Quantile-based Sensitivity Analysis on Structural Behavioral Models

Master Thesis Presented to the

Department of Economics at the

Rheinische Friedrich-Wilhelms-Universität Bonn

in Partial Fulfillment of the Requirements for the Degree of

Master of Science (M.Sc.)

Supervisor: Prof. Dr. Philipp Eisenhauer

Submitted in September 2021 by:
Yulei Li
Matriculation Number: 3209294

Contents

Abbreviations

Term	Meaning
CDF	Cumulative distribution function
DCDP	Discrete choice dynamic programming
EKW	Eckstein-Keane-Wolpin
GSA	Global sensitivity analysis
LSA	Local sensitivity analysis
PDF	Probability distribution function
QBSM	Quantile-based sensitivity measures
QoI	Quantity of interest
USD	US-Dollar

List of Figures

List of Tables

1 Introduction

Structural econometric models have been widely used to provide decision support and guidelines for policymakers. The Discrete Choice Dynamic Programming (DCDP) model, a structural model commonly used in the field of labor economics, plays a key role in evaluating the effects of various policy interventions. In particular, DCDP models study schooling and occupational choice and investigate the effects of policy interventions designed to boost human capital investment, such as parental and government subsidies and student loan programs (Keane & Wolpin, 1997; Sauer, 2004).

A typical structural model of optimal individual behavior consists of a number of structural parameters that capture the underlying preferences and constraints of the individual's decision process (Blundell, 2017). Such models require individuals to solve a sequential decision problem under given institutional constraints that are determined by a set of variables determined by policymakers (e.g., tax rates). So far, most structural models have been deterministic, with each model parameter fixed with a single numerical value or point estimate, ignoring their associated errors or uncertainties. (Adda, Dustmann, & Stevens, 2017; Attanasio, Meghir, & Santiago, 2012; Eckstein, Keane, & Lifshitz, 2019). In addition to using point estimates, policymakers can obtain more information about prediction errors by including uncertainty in the model parameters. For example, using confidence sets as set estimates (Manski, 2021). More recently, noticing the considerable parametric uncertainty in prediction, Eisenhauer, Gabler, and Janys (2021) suggested the use of global sensitivity analysis to identify which parameters are particularly responsive to prediction uncertainty.

Sensitivity analysis(SA) has been a well-established approach in a wide variety of scientific disciplines, such as biology (Zi, 2011), chemistry (Saltelli, Ratto, Tarantola, & Campolongo, 2005), environmental sciences (Campolongo & Saltelli, 1997) and so on. However, economists have had a very restricted contribution to the literature on sensitivity analysis. Although the importance of sensitivity analysis for quantitative models has been well acknowledged by some economists, very few studies have included any form of sensitivity analysis when building economic models, or if they do, they often rely on local techniques, in which case the sensitivity measures are only computed around a fixed point in the parameter space (Canova, 1994; Harenberg, Marelli, Sudret, & Winschel, 2019; Leamer, 1985). It is notable that SA in economics research is distinguished from other areas such as environmental science by the conditional distribution of the input parameters. As a result, using a one-factor-at-a-time Local SA without accounting for any parameter correlations would yield misleading results. Moreover, Local methods have the glaring drawback that they are invalid for nonlinear and nonmonotonic models (Saltelli, 2004), which are frequently used in econometric modeling. In contrast, global methods with a "model-free" setting are strongly recommended in systematic reviews (Saltelli et al., 2019). Unfortunately, a generally accepted practice of GSA has not yet been established in most of the existing economic literature.

I rely on quantile-based sensitivity measures(QBSM) (Kucherenko, Song, & Wang, 2019) to analyze the impact of the parametric uncertainty on prediction uncertainty. Such measures are based on quantiles of the model output, which allow us to identify the parametric uncertainty at quantile level. The quantiles of output Cumulative distribution(CDF) are also of interest in other domains such as reliability analysis and financial risk analysis. For such problems, conventional method such as variance-based sensitivity indices (Sobol, 1993) are in practical. Moreover, variance-based sensitivity measures are not able to identify variables that are most essential in achieving the extreme values of the model output. Alternatively, QBSM are designed for global sensitivity analysis of problems where the α-th quantile is a function of interest, and also for instances where the analyst is interested in the ranking of the inputs that cause the extreme values of the output. The quantile oriented sensitivity indices (Kala, 2019) share some similarities with QBSM, but these indices are based on contrast function instead of explicitly depending on quantiles.

As an application, I reanalyze the human capital investment decision model by Keane and Wolpin (1994). I estimate the model parameters and replicated the key results using the provided dataset. I generate counterfactual predictions based on the estimated parameters. I investigate the transition of uncertainty from key parameters to model output. I calculate the QBSM for a few parameters to identify the parameters that contribute most to the prediction uncertainty. I extend their framework to explain which parameter specifically contributes to the uncertainty.

This work complements the work by Eisenhauer et al. (2021), who documented significant prediction uncertainty in a Eckstein-Keane-Wolpin(EKW) model (Aguirregabiria & Mira, 2010). My work builds on their work by identifying which parameters account for most of the output uncertainty based on the quaniles of output. In their study, a 90% confidence set is used to represent the counterfactual uncertainty. As a result, the analyst is primarily concerned with the region of the output values range inside the confidence set. In other words, the critical quantiles of the cumulative distribution function (CDF) of the function are of interest, which fits well with the target problem that QBSM aim to addresses. Although including sensitivity analysis in model calibration has long been a standard technique in environmental and engineering modeling, there is a lack of application in econometric modelling.

This paper is organized as follows: Section 2 presents the model setting of Keane and Wolpin (1994). Section 3 gives an introduction to variance-based global sensitivity indices and QGSM. Algorithms and methods for numerical estimation of QGSM are discussed in Section 3. The results are presented in Section 5. Finally, conclusions are summarized in Section 6.

2 The Structural Behavioral Model

The SA approach is applied to the DCDP model of the human capital investment decisions presented in Keane and Wolpin (1994) which is briefly described in subsection 2.1. The uncertain input parameters of the model are described in subsection 2.2 and the quantity of interest is discussed in subsection 2.3.

2.1 Keane and Wolpin (1994) model

Structural models have been developed to understand individual behavior and inform policy design in labor economics, particularly for DCDP models. DCDP models, as a framework for policy evaluation, have been applied to study a range of sequential individual decisions. A notable example is a work by Keane and Wolpin (1994), who developed an approximate solution to a large-scale DCDP model of human capital occupational choice. This is an ideal demonstration of how a structural model can evaluate various policies and determine the optimal policy without running many treatments.

We assume that individuals are forward looking agents who make their occupational choice based on the principle of expected utility maximization. From age 15 to the age 55, the total period of decision is $T = 40$ years. As shown in Figure 1, in each period t, individual makes choice among $K = 4$ alternatives: occupation one($a_t = 1$), occupation two($a_t = 2$), schooling($a_t = 3$) and staying at home($a_t = 4$). We denote the agent's immediate rewards of choosing each alternative in a given period as:

$$r_t\left(s_t, a_t\right)=\begin{cases} w_{1t}=\exp\left\{\alpha_{10}+\alpha_{11} g_t+\alpha_{12} e_{1t}+\alpha_{13} e_{1t}^2+\alpha_{14} e_{2t}+\alpha_{15} e_{2t}^2+\epsilon_{1t}\right\} & \text{if } a_t=1 \\ w_{2t}=\exp\left\{\alpha_{20}+\alpha_{21} g_t+\alpha_{22} e_{1t}+\alpha_{23} e_{1t}^2+\alpha_{24} e_{2t}+\alpha_{25} e_{2t}^2+\epsilon_{2t}\right\} & \text{if } a_t=2 \\ \beta_0-\beta_1 \mathbb{I}\left[g_t \geq 12\right]-\beta_2 \mathbb{I}\left[a_{t-1} \neq 3\right]+\epsilon_{3t} & \text{if } a_t=3 \\ \gamma_0+\epsilon_{4t} & \text{if } a_t=4 \end{cases} \tag{1}$$

w_{1t} and w_{2t} is occupational-specific wage if individual choose either occupation one or two. α_1 and α_1 are parameters of corresponding wage functions. The primary factors that associated to wages are: years of completed schooling(g_t), years of working experience in two occupations respectively(e_{1t}, e_{2t}).If the individual does not choose to work($a_t = 3$ or $a_4 = 4$), although they are not paid during this period, they still receive rewards from their choices. The variables that influence the choice of schooling are: the consumption value of schooling(β_0), the tuition of post-secondary education(β_1), and the adjustment expenses involved with returning to school. The means return of staying at home that affects the home option is represented by γ_0.

The state space t can be represented by

$$\left\{g_t, e_{1t}, e_{2t}, a_{t-1}, \epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, \epsilon_{4t}\right\} \tag{2}$$

The part of state space observable to both individuals and researchers are $\{g_t, e_{1t}, e_{2t}, a_{t-1}\}$. The evolution of those variables follows a Markov chain, which means that the next state at period $t+1$ relies on the current state t and the decision maker's action:

$$\begin{aligned} e_{1,t+1} &= e_{1t}+\mathbb{I}\left[a_t=1\right] \\ e_{2,t+1} &= e_{2t}+\mathbb{I}\left[a_t=2\right] \\ g_{t+1} &= g_t+\mathbb{I}\left[a_t=3\right] \end{aligned} \tag{3}$$

The shocks$\{\epsilon_{1t}, \epsilon_{2t}, , \epsilon_{3t}, \epsilon_{4t}\}$ are serially independent and observed by individuals but not researchers.They are assumed to be jointly normally distributed.

The maximized value of the expected remaining lifetime utility at t is defined as:

$$V(S(t), t)=\max_{\left\{d_k(t)\right\}_{k \in K}} E\left[\sum_{\tau=t}^T \delta^{\tau-t} \sum_{k=1}^K R_k(\tau) d_k(\tau) \mid S(t)\right] \tag{4}$$

with $0<\beta<1$ the discount factor. $S(t)$ are state variables that contain information available before decision in period t is made. R_k is the rewards from choose alternative k at period t. d_k is individual's decision from among K=4 discrete alternatives.

Alternatively, this optimization problem can be expressed as a the Bellman equation, which is a necessary condition for optimality associated with dynamic programming(Bellman, 1966; Eisenhauer, 2019):

$$E \max(S(t+1))=E\left[V(S(t+1), t+1) \mid S(t), d_k(t)=1\right] \tag{5}$$

2.2 Model parameters

For model parametrization, our work takes advantage of the open-source Python package respy(Janos Gabler & Tobias Raabe, 2020), which has greatly simplified the process of simulation and estimation of finite-horizon

DCDP models. Table 1 shows the parameter values obtained from the simulation based on Data Set One in Keane and Wolpin (1994). According to this setting, occupation two requires greater expertise and thus is more beneficial from education. Experience in occupation one also increases the likelihood of increased earnings in another occupation. Thus, a general skill that is useful for both occupations can be obtained through either schooling or work experience, while occupation-specific skills are exclusively obtained through work. In this sense, we will refer to occupation two as white-collar and occupation one as blue-collar henceforth.

Once parameterization is completed, we are allowed to analyze the economic implications behind it. Firstly, we investigate the correlation between selective model parameters 3. In general, a large share correlation are prevalent, which emphasizes the importance of using GSA rather than LSA for this model. Because GSA takes into account the interaction between parameters. It is also worth noting that $\hat{\delta}$ and $\hat{\beta}_0$ and beta show a relatively strong negative correlation.

2.3 Quantity of Interest(QoI)

This paper measures the impact of a 500 USD tuition subsidy on the average number of years a individual spends on pursuing a higher education degree. Figure 2 shows the probability distribution of this quantity of interest. Our simulation results with respy(Janos Gabler & Tobias Raabe, 2020) showed a 1.55-year increase in average school years, which is slightly higher than the effect documented by Keane and Wolpin (1994), that is an increase of 1.44 years. More specifically, Figure 3 shows a comparison of the occupation shares over their relevant life cycle for a sample of 1000 individuals in different sectors. The left panel shows a baseline scenario that without any policy intervention, whereas the right panel shows a counterfactual scenario that all the 500 USD tuition subsidy is implemented. The yellow, green, dark blue and light blue bars represent the percentages of people stay-at-home, pursuing education, blue-collar, white-collar, respectively. With this setting, we are able to access the policy effect without running experiments at real world.

We noticed the change that, due to the fact that the white-collar industry values education, many agents begin their schooling early and continue to work in the white-collar sector. The positive impact and accumulated blue-collar experience pays off in the white-collar sector as well, leading to this shift from white- to blue-collar. It is nearly on impact due to the overall low level of participation in the household sector. The difference between the education shares for each age group in the right and left graphs, as represented by the red line, is quantified investment. This is because the vertical axis can be thought of as the percentage of time a typical agent spends in the education sector over the course of a year. When we look at both graphs, we can clearly see that tuition subsidies encourage younger people to stay in school longer and older people to pursue white collar works.

In section 3 and section for 4 I will present how uncertain in model input translate to the variations of model output, in this case, the average change of school years. However, due to the computational burden in the sampling procedure, I restrict my attention to three main parameters: return to an additional year of schoolingα_{11}, reward for going to collegeβ_1 and reward of non-market alternativesγ_0. As what we study is the impact of subsidy, the profit/cost trade-offs involved with the decision whether to pursue continuing education are fully reflected by the return on educational investment. Although γ_0 seems not explicitly related to our QoI like the previous chosen ones, we include is as a comparison.

3 Sensitivity Analysis for the Structural Behavioral Model

In this chapter, we introduce the basic concepts, terminologies, methods of GSA in Section 3.1, including the general model structure and the input factors. Then we demonstrate the mathematical descriptions of the variance-based and quantile-based sensitivity measures in Section 3.2 and Section 3.3.

3.1 Global sensitivity analysis: the framework

The sensitivity analysis is "The study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input"(Saltelli, 2004). Thus, the definition of sensitivity analysis includes models, model inputs and model outputs. Throughout this work, a structural economic model will be regarded as a general function that defines a relationship between inputs and output(s):

$$\mathcal{M} : \theta \mapsto y = \mathcal{M}(\theta) \tag{6}$$

where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ is a vector of model parameters. The output of interest is a vector y. To produce a counterfactual prediction, a policy $g \in \mathcal{G}$ changes the mapping to $\mathcal{M}_g(\theta)$(Eisenhauer et al., 2021). Consequently, the differences between the prediction before and after the policy intervention yield a structural estimate of policy effects.

Unlike LSA which analyzes how a tiny change near an input space value affects the scalar output, GSA identifies such effects in the whole input space. To be more precise, each input parameter is treated as a random variable assigned with a distribution of all possible values. Consequently, the uncertainty coming from model parameters is transmitted through the model to generate an empirical distribution of the output of interest $Y = M(\Theta)$ with a joint probability density function $f_Y(y)$. So far, this is how uncertainty propagation occurs.

Once uncertainty propagation characterizes the uncertainty of the model output, a sensitivity analysis then can be applied to identify which parameters are primarily responsible for the variability of output.

To date, a wide range of GSA methods has been developed. In this paper, we focus on variance-based sensitivity measures(Sobol, 1993) and quantile-based sensitivity measures(Kucherenko et al., 2019). A comprehensive discussion and comparison of the GSA methods can be found in Razavi et al. (2021).

3.2 Variance-based sensitivity measures

The model function $Y = f(\theta_1, \ldots, \theta_d)$ is defined in d-dimensional real coordinate space R^d with an input vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$.Note that θ is a random variable with a continuous probability distribution function(PDF). To quantify the effect of variations of θ on the variation of Y, let us consider the conditional expectation $\mathrm{E}[Y \mid \Theta_i = \theta_i]$. It is the mean value of output Y over the probability distribution of $\Theta_k (k \neq i)$ with the condition that Θ_i is fixed to θ_i. If we consider the variations of Θ_i, the associated random variable is $\mathrm{E}[Y \mid \Theta_i]$, whose variance quantifies the effect of Θ_i on the variation of Y.

According to the result by Sobol (1993), given k input variables mutually independent, the variance of the output can be decompose as the sum of variances of increasing order:

$$\operatorname{Var}(Y) = \sum_i V_i + \sum_{ij} V_{ij} + \sum_{ijh} V_{ijk} + \cdots + V_{1,2}, k \tag{7}$$

where $V_i = \mathrm{Var}(E(Y \mid \Theta_i))$ is first order variance and $V_{ij} = \mathrm{Var}(E(Y \mid \Theta_i, \Theta_j))$ is second order variance, etc. Note that in an additive model, this decomposition includes only first order variances.

Thus, *first-order indice*s of input parameter Θ_i is defined as:

$$S_i = \frac{V_i}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}(E(Y \mid \Theta_i))}{\mathrm{Var}(Y)} \tag{8}$$

Accordingly, *second-order indices* of input parameter Θ_i and Θ_j is defined as:

$$S_{ij} = \frac{V_{ij}}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}(E(Y \mid \Theta_i, \Theta_j))}{\mathrm{Var}(Y)} \tag{9}$$

First-order indices measures the proportion of the total variance which is due to the main effect of Θ_i on Y. Whereas, second-order index measures the proportion of the total variance which is explained by the interaction between the two inputs.

In 1996, Homma and Saltelli (1996) introduced the *total variance index* which measures the proportion of the total variance due to the main effect of Θ, and all its interactions with the other inputs:

$$S_{Ti} = \frac{\sum_i V_i + \sum_{jh} V_{ijh} + \cdots}{\mathrm{Var}(Y)} = \frac{E(Var(Y \mid \Theta_{\sim i}))}{Var(Y)} = 1 - \frac{Var(E(Y \mid \Theta_{\sim i})}{Var(Y)} \tag{10}$$

where $\Theta_{\sim i} = (\Theta_1, \Theta_2, \ldots, \Theta_{i-1}, \Theta_{i+1}, \ldots, \Theta_k)$

There are three features of total variance index to be aware of. Firstly, The condition $S_{Ti} = 0$ is necessary and sufficient for Θ_i to be a non-influential input(it can be treated as a fixed input). Secondly, if $S_{Ti} \approx S_i$ the interaction between Θ_i and the other inputs does not affect the variability of the output. Lastly, it is obvious that the sum of the total indexes is in general greater than Homma and Saltelli (1996).

3.3 Quantile-based sensitivity measures

Now we consider the scenarios where only a specific range of output is important to analyst. For instance, $Y = M((\Theta) \leq a)$ or $Y = M((\Theta) \geq b$. We reformulate such problems to α-th quantile of the output CDF $q_Y(\alpha)$:

$$\alpha = \int_{-\infty}^{q_Y(\alpha)} \rho_Y(y)dy = P\left\{Y \leq q_Y(\alpha)\right\} \tag{11}$$

alternatively, to be more formal:

$$q_Y(\alpha) = F_Y^{-1}(\alpha) = \inf\{y \mid F(Y \leq y) \geq \alpha\} \tag{12}$$

where $\rho_Y(y)$ denotes the PDF of the output Y and $F_Y(y)$ denotes to the CDF of the output Y.To solve such problems, Kucherenko et al. (2019) introduced QBSM $\bar{q}_i^{(1)}$ and $\bar{q}_i^{(2)}$

$$\bar{q}_i^{(1)}(\alpha) = E_{\theta_i}\left(\left|q_Y(\alpha) - q_{Y|\Theta_i}(\alpha)\right|\right) = \int \left|q_Y(\alpha) - q_{Y|\Theta_i}(\alpha)\right| dF_{\theta_i} \tag{13}$$

$$\bar{q}_i^{(2)}(\alpha) = E_{\theta_i}\left[\left(q_Y(\alpha) - q_{Y|\Theta_i}(\alpha)\right)^2\right] = \int \left(q_Y(\alpha) - q_{Y|\Theta_i}(\alpha)\right)^2 dF_{\theta_i} \tag{14}$$

Here, F_{Θ_i} denotes to the CDF of input variable, $q_{Y|\Theta_i}(\alpha)$ denotes to the conditional PDF with Θ_i being fixed at $X_i = x_i^*$ (Song, Bai, Kucherenko, Wang, & Yang, 2021)

$$q_{Y|\Theta_i}(\alpha) = F^{-1}_{Y|\Theta_i}(\alpha) = \inf\left\{P\left(Y \leq y \mid \Theta_i = \theta_i^*\right) \geq \alpha\right\} \tag{15}$$

Additionally, a normalized version of QBSM $Q_i^{(1)}(\alpha)$ and $Q_i^{(2)}(\alpha)$ was also presented in Kucherenko et al. (2019):

$$Q_i^{(1)}(\alpha) = \frac{\bar{q}_i^{(1)}(\alpha)}{\sum_{j=1}^{d} \bar{q}_j^{(1)}(\alpha)} \tag{16}$$

$$Q_i^{(2)}(\alpha) = \frac{\bar{q}_i^{(2)}(\alpha)}{\sum_{j=2}^{d} \bar{q}_j^{(2)}(\alpha)} \tag{17}$$

with $\left\{Q_i^{(1)}(\alpha), Q_i^{(2)}(\alpha)\right\} \in [0,1]$.

4 Monte Carlo Simulation

Because QBSM has no closed-form expression, they must be assessed numerically through Monte Carlo(MC) simulation(Fishman, 1996). MC simulation generates random samples from a probability distribution, and the problem becomes deterministic for each sample. Solving these deterministic problems allows us to obtain statistical information regarding the exact solutions, such as the mean or variance. However, MC approaches are renowned for their slow convergence, with an average coverage rate of $\sqrt{N}$(N stands for the sample size). This means that, to get more precise outcomes, a large number of samples are required. Although various improved MC methods have been developed, such as Latin sampling methods(Loh, 1996) and Quasi-Monte Carlo(QMC) methods(Niederreiter, 1992), there are still certain limits to it.

In Kucherenko and Song (2017), the performance of several numerical schemes based on sampling strategies was compared. It is documented that the double loop reordering(DLR) technique outperforms all other methods, especially when combined with QMC sampling. In this paper, by integrating the two MC estimators provided by Kucherenko et al. (2019) and the procedure improved by Song et al. (2021), we list the following steps to estimate QBSM in our study:

- **Step 1:** To obtain $q_{Y|\Theta_i}$:
  (1) Generate N points $\boldsymbol{\Theta}^{(k)} = \left\{\Theta_1^{(k)}, \Theta_2^{(k)}, \cdots, \Theta_d^{(k)}\right\}, k = 1, 2, \cdots, N$ according to the joint PDF $\rho(\boldsymbol{\Theta})$.
  (2) Compute the unconditional values of output $Y^{(k)} = g\left(\boldsymbol{\Theta}^{(k)}\right), k = 1, 2, \ldots, N$, then reorder them in ascending order $\left\{Y^{(1-st)}, Y^{(2-\mathrm{nd})}, \cdots, Y^{(N-\mathrm{th})}\right\}^{\mathrm{T}}$.
  (3) Decide the α-th output value as unconditional quantile $q_Y(\alpha)$: $q_Y^{(N)}(\alpha) = Y^{(\alpha N-\text{ th })}$.
- **Step 2:** To obtain $q_{Y|\Theta_i}(\alpha)$:
  (1) Generate M points $\left\{\theta_i^{(1)}, \theta_i^{(2)}, \cdots, x_i^{(M)}\right\}^{\mathrm{T}}$ according to the joint PDF $\rho(\Theta)$, which should be independent from $\boldsymbol{\Theta}^{(k)}$
  (2) Fix Θ_i st $\Theta_i = \theta_i^{(k)}, k = 1, \ldots, M$, and get N conditional points

$$\Theta^{(k)} = \left\{\Theta_1^k, \cdots, \Theta_{i-1}^k, x_i^{(k)}, \Theta_{i+1}^k, \cdots, X_d^k\right\}, k = 1, 2, \cdots, N$$

(3) Compute conditional values of $Y_{x_i^{(k)}}^{(k)} = g\left(\boldsymbol{X}^{(k)} \mid X_i = x_i^{(k)}\right)$, and reorder them in ascending order $\left\{Y_{x_i^{(k)}}^{(1-\mathrm{st})}, Y_{x_i^{(k)}}^{(2-\mathrm{nd})}, \cdots, Y_{x_i^{(k)}}^{(N-\mathrm{th})}\right\}$

(4) Decide the α-th output value as conditional quantile: $q_{Y|x_i^{(k)}}(\alpha) = Y_{x_i^{(k)}}^{(\alpha N-\mathrm{th})}$

- **Step 3**: To calculate QBSM accroding to (13) and (14)
- **Step 4**: To calculate normalized QBSM accroding to (16) and (17)

To perform DLR method, M typically be set between 50 to 100(Kucherenko & Song, 2017). To perform brutal force simulation, simply set M = N. The number function evaluation for a fixed i is equal to $N = (dM + 1)$. In general, MC simulation performs efficiently for the calculation of QBSM. However, when deal with extreme quantiles, for instance, $\alpha \leq 0.05$, the performane can be disappointing.

5 Results

The study is greatly governed by computer problems. On the one hand, DCDP models are computationally demanding, which is determined by their structure(Keane, Todd, & Wolpin, 2011). On the other hand, QBSM has no closed form expression to solve, thus they can only be numerically solved using MC simulation. For example, even DLR method is used, the number of function evaluations for computing QGSM for with d inputs are $N = N(dM + 1)$. To simplies the analyis, Figure 5 displays the estimate results of QBSM with 30 individuals and 3 chosen parameters. Even so, it is still computationally burdensome. However our aim with this quantile based sensitivity analysis of the structural econometric model is not to perform an in-depth analysis of the results, but rather to show how these statistical tools provide a better understanding of model complexity.

Table 6 displays QBSM of three carefully picked parameters, and through these poorly converged results we can still conclude that the parametric uncertainty in model prediction in previous literature has been overlooked. The red lines indicate the normalized QBSM of return to an additional year of schoolingα_{11}, while the blue lines indicate the normalized QBSM of return for collage education β_1. The red lines indicate the normalized QBSM of home rewards γ_0. $Q_i^{(1)}$ and $Q_i^{(2)}$ corresponds to the measure listed in (16) and (17), respectively. PDF of the output Y is symmetric(Figure 5) lead to a symmetric behavior of Q_i around $\alpha = 0.5$. The general ranking of variables determined by QBSM is: α_{11}, β_1, γ_0. Measures Q_i approximately reaches the minimum for β_1 and γ_0 and and maximum for α_{11} at $\alpha = 0.6$. When the quantities level α is very small or large, a larger sample size will required to achieve acceptable accuracy. However, the performance of these parameters around $\alpha = 0.5$ is quite stable, with α_{11} account for the greatest variability and γ_0 is nearly have no influence on the model uncertainty.

6 Conclusion

In this paper, we present an application of global sensitivity analysis that is based on quantifies of model output. Specifically, we apply quantile-base sensitivity measures to structural econometric models to identify which parameters contribute most to prediction uncertainty. We employ the Monte Carlo simulation and Double Loop Reordering approach to simulate the quantity of interest. We analyze the quantities of interest in a Discrete Choice Dynamic Programming model and compare it to traditional local sensitivity analysis approaches. The results show that the conventional practice of using point estimates to predict counterfactuals can be very misleading, whereas the proposed method is limited to computational burdens.

References

Adda, J., Dustmann, C., & Stevens, K. (2017, April). The Career Costs of Children. *Journal of Political Economy*, *125*(2), 293–337. Retrieved 2021-09-06, from https://www.journals.uchicago.edu/doi/abs/10.1086/690952 (Publisher: The University of Chicago Press)

Aguirregabiria, V., & Mira, P. (2010, May). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, *156*(1), 38–67. Retrieved 2021-09-13, from https://www.sciencedirect.com/science/article/pii/S0304407609001985

Attanasio, O. P., Meghir, C., & Santiago, A. (2012, January). Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. *The Review of Economic Studies*, *79*(1), 37–66. Retrieved 2021-09-12, from https://doi.org/10.1093/restud/rdr015

Bellman, R. (1966, July). Dynamic Programming. *Science*, *153*(3731), 34–37. Retrieved 2021-09-23, from https://www.sciencemag.org/lookup/doi/10.1126/science.153.3731.34

Blundell, R. (2017, May). What Have We Learned from Structural Models? *American Economic Review*, *107*(5), 287–292. Retrieved 2021-09-13, from https://pubs.aeaweb.org/doi/10.1257/aer.p20171116

Campolongo, F., & Saltelli, A. (1997, July). Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliability Engineering & System Safety*, *57*(1), 49–69. Retrieved 2021-02-24, from https://www.sciencedirect.com/science/article/pii/S0951832097000215

Canova, F. (1994). Statistical inference in calibrated models. *Journal of Applied Econometrics*, *9*(S1), S123–S144. Retrieved 2021-09-20, from https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.3950090508 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.3950090508)

Eckstein, Z., Keane, M., & Lifshitz, O. (2019). Career and Family Decisions: Cohorts Born 1935-1975. *Econometrica*, *87*(1), 217–253. Retrieved 2021-09-06, from https://www.econometricsociety.org/doi/10.3982/ECTA14474

Eisenhauer, P. (2019, January). The approximate solution of finite-horizon discrete-choice dynamic programming models. *Journal of Applied Econometrics*, *34*(1), 149–154. Retrieved 2021-04-11, from https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2648

Eisenhauer, P., Gabler, J., & Janys, L. (2021, March). *Structural models for policy-making: Coping with parametric uncertainty.* Retrieved 2021-08-05, from https://arxiv.org/abs/2103.01115v2

Fishman, G. (1996). *Monte Carlo: Concepts, Algorithms, and Applications.* New York: Springer-Verlag. Retrieved 2021-09-24, from https://www.springer.com/gp/book/9780387945279

Harenberg, D., Marelli, S., Sudret, B., & Winschel, V. (2019). Uncertainty quantification and global sensitivity analysis for economic models. *Quantitative Economics*, *10*(1), 1–41. Retrieved 2021-08-12, from https://onlinelibrary.wiley.com/doi/abs/10.3982/QE866 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE866)

Homma, T., & Saltelli, A. (1996, April). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, *52*(1), 1–17. Retrieved 2021-09-07, from https://www.sciencedirect.com/science/article/pii/0951832096000026

Janos Gabler, & Tobias Raabe. (2020). *respy - A Framework for the Simulation and Estimation of Eckstein-Keane-Wolpin Models* (Tech. Rep.). Retrieved from https://github.com/OpenSourceEconomics/respy

Kala, Z. (2019, April). Quantile-oriented global sensitivity analysis of design resistance. *Journal of Civil Engineering and Management*, *25*(4), 297–305. Retrieved 2020-09-10, from https://journals.vgtu.lt/index.php/JCEM/article/view/9627 (Number: 4)

Keane, M. P., Todd, P. E., & Wolpin, K. I. (2011, January). Chapter 4 - The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4, pp. 331–461). Elsevier. Retrieved 2021-09-06, from https://www.sciencedirect.com/science/article/pii/S0169721811004102

Keane, M. P., & Wolpin, K. I. (1994). The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *The Review of Economics and Statistics*, *76*(4), 648–672. Retrieved 2020-12-04, from https://www.jstor.org/stable/2109768 (Publisher: The MIT Press)

Keane, M. P., & Wolpin, K. I. (1997). The Career Decisions of Young Men. *Journal of Political Economy*, *105*(3), 473–522. Retrieved 2020-12-17, from https://www.jstor.org/stable/10.1086/262080 (Publisher: The University of Chicago Press)

Kucherenko, S., & Song, S. (2017, September). Different numerical estimators for main effect global sensitivity indices. *Reliability Engineering & System Safety*, *165*, 222–238. Retrieved 2020-08-30, from http://www.sciencedirect.com/science/article/pii/S0951832016301065

Kucherenko, S., Song, S., & Wang, L. (2019, May). Quantile based global sensitivity measures. *Reliability Engineering & System Safety*, *185*, 35–48. Retrieved 2020-07-17, from http://www.sciencedirect.com/science/article/pii/S0951832016304574

Leamer, E. E. (1985). Sensitivity Analyses Would Help. *The American Economic Review*, *75*(3), 308–313. Retrieved 2021-09-06, from https://www.jstor.org/stable/1814801 (Publisher: American Economic Association)

Loh, W.-L. (1996). On Latin Hypercube Sampling. *The Annals of Statistics*, *24*(5), 2058–2080. Retrieved 2021-09-24, from https://www.jstor.org/stable/2242641 (Publisher: Institute of Mathematical Statistics)

Manski, C. F. (2021, February). Econometrics For Decision Making: Building Foundations Sketched By Haavelmo And Wald. *arXiv:1912.08726 [econ]*. Retrieved 2021-09-06, from http://arxiv.org/abs/1912.08726 (arXiv: 1912.08726)

Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods.* USA: Society for Industrial and Applied Mathematics.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., … Maier, H. R. (2021, March). The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, *137*, 104954. Retrieved 2021-02-03, from http://www.sciencedirect.com/science/article/pii/S1364815220310112

Saltelli, A. (Ed.). (2004). *Sensitivity analysis in practice: a guide to assessing scientific models.* Hoboken, NJ: Wiley.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., … Wu, Q. (2019, April). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39. Retrieved 2021-09-01, from https://www.sciencedirect.com/science/article/pii/S1364815218302822

Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2005, July). Sensitivity Analysis for Chemical Models. *Chemical Reviews*, *105*(7), 2811–2828. Retrieved 2021-09-22, from https://pubs.acs.org/doi/10.1021/cr040659d

Sauer, R. M. (2004, October). Educational Financing and Lifetime Earnings. *The Review of Economic Studies*, *71*(4), 1189–1216. Retrieved 2021-09-23, from https://doi.org/10.1111/0034-6527.00319

Sobol, I. (1993). Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modelling and Computational Experiments*(4), 407–414. Retrieved 2021-09-06, from https://www.semanticscholar.org/paper/Sensitivity-Estimates-for-Nonlinear-Mathematical-Sobol/3e0b415213a580254226fdbcbfc9980b70dd0468

Song, S., Bai, Z., Kucherenko, S., Wang, L., & Yang, C. (2021, April). Quantile sensitivity measures based on subset simulation importance sampling. *Reliability Engineering & System Safety*, *208*, 107405. Retrieved 2021-05-19, from https://www.sciencedirect.com/science/article/pii/S0951832020308917

Zi, Z. (2011, November). Sensitivity analysis approaches applied to systems biology models. *IET systems biology*, *5*(6), 336–336.

6.1 Appendix A: Tables

Table 1. Parameterization of Keane and Wolpin (1994) model

Category	Parameter	Value	Standard Error	Definition
General	δ	0.95	0.000616	discount factor
Blue-collar	α_{10}	9.21	0.011858	log of rental price
	α_{11}	0.038	0.001045	return to an additional year of schooling
	α_{12}	0.033	0.000444	return to same sector experience
	α_{13}	-0.0005	0.000012	return to same sector, quadratic experience
	α_{14}	0.0	0.000738	return to other sector experience
	α_{15}	0.0	0.000029	return to other sector, quadratic experience
White collar	α_{20}	8.48	0.005998	log of rental price
	α_{21}	0.07	0.000344	return to an additional year of schooling
	α_{22}	0.067	0.000546	return to same sector experience
	α_{23}	-0.001	0.000016	return to same sector, quadratic experience
	α_{24}	0.022	0.00033	return to other sector experience
	α_{25}	-0.0005	0.000022	return to other sector, quadratic experience
Education	β_0	0.0	244.837595	constant reward for choosing education
	β_1	0.0	134.591673	reward for going to college (tuition, etc.)
	β_2	-4000	209.14159	reward for going back to school
Home	γ_0	17750	270.436557	constant reward of non-market alternative

6.2 Appendix B: Figures

Figure 1. Decision tree

Figure 2. Heat map of selective parameters

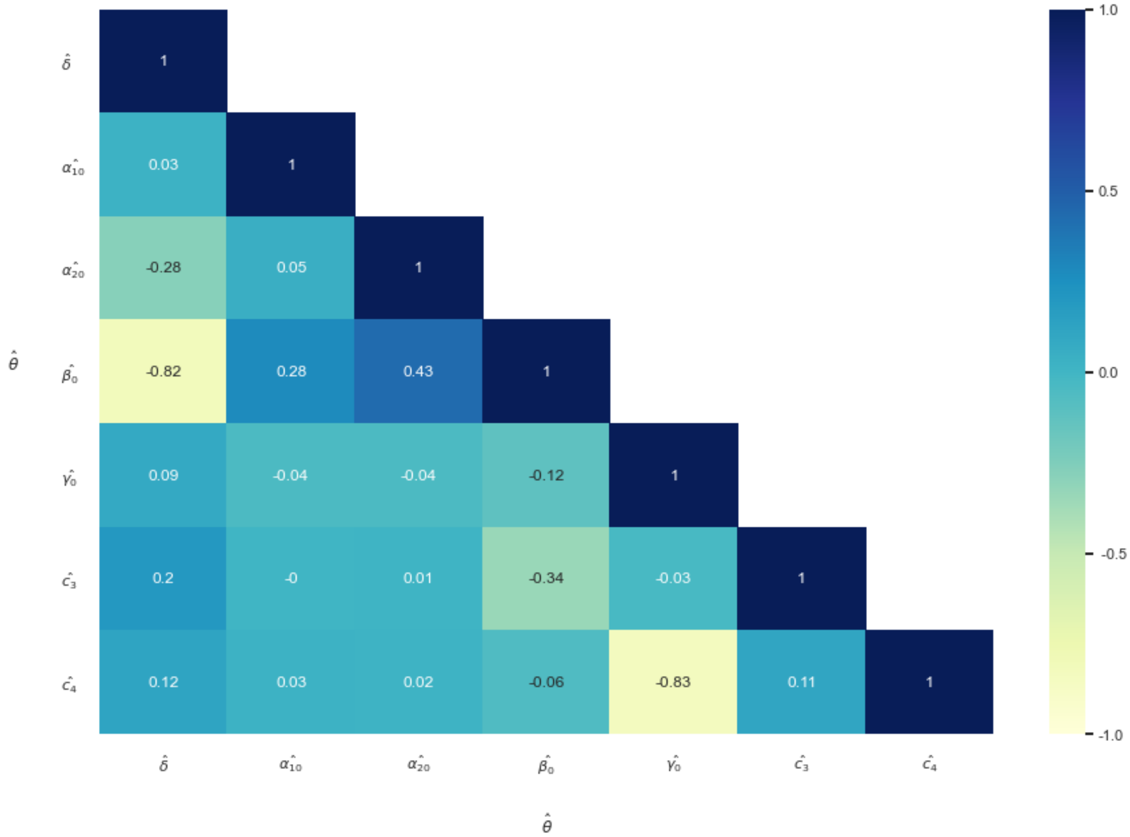


Figure 3. Comparison of shares of occupation decisions over time between scenarios

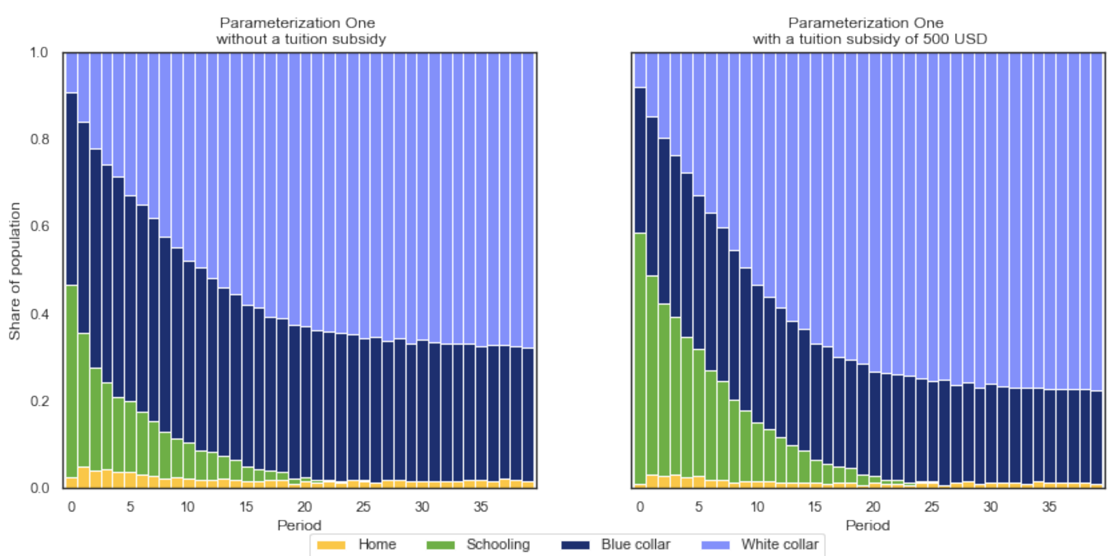


Figure 4. Probability distribution of QoI

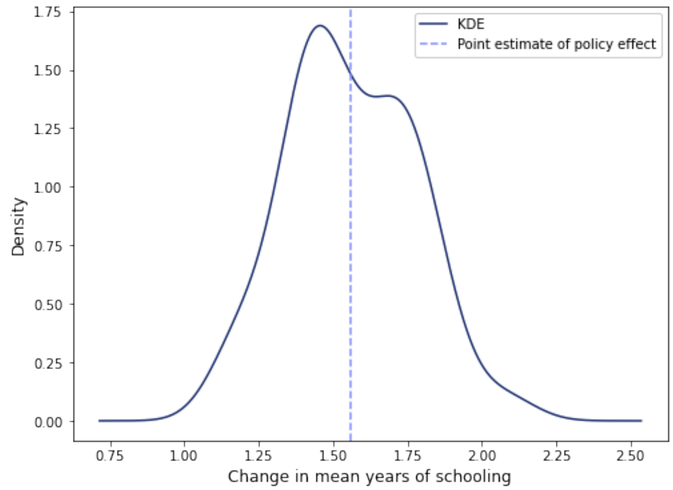


Figure 5. PDF of output for SA

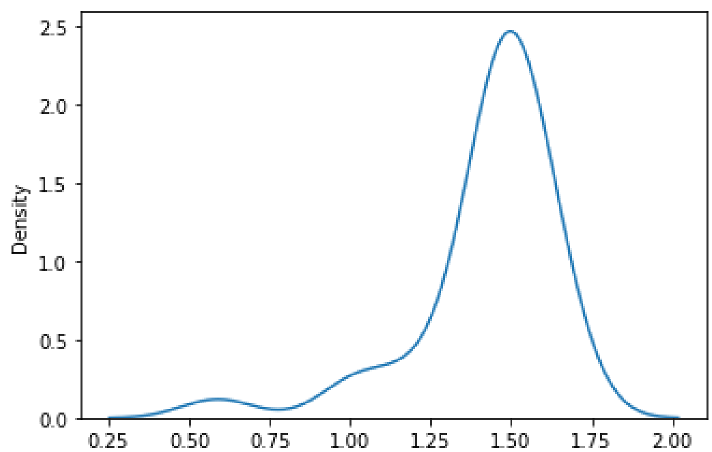
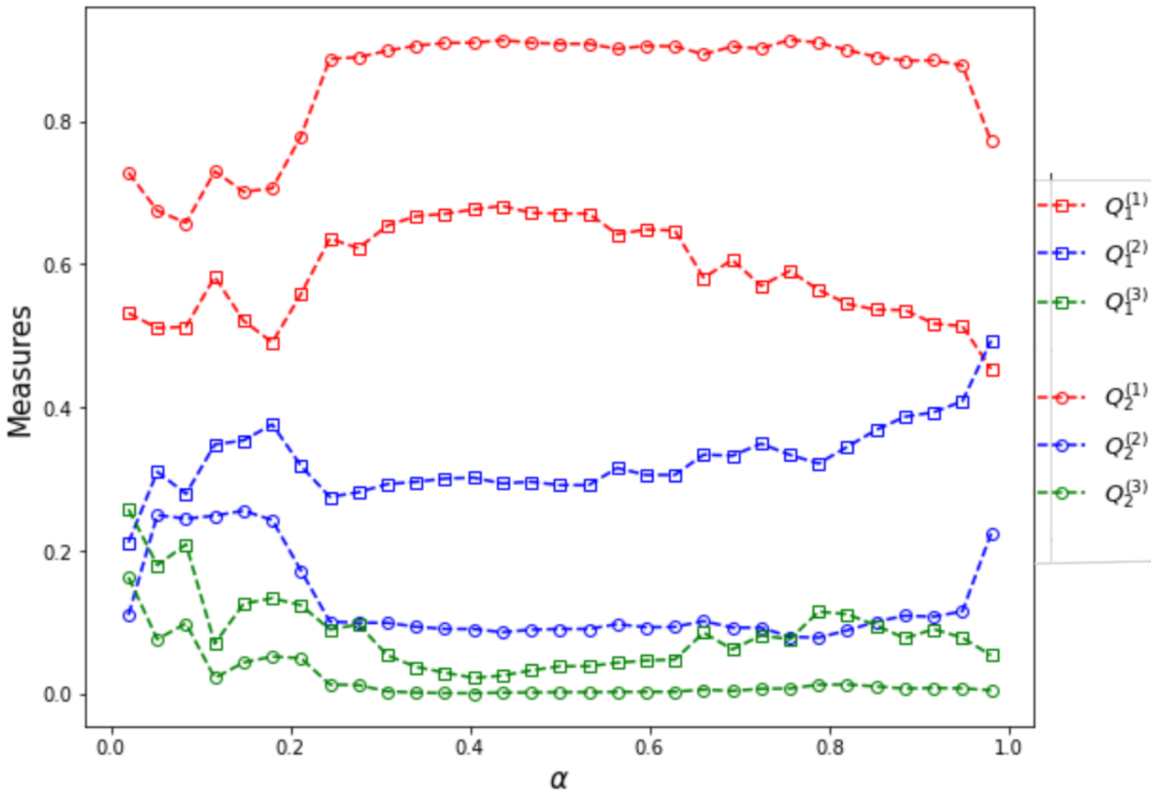


Figure 6. Quantile-base sensitivity measures on Keane and Wolpin (1994) model



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