

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

Contents

Abbreviations	I
List of Figures	II
List of Tables	III
1 Introduction	1
2 Uncertainty Quantification Framework	2
2.1 Overview of Uncertainty Quantification	2
2.2 Sensitivity Analysis	4
2.2.1 Quantitative Global Sensitivity Analysis	5
2.2.2 Qualitative Global Sensitivity Analysis	6
2.3 Correlated input parameters	8
2.4 Correlated input parameters	9
3 Uncertainty Quantification in the Economic Literature	10
4 Elementary Effects for functions with correlated input parameters	16
4.1 Sampling Schemes	16
4.2 The approach for correlated input parameters in Ge and Menendez (2017)	18
4.3 Implementation	22
4.4 Drawbacks in Ge and Menendez (2017) and Corrected Elementary Effects.	23
4.5 Replication and Validation	24
5 The Occupational Choice Model	27
5.1 Keane and Wolpin (1994)	28
5.2 Estimation Results	31
5.3 Quantity of Interest	33
6 Uncertainty Analysis	36
7 Qualitative Global Sensitivity Analysis	37
8 Discussion	39
9 Conclusion	39
References	40
Appendix	45

Abbreviations

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

USD, GS, UQ, QoI

List of Figures

1	Grid points for trajectory design in Ge and Menendez (2017)	26
2	Series of events	29
3	Correlations between estimates for important input parameters	32
4	Comparison of shares of occupation decision over time between scenarios .	34
5	Probability distribution of quantity of interest q	36
6	Comparison of shares of occupation decision over time between scenarios with cone plots	37
7	Sigma-normalized mean absolute Elementary Effects for trajectory design .	38
8	Sigma-normalized mean absolute Elementary Effects for radial design . . .	38

List of Tables

1	Overview of UQ literature	11
2	Replication and validation for test case 1 - trajectory design	25
3	Replication and validation for test case 1 - radial design	25
4	Model Parametrization	35

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. After general remarks, this part, in turn, 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 to make statements about the future. 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 may 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 the quantity of interest (henceforth QoI) 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

¹See page ix.

²The 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. The concept of model uncertainty is briefly explained in the next paragraph.

³See page 4-7.

⁴However, 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, \dots, 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 $f(X_1, X_2, \dots, X_k)$. Thus,

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

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 mainly 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 section 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 after this overview.

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, then variance-based UQ measures are meaningful. If the distribution has

⁵See page 8-10.

⁶See page 33-34.

a less tractable, for instance, a bimodal shape, then density-based measures are better suited (Plischke et al. (2013a)).

2.2 Sensitivity Analysis

This section draws largely 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 the 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 the data acquisition and estimation of a subset of parameters. The methods that meet 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 old 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, \bar{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 \bar{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 $\text{Var}(Y)$. For these reasons, the fields, its definition and its methods have evolved beyond the notion of local sensitivity analysis. Yet, until recently, the main part of applications in different fields, such as physics (Saltelli et al. (2004))⁷ and economics (Harenberg et al. (2019)), still uses local measures.

2.2.1 Quantitative Global Sensitivity Analysis

The quantitative GSA aims to determine the precise effect size of each random 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 $\text{Var}_{X_i}(Y|X_i)$ denote the variance of the model output Y conditional on input parameter X_i . Then,

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

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, $\mathbf{X}_{\sim i}$, of $\mathbf{X}_{\sim i}$ and a given value $X_i = x_i^*$. Then, we use $\mathbb{E}[f(X_i = x_i^*, \mathbf{X}_{\sim i})] \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{X}_{\sim i}}[Y|X_i]$ to write the first-order Sobol' index as the variance of $\mathbb{E}_{\mathbf{X}_{\sim i}}[Y|X_i]$ over all x_i^* s as

$$S_i = \frac{\text{Var}_{X_i}(\mathbb{E}_{\mathbf{X}_{\sim i}}[Y|X_i])}{\text{Var}(Y)}. \quad (3)$$

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 .

$$S_i^T = \frac{\text{Var}_{\mathbf{X}_{\sim i}}(\mathbb{E}_{X_i}[Y|\mathbf{X}_{\sim i}])}{\text{Var}(Y)} \quad (4)$$

⁷See page 42.

Computing these measures requires many function evaluations, even if an estimator is used as a shortcut (Saltelli et al. (2004))⁸. The more time-intense one function evaluation is, the more utility provides the factor fixing based on qualitative measures.

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, μ , the mean absolute EEs, μ^* , and the standard deviation of the EEs, σ . 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 Δ . The only restriction is that $X_i + \Delta$ is in the sample space of X_i . The Elementary Effect, or derivative, is denoted by

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

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 Δ has not to be infinitesimally small. To offset the third drawback of d_i or S_i , that base vector X_i does not represent the whole input space, one computes the mean EE, μ_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 . This measure is given by

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

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

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

⁸See page 124 -149.

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

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

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

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

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 Δ is somewhat arbitrary if the derivative is not analytic. To this date, the literature has not developed convincing solutions to 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

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

This expression makes use of the squared EE. In light of this characteristic, the use of σ_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 Δ . More precisely, depending on the sampling scheme, Δ might be random instead of arbitrary and constant. In both cases, Δ 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 Δ .

⁹See page 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 have been 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 the correlated setup has been 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 considerable¹¹, 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 the sigma-normalised mean absolute Elementary Effects. The measure in Kucherenko et al. 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. isregarded, is also offset because elements $X_{\sim i}$ are also resampled for each new X_i . This measure is given by

$$\mu_i = \frac{1}{r} \sum_{j=1}^r d_i^{(j)}. \quad (10)$$

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

$$\mu_i^* = \frac{1}{r} \sum_{j=1}^r |d_i^{(j)}|. \quad (11)$$

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

¹⁰See page 46.

¹¹On 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 not small.

¹²The fundamental reason is that the unique first- and higher-order Sobol' indices do not add to one if input parameters are correlated.

One last improvement is provided in Smith (2014)¹³. That is, the scaling of μ_i^* by $\frac{\sigma_{X_i}}{\sigma_Y}$. This measure is called the sigma-normalized mean absolute EE:

$$\mu_{i,\sigma}^* = \mu_i^* \frac{\sigma_{X_i}}{\sigma_Y}. \quad (12)$$

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

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 Δ is somewhat arbitrary if the derivative is not analytic. To this date, the literature has not developed convincing solutions to 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

$$S_i^T \leq \frac{\frac{1}{r} \sum_{j=1}^r |d_i^{2(j)}|}{\pi^2 \sigma_Y}. \quad (13)$$

This expression makes use of the squared EE. In light of this characteristic, the use of σ_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 Δ . More precisely, depending on the sampling scheme, Δ might be random instead of arbitrary and constant. In both cases, Δ 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 Δ .

2.4 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

¹³See page 332.

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 have been 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 the correlated setup has been made by Ge and Menendez (2017). Some authors, e.g. Saltelli et al. (2004)¹⁴ 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 considerable¹⁵, 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 the sigma-normalised mean absolute Elementary Effects. The measure in Kucherenko et al. is discarded because their results are not valid in the presence of correlations.¹⁶

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.¹⁷ Also GSA in particular has had strong advocates.¹⁸ 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 GSA because factor fixing is not the objective of the respective publications. Table 1 gives an overview of the major topics, analyses, measures and methods in the literature.

¹⁴See page 46.

¹⁵On 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 not small.

¹⁶The fundamental reason is that the unique first- and higher-order Sobol' indices do not add to one if input parameters are correlated.

¹⁷See Hansen and Heckman (1996), Kydland (1992) and Canova (1994), amongst others.

¹⁸See Canova (1995) and Gregory and Smith (1995).

Table 1. Overview of UQ literature

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

I find 14 contributions that meet the described criteria. Arguably, because UQ is more accomplished in climatology, a large share of research comes from climate economics. Another field where UQ finds some application is macroeconomics. Remarkably, no contribution computes their own estimates for the 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 Bilonis (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 \bar{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 \bar{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 δ -sensitivity measure, and correlation measures for paired QoIs in their GSA. The δ -sensitivity measure (see, e.g., Borgonovo (2006)) is density-based. It is given by half the expectation value of the absolute difference between the unconditional distribution of a QoI and the QoI distribution conditioned on

one specific, fixed input (group). Estimates for these measures are computed with the algorithm used in Plischke et al. (2013b) applied to a Monte-Carlo sample ($N=10,000$). In Anderson et al. (2014), the δ -sensitivity measure is the main measure of sensitivity and used to rank the parameters in terms of their contributions to the model uncertainty. The authors also use a surrogate model obtained through Cut-HDMR (cut-high dimensional model representation; see, e.g., Ziehn and Tomlin (2009)) for graphical analyses of the 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 δ -sensitivity measure. They compute the Sobol' indices based on Sobol' sequences (Sobol' (1967)) for the results and based on Latin Hypercube sampling (McKay et al. (1979)) as a check. The results in Butler et al. (2014) and Anderson et al. (2014) can not be compared as they analyse different QoIs.

Miftakhova (2018) applies the GSA procedure outlined by Harenberg et al. (2019). The importance ranking that she obtains from the polynomial-chaos-expansions-based Sobol' indices is different from the ranking that Anderson et al. (2014) obtain from the δ -sensitivity measure. Yet, this is not mentioned by Miftakhova.¹⁹ However, the author emphasizes that the standard procedure for obtaining Sobol' indices from a variance decomposition as used by Anderson et al. (2014) and Butler et al. (2014) is not feasible for the DICE model because a set of input parameters is calibrated jointly in order to let the model match some observables. Therefore, although these input parameters are not correlated in the classical sense, they are dependent. Hence, the variance-based Sobol decomposition is not applicable because the summands are not orthogonal to each other or, in other words, the input-specific variance terms contain a covariance component. Thus, they do not add to the total model variance and Sobol' indices cannot be computed directly. Miftakhova (2018) shows how the set of dependent input parameters can be changed to a set of independent parameters by changing the model structure: She includes uncertain 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.²⁰

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

¹⁹I do not have access to the numerical codes. Thus the reasons for the discrepancies remain unclear.

²⁰For 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.

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 Bilonis (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 Bilonis' 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 Bilonis, 2019, p. 4) from the value function interpolant. In other words, this step yields the QoI expressed by a surrogate model. Thus, to add an UQ, one has to, first, specify the uncertain parameters as continuous state variables, and second, assign a probability distribution to each of these parameters. Then (assuming the uncertain input parameters are independent), one provides a sample from each parameter's distribution as input to the Gaussian process regression to obtain a surrogate model. Following these steps, QoIs can be expressed as functions of the uncertain input parameters without much additional effort. Finally, by using a processed value function interpolant as a surrogate model, Scheidegger and Bilonis propagate the model uncertainty and depict univariate effects.

Building on the contributions by Harenberg et al. (2019) and Scheidegger and Bilonis (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 Bilonis' intrusive framework. However different to Scheidegger and Bilonis (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 next section describes the method in Ge and Menendez (2014) to compute EE-based measures for models with correlated input parameters. It also explains how the thesis improves these measures. In contrary to the quantitative measures in the literature review, these EE-based measures are qualitative and used to identify and fix input param-

eters that are irrelevant for the uncertainty in the QoI. The results can be used to apply more-involved, quantitative measures, as presented in this chapter, for an accelerated model with less uncertain input parameters. However, these measures must also be extended to allow correlations between the input.

4 Elementary Effects for functions with correlated input parameters

4.1 Sampling Schemes

According to several experiments by Campolongo et al. (2011) using common test functions, 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 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 $2, \dots, k + 1$ 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:

$$\mathbf{R}_{(k+1) \times k} = \begin{pmatrix} a_1 & a_2 & \dots & a_k \\ \mathbf{b}_1 & a_2 & \dots & a_k \\ a_1 & \mathbf{b}_2 & \dots & a_k \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & \mathbf{b}_k \end{pmatrix} \quad (14)$$

Note here, that each column consists only of the first row element, except of for 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 subtrahend in the formula for the EE. Then, $\Delta^{(i,j)} = b_i^{(j)} - a_i^{(j)}$. Tailored to the radial scheme, the derivative can be re-formulated as in Equation (3), where i represents the input parameter X_i . We abstract from multiple sample sets and do not use index j compared to Equation (3). The asterisk is an index for a complete vector.

$$d_i = \frac{Y(\mathbf{a}_{\sim i}, b_i) - Y(\mathbf{a})}{b_i - a_i} = \frac{Y(\mathbf{R}_{i+1,*}) - Y(\mathbf{R}_{1,*})}{b_i - a_i}. \quad (15)$$

If the number of radial subsamples is high, the draws from the quasi-random sequence lead

to a good and fast coverage of the input space (compared to a random sequence). The quasi-random sequence considered here is the Sobol' sequence. This sequence is comparably successful in covering the unit hypercube, but also conceptually more involved. Therefore, its presentation is beyond the scope of this work. Since this sequence is quasi-random, the sequence has to be drawn at once for all sets of radial matrices.

Next, I present the trajectory design. As we will see, it leads to a relatively representative coverage for a very small number of subsamples but also to frequent repetitions of similar draws. In this outline, I skip the equations that generate a trajectory and present the method verbally. There are different forms of trajectories. Here, I focus on the version presented in Morris (1991) that yields to 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 Δ to the lowest point in the lowest half results in the lowest point of the upper half of the grid points, and so on. The rest of the rows is constructed by, first, copying the row one above and, second, by adding Δ to the i -th element of the $i + 1$ -th row. The implied matrix scheme is depicted below.

$$\mathbf{T}_{(k+1) \times k} = \begin{pmatrix} a_1 & a_2 & \dots & a_k \\ \mathbf{b}_1 & a_2 & \dots & a_k \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & a_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_k \end{pmatrix} \quad (16)$$

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 randomly switching rows and 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 does sustain the stairs shape. Yet, this implies that $\Delta^{(i)}$ 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:²¹

²¹In contrary to most authors, I also denote the step as a subtraction instead of Δ when referring to the trajectory design. This provides additional clarity.

$$d_i = \frac{Y(\mathbf{b}_{\mathbf{f} \leq \mathbf{i}}, \mathbf{a}_{\mathbf{h} > \mathbf{i}}) - Y(\mathbf{b}_{\mathbf{f} < \mathbf{i}}, \mathbf{a}_{\mathbf{h} \geq \mathbf{i}})}{b_i - a_i} = \frac{Y(\mathbf{T}_{\mathbf{i}+1,*}) - Y(\mathbf{T}_{\mathbf{i},*})}{b_i - a_i}. \quad (17)$$

The trajectory design involves first, a fixed grid, and second and more importantly, a fixed step Δ . s.t. $\{\Delta\} = \{\pm\Delta\}$. This implies less step variety and less space coverage vis-à-vis the radial design for more than a small number of draws.

So far, we have only considered draws in $[0,1]$. For uncorrelated input parameters from arbitrary distributions with well-defined cumulative distribution function, Φ , one would simply evaluate each element (of course, potentially including the addition of the step) by the inverse cumulative distribution function, Φ^{-1} , of the respective parameter. A bit of intuition is, that Φ maps the sample space to $[0,1]$. Hence Φ^{-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 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 incomplete 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 (6) and Equation (8). The reason is that the correlations of parameter X_i , to which step Δ^i is added, implies that all other parameters in the same row with non-zero correlation in $\mathbf{X}_{\sim \mathbf{i}}$ are changed as well. Therefore, the rows cannot be denoted and compared as easily by a 's and b 's as in Equation (6) and (8). 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) fail to fully develop these measures. I will explain how their measures lead to arbitrary results 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 $\mathbf{u} = \{u_1, u_2, \dots, u_k\}$ from $[0, 1]$ for an input parameter vector to draws $\mathbf{x} = \{x_1, x_2, \dots, x_k\}$ from an arbitrary joint normal distribution. We will do this in three steps.

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

Creating potentially correlated draws \mathbf{x} from $\mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ is simple. Following Gentle (2006),

page 197, this can be achieved the following way: Draw a k -dimensional row vector of i.i.d standard normal deviates from the *univariate* $N_1(0, 1)$ distribution, such that $\mathbf{z} = \{z_1, z_2, \dots, z_k\}$, and compute the Cholesky decomposition of Σ , such that $\Sigma = \mathbf{T}^T \mathbf{T}$. The lower triangular matrix is denoted by \mathbf{T}^T . Then apply the operation in Equation (10) to obtain the correlated deviates from $\mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$.

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{T}^T \mathbf{z}^T \quad (18)$$

Intuition for the mechanics that underlie is provided in Appendix A (not contained in this draft).

The next step is to understand that we can split the operation in Equation (10) into two subsequent operations. 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 $\mathbf{z}_c = \mathbf{Q}^T \mathbf{z}^T$. In this equation, \mathbf{Q}^T is the lower matrix from the Cholesky decomposition $\mathbf{R}_k = \mathbf{Q}^T \mathbf{Q}$. 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: $\mathbf{z} = \mathbf{z}_c(\mathbf{i})\boldsymbol{\sigma}(\mathbf{i}) + \boldsymbol{\mu}_.$, where the i s indicate an element-wise multiplication.

The last step to construct the final approach is to recall the inverse transform sampling 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 operation in the above paragraph. The transformation from \mathbf{u} to \mathbf{z} is denoted by $F^{-1}(\Phi)$ and summarized by the following three points:

Step 3 is specific to the normal sample space in which we are interested in. A mapping to other sample spaces can, for instance, be achieved by, first, applying Φ and, second, the inverse cumulative distribution function of the desired distribution.

The procedure described in the three steps above is equivalent to an inverse Rosenblatt transformation, a linear inverse Nataf transformation²² and connects to Gaussian copulas. 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 varying assumptions. Notes on this relation are provided in Appendix B (not contained in this draft).

The one most important point to understand is that the transformation comprised by the three steps listed above is not unique for correlated input parameters. Rather, the transformation changes with the order of parameters in vector \mathbf{u} (see also the formal

²²For the first two transformations, see Lemaire (2013), page 78 - 113

definition of the Rosenblatt transformation). 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 $\boldsymbol{\rho}_i^{*,j} = (\rho_{1,j}, \rho_{2,j}, \dots, \rho_{i-1,j})$ 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

$$\mathbf{Q}^T = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \rho_{1,2} & \sqrt{1 - \rho_{1,2}^2} & 0 & \dots & 0 \\ \rho_{1,3} & \frac{\rho_{2,3} - \rho_{1,2}\rho_{1,3}}{\sqrt{1 - \rho_{1,2}^2}} & \sqrt{1 - \boldsymbol{\rho}_3 \mathbf{R}_2^{-1} \boldsymbol{\rho}_3^T} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1,k} & \frac{\rho_{2,k} - \rho_{1,2}\rho_{1,k}}{\sqrt{1 - \rho_{1,2}^2}} & \frac{\rho_{3,k} - \boldsymbol{\rho}_3^{*,k} \mathbf{R}_2^{-1} \boldsymbol{\rho}_3^T}{\sqrt{1 - \boldsymbol{\rho}_3 \mathbf{R}_2^{-1} \boldsymbol{\rho}_3^T}} & \dots & \sqrt{1 - \boldsymbol{\rho}_k \mathbf{R}_2^{-1} \boldsymbol{\rho}_k^T} \end{pmatrix}. \quad (19)$$

Equation (11), together with Step 2, implies that the order of the uncorrelated standard normal deviates \mathbf{z} constitute 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 $\mathbf{Q}^T \mathbf{z}^T \neq \mathbf{Q}^T (\mathbf{z}')^T$ 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. One could say that the correlations are transferred like the fall of dominoes.

Coming back to the EE-based measures for correlated inputs, Ge and Menendez (2017) attempt to construct two variations of derivatives / EEs, d , in Equation (3). They call one derivative the independent Elementary Effects, d_i^{ind} , and the other the full Elementary Effects, d_i^{full} . For each of these two EEs, they derive the aggregate measures μ , μ^* and σ . The difference between the two EEs is, that d^{ind} excludes and d^{full} includes the effect of the correlations from adding the step Δ^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 might possibly change the d_i^{full} s of $\mathbf{X}_{\sim i}$.

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

$$d_i^{\text{full},T} = \frac{f(\mathcal{T}(\mathbf{T}_{i+1,*}; i-1)) - f(\mathcal{T}(\mathbf{T}_{i,*}; i))}{\Delta}. \quad (20)$$

In Equation (12), $\mathcal{T}(\cdot; i) \stackrel{\text{def}}{=} \mathcal{T}_3\left(\mathcal{T}_2(\mathcal{T}_1(\cdot; i)); i\right)$. $\mathcal{T}_1(\cdot; i)$ orders the parameters / row elements to establish the right correlation hierarchy. \mathcal{T}_2 , or $F^{-1}(\Phi)$, 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 first row elements that is cut and moved to the back of the row in the same order and the reverse transformation. 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) yields the following two transformed trajectories:

$$\mathcal{T}_1(\mathbf{T}_{i+1,*}; i-1) = \begin{pmatrix} a_k & a_1 & \dots & \dots & a_{k-1} \\ \mathbf{b}_1 & a_2 & \dots & \dots & a_k \\ \mathbf{b}_2 & a_3 & \dots & \dots & \mathbf{b}_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_k & \mathbf{b}_1 & \dots & \dots & \mathbf{b}_{k-1} \end{pmatrix} \quad (21)$$

$$\mathcal{T}_1(\mathbf{T}_{i,*}; i-1) = \begin{pmatrix} a_1 & a_2 & \dots & \dots & a_k \\ a_2 & \dots & \dots & a_k & \mathbf{b}_1 \\ a_3 & \dots & \dots & \mathbf{b}_1 & \mathbf{b}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & \dots & \mathbf{b}_k \end{pmatrix} \quad (22)$$

Two points can be seen from Equation (13) and (14). 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. (13) and the (i) -th row in Eq. (14) only differ in the (i) -th element. The difference is $b_i - a_i$. Thus, these two rows can be used to compute the Elementary Effects as for the uncorrelated case in Equation (9). 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 Δ in $a_i + \Delta$ is included to transform all other elements that represent parameters $X_{\sim i}$. Importantly, to perform \mathcal{T}_2 , the mean vector \mathbf{x} and the 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 design is given by

$$d_i^{ind,T} = \frac{f(\mathcal{T}(\mathbf{T}_{i+1,*}; i)) - f(\mathcal{T}(\mathbf{T}_{i,*}; i))}{\Delta}. \quad (23)$$

Note that $\mathcal{T}(\mathbf{T}_{i+1,*}; i) = \mathcal{T}(\mathbf{T}_{i,*}; i-1)$ which is the function argument in the subtrahent in the definition of d_i^{full} in Equation (12). Therefore this transformation can be skipped for the trajectory design and the transformed trajectory in Equation (14) 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

$$\mathcal{T}_1(\mathbf{T}_{i,*}; i-1) = \begin{pmatrix} a_2 & a_3 & \dots & \dots & a_1 \\ a_3 & \dots & a_k & \mathbf{b}_1 & a_2 \\ a_4 & \dots & \mathbf{b}_1 & \mathbf{b}_2 & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_2 & \mathbf{b}_3 & \dots & \dots & \mathbf{b}_1 \end{pmatrix} \quad (24)$$

The transformation for the samples in radial design work equally except of 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. It is important to also reorder the minuend trajectories to account for the correlation hierarchy and the transformation to the sample space performed by \mathcal{T}_2 . The formulae of the Elementary Effects for the radial design are given by Equation (17) and (18).

$$d_i^{full,R} = \frac{f(\mathcal{T}(\mathbf{R}_{i+1,*}; i-1)) - f(\mathcal{T}(\mathbf{R}_{1,*}; i))}{b_i - a_i}. \quad (25)$$

$$d_i^{ind,R} = \frac{f(\mathcal{T}(\mathbf{R}_{i+1,*}; i)) - f(\mathcal{T}(\mathbf{R}_{1,*}; i))}{b_i - a_i}. \quad (26)$$

4.3 Implementation

I implemented the method by Ge and Menendez (2017) in Python 3. I wrote at least one test for each function. This is a coverage of 100 %. I am able to replicate their results for the normally distributed examples, i.e. Test Case 1 and Test Case 2 up to a reasonable degree given that they do not use a large number of samples. The replication can be found in the two notebooks at branch *replication_gm17*.

In the next section, I will present the shortcomings of the Elementary Effects in Ge and Menendez (2017). There, I will explain why it not possible to solve common tests function for uncorrelated, normally distributed parameters.

I will also present improved measures that I created. With these, I can solve all test functions for uncorrelated and correlated normalinput parameters and uncorrelated parameters in $[0,1]$. Test functions that I solved exactly are, first, the large test function with six paramters in $[0,1]$ based on the so-called g -function, presented in Saltelli et al.

(2008), page 123 - 129. The second test function is the linear function with two normally distributed input parameters and non-unit variance in Smith (2014), page 335. This function is similar to Test Case 1 and 2 and provides intuition on what the results for well-defined Elementary Effects for correlated input parameters should be. I am able to generate these results with the implementation of my measures.

4.4 Drawbacks in Ge and Menendez (2017) and Corrected Elementary Effects.

The drawback in the definition of the Elementary Effects is that the step is transformed in the numerator multiple times but not in the denominator. Therefore, these measures are not Elementary Effects in the sense of a derivation. The transformation in the numerator is performed by applying $F^{-1}(\Phi)$ to $w_i^j = a_i^j + \Delta^{(i,j)}$. The corrected and slightly renamed measures are 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. Let $Q_{k,*k-1}^T$ 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 $Q_{k,k}^T$ be the last element of the same matrix. Let F^{-1} be a transformation that maps standard normal deviates to an arbitrary sample space. Index $*k-1$ represent all elements except of the last one of a vector of length k . Index j is used to indicate an element-wise multiplication. Then

$$d_i^{c,T} = \frac{f(\mathcal{T}(\mathbf{T}_{i+1,*}; i-1)) - f(\mathcal{T}(\mathbf{T}_{i-1,*}; i))}{F^{-1}(\Phi(b_i)) - F^{-1}(\Phi(a_i))} \quad (27)$$

$$d_i^{u,T} = \frac{f(\mathcal{T}(\mathbf{T}_{i+1,*}; i)) - f(\mathcal{T}(\mathbf{T}_{i,*}; i))}{F^{-1}(Q_{k,*k-1}^T(j)T_{i+1,*k-1}^T(j) + Q_{k,k}^T\Phi(b_i)) - F^{-1}(Q_{k,*k-1}^T(j)T_{i,*k-1}^T(j) + Q_{k,k}^T\Phi(a_i))} \quad (28)$$

$$d_i^{c,R} = \frac{f(\mathcal{T}(\mathbf{R}_{i+1,*}; i-1)) - f(\mathcal{T}(\mathbf{R}_{i-1,*}; i-1))}{F^{-1}(\Phi(b_i)) - F^{-1}(\Phi(a_i))} \quad (29)$$

$$d_i^{u,R} = \frac{f(\mathcal{T}(\mathbf{R}_{i+1,*}; i)) - f(\mathcal{T}(\mathbf{R}_{i,*}; i))}{F^{-1}(Q_{k,*k-1}^T(j)R_{i+1,*k-1}^T(j) + Q_{k,k}^T\Phi(b_i)) - F^{-1}(Q_{k,*k-1}^T(j)R_{i,*k-1}^T(j) + Q_{k,k}^T\Phi(a_i))}. \quad (30)$$

In Equation (22), $Q_{k,*k-1}^T(j)R_{i,*k-1}^T(j)$ refers to the transformed trajectory of the copied

first rows, $\mathcal{T}(\mathbf{R}_{\mathbf{i}+1,*}; i)$, as described in the previous section.

The denominator of the correlated Elementary Effect, d_i^c , consists of Φ 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 $Q_{1,*}^T = (1, 0, 0, \dots, 0)$. Therefore, b_i and a_i are multiplied by one in the correlation step 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 $\Phi(b_i)$ by $Q_{k,*k-1}^T(j)T_{i+1,*k-1}^T(j)/\Phi(b_i) + Q_{k,k}^T$. The respective operation is done for $\Phi(a_i)$. 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 $X_1, \dots, X_k \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the denominator of $d_i^{u,\bullet}$ simplifies drastically to

$$\begin{aligned} & \left(\mu_i + \sigma_i \left(Q_{k,*k-1}^T(j)T_{i+1,*k-1}^T(j) + Q_{k,k}^T \Phi(b_i) \right) \right) \\ & - \left(\mu_i + \sigma_i \left(Q_{k,*k-1}^T(j)T_{i+1,*k-1}^T(j) + Q_{k,k}^T \Phi(a_i) \right) \right) \\ & = \sigma_i Q_{k,k}^T \left(\Phi(b_i) - \Phi(a_i) \right). \end{aligned} \quad (31)$$

4.5 Replication and Validation

Let $f(X_1, \dots, X_k) = \sum_{i=1}^k c_i X_i$ be an arbitrary linear function. Let $\rho_{i,j}$ be the linear correlation between X_i and X_j . Then, for all $i \in 1, \dots, k$,

$$d_i^{u,T} = d_i^{u,R} = c_i \quad (32)$$

$$d_i^{c,T} = d_i^{c,R} = \sum_{j=1}^k \rho_{i,j} c_j. \quad (33)$$

This implies that the mean and the mean absolute aggregate effects are equal to their respective (individual) uncorrelated and correlated Elementary Effect and that the respective standard deviation for both Elementary Effects is 0. This corresponds to the uncorrelated example in Saltelli et al. (2008), page 123.

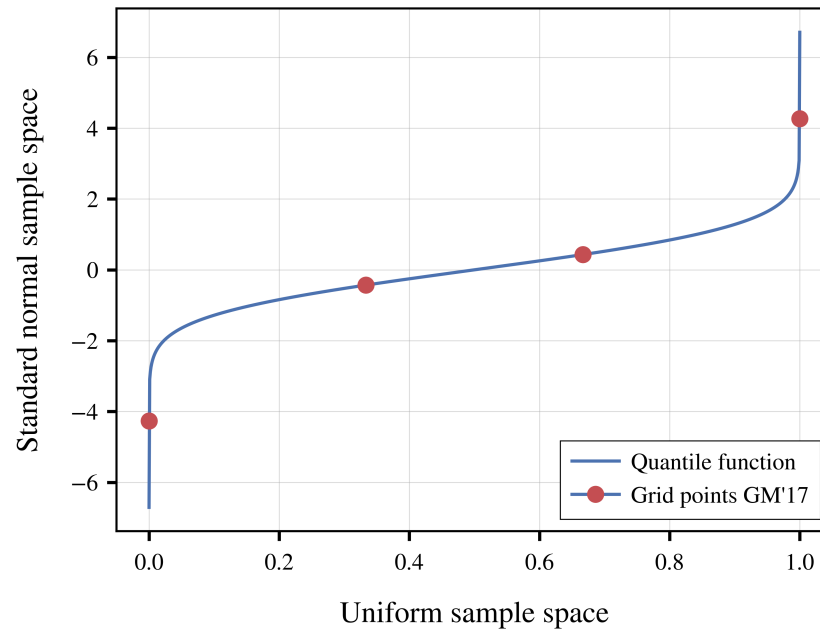
Table 2. Replication and validation for test case 1 - trajectory design

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

Table 3. Replication and validation for test case 1 - radial design

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

Figure 1. Grid points for trajectory design in Ge and Menendez (2017)



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)

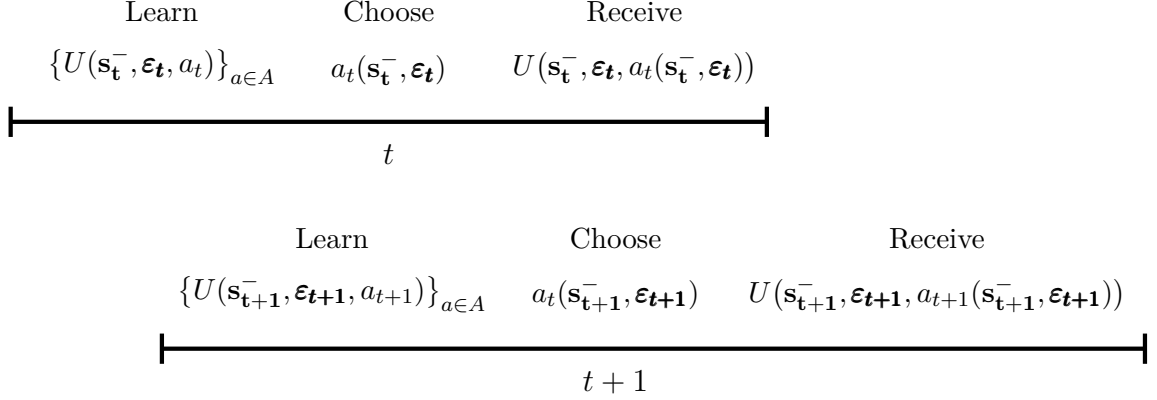
Aguirregabiria and Mira (2010) define Eckstein-Keane-Wolpin models by four characteristics. The first characteristic is that these models allow for permanent unobserved heterogeneity between agents. The simpler model by KW94 considered here does not use this option in contrast to Keane and Wolpin (1997).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

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

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

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

Figure 2. 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 (34). Like utility U , I also write value function V with different emphasis on occupation choice a_t .

$$V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t) = \max_{\{a\}_{t=0}^T} \left\{ \sum_{t=0}^T \delta^t \int_{\boldsymbol{\varepsilon}_t} U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t) f(\boldsymbol{\varepsilon}_t) d^{|A|} \boldsymbol{\varepsilon}_t \right\} \quad (34)$$

Value V depends directly on time t because T is finite. Together with the discount factor δ , this typically induces life-cycle behaviour. For example, agents invest more in the earlier time periods and work (and consume) more in the following periods. As $\{\varepsilon_{a,t}\}_{a \in A}$ are the only random parameters and serially independent, the expectation of $U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t)$ 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))²³ that breaks up the problem in (34) into more tractable sub-problems along the time dimension.

$$V(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t) = \max_{a_t} \left\{ U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t) + \delta \int_{\boldsymbol{\varepsilon}_t} \max_{a_{t+1}} V(\mathbf{s}_{t+1}^-, \boldsymbol{\varepsilon}_{t+1}, a_{t+1}) f(\boldsymbol{\varepsilon}_{t+1}) d^{|\mathcal{A}|} \boldsymbol{\varepsilon}_{t+1} \right\} \quad (35)$$

The Bellman equation shows, that solving for the whole sequence of policy functions $\{a\}_{t=0}^T$ is equivalent to solving iteratively for each optimal, period-specific policy function $a_t(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t)$. For this purpose, choose a_t for each period such that the current period utility and the discounted expected future lifetime utility (given the optimal choice of a_{t+1}) are maximized. The finite time horizon eases the problem as the value function for the last period T simplifies to $V(\mathbf{s}_T^-, \boldsymbol{\varepsilon}_T) = \max_{a_T} U(\mathbf{s}_T^-, \boldsymbol{\varepsilon}_T, a_T)$.²⁴ 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 side of (35) 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 $\{U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, a_t)\}_{a \in \mathcal{A}}$ 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 4.

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)):

$$\begin{aligned} U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, b) &= W_{b,t}^- = \exp \left\{ \beta^b + \beta_e^b x_{e,t} + \beta_b^b x_{b,t} + \beta_{bb}^b x_{b,t}^2 + \beta_w^b x_{w,t} + \beta_{ww}^b x_{w,t}^2 + \varepsilon_{b,t} \right\} \\ U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, w) &= W_{w,t}^- = \exp \left\{ \beta^w + \beta_e^w x_{e,t} + \beta_w^w x_{w,t} + \beta_{ww}^w x_{w,t}^2 + \beta_b^w x_{b,t} + \beta_{bb}^w x_{b,t}^2 + \varepsilon_{w,t} \right\} \end{aligned} \quad (36)$$

²³For more details, see Raabe (2019), p. 9-19.

²⁴More 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 finite 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.

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.²⁵ These coefficients are called covariates by many structural economists.

The utilities for education, or schooling, and staying at home are given by the functions in (37). These functions are also called non-pecuniary rewards.

$$\begin{aligned} U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, e) &= \beta^e + \beta_{he}^e \mathbf{1}(x_{e,t} \geq 12) + \beta_{re}^e (1 - \mathbf{1}(a_{t-1} = e)) + \varepsilon_{e,t} \\ U(\mathbf{s}_t^-, \boldsymbol{\varepsilon}_t, h) &= \beta^h + \varepsilon_{h,t} \end{aligned} \quad (37)$$

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

Write $\{\varepsilon_{a,t}\}_{a \in A}$ as vector $\boldsymbol{\varepsilon}_{a,t}$. It is assumed that $\boldsymbol{\varepsilon}_{a,t}$ follows a joint normal distribution, such that $\boldsymbol{\varepsilon}_{a,t} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$. $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ denotes the covariance matrix for shocks $\boldsymbol{\varepsilon}_{a,t}$. σ_a^2 and $\sigma_{a(j),a(k \neq j)}^2$ denote the alternative-specific variances and covariances in $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. 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 $\text{SE}(\hat{\boldsymbol{\theta}})$. It also shows the correlations between important estimates.

The second column in Table 4 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 KW94.²⁶ 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,

²⁵The 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.

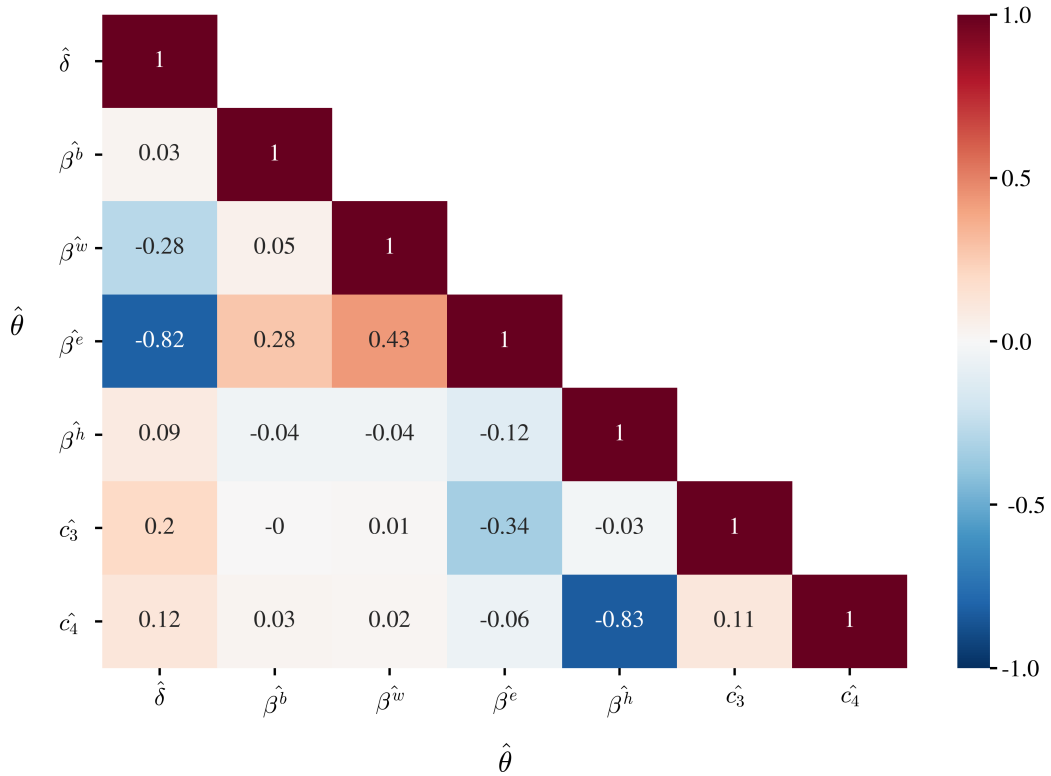
²⁶See table 1, p. 658 in Keane and Wolpin (1994); In contrary to the computation of variation measures for $\hat{\boldsymbol{\theta}}$, it is sufficient for obtaining $\hat{\boldsymbol{\theta}}$ to find the individual likelihood of the average agent instead of the sample likelihood in (46).

the diagonal elements of the lower triangular matrix c_i coincide with the standard deviations of the utility shocks $\varepsilon_{a,t}$ and the non-diagonal elements $c_{i,j}$ equal the correlations between different alternative-specific shocks $\varepsilon_{a,t}$. The parameter estimates $\hat{\theta}$ are precise. This means they equal the parameters with which the model is simulated.

The third column shows this thesis' estimates of the standard errors $\text{SE}(\hat{\theta})$. The fourth column shows the standard errors computed in KW94. Given the differences between both estimation specifications, namely the inclusion of β and correlations between standard errors in this thesis, the estimates are reasonably similar. However, the one exception that stands out is the results for the non-diagonal Choleksy factors $c_{i,j}$.

Figure 3 depicts the correlations between the estimates of important parameters in θ . 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.

Figure 3. Correlations between estimates for important input parameters



The intuition behind these results can be obtained from the following insight: Negative correlations imply similar effects, and positive correlations imply opposing effects on the likelihood of observed endogenous variables \mathcal{D} . For instance, consider an individual that decides for a long occupation in the education sector in the first years and then continues to work in the white-collar sector for the rest of his life. The likelihood to observe this individual increases when δ rises because all individuals get more patient, and therefore,

ceteris paribus, they invest more in education. However, the same likelihood also increases if the educational utility constant β^e rises. Hence, because they can compensate each other, the likelihood around the optimal parameter $\hat{\theta}$ decreases less for changes of both parameters in opposing directions than for changes in the same direction. Therefore, parameters δ and β^e are negatively correlated in terms of the score function in (47) around $\hat{\theta}$. It follows from (49) that their standard errors are negatively correlated.

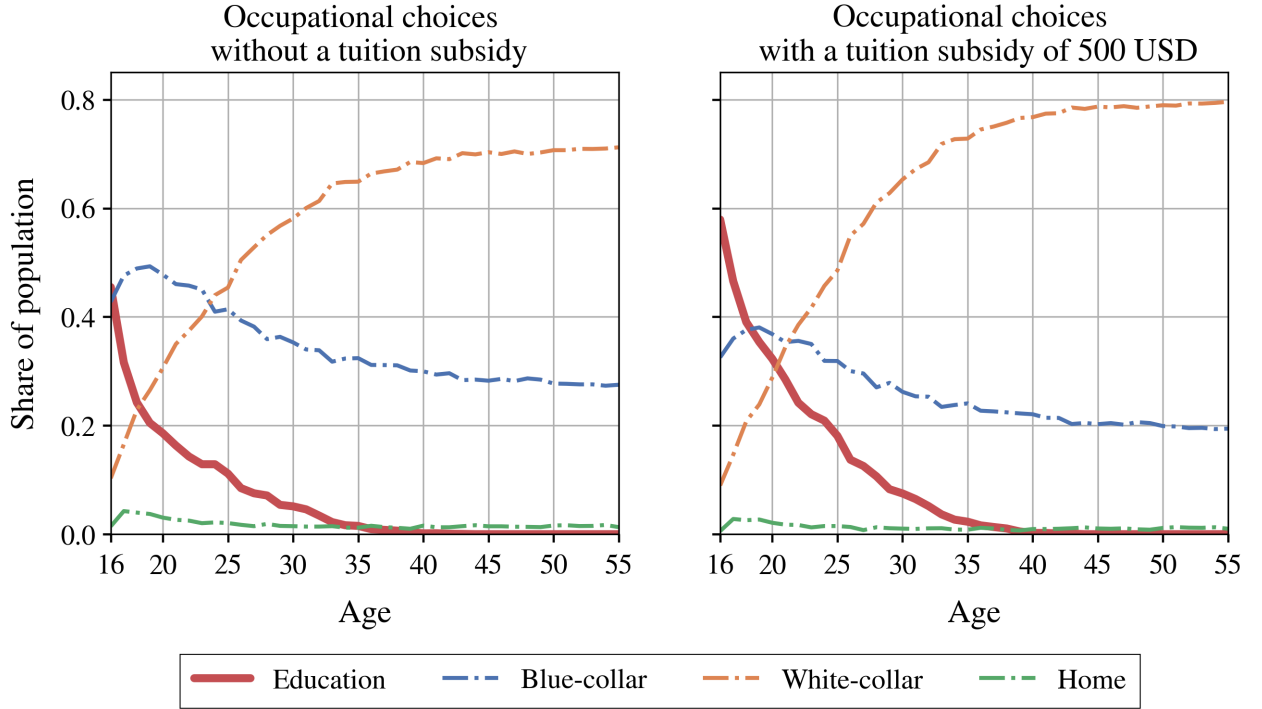
5.3 Quantity of Interest

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.²⁷ The same figure computed with *respy* is 1.5.

Figure 6 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.

²⁷See table 4, p. 668 in Keane and Wolpin (1994).

Figure 4. Comparison of shares of occupation decision over time between scenarios

The QoI is the sum of the differences between the education shares at each age in the right and in the left graph as depicted by the red lines.²⁸ This is because the vertical axis can also be interpreted as the share of one year that the average agent is occupied in the education sector. Comparing both graphs, we can see that the tuition subsidy incentivises younger individuals to stay in the education sector for a longer time and older individuals to work in the white-collar sector. The latter observation is a consequence of the first because the white-collar sector rewards education.

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

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

²⁸If the red lines would depict continuous functions instead of discrete points in time, the QoI would be the difference between the integrals of the education shares as a function of time in the policy and the base scenario.

Table 4. Estimates for the distribution of input parameters

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

6 Uncertainty Analysis

Figure 5. Probability distribution of quantity of interest q

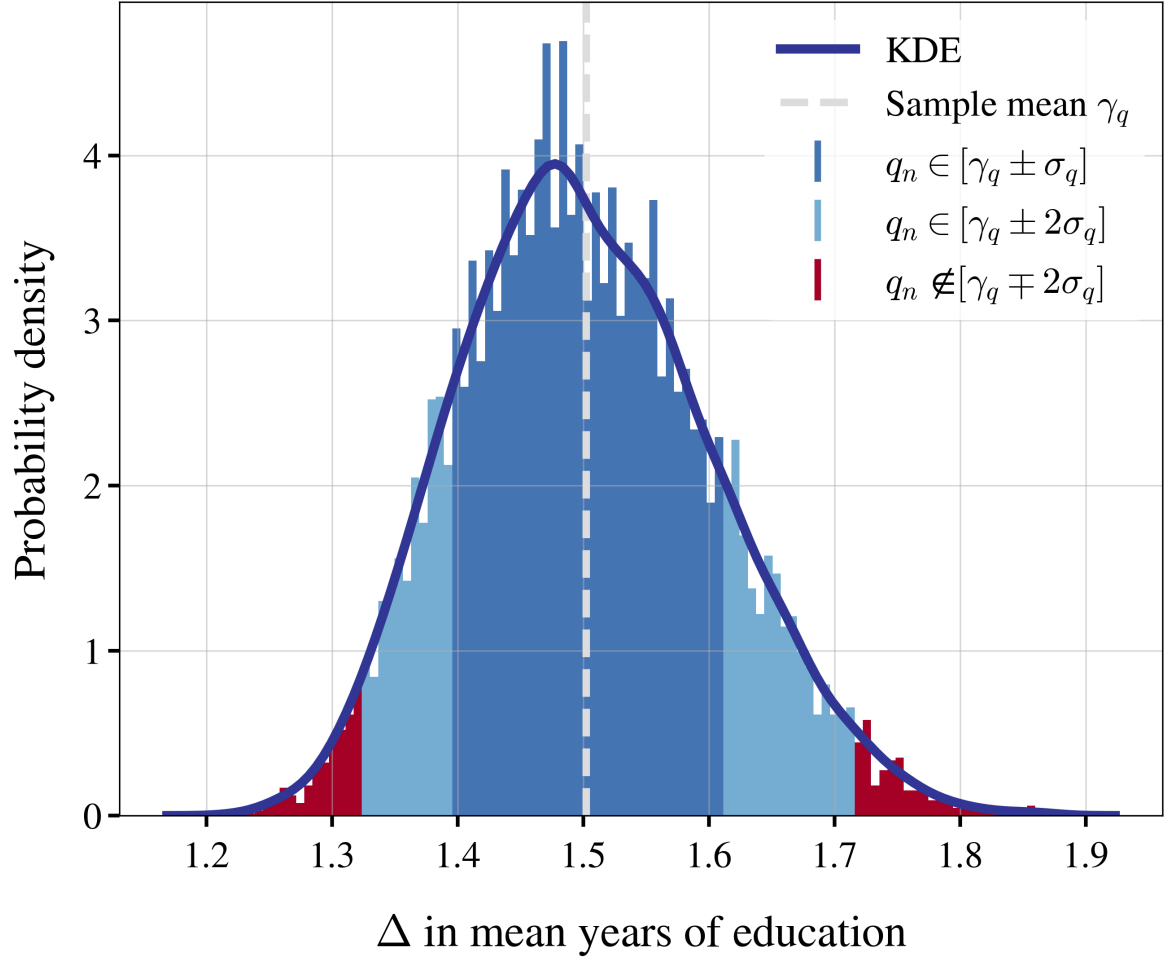
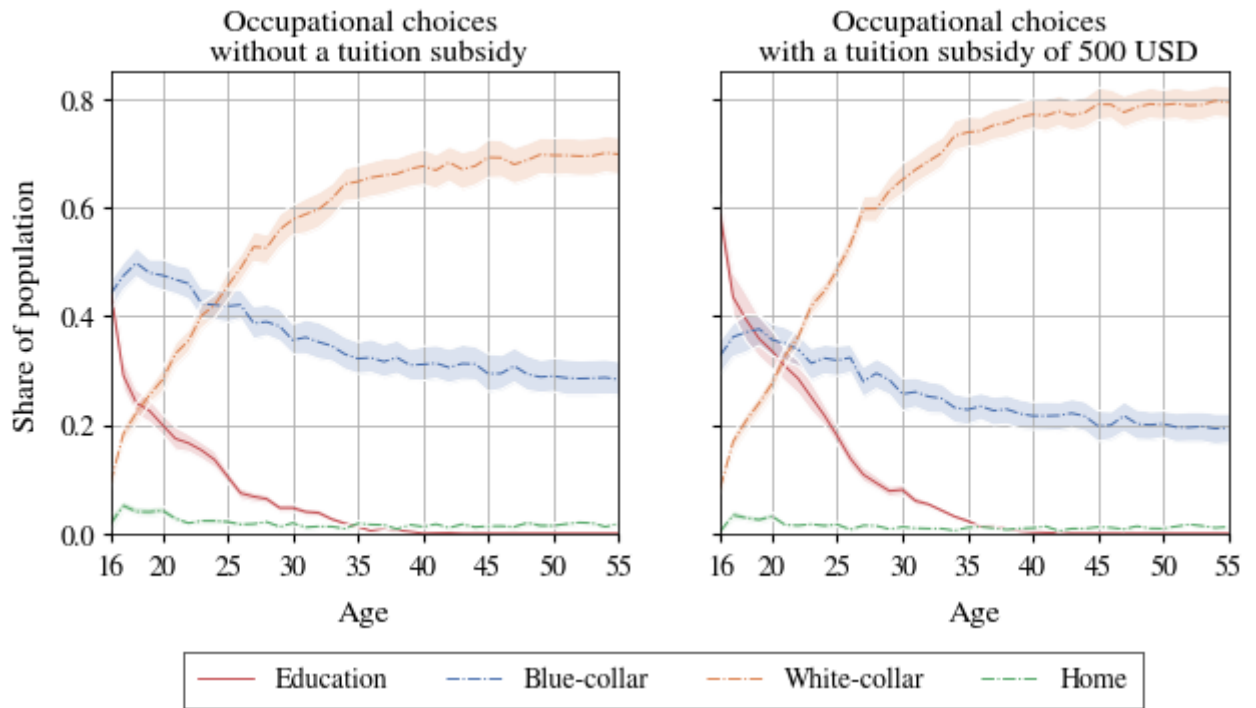


Figure 6. Comparison of shares of occupation decision over time between scenarios with cone plots



7 Qualitative Global Sensitivity Analysis

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

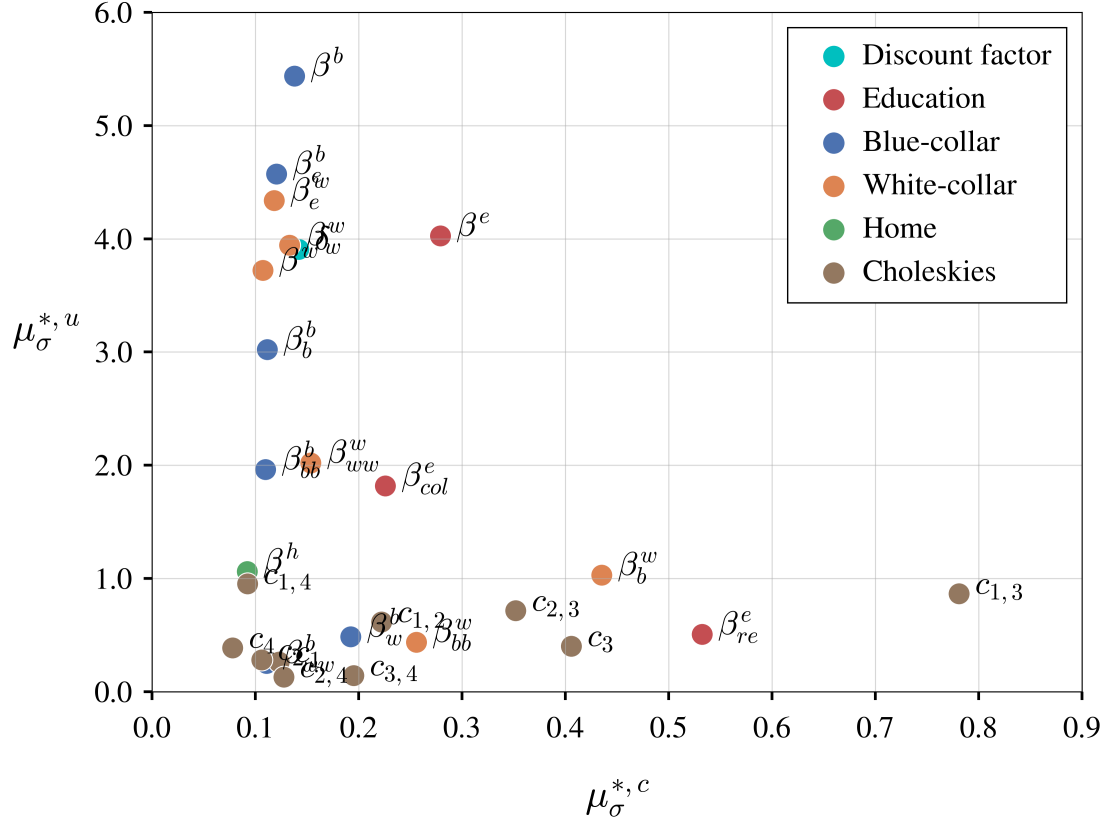
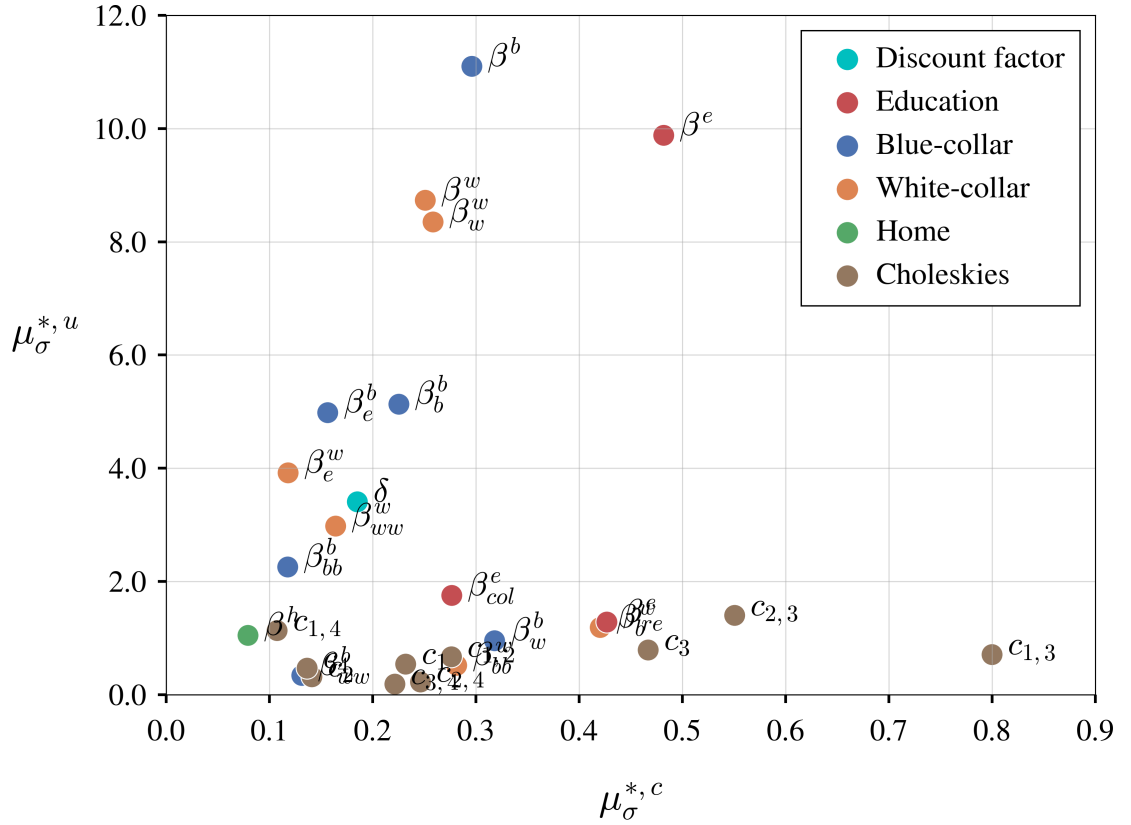


Figure 8. Sigma-normalized mean absolute Elementary Effects for radial design



8 Discussion

none

9 Conclusion

none

[Go over (especially capitalization 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., 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.
- 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.
- 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. Billionis (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.
- 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.

References

- 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.

Appendix A: Correlated standard normal deviates

Appendix B: Inverse transformations of probability distributions

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))²⁹.

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, \dots, 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 $\mathcal{D} \stackrel{\text{def}}{=} (\mathbf{s}_t^-, \mathcal{W}_{a,t}^-)$. Given this setup, the goal is to estimate the exogenous model parameters $\boldsymbol{\theta} \stackrel{\text{def}}{=} (\delta, \boldsymbol{\beta}, \boldsymbol{\Sigma}_\epsilon)$.³⁰ Thus, in the following, every probability is a function of the exogenous model parameters $\boldsymbol{\theta}$. The approach to compute the likelihood function $L_{\mathcal{D}}(\boldsymbol{\theta})$ of the observables in the data begins with the individual latent variable representation in period t .

$$a_t = \underset{a}{\operatorname{argmax}} V(\mathbf{s}_t^-, \boldsymbol{\epsilon}_t, a_t) \quad (38)$$

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

$$\boldsymbol{\epsilon}_t(a_t(j), \mathbf{s}_t^-) \stackrel{\text{def}}{=} \{\boldsymbol{\epsilon}_t | V(\mathbf{s}_t^-, \boldsymbol{\epsilon}_t, a_t(j)) = \max_{a_t} V(\mathbf{s}_t^-, \boldsymbol{\epsilon}_t, a_t)\}. \quad (39)$$

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 $\boldsymbol{\epsilon}_t(a_t(j), \mathbf{s}_t^-)$. This probability equals the integral of the probability distribution function $f(\boldsymbol{\epsilon}_t)$ over all elements of set $\boldsymbol{\epsilon}_t(a_t(j), \mathbf{s}_t^-)$ with respect to $\boldsymbol{\epsilon}_t$. Formally,

$$\mathbf{p}(a_t(j) | \mathbf{s}_t^-) = \int_{\boldsymbol{\epsilon}_t(a_t(j), \mathbf{s}_t^-)} f(\boldsymbol{\epsilon}_t) d^{|\mathcal{A}|} \boldsymbol{\epsilon}_t. \quad (40)$$

The second case is $a_t(k) \in [b, w]$. Assuming the dataset contains wages for the working

²⁹See Aguirregabiria and Mira (2010), p. 42-44 and Raabe (2019), p. 21-26 for more details.

³⁰Improvements in this thesis' estimation over KW94 are that, first, it is not assumed that the standard errors of the parameters estimates are uncorrelated, and, second, that β is not left out of the estimation.

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 alternative-specific shocks $\boldsymbol{\varepsilon}_{\mathbf{a},t}$ are log normally distributed. Second, in contrary to the non-working alternatives, using (36), the shocks can directly be expressed as a function of the alternative-specific model parameters $\boldsymbol{\beta}_{\mathbf{a}(k)}$ by inserting the inferred alternative-specific experiences $\mathbf{x}_{\mathbf{a}(k),t}$ into $W_{a(k),t}$ and subtracting the expression from the observed wage $\mathcal{W}_{a(k),t}^-$ for each individual. Both wages are logarithmized. Thus,

$$\varepsilon_{a(k),t} = \ln(\mathcal{W}_{a(k),t}^-) - \ln(W_{a(k),t}^-). \quad (41)$$

Third, the alternative-specific shocks $\boldsymbol{\varepsilon}_{\mathbf{a},t}$ 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

$$p(a_t(k) | \mathbf{s}_t^-, W_{a(k),t}^-) = \int_{\boldsymbol{\varepsilon}_t(a_t(k), \mathbf{s}_t^-)} f(\boldsymbol{\varepsilon}_t | \varepsilon_{a(k),t}) d^{|\mathbf{A}|} \boldsymbol{\varepsilon}_t. \quad (42)$$

Applying integration by substitution yields the following expression for the probability of the observed wage:³¹

$$p(\mathcal{W}_{a(k),t}^- | \mathbf{s}_t^-) = \omega_t^{-1} \frac{1}{\sigma_{a(k)}} \phi\left(\frac{\varepsilon_{a(k),t}}{\sigma_{a(k)}}\right). \quad (43)$$

Here, ω_t^{-1} is the Jacobian of the transformation from observed wage $\mathcal{W}_{a(k),t}^-$ to error $\varepsilon_{a(k),t}$ in (41) and ϕ is the standard normal probability density function. Finally, the joint probability of observing choice $a_t(k)$ and wage $\mathcal{W}_{a(k),t}^-$ conditional on the observed states is given by the product of the two probabilities in (42) and (43):

$$p(a_t(k), \mathcal{W}_{a(k),t}^- | \mathbf{s}_t^-) = p(a_t(k) | \mathbf{s}_t^-, \mathcal{W}_{a(k),t}^-) p(\mathcal{W}_{a(k),t}^- | \mathbf{s}_t^-) \quad (44)$$

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:

$$L_{\mathcal{D}}^i(\boldsymbol{\theta}) = P(\{a_t^i, \mathcal{W}_{a,t}^{-,i}\}_{t=0}^T) = \prod_{t=0}^T p(a_t^i, \mathcal{W}_{a,t}^{-,i} | \mathbf{s}_t^{-,i}) \quad (45)$$

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

³¹See Raabe (2019), p. 29, 39-40 for the complete derivation.

the whole sample of individuals:

$$L_{\mathcal{D}}(\boldsymbol{\theta}) = P\left(\left\{\{a_{t,i}^i, \mathcal{W}_{a,t,i}^{-,i}\}_{t=0}^T\right\}_{i \in I}\right) = \prod_{i \in I} \prod_{t=0}^T p(a_{t,i}^i, \mathcal{W}_{a,t,i}^{-,i} | \mathbf{s}_t^{-,i}) \quad (46)$$

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 (46). As maximum likelihoods estimates are asymptotically normal³², these results are taken as the mean vector for the input parameters in the uncertainty quantification.

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

Finally, the calculation of the estimator's covariance is described.³³ 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: $\text{Var}(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})^{-1}$. In this thesis, the information matrix $\mathcal{I}(\boldsymbol{\theta})$ is given by the variance of the scores of the parameters.³⁴ The scores $\mathbf{s}(\boldsymbol{\theta})$ 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

$$\mathbf{s}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\partial L_{\mathcal{D}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i \in I} \frac{\partial L_{\mathcal{D}}^i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \sum_{i \in I} \mathbf{s}_i(\boldsymbol{\theta}). \quad (47)$$

Having multiple individual likelihood contributions, the scores are in the form of the Jacobian matrix. Using the property that the expected values of scores, $\mathbb{E}[\mathbf{s}(\boldsymbol{\theta})]$, are zero at the maximum likelihood estimator, the variance of the scores is given by (48). It is equal to the inverse of the Fisher information matrix.

$$\mathcal{I}^{-1}(\boldsymbol{\theta}) = \text{Var}(\mathbf{s}(\boldsymbol{\theta})) = \mathbb{E}[\mathbf{s}(\boldsymbol{\theta})\mathbf{s}(\boldsymbol{\theta})']. \quad (48)$$

Hence, the estimator for the asymptotic covariance of the maximum likelihood estimator

³²This 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.

³³See Verbeek (2012), p. 184-186.

³⁴The 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.

is given by

$$\hat{\text{Cov}}_J(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{|I|} \sum_{i \in I} \mathbf{s}_i(\hat{\boldsymbol{\theta}}) \mathbf{s}_i(\hat{\boldsymbol{\theta}})' \right)^{-1}. \quad (49)$$

$|I|$ 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 (47) and (48) then imply that the inverse of the variance of the individual scores is equivalent to the variance of the maximum likelihood estimator.

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