

Sensitivity Analysis for Dynamic Discrete Choice Models

Chun Pong Lau*

August 29, 2024

Abstract

In dynamic discrete choice models, some parameters, such as the discount factor, are being fixed instead of being estimated. This paper proposes two sensitivity analysis procedures for dynamic discrete choice models with respect to the fixed parameters. First, I develop a local sensitivity measure that estimates the change in the target parameter for a unit change in the fixed parameter. This measure is fast to compute as it does not require model re-estimation. Second, I propose a global sensitivity analysis procedure that uses model primitives to study the relationship between target parameters and fixed parameters. I show how to apply the sensitivity analysis procedures of this paper through two empirical applications.

Keywords: Dynamic discrete choice models, sensitivity analysis, discount factor

*Kenneth C. Griffin Department of Economics, The University of Chicago, ccplau@uchicago.edu. I am grateful to Alex Torgovitsky for his continuous guidance and advice. I also thank Michael Din-
erstein, Jean-Pierre Dubé, Jacob Leshno, Ali Hortaçsu, Kirill Ponomarev, and Azeem Shaikh, as well as
participants of the Econometrics Advising Group, the Econometrics Student Group, the Industrial Organi-
zation Working Group, and the Third Year Microeconomics Research Seminar at the University of Chicago
for valuable comments. All errors are my own.

1 Introduction

In dynamic discrete choice (DDC) models, some structural parameters are often fixed rather than being estimated. A leading example of such parameters is the discount factor, which is nonparametrically unidentified without further restrictions (Rust, 1994; Magnac and Thesmar, 2002). Apart from DDC models, it is also common to have model parameters being fixed exogenously in calibrating general equilibrium models (Dawkins et al., 2001). In the rest of this paper, the parameters that are fixed in the estimation procedure are referred to as fixed parameters.

The choice of the discount factor in DDC models is usually based on the value used in related papers, some relevant rates of return, or some values larger than 0.9. Researchers may conduct sensitivity analysis by repeating part of the estimation at a few other values of the discount factor. But this can be time-consuming because estimating the full model once can take days or weeks. Hence, the current practice can only offer limited information on how the conclusions are affected by the discount factor due to the high computational cost.

In this paper, I propose new approaches to conduct local and global sensitivity analysis for DDC models with respect to the fixed parameters.

To begin with, I develop a local sensitivity measure that examines the change of the target parameter due to a small change in the fixed parameter. I show that the local sensitivity measure is low-cost to compute because it can be obtained by solving a system of linear equations. Researchers do not need to re-estimate the model in order to compute the local sensitivity measure. In addition, reporting the local sensitivity measure can be more informative than re-estimating the model at a few other values of the fixed parameters because readers can estimate the target parameter at their chosen values of the fixed parameters. I show that the local sensitivity measure can serve as a good local approximation through two empirical applications.

If the literature has some consensus that a certain fixed parameter lies in a tight interval, then local sensitivity analysis can already be informative to approximate how the conclusion changes in such an interval. However, this may not always be the case. For the discount factor, evidence from behavioral economics shows that there can be a lot of variation in the discount factor depending on the context and the sample (Frederick et al., 2002). A recent study by Kong et al. (2022) estimates the discount factor for various consumer goods and reports a wide range of discount factors among different products, from 0.357 (for mayonnaise) to 0.999 (for peanut butter). As a result, researchers may be

concerned that their conclusion does not hold at another value of the discount factor.

This motivates the global sensitivity analysis in the current paper that examines the structure of DDC models and finds conditions on the model primitives under which the target parameter is monotone in the fixed parameter. Monotonicity can be a useful property because of two computational benefits. First, parameters that are monotone in the fixed parameter are bounded by the endpoints. Second, researchers can easily estimate the breakdown point at which the conclusion changes (Horowitz and Manski, 1995; Kline and Santos, 2013; Masten and Poirier, 2020). In the current practice of sensitivity analysis, researchers typically re-estimate the model at a few (e.g., three) neighboring values of the fixed parameter used for the main analysis because estimating the model once is costly. For instance, Barwick and Pathak (2015), Fowlie et al. (2016), and Igami (2017) repeat the estimation using discount factors around the one used for the main analysis, and examine how the parameter estimates change with the discount factor. Although monotonic patterns are usually shown, they might not necessarily generalize to the entire support of the discount factor. Using the discount factor as a leading case of fixed parameters in DDC models, I show that utility can be monotone in the discount factor under some conditions on the transition matrices and conditional choice probabilities. However, counterfactuals are not necessarily monotone in the discount factor, even if utility is monotone in the discount factor. Therefore, I also propose a constrained optimization approach for global sensitivity analysis for more general target parameters and fixed parameters.

As will be discussed in Section 3, the methodology of this paper is not specific to single-agent DDC models nor the discount factor. The procedures proposed in this paper can be applied to other constrained optimization problems (see problem (8) ahead) with a unique solution and fixed parameters. Solving the single-agent DDC model using the full solution method is just an example with such a structure. Some other potential economic applications that contain a constrained optimization structure and a subset of parameters being fixed include: dynamic games (Egedal et al., 2015), dynamic matching (Verdier and Reeling, 2021; Chen and Choo, 2023), international trade (Ossa, 2014), and productivity (Yang, 2021). See Aguirregabiria et al. (2021) for a recent comprehensive review that contains many DDC examples related to industrial organization.

1.1 Related literature

This paper contributes to several strands of literature in economics. First, it is related to the literature on sensitivity analysis. Andrews et al. (2017) and Honoré et al. (2020) de-

velop measures to analyze the sensitivity of parameter estimates to the moments. More closely related are the papers by [Iskrev \(2019\)](#) and [Jørgensen \(2023\)](#) that conduct sensitivity analysis with respect to calibrated parameters. The former paper focuses on Bayesian approaches to macroeconomic models. The latter proposes a local sensitivity measure to study the sensitivity with respect to calibrated parameters when the target parameters are estimated from minimizing an unconstrained optimization problem with a Generalized Method of Moments (GMM) objective. The first contribution of the current paper is to provide computationally attractive tools for conducting local sensitivity analysis for estimators obtained from constrained optimization problems. Fixed-point constraints are common in economic problems to represent equilibrium conditions. The current paper focuses on sensitivity analysis with respect to the fixed parameters and is different from some other recent papers in the sensitivity analysis literature such as [Armstrong and Kolesár \(2021\)](#) which propose confidence intervals robust to local misspecification for overidentified moment condition models, [Bonhomme and Weidner \(2022\)](#) which focus on robustness to misspecification within a larger class of models, and [Christensen and Connault \(2022\)](#) which examine sensitivity to the distribution of the latent variables.

The global sensitivity analysis section of this paper shares a similar theme as the literature on monotone comparative statics (e.g., [Topkis \(1998\)](#)). [Light \(2021\)](#) is a recent paper that provides conditions for the policy function to be monotone in the discount factor, the parameters in the payoff function, or the transition probability function for Markov decision processes. But [Light \(2021\)](#) considers a different model than the one in the current paper and does not consider estimation of model parameters. I also do not impose conditions like increasing differences that are required in [Light \(2021\)](#).

1.2 Outline

The rest of the paper is organized as follows. Section 2 describes the setup and notation. Sections 3 and 4 describe the methodology for local and global sensitivity analysis, respectively. Section 5 contains two empirical applications in which I apply the methodology to the seminal bus engine replacement example in [Rust \(1987\)](#) and to a recent dynamic matching model in [Chen and Choo \(2023\)](#). Section 6 concludes. All proofs can be found in the appendix.

2 Model

In this paper, the [Rust \(1987\)](#) model is used as the running example to illustrate the local and global sensitivity analysis procedures. In order to introduce the relevant notations, this section starts by describing the canonical single-agent DDC model and common assumptions. Then, I outline some common solution methods for DDC models and their connection with constrained and unconstrained optimization problems that are relevant for the sensitivity analysis procedures in Sections [3](#) and [4](#).

2.1 Notations

Consider a DDC model, where time is indexed by $t = 1, \dots, T$. In each period t , each agent $i \in \mathcal{I} \equiv \{1, \dots, N\}$ chooses an action $a_{it} \in \mathcal{A} \equiv \{0, 1, \dots, A\}$ to maximize discounted future utility based on the state variables s_{it} . The vector of state variables can be decomposed as $s_{it} \equiv (x_{it}, \epsilon_{it})$, where $x_{it} \in \mathcal{X} \equiv \{1, \dots, X\}$ is observable by the agents and the researcher and $\epsilon_{it} \equiv (\epsilon_{0it}, \dots, \epsilon_{Ait}) \in \mathbb{R}^{A+1}$ is unobservable to the researcher. In addition, assume that X is finite, ϵ_{ait} is i.i.d. across agents, choices, and states, and that ϵ_{it} is continuously distributed and has full support over $\mathbb{R}^{A+1}$.

Each agent chooses the sequence of actions to maximize discounted future utility:

$$\max_{\{a_{it}\}} \mathbb{E} \left[\sum_{t=1}^{\infty} \beta^t \tilde{\pi}_i(a_{it}, s_{it}; \theta) \right],$$

where $\beta \in [0, 1)$ is the discount factor common across agents and $\tilde{\pi}_i(a_{it}, s_{it}; \theta)$ is the utility function of choosing action a_{it} at state s_{it} and parameterized by $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$.

2.2 Single-agent dynamic discrete choice model

In this section, I consider a single-agent stationary DDC model, with the following standard assumptions (see, e.g., [Hortaçsu and Joo \(2023\)](#)). I omit the i subscript because there is only one agent.

Assumption 2.1 (Additive separability). *The utility function can be written as*

$$\tilde{\pi}(a_t, s_t; \theta) = \pi(a_t, x_t; \theta) + \epsilon_t,$$

where $\pi(a_t, x_t; \theta)$ is bounded and monotone in x_t .

Assumption 2.2 (i.i.d. error terms). *For any ϵ_t , $\mathbb{P}[\epsilon_{t+1} | \epsilon_t] = \mathbb{P}[\epsilon_{t+1}]$.*

Assumption 2.3 (Conditional independence). *Given (x_t, a_t) observed, $x_{t+1} \perp\!\!\!\perp (\epsilon_t, \epsilon_{t+1})$.*

Let $\bar{V}(x_t, \epsilon_t)$ be the agent's value function. By Bellman's principle of optimality, the value equation can be written as

$$\bar{V}(x_t, \epsilon_t) = \max_{a \in \mathcal{A}} \{ \pi(a, x_t; \theta) + \epsilon_{at} + \beta \mathbb{E}[\bar{V}(x_{t+1}, \epsilon_{t+1}) | a, x_t] \}.$$

Define the choice-specific value function as

$$v(a, x) \equiv \pi(a, x; \theta) + \beta \mathbb{E}[\bar{V}(x_{t+1}, \epsilon_{t+1}) | a_t = a, x_t = x],$$

for any $a \in \mathcal{A}$ and $x \in \mathcal{X}$. The conditional choice probability (CCP) of choosing action $a \in \mathcal{A}$ at state $x \in \mathcal{X}$ is given by

$$\mathbb{P}[a_t = a | x_t = x] \equiv \mathbb{P} \left[a \in \arg \max_{j \in \mathcal{A}} \{ v(j, x) + \epsilon_{jt} \} \middle| x_t = x \right].$$

The following assumption on the distributions of the unobservables is standard in the literature.

Assumption 2.4 (Distribution of the unobservables). *ϵ_{at} follows a mean-zero type-1 extreme value (T1EV) distribution for each $a \in \mathcal{A}$ and time t .*

Under Assumption 2.4, the CCP can be written as

$$\mathbb{P}[a_t = a | x_t = x] = \frac{\exp\{v(a, x)\}}{\sum_{j \in \mathcal{A}} \exp\{v(j, x)\}}, \quad (1)$$

for each $a \in \mathcal{A}$ and $x \in \mathcal{X}$.

Next, let the state transition be governed by the Markov transition matrix Q_a , where the (x, x') -entry of Q_a is the probability of transitioning from state x in period t to state x' in period $t + 1$ when action a is chosen in period t , i.e., $q(x' | x, a) \equiv \mathbb{P}[x_{t+1} = x' | x_t = x, a_t = a]$. Let $Q_a(x) \equiv (q(1 | x, a), \dots, q(X | x, a))'$ be the x -th row of the transition matrix Q_a for any $a \in \mathcal{A}$. Define the ex ante value function as $V(x) \equiv \mathbb{E}[\bar{V}(x, \epsilon_t)]$, and write $V \equiv (V(1), \dots, V(X))'$. Then, V satisfies the following fixed-point relationship using Assumption 2.4 that the unobservables follow a mean-zero T1EV distribution:

$$V(x) = \log \left\{ \sum_{a \in \mathcal{A}} \exp [\pi(x, a; \theta) + \beta Q_a(x)' V] \right\}, \quad (2)$$

for any $x \in \mathcal{X}$. Let Ψ^V represents the Bellman operator on the right hand side of (2), then the fixed-point relationship (2) can be summarized as follows

$$V = \Psi^V(\theta, V; \beta). \quad (3)$$

Finally, this section ends with another representation of the flow utility and its connection with CCP. This representation is useful for global sensitivity analysis in Section 4. Using Lemma 1 of [Arcidiacono and Miller \(2011\)](#), there exists a real-valued function $\psi_a(\cdot)$ such that

$$V(x) = v(x, a) + \psi_a(p(x)),$$

for any $a \in \mathcal{A}$ and $x \in \mathcal{X}$. Let $\pi_a \equiv (\pi(1, a; \theta), \dots, \pi(X, a; \theta))'$ be the vector of utility functions at action $a \in \mathcal{A}$. Following the discussion in [Kalouptsi et al. \(2021a,b\)](#), the vector π_a can be expressed as

$$\pi_a = A_a \pi_A + b_a(p), \quad (4)$$

for any $a \in \mathcal{A} \setminus \{A\}$, where

$$\begin{aligned} A_a &\equiv (I_X - \beta Q_a)(I_X - \beta Q_A)^{-1}, \\ b_a(p) &\equiv A_a \psi_A(p) - \psi_a(p). \end{aligned}$$

Under Assumption 2.4 that the unobservables follow a mean zero T1EV distribution, it follows that $\psi_a(p) = -\log p_a(x)$ for any $a \in \mathcal{A}$ and $x \in \mathcal{X}$ (see also [Hotz and Miller \(1993\)](#)), and $b_a(p)$ can be written as

$$b_a(p) = -A_a \log p_A(x) + \log p_a(x),$$

for any $x \in \mathcal{X}$ and $a \in \mathcal{A} \setminus \{A\}$.

2.3 Solution methods

There are different methods to solve DDC models (see [Aguirregabiria et al. \(2021\)](#) and [Hortaçsu and Joo \(2023\)](#) for details). I briefly outline three common approaches in this section in order to emphasize their structure as constrained or unconstrained optimization problems that would fit into the local sensitivity analysis framework in the next section.

Let $L(\theta, V; \beta)$ be the likelihood function. Here, I introduce β as the argument of the

likelihood function after the semicolon to indicate that it is a parameter that researchers need to specify in advance, and is fixed throughout the estimation procedure.

The nested fixed-point method (NFXP) by [Rust \(1987\)](#) involves value function iteration and contains two loops. The inner loop takes the parameter θ as given and finds the fixed point that solves equation (3). The outer loop finds the parameter θ that maximizes the likelihood function. [Su and Judd \(2012\)](#) show that NFXP is equivalent to solving it by a mathematical program with equilibrium constraints (MPEC):

$$\begin{aligned} \max_{\theta, V} \quad & L(\theta, V; \beta) \\ \text{s.t.} \quad & V = \Psi^V(\theta, V; \beta). \end{aligned} \tag{5}$$

Upon convergence, the solution must satisfy the Bellman equations and maximize likelihood. As a result, the system (5) can be used as a starting point for local sensitivity analysis if the researcher uses NFXP or MPEC to solve the DDC model.

Two-step CCP methods can also be written in a similar manner. Using the [Hotz and Miller \(1993\)](#) inversion, the ex ante value function can be written as

$$V(x) = v(a, x) - \log \mathbb{P}[a_t = a | x_t = x].$$

By a suitable normalization, such as $v(A, x) = 0$ for all $x \in \mathcal{X}$ ([Hortaçsu and Joo, 2023](#)), the choice-specific value function can be written as

$$v(a, x) = \log \mathbb{P}[a_t = a | x_t = x] - \log \mathbb{P}[a_t = A | x_t = x], \tag{6}$$

for each $a \in \mathcal{A} \setminus \{A\}$ and $x \in \mathcal{X}$. Hence, with a given estimator of the CCP, the ex ante value functions can be estimated via (6) over all choices and states. With the estimated V and substituting the constraint into the objective, it becomes an unconstrained optimization problem.

[Aguirregabiria and Mira \(2002\)](#) propose the nested pseudo-likelihood method that iterates on the policy function instead. Their fixed-point equation is written as

$$P = \Psi^P(\theta, P; \beta), \tag{7}$$

where Ψ^P is the policy function operator. The K -stage policy iteration estimator takes the estimator of the policy function from the previous stage $\hat{P}^{K-1}$ and solves the following

problem that updates the policy function via equation (7):

$$\begin{aligned} \max_{\theta} \quad & L(\theta, P^K; \beta) \\ \text{s.t.} \quad & P^K = \Psi^P(\theta, \hat{P}^{K-1}; \beta). \end{aligned}$$

2.4 Running example: The bus engine replacement problem

For the rest of this paper, I provide examples in terms of the seminal bus engine replacement problem in Rust (1987). In this problem, Harold Zurcher, the manager, observes the bus mileage since the last engine replacement. The bus mileage for each bus $i = 1, \dots, M$ is denoted by x_{it} and the unobservable state variable is ϵ_{it} . In each period, Zurcher chooses to replace ($a_{it} = 1$) or maintain ($a_{it} = 0$) the bus engine. Assume that the utility functions are the same across the buses. Let $\theta \equiv (\text{MC}, \text{RC})$, MC be the maintenance cost, RC be the replacement cost, and $c(x, \text{MC})$ be the cost of maintaining engine at mileage $x = 1, \dots, X$. The utility function is given by

$$\tilde{\pi}(a, x, \epsilon_{it}; \theta) = \pi(a, x; \theta) + \epsilon_{ait},$$

where $\pi(0, x; \theta) = \text{RC} + c(x, \text{MC})$ and $\pi(1, x; \theta) = c(0, \text{MC})$ for any $x \in \mathcal{X}$. Here, Q_0 and Q_1 are the two Markov transition matrices that correspond to the actions that choose to maintain and replace the engine, respectively. Bus mileage is reset to 1 if $a_{it} = 1$. Otherwise, the transition probability of mileage follows a multinomial distribution as below:

- If $x \leq X - 2$, the transition probability is

$$\mathbb{P}[x_{it+1} = x' | x_{it} = x, a_{it} = 0] = \begin{cases} \phi_1 & , x' = x + 1 \\ \phi_2 & , x' = x + 2 \\ 1 - \phi_1 - \phi_2 & , x' = x \end{cases}.$$

- If $x \leq X - 1$, the transition probability is

$$\mathbb{P}[x_{it+1} = x' | x_{it} = x, a_{it} = 0] = \begin{cases} \phi_1 & , x' = x + 1 \\ 1 - \phi_1 & , x' = x \end{cases}.$$

- $\mathbb{P}[x_{it+1} = X | x_{it} = X, a_{it} = 0] = 1$.

The fixed-point relationship for the Bellman equation in this example can be written explicitly as follows

$$V(x) = \log \{ \exp[RC + c(x, MC) + \beta Q_0(x)'V] + \exp[c(0, MC) + \beta Q_1(x)'V] \},$$

for each $x \in \mathcal{X}$.

3 Local sensitivity analysis

Let $V \in \mathcal{V} \subseteq \mathbb{R}^{d_V}$ be a vector of auxiliary parameters and $\gamma \in \Gamma \subseteq \mathbb{R}^{d_\gamma}$ be a vector of fixed parameters. Assume that a researcher is interested in estimating θ through the following constrained optimization problem by first fixing the parameter γ as follows:

$$\begin{aligned} \min_{\theta \in \Theta, V \in \mathcal{V}} \quad & L(\theta, V; \gamma) \\ \text{s.t.} \quad & V = F(\theta, V; \gamma), \end{aligned} \tag{8}$$

where $L(\theta, V; \gamma)$ is the criterion function, and the constraint $V = F(\theta, V; \gamma)$ describes some fixed-point relationship that captures the equilibrium constraints.

In terms of the DDC model in Section 2, θ is the utility parameter, V is the value function, and γ is the discount factor. For the NFXP, $L(\theta, V; \gamma)$ corresponds to the likelihood function, and $V = F(\theta, V; \gamma)$ corresponds to the fixed-point equation (2) based on Bellman optimality. Here, I allow γ to be a vector because researchers might fix multiple parameters. For instance, [Igami \(2017\)](#) calibrates the discount factor, the rate of change of innovation cost, and the number of potential entrants.

Let $(\hat{\theta}(\gamma), \hat{V}(\gamma))$ be the solution obtained from solving the constrained optimization problem (8). Note that the optimal solution has γ as an argument because the constrained optimization problem is solved with the pre-specified γ that is fixed throughout the estimation procedure.

The following assumptions on the constrained optimization problem (8) are maintained throughout the paper:

Assumption 3.1.

1. L and F are continuously differentiable in θ , V , and γ around $(\hat{\theta}(\gamma), \hat{V}(\gamma))$.
2. $(\hat{\theta}(\gamma), \hat{V}(\gamma))$ is a regular point and is the unique solution to the optimization problem (8) and belongs to the interior of $\Theta \times \mathcal{V}$ for each $\gamma \in \Gamma$.

Assumption 3.1.1 ensures that the derivatives in the local sensitivity measure exist. See Rust (1988) and Norets (2010) for results on differentiability results related to DDC models. Assumption 3.1.2 ensures that the first-order condition holds as $(\hat{\theta}(\gamma), \hat{V}(\gamma))$ is a regular point (Bertsekas, 1999, Chapter 3) and that there is a unique solution to the optimization problem regardless of the value of the fixed parameter $\gamma \in \Gamma$.

3.1 Sensitivity measure

The gradient of $\hat{\theta}(\gamma)$ with respect to γ , i.e.,

$$\frac{\partial \hat{\theta}(\gamma)}{\partial \gamma'}, \quad (9)$$

can be used as a measure of sensitivity. The (i, j) -component of the above matrix measures the change in the i -th target parameter for a unit change in the j -th fixed parameter. Depending on the parameter and the context, the following sensitivity measures may be easier to interpret:

1. The elasticity gives the percentage change in $\hat{\theta}(\gamma)_i$ for one percentage change in γ_j . It is defined by

$$\frac{\partial \hat{\theta}(\gamma)_i}{\partial \gamma_j} \frac{\gamma_j}{\hat{\theta}(\gamma)_i}, \quad (10)$$

when $\hat{\theta}(\gamma)_i, \gamma_j \neq 0$.

2. The semi-elasticity gives a unit change in $\hat{\theta}(\gamma)_i$ for a percentage change in γ_j . It is defined by

$$\frac{\partial \hat{\theta}(\gamma)_i}{\partial \gamma_j} \gamma_j, \quad (11)$$

when $\gamma_j \neq 0$.

The sensitivity measure for $\hat{V}(\gamma)$ can be defined analogously.

I show that computing the local sensitivity measures defined above amounts to solving a linear system of equations. The coefficients and constants in the linear system are evaluated at $(\hat{\theta}(\gamma), \hat{V}(\gamma))$ at the original γ . Hence, there is no need to re-estimate the model at another value of γ .

The following proposition summarizes the main result of the local sensitivity analysis procedure. For notational simplicity, I write $(\theta^*, V^*) \equiv (\hat{\theta}(\gamma), \hat{V}(\gamma))$.

Proposition 3.2. *Let Assumption 3.1 hold. Denote λ^* as the Lagrange multiplier for the con-*

strained optimization problem (8), evaluated at the optimal solution. The local sensitivity measure of $(\theta^*, V^*, \lambda^*)'$ with respect to γ , i.e., $\frac{\partial \theta}{\partial \gamma'}$, $\frac{\partial V}{\partial \gamma'}$, and $\frac{\partial \lambda}{\partial \gamma'}$, can be obtained by solving the following system of $(2d_V + d_\theta)$ equations in $(2d_V + d_\theta)$ unknowns:

$$\begin{aligned} A_{\theta,\theta} \frac{\partial \theta}{\partial \gamma'} + A_{\theta,V} \frac{\partial V}{\partial \gamma'} - (F_\theta)' \frac{\partial \lambda}{\partial \gamma'} &= -A_{\theta,\gamma} \\ A_{V,\theta} \frac{\partial \theta}{\partial \gamma'} + A_{V,V} \frac{\partial V}{\partial \gamma'} + (I_{d_V} - F_V)' \frac{\partial \lambda}{\partial \gamma'} &= -A_{V,\gamma} \\ F_\theta \frac{\partial \theta}{\partial \gamma'} + (F_V - I_{d_V}) \frac{\partial V}{\partial \gamma'} &= -F_\gamma, \end{aligned} \quad (12)$$

if the system (12) has a unique solution, where

- $A_{x,y} \equiv \frac{\partial^2 L}{\partial x \partial y'} - R_{d_x} \frac{\partial \text{vec}[(\frac{\partial F}{\partial x'})']}{\partial y'}$ and $F_x \equiv \frac{\partial F}{\partial x'}$ for $x \in \mathbb{R}^{d_x}$ and $y \in \mathbb{R}^{d_y}$.
- $\text{vec}(B)$ stacks the columns of the matrix $B \in \mathbb{R}^{m \times n}$ into a column vector, i.e., $\text{vec}(B) \equiv (b_{1,1}, \dots, b_{1,m}, b_{2,1}, \dots, b_{2,m}, \dots, b_{n,1}, \dots, b_{n,m})'$.
- $R_d \equiv (\lambda_1^* I_d, \lambda_2^* I_d, \dots, \lambda_{d_V}^* I_d) = \lambda^{*'} \otimes I_d$, where $\otimes$ denotes Kronecker product, and I_d is the $d \times d$ identity matrix.
- All the terms above are evaluated at the optimal solution, e.g., $\frac{\partial^2 L}{\partial \theta \partial \theta'} \equiv \frac{\partial^2 L}{\partial \theta \partial \theta'} \Big|_{(\theta,V)=(\theta^*,V^*)}$.

The proof of Proposition 3.2 can be found in the appendix. Note that the quantities required to compute the sensitivity measures are either already computed in the model estimation procedure or are fast to compute. Other quantities that are not immediately available can be obtained analytically or numerically without model re-estimation. If one wishes to compute the coefficient terms by numerical derivatives, model re-estimation is not required because the derivatives are evaluated around the optimal solution for the fixed value of γ .

As already mentioned in the introduction, Proposition 3.2 can be applied to other economic problems with a constrained optimization structure and unique optimum. I show in Section 3.4 that Proposition 3.2 nests local sensitivity analysis for unconstrained optimization problems. Thus, the framework in this section is not specific to the running example of DDC models, the T1EV assumption, or the discount factor.

Comparing and reporting the local sensitivity measures have two benefits. First, it can be used to compare the sensitivity of the results with respect to the fixed parameters. Researchers can use this to find the fixed parameters that affect the results the most or determine which of the main results are more sensitive to the fixed parameters. Thus, this can also serve as a guide for more extensive sensitivity analysis.

Second, it can be used as a local approximation of the target parameter at another value of the fixed parameter. The empirical applications in Section 5 conduct sensitivity

analysis and examine the performance of local approximation through two empirical applications.

The following example demonstrates how the quantities in Proposition 3.2 can be computed in the context of the Rust (1987) model.

Example 3.3. In this example, I show the analytical expressions for the terms in the Rust (1987) model that are relevant for the linear system in Proposition 3.2. The following derivatives have to be evaluated at the optimal solution with the pre-specified discount factor. I follow the notations introduced in Section 2.4 and assume the cost function is given by $c(x, MC) = -MCx$. In addition, let $x, y, z \in \mathcal{X}$ and denote $p(x) \equiv \mathbb{P}[a_{it} = 1 | x_{it} = x]$ for all $x \in \mathcal{X}$. The likelihood function can be written as

$$L = \sum_{i=1}^M \sum_{t=1}^T \{a_{it} \log p(x_{it}) + (1 - a_{it}) \log[1 - p(x_{it})]\}.$$

The relevant second derivatives for the likelihood function are as follows:

$$\begin{aligned} \frac{\partial^2 L}{\partial \theta \partial \theta'} &= - \sum_{i=1}^M \sum_{t=1}^T p(x_{it}) [1 - p(x_{it})] \begin{pmatrix} x_{it}^2 & -x_{it} \\ -x_{it} & 1 \end{pmatrix}, \\ \frac{\partial^2 L}{\partial \theta \partial V(y)} &= - \sum_{i=1}^M \sum_{t=1}^T \beta p(x_{it}) [1 - p(x_{it})] [q(y|x_{it}, 0) - q(y|x_{it}, 1)] \begin{pmatrix} -x_{it} \\ 1 \end{pmatrix}, \\ \frac{\partial^2 L}{\partial \theta \partial \beta} &= - \sum_{i=1}^M \sum_{t=1}^T \beta p(x_{it}) [1 - p(x_{it})] [Q_0(x_{it}) - Q_1(x_{it})]' V \begin{pmatrix} -x_{it} \\ 1 \end{pmatrix}, \\ \frac{\partial^2 L}{\partial V(y) \partial V(z)} &= - \sum_{i=1}^M \sum_{t=1}^T \beta^2 p(x_{it}) [1 - p(x_{it})] [q(y|x_{it}, 0) - q(y|x_{it}, 1)] [q(z|x_{it}, 0) - q(z|x_{it}, 1)], \\ \frac{\partial^2 L}{\partial V(y) \partial \beta} &= - \sum_{i=1}^M \sum_{t=1}^T [q(y|x_{it}, 0) - q(y|x_{it}, 1)] \left\{ [a_{it} - p(x_{it})] + \beta^2 p(x_{it}) [1 - p(x_{it})] \right\}. \end{aligned}$$

The relevant first derivatives for the fixed-point equation are as follows:

$$\begin{aligned} \frac{\partial F(x)}{\partial \theta} &= [1 - p(x)] \begin{pmatrix} -x \\ 1 \end{pmatrix}, \\ \frac{\partial F(x)}{\partial V(y)} &= \beta \{q(y|x, 0)[1 - p(x)] + q(y|x, 1)p(x)\}, \\ \frac{\partial F(x)}{\partial \beta} &= Q_0(x)' V [1 - p(x)] + Q_1(x)' V p(x). \end{aligned}$$

The relevant second derivatives for the fixed-point equation are as follows:

$$\begin{aligned}
\frac{\partial^2 F(x)}{\partial \theta \partial \theta'} &= p(x)[1 - p(x)] \begin{pmatrix} x^2 & -x \\ -x & 1 \end{pmatrix}, \\
\frac{\partial^2 F(x)}{\partial \theta \partial V(y)} &= \beta[q(y|x, 0) - q(y|x, 1)]p(x)[1 - p(x)] \begin{pmatrix} -x \\ 1 \end{pmatrix}, \\
\frac{\partial^2 F(x)}{\partial \theta \partial \beta} &= p(x)[1 - p(x)][Q_0(x) - Q_1(x)]'V \begin{pmatrix} -x \\ 1 \end{pmatrix}, \\
\frac{\partial^2 F(x)}{\partial V(y) \partial V(z)} &= \beta^2[q(y|x, 0) - q(y|x, 1)][q(z|x, 0) - q(z|x, 1)]p(x)[1 - p(x)], \\
\frac{\partial^2 F(x)}{\partial V(y) \partial \beta} &= q(y|x, 0)[1 - p(x)] + q(y|x, 1)p(x) \\
&\quad + \beta[q(y|x, 0) - q(y|x, 1)]p(x)[1 - p(x)][Q_0(x) - Q_1(x)]'V.
\end{aligned}$$

3.2 Sensitivity of counterfactuals

In practice, researchers typically first estimate the parameters and then perform counterfactual analysis. Some common types of counterfactuals include changing the utility parameters (e.g., college subsidy program in [Keane and Wolpin \(1997\)](#)), changing the transition probabilities (e.g., demand volatility in [Collard-Wexler \(2013\)](#)), and changing the action and state space (e.g., eliminating patients' actions in [Crawford and Shum \(2005\)](#)). See [Kalouptsi et al. \(2021b\)](#) for a detailed discussion on various types of counterfactuals and some identification results for DDC models.

The local sensitivity of the above counterfactuals can be computed based on the estimates from Proposition 3.2. In particular, $\frac{\partial \theta}{\partial \gamma'}$ have already been computed. Then, the sensitivity for the counterfactual parameters $\tilde{\theta}$ and the counterfactual value function $\tilde{V}$ can be obtained as follows.

1. Suppose the counterfactual changes the utility parameters to $\tilde{\theta} \equiv H(\theta)$. Then, the sensitivity of the utility parameters under the counterfactual can be computed as

$$\frac{\partial \tilde{\theta}}{\partial \gamma'} = \frac{\partial H}{\partial \theta'} \frac{\partial \theta}{\partial \gamma'}. \quad (13)$$

The elasticity or semi-elasticity measures similar to the ones introduced at the beginning of Section 3.1 can also be easily computed for the counterfactuals. They can be computed by replacing $\hat{\theta}$ with $\tilde{\theta}$ in (10) and (11), and using (13).

2. For other types of counterfactuals where the utility parameters are unchanged, $\frac{\partial \tilde{\theta}}{\partial \gamma'} = \frac{\partial \theta}{\partial \gamma'}$ holds automatically.
3. Let $\tilde{V} = \tilde{F}(\tilde{\theta}, \tilde{V}; \gamma)$ be the updated equilibrium constraints under the counterfactuals. Depending on the counterfactual, there may be a different number of equations. Thus, the sensitivity of $\tilde{V}$ can be computed by substituting $\frac{\partial \tilde{\theta}}{\partial \gamma'}$, and solving the following system of equations:

$$\frac{\partial \tilde{F}}{\partial \theta'} \frac{\partial \tilde{\theta}}{\partial \gamma'} + \left(\frac{\partial \tilde{F}}{\partial V'} - I_{\tilde{d}_V} \right) \frac{\partial \tilde{V}}{\partial \gamma'} = - \frac{\partial \tilde{F}}{\partial \gamma'}.$$

3.3 Other economic applications

While this paper focuses on DDC models, many other economic applications also have a constrained optimization structure as in equation (8) with some parameters being fixed in the estimation procedure. Some details of the examples mentioned in Section 1 are as follows.

[Ossa \(2014\)](#) studies the effect of tariffs on welfare. The main optimization problem maximizes welfare subject to equilibrium constraints, such as budget constraints and market clearing conditions. Unlike the other applications, the objective is linear in the variables. The calibrated parameter is the elasticity of substitution of industry varieties σ_s . Sensitivity analysis with respect to calibrated parameters is conducted by resolving the optimization problem with different values of σ_s .

[Igami \(2017\)](#) studies creative destruction in the hard disk drive industry using a dynamic discrete game model. The model is solved using the nested fixed-point approach. Apart from the discount factor, the rate of change of innovation costs and the number of potential entrants are also fixed in the model. Sensitivity analysis is performed by estimating the full model again at different values of the three fixed parameters.

[Yang \(2021\)](#) studies aggregate productivity losses due to misallocation. The main optimization problem maximizes likelihood subject to firm optimality conditions. There are three calibrated parameters. They are the span of control, sectoral capital share, and conversion factor. Again, sensitivity analysis with respect to calibrated parameters is conducted by resolving the optimization problem at different calibrated parameters.

[Chen and Choo \(2023\)](#) study dynamic matching problems. Similar to dynamic discrete choice problems, the main optimization problem maximizes likelihood subject to fixed-point constraints. There are additional constraints that characterize matching equi-

libria. The discount factor is fixed at 0.95.

3.4 Connection with unconstrained optimization problems

In this section, I explain how Proposition 3.2 is related to unconstrained optimization problems with fixed parameters. This is relevant for solution methods such as Hotz and Miller (1993) and Aguirregabiria and Mira (2002) introduced in Section 2.3.

Write $V = \bar{F}(\theta; \gamma)$, so that V can be expressed as a known function of θ and γ . As a result, substituting V into the criterion function yields

$$L(\theta, V; \gamma) = L(\theta, \bar{F}(\theta; \gamma); \gamma) \equiv \bar{L}(\theta; \gamma).$$

Hence, optimization problem (8) becomes

$$\min_{\theta \in \Theta} \bar{L}(\theta; \gamma). \quad (14)$$

Here, the parameter of interest in problem (14) is θ and the relevant local sensitivity measure is $\frac{\partial \theta}{\partial \gamma}$. The following proposition shows how the local sensitivity measure can be computed and the connection with Proposition 3.2.

Proposition 3.4. *Consider the optimization problem (14). Assume that*

- $\bar{L}$ is continuously differentiable in θ and γ around $\hat{\theta}(\gamma)$.
- $\hat{\theta}(\gamma)$ is the unique solution to the optimization problem (14) and belongs to the interior of Θ for each $\gamma \in \Gamma$.

Then, the following statements hold:

1. If $\frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'}$ is invertible, the local sensitivity measure $\frac{\partial \theta}{\partial \gamma}$ can be obtained by solving

$$\frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'} \frac{\partial \theta}{\partial \gamma'} = - \frac{\partial^2 \bar{L}}{\partial \theta \partial \gamma'}. \quad (15)$$

2. In terms of the notations in Proposition 3.2,

$$\begin{aligned} \frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'} &= A_{\theta, \theta'} + \frac{\partial^2 L}{\partial \theta \partial V'} \frac{\partial F}{\partial \theta'} + \frac{\partial F}{\partial \theta'} \frac{\partial^2 L}{\partial V \partial \theta'} + \frac{\partial F}{\partial \theta'} \frac{\partial^2 L}{\partial V \partial V'} \frac{\partial F}{\partial \theta'}, \\ \frac{\partial^2 \bar{L}}{\partial \theta \partial \gamma'} &= A_{\theta, \gamma'} + \frac{\partial^2 L}{\partial \theta \partial V'} \frac{\partial F}{\partial \gamma'} + \frac{\partial F}{\partial \theta'} \frac{\partial^2 L}{\partial V \partial \gamma'} + \frac{\partial F}{\partial \theta'} \frac{\partial^2 L}{\partial V \partial V'} \frac{\partial F}{\partial \gamma'}. \end{aligned}$$

All the terms above are evaluated at the optimal solution, e.g., $\frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'} \equiv \frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^*}$.

The assumptions of Proposition 3.4 are similar to Assumption 3.1 but for unconstrained optimization problems. Part 1 of Proposition 3.4 follows immediately from total differentiating the first-order conditions of optimization problem (14). Part 2 of Proposition 3.4 shows that the local sensitivity measure for the unconstrained problem (14) can be expressed using the terms in Proposition 3.2 for constrained optimization problem (8). This shows Proposition 3.2 nests unconstrained optimization problems as a special case.

Remark 3.5. Jørgensen (2023) focuses on the following unconstrained optimization problem with a GMM objective:

$$\min_{\theta \in \Theta} g_n(\theta; \gamma)' W_n g_n(\theta; \gamma), \quad (16)$$

where $g_n(\theta; \gamma)$ is some vector-valued functions and W_n is a symmetric positive definite weighting matrix. The unconstrained optimization problem (16) can be written in terms of optimization problem (8) based on the preceding discussion. Let $\bar{L}(\theta; \gamma)$ be the objective function in equation (16). Since $\frac{\partial \bar{L}}{\partial \theta} = \left[\frac{\partial g_n(\theta; \gamma)}{\partial \theta} \right]' W_n g_n(\theta; \gamma)$, the expressions in equation (15) can be written as

$$\begin{aligned} \frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'} &= \left[\frac{\partial g_n(\theta; \gamma)}{\partial \theta} \right]' W_n \frac{\partial g_n(\theta; \gamma)}{\partial \theta'} + [g_n(\theta; \gamma)' W_n \otimes I_{d_\theta}] \frac{\partial \text{vec}\left\{ \left[\frac{\partial g_n(\theta; \gamma)}{\partial \theta} \right]' \right\}}{\partial \theta'}, \\ \frac{\partial^2 \bar{L}}{\partial \theta \partial \gamma'} &= \left[\frac{\partial g_n(\theta; \gamma)}{\partial \theta} \right]' W_n \frac{\partial g_n(\theta; \gamma)}{\partial \gamma'} + [g_n(\theta; \gamma)' W_n \otimes I_{d_\theta}] \frac{\partial \text{vec}\left\{ \left[\frac{\partial g_n(\theta; \gamma)}{\partial \theta} \right]' \right\}}{\partial \gamma'}, \end{aligned}$$

and they can be evaluated at the optimal solution of (16) for a given γ .

The above terms match the expressions in Proposition 1 of Jørgensen (2023). Thus, the local sensitivity measure in Proposition 3.2 also nests the unconstrained version in Jørgensen (2023). As discussed in Section 3.3, it is common for economic models to be estimated by constrained optimization with equilibrium constraints. The objective function is not necessarily a GMM objective function in empirical applications, as in Ossa (2014).

4 Global sensitivity analysis

Section 3 has provided a local sensitivity measure that is easy to compute and can be used as a good local approximation of the target parameter around its optimal value at the given fixed parameter. As mentioned in the introduction, researchers usually conduct sensitivity analysis by re-estimating the model at a few other values of the fixed parameter. Even though monotonic patterns are usually shown, they do not necessarily extend to the entire support of the fixed parameters.

Since the discount factor is the most commonly fixed parameter in DDC models, I start by studying what conditions on model primitives can imply that the flow utility is monotone in the discount factor in DDC models. Then, I focus on the linear-in-parameter utility specification that is common in practice. If the parameters are monotone in the discount factor, it is sufficient to estimate the model at the endpoints of the discount factor to obtain the bounds on the parameter. For more general models, the target parameter may not necessarily be monotone in the fixed parameter. As a result, I propose a constrained optimization approach to compute the bounds of the target parameter over a range of fixed parameters. In the context of DDC models and discount factors, the range of discount factors may be obtained by exclusion restrictions. See [Magnac and Thesmar \(2002\)](#) and [Abbring and Daljord \(2020\)](#) for exclusion restrictions and [Kong et al. \(2022\)](#) for a recent application that estimates the discount factor for various consumer goods.

4.1 Monotonicity of utility

In this section, I start from the representation in equation (4) that connects flow utility and CCP. This is motivated by two-step approaches that estimate CCP in the first step and the utility parameters in the second step.

Assume that the transition matrices Q_a and CCP p_a are available to the researchers for all $a \in \mathcal{A}$ in equation (4). In addition, consider the following normalization assumption that is standard in many applications.

Assumption 4.1. $\pi_A(x) = 0$ for any $x \in \mathcal{X}$.

Under Assumptions 2.4 and 4.1, the utility vector at state $a \in \mathcal{A} \setminus \{A\}$ is given by

$$\begin{aligned}\pi_a &= A_a \psi_A(p) - \psi_a(p) \\ &= (I_X - \beta Q_a)(I_X - \beta Q_A)^{-1}(-\log p_A) + \log p_a.\end{aligned}\tag{17}$$

The following proposition shows the sign of the flow utility with respect to the discount factor.

Proposition 4.2. *Let Assumptions 2.1 to 2.4 and 4.1 hold. The derivative of π_a in equation (17) with respect to β is*

$$\frac{\partial \pi_a}{\partial \beta} = -[Q_a(I_X - \beta Q_A)^{-1} - (I_X - \beta Q_A)^{-1}Q_a](I_X - \beta Q_A)^{-1}(-\log p_A),$$

for any $a \in \mathcal{A} \setminus \{A\}$.

The following corollary shows that normalizing the utility to another value gives a similar result.

Corollary 4.3. *Let Assumptions 2.1 to 2.4 hold and normalize the utility for choice A as $\pi_A = \bar{\pi}_A$. Then,*

$$\frac{\partial \pi_a}{\partial \beta} = -[Q_a(I_X - \beta Q_A)^{-1} - (I_X - \beta Q_A)^{-1}Q_a](I_X - \beta Q_A)^{-1}(\bar{\pi}_A - \log p_A),$$

for any $a \in \mathcal{A} \setminus \{A\}$.

The above corollary shows that a normalization that is different from the one in Assumption 4.1 only replaces $(-\log p_A)$ in Proposition 4.2 by $(\bar{\pi}_A - \log p_A)$ in Corollary 4.3. Therefore, the remaining results in this section will maintain Assumption 4.1.

Proposition 4.2 can be simplified under ρ -period finite dependence between actions a and A . For examples and discussion on finite dependence, see Arcidiacono and Miller (2011, 2019) and references therein.

Proposition 4.4. *Let Assumptions 2.1 to 2.4 and 4.1 hold. Suppose there exists $a \in \mathcal{A} \setminus \{A\}$ and $\rho \in \mathbb{N}$ such that $Q_a(x)Q_A^\rho = Q_A(x)Q_A^\rho$ for all $x \in \mathcal{X}$. Then,*

$$\frac{\partial \pi_a}{\partial \beta} = -(Q_a - Q_A)(I_X + \beta Q_A + \cdots + \beta^{\rho-1}Q_A^{\rho-1})(I_X - \beta Q_A)^{-1}(-\log p_A). \quad (18)$$

The following corollary shows that if $\rho = 1$, the RHS of equation (18) does not depend on β . As a result, the slope of flow utility for each state is constant across $\beta \in [0, 1)$.

Corollary 4.5. *Let Assumptions 2.1 to 2.4 and 4.1 hold. Suppose that there is one-period dependence between choices $a \in \mathcal{A} \setminus \{A\}$ and A for all states $x \in \mathcal{X}$. Then,*

$$\frac{\partial \pi_a}{\partial \beta} = (-Q_a + Q_A)(-\log p_A),$$

for any $a \in \mathcal{A} \setminus \{A\}$. As a result, for each $a \in \mathcal{A} \setminus \{A\}$ and $x \in \mathcal{X}$, $\pi_a(x)$ is either increasing, decreasing, or constant across in $\beta \in [0, 1)$.

One-period finite dependence is common in economics. This holds in models where one of the actions is a renewal or terminal action. A renewal action is an action that resets the states, such as the option to replace the engine in Rust (1987)'s bus engine model. A terminal action is a choice where the optimization problem is ended with no more future actions, such as Pakes et al. (2007)'s firm exit model. The following corollary summarizes the conditions for renewal actions.

Corollary 4.6. *Let Assumptions 2.1 to 2.4 and 4.1 hold. Suppose that $p_A(1) > 0$ and action A is a renewal action so that the corresponding transition matrix Q_A is given by*

$$Q_A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}.$$

1. If $p_A(1) \leq p_A(x)$ for any $x \in \mathcal{X}$, then π_a is nondecreasing in β for any $a \in \mathcal{A} \setminus \{A\}$.
2. If $p_A(1) \geq p_A(x)$ for any $x \in \mathcal{X}$, then π_a is nonincreasing in β for any $a \in \mathcal{A} \setminus \{A\}$.

The condition in Corollary 4.6 is easy to verify because it only depends on the CCP of action A and the presence of a renewal action.

While Corollaries 4.5 and 4.6 provide some convenient conditions under which the utility is monotone in β under some special conditions, it is important to note that the utilities can still be monotone in the discount factor whenever the transition probabilities and the CCP are such that the derivative in Proposition 4.2 is always positive or negative, even if the conditions in Corollaries 4.5 and 4.6 do not hold.

Although it is possible to obtain the monotonicity of utility under some additional conditions on the model primitives, characterizing the monotonicity for counterfactuals is more challenging due to the nonlinearity of the model. Let $\tilde{V}$ be the vector of counterfactual ex ante value functions. Let $\tilde{p}_a$ be the counterfactual CCP, $\tilde{Q}_a$ be the counterfactual transition matrix, $\tilde{\pi}_a$ be the counterfactual utility at action $a \in \mathcal{A}$. Assume that $\tilde{\pi}_a$ is related to π_a through the following affine transformation

$$\tilde{\pi}_a = H_a \pi_a + g_a,$$

for some $X \times X$ matrix H_a and length X vector g_a that are independent of β . If $\tilde{\pi}_A = 0$,

then $\tilde{\pi}_a$ can be written as

$$H_a \pi_a + g_a = \tilde{\pi}_a = (I_X - \beta \tilde{Q}_a)(I_X - \beta \tilde{Q}_A)^{-1}(-\log \tilde{p}_A) + \log \tilde{p}_a, \quad (19)$$

for any $a \in \mathcal{A} \setminus \{A\}$ using (17). Taking the derivative with respect to β on both sides yields

$$\begin{aligned} H_a \frac{\partial \pi_a}{\partial \beta} = & -[\tilde{Q}_a(I_X - \beta \tilde{Q}_A)^{-1} - (I_X - \beta \tilde{Q}_A)^{-1} \tilde{Q}_a](I_X - \beta \tilde{Q}_A)^{-1}(-\log \tilde{p}_A) \\ & + (I_X - \beta \tilde{Q}_a)(I_X - \beta \tilde{Q}_A)^{-1} \left(-\frac{1}{\tilde{p}_A} \odot \frac{\partial \tilde{p}_A}{\partial \beta} \right) + \frac{1}{\tilde{p}_a} \odot \frac{\partial \tilde{p}_a}{\partial \beta}, \end{aligned} \quad (20)$$

for any $a \in \mathcal{A} \setminus \{A\}$, where $\odot$ denotes Hadamard product (entrywise product).

Example 4.7. Consider the Rust (1987) model again, so that $\mathcal{A} = \{0, 1\}$, $A = 1$, and $\tilde{p}_1(x) = 1 - \tilde{p}_0(x)$ for each $x \in \mathcal{X}$. Then, equation (19) can be used to characterize the relationship between counterfactual CCP as follows

$$H_0 \pi_0 + g_0 = [I_X + \beta(\tilde{Q}_1 - \tilde{Q}_0)](-\log \tilde{p}_1) + \log \tilde{p}_0.$$

Thus, equation (20) can be written as

$$H_0 \frac{\partial \pi_0}{\partial \beta} = (\tilde{Q}_1 - \tilde{Q}_0)(-\log \tilde{p}_1) - [I_X + \beta(\tilde{Q}_1 - \tilde{Q}_0)] \left(\frac{1}{\tilde{p}_1} \odot \frac{\partial \tilde{p}_1}{\partial \beta} \right) - \frac{1}{\tilde{p}_0} \odot \frac{\partial \tilde{p}_1}{\partial \beta}.$$

where $\tilde{p}_0 \equiv 1_X - \tilde{p}_1$.

The above shows the relationship between the derivatives of counterfactuals with respect to the discount factor in DDC models is nonlinear and may not be easy to characterize.

Remark 4.8. Although the above results maintain the T1EV assumption on the unobservables due to Assumption 2.4, the same analysis can be applied to other distributions of unobservables. This is because the mapping $\psi_a(p)$ in equation (17) depends on the distribution of the unobservables and the CCP p . The distribution of unobservables is independent of the discount factor, and the CCP is assumed to be known. Hence, $\psi_a(p)$ does not depend on the discount factor. As a result, the results can still be applied by replacing the vector $(-\log p_A)$ by $\psi_A(p)$ for other distributions of the unobservables.

4.2 Monotonicity with linear-in-parameters utility

In this section, I connect the results in the previous section to the linear-in-parameters specification of the utility function that is common in empirical applications.

Assumption 4.9 (Linear-in-parameters). *For each $a \in \mathcal{A} \setminus \{A\}$, the flow utility at state $x \in \mathcal{X}$ is given by*

$$\pi_a(x) = \omega_a(x)' \theta,$$

where $\theta \in \mathbb{R}^{d_\theta}$ and $\omega_a(x) \equiv (\omega_{a,1}(x), \dots, \omega_{a,d_\theta}(x))'$ are the corresponding coefficients on θ .

Let Π_a be the $(X \times d_\theta)$ matrix that collects the coefficients on θ over all the states at $a \in \mathcal{A} \setminus \{A\}$, so the x -th row represents $\omega_a(x)'$ for each $a \in \mathcal{A} \setminus \{A\}$. Then, the vector of utilities for each $a \in \mathcal{A} \setminus \{A\}$ can be written as

$$\pi_a = \Pi_a \theta, \tag{21}$$

based on Assumption 4.9.

Let $\hat{p}_a$ be the estimator of the CCP for action $a \in \mathcal{A} \setminus \{A\}$. Then, $\hat{\pi}_a$ can be estimated via equation (17). Let $\hat{\pi}$ and Π be the matrices that stack $\hat{\pi}_a$ and Π_a over all $a \in \mathcal{A} \setminus \{A\}$. Then, the utility parameters can be obtained by the following minimum distance problem

$$\min_{\theta \in \Theta} (\hat{\pi} - \Pi \theta)' W (\hat{\pi} - \Pi \theta), \tag{22}$$

where W is a positive definite weighting matrix independent of the discount factor.

The following proposition shows the sensitivity of utility parameters estimated from optimization problem (22) with respect to the discount factor.

Proposition 4.10. *Let Assumptions 2.1 to 2.4, 4.1 and 4.9 hold. Let $\hat{\theta}$ be the parameter estimated from problem (22) and assume that the matrix $\Pi' W \Pi$ is invertible. Then, the derivative of $\hat{\theta}$ with respect to β is given by*

$$\frac{\partial \hat{\theta}}{\partial \beta} = (\Pi' W \Pi)^{-1} \Pi' W \frac{\partial \hat{\pi}}{\partial \beta}.$$

Proposition 4.10 shows that the sensitivity of the utility parameters with respect to β depends on Π and W , in addition to $\frac{\partial \hat{\pi}}{\partial \beta}$. Hence, even if $\hat{\pi}$ is monotone in β , the choice of the parameterization and the weighting matrix can also affect whether the utility parameters are monotone in β . The following corollary shows a special case where the weighting matrix is an identity matrix and when the same parameter does not appear in more than one choice.

Corollary 4.11. *Let Assumptions 2.1 to 2.4, 4.1 and 4.9 hold. Suppose that*

1. θ is partitioned as $\theta = (\theta'_0, \dots, \theta'_{A-1})'$ with $\theta_a \in \mathbb{R}^{d_a}$ for each $a \in \mathcal{A} \setminus \{A\}$ and $d_\theta = \sum_{a \in \mathcal{A} \setminus \{A\}} d_a$.
2. $\pi_a = \Pi_a \theta_a$ and Π_a has full rank.
3. W is an identity matrix.

Then,

$$\frac{\partial \hat{\theta}_a}{\partial \beta} = (\Pi'_a \Pi_a)^{-1} \Pi'_a \frac{\partial \hat{\pi}_a}{\partial \beta},$$

for each $a \in \mathcal{A} \setminus \{A\}$.

The above discussion has focused exclusively on the global sensitivity analysis for the discount factor. In practice, researchers may fix some other parameters in the utility function in the estimation procedure. An example is the rate of change of innovation cost that Igami (2017) calibrates in the coefficient for the sunk cost parameter that he estimates. Let $\delta \in \mathbb{R}$ be another parameter that the researcher calibrates apart from the discount factor. The following proposition shows the sensitivity of the utility parameters with respect to δ estimated from problem (22) when part of Π contains δ .

Proposition 4.12. *Let Assumptions 2.1 to 2.4, 4.1 and 4.9 hold. Suppose that part of the matrix Π contains fixed parameter $\delta \in \mathbb{R}$, the matrix $\Pi' W \Pi$ is invertible, and that $\hat{\theta}$ is estimated from problem (22). Then, the derivative of $\hat{\theta}$ with respect to δ is given by*

$$\frac{\partial \hat{\theta}}{\partial \delta} = -(\Pi' W \Pi)^{-1} \left[\Pi' (W' + W) \frac{\partial \Pi}{\partial \delta} \hat{\theta} - \left(\frac{\partial \Pi}{\partial \delta} \right)' W \hat{\pi} \right].$$

Whether monotonicity can be established for other fixed parameters depends on the exact structure of the parameterization. This motivates the estimation approach for global sensitivity analysis in the next section.

4.3 Other approaches

Researchers can still estimate the bounds on the target parameter over a certain interval of the fixed parameter through a constrained optimization problem. Following Assumption 3.1 that the solution is in the interior and that the optimal solution is a regular point, the first-order condition is satisfied at the optimum. Let $\tau(\theta, V; \gamma)$ be the target parameter the researcher wishes to estimate. Then, the bounds on the target parameter over a pre-specified range of fixed parameters $\gamma \in \bar{\Gamma} \subseteq \Gamma$ can be estimated by the following

optimization problem that estimates the bounds subject to the first-order conditions

$$\begin{aligned}
& \min/\max_{\theta \in \Theta, V \in \mathcal{V}, \gamma \in \bar{\Gamma}, \lambda \in \mathbb{R}^{d_V}} \tau(\theta, V; \gamma) \\
& \text{s.t.} \quad \frac{\partial Q}{\partial \theta} - \left(\frac{\partial F}{\partial \theta'} \right)' \lambda = 0 \\
& \quad \frac{\partial Q}{\partial V'} + \left(I_{d_V} - \frac{\partial F}{\partial V'} \right)' \lambda = 0 \\
& \quad V = F(\theta, V; \gamma).
\end{aligned} \tag{23}$$

If the researcher is interested in estimating the bounds on the target parameter that depends on the counterfactuals with $\tilde{\theta} = H(\theta)$, then optimization problem (23) can be updated as

$$\begin{aligned}
& \min/\max_{\theta \in \Theta, V \in \mathcal{V}, \tilde{V} \in \tilde{\mathcal{V}}, \gamma \in \bar{\Gamma}, \lambda \in \mathbb{R}^{d_V}} \tau(\theta, V, \tilde{V}; \gamma) \\
& \text{s.t.} \quad \frac{\partial Q}{\partial \theta} - \left(\frac{\partial F}{\partial \theta'} \right)' \lambda = 0 \\
& \quad \frac{\partial Q}{\partial V'} + \left(I_{d_V} - \frac{\partial F}{\partial V'} \right)' \lambda = 0 \\
& \quad V = F(\theta, V; \gamma) \\
& \quad \tilde{V} = \tilde{F}(H(\theta), \tilde{V}; \tilde{\gamma}).
\end{aligned}$$

The computational time by treating γ as a variable depends on $\bar{\Gamma}$ and the structure of the problem. The next section explores this using the Rust (1987) model.

Example 4.13. In terms of the Rust (1987) model, suppose the researcher is interested in the bounds on the target parameter $\tau(\theta, V; \beta)$ over $\beta \in [\underline{\beta}, \bar{\beta}] \equiv \mathcal{B}$. Then, optimization problem (23) can be implemented as follows:

$$\begin{aligned}
& \min/\max_{(\text{MC,RC}) \in \Theta, V \in \mathcal{V}, \beta \in \mathcal{B}, \lambda \in \mathbb{R}^{d_V}} \tau(\theta, V; \beta) \\
& \text{s.t.} \quad - \sum_{i=1}^M \sum_{t=1}^T [a_{it} - p(x_{it})] x_{it} - \sum_{j=1}^{d_V} \lambda_j [1 - p(j)] (-j) = 0 \\
& \quad - \sum_{i=1}^M \sum_{t=1}^T [a_{it} - p(x_{it})] - \sum_{j=1}^{d_V} \lambda_j [1 - p(j)] = 0 \\
& \quad - \beta \sum_{i=1}^M \sum_{t=1}^T [a_{it} - p(x_{it})] [q(y|x_{it}, 0) - q(y|x_{it}, 1)]
\end{aligned}$$

$$\begin{aligned}
& + \lambda_y - \beta \sum_{j=1}^{d_Y} \lambda_j \{q(y|j, 0)[1 - p(j)] + q(y|j, 1)p(j)\} \\
& \hspace{25em} \text{for all } y \in \mathcal{X} \\
p(x) &= \frac{1}{1 + \exp\{\text{RC} - \text{MC}x + \beta(Q_0(x) - Q_1(x))'V\}} \\
& \hspace{25em} \text{for all } x \in \mathcal{X} \\
V(x) &= \log \{ \exp[\text{RC} - \text{MC}x + \beta Q_0(x)'V] + \exp[\beta Q_1(x)'V] \} \\
& \hspace{25em} \text{for all } x \in \mathcal{X}
\end{aligned}$$

The second and third equations above correspond to the first-order conditions with respect to the two utility parameters. The fourth equation corresponds to the first-order condition with respect to the x -th component of the value function. The fifth equation is the equation of the choice probability (1). The last equation corresponds to the fixed-point equation for the Bellman equation as in (3). The above problem can be implemented in standard software such as Knitro. Researchers can also easily incorporate multi-start in Knitro to optimize the bounds to search for the global optimum.

Another avenue for conducting global sensitivity analysis is to conduct breakdown analysis (Horowitz and Manski, 1995; Kline and Santos, 2013; Masten and Poirier, 2020). In terms of the fixed parameters in DDC models, breakdown analysis can be used to find the range of fixed parameters such that a certain conclusion holds. More precisely, suppose that the researcher is interested in knowing whether the target parameter is above a threshold, i.e., $\hat{\tau}(\gamma) \geq \tau^*$. Then, the robust region (RR) is the region such that the conclusion holds and is defined as

$$\text{RR} \equiv \{\gamma \in \Gamma : \hat{\tau}(\gamma) \geq \tau^*\},$$

and the breakdown frontier is the boundary of the robust region. If one further restricts attention to the discount factor so that $\gamma = \beta$, and that $\hat{\tau}(\beta)$ is monotone in β , the breakdown frontier can be found by the bisection method.

5 Empirical applications

In this section, I perform sensitivity analysis of two empirical applications with respect to the discount factor. The first application is the seminal bus engine replacement model in Rust (1987). The second application is a recent dynamic marriage matching model from Chen and Choo (2023).

5.1 Application 1: Rust (1987)

In this section, I conduct local and global sensitivity analysis for the Rust (1987) model using the data from the group 4 buses. I assume that the utility function is the same as in Section 2.4, and I specify the cost function as $c(x, MC) = -MCx$.

5.1.1 Local analysis

This section conducts local sensitivity analysis of various parameters with respect to the discount factor. All estimates in this section are obtained by the NFXP. The parameters I consider are as follows:

1. Replacement cost.
2. Maintenance cost.
3. Counterfactual conditional choice probability at state 90.
4. Change in average welfare as defined by $W \equiv \frac{1}{X} \sum_{x=1}^X [\tilde{V}(x) - V(x)]$.

For parameters 3 and 4 above, the counterfactual I consider is a reduction in maintenance cost of 10%. To analyze the sensitivity of the parameters at different values of the discount factor, I estimate each parameter using the following discount factors β : 0.85, 0.9, 0.95, 0.99, 0.999, and 0.9999. Rust (1987) uses $\beta = 0.9999$ in the main results. Then, I estimate the local sensitivity measure using the elasticity measure as in (10) so that the sensitivity measure can be comparable across different target parameters. Using the point estimate and the local sensitivity measure, I approximate the parameter at another discount factor and report the percentage approximation error. The results are reported in Table 1.

The local sensitivity measure in Table 1 shows that not all parameters are equally sensitive to the choice of the discount factor. On the other hand, the magnitude of the elasticity is increasing in the discount factor for all four parameters. The last three columns of Table 1 report the performance when the estimate and the local sensitivity measure obtained at β is used for local approximation for the target parameter when the discount factor is changed to $\beta' = \beta - \Delta\beta$. Except for the change in average welfare when the discount factor is close to 1, local approximation gives a less than 1% approximation error in most cases. An intuition for the larger error in the change in average welfare is that the value function can be interpreted as an infinite sum of the discounted present payoffs.

Table 1 suggests that researchers should be cautious with the sensitivity of the param-

Table 1: Local sensitivity analysis for Rust (1987).

β	Estimate	Elasticity	Approx. error (in %) for $\beta' = \beta - \Delta\beta$		
			$\Delta\beta = 0.0001$	$\Delta\beta = 0.001$	$\Delta\beta = 0.01$
Replacement cost					
0.8500	7.982	0.246	-1.93×10^{-5}	-1.68×10^{-4}	-0.018
0.9000	8.151	0.540	-1.55×10^{-5}	-0.001	-0.049
0.9500	8.572	1.566	8.00×10^{-6}	-0.002	-0.193
0.9900	9.640	4.893	-1.28×10^{-4}	-0.009	-0.800
0.9990	10.147	6.541	-1.50×10^{-4}	-0.012	-1.146
0.9999	10.208	6.744	-8.85×10^{-5}	-0.013	-1.195
Maintenance cost					
0.8500	0.012	-4.788	-5.22×10^{-5}	-5.76×10^{-6}	-0.004
0.9000	0.009	-6.973	-2.41×10^{-5}	-1.29×10^{-4}	-0.010
0.9500	0.005	-11.597	5.08×10^{-5}	-0.001	-0.052
0.9900	0.003	-20.712	-1.42×10^{-4}	-0.005	-0.379
0.9990	0.002	-24.235	-1.11×10^{-4}	-0.009	-0.633
0.9999	0.002	-24.619	-9.66×10^{-5}	-0.010	-0.675
Counterfactual choice probability at state 90					
0.8500	0.097	-1.151	-1.06×10^{-4}	4.51×10^{-4}	0.040
0.9000	0.089	-1.856	-2.72×10^{-5}	0.001	0.072
0.9500	0.078	-3.355	8.37×10^{-5}	0.002	0.142
0.9900	0.064	-6.175	-7.59×10^{-5}	0.003	0.298
0.9990	0.061	-7.271	3.63×10^{-5}	0.004	0.339
0.9999	0.060	-7.382	-3.73×10^{-5}	0.003	0.333
Change in average welfare					
0.8500	0.310	0.332	-3.78×10^{-5}	-3.91×10^{-4}	-0.039
0.9000	0.322	1.127	-3.41×10^{-5}	-0.002	-0.156
0.9500	0.373	5.937	-1.02×10^{-4}	-0.014	-1.277
0.9900	0.927	76.967	-0.008	-0.769	-63.845
0.9990	7.480	973.503	-0.972	-95.023	-7693.629
0.9999	73.086	9973.111	-99.484	-9720.724	-784775.081

eters when the discount factor is large. The discount factor is typically higher when the time between two consecutive periods is lower, or when agents discount the future less. Nevertheless, unless the discount factor is very close to 1, local approximation provides a good approximation of the model parameters without the need to re-estimate the full model.

5.1.2 Global analysis

In this section, I study the global properties of various target parameters. Figure 1 shows the estimates of the maintenance and replacement cost against the discount factor using NFXP and the minimum distance two-step method as in equation (22). The two-step method in equation (22) requires the CCP to be estimated by data. If the data is rich enough, then the CCP can be estimated by the simple frequency estimator. Since not all 90 states can be observed in the data, I estimate the CCP using a simple logistic regression. I regress the observed action on the state and the square of the state.

The lines in Figure 1 are obtained by estimating the utility parameters for each discount factor in the interval $[0, 0.9999]$ with grid points of size 0.0001. The maintenance cost is strictly decreasing in the discount factor, while the replacement cost is strictly increasing in the discount factor.

Figure 1: Utility parameter estimates against the discount factor.

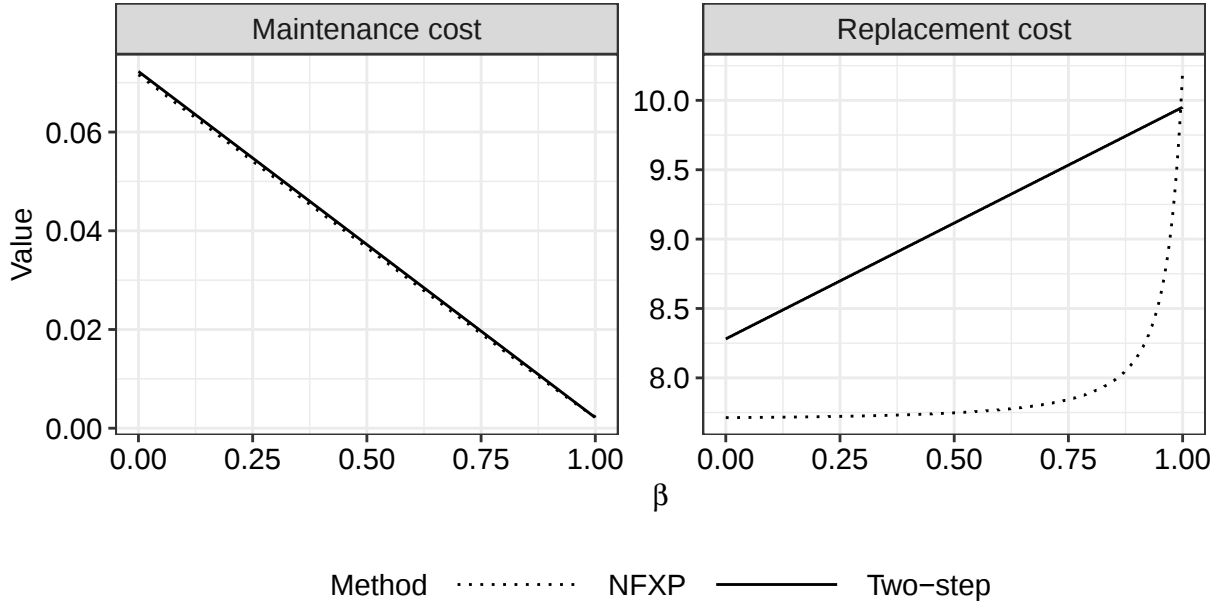


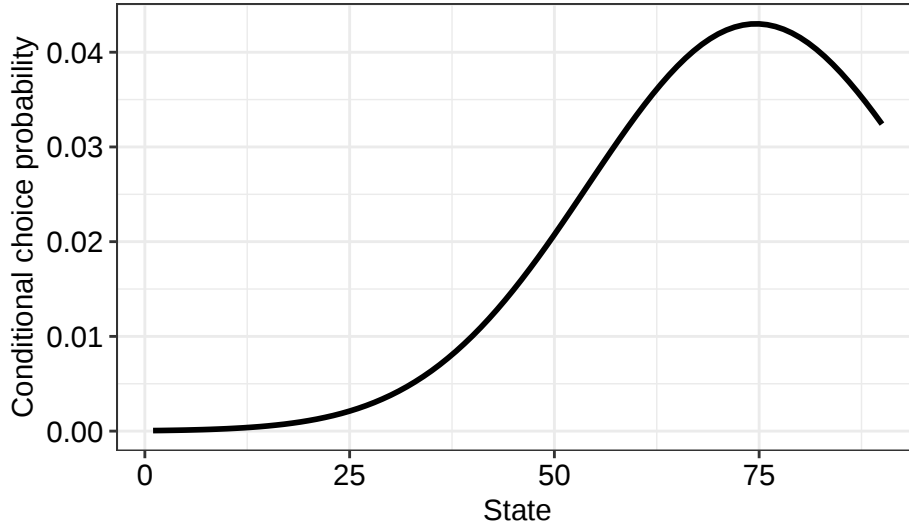
Figure 2 shows that the estimated CCP satisfies the conditions of Corollary 4.6. On the other hand, the option to replace the bus engine is a renewal action. Hence, the flow utility is monotonically increasing in the discount factor. With the cost function as specified at the beginning of this section, let $\theta = (MC, RC)$. Then, the matrix of utility

coefficients can be written as

$$\Pi_0 = \begin{pmatrix} -1 & 1 \\ -2 & 1 \\ \vdots & \vdots \\ -90 & 1 \end{pmatrix}.$$

Applying Corollary 4.11, it follows that $\frac{\partial \widetilde{MC}}{\partial \beta} < 0$ and $\frac{\partial \widetilde{RC}}{\partial \beta} > 0$.

Figure 2: Estimated probability of engine replacement.



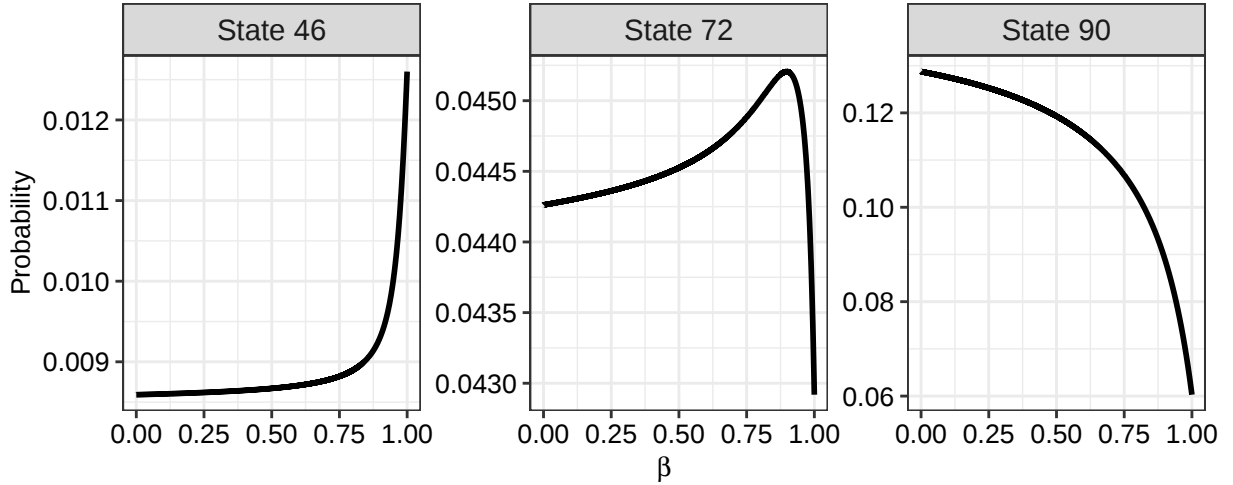
Although the utility parameters are monotone in the discount factor, the counterfactuals may not necessarily be monotone in the discount factor. Consider again the counterfactual of reducing the maintenance cost by 10%. Figure 3 shows the counterfactual CCP at three different states. It can be seen that the probabilities can exhibit different properties against the discount factor in Figure 3. Indeed, this can be seen by total differentiating the counterfactual CCP at state x , denoted as $\tilde{p}(x)$, with respect to the discount factor as follows

$$\frac{\partial \tilde{p}(x)}{\partial \beta} = \underbrace{-\tilde{p}(x)(1 - \tilde{p}(x))}_{(a)} \left\{ \underbrace{-\frac{\partial \widetilde{MC}}{\partial \beta} x + \frac{\partial \widetilde{RC}}{\partial \beta}}_{(b)} + \underbrace{[Q_0(x) - Q_1(x)]' \left(\tilde{V} + \beta \frac{\partial \tilde{V}}{\partial \beta} \right)}_{(c)} \right\}, \quad (24)$$

for each $x \in \mathcal{X}$. In equation (24), (a) is positive. However, (b) and (c) have opposite signs in this model and data. As a result, $\frac{\partial \tilde{p}(x)}{\partial \beta}$ is an average of positive and negative terms. Note that β also enters the expression in (24) explicitly. Hence, the CCP is not necessarily monotone in β . Figure 3 demonstrates the counterfactual CCPs at three different states.

Although the counterfactual CCPs demonstrate different shapes, their range is rather small.

Figure 3: Counterfactual conditional choice probabilities at states 46, 72, and 90.



Finally, I explore the performance of using the constrained optimization problem in Example 4.13 to find the bounds on the target parameter over a range of discount factors. Table 2 shows the bounds on the utility parameters and counterfactuals over different intervals of β . For all rows, I set the lower bound of the interval of β as 0.7 with different upper bounds. The computational time is the total time in running the optimization problem for the lower and upper bounds using Knitro. An attractive feature of the constrained optimization approach is that the optimizer would search for the optimal values so researchers do not need to choose the grid points over the support of the discount factor and re-estimate the model at each of the points.

5.2 Application 2: Chen and Choo (2023)

Chen and Choo (2023) develop a dynamic marriage matching framework that extends the methodology in Choo (2015). The empirical example in Chen and Choo (2023) studies how China's one-child policy affects the marriage distribution. The model has a constrained optimization structure, and they estimate their model using the nested fixed-point algorithm. The discount factor is fixed at $\beta = 0.95$ throughout their analysis. I illustrate that the methodology in Section 3 can be applied in this dynamic matching application, although it contains several additional components when compared to the DDC model described in Section 2.

Table 2: Bounds on two counterfactual target parameters as in Example 4.13 (the lower bound on β is 0.7).

Upper bound on β	Bounds	Computational time (in seconds)
Replacement cost		
0.8	[7.810, 7.895]	23.738
0.9	[7.810, 8.151]	19.701
0.95	[7.810, 8.572]	24.095
Counterfactual choice probability at state 90		
0.8	[0.102, 0.110]	20.817
0.9	[0.089, 0.110]	49.685
0.95	[0.078, 0.110]	57.942

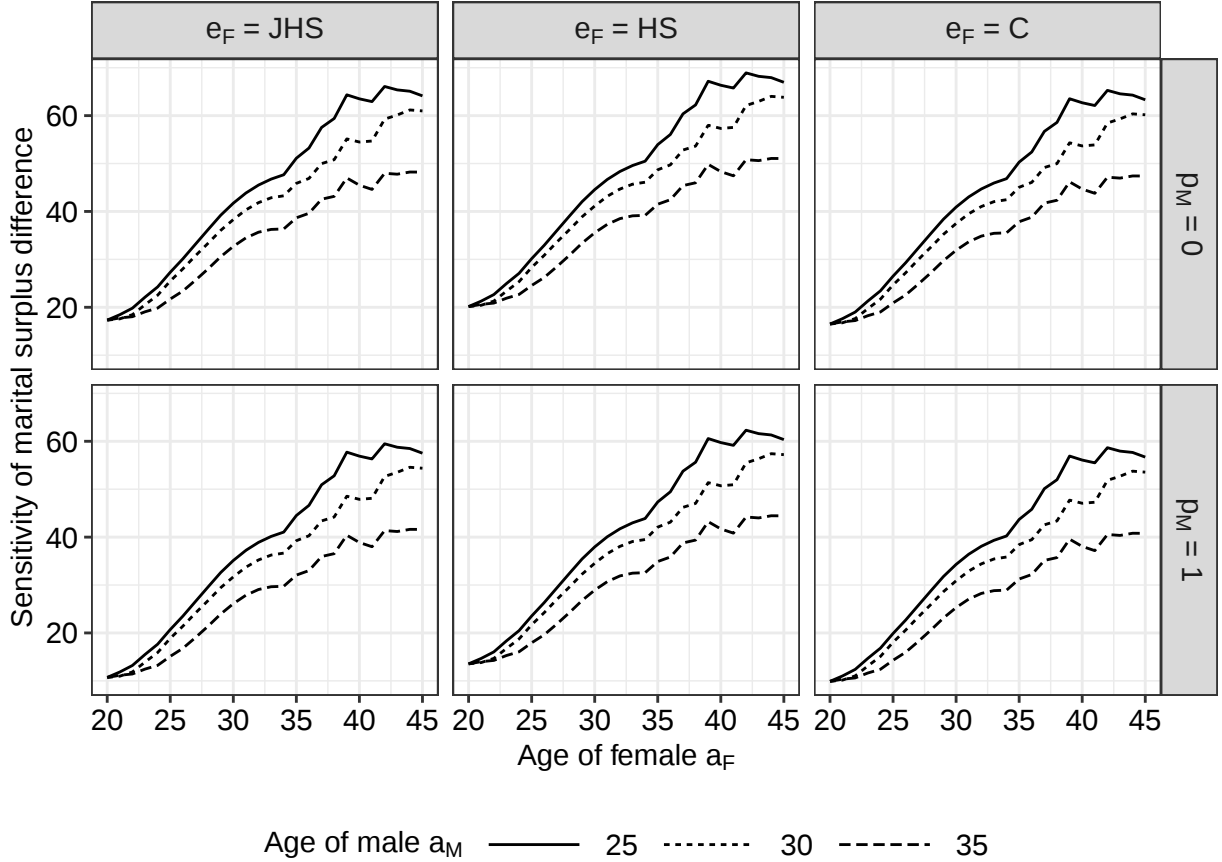
I focus on the analysis in Figure 2 of [Chen and Choo \(2023\)](#), where they examine how the types of couples affect marital surplus. In their model, individuals differ by education, previous marital status, and age. The education level can be junior high school (JHS), high school (HS), or college (C). Individuals can be previously married and divorced or not married. They report results of ages from 20 to 45 in their Figure 2.

Table 2 of [Chen and Choo \(2023\)](#) compares the surplus differences $\Pi_1(a_M, p_M) - \Pi_0(a_M, p_M, a_F, e_F)$. $\Pi_1(a_M, p_M)$ is the marriage surplus for an age a_M male with education level JHS, and marital status p_M and an age 20 female with education level JHS, and previously not married. $\Pi_0(a_M, p_M, a_F, e_F)$ is the marriage surplus for a male of the same type as in $\Pi_1(a_M, p_M)$ and an age a_F female with education level e_F and previously married and divorced. The support of the parameters they consider are $a_F \in \{20, 21, \dots, 45\}$, $a_M \in \{25, 30, 35\}$, $e_F \in \{JHS, HS, C\}$, and $p_M \in \{0, 1\}$ that equals 1 if the male is previously married and divorced. [Chen and Choo \(2023\)](#) find that couples with similar types generally have a similar surplus.

Figure 4 shows the sensitivity measure of marital surplus differences as in equation (9) for different types of individuals, arranged in the same format as in Figure 2 of [Chen and Choo \(2023\)](#). The terms required in Proposition 3.2 are computed using automatic differentiation. The sensitivity measure can be interpreted as the unit change in marital surplus for a unit change in the discount factor β . It can be seen that the sensitivity of the parameter estimates is not constant across different specifications and is generally increasing in a_F .

Finally, I examine the performance of the sensitivity measure as a local approxima-

Figure 4: Sensitivity measure for $\Pi_1(a_M, p_M) - \Pi_0(a_M, p_M, a_F, e_F)$ estimated at $\beta = 0.95$.



tion. To do this, I re-estimate the full model at other values of the discount factor and estimate the surplus difference as in Figure 2 of [Chen and Choo \(2023\)](#). Then, I approximate the surplus difference at another discount factor using the local sensitivity measure estimated at $\beta = 0.95$. Table 3 reports the summary statistics on the absolute error when local approximation is used to approximate the surplus differences at other discount factors. In the table, P_x refers to the x -th percentile. The magnitudes of the absolute errors remain small across different β and are mostly less than 1. On the other hand, Table 4 reports the summary statistics of the absolute percentage error for the approximation. It is important to note that the large percentage errors are due to the actual marital surplus differences being very close to 0. See Appendix Figure B.1 that shows the large percentage errors are all associated with actual marital surplus differences very close to 0. This exercise shows that the local approximation can be accurate in estimating the surplus without estimating the entire model again at another discount factor.

Table 3: Absolute approximation error of approximating marital surplus using the point estimates at $\beta = 0.95$ and the sensitivity measure.

β	Summary statistics							
	Min.	P_{25}	P_{50}	P_{75}	P_{90}	P_{95}	P_{99}	Max.
0.930	0.001	0.049	0.066	0.080	0.090	0.094	0.102	0.107
0.935	0.007	0.027	0.035	0.042	0.046	0.050	0.053	0.057
0.940	7.36×10^{-5}	0.023	0.037	0.054	0.072	0.110	0.189	0.207
0.945	1.13×10^{-5}	0.002	0.004	0.006	0.008	0.009	0.014	0.015
0.955	1.32×10^{-4}	0.005	0.013	0.036	0.065	0.089	0.128	0.131
0.960	5.96×10^{-5}	0.010	0.021	0.060	0.111	0.149	0.206	0.213
0.965	3.72×10^{-4}	0.019	0.032	0.053	0.085	0.121	0.184	0.201
0.970	1.65×10^{-4}	0.070	0.157	0.423	0.826	1.043	1.500	1.588

Table 4: Absolute approximate percentage error (in %) of approximating marital surplus using the point estimates at $\beta = 0.95$ and the sensitivity measure.

β	Summary statistics							
	Min.	P_{25}	P_{50}	P_{75}	P_{90}	P_{95}	P_{99}	Max.
0.930	0.004	0.342	0.674	1.500	3.916	5.025	20.027	337.273
0.935	0.023	0.192	0.330	0.705	1.848	2.477	12.738	43.120
0.940	0.001	0.187	0.420	0.879	1.919	3.186	19.946	87.474
0.945	1.03×10^{-4}	0.017	0.038	0.105	0.274	0.389	1.573	10.163
0.955	0.004	0.089	0.189	0.300	0.554	0.820	1.704	7.955
0.960	0.002	0.124	0.295	0.485	0.821	1.333	3.737	10.974
0.965	0.002	0.171	0.335	0.680	1.679	3.462	12.618	760.068
0.970	0.002	0.955	2.204	3.482	5.982	12.521	49.827	919.764

6 Conclusion

In this paper, I propose two procedures to conduct sensitivity analysis for parameter estimates with respect to fixed parameters in DDC models. First, the local sensitivity measure reports the change in the target parameter estimated from a constrained optimization problem for a unit change in the fixed parameter. This measure is fast to

compute, does not require model re-estimation at another value of the fixed parameter, and nests unconstrained estimation problems. The methodology in this paper can be applied to more general estimation problems, and not necessarily DDC models. For global sensitivity analysis, I examine whether target parameters are monotone in the fixed parameters. Using the discount factor as the leading example of fixed parameters in DDC models, I provide conditions under which utility is monotone in the discount factor.

From the empirical examples, I find that the estimates are typically more sensitive when the discount factor is closer to 1. On the other hand, I also show that the local sensitivity measure can serve as a good local approximation to the target parameter without re-estimating the full model. Hence, researchers can report the local sensitivity measure in addition to the point estimates to increase the transparency of structural research.

Appendix

The appendix contains all the proofs of the main text, and supplemental details on the empirical example on [Chen and Choo \(2023\)](#).

A Proofs

Proof of Proposition 3.2. Consider the optimization problem (8). Let λ be the Lagrange multiplier. Then, the Lagrangian can be written as

$$\mathcal{L} = L(\theta, V; \gamma) + \lambda' [V - F(\theta, V; \gamma)]. \quad (\text{A.1})$$

By Assumption 3.1, the following first-order conditions must hold at the optimum:

$$0 = \frac{\partial \mathcal{L}}{\partial \theta'} = \frac{\partial L}{\partial \theta} - \left(\frac{\partial F}{\partial \theta'} \right)' \lambda, \quad (\text{A.2})$$

$$0 = \frac{\partial \mathcal{L}}{\partial V'} = \frac{\partial L}{\partial V'} + \left(I_{d_V} - \frac{\partial F}{\partial V'} \right)' \lambda, \quad (\text{A.3})$$

$$0 = \frac{\partial \mathcal{L}}{\partial \lambda'} = V - F. \quad (\text{A.4})$$

The following result from [Magnus and Neudecker \(1985, Theorem 9\)](#) is useful for the remainder of the proof. Let $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{r \times p}$. Then, the Jacobian with respect to X can be written as

$$\frac{\partial \text{vec}(UV)}{\partial [\text{vec}(X)]'} = (V' \otimes I_m) \frac{\partial \text{vec}(U)}{\partial [\text{vec}(X)]'} + (I_p \otimes U) \frac{\partial \text{vec}(V)}{\partial [\text{vec}(X)]'}, \quad (\text{A.5})$$

where $\otimes$ denotes the Kronecker product.

The next step is to total differentiate equations (A.2) to (A.4) with respect to γ . Using equation (A.5) gives

$$\begin{aligned} 0 &= \frac{\partial^2 L}{\partial \theta \partial \theta'} \frac{\partial \theta}{\partial \gamma'} + \frac{\partial^2 L}{\partial \theta \partial V'} \frac{\partial V}{\partial \gamma'} + \frac{\partial^2 L}{\partial \theta \partial \gamma'} - \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial \lambda}{\partial \gamma'} \\ &\quad - (\lambda' \otimes I_{d_\theta}) \left\{ \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial \theta'} \frac{\partial \theta}{\partial \gamma'} + \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial V'} \frac{\partial V}{\partial \gamma'} + \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial \gamma'} \right\}, \\ 0 &= \frac{\partial^2 L}{\partial V \partial \theta'} \frac{\partial \theta}{\partial \gamma'} + \frac{\partial^2 L}{\partial V \partial V'} \frac{\partial V}{\partial \gamma'} + \frac{\partial^2 L}{\partial V \partial \gamma'} + \left(I_{d_V} - \frac{\partial F}{\partial V'} \right)' \frac{\partial \lambda}{\partial \gamma'} \end{aligned}$$

$$\begin{aligned}
& -(\lambda' \otimes I_{d_V}) \left\{ \frac{\partial \text{vec}[(\frac{\partial F}{\partial V'})']}{\partial \theta'} \frac{\partial \theta}{\partial \gamma'} + \frac{\partial \text{vec}[(\frac{\partial F}{\partial V'})']}{\partial V'} \frac{\partial V}{\partial \gamma'} + \frac{\partial \text{vec}[(\frac{\partial F}{\partial V'})']}{\partial \gamma'} \right\}, \\
0 &= \frac{\partial V}{\partial \gamma'} - \left(\frac{\partial F}{\partial \theta'} \frac{\partial \theta}{\partial \gamma'} + \frac{\partial F}{\partial V'} \frac{\partial V}{\partial \gamma'} + \frac{\partial F}{\partial \gamma'} \right).
\end{aligned}$$

The results follow from rearranging the above equations. $\square$

Proof of Proposition 3.4. Under the assumptions of the proposition, the following first-order condition for problem (14) hold at the optimum

$$\frac{\partial \bar{L}}{\partial \theta'} = 0. \quad (\text{A.6})$$

Total differentiating equation (A.6) with respect to γ gives

$$\frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'} \frac{\partial \theta}{\partial \gamma'} = - \frac{\partial^2 \bar{L}}{\partial \theta \partial \gamma'}.$$

This proves part 1 of the proposition.

Next, note that V is no longer an argument of $\bar{F}$. Hence, the linear system in Proposition 3.2 becomes

$$\begin{aligned}
\left\{ \frac{\partial^2 L}{\partial \theta \partial \theta'} - R_{d_\theta} \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial \theta'} \right\} \frac{\partial \theta}{\partial \gamma'} + \frac{\partial^2 L}{\partial \theta \partial V'} \frac{\partial V}{\partial \gamma'} - \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial \lambda}{\partial \gamma'} &= R_{d_\theta} \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial \gamma'} - \frac{\partial^2 L}{\partial \theta \partial \gamma'} \\
\frac{\partial^2 L}{\partial V \partial \theta'} \frac{\partial \theta}{\partial \gamma'} + \frac{\partial^2 L}{\partial V \partial V'} \frac{\partial V}{\partial \gamma'} + \frac{\partial \lambda}{\partial \gamma'} &= - \frac{\partial^2 L}{\partial V \partial \gamma'} \\
\frac{\partial F}{\partial \theta'} \frac{\partial \theta}{\partial \gamma'} - \frac{\partial V}{\partial \gamma'} &= - \frac{\partial F}{\partial \gamma'}.
\end{aligned} \quad (\text{A.7})$$

Combining the first two equations in system (A.7), the variable $\frac{\partial \lambda}{\partial \gamma'}$ can be eliminated. Rearranging the terms after the combination gives

$$\begin{aligned}
& \left\{ \frac{\partial^2 L}{\partial \theta \partial \theta'} - R_{d_\theta} \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial \theta'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial \theta'} \right\} \frac{\partial \theta}{\partial \gamma'} + \left[\frac{\partial^2 L}{\partial \theta \partial V'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial V'} \right] \frac{\partial V}{\partial \gamma'} \\
&= R_{d_\theta} \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial \gamma'} - \frac{\partial^2 L}{\partial \theta \partial \gamma'} - \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial \gamma'}.
\end{aligned} \quad (\text{A.8})$$

Substituting the last equation in system (A.7) to equation (A.8), the $\frac{\partial V}{\partial \gamma'}$ term can be

eliminated. This yields

$$\begin{aligned} & \left\{ \frac{\partial^2 L}{\partial \theta \partial \theta'} - R_{d_\theta} \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial \theta'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial \theta'} + \left[\frac{\partial^2 L}{\partial \theta \partial V'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 Q}{\partial V \partial V'} \right] \frac{\partial F}{\partial \theta'} \right\} \frac{\partial \theta}{\partial \gamma'} \\ &= R_{d_\theta} \frac{\partial \text{vec}[(\frac{\partial F}{\partial \theta'})']}{\partial \gamma'} - \frac{\partial^2 L}{\partial \theta \partial \gamma'} - \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial \gamma'} - \left[\frac{\partial^2 L}{\partial \theta \partial V'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial V'} \right] \frac{\partial F}{\partial \gamma'}. \end{aligned} \quad (\text{A.9})$$

Using equation (A.3), the Lagrange multiplier can be written as $\lambda = -\frac{\partial L}{\partial V}$, so $-R_{d_\theta} = (\frac{\partial L}{\partial V})' \otimes I_{d_\theta}$. In addition, since F is not a function of V under the present setup, $\frac{\partial V}{\partial \theta'} = \frac{\partial F}{\partial \theta'}$ holds. It follows that

$$\begin{aligned} \frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'} &= \frac{\partial}{\partial \theta'} \left(\frac{\partial L}{\partial \theta} + \frac{\partial V}{\partial \theta'} \frac{\partial L}{\partial V} \right) \\ &= \frac{\partial^2 L}{\partial \theta \partial \theta'} + \frac{\partial^2 L}{\partial \theta \partial V'} \frac{\partial V}{\partial \theta'} + \left[\left(\frac{\partial L}{\partial V} \right)' \otimes I_{d_\theta} \right] \frac{\partial \text{vec}[(\frac{\partial V}{\partial \theta'})']}{\partial \theta'} + \left(\frac{\partial V}{\partial \theta'} \right)' \left(\frac{\partial^2 L}{\partial V \partial \theta'} + \frac{\partial^2 L}{\partial V \partial V'} \frac{\partial V}{\partial \theta'} \right) \\ &= \frac{\partial^2 L}{\partial \theta \partial \theta'} + \frac{\partial^2 L}{\partial \theta \partial V'} \frac{\partial F}{\partial \theta'} - R_{d_\theta} \frac{\partial \text{vec}[(\frac{\partial V}{\partial \theta'})']}{\partial \theta'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial \theta'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial V'} \frac{\partial F}{\partial \theta'}, \end{aligned} \quad (\text{A.10})$$

and similarly,

$$\begin{aligned} \frac{\partial^2 \bar{L}}{\partial \theta \partial \gamma'} &= \frac{\partial}{\partial \gamma'} \left(\frac{\partial L}{\partial \theta} + \frac{\partial V}{\partial \theta'} \frac{\partial L}{\partial V} \right) \\ &= \frac{\partial^2 L}{\partial \theta \partial \gamma'} + \frac{\partial^2 L}{\partial \theta \partial V'} \frac{\partial V}{\partial \gamma'} + \left[\left(\frac{\partial L}{\partial V} \right)' \otimes I_{d_\theta} \right] \frac{\partial \text{vec}[(\frac{\partial V}{\partial \theta'})']}{\partial \gamma'} + \left(\frac{\partial V}{\partial \theta'} \right)' \left(\frac{\partial^2 L}{\partial V \partial \gamma'} + \frac{\partial^2 L}{\partial V \partial V'} \frac{\partial V}{\partial \gamma'} \right) \\ &= \frac{\partial^2 L}{\partial \theta \partial \gamma'} + \frac{\partial^2 L}{\partial \theta \partial V'} \frac{\partial F}{\partial \gamma'} - R_{d_\theta} \frac{\partial \text{vec}[(\frac{\partial V}{\partial \theta'})']}{\partial \gamma'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial \gamma'} + \left(\frac{\partial F}{\partial \theta'} \right)' \frac{\partial^2 L}{\partial V \partial V'} \frac{\partial F}{\partial \gamma'}. \end{aligned} \quad (\text{A.11})$$

Substituting equations (A.10) and (A.11) into equation (A.9) gives

$$\frac{\partial^2 \bar{L}}{\partial \theta \partial \theta'} \frac{\partial \theta}{\partial \gamma'} = -\frac{\partial^2 \bar{L}}{\partial \theta \partial \gamma'}.$$

This proves part 2 of the proposition. $\square$

Proof of Proposition 4.2. To begin with, note that p_a and Q_a is not a function of β for any $a \in \mathcal{A}$. They are assumed to be known to the researcher. Therefore, the derivative

of π_a with respect to β is

$$\frac{\partial \pi_a}{\partial \beta} = \frac{\partial}{\partial \beta} [(I_X - \beta Q_a)(I_X - \beta Q_A)^{-1}] (-\log p_A), \quad (\text{A.12})$$

for any $a \in \mathcal{A} \setminus \{A\}$. The derivative can be computed as follows

$$\begin{aligned} & \frac{\partial}{\partial \beta} [(I_X - \beta Q_a)(I_X - \beta Q_A)^{-1}] \\ &= -Q_a(I_X - \beta Q_A)^{-1} + (-1)(I_X - \beta Q_a)(I_X - \beta Q_A)^{-1}(-Q_A)(I_X - \beta Q_A)^{-1} \\ &= -[Q_a - (I_X - \beta Q_a)(I_X - \beta Q_A)^{-1}Q_A](I_X - \beta Q_A)^{-1} \\ &= -[Q_a - (I_X - \beta Q_A)^{-1}Q_A + \beta Q_a(I_X - \beta Q_A)^{-1}Q_A](I_X - \beta Q_A)^{-1} \\ &= -[Q_a(I_X - \beta Q_A)^{-1}(I_X - \beta Q_A + \beta Q_A) - (I_X - \beta Q_A)^{-1}Q_A](I_X - \beta Q_A)^{-1} \\ &= -[Q_a(I_X - \beta Q_A)^{-1} - (I_X - \beta Q_A)^{-1}Q_A](I_X - \beta Q_A)^{-1}. \end{aligned} \quad (\text{A.13})$$

Therefore, the result follows by substituting equation (A.13) into equation (A.12). $\square$

Proof of Corollary 4.3. Under the normalization that $\pi_A = \bar{\pi}_A$,

$$\pi_a = A_a \bar{\pi}_A + A_a \psi_A(p) - \psi_a(p) = A_a(\bar{\pi}_A - \log p_A) + \log p_a, \quad (\text{A.14})$$

for any $a \in \mathcal{A} \setminus \{A\}$ by equation (4).

Note that in equation (A.14), only A_a is a function of β . Hence, equation (A.13) can be applied to show that

$$\frac{\partial \pi_a}{\partial \beta} = -[Q_a(I_X - \beta Q_A)^{-1} - (I_X - \beta Q_A)^{-1}Q_A](I_X - \beta Q_A)^{-1}(\bar{\pi}_A - \log p_A),$$

for any $a \in \mathcal{A} \setminus \{A\}$. $\square$

Proof of Proposition 4.4. Let $\{\nu_i\}_{i=1}^X$ be the eigenvalues of Q_A . Since Q_A is a Markov matrix, the largest eigenvalue of Q_A is at most 1. Thus, the eigenvalues of $I_X - \beta Q_A$ are $\{1 - \beta \nu_i\}_{i=1}^X$. Since $\beta < 1$, it follows that $1 - \beta \nu_i > 0$ for any $i = 1, \dots, X$. Hence, $I_X - \beta Q_A$ is invertible. Next, note that

$$(I_X - \beta Q_A)^{-1} = I_X + \sum_{r=1}^{\infty} \beta^r Q_A^r. \quad (\text{A.15})$$

It follows that the two matrix products can be written as

$$Q_a(I_X - \beta Q_A)^{-1} = Q_a \left(I_X + \sum_{r=1}^{\infty} \beta^r Q_A^r \right) = Q_a + \sum_{r=1}^{\infty} \beta^r Q_a Q_A^r, \quad (\text{A.16})$$

and

$$(I_X - \beta Q_A)^{-1} Q_A = \left(I_X + \sum_{r=1}^{\infty} \beta^r Q_A^r \right) Q_A = Q_A + \sum_{r=1}^{\infty} \beta^r Q_A^{r+1}. \quad (\text{A.17})$$

Under ρ -period finite dependence for choices $a \in \mathcal{A} \setminus \{A\}$ and A for any states $x \in \mathcal{X}$, $Q_a Q_A^\rho = Q_A Q_A^\rho$ holds. As a result, equation (A.16) becomes

$$Q_a(I_X - \beta Q_A)^{-1} = Q_a + \sum_{r=1}^{\rho-1} \beta^r Q_a Q_A^r + \sum_{r=\rho}^{\infty} Q_A^{r+1}. \quad (\text{A.18})$$

Substituting equations (A.17) and (A.18) into the expression in Proposition 4.2 gives

$$\frac{\partial \pi_a}{\partial \beta} = -(Q_a - Q_A)(I_X + \beta Q_A + \cdots + \beta^{\rho-1} Q_A^{\rho-1})(I_X - \beta Q_A)^{-1}(-\log \hat{p}_A),$$

as desired. □

Proof of Corollary 4.5. Suppose that choice A is a renewal action. Following the same argument as in the proof of Proposition 4.4, $I_X - \beta Q_A$ is invertible. Using equation (A.15) and the fact that $Q_a Q_A = Q_A^2$ due to one-period finite dependence, the derivative can be written as

$$\begin{aligned} \frac{\partial \pi_a}{\partial \beta} &= -(Q_a - Q_A)(I_X - \beta Q_A)^{-1}(-\log p_A) \\ &= -(Q_a - Q_A) \left(I_X + \sum_{r=1}^{\infty} \beta^r Q_A^r \right) (-\log p_A) \\ &= \left[-Q_a + Q_A - \sum_{r=1}^{\infty} \beta^r (Q_a Q_A^r - Q_A Q_A^r) \right] (-\log p_A) \\ &= (-Q_a + Q_A)(-\log p_A), \end{aligned} \quad (\text{A.19})$$

for any $a \in \mathcal{A} \setminus \{A\}$. Note that the RHS of equation (A.19) is independent of β . In particular, the x -th row of equation (A.19) can be written as

$$\frac{\partial \pi_a(x)}{\partial \beta} = [-Q_a(x) + Q_A(x)]'(-\log p_A) \equiv \xi_x.$$

Thus, $\pi_a(x)$ is increasing, decreasing, or constant in β if $\xi_x > 0$, $\xi_x < 0$, or $\xi_x = 0$ respectively. $\square$

Proof of Corollary 4.6. To begin with, the x -th row in equation (A.19) can be written in summation form as

$$[-Q_a(x) + Q_A(x)]'(-\log p_A) = \sum_{y=1}^X [q(y|x, A) - q(y|x, a)][-\log p_A(y)]. \quad (\text{A.20})$$

Since A is a renewal action, the transition matrix is given by

$$Q_A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix},$$

i.e., all states are reset to state 1 if choice A is chosen. Hence, equation (A.20) can be further written as

$$[-Q_a(x) + Q_A(x)]'(-\log p_A) = [-\log p_A(1)] - \sum_{y=1}^X q(y|x, a)[-\log p_A(y)]. \quad (\text{A.21})$$

Under the assumption that $0 < p_A(1) \leq p_A(y)$ for any $y \in \mathcal{X}$, it follows that $-\log p_A(1) \geq -\log p_A(y)$. Hence, equation (A.21) can be written as

$$\begin{aligned} [-Q_a(x) + Q_A(x)]'(-\log p_A) &= [-\log p_A(1)] - \sum_{y=1}^X q(y|x, a)[-\log p_A(y)] \\ &\geq [-\log p_A(1)] - \sum_{y=1}^X q(y|x, a)[-\log p_A(1)] \\ &= 0, \end{aligned} \quad (\text{A.22})$$

where the last line follows from the fact that each row of the transition matrix sums to 1. Since equation (A.22) holds for any $x \in \mathcal{X}$, it follows that π_a is nondecreasing in β .

Similarly, if $p_A(1) \geq p_A(y)$ for any $y \in \mathcal{X}$, then $-\log p_A(1) \leq -\log p_A(y)$ for any $y \in \mathcal{X}$. If $p_A(1) > 0$, equation (A.21) can be written as follows

$$[-Q_a(x) + Q_A(x)]'(-\log p_A) = [-\log p_A(1)] - \sum_{y=1}^X q(y|x, a)[-\log p_A(y)]$$

$$\begin{aligned}
&\leq [-\log p_A(1)] - \sum_{y=1}^X q(y|x, a) [-\log p_A(1)] \\
&= 0,
\end{aligned}$$

which implies that the utility is nonincreasing in β . $\square$

Proof of Proposition 4.10. The optimization problem in (22) is analogous to least squares estimation. It follows that the solution is given by

$$\hat{\theta} = (\Pi' W \Pi)^{-1} (\Pi' W \hat{\pi}). \quad (\text{A.23})$$

Note that Π and W are independent of β by construction. Therefore, the derivative of $\hat{\theta}$ with respect to β can be obtained as follows:

$$\frac{\partial \hat{\theta}}{\partial \beta} = (\Pi' W \Pi)^{-1} \Pi' W \frac{\partial \hat{\pi}}{\partial \beta}.$$

$\square$

Proof of Corollary 4.11. By the given construction, the matrix Π can be written as

$$\Pi = \begin{pmatrix} \Pi_0 & 0 & \cdots & 0 \\ 0 & \Pi_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Pi_{A-1} \end{pmatrix}.$$

By definition, $\hat{\pi} = (\hat{\pi}'_0, \dots, \hat{\pi}'_{A-1})'$. Together with the assumption that W is an identity matrix, the solution in equation (A.23) can be written as

$$\begin{aligned}
\hat{\theta} &= \begin{pmatrix} \Pi'_0 \Pi_0 & 0 & \cdots & 0 \\ 0 & \Pi'_1 \Pi_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Pi'_{A-1} \Pi_{A-1} \end{pmatrix}^{-1} \begin{pmatrix} \Pi'_0 \hat{\pi}_0 \\ \Pi'_1 \hat{\pi}_1 \\ \vdots \\ \Pi'_{A-1} \hat{\pi}_{A-1} \end{pmatrix} \\
&= \begin{pmatrix} (\Pi'_0 \Pi_0)^{-1} \Pi'_0 \hat{\pi}_0 \\ (\Pi'_1 \Pi_1)^{-1} \Pi'_1 \hat{\pi}_1 \\ \vdots \\ (\Pi'_{A-1} \Pi_{A-1})^{-1} \Pi'_{A-1} \hat{\pi}_{A-1} \end{pmatrix}.
\end{aligned}$$

Taking the derivative of the a th block above with respect to β gives

$$\frac{\partial \hat{\theta}_a}{\partial \beta} = (\Pi'_a \Pi_a)^{-1} \Pi'_a \frac{\partial \hat{\pi}_a}{\partial \beta},$$

for $a \in \mathcal{A} \setminus \{A\}$. □

Proof of Proposition 4.12. To begin with, note that

$$\begin{aligned} \frac{\partial (\Pi' W \Pi)^{-1}}{\partial \delta} &= -(\Pi' W \Pi)^{-1} \frac{\partial (\Pi' W \Pi)}{\partial \delta} (\Pi' W \Pi)^{-1} \\ &= -(\Pi' W \Pi)^{-1} \Pi' (W' + W) \frac{\partial \Pi}{\partial \delta} (\Pi' W \Pi)^{-1}. \end{aligned} \quad (\text{A.24})$$

Hence, taking the derivative of equation (A.23) with respect to δ gives

$$\begin{aligned} \frac{\partial \hat{\theta}}{\partial \delta} &= -(\Pi' W \Pi)^{-1} \frac{\partial (\Pi' W \Pi)}{\partial \delta} (\Pi' W \Pi)^{-1} (\Pi' W \hat{\pi}) + (\Pi' W \Pi)^{-1} \left(\frac{\partial \Pi}{\partial \delta} \right)' W \hat{\pi} \\ &= -(\Pi' W \Pi)^{-1} \Pi' (W' + W) \frac{\partial \Pi}{\partial \delta} \hat{\theta} + (\Pi' W \Pi)^{-1} \left(\frac{\partial \Pi}{\partial \delta} \right)' W \hat{\pi} \\ &= -(\Pi' W \Pi)^{-1} \left[\Pi' (W' + W) \frac{\partial \Pi}{\partial \delta} \hat{\theta} - \left(\frac{\partial \Pi}{\partial \delta} \right)' W \hat{\pi} \right], \end{aligned}$$

where the second equality uses equations (A.23) and (A.24). □

B More details on Chen and Choo (2023)

This appendix provides additional details on the approximation error for the empirical application in Section 5.2. Figure B.1 shows the scatter plot of actual marital surplus differences against the approximate error (in percentage) for different discount factors β . It can be seen that the large percentage of approximation errors are associated with actual marital surplus differences being very close to 0 for various β .

Figure B.2 shows the scatter plot of actual marital surplus differences against the approximate error for different β . This figure shows that the approximate error remains small for various β .

Figure B.1: Percentage error of local approximation against actual marital surplus differences for various discount factors β .

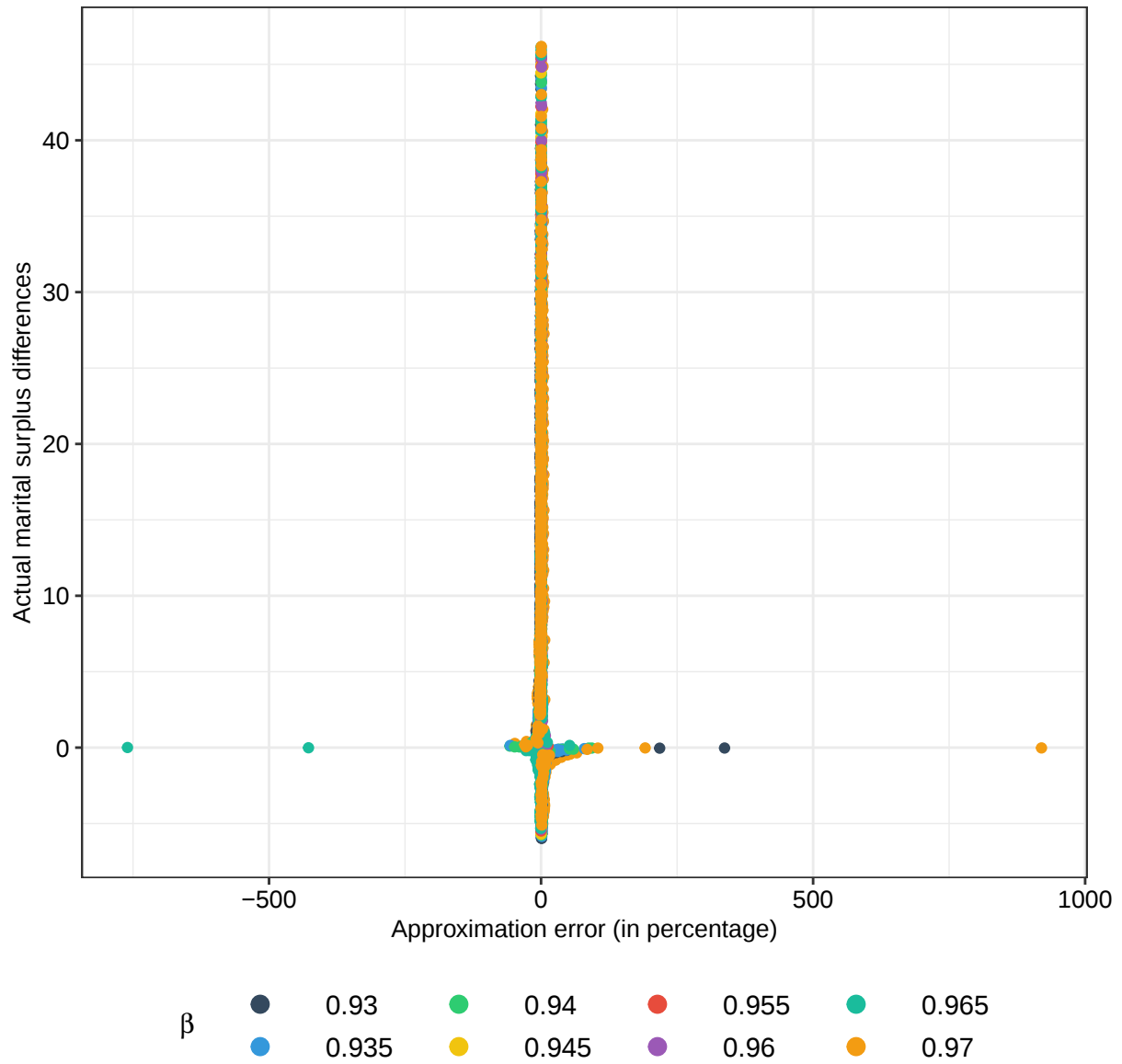
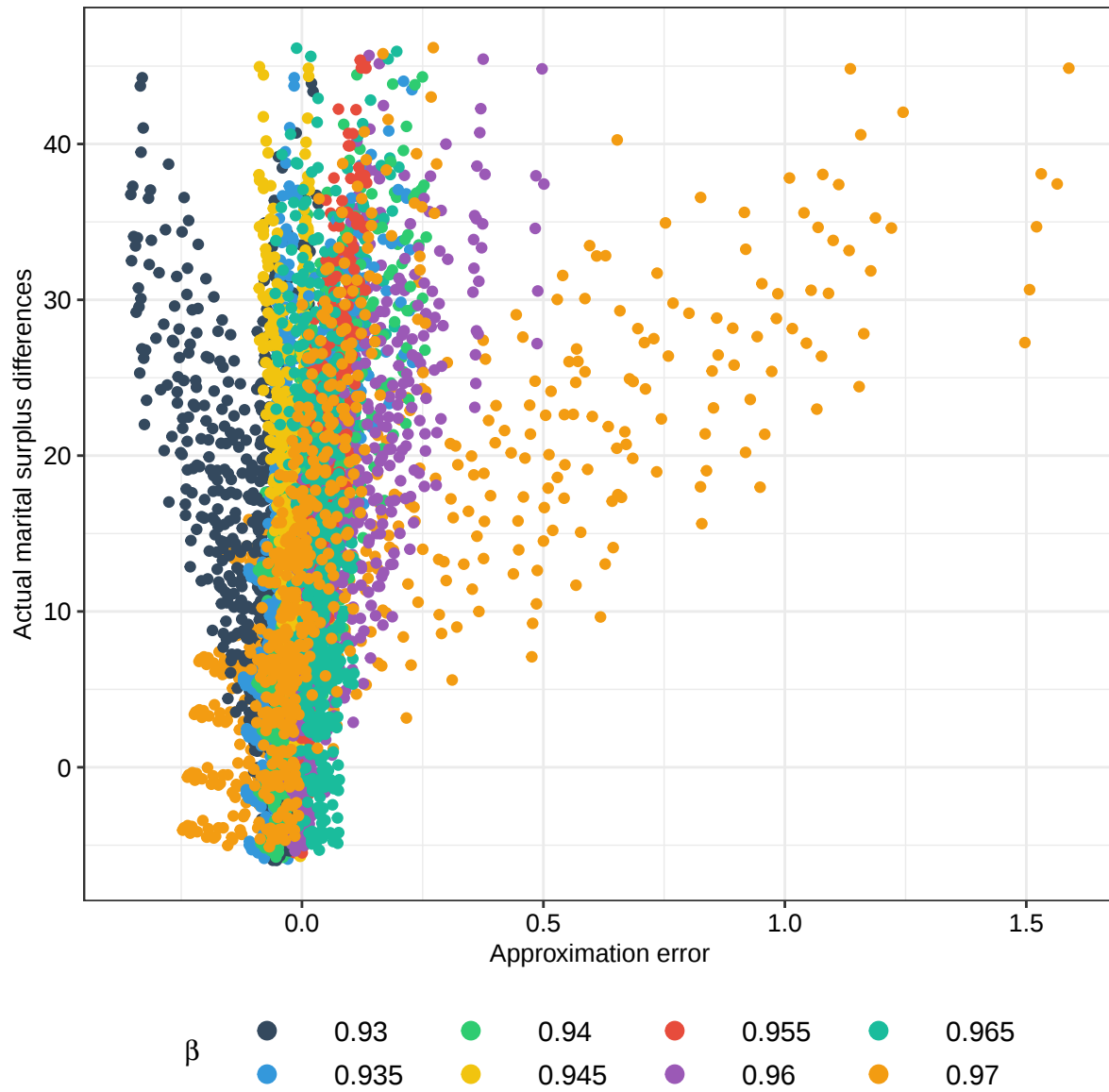


Figure B.2: Approximation error of local approximation against actual marital surplus differences for various discount factors β .



References

- ABBRING, J. H. AND O. DALJORD (2020): "Identifying the discount factor in dynamic discrete choice models," *Quantitative Economics*, 11, 471–501.
- AGUIRREGABIRIA, V., A. COLLARD-WEXLER, AND S. P. RYAN (2021): "Chapter 4 - Dynamic games in empirical industrial organization," in *Handbook of Industrial Organization, Volume 4*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, Elsevier, vol. 4 of *Handbook of Industrial Organization*, 225–343.
- AGUIRREGABIRIA, V. AND P. MIRA (2002): "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70, 1519–1543.
- ANDREWS, I., M. GENTZKOW, AND J. M. SHAPIRO (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments*," *The Quarterly Journal of Economics*, 132, 1553–1592.
- ARCIDIACONO, P. AND R. A. MILLER (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79, 1823–1867.
- (2019): "Nonstationary dynamic models with finite dependence," *Quantitative Economics*, 10, 853–890.
- ARMSTRONG, T. B. AND M. KOLESÁR (2021): "Sensitivity analysis using approximate moment condition models," *Quantitative Economics*, 12, 77–108.
- BARWICK, P. J. AND P. A. PATHAK (2015): "The costs of free entry: an empirical study of real estate agents in Greater Boston," *The RAND Journal of Economics*, 46, 103–145.
- BERTSEKAS, D. (1999): *Nonlinear Programming*, Athena scientific optimization and computation series, Athena Scientific.
- BONHOMME, S. AND M. WEIDNER (2022): "Minimizing sensitivity to model misspecification," *Quantitative Economics*, 13, 907–954.
- CHEN, L. AND E. CHOO (2023): "Identification and Parametric Estimation of Empirical Dynamic Marriage Matching Models," *The Economic Journal*, 133, 1005–1035.
- CHOO, E. (2015): "Dynamic Marriage Matching: An Empirical Framework," *Econometrica*, 83, 1373–1423.
- CHRISTENSEN, T. AND B. CONNAULT (2022): "Counterfactual Sensitivity and Robustness," *Working paper*.
- COLLARD-WEXLER, A. (2013): "Demand Fluctuations in the Ready-Mix Concrete Industry," *Econometrica*, 81, 1003–1037.
- CRAWFORD, G. S. AND M. SHUM (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73, 1137–1173.
- DAWKINS, C., T. SRINIVASAN, AND J. WHALLEY (2001): "Chapter 58 - Calibration," Elsevier, vol. 5 of *Handbook of Econometrics*, 3653–3703.

- EGESDAL, M., Z. LAI, AND C.-L. SU (2015): "Estimating dynamic discrete-choice games of incomplete information," *Quantitative Economics*, 6, 567–597.
- FOWLIE, M., M. REGUANT, AND S. P. RYAN (2016): "Market-Based Emissions Regulation and Industry Dynamics," *Journal of Political Economy*, 124, 249–302.
- FREDERICK, S., G. LOEWENSTEIN, AND T. O'DONOGHUE (2002): "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 40, 351–401.
- HONORÉ, B., T. JØRGENSEN, AND A. DE PAULA (2020): "The informativeness of estimation moments," *Journal of Applied Econometrics*, 35, 797–813.
- HOROWITZ, J. AND C. MANSKI (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.
- HORTAÇSU, A. AND J. JOO (2023): *Structural Econometric Modeling in Industrial Organization and Quantitative Marketing: Theory and Applications*, Princeton University Press.
- HOTZ, V. J. AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60, 497–529.
- IGAMI, M. (2017): "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981–1998," *Journal of Political Economy*, 125, 798–847.
- ISKREV, N. (2019): "What to expect when you're calibrating: Measuring the effect of calibration on the estimation of macroeconomic models," *Journal of Economic Dynamics and Control*, 99, 54–81.
- JØRGENSEN, T. H. (2023): "Sensitivity to Calibrated Parameters," *The Review of Economics and Statistics*, 105, 474–481.
- KALOUPSIDIS, M., Y. KITAMURA, L. LIMA, AND E. SOUZA-RODRIGUES (2021a): "Counterfactual Analysis for Structural Dynamic Discrete Choice Models," *Working Paper*.
- KALOUPSIDIS, M., P. T. SCOTT, AND E. SOUZA-RODRIGUES (2021b): "Identification of counterfactuals in dynamic discrete choice models," *Quantitative Economics*, 12, 351–403.
- KEANE, M. P. AND K. I. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.
- KLINE, P. AND A. SANTOS (2013): "Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure," *Quantitative Economics*, 4, 231–267.
- KONG, X., J.-P. DUBÉ, AND O. DALJORD (2022): "Non-parametric Estimation of Habitual Brand Loyalty," *Working Paper*.
- LIGHT, B. (2021): "Stochastic Comparative Statics in Markov Decision Processes," *Mathematics of Operations Research*, 46, 797–810.
- MAGNAC, T. AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816.
- MAGNUS, J. R. AND H. NEUDECKER (1985): "Matrix differential calculus with applications to simple, hadamard, and kronecker products," *Journal of Mathematical Psychology*, 29, 474–492.

- MASTEN, M. A. AND A. POIRIER (2020): "Inference on breakdown frontiers," *Quantitative Economics*, 11, 41–111.
- NORETS, A. (2010): "Continuity and differentiability of expected value functions in dynamic discrete choice models," *Quantitative Economics*, 1, 305–322.
- OSSA, R. (2014): "Trade Wars and Trade Talks with Data," *American Economic Review*, 104, 4104–46.
- PAKES, A., M. OSTROVSKY, AND S. BERRY (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *The RAND Journal of Economics*, 38, 373–399.
- RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.
- (1988): "Maximum Likelihood Estimation of Discrete Control Processes," *SIAM Journal on Control and Optimization*, 26, 1006–1024.
- (1994): "Chapter 51 Structural estimation of markov decision processes," Elsevier, vol. 4 of *Handbook of Econometrics*, 3081–3143.
- SU, C.-L. AND K. L. JUDD (2012): "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, 80, 2213–2230.
- TOPKIS, D. M. (1998): *Supermodularity and complementarity*, Princeton university press.
- VERDIER, V. AND C. REELING (2021): "Welfare Effects of Dynamic Matching: An Empirical Analysis," *The Review of Economic Studies*, 89, 1008–1037.
- YANG, M.-J. (2021): "Micro-level Misallocation and Selection," *American Economic Journal: Macroeconomics*, 13, 341–68.