



UvA-DARE (Digital Academic Repository)

Essays on empirical likelihood in economics

Gao, Z.

Publication date
2012

[Link to publication](#)

Citation for published version (APA):

Gao, Z. (2012). *Essays on empirical likelihood in economics*. Thela Thesis.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

ROBUST DECISIONS IN DYNAMIC CHOICE MODELS

3.1 INTRODUCTION

3.1.1 *Relevant Literature*

Dynamic discrete choice models build on the assumption of rational behavior. The rationality appears in the expectation form and utility maximization when an agent makes his/her decisions dynamically. This model exploits the intrinsic evolution structure and captures the endogenous effect of agents' actions. Self-expectations, however, are difficult to be characterized completely in practice because of the uncertainty of expectations. If an agent has an incomplete information set, his expectation based on such a set will be incomplete as well. The incompleteness will induce stochastic characteristics of his/her expectation. On the other hand, if this agent is aware of the incompleteness and will make a decision accounting for this effect rather than insisting on the maximal principle of utility, he/she may follow a strategy called robust decision. Such an agent is often referred to as a boundedly rational agents. Incomplete information has been addressed in many dynamic discrete choice papers (Aguirregabiria and Mira, 2002, 2007; Bajari et al., 2007) while boundedly rational agent and robust decision are often addressed in the dynamic control literature in especially for macroeconomic models (Hansen and Sargent, 2007). This chapter will consider the robust decision problem in the framework of dynamic discrete choice models¹. We will identify robustness issues in dynamic discrete choice models and then develop a tractable prediction approach under a more flexible assumption on consumer's optimal behavior and a more flexible model specification.

¹ This setup is often referred to as a (mathematical) dynamic game in control related subjects.

Estimation for single agent dynamic discrete choice models was initiated by Wolpin (1984), Pakes (1986) and Rust (1987). Estimating the parameters of these structural models requires solving an optimization problem with a nested dynamic programming problem. The computation is non-trivial for both the likelihood function evaluation (outer-loop) and the fixed point iteration (inner-loop). The difficulties mainly come from two sources, the exponential growth of the state space and the existence of multiple equilibria (Aguirregabiria and Mira, 2010).

A common approach for breaking down the complexity growth is to specify the expectation. When the expectation is evaluated with respect to a specified distribution, like the private information, the expectation will be computed in terms of analytic algebra function instead of numerical integration. In this way, randomization can break down the curse of dimensionality caused by the multivariate integration in the dynamic programming. Keane and Wolpin (1994) suggest solving the dynamic programming problem by Monte Carlo integration and interpolation. Later, Rust (1997) proves that a certain random Bellman operator with particular parametric structure can break the curse of dimensionality. This method works well for unique equilibrium models. But if there are multiple equilibria, the expectation will depend on the equilibrium that gives locally optimal utility. Estimation for multiple equilibrium may need to compute the full set of equilibria in order to select an optimal expression. Usually, this is a two-step procedure. In the first step, with some estimated or approximated components, one will establish an objective function of the expected utility and then use this function in the second step. In the second step, one will solve the optimization problem of this objective function with respect to the structural parameters. For this kind of two-step approach (Aguirregabiria and Mira, 2007; Pesendorfer and Schmidt-Dengler, 2003; Bajari et al., 2007; Pakes et al., 2008), the final estimation result depends heavily on the constructed objective function in the first step. Two issues may complicate the inferential result. 1. If the expectation is accompanied by parametric assumptions, then the true model should not be far from the parametric model, otherwise the bias in the first step will be accumulated afterward. 2. If the expected function involves non/semi-parametric components, the associated curse of dimensionality problem will occur.

For inferential or prediction purposes, the most important criterion is minimum mean square error (MSE). But two essential

reasons make it difficult to incorporate the MSE criterion in the two-step procedure. The *first* reason is the concern of the misspecification problems at the second step of the estimate. The second step objective functions is either a pseudo likelihood or an approximating moment constraint. They are not exact parametric functions or constraints but approximating ones. Fernandez-Villaverde et al. (2006) show that the higher order bias from the first step dynamic programming solution will give a first-order biased-effect on the likelihood function in the second step. Therefore, for the concern of consistency, achieving an exact fit in the first step is more important than minimizing MSE. But this may lead to a potential issue of over-fitting.

The *second* reason is the small sample problem. In agent-based modeling, data resources are often very limited and it may be unwise to solely rely on pure statistical inferential decisions to judge the correctness of the models. “The statistical mentality that ‘all structural models are rejected, therefore none of them are any good’ does a disservice and contributes to a radicalization of some members of the profession”-Rust (2008). Therefore, it seems that statistical prediction or inference should not be ranked in the first place for structural estimation and modeling.

In sum, given the complicated nesting dynamic programming structure, people often prefer to “fit” a given model (non)- parametrically rather than to obtain a (non)-parametric function with prediction ability. However, in practice, there is no universally acceptable true model. It means that in principle even an unbiased estimator for a hypothetical true model does not necessarily give us a good estimate for the underlying model. In this chapter, we give up maintaining the aim of consistent approximation. We look for an alternative method which allows for little bias but will yield more flexibility in return. In dynamic models, this purpose can be explained in terms of bounded rationality, to be more specific, *the worst case* of Bellman’s optimality principle, where the numerically approximated solution is a distance ϵ away from the fixed point value function.

3.1.2 Contributions

Our approach is motivated by the constrained maximum likelihood (Rust (1987, Nested Fixed Points - NXFP), Aguirregabiria and Mira (NPML 2007, -Nested Pseudo Maximum Likelihood)) and the constrained optimization (Su and Judd (2011, Mathematical Programming with Equilibrium Constraints-MPEC)). Our

approach does not rely on strong parametric assumptions on the specification of the incomplete information and addresses the robustness concern in the model. We also provide some techniques to solve the problems that accompany with the appearance of boundedly rational agents.

The first contribution of this paper is proposing an alternative approach to estimate a dynamic discrete choice model with boundedly rational agents. This is a game-theoretic approach. The departure from the rationality induces the so called ϵ -equilibrium or *robust equilibrium* (Radner, 1981, 1986). In ϵ -equilibrium, each player is satisfied to get within a small distance of his best response function. In other words, given the ϵ -equilibrium strategy profile, the player will not gain more than ϵ expected payoff by altering his strategy. In our context, ϵ -equilibrium will be converted to an ϵ -variation of a fixed point condition. Suppose one fixed point condition is equivalent to one equilibrium model, then ϵ -variation will induce a set of multiple equilibria where dynamic programming is difficult to implement. This chapter will consider how to select a single representative model from the set of equilibria. The selected model is able to explain those alternative models that satisfy the ϵ -fixed point condition.

The second contribution is to exploit the asymptotic properties of the solution of the dynamic programming problem with boundedly rational agents. The asymptotic behavior is not standard since the solution is inconsistent within a small neighborhood but the solution will not get out of this neighborhood in probability. This non-trivial consistency result forms the quantitative intuition of boundedly rational agent in the ϵ -equilibrium.

The third contribution is to apply a localization method to handle the moment constraints with non-smooth population functions. Due to the “bias” or “inconsistency” from the first step, one should expect that the population function is non-smooth even at the true parameter value. Thus the usual sample smoothing methods may not qualify for the handling of a non-differentiable population moment function. The localization method transfers this global infeasible problem about structural parameters into a feasible problem about local parameters.

3.2 STOCHASTIC DYNAMICS

The agent's action is taken in a given period and may affect both current and future profits. The evolution of an agent is modeled via the dynamic programming problem with an infinite discrete time horizon $t = 1, 2, \dots, \infty$. Dynamic programming is able to exploit the intrinsic evolution process and construct a controllable scheme. In period t the state variable is denoted as s_t . The vector of all states for all N -agents is commonly observed as $\mathbf{s}_t = (s_{1t}, \dots, s_{Nt}) \in \mathcal{S}$, where $\mathcal{S} \subset \mathbb{R}^N$ is the entire state space. Given the current common information \mathbf{s}_t , an agent will choose his/her action(s) $\mathbf{a}_t = (a_{1t}, \dots, a_{Nt}) \in \mathcal{A}$ in each period to adjust his/her expected utility. The action space \mathcal{A} includes only a finite number of actions. The expectation is driven by private shocks. These shocks, collected in $\Xi_t = (\xi_{1t}, \dots, \xi_{Nt})$, are unobservable information for the analysts and are observed only by the agents. They are treated as an N -dimensional independent *random vector*. The distribution of ξ_{it} is denoted $G(\cdot)$.

For simplicity, the current action and state are denoted by (\mathbf{a}, \mathbf{s}) and $(\mathbf{a}', \mathbf{s}')$ denote the next period action and state. If the evolution pattern is first order Markovian, then given an action \mathbf{a} , the state \mathbf{s} will transit to \mathbf{s}' with a transition density $\mathbf{p}(\mathbf{s}'|\mathbf{s}, \mathbf{a})$. With this transition density, we can construct the transition density of the full Markov chain from \mathbf{s}_1 to \mathbf{s}_T as:

$$\mathbf{p} = (p(\mathbf{s}_2|\mathbf{s}_1, \mathbf{a}_1), \dots, p(\mathbf{s}_{T+1}|\mathbf{s}_T, \mathbf{a}_T)).$$

We assume the Markov transition density is *homogeneous*, meaning that given action \mathbf{a} , states \mathbf{s} and \mathbf{s}' , the transition density $p(\mathbf{s}'|\mathbf{s}, \mathbf{a})$ is homogeneous in time.

Let $\Pi(\mathbf{a}_t, \mathbf{s}_t, \Xi_t)$ be the N -vector profit for all agents in period t . Because we consider discrete choice models, we assume that $\Pi(\mathbf{a}_t, \mathbf{s}_t, \Xi_t)$ contains independent elements. Then by implication we can consider operations of individual elements independent of each other. Let $\pi(\mathbf{a}_t, \mathbf{s}_t, \xi_t)$ be a single representative of an element in $\Pi(\mathbf{a}_t, \mathbf{s}_t, \Xi_t)$. A *rational* agent who makes his decisions by maximizing the expected current and discounted future profits,

$$\mathbb{E} \left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi(\mathbf{a}_\tau, \mathbf{s}_\tau, \xi_\tau) | \mathbf{s}_t \right], \quad (3.1)$$

where $\beta \in (0, 1)$ is the discount factor. The expectation is taken w.r.t. the distribution constructed by Markov chain \mathbf{p} and $G(\cdot)$. The shock ξ_t depends on the state \mathbf{s}_t therefore $G(\xi)$ is a condi-

tional distribution given \mathbf{s} . The primitives of the model include the discount factor β and the transition densities $p(\mathbf{s}'|\mathbf{s}, \mathbf{a})$.²

If the distribution $G(\cdot)$ can be associated with some parameter θ , then it is also a primitive factor. The parametric expected profit function is:

$$\mathbb{E}_\theta \left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi(\mathbf{a}_\tau, \mathbf{s}_\tau, \xi_\tau) | \mathbf{s}_t \right], \quad (3.2)$$

where θ is the unknown parameter. The difference between (3.1) and (3.2) is whether or not the value function can be parametrized by θ .

Markov Perfect Equilibria (MPE) i.e. equilibria where the optimal actions of agents will depend only on the state at time t , are frequently used in various dynamic discrete choice models. By MPE condition, each action depends on the current state and its current private shock. Let $\sigma(\mathbf{s}, \xi)$ be a Markov policy, or strategy function or decision rule that maps the current state and shock onto an action $\sigma(\cdot, \cdot) : \mathcal{S} \times \mathbb{R}^N \rightarrow \mathcal{A}$. The Bellman equation for the dynamic model (3.1) is

$$V(\mathbf{s}, \xi) = \max_{\mathbf{a} \in \mathcal{A}} \left\{ \pi(\mathbf{a}, \mathbf{s}, \xi) + \beta \int \int V(\mathbf{s}', \xi') dP(\mathbf{s}' | \mathbf{s}, \mathbf{a}) dG(\xi') \right\}, \quad (3.3)$$

where $V(\mathbf{s}, \xi)$ is the value function. Since the expression may involve discrete- or real-valued ξ , the notation of the underlying value function is set up in terms of integrals rather than sums. The associated policy function or decision function is:

$$\sigma(\mathbf{s}, \xi) := \arg \max_{\mathbf{a} \in \mathcal{A}} \left\{ \pi(\mathbf{a}, \mathbf{s}, \xi) + \beta \int \int V(\mathbf{s}', \xi') dP(\mathbf{s}' | \mathbf{s}, \mathbf{a}) dG(\xi') \right\}.$$

The above dynamic programming problem suffers from a serious curse of dimensionality if the dimension of ξ is high. In this case, the numerical integral w.r.t. $G(\xi)$ is difficult to evaluate and thus the expectation is poorly approximated. By Bellman's principle of optimality, Rust (1987) suggests to substitute the

² Publicly known transition densities are not standard in the literature, but it is a compromise by relaxing the specification of probability $G(\cdot)$. In the literature, $G(\cdot)$ is often assumed being multinomial logit which leads to a conditional choice probability $F(\mathbf{a}|\mathbf{s})$ which is also a multinomial logit see Hotz and Miller (1993, Lemma 3.1). This $F(\mathbf{a}|\mathbf{s})$ in turn leads to $p(\cdot|\cdot, \mathbf{a})$. Both are extremely sensitive to the choice of $G(\cdot)$. In the boundedly rational case, G is not consistently estimated and leads to a very unrealistic $p(\cdot|\cdot, \mathbf{a})$.

decision rule into equation (3.3) and to integrate out the private shocks, then:

$$V(\mathbf{s}, \sigma(\mathbf{s})) = \mathbb{E}_{\xi} \left[\pi(\sigma(\mathbf{s}), \mathbf{s}, \xi) + \beta \int V(\mathbf{s}', \sigma(\mathbf{s}')) dP(\mathbf{s}' | \mathbf{s}, \sigma(\mathbf{s})) \right]. \quad (3.4)$$

The function $V(\mathbf{s}, \sigma(\mathbf{s}))$ illustrates that the expected profit at the beginning of a period satisfies Bellman's optimality principle. Equation (3.4) is called the *integrated Bellman equation* and $V(\mathbf{s}, \sigma(\mathbf{s}))$ is called the *ex ante value function* (Bajari et al., 2007). Because the ex ante value function $V(\mathbf{s}, \sigma(\mathbf{s}))$, or $V(\mathbf{s})$ in short, is not a function of ξ but of \mathbf{s} only, the solution of the problem (3.4) will be a function of \mathbf{s} only. This is a way of dealing with the unknown distribution of ξ .

We assume the following standard conditions:

Condition 3.1. Conditional Independence (CI): Given \mathbf{s}_t , the random variables ξ_{it} and ξ_{js} are independent for any $s > t$ and $i \neq j$.

Condition 3.2. Additive Separability (AS): A private shock appears additively in the profit function. $\pi(\mathbf{a}, \mathbf{s}, \xi) = \bar{\pi}(\mathbf{a}, \mathbf{s}) + \xi$.

3.3 ROBUST DECISION

In this section, we will use an ϵ -approximating value function \hat{V} to solve dynamic programming problem dealing with the robustness concern. This problem is developed from equation (3.4). Also we will give a policy function iteration method for this problem. All the results in this section focus on the first step of a 2-step estimation procedure. The second step will be discussed in the next section.

3.3.1 The Kernel-based Constrained Optimization

The motivation of the constrained optimization procedure is to put the thorny part of the problem into the constraints and set up an objective function of controllable coefficients of the constraints. We will consider the ϵ -variation fixed point condition defined below as a single inequality constraint. The set of MPEs will satisfy this constraint. Then we will use a local basis function to approximate the unknown ex ante value function.

Given the MPE assumption, we use a vector of *Radial Basis Functions (RBF)* $\Phi(\cdot)$ to approximate the ex ante value function

in (3.4) since we assume a permutation structure of the elements of \mathbf{s} so that the $V(\cdot)$ depends on the length of \mathbf{s} , not the ordering of its elements:

$$\hat{V}(\mathbf{s}) = \rho^T \Phi(\mathbf{s}). \quad (3.5)$$

The standard inner product for a functional space of Φ defines a kernel function as follows. Select T points $(\mathbf{s}_1, \dots, \mathbf{s}_T)$, and calculate the kernel:

$$K_{ij} := k(\mathbf{s}_i, \mathbf{s}_j) := \langle \Phi(\mathbf{s}_i), \Phi(\mathbf{s}_j) \rangle.$$

This kernel is a positive definite kernel (Gram) matrix. By letting the dimension of ρ and Φ go to infinity, we can approximate the function $V(\cdot)$ arbitrarily well. But in practice, a finite number of basis functions has to be chosen. Similar approximations can also be found in Bajari et al. (2007) who assume a profit function and a value function that are linear in the unknown parameters in order to simplify the computation. However, equation (3.5) has a different meaning. The coefficient ρ in (3.5) does not necessarily correspond to a structural parameter as in Bajari et al. (2007) at all, but the coefficient ρ itself can be used to describe how complex³ the approximation is. The norm of ρ , $\|\rho\|$, is a regularizer for the complexity of the solution \hat{V} . Thus we are trying to *learn* the unknown function $\hat{V}(\cdot)$ and we can adjust the *learning* ability of function $\rho^T \Phi(\mathbf{s})$ by tuning ρ . We will include $\|\rho\|^2/2$ in the objective function in order to penalize the complexity of potential approximations.

Condition 3.3. (ϵ -equilibrium) The ϵ -equilibrium set includes Markov Perfect Equilibria and can be represented in terms of $\hat{V}(\mathbf{s})$ in equation (3.5).

Condition 3.3 is to convert a dynamic programming problem with robust decisions to an expression in ϵ -equilibria. Here ϵ is a *deterministic* value determined by prior deliberation. The ϵ -equilibria imply that agents are indifferent or cannot distinguish within a set that covers their optimal action functions, for example any V :

$$V(\mathbf{s}, \sigma(\mathbf{s})) - \frac{1}{2}\epsilon \leq V \leq V(\mathbf{s}, \sigma(\mathbf{s})) + \frac{1}{2}\epsilon,$$

would be a feasible solution for equation (3.4). Therefore, by Condition 3.3 the profit $\mathbb{E}_{\xi} \pi(\sigma(\mathbf{s}, \xi), \mathbf{s}, \xi)$ based on the decision

³ The complexity refers the number of implementing basis functions i.e. see Vapnik (1998).

$\sigma(\mathbf{s})$ for fully rational agents is in fact acceptable for boundedly rational agents. Mathematically, we relax the rational expectation condition based on $\mathbb{E}_{\xi}[\cdot]$ and require only that the solutions lie within an ϵ -distance of the exact rational expectation solution. We will refer to the resulting solution set as the ϵ -tube defined as follows:

$$\left| V(\mathbf{s};\sigma(\mathbf{s})) - \pi(\sigma(\mathbf{s},\xi),\mathbf{s},\xi) - \beta \int V(\mathbf{s}',\sigma(\mathbf{s}))dP(\mathbf{s}'|\mathbf{s},\sigma(\mathbf{s})) \right| \leq \epsilon. \quad (3.6)$$

Obviously, when ϵ goes to zero, equation (3.6) corresponds to the exact rational expectation solution in (3.4). The $\mathbb{E}_{\xi}[\cdot]$ in (3.4) corresponds to the exact fixed point and the principle of optimal decision in Bellman's equations, however, due to the uncertainty of the functional form of $G(\xi)$, it is quite common for an agent to realize that his/her profits will differ from the expected values and make decisions allowing for uncertainty. Mathematically, the difference between (3.6) and the integrated Bellman equation is that (3.6) accommodates the stochastic effects using a deterministic inequality. Relations between deterministic deviation ϵ and uncertainty can be traced back to the work of Debreu (1972) and Simon (1957).

Therefore, instead of using the equality function (3.4), we establish an inequality where the solution of the integrated Bellman equation (3.4) is included. Equation (3.6) can be interpreted as a so-called ϵ -insensitive loss function or a pseudo-metric by Vapnik (1998) such that:

$$|\hat{V} - \Gamma\hat{V}|_{\epsilon} = \max \{0, |\hat{V} - \Gamma\hat{V}| - \epsilon\}, \quad (3.7)$$

where Γ is an operator on $V(\mathbf{s}) \in \mathcal{V}$ such that $\Gamma : \mathcal{V} \mapsto \mathcal{V}$ and defined as

$$\Gamma V(\mathbf{s}) := \pi(\sigma(\mathbf{s}),\mathbf{s},\xi) + \beta \int V(\mathbf{s}',\sigma(\mathbf{s}))dP(\mathbf{s}'|\mathbf{s};\sigma(\mathbf{s})).$$

The stochastic feature of $\pi(\sigma(\mathbf{s}),\mathbf{s},\xi)$ implies that Γ is a *random* Bellman's operator (Rust, 1997) so Γ still depends on ξ . For the space \mathcal{V} which in general is a sub-space of a Banach space, we will only consider a special discretized version such that Γ shares similar properties with the *deterministic* Bellman's operator. This approach has been used previously e.g. in (Rust et al., 2002) where $\pi(\sigma(\mathbf{s},\xi),\mathbf{s},\xi)$ is discretized into finite states as non-deterministic lattices.

In this chapter, we let equation (3.6) suggest a different approach. The random operator is limited within a deterministic

ϵ tube. All the randomness of ΓV will be treated equivalent to the single deterministic counterpart inside the ϵ -tube⁴. We also need a condition for this random operator:

Condition 3.4. (E) Ergodic distribution $G(\xi)$: Γ is a measure-preserving mapping on the measurable space \mathcal{V} .

By Condition 3.3, the solution set of (3.6) will be represented in terms of \hat{V} . We call the equation (3.6) ϵ -tube which is a ϵ -fixed point condition in the optimization procedure. The primal problem of minimizing the complexity of the approximating value function subject to the ϵ -fixed point condition of (3.6) is:

$$\begin{aligned} \min_{\rho, \zeta, \zeta^*} \quad & \frac{1}{2} \|\rho\|^2 + \sum_{t=1}^T \left[\frac{C}{|\mathcal{S}|} (\zeta_t + \zeta_t^*) \right], \\ \text{s.t.} \quad & \rho^T \Phi(\mathbf{s}_t) - \Gamma \hat{V}(\mathbf{s}_t) \leq \epsilon + \zeta_t, \quad -\rho^T \Phi(\mathbf{s}_t) + \Gamma \hat{V}(\mathbf{s}_t) \leq \epsilon + \zeta_t^*, \\ & \zeta_t \geq 0, \quad \zeta_t^* \geq 0 \text{ for all } 0 \leq t \leq T, \\ \text{with } \Gamma \hat{V}(\mathbf{s}_t) := & \pi(\mathbf{a}_t, \mathbf{s}_t, \xi_t) + \beta \sum_{\mathbf{s}_{t+1} \in \mathcal{S}} (\rho^T \Phi(\mathbf{s}_{t+1})) \mathbf{p}(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}), \end{aligned} \tag{3.8}$$

ζ and ζ^* are slack variables, $|\mathcal{S}|$ is the number of states and C is some constant. The ϵ -fixed point condition sets up a 2ϵ -wide “tube” for the fitting curve. The slack variables ζ and ζ^* play roles as soft margins for that tube, like dual variables in the L^1 -penalty⁵.

Remark 3.5. If ξ is an additive term in $\pi(\sigma(\mathbf{s}, \xi), \mathbf{s}, \xi)$ in (3.6), equation (3.6) and equation (3.4) associated with condition 3.2 imply that $|\mu| \leq 2\epsilon$ where μ is the expectation of ξ , we have:

$$\left| \rho^T \Phi(\mathbf{s}_t) - \left[\bar{\pi}(\sigma_t, \mathbf{s}_t) + \mu + \beta \sum_{\mathbf{s}_{t+1} \in \mathcal{S}} (\rho^T \Phi(\mathbf{s}_{t+1})) \mathbf{p}(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}) \right] \right|$$

less than 2ϵ . This inequality implicitly states that ϵ -deviation implies MPE as in the deterministic Bellman equation when ϵ goes to zero.

Remark 3.6. Minimization in (3.8) captures the main feature of constructing a robust procedure. It states that in order to obtain a small risk (the expected value of a loss function) in the prediction, we need to control both empirical risk, via the

⁴ The idea of ϵ -tube is similar to the idea of averaging in statistics in sense that it maps a stochastic feature into deterministic parameter values.

⁵ The dual problem of $\min |f|$ is $\min \zeta$ s.t. $-\zeta \leq f \leq \zeta$.

ϵ -insensitive loss function, and model complexity, via penalties ρ . The parameter C trades off the complexity and the prediction ability of the approximation.

The key idea of solving problem (3.8) is to construct a Lagrangian from the objective function and setup dual constraints. The Lagrangian function is the following:

$$\begin{aligned} L := & \frac{1}{2} \|\rho\|^2 + \frac{C}{|\mathcal{S}|} \sum_{t=1}^T (\zeta_t + \zeta_t^*) - \sum_{t=1}^T (\eta_t \zeta_t + \eta_t^* \zeta_t^*) \quad (3.9) \\ & - \sum_{t=1}^T \alpha_t \cdot \left(\epsilon + \zeta_t + \rho^T \Phi(\mathbf{s}_t) - \Gamma \left(\rho^T \Phi(\mathbf{s}_t) \right) \right) \\ & + \sum_{t=1}^T \alpha_t^* \cdot \left(\epsilon + \zeta_t^* - \rho^T \Phi(\mathbf{s}_t) + \Gamma \left(\rho^T \Phi(\mathbf{s}_t) \right) \right), \end{aligned}$$

where α_t^* and α_t are the multipliers for ϵ -inequalities and η_t^* and η_t are the multipliers for the positive constraints $\zeta_t^* \geq 0$ and $\zeta_t \geq 0$ respectively, and $\alpha_t^*, \alpha_t, \eta_t^*, \eta_t \geq 0$. For the *fitting error* $\rho^T \Phi(\mathbf{s}_t) - \Gamma(\rho^T \Phi(\mathbf{s}_t))$, we use a regression-type expression to separate the approximating value function $\rho^T \Phi(\cdot)$:

$$\rho^T \Phi(\mathbf{s}_t) - \Gamma \left(\rho^T \Phi(\mathbf{s}_t) \right) = \rho^T \Psi(\mathbf{s}_t) - \pi(\mathbf{a}_t, \mathbf{s}_t, \zeta_t), \quad (3.10)$$

where $\Psi(\mathbf{s}) := \Phi(\mathbf{s}) - \beta \sum_{\mathbf{s}' \in \mathcal{S}} \Phi(\mathbf{s}') \mathbf{p}(\mathbf{s}' | \mathbf{s}, \mathbf{a})$. The partial derivatives of L with respect to the primal variables (ρ, ζ, ζ^*) equal to zero for optimality:

$$\begin{aligned} \partial L / \partial \rho &= \rho - \sum_{t=1}^T (\alpha_t^* - \alpha_t) \Psi(\mathbf{s}_t) = 0, \\ \partial L / \partial \epsilon &= - \sum_{t=1}^T (\alpha_t^* - \alpha_t) = 0, \quad (3.11) \\ \partial L / \partial \zeta_t^{(*)} &= \frac{C}{|\mathcal{S}|} - \alpha_t^{(*)} - \eta_t^{(*)} = 0, \end{aligned}$$

where $(*)$ indicates that analogous results hold for both ζ_t and ζ_t^* . Substituting (3.11) into (3.9) yields the dual optimization problem,

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & - \frac{1}{2} \sum_{i,j=1}^T (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) \langle \Psi(\mathbf{s}_i), \Psi(\mathbf{s}_j) \rangle \\ & - \sum_{t=1}^T (\alpha_t^* + \alpha_t) \epsilon + \sum_{t=1}^T (\alpha_t^* - \alpha_t) \pi(\mathbf{a}_t, \mathbf{s}_t, \zeta), \quad (3.12) \\ \text{s.t.} \quad & 0 \leq \alpha^*, \alpha \leq C/|\mathcal{S}| \text{ and } \sum_{t=1}^T (\alpha_t^* - \alpha_t) = 0, \end{aligned}$$

where the quadratic term is the simplification of

$$\frac{1}{2} \|\rho\|^2 - \sum_{t=1}^T (\alpha_t^* - \alpha_t) \left[\rho^T \Phi(\mathbf{s}_t) - \Gamma \left(\rho^T \Phi(\mathbf{s}_t) \right) + \pi(\mathbf{a}_t, \mathbf{s}_t, \xi_t) \right].$$

The dual optimization problem (3.12) is a convex linear quadratic programming. Many available packages can solve this standard problem (3.12). Given the numerical values α and α^* , coefficients ρ and approximating value function \hat{V} are:

$$\begin{aligned} \rho &= \sum_{t=1}^T (\alpha_t^* - \alpha_t) \Psi(\mathbf{s}_t), \\ \hat{V}(\mathbf{s}_t) &= \rho^T \Phi(\mathbf{s}_t) = \sum_{t=1}^T (\alpha_t^* - \alpha_t) \times \\ &\quad \left[K(\mathbf{s}_t) - \beta \sum_{\mathbf{s}_{t+1} \in \mathcal{S}} \mathbf{p}(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}) K(\mathbf{s}_t) \right] \end{aligned} \quad (3.13)$$

where $K(\mathbf{s}_t) := \langle \Phi(\mathbf{s}_t), \Phi(\mathbf{s}_t) \rangle$. The coefficient vector ρ consists of the dual slack variables α , α^* and basis functions $\Phi(\cdot)$. As a function of α, α^* that dually represent inequality conditions, the coefficient ρ captures the influence from bounded rationality via ϵ . The optimal α, α^* also implies that the estimated ρ should trade off the complexity that preserves the shape of the approximating solution and the feasibility of implementations. When ϵ is changed, the approximation \hat{V} will change as reflected in a change of ρ .

The constraint (3.6) restricts the potential solution set of \hat{V} as a regularized operator. We denote this regularized operator Y such that it maps from a kernel solution $K(\cdot)$ into a restricted inner product $\Omega(K) := \langle Y\Phi, Y\Phi \rangle$ where $\Omega(K)$ is the regularization term. The operator Y induces an invariant subspace, because it imposes that interesting “potential candidate” functions should “almost” satisfy the constraint and should be stable under Y transformation.

Before giving the main result of this chapter, we need three technical conditions. Condition 3.7 is the *Lipschitz condition* for the profit function and transition density function. Condition 3.8 is to abstract from unboundedness issues in basis and kernel functions. Both of them will restrict us to a “stable” functional class. Condition 3.9 is the necessary condition for applying empirical process results in Bousquet and Elisseeff (2002). This condition basically implies independent increments.

Condition 3.7. (i) There exists a constant C_π such that for any $\mathbf{a} \in \mathcal{A}$ and $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$ such that $|\pi(\mathbf{s}, \mathbf{a}, \xi) - \pi(\mathbf{s}', \mathbf{a}, \xi)| \leq C_\pi \|\mathbf{s} - \mathbf{s}'\|$.

(ii) For any $\mathbf{a} \in \mathcal{A}$, $\mathbf{s}, \mathbf{s}', \mathbf{t} \in \mathcal{S}$ and $G(\xi)$ with Radon–Nikodym density $g(\xi)$, $|p(\mathbf{s}'|\mathbf{s}, \mathbf{a}) - p(\mathbf{s}'|\mathbf{t}, \mathbf{a})| \cdot |g(\xi)| \leq C_p(\mathbf{s}') \|\mathbf{s} - \mathbf{t}\|$ where $C_p(\mathbf{s}')$ is an envelop function in L^2 .

Condition 3.8. There is an upper bound κ such that $\|k(\mathbf{s}, \mathbf{s}')\| < \kappa$ for any $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$. There is an associated upper bound M^* such that $|\hat{V}| = |\rho^T \Phi(\mathbf{s})| < M^*$ for all $\mathbf{s} \in \mathcal{S}$.

Let π_t denote $\pi(\mathbf{a}_t, \mathbf{s}_t, \xi_t)$. We define a fitting penalty function $\tilde{c}(\mathbf{s}_j, \pi_j, \hat{V})$ in terms of a fitting error function in (3.10):

$$\hat{V}(\mathbf{s}_j) - \Gamma \hat{V}(\mathbf{s}_j) = \sum_{j \in \mathcal{S}_i} (\alpha_j^* - a_j) \langle \Psi(\mathbf{s}_j), \Psi(\mathbf{s}_j) \rangle - \pi(\mathbf{a}, \mathbf{s}_j, \xi)$$

and for exact fixed point condition $\tilde{c}(\mathbf{s}_j, \pi_j, V)$.

Condition 3.9. The form $\tilde{c}(\mathbf{s}_t, \xi_t, \hat{V}) = \hat{V}(\mathbf{s}_t) - \Gamma \hat{V}(\mathbf{s}_t)$ in (3.10) is independent across t .

Theorem 3.10. *Given conditions 3.7 to 3.9, for any action \mathbf{a} given state \mathbf{s} based on T observations, the limit*

$$\lim_{T \rightarrow \infty} \Pr \left\{ \frac{1}{T} \sum_{t=1}^T \left| \tilde{c}(\mathbf{s}_t, \pi(\mathbf{a}, \mathbf{s}_t, \xi_t), \hat{V})|_c - \tilde{c}(\mathbf{s}_t, \pi(\mathbf{a}, \mathbf{s}_t, \xi_t), V) \right| > 4\epsilon \right\}$$

equals zero, where \hat{V} is $\rho^T \Phi(\mathbf{s})$ in equation (3.13) and V is the solution in the exact fixed point condition. We say that the \hat{V} is non-trivially consistent to V .

The concept of *nontrivial* consistency is proposed by Vapnik and Chervonenkis (Vapnik, 1998). This is the concept of consistency over all tractable candidate functions where classic consistency requires selection over all admissible functions for a given sample. While nontrivial consistency requires that the induction principle be consistent even after the “best” functions have been removed. Because these “best” functions are usually the best in some given samples but not the best in other test samples, moreover, these “best” functions are often much more complicated than the second best functions. Theorem (3.10) is another version of law of large numbers which is uniform over all second-best candidate functions. Vapnik and Chervonenkis propose this concept to rule out any function which uniformly does better than all other functions within sample but has little predictive power out of sample.

Remark 3.11. Theorem 3.10 is also a very useful condition for capacity control. The capacity term is a property of the function class which can be measured by entropy numbers. The robustness concern depresses the growth rate of entropy numbers⁶ so that it avoids using functional forms that are too complicated. Since the ϵ -approximation makes the fixed point condition converge to a small neighborhood of the exact fixed point condition, the simplification of introducing ϵ does not sacrifice too much consistency. In addition we have a computationally more efficient expression, a linear-quadratic programming problem.

3.3.2 Iterative Policy Algorithm

Finding the optimal policy function σ in (3.3) is often considered as a dual problem of (3.3) itself. The common approach of defining policy functions in the dynamic discrete choice literature is to find out a specified conditional choice probability. However, the use of ϵ -fixed point conditions in our framework will make the computation of conditional choice probability *invalid*.

The conditional choice probability is used to set up the transition density of the likelihood function (Rust, 1987; Aguirregabiria and Mira, 2007) or to derive a closed form policy function (Hotz and Miller, 1993; Bajari et al., 2007) or both. In the first step of our approach, we use the ϵ -approximation to nonparametrically solve the dynamic programming problem. If we derive the choice probability from some parametric likelihood, then the bias from ϵ -approximations will be accumulated. Thus it is impossible to derive a closed form policy function from the approximating value function in our framework.

The use of a conditional choice probability is to recover the policy function of (3.3). But it often causes a huge computational burden because of specifying the relation between the policy rule and structural parameters. Su and Judd (2011) give an extensive discussion of this issue and propose a computational alternative. Su and Judd (2011) pointed out that computing a conditional choice probability is not a necessary nor a sufficient way to find out the optimal policy rule in the dynamic discrete choice problem. A good solver in optimization programs will implicitly define the policy function and will implement the augmented likelihood through updating the optimization process.

⁶ Uniform convergence often requires an exponential bound which has a factor in terms of an entropy number. By Theorem 3.10, Y is a scale operator so that the entropy number changes by scaling.

We use an algorithm oriented method to search the optimal policy rule. With the explicit kernel-based expression (3.13), we can set up an iteration algorithm to obtain the estimated policy function $\hat{\sigma}$. Note that due to the ϵ -loss function, the estimated policy function is not necessarily equivalent to that in the recursive forward iteration method. The ϵ -tube avoids “over-fitting” the model thus $\Gamma\hat{V}$ will not fluctuate within the ϵ -wide tube. Also, an ϵ -optimal policy rule may be different from the one in the exact fixed point problem.

The algorithm is given below:

1. Set the initial policy \mathbf{a}_0 , select a basis function and its corresponding kernel K .
2. Choose a subset \mathcal{S}_a of states space and ensure that the transition density between any two of them are strictly positive.
3. Given the action \mathbf{a}_t (\mathbf{a}_0 for the first evaluation), calculate the kernel matrix $K^a := \{K_{kj} = \langle \Phi(\mathbf{s}_k), \Phi(\mathbf{s}_j) \rangle | \mathbf{s} \in \mathcal{S}_a\}$.
4. Given the profit function π evaluated at policy \mathbf{a} and kernel matrix K , solve the optimization problem (3.12) for α^* and α .
5. Apply (3.13) to obtain the ex ante value function $V(\mathbf{s})$, and then calculate the one-step policy improvement:

$$\mathbf{a}_{t+1} = \arg \max_{\mathbf{a}} (\alpha^* - \alpha) [K^a(\mathbf{a}_t) - \beta \sum \mathbf{p}(\mathbf{s}' | \mathbf{s}, \mathbf{a}) K^a(\mathbf{a}_t)].$$

Set next period action to \mathbf{a}_{t+1} , update the policy rule and then go back to Step-3.

The algorithm procedure is similar to the inner iteration of the NXFP (Rust, 1987) except that we implement the kernel-based optimization. Thus one might be tempted to compare the results of NXFP with (3.12). However, we have to emphasize that it is inappropriate to use the policy function from the original Bellman problem to judge the correctness of the policy function based on the ϵ -deviated Bellman problem. The two methods focus on different aspects. People who use the approximation to solve the exact Bellman equations have a lot of faith in the fixed point condition and are convinced the equilibria satisfy this condition. A standard method for obtaining the optimal policy function is by solving the following system of linear equations:

$$V^*(\mathbf{s}) = \sum_{\mathbf{a} \in \mathcal{A}} F^*(\mathbf{a}) [\pi^*(\mathbf{a}) + \mu^*(\mathbf{a})] + \beta \sum_{\mathbf{s}'} V^*(\mathbf{s}') \mathbf{p}^*(\mathbf{s}' | \mathbf{s}), \quad (3.14)$$

where μ^* is the expectation of ξ conditional on \mathbf{a} and $F^*(\mathbf{a})$ is the conditional choice probability. P^*, π^*, e^* are vectors that stack the corresponding state-specific elements. Star $*$ represents the elements associated with an equilibrium conditional choice probability. This method has been used for example by Aguirregabiria and Mira (2007).

People using the robust kernel-based optimization are less optimistic and are willing to consider actions that are less than perfect. It is quite intuitive that the solution will converge to (3.14) when ϵ is set to zero in (3.8). The ϵ can therefore be used to express the degree of confidence in Bellman's principle of optimality. We give the following Theorem to formalize this intuition.

Theorem 3.12. *When ϵ goes to zero, the approximating value function obtained by (3.13) is equivalent to the solution $V^*(\mathbf{s})$ in (3.14). The iterative policy algorithms are also equivalent.*

The iterative policy algorithm may not be satisfactory even if the value function approximation is good. A small ϵ might give a highly fluctuating policy function and then the wiggling policy function induces many local equilibria while a large ϵ may give a poor approximation for the value function. Since the iterative policy algorithm depends on the approximating value functions, choosing ϵ too small or too large leads to serious consequences. One may expect the policy iteration not to be a robust approach, in other words, policy iteration is sensitive to the selection of ϵ . To obtain a stable policy function, either one needs to have very precise knowledge on ϵ or impose constraints on the shape of the policy function. Cai and Judd (2010) show that monotonic and concave constraints on policy functions significantly improve the stability of solutions.

3.4 THE SECOND STEP ESTIMATION

In the preceding section, θ is treated as a structural parameter in Equation (3.2) and thus for the integrated Bellman equation (3.4) θ is incorporated in the integrated value function and ρ accommodates the robustness concerns in the ϵ -fixed point condition (3.6). We need to establish the relation between these two systems in order to estimate θ . However, there are two obstacles. First, one has no prior information about the form of the distribution of the random variable ξ , nor of the underlying parameter space. Second, the pseudo parameter ρ is irrelevant

for θ in the first step. In this section, we formulate these obstacles as another constrained optimization problem and then solve it.

The nonparametric smoothing densities are feasible solutions for the first obstacle. Empirical Likelihood (EL Owen, 1988, 1990, 2001) generates a so-called implied density function based on model constraints. EL distributes weights according to the imposed constraints. The input values satisfying the model constraints will be assigned to higher weights while those violating the model constraints will be assigned to lower weights or be penalized to zero if it is an outlier.

We handle the second obstacle by constructing a relation between nonparametric solutions \hat{V} and parametric functions of θ . However, the construction will render the EL or GMM method infeasible, because of a *non-differentiable* moment constraint. In this section, we use a local type EL estimator to solve this problem. For theoretical properties of this method, please refer to Chapter 2.

In Section 3.4.1, we will replace observations $\mathbb{E}_{\xi}\pi(\cdot, \xi)$ in (3.3) with a parametric function $\mathbb{E}_{\theta}\pi(\cdot, \xi)$ in order to combine the expression of θ and ρ . The new expression of the fitting error is a combination of nonparametric and parametric forms:

$$\rho^T \Phi(\mathbf{s}) - \mathbb{E}_{\theta}\pi(\sigma(\mathbf{s}, \xi), \mathbf{s}, \xi) + \beta \int \left[\rho^T \Phi(\mathbf{s}') \right] dP(\mathbf{s}' | \mathbf{s}, \mathbf{a}),$$

where for each observation π_t , $\sigma(\mathbf{s}_t, \theta) = \mathbf{a}_t$ is the sample action. In boundedly rational cases, there will be a set of strategies σ satisfying the ϵ -equilibrium condition, but here we will only consider a local optimal case when the fitting error is zero for each σ in this class. The optimized $\hat{\theta}$ for the zero-fitting error will be interpreted as the structural estimator. The problem in this new construction, as well as in many control oriented problems, is the discontinuity of the policy function $\sigma(\theta, \mathbf{s})$ over θ , e.g. kinks in choice functions. This issue will be enlarged when we plug in a “biased” approximating representation $\rho^T \Phi(\mathbf{s})$. Thus, in subsection 4.2 we need to consider an estimation that is robust towards non-differentiability.

3.4.1 The Semi-parametric Constraints

In parametric models, the constraint for the second step estimation usually forms a fixed point condition via the conditional choice probability (Pesendorfer and Schmidt-Dengler, 2003; Aguirregabiria and Mira, 2007) or an equilibrium condition of

the policy function and structural parameters (Su and Judd, 2011; Bajari et al., 2007). A nonparametric model does not specify the functional form of the distribution of ξ nor the dependence on θ , thus we have to seek an alternative constraint.

We recall that the approximating integrated Bellman equation for parametric models is:

$$\hat{V}(\mathbf{s}) = \mathbb{E}_\theta \left[\pi(\mathbf{a}, \mathbf{s}, \xi) + \beta \int \hat{V}(\mathbf{s}') dP(\mathbf{s}' | \mathbf{s}, \mathbf{a}) \right] = \rho^T \Phi(\mathbf{s}). \quad (3.15)$$

If we replace $\pi(\mathbf{a}, \mathbf{s}, \xi)$ with $\bar{\pi}(\mathbf{s}, \mathbf{a}) + \xi$ by Condition 3.2, the expression becomes

$$\begin{aligned} \rho^T \Phi(\mathbf{s}) &= \mathbb{E}_\theta \left[\bar{\pi}(\mathbf{a}, \mathbf{s}) + \xi(\mathbf{a}) + \beta \int \rho^T \Phi(\mathbf{s}) dP(\mathbf{s}' | \mathbf{s}, \mathbf{a}) \right], \\ &= \bar{\pi}(\mathbf{a}, \mathbf{s}) + \mathbb{E}_\theta \xi(\mathbf{a}) + \beta \int \rho^T \Phi(\mathbf{s}) dP(\mathbf{s}' | \mathbf{s}, \mathbf{a}). \end{aligned} \quad (3.16)$$

The equality follows from the ϵ -equilibrium condition 3.3. Rewrite equation (3.16) as:

$$\rho^T \Phi(\mathbf{s}) - \left[\bar{\pi}(\mathbf{a}, \mathbf{s}) + \beta \int \rho^T \Phi(\mathbf{s}) dP(\mathbf{s}' | \mathbf{s}, \mathbf{a}) \right] = \int \xi dG_\theta(\xi | \mathbf{a}) \quad (3.17)$$

With the ergodicity condition on ξ_t and Fubini's Theorem, the sample average of equation (3.17) has the following property:

$$\begin{aligned} & \frac{\sum_t^T [\rho^T \Phi(\mathbf{s}_t) - (\bar{\pi}(\mathbf{a}, \mathbf{s}) + \beta \int \rho^T \Phi(\mathbf{s}_t) dP(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}_t))]}{T} \\ &= \int \left[\sum_t \xi_t / T \right] dG_\theta(\xi | \mathbf{a}) \rightarrow \int \xi dG_\theta(\xi | \mathbf{a}). \end{aligned} \quad (3.18)$$

where G_θ is the parametric distribution with unknown parameter.

The ϵ -equilibrium corresponds to a class of policy functions $\sigma(\mathbf{s}, \theta)$. We will only consider the robust decision function, where value function of MPE $\sigma(\mathbf{s}, \theta)$ will be covered by the ϵ -tube (3.6) such that $\mathbb{E}_\theta \xi(\mathbf{a}) < \epsilon$. Given fixed ϵ , we only need to minimize the following fitting error $m_t(\theta | \mathbf{a})$ w.r.t. θ :

$$\left[\rho^T \Phi(\mathbf{s}_t) - \left(\bar{\pi}(\mathbf{a}, \mathbf{s}) + \beta \int \rho^T \Phi(\mathbf{s}_{t+1}) dP(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}_t) \right) \right] - \mathbb{E}_\theta \xi(\mathbf{a}).$$

As in equation (3.16), the limit of the sample average of $m_t(\theta | \mathbf{a})$ over t is an integration w.r.t. G . When $\theta = \theta_0$, $dG_0(\xi | \mathbf{a})$ implies

that $\int m(\theta_0)dG_0(\xi|\mathbf{a})$ conditional on the action trajectory \mathbf{a} . The empirical counterpart of G can be estimated by EL.

In practice, heterogeneous action trajectories of \mathbf{a} brings a great difficulty of obtaining an empirical counterpart of $G_0(\xi|\mathbf{a})$. Beside using $G_0(\xi|\mathbf{a})$, one may prefer to consider $dG_\theta(\xi) = dG_\theta(\xi|\mathbf{a})dP(\mathbf{a})$ where $P(\mathbf{a})$ is the choice probability. Then the conditional moment constraint function $m_t(\theta|\mathbf{a})$ is replaced by $m_t(\theta|\mathbf{a})p(\mathbf{a})$ which we denote $m_t(\theta)$. However, in this case, one will confront an issue of executing the derivative of

$$\int m_t(\theta|\mathbf{a})dG_\theta(\xi|\mathbf{a})dP(\mathbf{a})$$

when \mathbf{a} is discrete. To see this problem, we need to point out that $dG_\theta(\xi) = dG_\theta(\xi|\mathbf{a})dP(\mathbf{a})$ is a mixture distribution. Mixtures of log-concave densities may be log-concave, but in general they are not. For instance, if $\mathbf{a} = \{0,1\}$ and $p(\mathbf{a}) = P(\mathbf{a} = 0)$, the location mixture of standard univariate normal densities $dG_\theta(\xi) = \phi(\xi)$

$$\phi(\xi)p(\mathbf{a}) + \phi(\xi - c\mathbf{a})(1 - p(\mathbf{a}))$$

is log-concave if and only if $c\mathbf{a} \leq 2$. Bi-modal or multi-modal distributions lead to non-continuous derivative and zero-Hessian of the log-likelihood of $G_\theta(\xi)$. The worst scenario is the non-differentiability⁷ of $\int m_t(\theta|\mathbf{a})dG_\theta(\xi)$ at the extrema. In the following sub-section, we propose an approach to approximate the derivative of the likelihood ratio of $G_\theta(\xi)$ without using differentiation techniques.

3.4.2 Empirical Likelihood and Local Empirical Likelihood

Let $m_t(\theta)$ denote $\int m_t(\theta|\mathbf{a})dP(\mathbf{a})$. The constraint (3.18) is used to identify the unknown distribution function $G(\xi)$ and speed up the computation convergent rate. We apportion the probabilities $g = (g_1, \dots, g_T)$ for the distribution G . The sum of empirical log-likelihood contribution is $\sum_t \log Tg_t$. In addition, g should satisfy

⁷ The non-differentiability implies that there is no closed form derivative on \mathbb{R} . It is possible that the functional derivative is feasible and tractable.

the common requirements of probabilities such that $g_t \geq 0$ and that $\sum_t g_t = 1$. The EL criterion is:

$$\begin{aligned} & \max_{\theta} \sum_{t=1}^T \log T g_t, \\ & \text{s.t. } g_t \geq 0, \quad \sum_{t=1}^T g_t = 1, \\ & \quad \sum_{t=1}^T g_t m_t(\theta) = 0, \end{aligned} \quad (3.19)$$

where the constraint $\sum_t g_t m_t(\theta) = 0$ expresses the fact that in the second stage we pursue zero fitting error for the ϵ -equilibrium. The Lagrangian is:

$$L' := \sum_{t=1}^T \log T g_t - T\lambda \sum_{t=1}^T m_t(\theta) + \gamma \left(\sum_{t=1}^T g_t - 1 \right). \quad (3.20)$$

With the help of the Lagrange multiplier λ , we have:

$$\tilde{g}_t(\theta) = \frac{1}{T} \frac{1}{1 + \lambda m_t(\theta)}, \quad (3.21)$$

the so-called implied density g . Substituting (3.21) into (3.19), we have a dual representation of problem (3.19):

$$\begin{aligned} & \min_{\theta} - \sum_{t=1}^T \log T (1 + \lambda m_t(\theta)), \\ & \text{s.t. } \frac{1}{T} \sum_{t=1}^T \frac{m_t(\theta)}{1 + \lambda m_t(\theta)} = 0. \end{aligned} \quad (3.22)$$

Problem (3.22) is a standard nonlinear optimization problem. The outer-loop of the optimization is to minimize minus the empirical log-likelihood with respect to θ and the inner-loop is to obtain numerical solution of λ .

The outer-loop optimization in (3.22) could be discontinuous due to the discrete feature of $\sigma(\mathbf{s}, \theta)$. Thus the outer-loop search in optimization is unstable. Let H and s denote the Hessian and gradient function of $\sum_t \log T (1 + \lambda m_t(\theta))$ with respect to θ . The $k + 1$ -step Newton iteration used in most computational methods is

$$\theta^{(k+1)} = \theta^{(k)} - H(\theta^{(k)})^{-1} s(\theta^{(k)}). \quad (3.23)$$

The evaluation of Hessian matrix $H(\theta^{(k)})$ requires the second derivative of the log-likelihood function. The numerical derivative is difficult to implement and time consuming because of

nonlinearity and non-differentiability. The Hessian matrix in this region is so flat that singularity problems may occur.

By the localization technique, we have a linear-quadratic representation for the log-likelihood ratios:

$$-\min_{\tau} \sum_{t=1}^T \left[\tau^T S_t - \frac{1}{2} \tau^T M_t \tau \right]. \quad (3.24)$$

The construction of S and M of this representation are computed without using derivative argument. The construction of this representation will be described in the next subsection.

The term “local” here is a counterpart of “differential”. One fixes a particular θ_0 and investigates what happens to likelihood ratio functions with parameter values of $\theta = \theta_0 + \delta_T \tau$, with $\delta_T \rightarrow 0$. Here, δ_T is a kind of “differentiation rate”. We will fix the differential rate to $T^{-1/2}$ such that $\theta + T^{-1/2} \tau$ and we will call τ the local parameter.

Remark 3.13. In the parametric case, people usually assume that the unknown private shocks have multinomial distributions. The implied density (3.22) is far more flexible than a parametric density because it can assign the weights arbitrarily. In addition, if the log-likelihood ratio in (3.19) is replaced by entropy $\sum_t T g_t \log T g_t$, Kitamura and Stutzer (1997) show that the implied density will become

$$\tilde{g}_{t,ET}(\theta) = \frac{\exp \lambda m_t(\theta)}{\mathbb{E}_t \exp \lambda m_t(\theta)}, \quad (3.25)$$

which is similar to the multinomial choice probability.

3.4.3 Construction of the Second-Step Estimator

Partition the parameter space of θ into several multi-dimensional grids and select one particular value on this grid. Denote the selected value θ^* and let

$$\Lambda_T(\theta_1, \theta_2) = \sum_{t=i}^T \log(\tilde{g}_{it}(\theta_1) / \tilde{g}_{it}(\theta_2)).$$

Now run the following scheme at every grid point θ^*

1. M_T with $M_{T,q,p} = u_q^T M_T u_p$, $q, p = 1, 2, \dots, l$, given by

$$M_{t,q,p} = - \left\{ \Lambda_T[\theta^* + \sqrt{T}(u_q + u_p), \theta^*] - \Lambda_T[\theta^* + \sqrt{T}u_q, \theta^*] - \Lambda_T[\theta^* + \sqrt{T}u_p, \theta^*] \right\}$$

where $\{u_1, \dots, u_l\}$ is a linearly independent set of directional vectors in \mathbb{R}^l selected in advance.

2. Construct a linear term $S_T = \{S_{T,q}\}$, $q = 1, 2, \dots, l$, by the linear-quadratic approximation function:

$$S_{T,q} = \Lambda_T[\theta^* + \sqrt{T}u_q, \theta^*] + \frac{1}{2}M_{T,q,q}.$$

Since all the values on the RHS are known.

3. Construct a one-step improved estimator:

$$\tilde{\theta} = \theta^* + \sqrt{T}M_T^{-1}S_T,$$

4. Obtain the value of $\sum_t \log n \tilde{g}(\tilde{\theta} + T^{-1/2}\tau)$ and compare it with the values from other grid points. If $\tilde{\theta}$ returns a higher likelihood value, it will be selected.

The constructed Hessian type matrix M_T in step 2 is invertible. The advantage of the local method is that the gradient vectors and Hessian matrices are available without any differentiation operation. As we mention earlier, the mixing parametric choice distribution implies an irregular shape of the likelihood function so that gradient and Hessian is less tractable than in the usual case. As one can see, the global irregular shape plays no role in our local quadratic construction and the estimator construction. The calculation of M at every fixed θ^* is independent of the numerical second-order derivatives.

The theoretical properties of this local EL estimator has been derived in Chapter 2. It is shown there that the estimator has a normal limiting distribution within the local neighborhood of the true θ and is asymptotically optimal. The result does not require globally smooth functions and the optimization only depends on local EL values. The estimation concerns local parameters rather than θ directly, thus it avoids evaluating non-differentiable or highly nonlinear functions of θ .

3.5 NUMERICAL ILLUSTRATION

We consider a simple application of Rust's model: optimal replacement of bus engines (Rust, 1987). This is a single agent dynamic discrete choice model, but it is a useful starting example to illustrate how the semiparametric algorithm works. We will use 117 observations of monthly data for the 1975 GMC

model 5308 buses, 37 in total. This is data set a530875 of Rust (1987).

In this setting, the maintenance manager of this fleet of 37 buses has to decide for each bus i how long to operate it before replacing its engine with a new one. The state s_{it} variable is the accumulated miles of the engine in bus i at time t . The decision variable a_{it} is whether to replace the engine $a_{it} = 1$ or maintain the engine $a_{it} = 0$ in bus i at time t . When a bus engine is replaced, it is as good as a new, so the state of the system regenerates to $s_{it} = 0$ when $a_{it} = 1$. The private shock $\zeta_{it}(a_{it})$ in this case is the unforeseen cost to bus i in period t given decision a_{it} . The profit function π in our notation for the current example is equal to minus the cost since the maintenance manager takes the number of miles as given. Rust (1987) refers to this as the utility. This π is assumed to be additively separable and is given by:

$$\pi(a_{it}, s_{it}, \theta_1, \theta_2) = \begin{cases} -RC - C(0, \theta_1) - \zeta_{it}(1) & \text{if } a_{it} = 1 \\ -C(s_{it}, \theta_1) - \zeta_{it}(0) & \text{if } a_{it} = 0, \end{cases}$$

where $C(\cdot)$ is an engine's operating and maintenance cost function. The transition density for s_{it} depends only on miles driven at the beginning of time period $t + 1$:

$$p(s_{it+1} | s_{it}, a_{it}) = \begin{cases} g(s_{it+1} - 0) & \text{if } a_{it} = 1, \\ g(s_{it+1} - s_{it}) & \text{if } a_{it} = 0, \end{cases}$$

where $g(\cdot)$ is a known probability density function that may depend on a parameter θ_2 . We will use $g(\cdot) = \theta_2 \exp[-\theta_2(\cdot)]$. In our approximation step (step-1), θ_2 is treated as a fixed value. The specification of cost function is one of the following:

$$\begin{aligned} \text{Quadratic:} & \quad C(s, \theta_1) = \theta_{11}s + \theta_{12}s^2, \\ \text{Power:} & \quad C(s, \theta_1) = \theta_{11}s^{\theta_{12}}, \\ \text{Mixed:} & \quad C(s, \theta_1) = \theta_{11}/(1.1 - s) + \theta_{12}s^{1/2}. \end{aligned} \tag{3.26}$$

Rust (1987) estimates θ_2 and θ_1 separately because only θ_2 affects the transition function. In step-1, we use the estimated result of θ_2 in Rust (1987) as the initial value of θ_2 and then simulate the transition density $p(s', s | \theta_2, a)$. In step-2 we update the values of θ_1 via (3.24). We do this iteratively until the algorithm converges. Thus in step-1, the transition density is taken as known and then new status is given in step-2.

The last equation of (3.26) is slightly different from the original one where the constant is set to 91. The reason is that

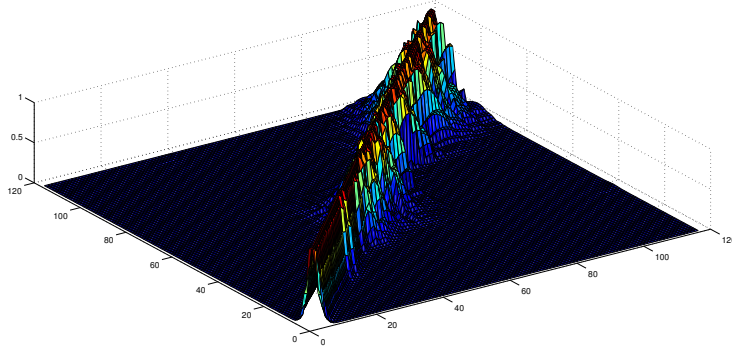


Figure 2: A realization of the kernel function $K(\cdot, \cdot)$ of the cumulated miles $\sum_i s_{it}$ and $\sum_i s_{it'}$ over 117 periods. The x -axis and the y -axis denote time periods t and t' . The z -axis denotes the transition kernel value of $K(\sum_i s_{it}, \sum_i s_{it'})$ for different t and t' .

we re-scale the state variables to the $[0, 1]$ -interval rather than discretize them into 90 states. The advantage of scaling is that it avoids the case where states in greater numeric ranges dominate those in smaller numeric ranges. Another advantage is that it avoids numerical difficulties in kernel calculation. In kernel calculations, the inner products of basis functions may generate large values which cause problems in the numerical operation.

The kernel matrix used in the approximation of the value function (see Equation (3.13)) is evaluated via $\langle \Psi(\mathbf{s}), \Psi(\mathbf{s}) \rangle$. It can be calculated in each period t as is done in Figure 2. The $\Psi(\mathbf{s})$ equals $\Phi(\mathbf{s}) - \beta \sum \Phi(\mathbf{s}') \mathbf{p}(\mathbf{s}' | \mathbf{s}, \mathbf{a})$ with a fixed discount factor $\beta = 0.98$. The transition density $\mathbf{p}(\mathbf{s}' | \mathbf{s}, \mathbf{a})$ changes over time simply because the action \mathbf{a} changes and this causes the kernel defined by $\langle \Psi(\mathbf{s}), \Psi(\mathbf{s}) \rangle$ to change over time. Figure 2 shows that at the beginning stage the distribution is quite homogeneous but when the odometers accumulate to certain miles the transition distribution among states becomes divergent. Suppose that \bar{s} is a standard mileage for engine replacement such that it satisfies $V(\bar{s}, \mathbf{a} = 1) = V(0, \mathbf{a} = 0)$. When the engine is close to this stage, the manager has a greater tendency to replace it. Therefore, the transition among states has significant dissimilarities at the later stage.

The sensitivity loss ϵ controls the goodness of fit of the approximation and furthermore affects the performance of the predictor. To analyze the effects of ϵ , we use different ϵ in the power cost function case. We compare approximation and

ϵ	MSE	r^2	Basis Num
0.005	0.609	0.736	108
0.050	0.614	0.739	63
0.500	0.960	0.690	28

Table 1: Sensitivity analysis of the approximation. By tuning the value of ϵ , we can examine the MSE of the approximation. Changes of the width of the ϵ -tube also affect the complexity of the approximation via the number of used basis functions.

inference ability of this kernel function based on test and estimation samples. With calibrated C , b and θ , Table 1 and Figure 3 give the results of goodness of fit and the predictor ability of ϵ -approximation. The first 50% data is used for estimation while the other is for prediction. The prediction ability is measured in two terms, Mean Squared Error (MSE) and squared correlation coefficient (r^2) which are $T^{-1} \sum_t (\rho^T \Phi(s_t) - \pi_t)^2$ and

$$r^2 = \frac{(T \sum_t \rho^T \Phi(s_t) \pi_t - \sum_t \rho^T \Phi(s_t) \sum_t \pi_t)^2}{(T \sum_t (\rho^T \Phi(s_t))^2 - (\sum_t \rho^T \Phi(s_t))^2) (T \sum_t \pi_t^2 - (\sum_t \pi_t)^2)} \quad (3.27)$$

respectively. MSE for test sample can be considered as the information loss due to the inaccurate prediction. The quantity r^2 measures the linear relationship between the approximating value function and the parametric profit function. It is obvious that a small ϵ makes the kernel over-fit and thus ask for more information to construct the fitting. MSE and r^2 are not significantly different for $\epsilon = 0.005$ and $\epsilon = 0.05$, the former case nearly uses the whole 117 sample points while the later case only asks for 63 out of 117. One can find out there is almost no loss by cutting 50% evaluations of samples. The simpler approximation is favored because a complicated basis function leads to intractable kernel computations and optimizations.

From $\epsilon = 0.05$ to $\epsilon = 0.5$, the prediction accuracy decreases slightly, which, however, is caused by an unpredictable downside shift in the averse direction of the trend. From Figure 5, we can realize that the curve shape of $\epsilon = 0.5$ is more flexible than that of $\epsilon = 0.05$. In addition, the basis number falls to 28, a big gain in simplicity with an insignificant cost of the prediction power. Such a benefit does not appear by increasing ϵ to 1. When $\epsilon = 1$, the fitting has poor MSE and r^2 values. Figure 4 shows the values of $(\alpha - \alpha^*)$ for the sub-state space.

3.6 CONCLUSION

Table 2 describes the performance of different cost functions. The cost function with power functional form shares similar behaviors with that of the quadratic form. The mixed functional form gives unsatisfied outputs for larger structural parameter values. The reason is that the power function and the quadratic function can be easily approximated via linear basis functions but not for the mixed function. To obtain a satisfied fitting of the mixed function, a bigger number of basis functions is required. Therefore the vectors used in the mixed function case is significant larger than the other cases. From Figure 3, we can find out that the value function curves do not seem to converge for large parameters. Because large parameter enlarges the fluctuation and thus causes divergent plots.

The average accuracy of predicting the validation sets is the cross validation accuracy. We also test the sensitivity of the choice of loss functions for the fixed point condition. It seems that, a small ϵ will lead to an exact classification policy rule. In this case, each decision is strictly isolated with the others and has a narrow inference interval. On the other hand, if we allow a flexible ϵ for the fixed point condition, we can use a simple policy rule for classification and have much larger confidence region for the whole data set.

3.6 CONCLUSION

In this chapter, we propose a new two-step estimation approach for dynamic discrete choice models with boundedly rational agents. In the first step, the method embeds robust decisions caused by bounded rationality into the Bellman's optimal principle. In the second step, the estimated model is accompanied by non-standard moment restrictions. We use the method developed in Chapter 2 to estimate this model. The solution of this problem is based on an ϵ -approximating value function with locally estimated structural parameter. We show that both the value function and the log-likelihood ratio of the local parameter have flexible representations. The result reflects the robustness concern of agents' decision processes.

3.6 CONCLUSION

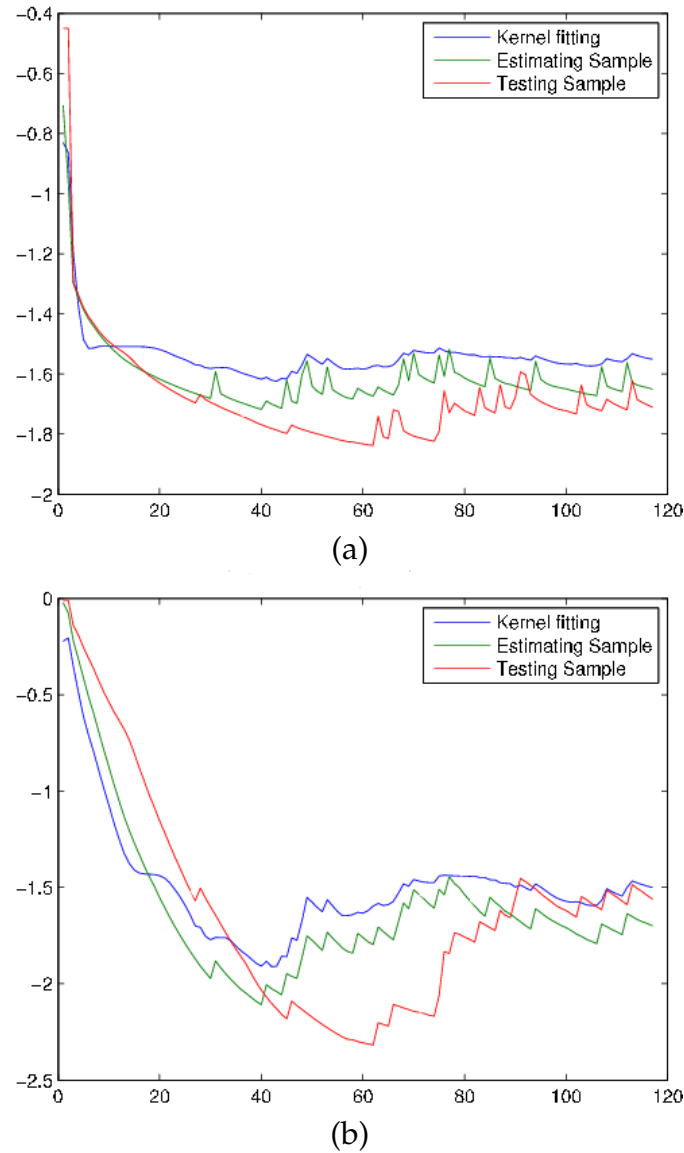


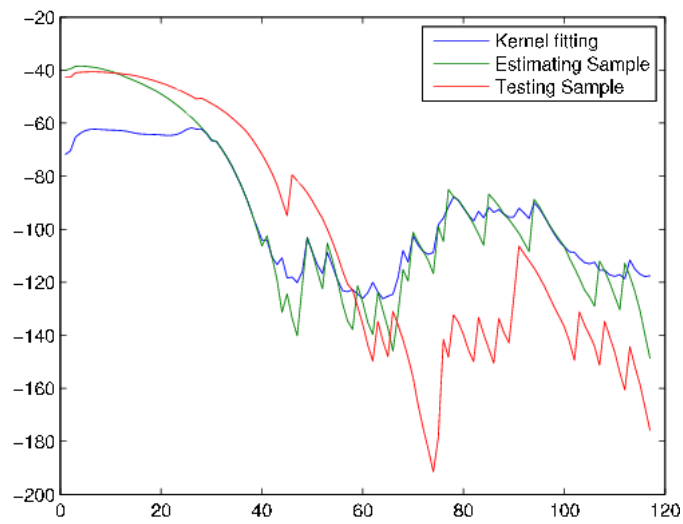
Figure 3: Prediction ability of the approximation $\rho^T \Phi(\mathbf{s})$ under different functional forms $C(s, \theta)$: (a) Power function $\theta_{11} = 0.1$, $\theta_{12} = 0.1$ (b) Quadratic function with $\theta_{11} = 0.32$, $\theta_{12} = 0.5$ (c) Mixed function $\theta_{11} = 2.5$, $\theta_{12} = 1.1$. One can compare the approximation $\rho^T \Phi(\mathbf{s})$ (in blue) with the value function of the used sample (in green) the testing sample (in red). The x -axis denotes the time period and the y -axis denotes the cost in the value function.

3.6 CONCLUSION

Function type	Param θ_{11}, θ_{12}	MSE	Nr Basis Func
Power	0.1, 0.1	0.031	21
Quadratic	0.3, 0.5	0.719	27
Mixed	2.5, 1.1	0.041	27

Function type	r^2	Likelihood
Power	0.840	-51.908
Quadratic	0.665	-17.957
Mixed	0.645	-20.909

Table 2: Likelihood Estimation: The results of the second step estimation.



(c)

Figure 3 Continued

3.6 CONCLUSION

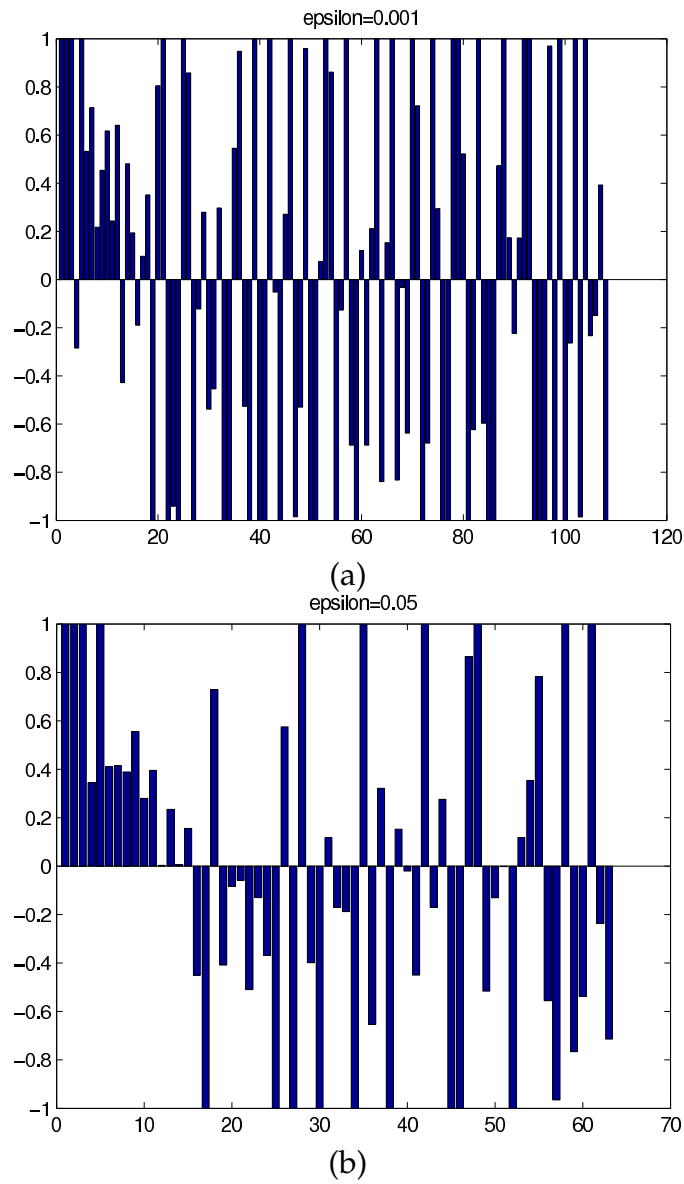
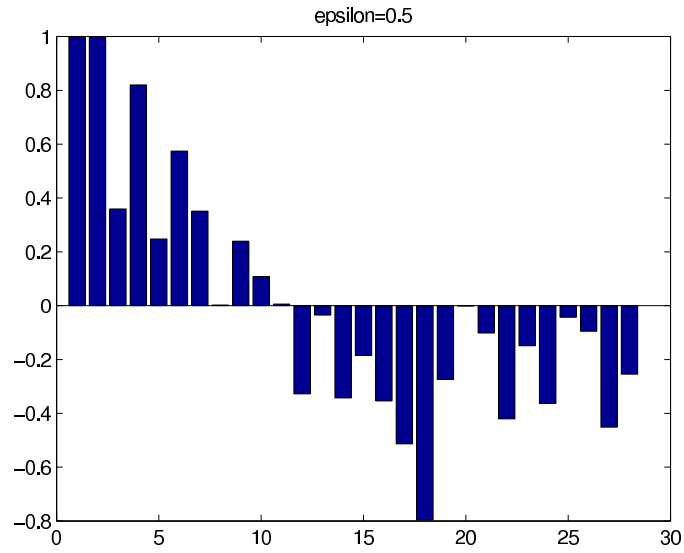
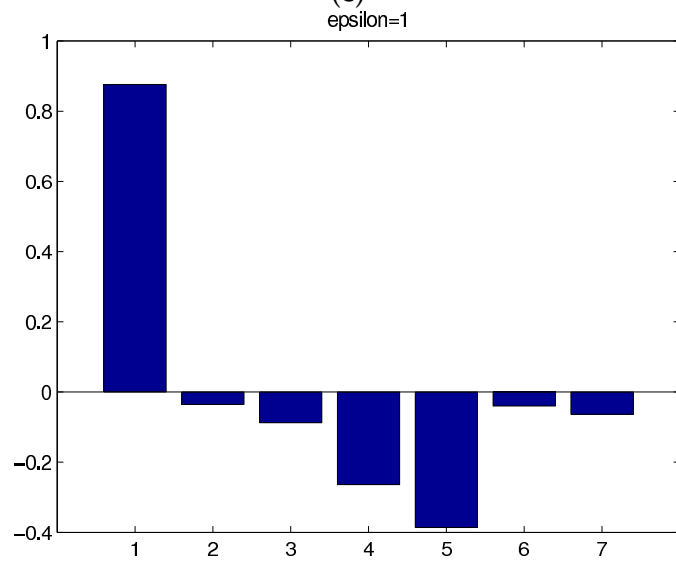


Figure 4: The sensitivity analysis of ϵ reflects on the change of $(\alpha - \alpha^*)$ since $(\alpha - \alpha^*)$ is the dual parameter of ϵ . We have the value of $(\alpha - \alpha^*)$ under (a) $\epsilon = 0.001$, (b) $\epsilon = 0.05$ (c) $\epsilon = 0.5$ and (d) $\epsilon = 1$. The y -axis denotes the value of $(\alpha - \alpha^*)$. The x -axis denotes the number of basis functions used.

3.6 CONCLUSION



(c)



(d)

Figure 4 Continued

3.6 CONCLUSION

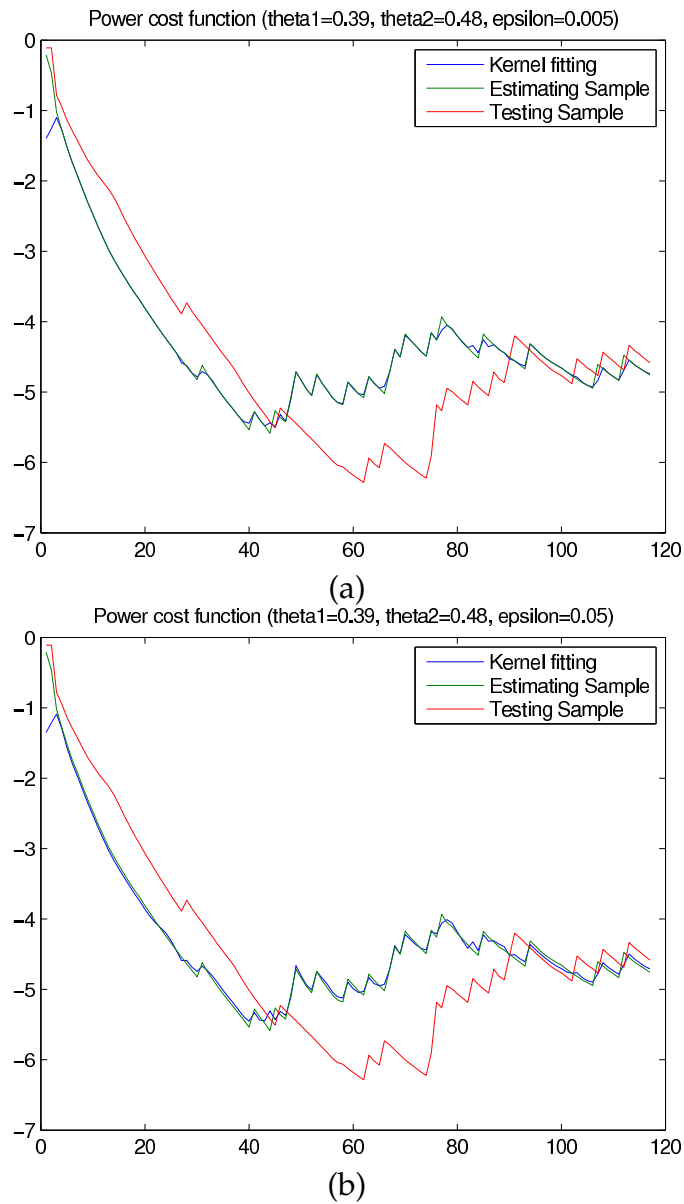


Figure 5: Different ϵ affect the results of the approximation $\rho^T \Phi(\mathbf{s})$. (a) $\epsilon = 0.001$, (b) $\epsilon = 0.05$ (c) $\epsilon = 0.5$ and (d) $\epsilon = 1$. The x -axis denotes the time period and the y -axis denotes the cost in the value function.

3.6 CONCLUSION



(c)



(d)

Figure 5 Continued

APPENDIX TO CHAPTER 3

PROOF OF THEOREMS

Proof of Theorem 3.10

First we need to introduce the definition of approximation stability. The purpose of introducing this definition is to restrict the function class of our potential solution sets.

Approximation Stability:

The approximation function \hat{V} is a function of \mathcal{S} , if the \mathcal{S} is changed to \mathcal{S}^m , a new approximation follows, $\hat{V}_{\mathcal{S}^m}$ say. We will now define a replacement perturbation by replacing the m -th observation with a new s . So $\mathcal{S}^m := (\mathcal{S} \setminus \{s_m\}) \cup \{s_{new}\}$, where s_{new} is drawn from a relevant distribution. In our case, this is the empirical distribution. The stability means the difference between $\hat{V}_{\mathcal{S}^m}$ and \hat{V} is bounded. In the proof, we will give bounds for $\hat{V}_{\mathcal{S}^m} - \hat{V}$ and show that the approximation is stable.

Sketch of the proof:

The purpose of the proof is to show the fixed point condition with regularized Bellman operator $\Gamma\hat{V}$ converges uniformly to a neighborhood of the fixed point condition with $\hat{\Gamma}V$. A straightforward idea is to check the validity of the following argument:

$$\lim_{T \rightarrow \infty} \left| \sum_{t=1}^T (\hat{V}_t - \Gamma\hat{V}_t) / T - \mathbb{E}(V - \hat{\Gamma}V) \right|_{\epsilon} \rightarrow 0.$$

If an empirical process Theorem (e.g. Glivenko-Cantelli) can be applied to the left handside of the argument, then it would be simple. However, there are two difficulties of using an empirical process Theorem to prove the uniform convergence here. The first problem is the ϵ -loss function (ϵ -fixed point constraints) make $\hat{V}_j - \Gamma\hat{V}_j$ not exactly zero and the second is the effect from regularization. Fortunately, Theorem 12 from Bousquet and Elisseeff (2002) (BE Theorem) states that if changing observations

in a given loss function only leads to a bounded variation then there is a Hoeffding type bound⁸ for this function.

In our setting, the proof that we will follow can be stated in this way:

If Condition 3.7 (i) holds and

$$|\tilde{c}(\mathbf{s}, \pi, \hat{V}) - \tilde{c}(\mathbf{s}, \pi, \hat{V}_{S^m})| \leq \zeta \quad (3.28)$$

for all (s, π) where ζ depends on T then we have

$$\Pr \left\{ \left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) \right| > \epsilon + \zeta \right\} \quad (3.29)$$

$$\leq 2 \exp \left(-\frac{2T\epsilon^2}{(T\zeta + M^*)^2} \right). \quad (3.30)$$

So if ζ is of order $1/T$, then when $T \rightarrow \infty$ the exponential bound (3.30) goes to zero. The standard empirical process of $\tilde{c}(\mathbf{s}_t, \pi_t, V)$ has a Hoeffding bound such that

$$\Pr \left\{ \left| \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, V)) \right| > \epsilon \right\} \quad (3.31)$$

$$\leq N(\epsilon) \exp \left(-\frac{2T\epsilon^2}{M^{*2}} \right),$$

where $N(\epsilon)$ is the covering number for the function \hat{V} which is smoothed by a RBF kernel. When T increases, $\exp(-2T\epsilon^2/M^{*2})$ will decrease to zero at a faster rate than the bound in (3.30). Therefore, we can incorporate equation (3.31) into (3.29). Since the probability rule implies $2\Pr(A) \geq \Pr(A) + \Pr(B) \geq \Pr(A \cup B)$

⁸ The bound slightly differs from original Hoeffding's bound but has similar convergence rate.

B), we can combine the probability $\Pr(A)$ in equation (3.29) and $\Pr(B)$ in equation (3.31) to obtain the following expression⁹:

$$\begin{aligned} & 2\Pr \left\{ \left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) \right| > \epsilon + \zeta \right\} \\ & \geq \Pr \left\{ \left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) \right| - \right. \\ & \quad \left. \left| \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, V)) \right| > 2\epsilon + \zeta \right\} \quad (3.32) \end{aligned}$$

From the right handside of equation (3.32) and using the bound in (3.30), $|\mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, V))|$ is the population of

$$\left| \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) - \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) \right|.$$

⁹ The explicit derivation is as follows: $2\Pr(A) \geq \Pr(A \cup B)$ implies

$$\begin{aligned} & 2\Pr \left\{ \left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) \right| > \epsilon + \zeta \right\} \\ & \geq \Pr \left\{ \left| \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) - \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon \right| + \right. \\ & \quad \left. \left| \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, V)) \right| > \epsilon + \zeta \right\} \end{aligned}$$

Using triangular inequality, we have:

$$|x - y| - |z - q| \leq |x - y + z - q| = |x - q + z - y| \leq |x - q| + |z - y|$$

where $|z - y|$ is $\left| \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) - \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon \right|$ and $|x - q|$ is $\left| \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, V)) \right|$.

We can conclude¹⁰

$$\Pr \left\{ \left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) \right| > 4\epsilon + 2\zeta \right\} \quad (3.33)$$

$$\leq 4 \exp \left(-\frac{T\epsilon^2}{2(T\zeta + M^*)^2} \right).$$

By Condition 3.3,

$$\left| \frac{1}{T} \sum_{t=1}^T c(\mathbf{s}_t, \pi_t, \hat{V}) - \frac{1}{T} \sum_{t=1}^T c(\mathbf{s}_t, \pi_t, V) \right| \leq 2\epsilon$$

for V satisfying the Markov Perfect Equilibrium. If $\zeta \rightarrow 0$ as $T \rightarrow \infty$ then the proof ends.

Thus we need to find the bound in (3.28). In Example 1 and Appendix C of Bousquet and Elisseff (2002), they give an instruction how to use the subgradient of a constructed convex function to derive a stable bound for a support vector regression model. We follow this idea.

Proof. Let's define the specific form of the regularized function:

$$\begin{aligned} Y\hat{V}(\mathcal{S}) &= \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon + \frac{1}{2} \|\hat{V}\|^2 \\ &= \frac{1}{T} \sum_{t=1}^T \left| \rho^T \Psi(\mathbf{s}_t) - \pi_t \right|_\epsilon + \frac{1}{2} \left\| \rho^T \Phi(\mathbf{s}) \right\|^2 \\ &= \frac{1}{T} \sum_{t=1}^T \left| \sum_{i=1}^T (\alpha_i^* - \alpha_i) \langle \Psi(\mathbf{s}_i), \Psi(\mathbf{s}_t) \rangle - \pi_t \right|_\epsilon \\ &\quad + \frac{1}{2} \left\| \sum_{i=1}^T (\alpha_i^* - \alpha_i) K_{it} \right\|^2, \end{aligned}$$

¹⁰ The explicit derivation is as follows: Let

$\left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) \right| > \epsilon + \zeta$ be an event A and

$|\mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, V))| > \epsilon$ be an event B . The event A'

$$\left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) \right| - |\mathbb{E}(\tilde{c}(\mathbf{s}, \pi, \hat{V})) - \mathbb{E}(\tilde{c}(\mathbf{s}, \pi, V))| > 2\epsilon + 2\zeta$$

implies the event B'

$$\left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) \right| > 4\epsilon + 2\zeta$$

so that $\Pr(B') < \Pr(A')$ while $\Pr(A') < \Pr(A \cap B)$.

where $\Psi(\mathbf{s}) = \Phi(\mathbf{s}) - \beta \sum \Phi(\mathbf{s}') \mathbf{p}(\mathbf{s}' | \mathbf{s}, \mathbf{a})$. In order to optimize this function, we need the functional derivative of $|\tilde{c}(\mathbf{s}_j, \pi_j, \hat{V})|_\epsilon$ w.r.t. \hat{V} , see for the definition of the functional derivative and its associated notation:

$$\begin{aligned} [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) &:= \partial \left| \rho^T \Psi(\mathbf{s}_t) - \pi_t \right|_\epsilon \\ &= \begin{cases} 0 & \text{if } |\rho^T \Psi(\mathbf{s}_t) - \pi_t| \leq \epsilon \\ \frac{(\pi_t - \rho^T \Psi(\mathbf{s}_t))}{|\rho^T \Psi(\mathbf{s}_t) - \pi_t|} & \text{otherwise.} \end{cases} \end{aligned}$$

It is obvious that $||[\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})| \leq 1$. The derivative of the regularized function for the original sample \mathcal{S} is:

$$\partial Y \hat{V}(\mathcal{S}) = \frac{1}{T} \sum_{t=1}^T [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) + \rho^T \Phi(\mathbf{s}) = 0, \quad (3.34)$$

and for the replacement sample \mathcal{S}^m is:

$$\partial Y \hat{V}(\mathcal{S}^m) = \frac{1}{T} \sum_{t=1}^T [\delta] \tilde{c}(\mathbf{s}_j, \pi_j, \hat{V}_{\mathcal{S}^m}) + \rho_{\mathcal{S}^m}^T \Phi(\mathbf{s}_m) = 0. \quad (3.35)$$

Next, we construct an auxiliary convex function:

$$\begin{aligned} \mathcal{A}(f) &= \left\langle \frac{1}{T} \sum_{t=1}^T [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) - \frac{1}{T} \sum_{t=1}^T [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}_{\mathcal{S}^m}), f - \rho_{\mathcal{S}^m}^T \Phi(\mathbf{s}) \right\rangle \\ &\quad + \frac{1}{2} \left\| f - \rho_{\mathcal{S}^m}^T \Phi(\mathbf{s}_m) \right\|^2. \end{aligned}$$

It is obvious that when $f = \rho_{\mathcal{S}^m}^T \Phi(\mathbf{s}_m)$, $\mathcal{A}(\rho_{\mathcal{S}^m}^T \Phi(\mathbf{s}_m)) = 0$. Furthermore, the functional derivative of $\mathcal{A}(f)$ w.r.t. f is

$$\begin{aligned} \partial \mathcal{A}(f) &= \frac{1}{T} \sum_{t=1}^T [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) - \frac{1}{T} \sum_{t=1}^T [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}_{\mathcal{S}^m}) \\ &\quad + \left(f - \rho_{\mathcal{S}^m}^T \Phi(\mathbf{s}_m) \right) \\ &= \frac{1}{T} \sum_{t=1}^T [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) + f. \end{aligned}$$

The second equation uses (3.35).

By (3.34), the minimum of $\mathcal{A}(f)$ is achieved at $\rho^T \Phi(\mathbf{s})$. because $\partial \mathcal{A}(\rho^T \Phi(\mathbf{s}_m)) = 0$ and $\mathcal{A}(f)$ is a convex function. Thus $\mathcal{A}(f) \leq 0$.

$$\begin{aligned}
 & T \times \left\langle \partial \Upsilon \Gamma \hat{V}(\mathcal{S}) - \partial \Upsilon \Gamma \hat{V}(\mathcal{S}^m), \rho^T \Phi(\mathbf{s}) - \rho_{\mathcal{S}^m}^T \Phi(\mathbf{s}_m) \right\rangle \\
 = & \sum_{t \neq m}^T \left[[\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) - [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}_{\mathcal{S}^m}) \right] \times \left[\rho^T \Psi(\mathbf{s}_t) - \rho_{\mathcal{S}^m}^T \Psi(\mathbf{s}_t) \right] \\
 & + [\delta] \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}) \times \rho^T \Psi(\mathbf{s}_m) - [\delta] \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}_{\mathcal{S}^m}) \times \rho_{\mathcal{S}^m}^T \Psi(\mathbf{s}_m)
 \end{aligned} \tag{3.36}$$

For a convex function $f(a) + (b - a)f'(a) \leq f(b)$ and $f(b) + (a - b)f'(b) \leq f(a)$, then one can obtain $(f'(a) - f'(b))(a - b) \geq 0$. If $\rho^T \Psi(\mathbf{s}_t) > \rho_{\mathcal{S}^m}^T \Psi(\mathbf{s}_t)$ then $[\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}) > [\delta] \tilde{c}(\mathbf{s}_t, \pi_t, \hat{V}_{\mathcal{S}^m})$ in (3.36) by this convex property of ϵ -loss functions. So the first term of equation (3.36) is positive. Therefore, from the regularized auxiliary function, we have:

$$\begin{aligned}
 & T \times \left\{ 0 - \frac{1}{2} \|\hat{V} - \hat{V}_{\mathcal{S}^m}\|^2 \right\} \geq T \times \left\{ \mathcal{A}(\hat{V}) - \frac{1}{2} \|\hat{V} - \hat{V}_{\mathcal{S}^m}\|^2 \right\} \\
 & \geq [\delta] \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}) \times \rho^T \Psi(\mathbf{s}_m) - [\delta] \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}_{\mathcal{S}^m}) \times \rho_{\mathcal{S}^m}^T \Psi(\mathbf{s}_m).
 \end{aligned}$$

By Condition 3.7 (ii) and $|[\delta] \tilde{c}(\mathbf{s}, \pi, \hat{V})| \leq 1$,

$$\begin{aligned}
 & \frac{T}{2} \|\hat{V} - \hat{V}_{\mathcal{S}^m}\|^2 \leq [\delta] \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}) \times \rho^T \Psi(\mathbf{s}_m) \\
 & - [\delta] \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}_{\mathcal{S}^m}) \times \rho_{\mathcal{S}^m}^T \Psi(\mathbf{s}_m) \leq 4M^*.
 \end{aligned} \tag{3.37}$$

Note that Condition 3.7 implies that the fitting error is Lipschitz continuous, so there is

$$|\tilde{c}(\mathbf{s}, \pi, \hat{V}) - \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}_{\mathcal{S}^m})| \leq C_\pi |\hat{V} - \hat{V}_{\mathcal{S}^m}|.$$

Now we need to derive a bound for $|\hat{V} - \hat{V}_{\mathcal{S}^m}|$ using equation (3.37). By Condition 3.8 and 3.7 (ii), the Cauchy-Schwarz inequality and the kernel property, we have

$$|\hat{V} - \hat{V}_{\mathcal{S}^m}| \leq \|\hat{V} - \hat{V}_{\mathcal{S}^m}\| \times \|k\| \leq \kappa \sqrt{\frac{8M}{T}}. \tag{3.38}$$

Therefore, for any $\mathbf{s}_m \in \mathcal{S}$, $|\tilde{c}(\mathbf{s}, \pi, \hat{V}) - \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}_{\mathcal{S}^m})|$ will be bounded by $C_\pi \kappa \sqrt{\frac{8M}{T}}$. Use this bound in (3.37), we will have

$$\frac{T}{2} \|\hat{V} - \hat{V}_{\mathcal{S}^m}\|^2 \leq C_\pi \|\hat{V} - \hat{V}_{\mathcal{S}^m}\|$$

thus $\|\hat{V} - \hat{V}_{S^m}\| \leq 2C_\pi\kappa/T$. Substituting this into (3.38), we have

$$|\tilde{c}(\mathbf{s}, \pi, \hat{V}) - \tilde{c}(\mathbf{s}_m, \pi_m, \hat{V}_{S^m})| \leq \frac{2C_\pi^2\kappa^2}{T}.$$

The factor decreases to zero when $T \rightarrow \infty$. By the BE Theorem, equation (3.33) turns to

$$\Pr \left\{ \left| \frac{1}{T} \sum_{t=1}^T |\tilde{c}(\mathbf{s}_t, \pi_t, \hat{V})|_\epsilon - \frac{1}{T} \sum_{t=1}^T \tilde{c}(\mathbf{s}_t, \pi_t, V) \right| > 4\epsilon + 2\zeta \right\} \\ \leq 4 \exp \left(-\frac{T}{2} \left(\frac{\epsilon}{M^*} \right)^2 \left(1 + \frac{1}{M^*} (C_\pi\kappa) \right)^{-2} \right),$$

where $\zeta = 2C_\pi^2\kappa^2/T$. \square

Proof of Theorem 3.12

When ϵ goes to zero, the soft-margins of the kernel-based approximation ζ and ζ^* are slack. We set up the new programming problem and then show that the kernel-based policy iteration algorithm coincides with the exact policy iteration in the limit.

Proof. Let ζ and ζ^* be equal to zero in (3.8), the optimization problem is then reduced to

$$\min_{\rho} \frac{1}{2} \|\rho\|^2 \\ \text{s.t. } \pi = \rho^T \Psi(\mathbf{s}) \quad (3.39)$$

where $\Psi(\mathbf{s}) = \Phi(\mathbf{s}) - \beta \sum \Phi(\mathbf{s}') \mathbf{p}(\mathbf{s}' | \mathbf{s}, \mathbf{a})$ and the constraint is the fitting error in (3.10). As in (3.11), taking the partial derivative of the Lagrangian with respect to ρ , we have

$$\partial L / \partial \rho = \rho - \sum_{t=1}^T \alpha_t^0 \Psi(\mathbf{s}_t) = 0,$$

or in the matrix expression $\rho = \boldsymbol{\alpha}^T \Psi$ by stacking α_t and $\Psi(\mathbf{s}_t)$. Similarly, we can set up the dual problem such that

$$\max_{\boldsymbol{\alpha}} -\frac{1}{2} \boldsymbol{\alpha}^T K^* \boldsymbol{\alpha} + \boldsymbol{\alpha}^T \boldsymbol{\pi}, \quad (3.40)$$

with $\boldsymbol{\alpha} \geq 0$ and kernel matrix $K^* = \langle \Psi, \Psi \rangle$. This is a standard linear quadratic objective function. The derivative of (3.40) with respect to $\boldsymbol{\alpha}$ gives linear system $K^T \boldsymbol{\alpha} = \boldsymbol{\pi}$ and $\boldsymbol{\alpha}$ can be solved by a simple inversion.

When α has a unique solution and does not have a zero element, by the dual space property (Slater condition) we know the constraint $\pi = \rho^T \Psi(s)$ is strictly satisfied.

The uniqueness and non-zero α can be proved as follows. Note that K is invertible and positive definite. In the expression of Ψ , $\beta < 1$ and $|\mathbf{p}(\mathbf{s}'|\mathbf{s}, \mathbf{a})| < 1$ imply that $\Psi^T \Psi$ has positive eigenvalues. Hence α is unique. Substitute $K^T \alpha = \pi$ into (3.40), equation (3.40) reduces to $\alpha^T K \alpha / 2$. If we suppose $\alpha_1 = 0$, it implies (3.40) has $\pi_1 \leq 0$. Since $\pi_t = [\sum \alpha_t \Psi(\mathbf{s}_t)] \Psi(\mathbf{s}) > 0$ for any t , contradiction. Hence all elements in α should be non-zero.

By the result of Theorem 1 in Aguirregabiria and Mira (2007), we have

$$V^*(\mathbf{s}) = \sum_{\mathbf{a} \in \mathcal{A}} F^*(\mathbf{a})[\pi^*(\mathbf{a}) + e^*(\mathbf{a})] + \beta \sum_{\mathbf{s}'} V^*(\mathbf{s}') \mathbf{p}^*(\mathbf{s}'|\mathbf{s}).$$

The solution of the exact Bellman equations fits our request. \square