing one uncertain parameter while keeping the others constant. The choice of parameter combinations tends to be arbitrary. These methods are typically used in economics. The authors conclude that LSA is neither adequate for identifying the inputs that drive the uncertainty, nor does it allow to analyse interactions. On the other hand, a GSA can provide profound insights, and polynomial chaos expansions are a fast way to compute approximations for the respective global sensitivity measures.

Ratto (2008) presents global sensitivity measures for multiple variants of DSGE models computed by Monte Carlo methods and surrogate models. The first measure is density based and derived from the Smirnov test (see, e.g., Hornberger and Spear (1981)): The QoI range is partitioned into a desired set S, and an undesired set $\overline{S}$. Then a Monte Carlo sample of parameter vectors from the input distribution is propagated through the model. From the QoI realizations for each set, two cumulative distribution functions for each input parameter, one conditioned on QoI realizations in set S, and the other conditioned on realizations in set $\overline{S}$, are generated. For each input independently, it is tested whether the distributions differ. If they do, the parameters and their specific regions that lead to the undesired QoI realizations can directly be identified. The second measure is first-order Sobol' indices. Ratto computes them by employing two different surrogate models. The first surrogate is obtained by state-dependent regression. The idea is to regress the QoI on (combinations of) input parameters. The second surrogate is a polynomial representation of the first one. The author finds that the surrogates provide a good fit for the Monte Carlo sample except for the distribution tails. The fit varies conditional on different input parameters. Ratto compares his results for the first-order Sobol' indices computed by both surrogates. The results show some differences in size but not in ranking.

Saltelli and D'Hombres (2010) criticise the arbitrary input value choices in the sensitivity analysis design of the influential Stern (2007) report about the consequences of climate change. Particularly, Stern argue that their cost-benefit analysis' results about the economic impact of climate change are robust towards the uncertainty in their input parameters. Yet, Saltelli and D'Hombres (2010) contradict Stern's assertion by presenting a more thorough sensitivity analysis with parameter choices that better represent the original input distribution.

A series of papers (Anderson et al. (2014), Butler et al. (2014), Miftakhova (2018)) conducts sensitivity analyses for the Dynamic Integrated Climate-Economy model in Nordhaus (2008), in short DICE model. Each work concludes that a GSA is superior to a LSA for the same reasons as Harenberg et al. (2019). Furthermore, all contributions find that leaving some hypothetically low-impact parameters out of the sensitivity analyses lead Nordhaus to neglect the uncertainty in important parameters.

Anderson et al. (2014) use Sobol' indices, the δ-sensitivity measure, and correlation measures for paired QoIs in their GSA. The δ-sensitivity measure (see, e.g., Borgonovo (2006)) is density-based. It is given by half the expectation value of the absolute difference between the unconditional distribution of a QoI and the QoI distribution conditioned on one specific, fixed input (group). Estimates for these measures are computed with the algorithm used in Plischke et al. (2013) applied to a Monte-Carlo sample (N=10,000). In Anderson et al. (2014), the δ-sensitivity measure is the main measure of sensitivity and used to rank the parameters in terms of their contributions to the model uncertainty. The authors also use a surrogate model obtained through Cut-HDMR (Cut-High Dimensional Model Representation; see, e.g., Ziehn and Tomlin (2009)) for graphical analyses of the interaction between input parameters.

Butler et al. (2014) also generate importance rankings for the uncertainty in input parameters. However, they use first, second and total order Sobol' indices instead of the δ-sensitivity measure. They compute the Sobol' indices based on Sobol' sequences (Sobol' (1967)) for the results and based on Latin Hypercube sampling (McKay et al. (1979)) as a check. The results in Butler et al. (2014) and Anderson et al. (2014) can not be compared as they analyse different QoIs.

Miftakhova (2018) applies the GSA procedure outlined by Harenberg et al. (2019). The importance ranking that she obtains from the polynomial-chaos-expansions-based Sobol' indices is different from the ranking that Anderson et al. (2014) obtain from the δ-sensitivity measure. Yet, this is not mentioned by Miftakhova.[4] However, the author emphasizes that the standard procedure for obtaining Sobol' indices from a variance decomposition as used by Anderson et al. (2014) and Butler et al. (2014) is not feasible for the DICE model because a set of input parameters is calibrated jointly in order to let the model match some observables. Therefore, although these input parameters are not correlated in the classical sense, they are dependent. Hence, the variance-based Sobol decomposition is not applicable because the summands are not orthogonal to each other or, in other words, the input-specific variance terms contain a covariance component. Thus, they do not add to the total model variance and Sobol' indices cannot be computed directly. Miftakhova (2018) shows how the set of dependent input parameters can be changed to a set of independent parameters by changing the model structure: She includes uncertain

[4] I do not have access to the numerical codes. Thus the reasons for the discrepancies remain unclear.

observables as independent parameters and reformulates dependent input parameters as endogenous variables. These endogenous variables are functions of the remaining, formerly dependent parameters and the new input parameters.[5]

Gillingham et al. (2015) conduct an UQ for six major climate models. They select three input parameters that are present in each model. The authors generate a surrogate model from regressing several model outputs separately on a linear-quadratic-interaction specification of the three input parameters on 250 grid points. Then they draw 1,000,000 parameter vectors randomly from the probability density function of the input parameters and evaluate the sample with the surrogate model. They find that the parametric uncertainty contributes to more than 90% whereas the differences in the six models contribute to less than 10% of the QoI variances for the year 2100. They also present QoI values for multiple percentiles of each input parameter.

Most recently, Scheidegger and Bilionis (2019) made a noteworthy contribution that naturally connects the solution process of economic models to UQ with surrogate models. The difference to the prior contributions is that their method is intrusive instead of non-intrusive (see page XX). In particular, they conduct an uncertainty propagation and compute univariate effects. Scheidegger and Bilionis' approach is to solve very-high-dimensional dynamic programming problems by approximating and interpolating the value function with a combination of the active subspace method (see, e.g., Constantine (2015)) and Gaussian process regression (see, for example, Rasmussen and Williams (2005)) within each iteration of the value function iteration algorithm. The authors can apply their method up to a 500-dimensional stochastic growth model. Therefore, they can solve models that contain substantial parameter heterogeneity. The link to UQ is that one can also "directly solve for all possible steady state policies as a function of the economic states and parameters in a single computation" (Scheidegger and Bilionis, 2019, p. 4) from the value function interpolant. In other words, this step yields the QoI expressed by a surrogate model. Thus, to add an UQ, one has to, first, specify the uncertain parameters as continuous state variables, and second, assign a probability distribution to each of these parameters. Then (assuming the uncertain input parameters are independent), one provides a sample from each parameter's distribution as input to the Gaussian process regression to obtain a surrogate model. Following these steps, QoIs can be expressed as functions of the uncertain input parameters without much additional effort. Finally, by using a processed value function interpolant as a surrogate model, Scheidegger and Bilionis propagate the model uncertainty and depict univariate effects.

Building on the contributions by Harenberg et al. (2019) and Scheidegger and Bilionis (2019), Usui (2019) conducts a GSA based on Sobol' indices and univariate effects

[5] For a discussion of more general methods to compute Sobol' indices in the presence of dependent input parameters see, e.g., Chastaing et al. (2015) and Wiederkehr (2018), with references therein.

to study rare natural disasters in a dynamic stochastic economy. Because the repeated model evaluations required to construct an adequate surrogate model are too computationally expensive, they choose to apply a method similar to Scheidegger and Bilionis' intrusive framework. However different to Scheidegger and Bilionis (2019), they generate numerical approximates for their policy functions by time iteration collocation (see, e.g., Judd (1998)) with adaptive sparse grid (see Scheidegger et al. (2018)) instead of Gaussian machine-learning.

The remaining section explains where the thesis places within the literature and what it adds to it.

The first contribution is that it quantifies the uncertainty of a model in a merely microeconomic field, namely the model of occupational choice in Keane and Wolpin (1994).

The second contribution is that the estimates for the parametric uncertainty are computed in the same contribution. This adds an important layer of transparency to the uncertainty quantification.

In line with the literature, the thesis' objective is to quantify the parametric uncertainty in an economic model and to attribute shares of this uncertainty to individual input parameters by means of a GSA. For this purpose, I carry out an uncertainty propagation to get an overview of the QoI's probability distribution, given the joint uncertainty in all input parameters. It is shown that this distribution is bell-shaped. Therefore, the variance-based Sobol' indices are a suitable GSA measure. The density-based measures are discarded because they are tailored to less simple distributions.

The specific methods are derived from the following three properties of the computational model. These are: First, the input parameters are correlated. Second, 27 input parameters is a moderately large number. Third, the computation time of the computational model is not rapidly short. The second point mainly amplifies the third point. The combination of these properties is also a novelty in the economic literature. Therefore, the approach that is explained in the following is another contribution to this literature.

There are multiple ways to deal with the moderate computational costs and the parameter dependencies. Each has its advantages and disadvantages. However, the general approach of this thesis is to circumvent unnecessary degrees of methodical complexity to avoid errors. Thus, I apply the following methods: First, I decrease the number of uncertain input parameters through Morris screening as part of an intermediate GSA. This approach is chosen over the alternative of constructing a surrogate model. Second, a quasi-Monte Carlo scheme is used to compute the Sobol' indices in the presence of parameter dependencies. The discarded alternatives are the decorrelation techniques Rosenblatt and Nataf transformations.

In summary, the thesis makes the following contributions. First, it quantifies the parametric uncertainty for a merely microeconomic model. Second, it computes own estimates for this uncertainty. The last contributions emphasize that the thesis is the first contribution that carries out a GSA for a model with a moderate number of correlated input parameters: Third, it reduces the number of uncertain, correlated input parameters through Morris screening to prepare the next GSA step. Fourth, it computes Sobol' indices for correlated input parameters using a quasi-Monte Carlo scheme.

The next section describes the model in Keane and Wolpin (1994) and the estimation of the joint distribution of its input parameters.

4 Model and Estimation

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

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

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

The section proceeds as follows: First, I introduce the KW94 model specification embedded in the more general Eckstein-Keane-Wolpin framework. This embedding provides additional context to the reader. In the next step, the estimation method simulated maximum likelihood is presented. This approach is used for the structural estimation of the exogenous model parameters. After remarks on the numerical implementation, I show the estimation results. They include the estimates, the standard errors, and the correlations for all parameters. These results constitute the mean vector and the covariance matrix, which are used to characterize the joint input distribution for the UQ in the next section. The section ends by describing the QoI choice.

4.1 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:

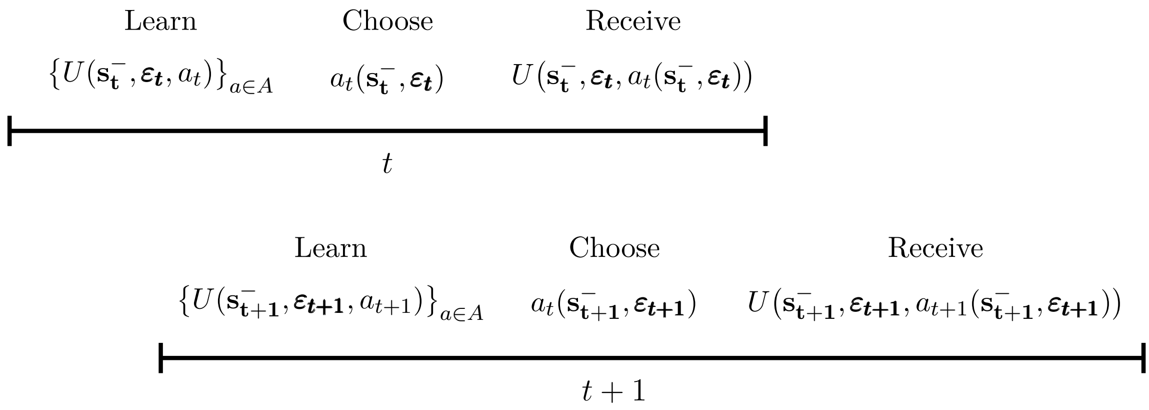
1. Unobservable shocks ε_t do not have to be additively separable from the remainder of the utility functions.
2. Shocks ε_t can be correlated across choices a_t.
3. Observable payoffs, or wages, $W^{-}_{a,t}$ are not conditionally independent from the unobservable shocks ε_t given the observable choices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

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

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

For some occupation alternatives, utility and prior decisions may be intertemporally connected: Agents receive a higher utility if they accumulated skills in past occupations that are useful for these alternatives. Other occupations may not reward experience. The observable part of the state space comprises the period, the work experience and the choice in the previous period. The unobservable part of the state space consists of the alternative-specific shocks $\{\varepsilon_{a,t}\}_{a\in A}$. Figure 1 depicts the series of events.

Figure 1. Series of events



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

Agents are rational and forward-looking. Future utilities are subject to time discount factor $\delta \in [0, 1]$. Hence, they choose their optimal sequence of occupations by maximizing the remaining expected, discounted life-time utility. This maximal value is given by value function $V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t})$ in (21). Like utility U, I also write value function V with different emphasis on occupation choice a_t.

$$V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) = \max_{\{a\}_{t=0}^{T}} \left\{ \sum_{t=0}^{T} \delta^t \int_{\boldsymbol{\varepsilon_t}} U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t) f(\boldsymbol{\varepsilon_t}) d^{|A|}\boldsymbol{\varepsilon_t} \right\} \tag{21}$$

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 $\{\varepsilon_{a,t}\}_{a\in A}$ 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))[6] that breaks up the problem in (21) into more tractable sub-problems along the time dimension.

$$V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) = \max_{a_t} \left\{ U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t) + \delta \int_{\boldsymbol{\varepsilon_t}} \max_{a_{t+1}} V(\mathbf{s_{t+1}^-}, \boldsymbol{\varepsilon_{t+1}}, a_{t+1}) f(\boldsymbol{\varepsilon_{t+1}}) d^{|A|} \boldsymbol{\varepsilon_{t+1}} \right\} \tag{22}$$

The Bellman equation shows, that solving for the whole sequence of policy functions $\{a\}_{t=0}^{T}$ is equivalent to solving iteratively for each optimal, period-specific policy function $a_t(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t})$. For this purpose, choose a_t for each period such that the current period utility and the discounted expected future lifetime utility (given the optimal choice of a_{t+1}) are maximized. The finite time horizon eases the problem as the value function for the last period T simplifies to $V(\mathbf{s_T^-}, \boldsymbol{\varepsilon_T}) = \max_{a_T} U(\mathbf{s_T^-}, \boldsymbol{\varepsilon_T}, a_T)$.[7] With this condition, the problem can be solved for all states by iterating backwards: First, one solves for the final period policy $a_T(\mathbf{s_T^-}, \boldsymbol{\varepsilon_T})$. Then this sub-result is plugged in the future value function on the right hand sight of (22) to solve for $a_{T-1}(\mathbf{s_{T-1}^-}, \boldsymbol{\varepsilon_{T-1}})$, and so forth until $t = 1$. By emphasizing time as a dimension, the policies can also be summarized as one function $a(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t})$. Given random draws for the unobservable shocks $\boldsymbol{\varepsilon_t}$ for each period, this policy is used to simulate the occupational paths for a number of agents.

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

There are four different occupations, b, w, e and h, of which occupations b and w are defined by the same type of utility function. In the following, I will roughly explain how the first two utility functions model characteristics for working in the blue- and in the white-collar sector and how the latter two equations sketch receiving institutional education and staying at home. The parametrization that distinguishes the blue from the white-collar sector and additional intuition is given later in subsection Estimation Results and in Table 2.

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

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

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

(Mincer (1958)):

$$\begin{aligned} U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, b) = W_{b,t}^- &= \exp\Big\{\beta^b + \beta_e^b x_{e,t} + \beta_b^b x_{b,t} + \beta_{bb}^b x_{b,t}^2 + \beta_w^b x_{w,t} + \beta_{ww}^b x_{w,t}^2 + \varepsilon_{b,t}\Big\} \\ U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, w) = W_{w,t}^- &= \exp\Big\{\beta^w + \beta_e^w x_{e,t} + \beta_w^w x_{w,t} + \beta_{ww}^w x_{w,t}^2 + \beta_b^w x_{b,t} + \beta_{bb}^w x_{b,t}^2 + \varepsilon_{w,t}\Big\} \end{aligned} \tag{23}$$

Both equations comprise of a constant term, years of schooling $x_{e,t}$, linear and quadratic terms of occupation experience, and cross-occupational experience and the respective shocks in $\boldsymbol{\varepsilon_t}$. $\boldsymbol{\beta}$ is the vector of coefficients that multiply the previously defined terms.[8] These coefficients are called covariates by many structural economists.

The utilities for education, or schooling, and staying at home are given by the functions in (24). These functions are also called non-pecuniary rewards.

$$\begin{aligned} U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, e) &= \beta^e + \beta_{he}^e \mathbf{1}(x_{e,t} \geq 12) + \beta_{re}^e (1 - \mathbf{1}(a_{t-1} = e)) + \varepsilon_{e,t} \\ U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, h) &= \beta^h + \varepsilon_{h,t} \end{aligned} \tag{24}$$

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.

Write $\{\varepsilon_{a,t}\}_{a \in A}$ as vector $\boldsymbol{\varepsilon_{a,t}}$. It is assumed that $\boldsymbol{\varepsilon_{a,t}}$ follows a joint normal distribution, such that $\boldsymbol{\varepsilon_{a,t}} \sim \mathcal{N}(0, \boldsymbol{\Sigma_\varepsilon})$. $\boldsymbol{\Sigma_\varepsilon}$ denotes the covariance matrix for shocks $\boldsymbol{\varepsilon_{a,t}}$. 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 $\{j, k\} \in \mathbb{N}$ are used to denote subsets of a.

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

4.2 Simulated Maximum Likelihood Estimation

To estimate the exogenous model parameters, the approach that this thesis and also KW94 use is the simulated maximum likelihood method (Albright et al. (1977))[9].

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 $\boldsymbol{\mathcal{D}} \stackrel{\text{def}}{=} (\mathbf{s_t^-}, \mathcal{W}_{a,t}^-)$. Given this setup, the goal

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

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

is to estimate the exogenous model parameters $\boldsymbol{\theta} \stackrel{\text{def}}{=} (\delta, \boldsymbol{\beta}, \boldsymbol{\Sigma_\varepsilon})$.[10] Thus, in the following, every probability is a function of the exogenous model parameters $\boldsymbol{\theta}$. The approach to compute the likelihood function $L_{\mathcal{D}}(\boldsymbol{\theta})$ of the observables in the data begins with the individual latent variable representation in period t.

$$a_t = \underset{a}{\operatorname{argmax}}\, V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t) \tag{25}$$

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

$$\boldsymbol{\varepsilon_t}(a_t(j), \mathbf{s_t^-}) \stackrel{\text{def}}{=} \{\boldsymbol{\varepsilon_t} | V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t(j)) = \max_{a_t} V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t)\}). \tag{26}$$

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,

$$\mathrm{p}\Big(a_t(j) | \mathbf{s_t^-}\Big) = \int_{\boldsymbol{\varepsilon_t}(a_t(j), \mathbf{s_t^-})} f(\boldsymbol{\varepsilon_t}) d^{|A|} \boldsymbol{\varepsilon_t}. \tag{27}$$

The second case is $a_t(k) \in [b, w]$. Assuming the dataset contains wages for the working alternatives $a_t(k)$, the probabilities of choosing $a_t(k)$ take a few steps more to compute. First, note from the wage equations that the the alternative-specific shocks $\boldsymbol{\varepsilon_{a,t}}$ are log normally distributed. Second, in contrary to the non-working alternatives, using (23), 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 wages are logarithmized. Thus,

$$\varepsilon_{a(k),t} = \ln(\mathcal{W}_{a(k),t}^-) - \ln(W_{a(k),t}^-). \tag{28}$$

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 $\mathcal{W}_{a(k),t}^-$, this information can be used to form the expectation about the whole error distribution. Therefore, using the conditional probability density function $f(\boldsymbol{\varepsilon_t} | \varepsilon_{a(k),t})$, the probability of choosing occupation $a_t(k)$ conditional on observed states and wages writes

$$\mathrm{p}\Big(a_t(k) | \mathbf{s_t^-}, W_{a(k),t}^-\Big) = \int_{\boldsymbol{\varepsilon_t}(a_t(k), \mathbf{s_t^-})} f(\boldsymbol{\varepsilon_t} | \varepsilon_{a(k),t}) d^{|A|} \boldsymbol{\varepsilon_t}. \tag{29}$$

[10] 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.

Applying integration by substitution yields the following expression for the probability of the observed wage:[11]

$$\mathrm{p}\Big(\mathcal{W}^{-}_{a(k),t}|\mathbf{s}^{-}_{\mathbf{t}}\Big) = \omega_t^{-1}\frac{1}{\sigma_{a(k)}}\phi\Bigg(\frac{\varepsilon_{a(k),t}}{\sigma_{a(k)}}\Bigg). \tag{30}$$

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 (28) 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 (29) and (30):

$$\mathrm{p}\Big(a_t(k), \mathcal{W}^{-}_{a(k),t}|\mathbf{s}^{-}_{\mathbf{t}}\Big) = \mathrm{p}\Big(a_t(k)|\mathbf{s}^{-}_{\mathbf{t}}, \mathcal{W}^{-}_{a(k),t}\Big)\mathrm{p}\Big(\mathcal{W}^{-}_{a(k),t}|\mathbf{s}^{-}_{\mathbf{t}}\Big) \tag{31}$$

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:

$$L^i_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) = P\Big(\{a^i_{t,}, \mathcal{W}^{-;i}_{a,t,}\}^T_{t=0}\Big) = \prod_{t=0}^{T}\mathrm{p}\Big(a^i_t, \mathcal{W}^{-;i}_{a,t}|\mathbf{s}^{-,\mathbf{i}}_{\mathbf{t}}\Big) \tag{32}$$

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

$$L_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) = P\Big(\Big\{\{a^i_{t,}, \mathcal{W}^{-;i}_{a,t,}\}^T_{t=0}\Big\}_{i\in I}\Big) = \prod_{i\in I}\prod_{t=0}^{T}\mathrm{p}(a^i_t, \mathcal{W}^{-;i}_{a,t}|\mathbf{s}^{-,\mathbf{i}}_{\mathbf{t}}) \tag{33}$$

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 (33). As maximum likelihoods estimates are asymptotically normal[12], these results are taken as the mean vector for the input parameters in the uncertainty quantification.

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

Finally, the calculation of the estimator's covariance is described.[13] 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

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

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

[13]See Verbeek (2012), p. 184-186.

the Fisher information matrix: $\text{Var}(\theta) = \mathcal{I}(\theta)^{-1}$. In this thesis, the information matrix $\mathcal{I}(\boldsymbol{\theta})$ is given by the variance of the scores of the parameters.[14] 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

$$\text{s}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\partial L_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i \in I} \frac{L^i_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \sum_{i \in I} \text{s}_i(\boldsymbol{\theta}). \tag{34}$$

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 (35). It is equal to the inverse of the Fisher information matrix.

$$\mathcal{I}^{-1}(\boldsymbol{\theta}) = \text{Var}(\text{s}(\boldsymbol{\theta})) = \mathbb{E}[\text{s}(\boldsymbol{\theta})\text{s}(\boldsymbol{\theta})']. \tag{35}$$

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

$$\hat{\text{Cov}}_J(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{|I|} \sum_{i \in I} \text{s}_i(\hat{\boldsymbol{\theta}})\text{s}_i(\hat{\boldsymbol{\theta}})' \right)^{-1}. \tag{36}$$

$|I|$ is the number of individuals in the data set. The intuition behind the above expression is the following: Estimator $\hat{\boldsymbol{\theta}}$ maximizes the sample likelihood. This is equivalent to $\hat{\boldsymbol{\theta}}$ setting the sample scores to zero. However, the individual likelihood may not be zero at the optimal parameter vector for the sample likelihood. This variation is captured by the variance of the individual scores evaluated at $\hat{\boldsymbol{\theta}}$. The relations in (34) and (35) then imply that the inverse of the variance of the individual scores is equivalent to the variance of the maximum likelihood estimator.

4.3 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 because it is possible to draw parameter values outside of their domain. Therefore, covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is written in terms of the lower triangular matrix $\boldsymbol{\Sigma}^{\boldsymbol{c}}_{\boldsymbol{\epsilon}}$ obtained from the Cholesky decomposition of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. The contained Cholesky factors are unrestricted and denoted by

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

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 $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$ in parameter vector $\boldsymbol{\theta}$ and its covariance matrix $\text{Cov}(\boldsymbol{\theta})$ that are presented in the previous subsection.

4.4 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 2 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.[15] 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' estimates 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 2. Nonetheless, they are conceptually different. Therefore, measures for their variation, which, by construction, also consider parameter values other than the mean estimates, have to differ. The arguments are: First, own estimates of $\text{SE}(\hat{\boldsymbol{\theta}})$ for the model in terms of standard deviations and correlations of shocks $\boldsymbol{\epsilon_{a,t}}$ are close to those in KW94. Second, the estimates in KW94 for the non-diagonal elements $c_{i,j}$, except of the estimate for $c_{1,2}$, would be unlikely corner solutions. For instance, fix the variance for the shocks in U_e, $\overline{\sigma_e^2}$, and write this variance in terms of the Cholesky factors[16] such that $\overline{\sigma_e^2} = c_{3,1}^2 + c_{3,2}^2 + c_3^2$. There are no restrictions that can force the maximum likelihood

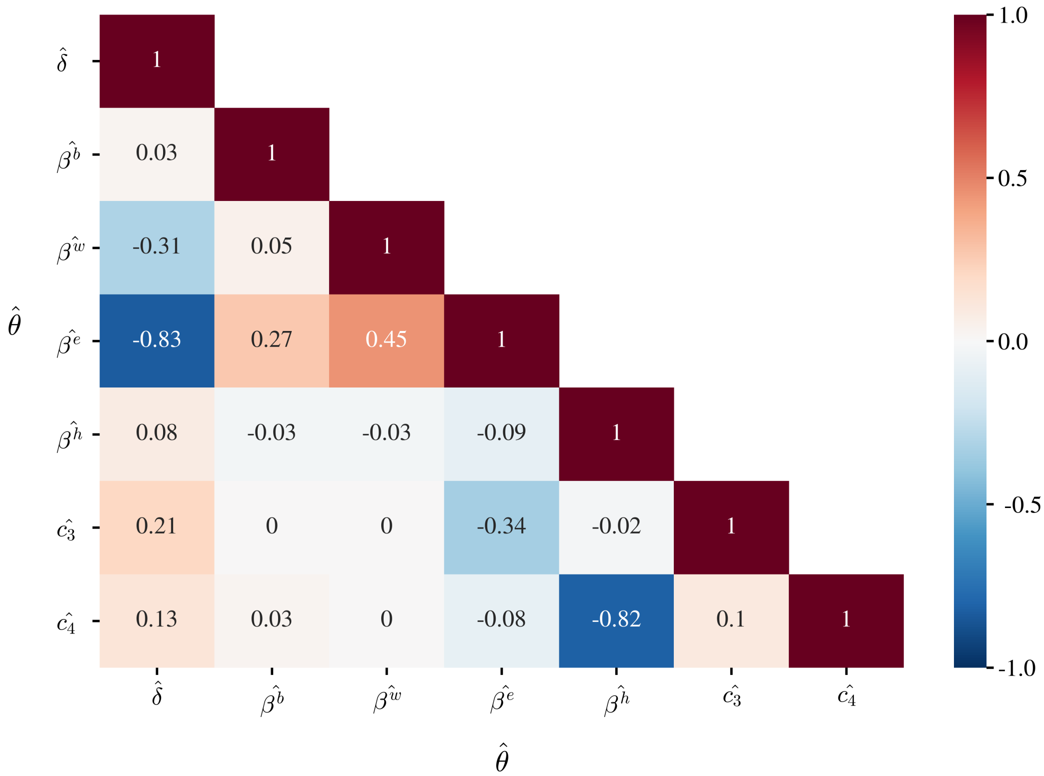
[15] See table 1, p. 658 in Keane and Wolpin (1994); In contrary to the computation of variation measures for $\hat{\boldsymbol{\theta}}$, it is sufficient for obtaining $\hat{\boldsymbol{\theta}}$ to find the individual likelihood of the average agent instead of the sample likelihood in (33).

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

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 and not necessarily for cholesky coefficients. These coefficients are unresticted, and therefore their standard errors have not to be in this specific size. The previous arguments undermine the credibility of the estimation results in KW94, and as a consequence, the thesis proceeds with its own estimates.

Figure 2 depicts the correlations between the estimates of important parameters in $\boldsymbol{\theta}$. In general, the share of high correlations is considerable. Thus, to allow for covariation between the standard errors of the parameter estimates is an important improvement over KW94. The coefficients that stand out are 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 $\boldsymbol{\mathcal{D}}$. For instance, consider an individual that decides for a long occupation in the education sector in the first years and then continues to work in the white-collar sector for the rest of his life. The likelihood to observe this individual increases when 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 (34) around $\hat{\boldsymbol{\theta}}$. It follows from (36) that their standard errors are negatively correlated.

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

4.5 Quantity of Interest

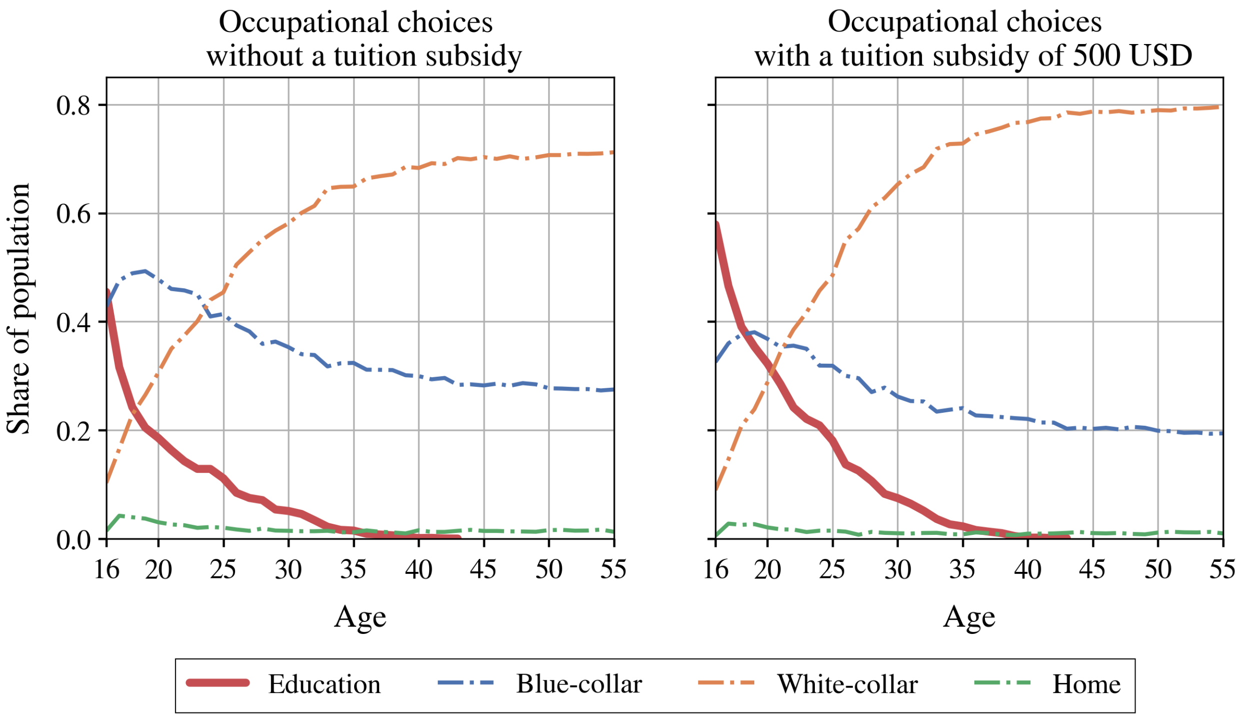
The QoI is the effect of a 500 USD subsidy on annual tuition costs for higher education on the average years of education. Formally, $\beta_{he}^{e,pol} = \beta_{he}^{e} - 500$, where $\beta_{he}^{e,pol}$ represents the subsidised tuition costs. In KW94, the effect is an increase of 1.44 years.[17] The same figure computed with *respy* is 1.5.

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

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

[17] See table 4, p. 668 in Keane and Wolpin (1994).

Figure 3. Comparison of occupation paths between scenarios



The QoI is the sum of the differences between the education shares at each age in the right and in the left graph as depicted by the red lines.[18] This is because the vertical axis can also be interpreted as the share of one year that the average agent is occupied in the education sector. Comparing both graphs, we can see that the tuition subsidy incentivises younger individuals to stay in the education sector for a longer time and older individuals to work in the white-collar sector. The latter observation is a consequence of the first because the white-collar sector rewards education.

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

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

[18]If the red lines would depict continuous functions instead of discrete points in time, the QoI would be the difference between the integrals of the education shares as a function of time in the policy and the base scenario.

Table 2. 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

5 Uncertainty Propagation

6 Global Sensitivity Analysis

[inlcude relative LOO error panel like in Miftakhova]

6.1 Morris screening

6.2 Sobol' Indices

[PCEs do not use Monte Carlo sampling, at least no converging one, i.e. a small number of evalations is enough]

7 Discussion

none

8 Conclusion

none

[Go over (especially capitalizazion of) References]

References

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

Albright, R. S., S. Lerman, and C. F. Manski (1977). *Report on the Development of an Estimation Program for the Multinomial Probit Model.* Cambridge Systematics.

Anderson, B., E. Borgonovo, M. Galeotti, and R. Roson (2014). Uncertainty in climate change modeling: can global sensitivity analysis be of help? *Risk Analysis 34*(2), 271–293.

Bellman, R. E. (1957). *Dynamic Programming.* Princeton, NJ: Princeton University Press.

Borgonovo, E. (2006). Measuring uncertainty importance: investigation and comparison of alternative approaches. *Risk analysis 26*(5), 1349–1361.

Butler, M. P., P. M. Reed, K. Fisher-Vanden, K. Keller, and T. Wagener (2014). Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis. *Environmental modelling & software 59*, 10–29.

Canova, F. (1994). Statistical inference in calibrated models. *Journal of Applied Econometrics 9*(1), 123–144.

Canova, F. (1995). Sensitivity analysis and model evaluation in simulated dynamic general equilibrium economies. *International Economic Review 36*(2), 477–501.

Chastaing, G., F. Gamboa, and C. Prieur (2015). Generalized sobol sensitivity indices for dependent variables: numerical methods. *Journal of Statistical Computation and Simulation 85*(7), 1306–1333.

Constantine, P. G. (2015). *Active subspaces: Emerging ideas for dimension reduction in parameter studies*, Volume 2. SIAM.

Gabler, J. (2019). *A Python Tool for the Estimation of (Structural) Econometric Models.* URL: https://github.com/OpenSourceEconomics/estimagic.

Gillingham, K., W. D. Nordhaus, D. Anthoff, G. Blanford, V. Bosetti, P. Christensen, H. McJeon, J. Reilly, and P. Sztorc (2015). Modeling uncertainty in climate change: A multi-model comparison. Technical report, National Bureau of Economic Research.

Gregory, A. W. and G. W. Smith (1995). Business cycle theory and econometrics. *The Economic Journal 105*(433), 1597–1608.

Hansen, L. P. and J. J. Heckman (1996). The empirical foundations of calibration. *Journal of economic perspectives 10*(1), 87–104.

Harenberg, D., S. Marelli, B. Sudret, and V. Winschel (2019). Uncertainty quantification and global sensitivity analysis for economic models. *Quantitative Economics 10*(1), 1–41.

Harrison, G. W. and H. Vinod (1992). The sensitivity analysis of applied general equilibrium models: Completely randomized factorial sampling designs. *The Review of Economics and Statistics 74*(2), 357–362.

Hope, C. (2006). The marginal impact of co2 from page2002: an integrated assessment model incorporating the ipcc's five reasons for concern. *Integrated assessment 6*(1).

Hornberger, G. M. and R. C. Spear (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management 12*, 7–18.

Judd, K. L. (1998). *Numerical Methods in Economics.* MIT Press.

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.

Kydland, F. E. (1992). On the econometrics of world business cycles. *European Economic Review 36*(2-3), 476–482.

Mattoo, A., A. Subramanian, D. Van Der Mensbrugghe, and J. He (2009). Reconciling climate change and trade policy.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 21*(2), 239–245.

Miftakhova, A. (2018). Global sensitivity analysis in integrated assessment modeling. *Working Paper*.

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

NLSY79 (1990). *National Longitudinal Survey of Youth 1979 Cohort, 1979–1990 (Rounds 1–11).* URL: https://www.nlsinfo.org/content/cohorts/nlsy79.

Nordhaus, W. D. (2008). *A question of balance: economic modeling of global warming.* Yale University Press New Haven.

Plischke, E., E. Borgonovo, and C. L. Smith (2013). Global sensitivity measures from given data. *European Journal of Operational Research 226*(3), 536–550.

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

Rasmussen, C. E. and C. K. I. Williams (2005). *Gaussian Processes for Machine Learning*. MIT press.

Ratto, M. (2008). Analysing dsge models with global sensitivity analysis. *Computational Economics 31*(2), 115–139.

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

Saltelli, A. and B. D'Hombres (2010). Sensitivity analysis didn't help. a practitioner's critique of the Stern review. *Global Environmental Change 20*(2), 298–302.

Scheidegger, S. and I. Bilionis (2019). Machine learning for high-dimensional dynamic stochastic economies. *Journal of Computational Science 33*, 68–82.

Scheidegger, S., D. Mikushin, F. Kubler, and O. Schenk (2018). Rethinking large-scale economic modeling for efficiency: optimizations for GPU and Xeon Phi clusters. In *2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 610–619. IEEE.

Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math Math. Phys. 7*(4), 784–802.

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

Stern, N. H. (2007). *The economics of climate change: the Stern review*. Cambridge University press.

Usui, T. (2019). Adaptation to rare natural disasters and global sensitivity analysis in a dynamic stochastic economy. *Working Paper*.

Verbeek, M. (2012). *A Guide to Modern Econometrics* (4 ed.). Wiley.

Webster, M., A. P. Sokolov, J. M. Reilly, C. E. Forest, S. Paltsev, A. Schlosser, C. Wang, D. Kicklighter, M. Sarofim, J. Melillo, et al. (2012). Analysis of climate policy targets under uncertainty. *Climatic change 112*(3-4), 569–583.

Wiederkehr, P. (2018). Global sensitivity analysis with dependent inputs. Master's thesis, ETH Zurich, Zurich, Switzerland.

Ziehn, T. and A. S. Tomlin (2009). Gui-hdmr - a software tool for global sensitivity analysis of complex models. *Environmental Modelling & Software 24*(7), 775–785.

Appendix

none

none

Affidavit

"I hereby confirm that the work presented has been performed and interpreted solely by myself except for where I explicitly identified the contrary. I assure that this work has not been presented in any other form for the fulfillment of any other degree or qualification. Ideas taken from other works in letter and in spirit are identified in every single case."