# 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 February 2020 by:

Tobias Stenzel

Matriculation Number: 2971049

# Acknowledgements

[Write acknowledgements here.]

# Contents

# Abbreviations

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

USD, GSA, UQ, QoI, LSA

# 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.

The next section introduces the reader to important concepts in the literature. Therefore, it is placed before the literature review.

# 2 Uncertainty Quantification Framework

This chapter consists of three parts. The first section gives an overview of UQ and introduces the basic notation. The second part describes the subdiscipline sensitivity analysis. This part is divided into quantitative and qualitative GSA. I explain the most common measures for both levels of GSA and how they relate. The third part concludes with remarks on the role of correlated input parameters.

## 2.1 Overview of Uncertainty Quantification

Model-based forecasting includes two main steps (Smith (2014))[1]: The first step is calibration. In this step, the input parameters of the model are estimated. The second step is the prediction. The prediction contains the model evaluation at the estimated parameters, This allows us t make statements about potential scenarios. These statements are made in a probabilistic way. Thereby, the uncertainty of these statements is emphasised.[2]

There are four sources of uncertainty in modern forecasting that are based on complex computational models (Smith (2014))[3]. The first source, the model uncertainty, is the uncertainty about whether the mathematical model represents the reality sufficiently.[4] 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 translation of a mathematical to a 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 UQ offers the most and also the strongest instruments. This might result from the fact that the estimation step produces standard errors as basic measures for the variation or uncertainty in the input parameter estimates. These can then be used to compute a variety of measures for the impact of the input uncertainty on the variation in the model output.

The following explains the basic notation. It is essential to define the quantity that one wants to predict with a model. This quantity is called output, or quantity of interest, and is denoted by $Y$. For instance, the QoI in the thesis is the impact of a 500 USD tuition subsidy for higher education on average schooling years. The uncertain model

[1] See p. ix.

[2] The general procedure of model-based forecasting can also include other steps. However, steps like model validation and model verification can also be viewed as belonging to the analysis of the so-called model uncertainty.

[3] See p. 4-7.

[4] However, apparently, there are not many powerful instruments to evaluate and improve the model uncertainty except comparing statements derived from the model to the data and then improving it where appropriate.

parameters $X_1, X_2, ..., X_k$ are denoted by vector $\mathbf{X}$. The function that computes QoI $Y$ by evaluating a model and, if necessary, post-processing the model output is denoted by $f(X_1, X_2, ..., X_k)$. Thus,

$$Y = f(\mathbf{X}). \tag{1}$$

From this point forward, I also refer to $f$ as the model. Large-scale UQ applications draw from various fields such as probability, statistics, analysis, and numeric. These disciplines are used in a combined effort for parameter estimation, surrogate model construction, parameter selection, uncertainty analysis, LSA, and GSA, amongst others. Drawing from Smith (2014)[5], I briefly sketch the first four components. The last two components, local and especially global sensitivity analysis, are discussed more extensively after that.

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, as detailed in the model chapter and in Appendix C.

If the run time of a model is too long to compute the desired UQ measures, surrogate models are constructed to substitute the original model $f$ (McBride and Sundmacher (2019)). These surrogate models are functions of the model input parameters which are faster to evaluate. The functions are also called interpolants because they are computed from a random sample of input vectors, drawn from the input distribution and evaluated by the model. Typically, a surrogate model is computed by minimising a distance measure between a predetermined type of function and the model evaluations at the sample points. Therefore, the surrogate model interpolates this sample. Some specifications, like orthogonal polynomials, have properties which can simplify the computation of UQ measures tremendously (Xiu (2010)).

Another way to reduce the computation time, not directly of the model but of UQ measures, is to reduce the number of uncertain input parameters as part of a parameter selection. The decision which parameters to fix is made based on sensitivity measures. This is called screening or factor fixing (Saltelli et al. (2008))[6]. This point will be taken up again.

Uncertainty analysis is the core of the prediction step. It comprises two steps. The first step is the construction of the QoI's probability distribution by propagating the input uncertainty through the model. For instance, this can be achieved by evaluating a sample of random input parameters (as also required for the construction of a surrogate model). The second step is the computation of descriptive statistics like the probabilities for a set of specific events in the QoI range using this distribution. Both steps are conceptually simple. The construction of the probability distribution is also important for designing subsequent steps like a sensitivity analysis. For example, if the distribution is unimodal and symmetric, variance-based UQ measures are meaningful. If the distribution has a less

[5]See p. 8-10.
[6]See page 33-34.

tractable, density-based measures are better suited (Plischke et al. (2013a)).

## 2.2 Sensitivity Analysis

This section draws from Saltelli et al. (2004) and Saltelli et al. (2008). They define sensitivity analysis as "the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" Saltelli et al. (2004, p. 42). This apportioning implies a ranking of input parameters in terms of their importance for the model output. Saltelli et al. (2004, p. 52) define the most important parameter as "the one that [if fixed to its true, albeit unknown value] would lead to the greatest reduction in the variance of the output Y." Therefore, a factor is not important if it influences output Y directly but rather its variance.

Sensitivity analysis includes different objectives. These have to be determined at first because the choice of methods depends on these objectives. Typically, the main and final goal is factor prioritisation. This is the aforementioned ranking of input parameters in terms of their importance. This ranking can be used to concentrate resources on data acquisition and estimation of a subset of parameters. The methods meeting the demands of factor prioritisation best are called quantitative. These typically require the highest computational effort.

There are multiple other objectives. The one additional objective featured in this thesis is screening/factor fixing. It is basically the same as factor prioritisation except that it only aims to identify the input parameters that can be fixed at a given value without significantly reducing the output variance. Therefore, it focuses on the lowest parameters in the potential importance ranking. The reason, why one would pursue factor fixing instead of factor prioritisation. is computational costs. As factor fixing generates less information than factor prioritisation, less powerful methods can be applied. These methods require less computational resources and are called qualitative. Factor fixing can be used to prepare factor prioritisation for models that are more costly to evaluate. In this sense, it serves the same purpose as surrogate modelling.

Another important distinction is local versus global sensitivity analysis. It essentially refers to the applied methods. In fact, the definition by Saltelli et al. (2004) is already tailored to a global sensitivity analysis. In contrast to the given definition, "until quite recently, sensitivity analysis was [...] defined as a local measure of the effect of a given input on a given output" Saltelli et al. (2004, p. 42). This older definition differs from the definition used here in two aspects. First, it emphasises the level of output rather than its variance. Second, it describes the measure as a local one. The drawback of this approach becomes clear by considering an example of a local sensitivity measure. This measure is the so-called system derivate $D_i = \frac{\partial Y}{\partial X_i}$ (Rabitz (1989)). The derivative is typically computed at the mean, $\overline{X_i}$, of the estimate for $X_i$. $D_i$ is a so-called one-at-a-time measure

because it changes only one factor. It has the following four drawbacks: First, it does not account for the interaction between multiple input parameters because it is one-at-a-time. Second, If the model derivation is not analytical, the choice of the (marginal) change in $X_i$ is arbitrary. Third, the local derivative at $\overline{X_i}$ is only representative for the whole sample space of a random input if the model is linear in $X_i$. Fourth, the measure does not relate to the output variance $Var(Y)$. For these reasons, the field, its definitions and its methods have evolved beyond the notion of local sensitivity analysis. Yet, until recently, the main part of applications in different disciplines, such as physics (Saltelli et al. (2004))[7] or economics (Harenberg et al. (2019)), still uses local measures.

### 2.2.1 Quantitative Global Sensitivity Analysis

Quantitative GSA aims to determine the precise effect size of each input parameter and its variation on the output variation. The most common measures in quantitative GSA are the Sobol' sensitivity indices. Equation (2) gives the general expression for the first order index. Let $\mathrm{Var}_{X_i}(Y|X_i)$ denote the variance of the model output $Y$ conditional on input parameter $X_i$. Then,

$$S_i = \frac{\mathrm{Var}_{X_i}(Y|X_i)}{\mathrm{Var}(Y)}. \tag{2}$$

The equation becomes clearer with the following equivalent expression in Equation (3). For this purpose, let $\sim i$ denote the set of indices except $i$. The expectation of $Y$ for one specific value of $X_i$ equals the average of the model evaluations from a sample, $\boldsymbol{\chi_{\sim i}}$, of $\mathbf{X_{\sim i}}$ and a given value $X_i = x_i^*$. Then, we use $\mathbb{E}[f(X_i = x_i^*, \boldsymbol{\chi_{\sim i}})] \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{X_{\sim i}}}[Y|X_i]$ to write the first-order Sobol' index as the variance of $\mathbb{E}_{\mathbf{X_{\sim i}}}[Y|X_i]$ over all $x_i^*$ as

$$S_i = \frac{\mathrm{Var}_{X_i}\Big(\mathbb{E}_{\mathbf{X_{\sim i}}}[Y|X_i]\Big)}{\mathrm{Var}(Y)}. \tag{3}$$

The first-order index does not include the additional variance in $Y$ that may occur from the interaction of $\mathbf{X_{\sim i}}$ with $X_i$. This additional variance is included in the total-order Sobol' index given by Equation (4). It is the same expression as in Equation (3) except that the positions for $X_i$ and $\mathbf{X_{\sim i}}$ are interchanged. Conditioning on $\mathbf{X_{\sim i}}$ accounts for the inclusion of the interaction effects of $X_i$.

$$S_i^T = \frac{\mathrm{Var}_{\mathbf{X_{\sim i}}}\Big(\mathbb{E}_{X_{\sim i}}[Y|\mathbf{X_{\sim i}}]\Big)}{\mathrm{Var}(Y)} \tag{4}$$

Computing these measures requires many function evaluations, even if an estimator is

[7] See p. 42.

used as a shortcut (Saltelli et al. (2004))[8]. The more time-intense one function evaluation is, the more utility does factor fixing based on qualitative measures provide.

### 2.2.2 Qualitative Global Sensitivity Analysis

Qualitative GSA deals with the computation of measures that can rank random input parameters in terms of their impact on the function output and the variability thereof. This is done to a degree of accuracy that allows distinguishing between influential and non-influential parameters. If the measures for some input parameters are negligibly small, these parameters can be fixed so that the number of random input parameters decreases for a subsequent quantitative GSA. This section explains the qualitative measures and the trade-off between computational costs and accuracy.

The most commonly used measures in qualitative GSA is the mean EE, $\mu$, the mean absolute EEs, $\mu^*$, and the standard deviation of the EEs, $\sigma$. The EE of $X_i$ is given by one individual function derivative with respect to $X_i$. The "change in", or the "step of" the input parameter, denoted by $\Delta$. The only restriction is that $X_i + \Delta$ is in the sample space of $X_i$. The Elementary Effect, or derivative, is denoted by

$$d_i^{(j)} = \frac{f(\mathbf{X}_{\sim \mathbf{i}}^{(\mathbf{j})}, X_i^{(j)} + \Delta^{(i,j)}) - f(\mathbf{X})}{\Delta^{(i,j)}}, \tag{5}$$

where $j$ is an index for the number of $r$ observations of $X_i$. Note, that the EE, $d_i^{(j)}$, is equal to the aforementioned local measure, the system derivate $S_i = \frac{\partial Y}{\partial X_i}$, except that the value $\Delta$ has not to be infinitesimally small. To offset the third drawback of $d_i$ and $S_i$, that base vector $X_i$ does not represent the whole input space, one computes the mean EE, $\mu_i$, based on a random sample of $X_i$ from the input space. The second drawback, that interaction effects are disregarded, is also offset because elements $X_{\sim i}$ are also resampled for each new $X_i$. The mean EE is given by

$$\mu_i = \frac{1}{r} \sum_{j=1}^{r} d_i^{(j)}. \tag{6}$$

Thus, $\mu_i$ is the global version of $d_i^{(j)}$. The standard deviation of the EEs writes $\sigma_i = \sqrt{\frac{1}{r} \sum_{j=1}^{r} (d_i^{(j)} - \mu_i)^2}$. The mean absolute EE, $\mu_i^*$, is used to prevent observations of opposite sign to cancel each other out:

$$\mu_i^* = \frac{1}{r} \sum_{j=1}^{r} \left| d_i^{(j)} \right|. \tag{7}$$

Step $\Delta^{(i,j)}$ may or may not vary depending on the sample design that is used to draw the

[8]See p. 124 -149.

input parameters.

One last variant is provided in Smith (2014)[9]. That is, the scaling of $\mu_i^*$ by $\frac{\sigma_{X_i}}{\sigma_Y}$. This measure is called the sigma-normalized mean absolute EE:

$$\mu_{i,\sigma}^* = \mu_i^* \frac{\sigma_{X_i}}{\sigma_Y}. \tag{8}$$

This improvement is necessary for a consistent ranking of $X_i$. Otherwise, the ranking would be distorted by differences in the level of the the input parameters. The reason is that the input space constrains $\Delta$. If the input space is larger, the base value of $X_i$ can be changed by a larger $\Delta$.

From the aforementioned set of drawbacks of the local derivate $D_i = \frac{\partial Y}{\partial X_i}$, two drawbacks are remaining for the EE method. The first drawback is the missing direct link to the variation in $Var(Y)$. The second drawback is that the choice of $\Delta$ is somewhat arbitrary if the derivative is not analytic. To this date, the literature has not developed convincing solutions for these issues.

In an attempt to establish a closer link between EE-based measures and Sobol' indices, Kucherenko et al. (2009) made two conclusions: The first conclusion is that there is an upper bound for the total index, $S_i^T$, such that

$$S_i^T \leq \frac{\frac{1}{r}\sum_{j=1}^{r} d_i^{2(j)}|}{\pi^2 \sigma_Y}. \tag{9}$$

This expression makes use of the squared EE. In light of this characteristic, the use of $\sigma_i$ as a measure that aims to include the variation of $d_i^j$ appears less relevant. Nevertheless, this rescaling makes the interpretation more difficult. The second conclusion is that the Elementary Effects method can lead to false selections for non-monotonic functions. This is also true if functions are non-linear. The reason is linked to the aforementioned second drawback, the arbitrary choice of step $\Delta$. More precisely, depending on the sampling scheme, $\Delta$ might be random instead of arbitrary and constant. In both cases, $\Delta$ can be too large to approximate a derivative. If, for example, the function is highly non-linear of varying degree with respect to the input parameters $\mathbf{X}$, $\Delta > \epsilon$ can easily distort the results. Especially if the characteristic length of function variation is much smaller than $\Delta$.

[9]See p. 332.

## 2.3 Correlated input parameters

So far, all measures described in this chapter assumed uncorrelated input parameters. Typically, this assumption is not met by practical applications as joint estimations of parameters tend to produce correlated estimates. This gives reason to expand sensitivity measures to a more general setting.

Today, several recent contributions deal with the extension of the Sobol' sensitivity measures to the setup with correlated inputs. For instance, estimators for two complementary sets of measures are developed by Kucherenko et al. (2012) and Mara et al. (2015). On the other hand, the only contribution to the computation of EE-based screening measures for correlated parameters is made by Ge and Menendez (2017). Some authors, e.g. Saltelli et al. (2004)[10] even negate the necessity for expanded Elementary Effects due to overcomplexity. Obviously, this can lead to false results.

As the computation time of the Keane and Wolpin (1994) model is considerable[11], the thesis reviews Ge and Menendez (2017) and computes measures derived from this contribution to the occupational choice model. The aim is to fix parameters as preparation for a potential quantitative GSA. The measures correspond to $\mu^*_{i,\sigma}$ and $\mu^*_i$. The measure that Kucherenko et al. (2009) link to the total Sobol' index is discarded because their results are not valid in the presence of correlations.[12]

The next chapter reviews uncertainty analyses and quantitative GSAs for economic models in the literature.

# 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.[13] Also GSA in particular has had strong advocates.[14] 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 two UQ subfields that are emphasized in the prior section. These are uncertainty analysis and quantitative GSA. The review excludes qualitative GSAs because factor fixing is not the objective of the respective publications. A qualitative GSA as a standalone GSA is not considered as best practice (Saltelli et al. (2004))[15]. Table 1 gives an overview of

[10] See p. 46.
[11] On my machine it takes approximately six seconds to compute the QoI. Given that Sobol' indices can require a five digits number of evaluations per input parameter, the computation time is relatively large.
[12] The fundamental reason is that the unique first- and higher-order Sobol' indices do not add to one if input parameters are correlated.
[13] See Hansen and Heckman (1996), Kydland (1992) and Canova (1994), amongst others.
[14] See Canova (1995) and Gregory and Smith (1995).
[15] See page 48.

the major topics, analyses, measures and methods in the literature.

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 model input uncertainty. The earlier publications tend to use the conceptually simple Monte Carlo uncertainty analysis. 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 to construct a surrogate model (as first applied in Harenberg et al. (2019)) or intrusive approaches (see, for instance, Scheidegger and Bilionis (2019)). Intrusive methods require essential changes to the model structure, for instance to the state space, whereas the usual non-intrusive methods leave the model untouched and treat it like a so-called black box.

Harrison and Vinod (1992) suggest to use uncertainty analysis 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 analysis 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 analysis 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. Univariate effects are the conditional expectation of a QoI as a function of one input parameter $X_i$, where the expectations are taken over $X_{\sim i}$. They are equal to the argument in the variance numerator of the first-order Sobol' index in Equation (3). The sensitivity indices and univariate effects 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 due to the reasons detailed in the previous chapter. LSA methods are typically used in economics. The authors conclude that these are neither adequate for identifying the inputs that drive the uncertainty, nor do they 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 argues that this cost-benefit analysis' results about the economic impact of climate change are robust towards the uncertainty in the 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. 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 $\delta$-sensitivity measure, and correlation measures for paired QoIs in their GSA. The $\delta$-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. (2013b) applied to a Monte-Carlo sample (N=10,000). In Anderson et al. (2014), the $\delta$-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 interactions 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 $\delta$-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 $\delta$-sensitivity measure. Yet, this is not mentioned by Miftakhova.[16] 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

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

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 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.[17]

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 input 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. In particular, they conduct an uncertainty analysis 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

[17] 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.

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 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 thesis is distinct from most of the literature in the following points. First, it analyses a labour economic model with a larger number of parameters. Moreover, it uses its own estimates for the input uncertainty. These estimates imply correlated input parameters. Furthermore, the model evaluation is relatively costly. Therefore, this work computes screening measures for factor fixing to prepare a subsequent quantitative GSA.

The next section describes the method in Ge and Menendez (2014) to compute EE-based measures for models with correlated input parameters. It also addresses important drawbacks of these measures and develops them further.

# 4 Qualitative GSA measures for functions with correlated input parameters

This chapter explains the computation of the EE-based screening measures that are applied to the model in Keane and Wolpin (1994). First, I outline two sampling schemes that are tailored to the computation of EEs. Second, I present a simplified version of the approach to extend EEs to models with correlated parameters by Ge and Menendez (2017). Third, I develop redesigned measures that sustain the EE's fundamental derivative characteristic. The section concludes with the analysis of a test function. This analysis shows, first, my ability to replicate the results in Ge and Menendez (2017), and second, validates my redesigned measures.

## 4.1 Sampling Schemes

The two presented sampling schemes are the trajectory and the radial design. Although the schemes are tailored to the computation of EEs, positional differences between them cause differences in their post-processing.

According to several experiments with common test functions by Campolongo et al. (2011), the best design is the radial design (Saltelli (2002)) and the most commonly used is the trajectory design (Morris (1991)). Both designs are comprised by a $(k+1) \times k$-dimensional matrix. The elements are generated in $[0, 1]$. Afterwards, they can potentially be transformed to the distributions of choice. The columns represent the different input parameters and each row is a complete input parameter vector. To compute the aggregate qualitative measures, a set of multiple matrices, or sample subsets, of input parameters has to be generated.

A matrix in radial design is generated the following way: Draw a vector of length $2k$ from a quasi-random sequence. The first row, or parameter vector, is the first half of the sequence. Then, copy the first row to the remaining $k$ rows. For each row $k'$ of the remaining 2, ..., $k+1$ rows, replace the $k'$-th element by the $k'$-th element of the second half of the vector. This generates a matrix of the following form:

$$\underset{(k+1)\times k}{\mathbf{R}} = \begin{pmatrix} a_1 & a_2 & ... & a_k \\ \mathbf{b_1} & a_2 & ... & a_k \\ a_1 & \mathbf{b_2} & ... & a_k \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & ... & \mathbf{b_k} \end{pmatrix} \tag{10}$$

Note here, that each column consists only of the respective first row element, except in one row. From this matrix, one EE can be obtained for each parameter $X_i$. This is achieved by using the $(i+1)$-th row as function argument for the minuend and the first row as the subtrahend in the EE formula in Equation (5). Then, $\Delta^{(i,j)} = b_i^{(j)} - a_i^{(j)}$. The asterisk is an index for all elements of a vector.

$$d_i = \frac{Y(\mathbf{a}_{\sim \mathbf{i}}, b_i) - Y(\mathbf{a})}{b_i - a_i} = \frac{Y(\mathbf{R_{i+1},_*}) - Y(\mathbf{R_{1},_*})}{b_i - a_i}. \tag{11}$$

If the number of radial subsamples is high, the draws from the quasi-random sequence lead to a fast coverage of the input space (compared to a random sequence). However, a considerable share of steps will be large – the maximum is $1 - \epsilon$ in a sample space of $[0, 1]$. This amplifies the aforementioned problem of EE-based measures with non-linear functions. The quasi-random sequence considered here is the Sobol' sequence. It is comparably successful in the dense coverage of the unit hypercube, but also conceptually more involved. Therefore, its presentation is beyond the scope of this work. As the first elements of each

Sobol' sequence, the direction numbers, are equal, I draw the sequence at once for all sets of radial matrices.

Next, I present the trajectory design. As we will see, it can lead to a relatively representative coverage for a very small number of subsamples but also to repetitions of similar draws. I skip the equations that generate a trajectory and instead present the method verbally. There are different forms of trajectories. I focus on the common version presented in Morris (1991) that generates equiprobable elements. The first step is to decide the number $p$ of equidistant grid points in interval $[0, 1]$. Then, the first row of the trajectory is composed of the lower half value of these grid points. Now, fix $\Delta = p/[2(p-1)]$. This function implies, that adding $\Delta$ to the lowest point in the lowest half results in the lowest point of the upper half of the grid points, and so on. It also implies that 0.5 is the lower bound of $\Delta$. The rest of the rows is constructed, first, by copying the row one above and, second, by adding $\Delta$ to the $i$-th element of the $i+1$-th row. The created matrix scheme is depicted below.

$$\underset{(k+1)\times k}{\mathbf{T}} = \begin{pmatrix} a_1 & a_2 & ... & a_k \\ \mathbf{b_1} & a_2 & ... & a_k \\ \mathbf{b_1} & \mathbf{b_2} & ... & a_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b_1} & \mathbf{b_2} & ... & \mathbf{b_k} \end{pmatrix} \tag{12}$$

In contrary to the radial scheme, each $b_i$ is copied to the subsequent row. Therefore, the EEs have to be determined by comparing each row with the row above instead of with the first row. Importantly, two random transformations are common. These are, first, randomly switching rows, and second, randomly interchanging the $i$-th column with the $(k-i)$-th column and then reversing the column. The first transformation is skipped as it does not add additional coverage and because we need the stairs-shaped design to facilitate later transformations which account for correlations between input parameters. The second transformation is adapted because it is important to also have negative steps and because it sustains the stairs shape. Yet, this implies that $\Delta$ is also parameter- and trajectory-specific. Let $f$ and $h$ be additional indices representing the input parameters. The derivative formula is adapted to the trajectory design as follows:[18]

$$d_i = \frac{Y(\mathbf{b_{f\leq i}}, \mathbf{a_{h>i}}) - Y(\mathbf{b_{f<i}}, \mathbf{a_{h\geq i}})}{b_i - a_i} = \frac{Y(\mathbf{T_{i+1,*}}) - Y(\mathbf{T_{i,*}})}{b_i - a_i}. \tag{13}$$

The trajectory design involves first, a fixed grid, and second and more importantly, a fixed

[18] In contrary to most authors, I also denote the step as a subtraction instead of $\Delta$ when referring to the trajectory design. This provides additional clarity.

step $\Delta$, s.t. $\{\Delta\} = \{\pm\Delta\}$. This implies less step variety and less space coverage vis-á-vis the radial design for a larger number of draws.

To improve the sample space coverage by the trajectory design, Campolongo et al. (2007) develop a post-selection approach based on distances. The approach creates enormous costs for more than a small number of trajectories. This problem is effectively mitigated by Ge and Menendez (2014). The following describes the main ideas of both contributions.

The objective of both works is to select $k$ trajectories from a set of $N$ matrices. Campolongo et al. (2007) assign a pair distance to each pair of trajectories in the start set. Thereafter, they identify each possible combination of $k$ from $N$ trajectories. Then, they compute an aggregate distance for each combination based on the single pair distances. Finally, the optimized trajectory set is the subset with the highest aggregate distance.

This is computationally very costly because each aggregate distance is a sum of a binomial number of pair distances[19]. To decrease the computation time, Ge and Menendez (2014) propose two improvements. First, in each iteration $i$, they select only $N(i) - 1$ matrices from a set containing $N(i)$ trajectories until the set size has decreased to $k$. Second, they compute the pair distances in each iteration based on the aggregate distances and the pair distances from the first set. Due to numerical imprecisions, their improvement does not always result in the same set as obtained from Campolongo et al. (2007). However, the sets are usually very similar in terms of the aggregate distance. This thesis only uses the first step in Ge and Menendez (2014) to post-select the trajectory set because the second step does not provide any gain.[20]

So far, we have only considered draws in [0,1]. For uncorrelated input parameters from arbitrary distributions with well-defined CDF, $\Phi$, one would simply evaluate each element (potentially including the addition of the step) by the inverse CDF, or quantile function, $\Phi^{-1}$, of the respective parameter. One intuition is, that $\Phi$ maps the sample space to [0,1]. Hence $\Phi^{-1}$ can be used to transform random draws in [0,1] to the sample space of the arbitrary distribution. This is a basic example of so-called inverse transform sampling (Devroye (1986)) which we will recall in the next section.

## 4.2 The approach for correlated input parameters in Ge and Menendez (2017)

This section describes the approach by Ge and Menendez (2017) to extend the EE-based measures to input parameters that are correlated. Their main achievement is to outline a transformation of samples in radial and trajectory design that incorporates the correlation between the input parameters. This implies, that the trajectory and radial samples cannot be written as in Equation (11) and Equation (13). The reason is that the correlations

[19] For example, $\binom{30}{15} = 155117520$.
[20] This refers only to my implementation.

of parameter $X_i$, to which step $\Delta^i$ is added, imply that all other parameters $\mathbf{X_{\sim i}}$ in the same row with non-zero correlation with $X_i$ are changed as well. Therefore, the rows cannot be denoted and compared as easily by $a$'s and $b$'s as in Equation (11) and (13). Transforming these matrices allows to re-define the EE-based measures accordingly, such that they sustain the main properties of the ordinary measures for uncorrelated parameters. The property is being a function of the mean derivative. Yet, Ge and Menendez (2017) do not fully develop these measures. I explain how their measures can lead to arbitrary rankings for correlated input parameters. This section covers their approach in a simplified form, focussing on normally distributed input parameters, and presents their measures.

The next paragraph deals with developing a recipe for transforming draws $\mathbf{u} = \{u_1, u_2, ..., u_k\}$ from $[0, 1]$ for an input parameter vector to draws $\mathbf{x} = \{x_1, x_2, ..., x_k\}$ from an arbitrary joint normal distribution. We will do this in three steps.

For this purpose, let $\mathbf{\Sigma}$ be a non-singular variance-covariance matrix and let $\boldsymbol{\mu}$ be the mean vector. The $k$-variate normal distribution is denoted by $\mathcal{N}_k(\boldsymbol{\mu}, \mathbf{\Sigma})$.

Creating potentially correlated draws $\mathbf{x}$ from $\mathcal{N}_k(\boldsymbol{\mu}, \mathbf{\Sigma})$ is simple. Following Gentle (2006)[21], this can be achieved the following way: Draw a $k$-dimensional row vector of i.i.d standard normal deviates from the univariate $N(0, 1)$ distribution, such that $\mathbf{z} = \{z_1, z_2, ..., z_k\}$, and compute the Cholesky decomposition of $\Sigma$, such that $\mathbf{\Sigma} = \mathbf{T^T T}$. The lower triangular matrix is denoted by $\mathbf{T^T}$. Then apply the operation in Equation (10) to obtain the correlated deviates from $\mathcal{N}_k(\boldsymbol{\mu}, \mathbf{\Sigma})$.

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{T^T z^T} \tag{14}$$

The next step is to understand that we can split the operation in Equation (10) into two subsequent operations. The separated first part allows to potentially map correlated standard normal deviates to other distributions than the normal one. For this, let $\boldsymbol{\sigma}$ be the vector of standard deviations and let $\mathbf{R_k}$ be the correlation matrix of $\mathbf{x}$.

The first operation is to transform the standard deviates $\mathbf{z}$ to correlated standard deviates $\mathbf{z_c}$ by using $\mathbf{z_c} = \mathbf{Q^T z^T}$. In this equation, $\mathbf{Q^T}$ is the lower matrix from the Cholesky decomposition $\mathbf{R_k} = \mathbf{Q^T Q}$. This is equivalent to the above approach in Gentle (2006) for the specific case of the multivariate standard normal distribution $\mathcal{N}_k(0, R_k)$. This is true because for multivariate standard normal deviates, the correlation matrix is equal to the covariance matrix.

The second operation is to scale the correlated standard normal deviates: $\mathbf{z} = \mathbf{z_c(i)}\boldsymbol{\sigma}\mathbf{(i)} + \boldsymbol{\mu}.$, where the $i$s indicate an element-wise multiplication. This equation is specific to normally distributed parameters.

The last step to construct the final approach is to recall the inverse transform sampling

[21] See p. 197

method. Therewith, we can transform the input parameter draws $\mathbf{u}$ to uncorrelated standard normal draws $\mathbf{z}$. Then we will continue with the two operations in the above paragraph. The transformation from $\mathbf{u}$ to $\mathbf{z}$ is denoted by $F^{-1}(\Phi^c)$, where the $c$ in $\Phi^c$ stands for the introduced correlations. This transformation is summarized by the following three steps:

$$\left.\begin{aligned} &\textbf{Step 1: } \mathbf{z} = \boldsymbol{\Phi^{-1}(u)} \\ &\textbf{Step 2: } \mathbf{z_c} = \mathbf{Q^T z^T} \\ &\textbf{Step 3: } \mathbf{x} = \boldsymbol{\mu} + \mathbf{z_c(i)}\boldsymbol{\sigma(i)} \end{aligned}\right\} F^{-1}(\Phi^c) \stackrel{\text{def}}{=} \mathcal{T}_2$$

To map $u$ to different sample spaces, Step 3 can be substituted. For instance, this could be achieved by applying $\Phi^u$ and the inverse CDF of the desired distribution to $z_c$.[22]

The one most important point to understand is that the transformation comprised by the three steps is not unique for correlated input parameters. Rather, the transformation changes with the order of parameters in vector $\mathbf{u}$[23]. This can be seen from the lower triangular matrix $\mathbf{Q^T}$. To prepare the next equation, let $\mathbf{R_k} = (\rho_{ij})_{ij=1}^{k}$ and sub-matrix $\mathbf{R_h} = (\rho_{ij})_{ij=1}^{h}$, $h \leq k$. Also let $\boldsymbol{\rho_i^{*,j}} = (\rho_{1,j}, \rho_{2,j}, ..., \rho_{i-1,j})$ for $j \geq i$ with the following abbreviation $\boldsymbol{\rho_i} \stackrel{\text{def}}{=} \boldsymbol{\rho_i^{*,i}}$. Following Madar (2015), the lower matrix can be written as

$$\mathbf{Q^T} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ \rho_{1,2} & \sqrt{1-\rho_{1,2}^2} & 0 & ... & 0 \\ \rho_{1,3} & \frac{\rho_{2,3}-\rho_{1,2}\rho_{1,3}}{\sqrt{1-\rho_{1,2}^2}} & \sqrt{1-\boldsymbol{\rho_3}\mathbf{R_2^{-1}}\boldsymbol{\rho_3^T}} & ... & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1,k} & \frac{\rho_{2,k}-\rho_{1,2}\rho_{1,k}}{\sqrt{1-\rho_{1,2}^2}} & \frac{\boldsymbol{\rho_{3,k}}-\boldsymbol{\rho_3^{*,k}}\mathbf{R_2^{-1}}\boldsymbol{\rho_3^T}}{\sqrt{1-\boldsymbol{\rho_3}\mathbf{R_2^{-1}}\boldsymbol{\rho_3^T}}} & ... & \sqrt{1-\boldsymbol{\rho_k}\mathbf{R_2^{-1}}\boldsymbol{\rho_k^T}} \end{pmatrix}. \tag{15}$$

Equation (15), together with Step 2, implies that the order of the uncorrelated standard normal deviates $\boldsymbol{z}$ constitutes a hierarchy amongst the correlated deviates $\mathbf{z_c}$ in the following manner: The first parameter is not subject to any correlations, the second parameter is subject to the correlation with the first parameter, the third parameter is subject to the correlations with the parameters before, etc. Therefore, if parameters are correlated, typically $\mathbf{Q^T z^T} \neq \mathbf{Q^T}(\mathbf{z'})^\mathbf{T}$ and $F^{-1}(\Phi)(\mathbf{u}) \neq F^{-1}(\Phi)(\mathbf{u'})$, where $\mathbf{z'}$ and $\mathbf{u'}$ denote $\mathbf{z}$ and $\mathbf{u}$ in different orders.

[22] The procedure described by the three steps above is equivalent to an inverse Rosenblatt transformation and a linear inverse Nataf transformation for parameters in normal sample space and connects to Gaussian copulas. For the first two transformations, see Lemaire (2013), p. 78 - 113. These concepts can be used to transform deviates in [0,1] to the sample space of arbitrary distributions by using the properties sketched above under different conditions.

[23] This point is more obvious in the formal definition of the Rosenblatt transformation.

Coming back to the EE-based measures for correlated inputs, Ge and Menendez (2017) attempt to construct two variations of $d_i$. They name one EE the independent Elementary Effect, $d_i^{ind}$, and the other one the full Elementary Effect, $d_i^{full}$. For each of these two EEs, they derive the aggreate measures $\mu$, $\mu^*$ and $\sigma$. The difference between the two EEs is that $d^{ind}$ excludes and $d^{full}$ includes the effect of the correlations from adding step $\Delta^i$ to $X_i$ on the other parameters $\mathbf{X_{\sim i}}$. Knowing $d_i^{ind}$ is important because the correlations can decrease $d_i^{full}$ (close to 0). However, if the independent-EE-based measures are not close to zero, $X_i$ is still important. Additionally, fixing this parameter can potentially change $d_i^{full}$ for $\mathbf{X_{\sim i}}$. Therefore, $d_i^{full}$ and $d_i^{ind}$ have to be interpreted together.

For the trajectory design, the formula for the full Elementary Effect, is given by

$$d_i^{full,T} = \frac{f\Big(\mathcal{T}(\mathbf{T_{i+1,*}}; i-1)\Big) - f\Big(\mathcal{T}(\mathbf{T_{i,*}}; i)\Big)}{\Delta}. \tag{16}$$

In Equation (12), $\mathcal{T}(\cdot; i) \stackrel{\text{def}}{=} \mathcal{T}_3\Big(\mathcal{T}_2\big(\mathcal{T}_1(\cdot; i)\big); i\Big)$. $\mathcal{T}_1(\cdot; i)$ orders the parameters, or row elements, to establish the right correlation hierarchy. $\mathcal{T}_2$, or $F^{-1}(\Phi^c)$, correlates the draws in $[0, 1]$ and transforms them to the sample space. $\mathcal{T}_3(\cdot; i)$ reverses the element order back to the start, to be able to apply the subtraction in the numerator of the EEs row-by-row. Index $i$ in $\mathcal{T}_1(\cdot; i)$ and $\mathcal{T}_3(\cdot; i)$ stands for the number of initial row elements that are cut and moved to the back of the row in the same order. Applying $\mathcal{T}(\mathbf{T_{i+1,*}}; i-1)$ and $\mathcal{T}(\mathbf{T_{i+1,*}}; i)$ to all rows $i$ of trajectory $\mathbf{T}$ as in Equation (12) gives the following two transformed trajectories:

$$\mathcal{T}_1(\mathbf{T_{i+1,*}}; i-1) = \begin{pmatrix} a_k & a_1 & ... & ... & a_{k-1} \\ \mathbf{b_1} & a_2 & ... & ... & a_k \\ \mathbf{b_2} & a_3 & ... & ... & \mathbf{b_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b_k} & \mathbf{b_1} & ... & ... & \mathbf{b_{k-1}} \end{pmatrix} \tag{17}$$

$$\mathcal{T}_1(\mathbf{T_{i,*}}; i-1) = \begin{pmatrix} a_1 & a_2 & ... & ... & a_k \\ a_2 & ... & ... & a_k & \mathbf{b_1} \\ a_3 & ... & ... & \mathbf{b_1} & \mathbf{b_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b_1} & \mathbf{b_2} & ... & ... & \mathbf{b_k} \end{pmatrix} \tag{18}$$

Two points can be seen from Equation (17) and (18). First, $\mathcal{T}_1(\mathbf{T_{i+1,*}}; i-1)$ and $\mathcal{T}_1(\mathbf{T_{i,*}}; i)$, i.e. the $(i+1)$-th row in Eq. (17) and the $(i)$-th row in Eq. (18), only differ in the $i$-th

element. The difference is $b_i - a_i$. Thus, these two rows can be used to compute the EEs like in the uncorrelated case in Equation (5). However, in this order, the parameters are in the wrong positions to be directly handed over to the function, as the $i$-th parameter is always in front. The second point is that in $\mathcal{T}_1(\mathbf{T_{i+1},*}; i-1)$, $b_i$ is in front of the $i$-th row. This order prepares the establishing of the right correlation hierarchy by $\mathcal{T}_2$, such that the $\Delta$ in $a_i + \Delta$ is included to transform all other elements representing $X_{\sim i}$. Importantly, to perform $\mathcal{T}_2$, mean vector $\mathbf{x}$ and covariance matrix $\mathbf{\Sigma}$ and its transformed representatives have always to be re-ordered according to each row. Then, $\mathcal{T}_3$ restores the original row order and $d_i^{full}$ can comfortably be computed by comparing function evaluations of row $i+1$ in $\mathcal{T}(\mathbf{T_{i+1},*}; i-1)$ with function evaluations of row $i$ in $\mathcal{T}(\mathbf{T_{i},*}; i-1)$. Now, the two transformed trajectories only differ in the $i$-th element in each row $i$.

The formula for the independent Elementary Effect for the trajectory samples is given by

$$d_i^{ind,T} = \frac{f\Big(\mathcal{T}(\mathbf{T_{i+1},*}; i)\Big) - f\Big(\mathcal{T}(\mathbf{T_{i},*}; i)\Big)}{\Delta}. \tag{19}$$

Note that $\mathcal{T}(\mathbf{T_{i+1},*}; i)$ equals $\mathcal{T}(\mathbf{T_{i},*}; i-1)$ which is the function argument in the subtrahend in Equation (16). Therefore this transformation can be skipped for the trajectory design and the transformed trajectory in Equation (19) can be recycled. Here, the $X_i$ that takes the step in the denominator of the Elementary Effect is moved to the back such that the step does not affect the other input parameters $\mathbf{X_{\sim i}}$ through the correlations. The left trajectory is constructed such that for each row $i$, $a_i$ does take the step that $b_i$ took in the aforementioned minuend trajectory. The argument of the subtrahend in the denominator in $d_i^{ind}$ is given by the rows in

$$\mathcal{T}_1(\mathbf{T_{i},*}; i-1) = \begin{pmatrix} a_2 & a_3 & ... & ... & a_1 \\ a_3 & ... & a_k & \mathbf{b_1} & a_2 \\ a_4 & ... & \mathbf{b_1} & \mathbf{b_2} & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b_2} & \mathbf{b_3} & ... & ... & \mathbf{b_1} \end{pmatrix} \tag{20}$$

The transformation for the samples in radial design work equally except that the minuend samples are composed of the first row copied to each row below because the steps in the radial design are always taken from the draws in the first row. To account for the correlation hierarchy and the transformation to the sample space performed by $\mathcal{T}_2$, it is important to also reorder the minuend trajectories. In fact, it is enough to compute one minuend trajectory if the order is regarded correctly. The formulae of the Elementary Effects for the radial design are given by Equation (17) and (18).

$$d_i^{full,R} = \frac{f\Big(\mathcal{T}(\mathbf{R_{i+1},*}; i-1)\Big) - f\Big(\mathcal{T}(\mathbf{R_{1},*}; i)\Big)}{b_i - a_i}. \tag{21}$$

$$d_i^{ind,R} = \frac{f\Big(\mathcal{T}(\mathbf{R_{i+1},*}; i)\Big) - f\Big(\mathcal{T}(\mathbf{R_{1},*}; i)\Big)}{b_i - a_i}. \tag{22}$$

The trajectory scheme requires $N(3k+1)$ and the radial scheme $N3k$ function evaluations to compute both EEs. The first row of the minuend and the last row of the subtrahend trajectories can be skipped. If one transformed trajectory contains information for two EEs, then all rows are used for the trajectory design. The radial scheme requires one row evaluation less because the subsamples that contain the repeated first rows include a pair of identical rows.

In the next section, I present the shortcomings of the EEs by Ge and Menendez (2017). Furthermore, I develop improved measures that finally correspond to the actual EE definition for uncorrelated parameters.

## 4.3 Drawbacks in Ge and Menendez (2017) and redesigned Elementary Effects.

For the following explanation, I refer to a normal sample space. The drawback in the EE definitions by Ge and Menendez (2017) is that $\Delta$ is transformed multiple times in the numerator but not in the denominator. Therefore, these measures are not Elementary Effects in the sense of a derivative. To understand the intuition, it is easier to view $\Delta$ as $b - a$. The transformation in the numerator is performed by applying $F^{-1}(\Phi^c)$ to $u_i^j = a_i^j + \Delta^{(i,j)}$. The implications of this approach is twofold. The first implication is that function $f$ is evaluated at arguments that are non-linearily transformed by Step 1 and Step 3. Then, the difference in the numerator is divided by the difference between the changed input parameter and the base input parameter – in unit space. Therefore, numerator and denominator refer to different sample spaces. This makes the results hard to interpret. It also increases the influence of more extreme draws in $[0, 1]$ because $\Phi^{-1}$ is very sensitive to these. Therefore, it will take a larger sample for the aggregate EE measures in Ge and Menendez (2017) to converge. Additionally, these problems are amplified if there are large differences between the inputs' standard deviation through the subsequent multiplicative scaling. The large sensitivity to more extreme draws implies also a higher sensitivity to larger differences in $\Delta = b - a$. Therefore, the results will differ in their level depending on the sampling scheme. The largest drawback, however, is that $b_i - a_i$ in the denominator of $d_i^{ind}$ does not account for the transformation of the $b_i - a_i$ in the nominator by the establishing of correlations in Step 2. This transformation decreases $b_i - a_i$ proportional

to the degree of correlations of the respective input parameter as can be seen by the last row of the lower Cholesky matrix in Eq. (15). Hence, $d_i^{ind}$ is inflated by the the input parameters' correlations even tough this measure is introduced as an independent effect. It actually is a dependent effect as well. Because the full EE has to be interpreted with regards to the independent EE, this problem spills over to $d_i^{full}$. For these reasons, I argue that these measures can not be used for screening. I redesign the measures in Ge and Menendez (2017) by scaling the $\Delta$ in the denominator according to the nominator. I refer to these measures as the correlated and the uncorrelated Elementary Effects, $d_i^c$ and $d_i^u$. They are given below for arbitrary input distributions and for samples in trajectory and radial design. For this purpose, let ${Q^T}_{k,*k-1}$ be the last row except of the last element of the lower triangular Cholesky matrix of the correlation matrix in the respective order of $\mathcal{T}_1(\mathbf{T_{i+1},}_*; i-1)$ and $\mathcal{T}_1(\mathbf{T_{i+1},}_*; i-1)$. Let ${Q^T}_{k,k}$ be the last element of the same matrix. Let $\Phi^u$ be the transformation from $[0,1]$ to standard normal space without regarding correlations unlike Step 2. Let $F^{-1}$ be the transformation that maps standard normal deviates to an arbitrary sample space like in Step 1. Index $*k-1$ represents all elements except of the last one of a vector of length $k$. Index $j$ is used to indicate an element-wise multiplication.

$$d_i^{c,T} = \frac{f\Big(\mathcal{T}(\mathbf{T_{i+1},}_*; i-1)\Big) - f\Big(\mathcal{T}(\mathbf{T_{i-1},}_*; i)\Big)}{F^{-1}\Big(\Phi^u(b_i)\Big) - F^{-1}\Big(\Phi^u(a_i)\Big)} \tag{23}$$

$$d_i^{u,T} = \frac{f\Big(\mathcal{T}(\mathbf{T_{i+1},}_*; i)\Big) - f\Big(\mathcal{T}(\mathbf{T_{i},}_*; i)\Big)}{F^{-1}\Big({Q^T}_{k,*k-1}(j){T^T}_{i+1,*k-1}(j) + {Q^T}_{k,k}\Phi^u(b_i)\Big) - F^{-1}\Big({Q^T}_{k,*k-1}(j){T^T}_{i,*k-1}(j) + {Q^T}_{k,k}\Phi^u(a_i)\Big)} \tag{24}$$

$$d_i^{c,R} = \frac{f\Big(\mathcal{T}(\mathbf{R_{i+1},}_*; i-1)\Big) - f\Big(\mathcal{T}(\mathbf{R_{1},}_*; i-1)\Big)}{F^{-1}\Big(\Phi^u(b_i)\Big) - F^{-1}\Big(\Phi^u(a_i)\Big)} \tag{25}$$

$$d_i^{u,R} = \frac{f\Big(\mathcal{T}(\mathbf{R_{i+1},}_*; i)\Big) - f\Big(\mathcal{T}(\mathbf{R_{1},}_*; i)\Big)}{F^{-1}\Big({Q^T}_{k,*k-1}(j){R^T}_{i+1,*k-1}(j) + {Q^T}_{k,k}\Phi^u(b_i)\Big) - F^{-1}\Big({Q^T}_{k,*k-1}(j){R^T}_{i,*k-1}(j) + {Q^T}_{k,k}\Phi^u(a_i)\Big)}. \tag{26}$$

In Equation (24) and (26), ${Q^T}_{k,*k-1}(j){R^T}_{i,*k-1}(j)$ refers to the transformed trajectory of the copied first rows, $\mathcal{T}(\mathbf{R_{i+1},}_*; i)$, as previously described.

The denominator of the correlated Elementary Effect, $d_i^c$, consists of $\Phi$ and $F^{-1}$. These functions account for the transformation from uniform to arbitrary sample space. In $d_i^c$, the denominator must not account for a correlating transformation of $a_i$ and $b_i$ because

${Q^T}_{1,*} = (1, 0, 0, ..., 0)$. Therefore, $b_i$ and $a_i$ are multiplied by one in the correlation step, Step 2, of the standard normal deviates.

The last point is not true for the uncorrelated Elementary Effect $d_i^{c,*}$. One accounts for the correlation step by multiplying $\Phi^u(b_i)$ by ${Q^T}_{k,*k-1}(j){T^T}_{i+1,*k-1}(j)/\Phi^u(b_i) + {Q^T}_{k,k}$. The respective operation is also done for $\Phi^u(a_i)$. One can also write the denominator equivalently but less explicitly for all four EEs as $\mathcal{T}(b_i) - \mathcal{T}(a_i)$[24]. Not accounting for $Q_t$ like Ge and Menendez (2017) leads to arbitrarily decreased independent Elementary Effects for input parameters with higher correlations. As $d_i^{full}$ and $d_i^{ind}$ are interpreted jointly, both measures are in fact useless.

For $X_1, ..., X_k \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the denominator of $d_i^{u,*}$ simplifies drastically to

$$\begin{aligned} &\Big(\mu_i + \sigma_i\Big({Q^T}_{k,*k-1}(j){T^T}_{i+1,*k-1}(j) + {Q^T}_{k,k}\Phi^u(b_i)\Big) \\ -&\Big(\mu_i + \sigma_i\Big({Q^T}_{k,*k-1}(j){T^T}_{i+1,*k-1}(j) + {Q^T}_{k,k}\Phi^u(a_i)\Big) \\ &= \sigma_i {Q^T}_{k,k}\Big(\Phi^u(b_i) - \Phi^u(a_i)\Big). \end{aligned} \tag{27}$$

The next section presents results for a test function to demonstrate a replication of Ge and Menendez (2017) and to validate this thesis' measures.

## 4.4 Replication and Validation

Let $g(X_1, ..., X_k) = \sum_{i=1}^k c_i X_i$ be an arbitrary linear function. Let $\rho_{i,j}$ be the linear correlation between $X_i$ and $X_j$. Then, for all $i \in 1, ..., k$, I expect:[25]

$$d_i^{u,*} = c_i \tag{28}$$

$$d_i^{c,*} = \sum_{j=1}^{k} \rho_{i,j} c_j .. \tag{29}$$

Both equations state that, conceptually, the result does not depend on the sampling scheme.

Equation (28) shows that the uncorrelated EE of each input equals the derivative of $f(\mathbf{X})$ with respect to $X_i$. This corresponds to the concept of the original EE as a derivative. As the function is linear, the derivative is equal for all values in the sample space of each $X_i$. This implies $\mu = \mu^* = d_i$ and $\sigma = 0$ for both sampling schemes and both EE versions.

Equation (29) states that $d_i^{c,*}$ also includes the effects of the other input parameters that are correlated with $X_i$. Then, the function's linearity implies that these effects are included by adding the respective coefficients multiplied by the correlation with $X_i$ to the uncorrelated effect $d_i^{u,*}$.

[24] From an implementation perspective, this is also the simplest and most general way. It only requires to transfer the right elements from the final, transformed samples to the denominator.

[25] These results correspond to the intuition provided by the example in Saltelli et al. (2008), p. 123.

Table 3 and Table 4 show results for a simple linear test function with 3 input parameters, where $X_i \sim \mathcal{N}(0,1)$ for $i \in \{1,2,3\}$, and with correlations $\rho_{1,2} = 0.9$, $\rho_{1,3} = 0.4$, $\rho_{2,3} = 0.01$. The first columns in both tables show the aggregate EE-based measures by Ge and Menendez (2017). The second (and third) columns show my replications for these measures. Table 3 consists of two replications because the trajectory design contains two numerical decisions. The details are explained in the following. The last columns in both tables depict the results for the redesigned measures. Ge and Menendez generate 240 trajectories and 260 radial samples to compute their results. My figures base on 10,000 samples for each scheme.

From the first columns, we can observe three things about the measures in Ge and Menendez (2017).

First the measures are not related to derivatives, as the non-linear transformation $F(\Phi^c(\cdot))$ is only applied to the EE numerator. This transformation leads to two additional observations.

The second result is that the measures differ between trajectory and radial design. This is because the non-linearity introduces variation to these measure according to the values of the draws in $[0,1]$. This variation is due to two reasons. The first reason are the general differences between both sampling schemes.

The last observation results from the fact that the trajectory design, in contrary to the radial design, includes two numeric parameters. These are, first, the approximation of 0 and 1 to prevent the inverse CDF from returning infinite values. It is enough to approximate 0 by $0^{num}$. The lower $0^{num}$, the more extreme are the transformed draws for 0 and 1 in normal space. Therefore, lower $0^{num}$ also increase the standard deviations. The second approximation is the number of levels, or grid points, $l$ in $[0,1]$. The more grid points, the smaller is the share of draws from set $0,1$ and the less imprecise are the results. The authors indicate that they use $0^{num} = 0.00001$ and $l$ for the trajectory design.[26]

My replications for the trajectory design are depicted in the second column of Table 3. The mean and mean absolute EEs, $\mu^{*,ind}$ and $\mu^{*,full}$ come reasonably close. However, the standard deviations, $\sigma^{*,ind}$ and $\sigma^{*,full}$ equal 0. The reason is that four grid points re-linearise the transformed numerator in the EEs. This argument is illustrated by Figure 4.

The grid points with four levels are $\{0^{num}, 1/3, 2/3, 1-0^{num}\}$ and $\Delta = 2/3$. The quantile function is non-linear but point-symmetric to $(0.5, 0)$. Thus, $\Phi(\frac{1/3-(1-0^{num})}{2/3}) = \Phi(\frac{0^{num}-2/3}{2/3})$. Hence, for each single $X_i$ the Elementary Effects are constant and the transformed $\Delta$-specific derivation is equal at each of the two base grid points.

Therefore, I use $0^{num} = 0.00000001$ and $l = 24$ to increase the variation amongst the EEs by Ge and Menendez (2017) in order to achieve similar results.

[26] I obtained this information from an unofficial email from Qiao Ge.

The replication in Table 4 for the radial design is relatively close for the mean Effects. However, there are considerable differences between the standard deviations. In my view, this results, first, from different setups for the generation of the Sobol' sequence, and second, from a too low number of samples in Ge and Menendez (2017). This indicates that the appropriate number of samples for correlated parameters has to be higher than what is common for functions with uncorrelated input parameters.

The last column in both Table 3 and Table 4 depicts the aggregate measures based on the improved EEs. These are equal to what is anticipated at the beginning of this section. Therefore, the correlated and uncorrealted Elementary Effects developed in this thesis can be regarded as a valid extension of Elementary Effects for models with correlated input parameters.

To prepare the analysis in the results chapter, I describe the link from EE-measures to Sobol' indices by means of another linear test function $h$. However, let us consider the case without any correlations between the inputs. Additionally, let $c_i = \{3, 2, 1\}$ and $\sigma^2_{X_i} = \{1, 4, 9\}$ for $i \in \{1, 2, 3\}$. The following results are derived from Saltelli et al. (2008)[27]. Let us first compute the Sobol' indices. As $g$ does not include any interactions, $S_i^T = S_i$. Additionally, we have $\text{Var}(Y) = \sum_{i=1}^{k} c_i^2 \sigma^2_{X_i}$ and $\text{Var}_{\mathbf{X}_{\sim \mathbf{i}}}\Big(\mathbb{E}_{X_{\sim i}}[Y|\mathbf{X}_{\sim \mathbf{i}}]\Big) = c_i^2 \sigma^2_{X_i}$. Table 5 compares three different sensitvity measures. These are the total Sobol' indices, $S_i^T$, the mean absolute EE, $\gamma_i^*$, and the *squared* sigma-normalized mean absolute EE, $(\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y})^2$.

In context of screening, $S_i^T$ is the objective measure that we would like to predict approximately. We observe that $\gamma_i^*$ ranks the parameters incorrectly. The reason, is that $\gamma_i^*$ is only a measure of the influence of $X_i$ on $Y$ and not of the influence of the variation and the level of $X_i$ on the variation of $Y$. We also see that $(\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y})^2$ is an exact predictor for $S_i^T$ as it does not only generate the correct ranking but also the right effect size. Importantly, this result is specific to a linear function without any interactions and correlations. However, it underlines the point that $\gamma_i^*$ alone is not sufficient for screening. Following Ge and Menendez (2017), one approach would be to additionally consider the EE variation, $\sigma_i$. However, analysing two measures at once is difficult for models with a large number of input parameters. Table 5 indicates that $(\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y})^2$ and also $\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y}$ can be an appropriate alternative. The actual derivative version of this measure is also recommended by guidelines of the Intergovernmental Panel for Climate Change (IPCC (1999, 2000)).

The next chapter presents the occupational choice model in Keane and Wolpin (1994).

[27] See p.15-25.

# 5 The Occupational Choice Model

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. 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 substantially 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.

## 5.1 Keane and Wolpin (1994)

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

1. Unobservable shocks $\varepsilon_t$ do not have to be additively separable from the remainder of the utility functions.

2. Shocks $\varepsilon_t$ can be correlated across choices $a_t$.

3. Observable payoffs, or wages, $W_{a,t}^{-}$ are not conditionally independent from the unobservable shocks $\varepsilon_t$ given the observable choices $a_t$ and the observable part of the state vector $\mathbf{s_t^-}$. The reason is that wage shocks enter the wage function directly. If the agent decides against choices with observable payoffs that depend on $\varepsilon_t$, these payoffs can not be observed. Therefore, positive shocks and the observation of payoffs for the same choice are positively correlated.

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

In each period, a representative agent receives utility $U$. This utility depends on the state space and on choice $a$ in period $t \in \{0, 1, 2, ..., T\}$. 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.

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 (30). Like utility $U$, I also write value function $V$ with different emphasis on occupation choice $a_t$.

$$V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) = \max_{\{a\}_{t=0}^{T}} \left\{ \sum_{t=0}^{T} \delta^t \int_{\boldsymbol{\varepsilon_t}} U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t) f(\boldsymbol{\varepsilon_t}) d^{|A|}\boldsymbol{\varepsilon_t} \right\} \tag{30}$$

Value $V$ depends directly on time $t$ because $T$ is finite. Together with the discount factor $\delta$, this typically induces life-cycle behaviour. For example, agents invest more in the earlier time periods and work (and consume) more in the following periods. As $\{\varepsilon_{a,t}\}_{a\in A}$ are the only random parameters and serially independent, the expectation of $U(\boldsymbol{s_t^-}, \boldsymbol{\varepsilon_t}, a_t)$ 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))[28] that breaks up the problem in (30) into more tractable sub-problems along the time dimension.

$$V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}) = \max_{a_t} \left\{ U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t) + \delta \int_{\boldsymbol{\varepsilon_t}} \max_{a_{t+1}} V(\mathbf{s_{t+1}^-}, \boldsymbol{\varepsilon_{t+1}}, a_{t+1}) f(\boldsymbol{\varepsilon_{t+1}}) d^{|A|}\boldsymbol{\varepsilon_{t+1}} \right\} \tag{31}$$

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 $a_{t+1}$)

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

are maximized. The finite time horizon eases the problem as the value function for the last period $T$ simplifies to $V(\mathbf{s_T^-}, \boldsymbol{\varepsilon_T}) = \max_{a_T} U(\mathbf{s_T^-}, \boldsymbol{\varepsilon_T}, a_T)$.[29] 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 (31) 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 (Mincer (1958)):

$$\begin{aligned} U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, b) = W_{b,t}^- &= \exp\Big\{\beta^b + \beta_e^b x_{e,t} + \beta_b^b x_{b,t} + \beta_{bb}^b x_{b,t}^2 + \beta_w^b x_{w,t} + \beta_{ww}^b x_{w,t}^2 + \varepsilon_{b,t}\Big\} \\ U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, w) = W_{w,t}^- &= \exp\Big\{\beta^w + \beta_e^w x_{e,t} + \beta_w^w x_{w,t} + \beta_{ww}^w x_{w,t}^2 + \beta_b^w x_{b,t} + \beta_{bb}^w x_{b,t}^2 + \varepsilon_{w,t}\Big\} \end{aligned} \tag{32}$$

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.[30] These coefficients are called covariates by many structural economists.

The utilities for education, or schooling, and staying at home are given by the functions

[29] 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.

[30] 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.

in (33). These functions are also called non-pecuniary rewards.

$$\begin{aligned} U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, e) &= \beta^e + \beta^e_{he}\mathbf{1}(x_{e,t} \geq 12) + \beta^e_{re}(1 - \mathbf{1}(a_{t-1} = e)) + \varepsilon_{e,t} \\ U(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, h) &= \beta^h + \varepsilon_{h,t} \end{aligned} \tag{33}$$

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

Write $\{\varepsilon_{a,t}\}_{a \in A}$ as vector $\boldsymbol{\varepsilon_{a,t}}$. It is assumed that $\boldsymbol{\varepsilon_{a,t}}$ follows a joint normal distribution, such that $\boldsymbol{\varepsilon_{a,t}} \sim \mathcal{N}(0, \boldsymbol{\Sigma_\varepsilon})$. $\boldsymbol{\Sigma_\varepsilon}$ denotes the covariance matrix for shocks $\boldsymbol{\varepsilon_{a,t}}$. $\sigma^2_a$ and $\sigma^2_{a(j),a(k\neq j)}$ 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.

The model parameters are estimated with the simulated maximum likelihood method. The method is outlined in Appendix C. The next subsection presents the results.

## 5.2 Estimation Results

This subsection presents estimates $\hat{\boldsymbol{\theta}}$ for the exogenous parameters and the standard errors 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.[31] 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 $\boldsymbol{\varepsilon_{a,t}}$ and the non-diagonal elements $c_{i,j}$ equal the correlations between different alternative-specific shocks $\boldsymbol{\varepsilon_{a,t}}$. 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 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 $\beta$ and correlations between standard errors in this thesis, the estimates are reasonably similar. However, the one exception that

[31] 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 (42).

stands out is the results for the non-diagonal Choleksy factors $c_{i,j}$.

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 $\text{corr}(\hat{\delta}, \hat{\beta}^e)$, $\text{corr}(\hat{\delta}, \hat{\beta}^h)$, $\text{corr}(\hat{\beta}^e, \hat{\beta}^w)$, $\text{corr}(\hat{c_3}, \hat{\beta}^e)$ and $\text{corr}(\hat{c_4}, \hat{\beta}^h)$ with $-0.83$, $-0.31$, $0.45$, $-0.34$ and $-0.82$, respectively.

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 $\delta$ rises because all individuals get more patient, and therefore, ceteris paribus, they invest more in education. However, the same likelihood also increases if the educational utility constant $\beta^e$ rises. Hence, because they can compensate each other, the likelihood around the optimal parameter $\hat{\boldsymbol{\theta}}$ decreases less for changes of both parameters in opposing directions than for changes in the same direction. Therefore, parameters $\delta$ and $\beta^e$ are negatively correlated in terms of the score function in (43) around $\hat{\boldsymbol{\theta}}$. It follows from (45) that their standard errors are negatively correlated.

## 5.3 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.[32] 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.

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

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.[33] 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 uncertainty quantification.

# 6 Results

The section is divided in two parts. First, I present the results of the uncertainty analysis for an overview of the general variation in QoI $Y$. Thereafter, I present the results of the qualitative GSA. The aim is to draw inferences about the contribution of an individual input $X_i$ and its uncertainty to the uncertainty in QoI $Y$ to a degree that allows the identification of non-influential parameters.

## 6.1 Uncertainty Analysis

The following results are obtained by evaluating each parameter vector from a random sample by the occupational choice model. The sample is drawn according to the estimates for the joint distribution of the input parameters. The number of draws is $10,000$.[34]

Figure 5 incorporates the input uncertainty into the shares of life-time occupation decisions in Figure 3. It depicts the sample mean and the intervals for 99% of the shares' probability distribution. We can see that the input uncertainty has an effect on the shares of white- and blue-collar occupation but almost none on the shares of occupation in the education and home sector. This suggests that, given the input distribution, individuals mainly tend to change their decisions between occupation in blue- and white-collar occupation.

[33] 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.

[34] A reasonable level of convergence is already achieved after approximately 800 evaluations. See the respective notebook in the *Master's Thesis Replication Repository*.

However, the uncertainty in the shares for both labour sectors is also not strikingly large. There is also no visible difference in the uncertainties between both scenarios.

Figure 6 depicts the probability distribution of QoI $Y$. The colorised bars show the realisations within one and two and outside of two standard deviations, $\sigma_Y$, from sample mean $\overline{Y}$. The distribution is almost normal but minimally skewed to the left. This leads to the first conclusion for a potential quantitative GSA. That is, Sobol' indices are a good choice of a quantitative GSA measure because the variance provides a good summary for the variation of normally distributed variables.

We have standard deviation $\sigma_Y$ equals 0.1 and variance $\sigma_Y^2$ equals 0.01. The final goal of a quantitative GSA is to compute the share that input parameter $X_i$ and its variation contribute to $\sigma_Y$ or $\sigma_Y^2$. We can expect that reasonable measures for the contribution of $X_i$ are not completely detached from the measures for the total variation in $Y$ if they are on the same scale. As previously seen, for a linear function without any interactions and correlations, we would expect that the contribution of $X_i$ is even smaller than the measure for the variation in $Y$.

The next section computes and analyses multiple measures for these contributions.

## 6.2 Qualitative Global Sensitivity Analysis

The analysis is divided in three parts. The first part computes the measures by Ge and Menendez (2017) to validate the conceptual analysis and the results derived from the linear test function. The second part computes measures based on the redesigned EEs. They are used for a comparison with the first part and also to compare the two sampling schemes. The third part presents results for a measure that combines the influence of $X_i$ on the level and on the variation of QoI $Y$.

Due to time restrictions, I present results based on 100 samples in trajectory and radial designs. This is equal to 8200 and 8100 model evaluations. I recommend a higher number for future research.

Table 6 presents the aggregate measures for the full and independent EEs as developed in Ge and Menendez (2014). I choose the trajectory design because it does not produce extreme steps $b - a$. Additionally, it allows for more control over the set of random draws in the unit space. Thus, one can adjust it to number of draws. The numerical parameters are $0^{num} = 0.005$ and $l = 100$. The 100 trajectories are selected based on 200 trajectories in standard normal space following the first improvement in Ge and Menendez (2014). I do not apply the post-selection in unit space to prevent a concentration of draws in the centre of the normal space. I use this procedure for all trajectory samples throughout the

section.

The second and third columns depict the mean absolute EEs, $\mu_T^{*,full}$ and $\mu_T^{*,ind}$, and the fourth and fifth columns show the standard deviations of the EEs, $\sigma_T^{ind}$ and $\sigma_T^{full}$. The authors state that mean EEs and standard deviations have to be interpreted jointly for inferences about the variance of $Y$. They also state that, if for one parameter $X_i$, the full measures are close to zero, the independent measures has also to be low to confirm that $X_i$ can be fixed.

We make four main observations. First, the level of mean effects and standard deviations is very similar for each variable. This suggests that there are no parameters with a low effect on the level of $Y$ but a high effect on its variation. Second, the highest values for $\mu^{*,full}$ and $\sigma_T^{*,full}$ are achieved by the coefficients for the quadratic (cross-sector) experiences for blue- and white-collar sector, $\beta_{bb}^b$, $\beta_{ww}^b$, $\beta_{ww}^w$ and $\beta_{bb}^w$. Additionally, the effects for education and home sector are below 1. This makes sense in so far as Figure 5 already indicated that home and education sector are less important for the variation in occupation choices. The third observation is that $\mu^{*,full}$ and $\sigma_T^{*,full}$ contain large variations among the different parameters and that some values are very high. These values seem to be detached from the level of the variation in $Y$. The fourth observation is that the results for the independent measures are all smaller than 0.005. Following Ge and Menendez (2014), this would mean, that each parameter with full measures close to zero, i.e. all parameters for home and the education sector, can be fixed because the independent measures are close to zero anyway.

The third and fourth observations are precisely what we would expect from the previous analyses of the measures in Ge and Menendez (2017). Let us first look at the fourth observation. The independent EEs are strongly deflated for parameters with large correlations because the denominator does not account for the second transformation step that involved the lower Cholesky matrix, $Q^T$. As the correlations between all parameters are relatively high, all parameters, even those for blue- and white-collar sector, are close to 0. The third observation, extreme values for the full measures, can be explained by the transformations in the EE's numerator that are unanswered in the EE's denominator, namely the transformation from uniform to normal space in Step 1 and Step 3. This non-linear transformation of the numerator implies that comparably high values are even higher. Both explanations are confirmed in the next part of the analysis.

Table 7 depicts the redesigned measures, $\mu^{*,c}$ and $\mu^{*,u}$. These correspond to the mean absolute relative EEs, $\mu^{*,full}$ and $\mu^{*,ind}$. The measures are computed for both sampling schemes. I do not show the standard deviations. However, as in Table 6, they are also relatively close to the mean Elementary Effects.

The analysis of Table 7 is divided in three parts. First, I compare the results from this to the findings in Table 6. Then, I analyse the results with regards to the different sampling schemes. Thereafter, I discuss three more general findings.

One observation from Table 6 is that the values for $\mu_T^{*,full}$ and $\sigma_T^{*,full}$ have a large variation. Comparing column $\mu_T^{*,full}$ in Table 6 with column $\mu_T^{*,c}$ in Table 7, we find that the values for $\mu_T^{*,c}$ are indeed more compressed while the ranking remains the same. This again confirms the first drawback in Ge and Menendez (2017). Still, the variation in $\mu_T^{*,c}$ is relatively large. One possible explanation is that the model itself is highly non-linear.

Another observation from Table 6 is that $\mu_T^{*,ind}$ and $\sigma_T^{*,ind}$ are both close to zero. Looking at $\mu_T^{*,c}$, we find that this drawback is also absent. This finally allows a joint interpretation of the correlated and uncorrelated measures for factor fixing.

Comparing $\mu_T^{*,c}$ and $\mu_T^{*,u}$ with $\mu_R^{*,c}$ and $\mu_R^{*,u}$, we find that, in general, the measures for the trajectory design are much smaller than the ones for the radial design. The reason is that in the radial samples, $b - a$ can be every element in $(0, 1)$. In the trajectory samples, the step is always 0.55. The two statements refer exclusively to draws in unit space. Assuming the model is non-linear, this difference can lead to more extreme values for the radial design.[35]

A striking finding is that the uncorrelated measures are considerably larger than the correlated measures. This observation leads to the first of two insights about the interaction between estimation method and sensitivity analysis. I suggest the following explanation: In the model section, we saw that negative correlations imply similar effects, and positive correlations imply opposing effects on the likelihood of observed endogenous variables $\boldsymbol{\mathcal{D}}$. This has the following implication for the correlated EE $d_i^c$: If we change parameter $X_i$, then the other parameters $X_{\sim i}$ will change as well with respect to their correlations to $X_i$. If the correlations are such that negative correlations imply similar effects, and positive correlations imply opposing effects, the influence of the change in $X_i$ on observables $\boldsymbol{\mathcal{D}}$ is mitigated by the change in $X_{\sim i}$. Therefore, given that the QoI is closely linked to observables $\boldsymbol{\mathcal{D}}$, $d_i^c$ will be smaller than the uncorrelated EE $d_i^u$. This carries over to the respective aggregate measures.

The last two points about Table 7 are that the measures remain detached from the level of $\sigma_Y$ as there are also parameters with very high values. Additionally, all parameters for education and home sector remain very low compared to the other parameters. However, a different view can be that they are very close to $\sigma_Y$. As depicted in Table 2, these are also

[35] In principle, one could adjust the numerical parameters for the trajectory design such that they come closer to the radial design. This can be achieved by, first, decreasing $l$ such that the share of zeros and ones increases. Then, we decrease $0_n um$, such that the zeros and ones are mapped to much more extreme values in the standard normal space. However, smaller steps and a good coverage of the sample are desired.

the parameters with the highest standard deviations. Because $\sigma_i$ does not only depend on $\sigma_{X_i}$ but also on the level of $X_i$, it might be a difficult measure for the influence of $X_i$ on the variation of $Y$. From the linear function example, we learn how misleading $\mu_T^{*,c}$ can be. Therefore, I present results for another measure.

Figure 7 and Figure 8 depict the sigma-normalized mean absolute EEs for the radial and the trajectory scheme. We observe four main results:

First, the sigma-normalized measures are lower and much more compressed. Therefore, they appear less detached from $\sigma_Y$.

Second, the parameter order for each measure differs between the two sampling schemes. This is consequence of the first observation: Because the parameters are very close to each other, differences between the schemes are more important. It would be interesting to see how much these differences can be decreased by a higher number of draws. However, the fact remains that the uncorrelated measures are much larger than the uncorrelated measures.

The third result is that the parameter order differs largely from the order based on previous measures. For example, parameters that are irrelevant in terms of the non-normalised measures are at the top of the rankings for $\mu_\sigma^{*,u}$ and $\mu_\sigma^{*,c}$. According to $\mu_\sigma^{*,u}$, the most important parameters are not the coefficients for the quadratic (cross) experiences in both labour sectors but the constants $\beta^b$, $\beta^w$ and also the constant for the education sector $\beta^w$. According to $\mu_\sigma^{*,c}$, the most important factors are the Cholesky parameters $c_{1,3}$, $c_{2,3}$ and $c_3$. Trying to provide much intuition for the results of particular parameters because the level of complexity is very high. The results do not only include the effects on the level of $Y$ according to the economic model but also the effects of standard deviations and correlations between twenty-seven input parameters. However, it is interesting that the Cholesky factors, which belong to the set of parameters with the highest standard deviations, are now at the top of ranking of the measures that include correlations. This makes sense, in so far, as they have high covariances. In my view, the result that the uncorrelated measures for the constant terms $\beta^b$, $\beta^w$ and $\beta^w$ are more important, is not unreasonable. Economically, they anchor the rewards for working in the respective sectors. Also, $\beta^b$, $\beta^w$ have by far the highest absolute mean amongst all coefficients for both sectors. Moreover, their standard deviations are also much higher than those for the coefficients of linear and quadratic returns to experience. It seems less reasonable to me that the coefficients for the quadratic (cross) experiences are the most important parameters given that their mean is close to zero and that their standard deviation is much lower than those of the coefficients of the constant terms.

Assuming that the sigma-normalised measures are indeed a sensible global sensitivity measure, a second inference about the link between estimation method and sensitivity analysis can be made. That is, the larger the effect of parameter $X_i$ on observables $\boldsymbol{\mathcal{D}}$,

the smaller will be the respective standard error. The reason is that larger changes in $X_i$ decrease the probability of observing $\boldsymbol{\mathcal{D}}$. The smaller the effect of parameter $X_j$, the larger will be its standard error because the probability of observing $\boldsymbol{\mathcal{D}}$ is less affected. This mechanism mitigates the differences between $X_i$ and $X_j$ with respect to their effects on the variation of $Y$. However, it does not mitigate the differences between $X_i$ and $X_j$ in the effects on the level of $Y$.

Finally, I do not provide a recommendation about which parameters to fix. To my mind, finding convincing measures and criteria that link low-cost sensitivity measures to, for example, Sobol' indices is still an open research question for non-linear models with correlated input parameters and interactions. Nonetheless, I provide additional measures and practical insights as a contribution to potential conceptual research. This might prepare the ground for serious screening applications in the future.

# 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.

Campolongo, F., J. Cariboni, and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models. *Environmental modelling & software 22*(10), 1509–1518.

Campolongo, F., A. Saltelli, and J. Cariboni (2011). From screening to quantitative sensitivity analysis. a unified approach. *Computer Physics Communications 182*(4), 978–988.

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.

Devroye, L. (1986). Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pp. 260–265.

Ge, Q. and M. Menendez (2014). An efficient sensitivity analysis approach for computationally expensive microscopic traffic simulation models. *International Journal of Transportation 2*(2), 49–64.

Ge, Q. and M. Menendez (2017). Extending morris method for qualitative global sensitivity analysis of models with dependent inputs. *Reliability Engineering & System Safety 100*(162), 28–39.

Gentle, J. E. (2006). *Random number generation and Monte Carlo methods*. Springer Science & Business Media.

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.

IPCC (1999). Ipcc expert meetings on good practice guidance and uncertainty management in national greenhouse gas inventories. Background papers.

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.

Kucherenko, S. et al. (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation 79*(10), 3009–3017.

Kucherenko, S., S. Tarantola, and P. Annoni (2012). Estimation of global sensitivity indices for models with dependent variables. *Computer physics communications 183*(4), 937–946.

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

Lemaire, M. (2013). *Structural reliability*. John Wiley & Sons.

Madar, V. (2015). Direct formulation to cholesky decomposition of a general nonsingular correlation matrix. *Statistics & probability letters 103*, 142–147.

Mara, T. A., S. Tarantola, and P. Annoni (2015). Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environmental modelling & software 72*, 173–183.

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

McBride, K. and K. Sundmacher (2019). Overview of surrogate modeling in chemical process engineering. *Chemie Ingenieur Technik 91*(3), 228–239.

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.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics 33*(2), 161–174.

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 (2013a). Global sensitivity measures from given data. *European Journal of Operational Research 226*(3), 536–550.

Plischke, E., E. Borgonovo, and C. L. Smith (2013b). 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.

Rabitz, H. (1989). Systems analysis at the molecular scale. *Science 246*(4927), 221–226.

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. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer physics communications 145*(2), 280–297.

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.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons.

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.

Smith, R. C. (2014). *Uncertainty Quantification: Theory, Implementation, and Applications*. Philadelphia: SIAM-Society for Industrial and Applied Mathematics.

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.

Xiu, D. (2010). *Numerical methods for stochastic computations: a spectral method approach*. Princeton university press.

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.

## 8.1 Appendix A: Tables

**Table 1.** Overview of UQ literature

| Content | Number of articles |
|---|---|
| *Topics* | |
| Climate economics | 8 |
| Macroeconomics | 4 |
| *Analyses* | |
| Uncertainty analysis | 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 |

**Table 2.** Estimates for the distribution of input parameters

| Parameter | Mean | Standard error (SE) | SE in KW94 |
|---|---|---|---|
| *General* | | | |
| $\delta$ | 0.95 | 0.000 84 | - |
| *Blue-collar* | | | |
| $\beta^b$ | 9.21 | 0.013 | 0.014 |
| $\beta^b_e$ | 0.038 | 0.0011 | 0.0015 |
| $\beta^b_b$ | 0.033 | 0.000 44 | 0.000 79 |
| $\beta^b_{bb}$ | −0.0005 | 0.000 013 | 0.000 019 |
| $\beta^b_w$ | 0.0 | 0.000 67 | 0.0024 |
| $\beta^b_{ww}$ | 0.0 | 0.000 029 | 0.000 096 |
| *White-collar* | | | |
| $\beta^w$ | 8.48 | 0.0076 | 0.0123 |
| $\beta^w_e$ | 0.07 | 0.000 47 | 0.000 96 |
| $\beta^w_w$ | 0.067 | 0.000 55 | 0.000 90 |
| $\beta^w_{ww}$ | −0.001 | 0.000 017 | 0.000 070 |
| $\beta^w_b$ | 0.022 | 0.000 33 | 0.0010 |
| $\beta^w_{bb}$ | −0.0005 | 0.000 021 | 0.000 030 |
| *Education* | | | |
| $\beta^e$ | 0.0 | 330 | 459 |
| $\beta^e_{he}$ | 0.0 | 155 | 410 |
| $\beta^e_{re}$ | −4000 | 202 | 660 |
| *Home* | | | |
| $\beta^h$ | 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 | 176 | 786 |
| $c_{1,2}$ | 0.0 | 0.0064 | 0.023 |
| $c_{1,3}$ | 0.0 | 145 | 0.412 |
| $c_{2,3}$ | 0.0 | 116 | 0.379 |
| $c_{1,4}$ | 0.0 | 235 | 0.911 |
| $c_{2,4}$ | 0.0 | 131 | 0.624 |
| $c_{3,4}$ | 0.0 | 178 | 0.870 |

**Table 3.** Replication and Validation - trajectory design

| Measure | GM'17 | Repl. $\mu^{*\dagger}$ | Repl. $\sigma^{\ddagger}$ | S'20 |
|---|---|---|---|---|
| | 1.20 | 1.36 | 0.83 | 1.00 |
| $\mu^{*,ind}$ | 1.30 | 1.48 | 0.91 | 1.00 |
| | 3.20 | 3.11 | 1.94 | 1.00 |
| | 0.55 | 0.00 | 0.56 | 0.00 |
| $\sigma^{ind}$ | 0.60 | 0.00 | 0.62 | 0.00 |
| | 1.30 | 0.00 | 1.32 | 0.00 |
| | 14.90 | 16.20 | 9.97 | 2.30 |
| $\mu^{*,full}$ | 12.50 | 13.45 | 8.31 | 1.91 |
| | 10.00 | 9.93 | 6.18 | 1.41 |
| | 6.50 | 0.00 | 6.74 | 0.00 |
| $\sigma^{full}$ | 5.50 | 0.00 | 5.63 | 0.00 |
| | 4.00 | 0.00 | 4.20 | 0.00 |

$^{\dagger}$ $0^{num} = 0.00001$ and $l = 4$.
$^{\ddagger}$ $0^{num} = 0.00000001$ and $l = 24$.

**Table 4.** Replication and Validation - radial design

| Measure | GM'17 | Replication | S'20 |
|---|---|---|---|
| | 0.60 | 0.57 | 1.00 |
| $\mu^{*,ind}$ | 0.75 | 0.85 | 1.00 |
| | 1.50 | 1.31 | 1.00 |
| | 0.20 | 0.10 | 0.00 |
| $\sigma^{ind}$ | 0.30 | 0.41 | 0.00 |
| | 0.85 | 0.22 | 0.00 |
| | 7.50 | 6.84 | 2.30 |
| $\mu^{*,full}$ | 6.80 | 7.77 | 1.91 |
| | 4.75 | 4.19 | 1.41 |
| | 2.90 | 1.15 | 0.00 |
| $\sigma^{full}$ | 2.65 | 3.68 | 0.00 |
| | 2.50 | 0.70 | 0.00 |

**Table 5.** Comparison of sensitivity measures for a linear function

| Parameters | $S_i^T$ | $\gamma_i^*$ | $(\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y})^2$ |
|---|---|---|---|
| $X_1$ | 9 | 3 | 9 |
| $X_2$ | 8 | 2 | 8 |
| $X_3$ | 9 | 1 | 9 |

**Table 6.** EE-based measures by Ge and Menendez (2017) for 100 trajectories

| Parameter | $\mu_T^{*,full}$ | $\mu_T^{*,ind}$ | $\sigma_T^{*,full}$ | $\sigma_T^{*,ind}$ |
|---|---|---|---|---|
| *General* | | | | |
| $\delta$ | 53.40 | 0.00 | 69.23 | 0.09 |
| *Blue-collar* | | | | |
| $\beta^b$ | 3.55 | 0.05 | 4.38 | 0.07 |
| $\beta_e^b$ | 39.84 | 0.05 | 49.69 | 0.07 |
| $\beta_b^b$ | 77.21 | 0.05 | 90.23 | 0.07 |
| $\beta_{bb}^b$ | 2616.50 | 0.05 | 3357.92 | 0.06 |
| $\beta_w^b$ | 94.74 | 0.05 | 113.49 | 0.06 |
| $\beta_{ww}^b$ | 1136.58 | 0.03 | 1405.94 | 0.04 |
| *White-collar* | | | | |
| $\beta^w$ | 5.07 | 0.05 | 6.42 | 0.06 |
| $\beta_e^w$ | 90.25 | 0.07 | 111.50 | 0.08 |
| $\beta_w^w$ | 82.88 | 0.05 | 103.66 | 0.07 |
| $\beta_{ww}^w$ | 2444.13 | 0.06 | 3044.69 | 0.07 |
| $\beta_b^w$ | 452.91 | 0.07 | 490.31 | 0.09 |
| $\beta_{bb}^w$ | 4317.58 | 0.05 | 4851.54 | 0.06 |
| *Education* | | | | |
| $\beta^e$ | 0.00 | 0.09 | 0.00 | 0.10 |
| $\beta_{he}^e$ | 0.00 | 0.11 | 0.00 | 0.13 |
| $\beta_{re}^e$ | 0.00 | 0.04 | 0.000 | 0.09 |
| *Home* | | | | |
| $\beta^h$ | 0.00 | 0.04 | 0.00 | 0.05 |
| *Lower Triangular Cholesky Matrix* | | | | |
| $c_1$ | 27.94 | 0.07 | 33.72 | 0.08 |
| $c_2$ | 31.89 | 0.05 | 38.58 | 0.06 |
| $c_3$ | 0.00 | 0.06 | 0.00 | 0.07 |
| $c_4$ | 0.00 | 0.04 | 0.00 | 0.09 |
| $c_{1,2}$ | 12.41 | 0.06 | 14.33 | 0.08 |
| $c_{1,3}$ | 0.00 | 0.09 | 0.00 | 0.10 |
| $c_{2,3}$ | 0.00 | 0.05 | 0.00 | 0.06 |
| $c_{1,4}$ | 0.00 | 0.04 | 0.00 | 0.05 |
| $c_{2,4}$ | 0.00 | 0.03 | 0.00 | 0.03 |
| $c_{3,4}$ | 0.00 | 0.04 | 0.00 | 0.05 |

**Table 7.** Mean absolute correlated and uncorrelated elementary effects (based on 100 subsamples in trajectory and radial design)

| Parameter | $\mu_T^{*,c}$ | $\mu_R^{*,c}$ | $\mu_T^{*,u}$ | $\mu_R^{*,u}$ |
|---|---|---|---|---|
| *General* | | | | |
| $\delta$ | 17 | 23 | 476 | 415 |
| *Blue-collar* | | | | |
| $\beta^b$ | 1 | 3 | 43 | 88 |
| $\beta_e^b$ | 11 | 14 | 406 | 443 |
| $\beta_b^b$ | 25 | 51 | 688 | 1169 |
| $\beta_{bb}^b$ | 871 | 934 | 15 540 | 17 860 |
| $\beta_w^b$ | 29 | 48 | 73 | 143 |
| $\beta_{ww}^b$ | 389 | 460 | 869 | 1183 |
| *White-collar* | | | | |
| $\beta^w$ | 1 | 3 | 50 | 117 |
| $\beta_e^w$ | 26 | 28 | 943 | 852 |
| $\beta_w^w$ | 24 | 47 | 718 | 1521 |
| $\beta_{ww}^w$ | 933 | 997 | 12 257 | 18 069 |
| $\beta_b^w$ | 131 | 127 | 309 | 356 |
| $\beta_{bb}^w$ | 1230 | 1352 | 2088 | 2477 |
| *Education* | | | | |
| $\beta^e$ | 0.0008 | 0.0002 | 0.001 | 0.003 |
| $\beta_{he}^e$ | 0.0001 | 0.0002 | 0.001 | 0.001 |
| $\beta_{re}^e$ | 0.0003 | 0.0002 | 0.0003 | 0.0006 |
| *Home* | | | | |
| $\beta^h$ | 0.0003 | 0.0003 | 0.000 02 | 0.000 02 |
| *Lower Triangular Cholesky Matrix* | | | | |
| $c_1$ | 8 | 16 | 18 | 37 |
| $c_2$ | 8 | 11 | 22 | 24 |
| $c_3$ | 0.0004 | 0.0004 | 0.0004 | 0.0007 |
| $c_4$ | 0.0004 | 0.000 08 | 0.0002 | 0.0003 |
| $c_{1,2}$ | 4 | 4 | 10 | 10 |
| $c_{1,3}$ | 0.0005 | 0.0006 | 0.0006 | 0.0005 |
| $c_{2,3}$ | 0.0003 | 0.0005 | 0.0006 | 0.001 |
| $c_{1,4}$ | 0.000 04 | 0.000 05 | 0.0004 | 0.0005 |
| $c_{2,4}$ | 0.0001 | 0.0002 | 0.0001 | 0.0002 |
| $c_{3,4}$ | 0.0001 | 0.0001 | 0.000 08 | 0.0001 |

## 8.2 Appendix B: Figures

**Figure 1.** Series of events

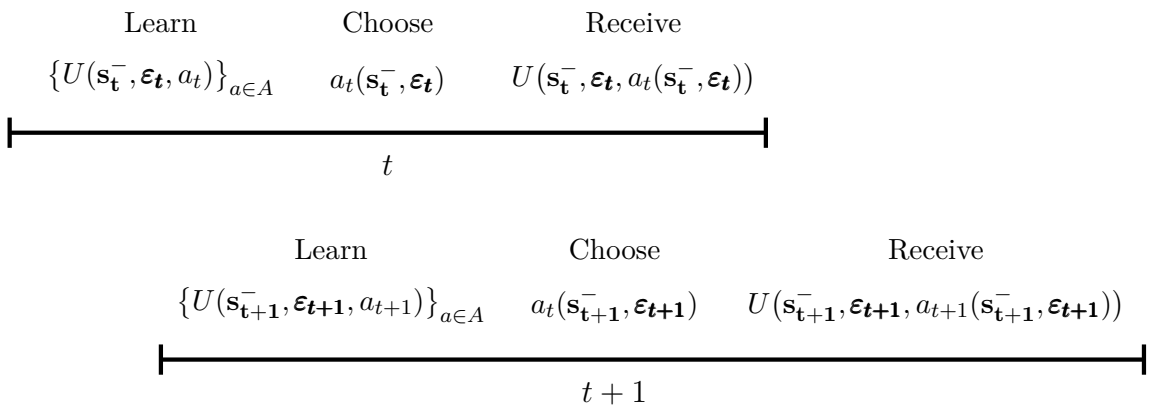



**Figure 2.** Correlations between estimates for important input parameters

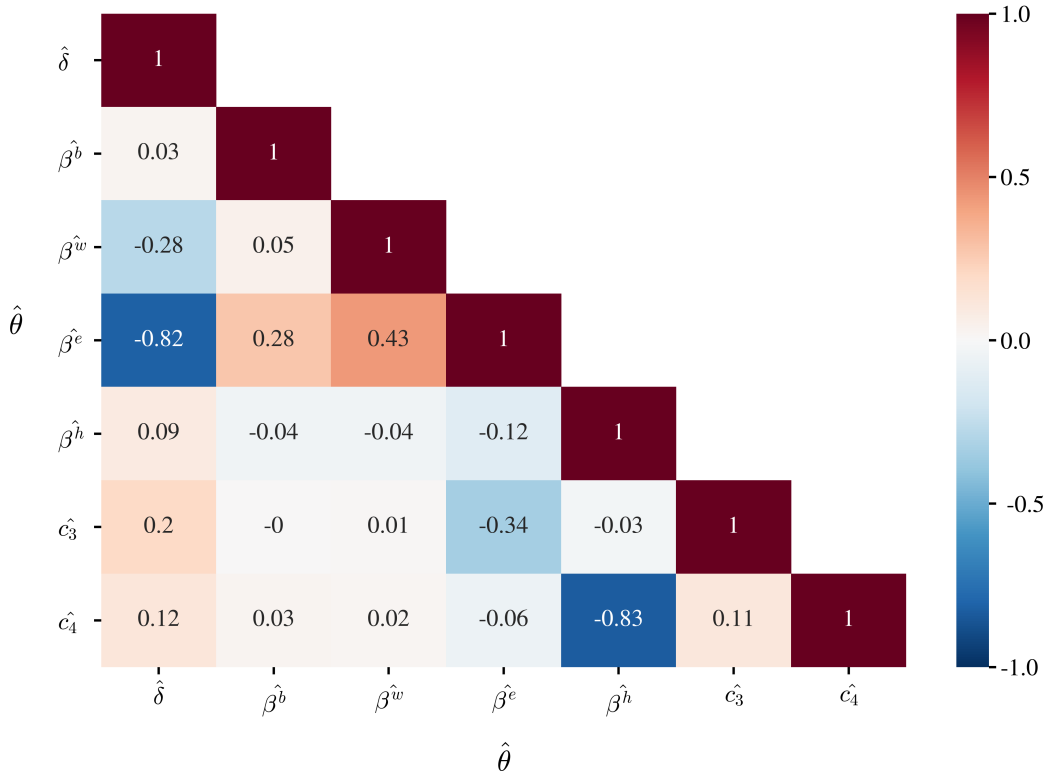


**Figure 3.** Comparison of shares of occupation decisions over time between scenarios

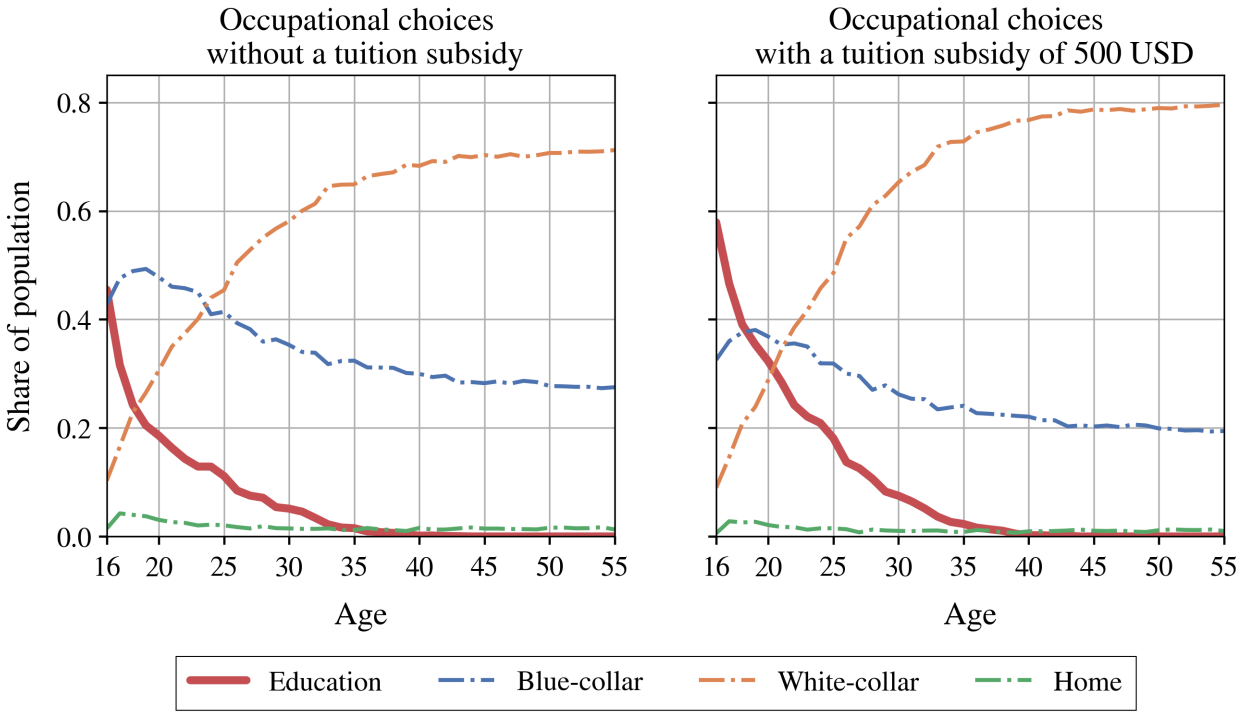



**Figure 4.** Grid points in standard normal sample space for trajectory design with $l = 4$

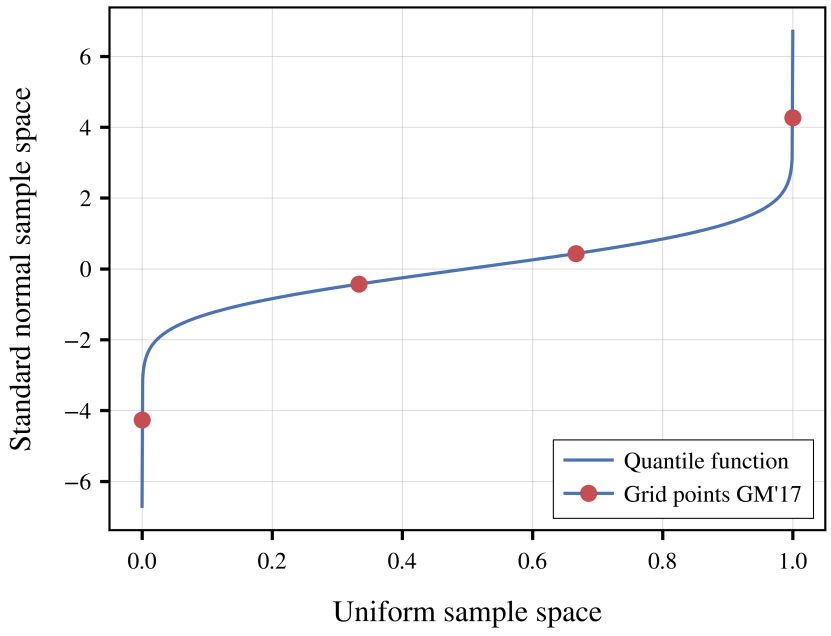


**Figure 5.** Comparison of shares of occupation decisions over time between scenarios including 99% confidence intervals

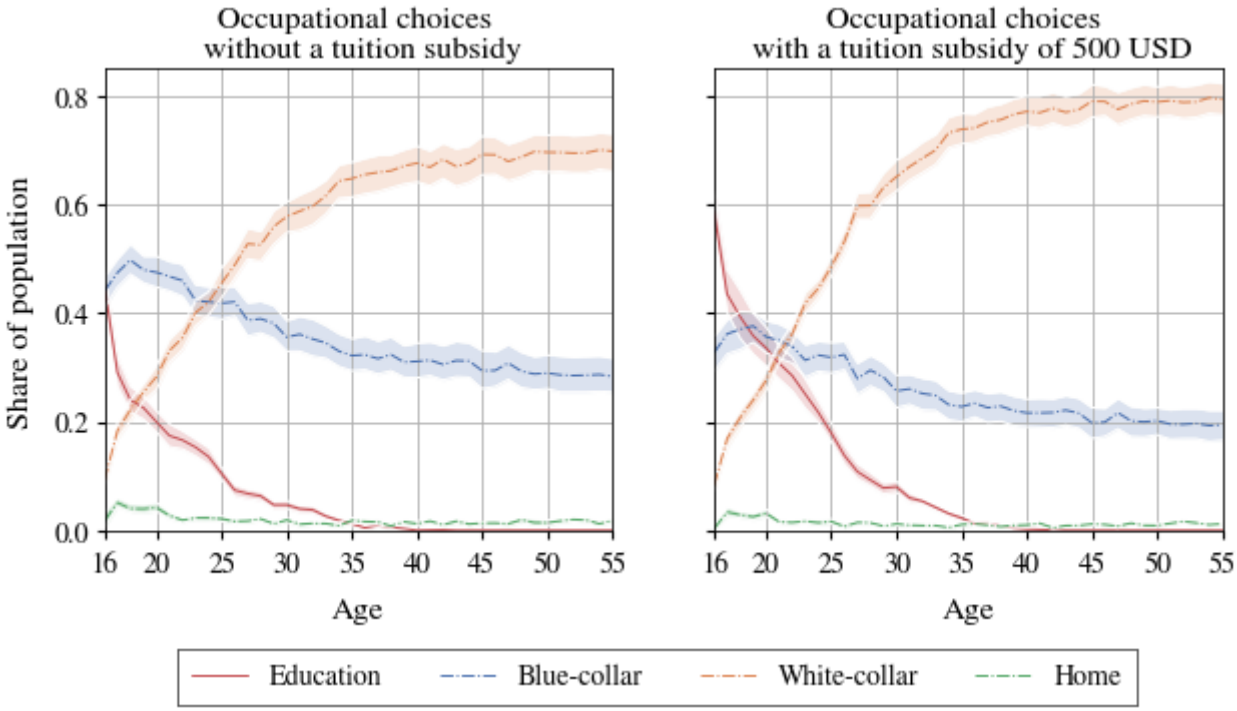



**Figure 6.** Probability distribution of QoI $Y$

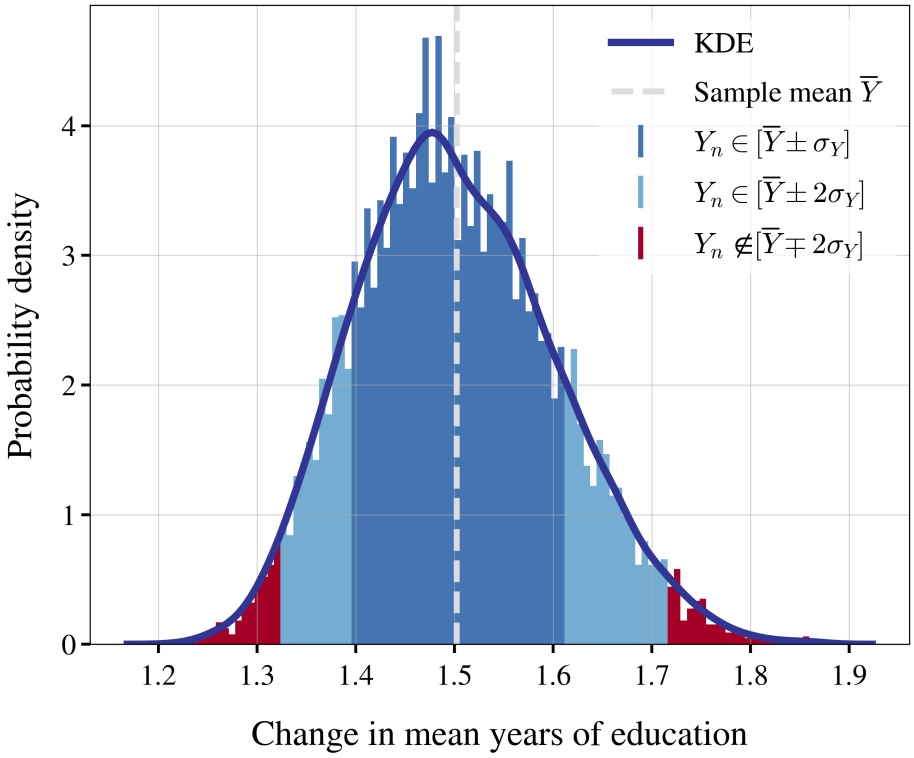


**Figure 7.** Sigma-normalized mean absolute Elementary Effects for radial design

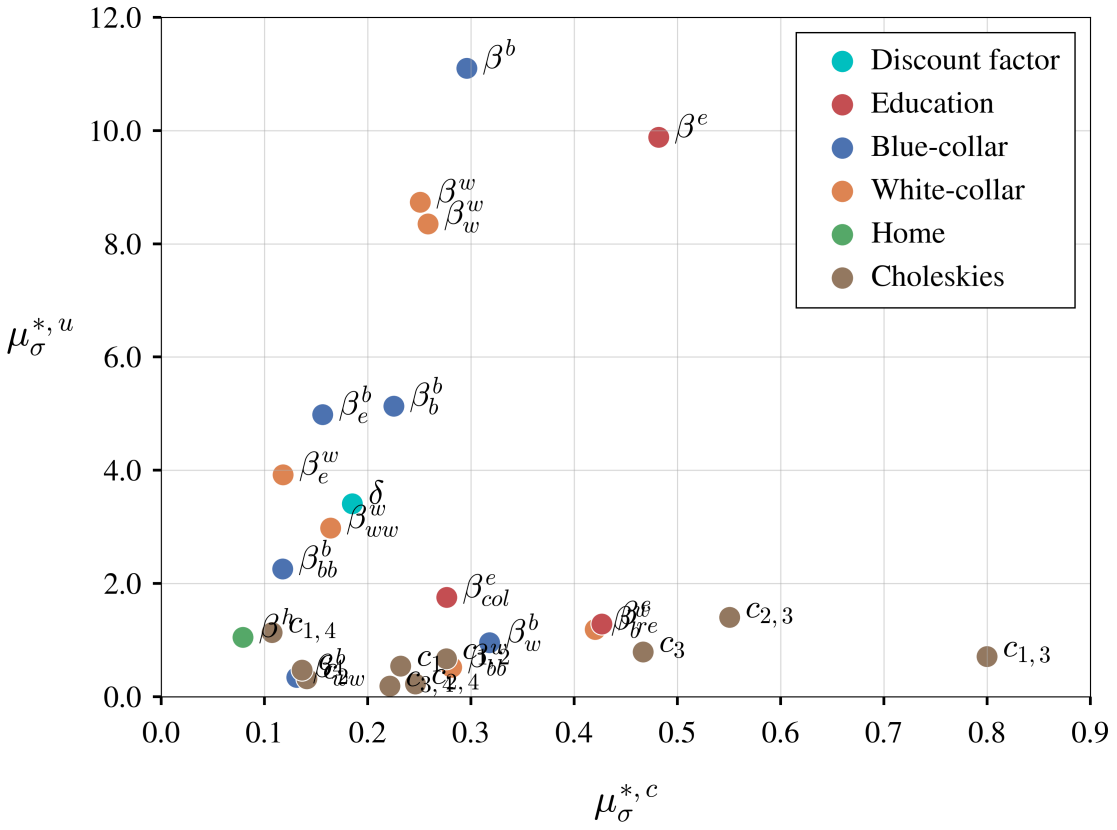
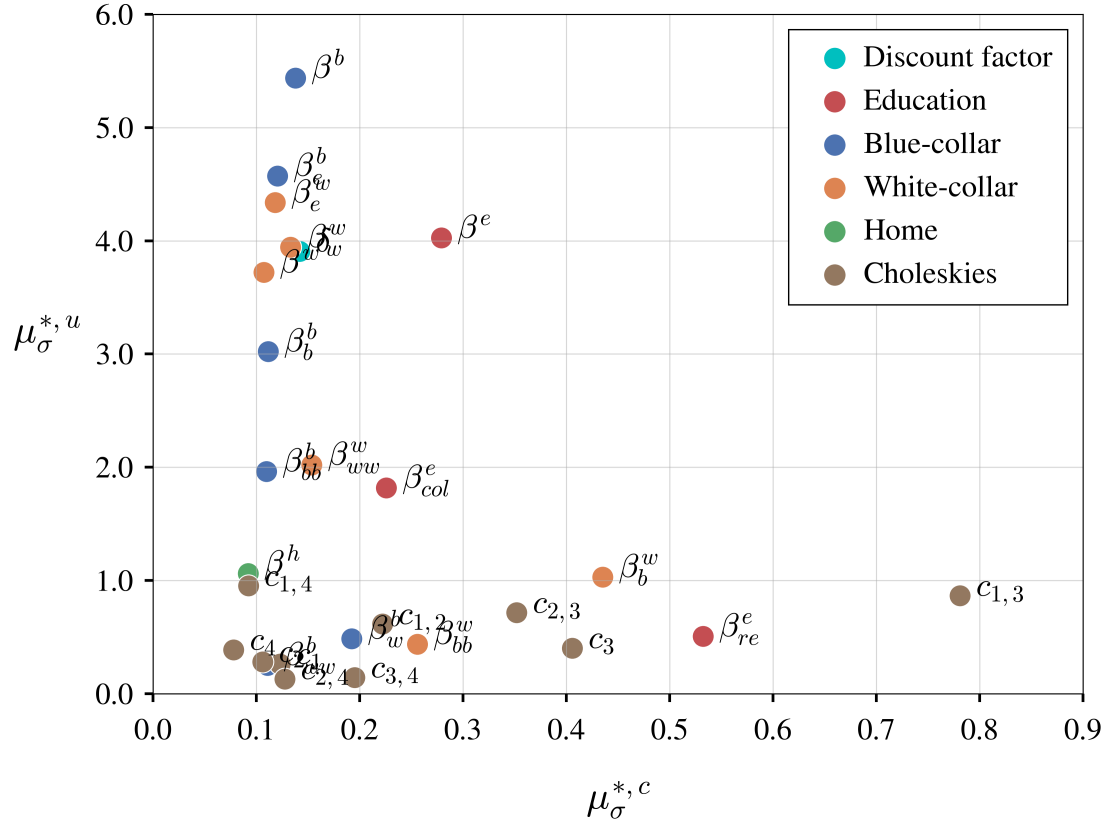



**Figure 8.** Sigma-normalized mean absolute Elementary Effects for trajectory design



## 8.3 Appendix C: 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))[36].

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 is to estimate the exogenous model parameters $\boldsymbol{\theta} \stackrel{\text{def}}{=} (\delta, \boldsymbol{\beta}, \boldsymbol{\Sigma_\varepsilon})$.[37] Thus, in the following, every probability is a function of the exogenous model parameters $\boldsymbol{\theta}$. The approach to compute the likelihood function $L_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})$ of the observables in the data begins with the individual latent variable representation in period $t$.

$$a_t = \underset{a}{\operatorname{argmax}}\, V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t) \tag{34}$$

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

$$\boldsymbol{\varepsilon_t}(a_t(j), \mathbf{s_t^-}) \stackrel{\text{def}}{=} \{\boldsymbol{\varepsilon_t} | V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t(j)) = \max_{a_t} V(\mathbf{s_t^-}, \boldsymbol{\varepsilon_t}, a_t)\}). \tag{35}$$

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,

$$\mathrm{p}\Big(a_t(j) | \mathbf{s_t^-}\Big) = \int_{\boldsymbol{\varepsilon_t}(a_t(j), \mathbf{s_t^-})} f(\boldsymbol{\varepsilon_t}) d^{|A|} \boldsymbol{\varepsilon_t}. \tag{36}$$

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 (32), 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

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

[37] 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 $\beta$ is not left out of the estimation.

wages are logarithmized. Thus,

$$\varepsilon_{a(k),t} = \ln(\mathcal{W}^{-}_{a(k),t}) - \ln(W^{-}_{a(k),t}). \tag{37}$$

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

$$\mathrm{p}\Big(a_t(k)|\mathbf{s_t^-}, W^{-}_{a(k),t}\Big) = \int_{\boldsymbol{\varepsilon_t}(a_t(k),\mathbf{s_t^-})} f(\boldsymbol{\varepsilon_t}|\varepsilon_{a(k),t}) d^{|A|}\boldsymbol{\varepsilon_t}. \tag{38}$$

Applying integration by substitution yields the following expression for the probability of the observed wage:[38]

$$\mathrm{p}\Big(\mathcal{W}^{-}_{a(k),t}|\mathbf{s_t^-}\Big) = \omega_t^{-1} \frac{1}{\sigma_{a(k)}} \phi\left(\frac{\varepsilon_{a(k),t}}{\sigma_{a(k)}}\right). \tag{39}$$

Here, $\omega_t^{-1}$ is the Jacobian of the transformation from observed wage $\mathcal{W}^{-}_{a(k),t}$ to error $\varepsilon_{a(k),t}$ in (37) and $\phi$ is the standard normal probability density function. Finally, the joint probability of observing choice $a_t(k)$ and wage $\mathcal{W}^{-}_{a(k),t}$ conditional on the observed states is given by the product of the two probabilities in (38) and (39):

$$\mathrm{p}\Big(a_t(k), \mathcal{W}^{-}_{a(k),t}|\mathbf{s_t^-}\Big) = \mathrm{p}\Big(a_t(k)|\mathbf{s_t^-}, \mathcal{W}^{-}_{a(k),t}\Big)\mathrm{p}\Big(\mathcal{W}^{-}_{a(k),t}|\mathbf{s_t^-}\Big) \tag{40}$$

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:

$$L^i_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) = P\Big(\{a^i_{t,}, \mathcal{W}^{-,i}_{a,t,}\}^T_{t=0}\Big) = \prod_{t=0}^{T} \mathrm{p}\Big(a^i_t, \mathcal{W}^{-,i}_{a,t}|\mathbf{s_t^{-,i}}\Big) \tag{41}$$

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

$$L_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) = P\Big(\Big\{\{a^i_{t,}, \mathcal{W}^{-,i}_{a,t,}\}^T_{t=0}\Big\}_{i\in I}\Big) = \prod_{i\in I}\prod_{t=0}^{T} \mathrm{p}(a^i_t, \mathcal{W}^{-,i}_{a,t}|\mathbf{s_t^{-,i}}) \tag{42}$$

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 (42). As maximum likelihoods estimates are asymptotically normal[39], these results are taken as the

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

[39] 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.

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.[40] The result is used as the covariance matrix for the input parameters in the UQ.

The asymptotic covariance of a maximum likelihood estimator equals the inverse of the Fisher information matrix: $\text{Var}(\theta) = \mathcal{I}(\theta)^{-1}$. In this thesis, the information matrix $\mathcal{I}(\boldsymbol{\theta})$ is given by the variance of the scores of the parameters.[41] 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

$$\text{s}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\partial L_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i \in I} \frac{L^i_{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \sum_{i \in I} \text{s}_i(\boldsymbol{\theta}). \tag{43}$$

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

$$\mathcal{I}^{-1}(\boldsymbol{\theta}) = \text{Var}(\text{s}(\boldsymbol{\theta})) = \mathbb{E}[\text{s}(\boldsymbol{\theta})\text{s}(\boldsymbol{\theta})']. \tag{44}$$

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

$$\hat{\text{Cov}}_J(\hat{\boldsymbol{\theta}}) = \left( \frac{1}{|I|} \sum_{i \in I} \text{s}_i(\hat{\boldsymbol{\theta}}) \text{s}_i(\hat{\boldsymbol{\theta}})' \right)^{-1}. \tag{45}$$

$|I|$ 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 (43) and (44) then imply that the inverse of the variance of the individual scores is equivalent to the variance of the maximum likelihood estimator.

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

[41] 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.

# 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.

| | |
|---|---|
| Place, Date | Signature |