

# Random Coefficients in Static Games of Complete Information

Fabian Dunker\*

University of Goettingen

Stefan Hoderlein<sup>†</sup>

Boston College

Hiroaki Kaido<sup>‡</sup>

Boston University

March 25, 2013

## Abstract

Individual players in a simultaneous equation binary choice model act differently in different environments in ways that are frequently not captured by observables and a simple additive random error. This paper proposes a random coefficient specification to capture this type of heterogeneity in behavior, and discusses nonparametric identification and estimation of the distribution of random coefficients. We establish nonparametric point identification of the joint distribution of all random coefficients, except those on the interaction effects, provided the players behave competitively in all markets. Moreover, we establish set identification of the density of the coefficients on the interaction effects, and provide additional conditions that allow to point identify this density. Since our identification strategy is constructive throughout, it allows to construct sample counterpart estimators. We analyze their asymptotic behavior, and illustrate their finite sample behavior in a numerical study. Finally, we discuss several extensions, like the semiparametric case, or correlated random coefficients.

**Keywords:** Games, Heterogeneity, Nonparametric Identification, Random Coefficients, Inverse Problems.

---

\*Institute for Numerical and Applied Mathematics, University of Goettingen, Lotzestr. 16-18, D-37083 Goettingen, Germany, dunker@math.uni-goettingen.de

<sup>†</sup>Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, email: stefan.hoderlein@yahoo.com.

<sup>‡</sup>Hiroaki Kaido: Boston University, Department of Economics, 270 Bay State Road, Boston, MA 02215, USA, Email: hkaido@bu.edu. Excellent research assistance by Michael Gechter is gratefully acknowledged. We also thank Andres Aradillas-Lopez, Arie Beresteanu, Iván Fernández-Val, Jeremy Fox, Eric Gautier, Yuichi Kitamura, Elie Tamer, Whitney Newey, seminar participants at Boston College, Boston University, Harvard, UCL, University of Pittsburgh, Yale, and conference participants at the Second CIREQ-CEMMAP Workshop on Incomplete Models and the 2013 North American Winter Meeting of the Econometric Society in San Diego.

# 1 Introduction

**Motivation.** Heterogeneity across cross sectional units is ubiquitous in situations of strategic interaction. The behavior of an airline, for instance, may vary dramatically across markets in ways that are only partially explicable by observable factors, like market size or average income. Similarly, there are profound differences in the work decisions of married couples that are not entirely reflected by, say, the number of children, ethnic background, or age. Many of the determinants for these different decisions are unobserved to the applied researcher. Yet, understanding the extent of these differences is crucially important as many important policy questions depend on them.

In this paper, we adopt a random coefficients approach to model such heterogeneity of players across different cross sectional units, like market environments or families. To start out with, we consider the most basic model of strategic interaction in a bivariate, two player one-shot complete (perfect) information game in its reduced form, as a linear dummy endogenous simultaneous equation model. This model has been extensively analyzed with nonrandom coefficients, see Amemiya (1974), Heckman (1978), and Bjorn and Vuong (1985). More recently, Bresnahan and Reiss (1990, 1991) and Tamer (2003) also analyze this model, but elaborate on the derivation from a simple static two player game of complete information.

In this game, there are two players, denoted player 1 and 2, each of which can choose among two actions, denoted 0 and 1. To fix ideas, think of the popular example where the two players are two firms, and the decision is whether or not to enter a market. Alternatively, the two players may be husband and wife in a married couple, and the decision in question would be whether or not to work. International trade may give other examples, where the two players are bilateral trade partners, and the players are a large, respectively a small, partner (e.g., USA - Costa Rica would be one observation, Japan - Bhutan another, etc).

Each of the two players bases her decision in parts on factors that are observed to the econometrician, denoted by  $\tilde{Z}_j$ , and indexed by the number of the player  $j \in \{1, 2\}$ , to indicate that these may not just include observable factors of the market (say, market size), but also variables that might be specific to the player. Moreover, each player takes the actions of the other player into account when making her decision. Importantly, she also bases her decision on variables that are unobserved to the econometrician, and that may impact the way in which she acts on observables.

Throughout this paper, we assume that in every cross section unit, from now on called “market”, each player forms a latent net utility  $Y_j^*$  of choosing action 0 or 1, and that they each pick action 1, provided this latent net utility is above a threshold (which we normalize to be 0). In each market, the players relate the net utility of being in the market in a linear fashion to

the determining variables  $(\tilde{Z}_j, Y_{-j})$ . The coefficients in this relationship are denoted  $\tilde{\beta}_j$  and  $\tilde{\Delta}_j$ , which we consider fixed in any given market. Moreover, we allow for a market and player specific intercept  $\tilde{u}_j$ . The key innovation in this paper is that we allow for all of these variables, including the coefficients  $(\tilde{\beta}_j, \tilde{\Delta}_j)$ , to vary across markets, and that we provide a framework in which we point identify and estimate the distribution of these random parameters<sup>1</sup>. Random coefficients models are commonly used to capture unobserved heterogeneity across cross-sectional units. Recent work on the identification of these models includes Ichimura and Thompson (1998), Berry and Haile (2009), Hoderlein, Klemelä, and Mammen (2010, HKM), Fox, Ryan, Bajari, and Kim (2012), Gautier and Hoderlein (2012), and Gautier and Kitamura (2013, GK). In our setting, the random coefficients allow us to flexibly model unobserved heterogeneity in firms' profit structure and strategic interactions between them across different markets.

Summarizing, the reduced form model is as follows

$$\begin{aligned} Y_1^* &= \tilde{Z}'_1 \tilde{\beta}_1 + Y_2 \tilde{\Delta}_1 + \tilde{u}_1 \\ Y_2^* &= \tilde{Z}'_2 \tilde{\beta}_2 + Y_1 \tilde{\Delta}_2 + \tilde{u}_2, \\ Y_j &= \begin{cases} 1 & \text{if } Y_j^* > 0 \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, 2, \end{aligned} \tag{1.1}$$

where we assume that the factors  $\tilde{Z} = (\tilde{Z}_1, \tilde{Z}_2)$  are fully independent of all unobserved random variables  $(\tilde{\beta}, \tilde{\Delta}, \tilde{u})$ . Because we think of the system (1.1) as a system of simultaneous equations, we refer to  $\tilde{Z}$  as instruments - variables that provide the exogenous variation which is needed to identify the object of interest, the distribution of random parameters.

As is well known in the literature, the properties of the model change fundamentally with the sign of  $\tilde{\Delta}_1$  and  $\tilde{\Delta}_2$ , see, e.g., Bresnahan and Reiss (1991), or Tamer (2003). Indeed, in the entry game,  $\tilde{\Delta}_1, \tilde{\Delta}_2 \leq 0$  is a natural choice arising from economic theory, while other specifications are difficult to reconcile with economic theory. This makes the aim to identify, say, the density of  $\tilde{\Delta}_1$  over the entire space  $\mathbb{R}$  both problematic and economically questionable. Therefore, we focus largely on subcases. In particular, we start out with the case of  $\tilde{\Delta}_1, \tilde{\Delta}_2 \leq 0$ , almost surely, i.e., for every single market. This case is called “strategic substitutes”, and is central to the literature on market entry. In our setup, this means that there is always a negative externality from a player entering the market on the net utility of the other player, but to varying degree across markets.

In this setup, we provide conditions under which the joint density of all random coefficients is point identified. An important point is that we identify the joint distribution of random coefficients, and hence also the marginals, using the entire distribution of the data, and not

---

<sup>1</sup>We sometimes loosely refer to distribution, when we actually mean probability density function.

just those observations for which one player’s entry decision is determined with probability 1. A key insight here is that point identification is satisfied if the sign of a linear combination of the random parameters *for each player* is given, which is for instance satisfied, if the sign of one of the random coefficients is the same across all markets. This generalizes identification in the exogenous binary random coefficients model as in Ichimura and Thompson (1998); the important insight is that no added constraint is required, even though more than just the two marginal distributions for each player are identified. The key identifying restriction is the aforementioned full independence assumption, and it allows to point identify the joint densities of  $(\tilde{\beta}, \tilde{\Delta} + \tilde{u})$  and  $(\tilde{\beta}, \tilde{u})$ , respectively. However, this implies that the joint density of the interaction effects,  $f_{\tilde{\Delta}_1, \tilde{\Delta}_2}$  is only set identified in general, unless one is willing to invoke an additional condition that we provide, and which allows to obtain point identification.

Since our aim is to recover the entire distribution of random coefficients, it is not surprising that we require a large support assumption on the distribution of  $\tilde{Z}$ . This is a common feature in all nonparametric random coefficient models that aim at recovering the density of parameters, see Ichimura and Thompson (1998), HKM (2010), Gautier and Hoderlein (2012), and GK (2013), and should not be confused with “identification at infinity”, as we are using the entire distribution of the data. However, we discuss the case where some of the instruments are discrete in another extension. Another important restriction required for identification in the baseline scenario is that all instruments are player specific. This restriction, too, is relaxed in an extension.

The identification principle put forward is constructive and based on the inversion of operators. Regularized versions of these inverses can be used to construct sample counterpart estimators, and an important part of the analysis in this paper is concerned with their asymptotic behavior. But our intention is not only to contribute to the understanding of these models on an abstract level, but also to provide feasible versions of our approach that are useful in applications. To this end, we discuss semiparametric versions of our approach where some of the coefficients are deterministic in another extension.

After clarifying what can be learned in the case where  $\tilde{\Delta}_1, \tilde{\Delta}_2 < 0$ , we consider various extensions. We first consider the scenario where  $\tilde{\Delta}_1, \tilde{\Delta}_2 > 0$  holds across all markets, a scenario called “strategic complements” and then discuss a more general setup where  $\tilde{\Delta}$  is allowed to have a point mass. We then discuss an extension of our analysis to games with more than two players. Further, we introduce and discuss a semiparametric version of our model, with fixed and random coefficients, which will be more relevant for applications, as it also allows to deal with discrete  $Z$ . We also discuss ways in which the interaction effects may depend on observable variables, as well as the case of correlated random coefficients that cause the covariates to be endogenous. We further explain how to obtain structural objects including average structural

functions and the probability of a specific action profile being a Nash equilibrium. Finally, we discuss the case that some or all of the instruments are the same for both players, i.e.  $Z_1 = Z_2$ .

**Contributions relative to the Literature.** Simultaneous discrete response models have been studied extensively. Much of the literature has focused on identification and estimation of structural parameters that are assumed to be fixed across markets. Ciliberto and Tamer (2009), for example, estimate an entry model of airline markets assuming that the parameters in the airlines' profit functions are either fixed or depend only on observable characteristics of the markets. A novel feature of our model is that the structural parameter may vary across markets following a distribution which is only assumed to satisfy mild assumptions.

A key challenge for the econometric analysis of this class of models is the presence of a region in which each value of payoff relevant variables may correspond to multiple outcomes. Tamer (2003) calls such a region the *region of incompleteness*. Early work in the literature including Amemiya (1974), Heckman (1978), and Bjorn and Vuong (1985) assume that a unique outcome is selected with a fixed probability. More recently, Bresnahan and Reiss (1990, 1991) and Tamer (2003) show that structural parameters can be identified without making such an assumption. The former treats the multiple outcome as a single event and identifies the structural parameters by analyzing the likelihood function (while our approach is non-parametric in nature, we follow this general approach). The latter treats multiple outcome as is, but requires the existence of special covariates that are continuously distributed with full supports, see also Berry and Tamer (2006) for extensions.

As already mentioned, we nonparametrically identify the distribution of the random coefficients without making any assumption on the equilibrium selection mechanism, but utilize the assumption that covariates are continuously distributed with full supports. Other recent work on identification in complete information games includes Bajari, Hong, and Ryan (2010), who establish identification of model primitives including an equilibrium selection mechanism using exclusion restrictions, Beresteanu, Molchanov, and Molinari (2011) and Chesher and Rosen (2012), who apply the theory of random sets to characterize the sharp identification region of structural parameters, and Kline and Tamer (2012), who derive sharp bounds on best response functions without parametric assumptions. Less closely related is the work on identification, estimation and testing in games of incomplete information, see e.g., Aradillas-Lopez (2010), de Paula and Tang (2012), and Lewbel and Tang (2012).

Our model is closely related to index models with random coefficients. In particular, as already discussed, it is widely related to the work on the linear model in Beran, Hall and Feuerverger (1994), HKM (2010), Gautier and Hoderlein (2012), and Masten (2012). Since we are considering a binary dependent variables, our approach is particularly close to the approach of GK (2013), who generalize the nonparametric approach of Ichimura and Thompson (1998).

However, to the best of our knowledge, nonparametric identification of the distribution of random coefficients in a simultaneous system of binary choice models has not been considered. This paper therefore also contributes to the literature of nonparametric identification in simultaneous equation models, see e.g., Matzkin (2008), Berry and Haile (2011), Matzkin (2012), and Masten (2012).

Recent developments on nonparametric identification and estimation in random coefficients models show that recovering the density of random coefficients can be viewed as solving an ill-posed inverse problem, see HKM (2010), GK (2013), and Gautier and Le Penneç (2012). We show that recovering the joint density of random coefficients in a complete information game is also a linear inverse problem. Our identification strategy is more closely related to GK (2013): To recover the joint distribution of random coefficients including the strategic interaction effects, we develop a procedure to invert tensor products of hemispherical transforms. We further provide a conditional deconvolution method to disentangle the distribution of the strategic interaction effects from the distribution of the remaining coefficients.

Empirical studies have shown that the firm heterogeneity plays an important role in entry decisions (See Reiss and Spiller (1989), Berry (1992), Ciliberto and Tamer (2009) among others). This paper considers heterogeneity in the variable cost and the interaction effects. In particular, we allow for unobserved heterogeneity in both of them. There have been recent independent attempts to introduce unobserved heterogeneity into the interaction effects. To the best of our knowledge, Kline (2011) is the first paper that has explicitly allowed for one-dimensional unobservable heterogeneity in the interaction effects. Fox and Lazatti (2012) consider a complete information game with multiple players and study its relation to the demand of bundles, while allowing for unobservable heterogeneity as in Kline (2011). In contrast, we focus on the two player game with possibly multidimensional unobservable heterogeneity.

**Structure of the Paper.** In the second section, we first define the baseline setup considered in this paper, a heterogeneous game of complete information, in the case where the interaction effects  $\tilde{\Delta}_1, \tilde{\Delta}_2$  are negative, as is typical for entry models. We show that the marginal distribution of each player's random coefficients is nonparametrically identified. In the third section, we extend this analysis to recover the joint distribution of random coefficients of both players in the same setup, and establish how to identify the joint density of  $\tilde{\Delta}_1, \tilde{\Delta}_2$ . This section is arguably the main innovation in this paper, and requires new functional analytic tools. In the fourth section, we discuss estimation by sample counterparts. More specifically, we suggest an estimator, and discuss its large sample properties. The fifth section discusses extensions. The sixth section provides a numerical study that illustrates the applicability of the tools introduced in this paper. Finally, an outlook concludes.

## 2 The general structural model and preliminaries

In this section we introduce the basic building blocks of our model. We start by providing formal notation, and clarify and discuss the assumptions. One key assumption is that the interaction effects are negative. Based on the insight of Bresnahan and Reiss (1991), we separate the outcome space into three cases, no entry, duopoly and monopoly. This provides us with two separate conditional probabilities - the third is determined once we know the first two - which we invert to obtain the joint distribution of  $(\tilde{\beta}_1, \tilde{u}_1, \tilde{\beta}_2, \tilde{u}_2)$  and that of  $(\tilde{\beta}_1, \tilde{u}_1 + \tilde{\Delta}_1, \tilde{\beta}_2, \tilde{u}_2 + \tilde{\Delta}_2)$ . From these individual pieces we recover the joint density of  $(\tilde{\Delta}_1, \tilde{\Delta}_2)'$  by deconvolution. We conjecture that it is possible to incorporate Tamer's (2003) insight and use at least some of the information in the monopoly case by distinguishing between the players. However, this would lead to a very different approach that we are pursuing in a separate paper.

### 2.1 Basic definitions and assumptions

We consider a simultaneous game of complete information with two players. Our first assumption describes the implied data generating process (DGP).

**Assumption 2.1.** *Let  $(\Omega, \mathfrak{F}, P)$  be a complete probability space. Let  $k_1, k_2 \in \mathbb{N}$ . For each  $j = 1, 2$ , let  $\tilde{Z}_j : \Omega \rightarrow \mathbb{R}^{k_j}$  be a Borel measurable map. Further, for each  $j = 1, 2$ , let  $\tilde{\beta}_j : \Omega \rightarrow \mathbb{R}^{k_j}$ ,  $\tilde{\Delta}_j : \Omega \rightarrow \mathbb{R}$ , and  $\tilde{u}_j : \Omega \rightarrow \mathbb{R}$  be Borel measurable maps.*

For each  $j = 1, 2$ ,  $\tilde{Z}_j$  is player  $j$ 's observable characteristics. The binary outcome variables  $Y_1, Y_2$  are generated as follows.

$$Y_1^* = \tilde{Z}'_1 \tilde{\beta}_1 + Y_2 \tilde{\Delta}_1 + \tilde{u}_1, \quad (2.1)$$

$$Y_2^* = \tilde{Z}'_2 \tilde{\beta}_2 + Y_1 \tilde{\Delta}_2 + \tilde{u}_2, \quad (2.2)$$

$$Y_j = \begin{cases} 1 & \text{if } Y_j^* > 0 \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, 2. \quad (2.3)$$

For each player, the coefficient  $\tilde{\beta}_j$  captures the marginal impact of player  $j$ 's own covariates  $\tilde{Z}_j$  on the latent variable  $Y_j^*$ , while  $\tilde{u}_j$  captures the effect of other unobservable characteristics. The strategic interaction effect  $\tilde{\Delta}_j$  captures the impact of the other player's decision on the net utility player  $j$  obtains. Assumption 2.1 allows  $(\tilde{\beta}_j, \tilde{\Delta}_j, \tilde{u}_j)'$  to vary across markets. This allows us to flexibly model unobserved heterogeneity in strategic interactions across different markets. In what follows, we let  $Z_j^* \equiv (1, \tilde{Z}'_j)'$ ,  $\beta_j^* \equiv (\tilde{u}_j, \tilde{\beta}'_j)'$ , and  $\theta_j^* \equiv (\tilde{\Delta}_j + \tilde{u}_j, \tilde{\beta}'_j)'$  for  $j = 1, 2$ .

We start with the case in which firms compete across markets, i.e., the utility of each player is adversely affected by the other players choosing action 1:

**Assumption 2.2.** (i)  $\tilde{\Delta}_1 \leq 0, \tilde{\Delta}_2 \leq 0$ ,  $P$ -almost surely; (ii) The distribution of  $\tilde{\Delta} \equiv (\tilde{\Delta}_1, \tilde{\Delta}_2)'$  have the density  $f_{\tilde{\Delta}}$  with respect to Lebesgue measure.

We here assume that  $\tilde{\Delta}$  is continuously distributed for simplicity. It is, however, possible to allow  $\tilde{\Delta}$  to have a point mass at some point. For example,  $\tilde{\Delta}_j$  can be 0 for one of the players with positive probability, in which case the well-known *coherency condition* holds. We will discuss this possibility in Section 6.

Table 1 summarizes the payoffs of the game. In each market, the primitives of the game  $\{(\tilde{Z}_j, \tilde{\beta}_j, \tilde{\Delta}_j, \tilde{u}_j)\}_{j=1,2}$  are assumed to be common knowledge among the players. Our solution concept for this complete information game is the pure strategy Nash equilibrium. Depending on the realizations of  $\{(\tilde{Z}_j, \tilde{\beta}_j, \tilde{\Delta}_j, \tilde{u}_j)\}_{j=1,2}$ , there exist four possible equilibrium outcomes:  $(Y_1, Y_2) = (0, 0)$ , *no entry*;  $(0, 1), (1, 0)$ , *monopoly*; and  $(1, 1)$ , *duopoly*. In case of multiple equilibria, we assume that one of them is selected by some equilibrium selection mechanism, which we do not explicitly specify. Each player's decision  $Y_j$  and instruments  $Z_j$  are assumed to be observable. Our goal is then to recover the distribution of the random coefficients  $(\tilde{\beta}'_j, \tilde{\Delta}_j, \tilde{u}_j)'$  nonparametrically from the observables.

|                      | $Y_2 = 0$ (no entry)                              | $Y_2 = 1$ (entry)  |
|----------------------|---|--|
| $Y_1 = 0$ (no entry) | $(0, 0)$  | $(0, \tilde{Z}'_2 \tilde{\beta}_2 + \tilde{u}_2)$  |
| $Y_1 = 1$ (entry)    | $(\tilde{Z}'_1 \tilde{\beta}_1 + \tilde{u}_1, 0)$ | $(\tilde{Z}'_1 \tilde{\beta}_1 + \tilde{\Delta}_1 + \tilde{u}_1, \tilde{Z}'_2 \tilde{\beta}_2 + \tilde{\Delta}_2 + \tilde{u}_2)$ |

Table 1: The Entry Game Payoff Matrix

Since only the angles of  $\beta_j^*$  and  $Z_j^*, j = 1, 2$  and  $(\theta_j^*$  and  $Z_j^*, j = 1, 2)$  matter for the binary decisions, we define the normalized coefficients and instruments by  $\beta_j \equiv \beta_j^* / \|\beta_j^*\|$ ,  $\theta_j \equiv \theta_j^* / \|\theta_j^*\|$  and  $Z_j \equiv Z_j^* / \|Z_j^*\|$ . Hence, the normalized random coefficients and instruments take values in a unit sphere. This normalization will be also instrumental for us to analyze identification from a linear inverse problem perspective, which we elaborate in the next section.

Below, we introduce additional notation. Let  $\ell \in \mathbb{N}$ . Let  $\mathbb{S}^\ell$  denote the unit sphere in  $\mathbb{R}^{\ell+1}$ . For each  $c \in \mathbb{S}^\ell$ , let  $H_c \equiv \{x \in \mathbb{S}^\ell : c'x \geq 0\}$  be the  $\ell$ -dimensional hemisphere. Let  $\sigma_\ell$  denote the spherical Lebesgue measure on  $\mathbb{S}^\ell$  and let  $L^2(\mathbb{S}^\ell)$  denote the set of square integrable functions on  $\mathbb{S}^\ell$ . The product measure on  $\mathbb{S}^{\ell_1} \times \mathbb{S}^{\ell_2}$  is denoted by  $\sigma \equiv \sigma_{\ell_1} \otimes \sigma_{\ell_2}$ . Let  $L^2(\mathbb{S}^{\ell_1} \times \mathbb{S}^{\ell_2})$  denote square integrable functions on  $\mathbb{S}^{\ell_1} \times \mathbb{S}^{\ell_2}$  with respect to  $\sigma$ .

Throughout, we assume that  $\beta$  and  $\theta$  have well-defined densities with respect to  $\sigma$ .

**Assumption 2.3.** The distributions of  $\beta = (\beta'_1, \beta'_2)'$  and  $\theta = (\theta'_1, \theta'_2)'$  are absolutely continuous with respect to  $\sigma$  with densities  $f_\beta \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  and  $f_\theta \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$ .



We let  $f_{\beta_1}, f_{\beta_2}$  denote the marginal probability density functions of  $\beta_1$  and  $\beta_2$  with respect to  $\sigma_{k_1}$  and  $\sigma_{k_2}$  respectively. The marginal densities  $f_{\theta_1}, f_{\theta_2}$  are similarly defined. One of our key identification assumptions is the following exogeneity of covariates.

**Assumption 2.4.**  $(\beta'_1, \beta'_2, \Delta_1, \Delta_2)'$  is independent of  $Z \equiv (Z'_1, Z'_2)'$ .

This is the central exogeneity assumption we employ. It states that the instruments  $Z$  are fully independent of all unobservables in the system. This is a natural extension of assumptions made in the literature in the fixed coefficients case (e.g., Bresnahan and Reiss (1991), Tamer (2003)). Since we are explicitly considering random coefficients, in our case this is less restrictive than the commonly assumed full independence of a scalar additive unobservable from the instruments, as we allow for this leading type of heteroskedasticity. However, this assumption rules out a heteroskedastic measurement error, and correlation between  $Z$  and the random unobservables. We remark that this correlation could be handled through a control function approach as in Blundell and Powell (2004), but we defer the discussion of this complication to a later section, and focus on the core innovation here.

Let  $r_{(y_1, y_2)}(z) \equiv P((Y_1, Y_2) = (y_1, y_2) | Z = z)$  be the probability of observing  $(y_1, y_2)$  conditional on  $Z = z$ . Under Assumption 2.4, we may write

$$r_{(1,1)}(z) = \mathcal{T}(f_\theta)(z) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} 1\{z'_1 t_1 > 0\} 1\{z'_2 t_2 > 0\} f_\theta(t_1, t_2) d\sigma(t_1, t_2) \quad (2.4)$$

$$r_{(0,0)}(z) = \mathcal{S}(f_\beta)(z) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} 1\{z'_1 b_1 \leq 0\} 1\{z'_2 b_2 \leq 0\} f_\beta(b_1, b_2) d\sigma(b_1, b_2). \quad (2.5)$$

Here,  $\mathcal{T}$  and  $\mathcal{S}$  are integral transforms that map the joint densities  $f_\theta, f_\beta \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  to the conditional entry probabilities  $r_{(1,1)}(z), r_{(0,0)}(z) \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$ , respectively. As we show below, these transforms are closely related to an integral transform called the *hemispherical transform*. This is helpful for analyzing the properties of  $\mathcal{T}$  and  $\mathcal{S}$ .

## 2.2 The hemispherical transform and random coefficients binary choice model

We briefly review the hemispherical transform and its properties relevant for studying identification issues in a random coefficients binary choice model. Details can be found in Groemer (1996), Rubin (1999), and GK (2013). Toward this end, we introduce additional notation. For each real valued function  $\varphi$  on  $\mathbb{S}^\ell$ , let the *odd part* and *even part* of  $\varphi$  be defined by  $\varphi^-(x) \equiv (\varphi(x) - \varphi(-x))/2$  and  $\varphi^+(x) \equiv (\varphi(x) + \varphi(-x))/2$ . Similarly, for each real valued

function  $\varphi$  on  $\mathbb{S}^{\ell_1} \times \mathbb{S}^{\ell_2}$ , let the *component-wise odd part* of  $\varphi$  be defined by<sup>2</sup>:

$$\varphi^{--}(x_1, x_2) \equiv \frac{1}{4} \{ \varphi(x_1, x_2) - \varphi(-x_1, x_2) - \varphi(x_1, -x_2) + \varphi(-x_1, -x_2) \}. \quad (2.6)$$

For each  $\ell \in \mathbb{N}$ , the *hemispherical transform*  $\mathcal{H}_{\mathbb{S}^\ell} : L^2(\mathbb{S}^\ell) \rightarrow L^2(\mathbb{S}^\ell)$  is defined pointwise by

$$\mathcal{H}_{\mathbb{S}^\ell}(s)(z) \equiv \int_{\mathbb{S}^\ell} 1\{z'b > 0\} s(b) d\sigma_\ell(b). \quad (2.7)$$

Let  $\alpha : \Omega \rightarrow \mathbb{R}^\ell$  be random coefficients with a density function  $f_\alpha$  with respect to  $\sigma_\ell$ . Let  $Z : \Omega \rightarrow \mathbb{R}^\ell$  be a vector of instruments and let  $Y$  be generated as

$$Y = 1\{Z'\alpha > 0\}.$$

When  $\alpha$  is independent of  $Z$ , the conditional choice probability is given by:  $P(Y = 1|Z = z) = \mathcal{H}_{\mathbb{S}^\ell} f_\alpha(z)$ . This implies  $f_\alpha$  is identified if  $\mathcal{H}_{\mathbb{S}^\ell}$  is injective. However, the hemispherical transform is known to have a nontrivial null space. Rubin (1999) shows that its null space is

$$\mathcal{N}(\mathcal{H}_{\mathbb{S}^\ell}) = \left\{ f \in L^2(\mathbb{S}^\ell) : f \text{ is an even function, } \int_{\mathbb{S}^\ell} f(a) d\sigma_\ell(a) = 0 \right\}. \quad (2.8)$$

Therefore, restrictions have to be imposed to identify  $f_\alpha$ . GK (2013) show that  $f_\alpha$  is fully determined by its odd-part and therefore can be identified by inverting  $\mathcal{H}_{\mathbb{S}^\ell}$  if the support of  $f_\alpha$  is contained in some hemisphere, i.e., there is a vector  $c \in \mathbb{S}^\ell$  such that  $P(c'\alpha > 0) = 1$ .

A direct application of GK (2013) to our setting would allow to recover the marginal distributions of the random coefficients. We illustrate this for the case of duopoly. Suppose that for each  $j$ , there is a known  $c_j \in \text{supp}(Z_j)$  such that  $P(c_j'\theta_j > 0) = 1$ . Then, we may reduce (2.4) to two separate binary choice equations. To see this, consider conditioning on the event  $\{\omega : Z_1 = z_1, Z_2 = c_2\}$  or  $\{\omega : Z_1 = c_1, Z_2 = z_2\}$  for some  $z_1 \in \mathbb{S}^{k_1}, z_2 \in \mathbb{S}^{k_2}$ . Assumption 2.4 then implies:

$$r_{(1,1)}(z_1, c_2) = \int_{\mathbb{S}^{k_1}} 1\{z_1't_1 > 0\} f_{\theta_1}(t_1) d\sigma_{k_1}(t_1) = \mathcal{H}_{\mathbb{S}^{k_1}}(f_{\theta_1})(z_1), \quad (2.9)$$

$$r_{(1,1)}(c_1, z_2) = \int_{\mathbb{S}^{k_2}} 1\{z_2't_2 > 0\} f_{\theta_2}(t_2) d\sigma_{k_2}(t_2) = \mathcal{H}_{\mathbb{S}^{k_2}}(f_{\theta_2})(z_2). \quad (2.10)$$

We may invert the hemispherical transforms to recover the odd parts of  $f_{\theta_j}, j = 1, 2$ , which determine the marginals  $f_{\theta_j}, j = 1, 2$ . The analysis of no entry is similar. This suggests

---

<sup>2</sup>Similarly, other parts of  $\varphi$ , including the component-wise even part, the part of  $\varphi$  that is odd in the first argument and even in the second argument and vice versa can be defined, but they will not be used in our analysis.

that, employing the results of GK (2013), we may (only) identify the marginals but not the joint distribution  $f_\theta$ . Hence, we will develop an extended framework that allows us to study identification of  $f_\theta$ . We also note that an identification strategy as the one outlined above would use only a subset of the data, and be akin to identification at infinity. In contrast, we will show in the next section how to identify the joint distribution of random coefficients, and hence also the marginals, using the entire distribution of the data, and not just those observations for which one player's entry decision is determined with probability 1.

### 3 Identification of the joint densities in the case of strategic substitutes

In this section we show that the joint density of all random coefficients can be recovered. We present the result for the case of duopoly, from which we can recover  $f_\theta$ . We then employ the case of no entry to recover  $f_\beta$ . From a combination of both objects, the density  $f_\Delta$  can be partially identified generally and point identified under additional assumptions. There is an important technical innovation: The analysis of tensor products of linear operators, a key steps in our identification analysis.

#### 3.1 Duopoly

In this section, we establish that  $f_\theta$  is identified from the conditional probability of duopoly outcomes. Our analysis proceeds in two steps. In a first step, we assume that we know the function  $r_{(1,1)}$  on the whole domain  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ . We show that  $f_\theta$  is identified by (2.4), through a more general form of operator inversion. As we will see shortly,  $r_{(1,1)}$  can only be observed on a part of the domain. Hence, it must be extended to rest of it. How this can be done in a way that is consistent with identification is shown in a second step.

##### 3.1.1 Identification of $f_\theta$ given knowledge of $r_{(1,1)}$ on the whole domain $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$

We start by considering the operator equation (2.4), which can be written as

$$r_{(1,1)} = \mathcal{T}f_\theta.$$

We assume that the function  $r_{(1,1)}$  in (2.4) is known on  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$  and lies in the range of  $\mathcal{T}$ . The first step of the identification analysis is to show that  $\mathcal{T}$  is a *tensor product* of two hemispherical transforms. This allows for a convenient characterization of its null space.

To this end, let  $p \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  be a function which can be written as a product  $p(t_1, t_2) = p_1(t_1)p_2(t_2)$ . Then,  $\mathcal{T}$  becomes a product of hemispherical transforms, i.e.,

$$\mathcal{T}p(z_1, z_2) = \int_{\mathbb{S}^{k_1}} 1\{z_1' t_1 > 0\} p_1(t_1) d\sigma_{k_1}(t_1) \int_{\mathbb{S}^{k_2}} 1\{z_2' t_2 > 0\} p_2(t_2) d\sigma_{k_2}(t_2).$$

As  $L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}) = L^2(\mathbb{S}^{k_1}) \otimes L^2(\mathbb{S}^{k_2})$ , this implies  $\mathcal{T} = \mathcal{H}_{\mathbb{S}^{k_1}} \otimes \mathcal{H}_{\mathbb{S}^{k_2}}$ , where  $\mathcal{H}_{\mathbb{S}^{k_1}}$  and  $\mathcal{H}_{\mathbb{S}^{k_2}}$  are hemispherical transforms as defined in (2.7).<sup>3</sup> The null space of  $\mathcal{T}$  is then given by

$$\begin{aligned} \mathcal{N}(\mathcal{T}) = \{ \tilde{f} \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}) \mid \tilde{f} = \tilde{f}_1 + \tilde{f}_2 \text{ with } \tilde{f}_1(\cdot, t_2) \in \mathcal{N}(\mathcal{H}_{\mathbb{S}^{k_1}}) \text{ for all } t_2 \\ \text{and } \tilde{f}_2(t_1, \cdot) \in \mathcal{N}(\mathcal{H}_{\mathbb{S}^{k_2}}) \text{ for all } t_1 \}. \end{aligned}$$

To see this, let  $\varphi_i$  be the Hilbert basis of spherical harmonics of  $L^2(\mathbb{S}^{k_1})$  and  $\psi_j$  the same for  $L^2(\mathbb{S}^{k_2})$ . For any function  $f \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$ , there exist uniquely determined coefficients  $a_{i,j}$  such that  $f = \sum a_{i,j} \varphi_i \psi_j$ . By Lemma 2.3. in Rubin (1999)  $\varphi_i$  is either in  $\mathcal{N}(\mathcal{H}_{\mathbb{S}^{k_1}})$  or orthogonal to it. The same is true for  $\psi_j$  and  $\mathcal{N}(\mathcal{H}_{\mathbb{S}^{k_2}})$ . Now if  $f \in \mathcal{N}(\mathcal{T})$  and  $a_{i,j} \neq 0$  then  $\mathcal{T}(\varphi_i \psi_j) = 0$ . Hence,  $\mathcal{H}_{\mathbb{S}^{k_1}}(\varphi_i) = 0$  or  $\mathcal{H}_{\mathbb{S}^{k_2}}(\psi_j) = 0$ . For additional information on tensor products of Hilbert spaces, we refer to Reed and Simon (1980).

The spaces  $\mathcal{N}(\mathcal{H}_{\mathbb{S}^{k_1}})$  and  $\mathcal{N}(\mathcal{H}_{\mathbb{S}^{k_2}})$  are determined by (2.8). It implies that every  $f \in \mathcal{N}(\mathcal{T})$  is the sum of two functions  $\tilde{f}_1$  and  $\tilde{f}_2$ , such that  $\tilde{f}_1(\cdot, t_2)$  is even and integrates to 0 for all  $t_1$ , and  $\tilde{f}_2(t_1, \cdot)$  is even and integrates to 0 for all  $t_2$ . Both kinds of functions are orthogonal to a function, which is odd in both variables like  $\varphi^{--}$  in (2.6). Furthermore, we can write the orthogonal complement of the null space as

$$\begin{aligned} \mathcal{N}(\mathcal{T})^\perp = \{ f \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}) \mid f(t_1, t_2) = f^{--}(t_1, t_2) + f_1^-(t_1) + f_2^-(t_2) + c \\ \text{with } f^{--}, f_1^- \text{ and } f_2^- \text{ odd in } t_1 \text{ and } t_2 \}, \end{aligned} \quad (3.1)$$

where  $c$  is a constant, and  $f^{--}$ ,  $f_1^-$ , and  $f_2^-$  satisfy the following equations

$$f_1^-(t_1) = -f_1^-(-t_1) \quad (3.2)$$

$$f_2^-(t_2) = -f_2^-(-t_2) \quad (3.3)$$

$$f^{--}(t_1, t_2) = -f^{--}(-t_1, t_2) = -f^{--}(t_1, -t_2) = f^{--}(-t_1, -t_2). \quad (3.4)$$

It is easy to verify that  $f^{--}$ ,  $f_1^-$ ,  $f_2^-$ , and a constant function  $c$  are pairwise orthogonal. Hence, the representation of  $f \in \mathcal{N}(\mathcal{T})^\perp$  as  $f^{--}(t_1, t_2) + f_1^-(t_1) + f_2^-(t_2) + c$  is unique.

---

<sup>3</sup>We are grateful to Thorsten Hohage who suggested to make use of the tensor product structure of the operator  $\mathcal{T}$ .

Let  $\mathcal{P}_{N(\mathcal{T})^\perp}$  be the orthogonal projection on  $N(\mathcal{T})^\perp$ . Clearly, the operator  $\mathcal{T}$  is injective on  $\mathcal{N}(\mathcal{T})^\perp$ . As a consequence,  $\bar{f}_\theta \equiv \mathcal{P}_{N(\mathcal{T})^\perp}(f_\theta) \in \mathcal{N}(\mathcal{T})^\perp$  is identified by Equation (2.4).

In order to identify  $f_\theta$ , we have to determine  $\tilde{f}_\theta \equiv f_\theta - \mathcal{P}_{N(\mathcal{T})^\perp}(f_\theta) \in \mathcal{N}(\mathcal{T})$  given  $\bar{f}_\theta$ . For this purpose, some additional information about  $f_\theta$  is needed. If we know that  $f_\theta$  is in a subspace  $X_\theta \subset L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  for which the intersection  $\mathcal{P}_{N(\mathcal{T})^\perp}^{-1}(f) \cap X_\theta$  is a singleton for every  $f \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$ ,  $f_\theta$  is identified. Our main example for such a set is given by the following support assumption.

**Assumption 3.1.** *There exists  $(c_1, c_2) \in \text{supp}(Z)$  such that  $\text{supp}(f_\theta) \subseteq H_{c_1} \times H_{c_2}$ .*

This assumption requires that the support of  $\theta_j$  is contained in some known hemisphere  $H_{\theta_j}$ . As Ichimura and Thompson (1998) argue, this assumption is sensible in many applications. For example, if one of the coefficients has a known sign and if  $Z_j$  has a full support, the assumption is satisfied. A slight difference from their assumption (See Theorem 1.(iii) in Ichimura and Thompson (1998)) is that Assumption 3.1 requires the normal vector  $c_j$  of the hemisphere to be in the support of  $Z_j$ . This requirement is not necessary for the following result, but will be used in the next section.

Lemma A.2 in the appendix shows that Assumption 3.1 implies

$$f_\theta(t_1, t_2) = 4f_\theta^{--}(t_1, t_2)1\{f_\theta^{--}(t_1, t_2) > 0, f_{\theta_1}^-(t_1) > 0, f_{\theta_2}^-(t_2) > 0\}, \quad (3.5)$$

for all  $(t_1, t_2) \in \mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ . This allows us to recover  $f_\theta$  from functions in  $\mathcal{N}(\mathcal{T})^\perp$ .

**Remark 3.1.** *Restrictions on the support are not the only possibility to guarantee identification of  $f_\theta$ . Some function classes are as well uniquely determined by its component-wise odd part. The most obvious example are component-wise odd functions. Further examples are functions which are symmetric to some hyperplanes through the origins, e.g., symmetric densities.*

### 3.1.2 Extending $r_{(1,1)}$ to $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$

The argumentation of the last section has still a small gap. The operator equation (2.4) can only identify  $\bar{f}_\theta$ , if the function  $r_{(1,1)}$  is uniquely determined on  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ , but  $r_{(1,1)}$  is not well defined outside the support of  $Z$ . For this reason, we make the following assumption. For each  $j$ , let  $\mathbf{n}_j \equiv (1, 0, \dots, 0)' \in \mathbb{S}^{k_j}$ .

**Assumption 3.2.** *The support of  $Z$  is  $H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}$ .*

This is equivalent to the assumption that the support of  $\tilde{Z}_j$  is  $\mathbb{R}^{k_j}$  for  $j = 1, 2$ . A similar assumption is also invoked in Ichimura and Thompson (1998) and GK (2013) for the simple

binary choice model. This assumption requires that the distribution of  $k_j$  non-constant instruments is supported on  $\mathbb{R}^{k_j}$  and does not degenerate on a set of smaller dimension. This, for example, excludes the case in which  $Z_1$  and  $Z_2$  have a variable in common, or some variables are discrete. We discuss how this assumption can be relaxed in Section 5.

Under Assumptions 3.1 and 3.2, there is a unique extension  $R_{(1,1)}$  of  $r_{(1,1)}$ , which is given by

$$R_{(1,1)}(z_1, z_2) = \begin{cases} r_{(1,1)}(z_1, z_2) & \text{for } (z_1, z_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2} \\ r_{(1,1)}(c_1, z_2) - r_{(1,1)}(-z_1, z_2) & \text{for } (z_1, z_2) \in H_{\mathbf{n}_1}^c \times H_{\mathbf{n}_2} \\ r_{(1,1)}(z_1, c_2) - r_{(1,1)}(z_1, -z_2) & \text{for } (z_1, z_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}^c \\ 1 - (r_{(1,1)}(-z_1, c_2) + r_{(1,1)}(c_1, -z_2)) + r_{(1,1)}(-z_1, -z_2) & \text{for } (z_1, z_2) \in H_{\mathbf{n}_1}^c \times H_{\mathbf{n}_2}^c. \end{cases} \quad (3.6)$$

In addition, we note that  $\mathcal{T}$  maps  $f_\theta^{--}$  to the component-wise odd part of  $R_{(1,1)}$ . That is, it holds that

$$R_{(1,1)}^{--}(z_1, z_2) = \mathcal{T}f_\theta^{--}(z_1, z_2).$$

This suggests that we may apply  $\mathcal{T}^{-1}$  to  $R_{(1,1)}^{--}$  to recover  $f_\theta^{--}$ . Further,  $f_{\theta_1}^-$ , and  $f_{\theta_2}^-$  can be recovered by inverting hemispherical transforms in (2.4)-(2.5) using the results of GK (2013). The joint density  $f_\theta$  can be recovered by (3.5). This closes the gap in the argumentation mentioned at the beginning of this section and gives therefore the following theorem.

**Theorem 3.1.** *In the entry model defined by Equations (2.1) and (2.2), the density  $f_\theta$  is point identified, if Assumptions 2.1-3.2 hold.*

### 3.2 No entry

Identification of  $f_\beta$  in the case of a no entry can be shown by exactly the same argument as above. In this case we have to consider the operator equation  $r_{(0,0)} = \mathcal{S}f_\beta$  defined in (2.5). It follows immediately from the definitions of  $\mathcal{S}$  and  $\mathcal{T}$ , that

$$\mathcal{S} = \mathcal{T} \circ \mathcal{M}_{-1}.$$

Here  $\mathcal{M}_{-1}$  is the operator which multiplies every function pointwise with  $-1$ . As the null space of  $\mathcal{T}$  is invariant under  $\mathcal{M}_{-1}$  we have

$$\mathcal{N}(\mathcal{T}) = \mathcal{N}(\mathcal{S}) \quad \text{and} \quad \mathcal{N}(\mathcal{T})^\perp = \mathcal{N}(\mathcal{S})^\perp.$$

Hence,  $\mathcal{S}$  is injective on the same subspaces as  $\mathcal{T}$  and the operator equation (2.5) can identify the same class of functions as (2.4). The following Assumption is made to identify  $f_\beta$  through functions in  $\mathcal{N}(\mathcal{T})$ .

**Assumption 3.3.** *There exists  $(e_1, e_2) \in \text{supp}(Z)$  such that  $\text{supp}(f_\beta) \subseteq H_{-e_1} \times H_{-e_2}$ .*

By using Assumption 3.3 instead of 3.1, the argumentation in Section 3.1.1 can be applied to  $f_\beta$  as well. When we follow the argumentation of Section 3.1.2 for extending  $r_{(0,0)}$  to  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$  and substitute again Assumption 3.1 by 3.3 we get

$$R_{(0,0)}(z_1, z_2) = \begin{cases} r_{(0,0)}(z_1, z_2) & \text{for } (z_1, z_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2} \\ r_{(0,0)}(e_1, z_2) - r_{(0,0)}(-z_1, z_2) & \text{for } (z_1, z_2) \in H_{\mathbf{n}_1}^c \times H_{\mathbf{n}_2} \\ r_{(0,0)}(z_1, e_2) - r_{(0,0)}(z_1, -z_2) & \text{for } (z_1, z_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}^c \\ 1 - (r_{(0,0)}(-z_1, e_2) + r_{(0,0)}(e_1, -z_2)) + r_{(0,0)}(-z_1, -z_2) & \text{for } (z_1, z_2) \in H_{\mathbf{n}_1}^c \times H_{\mathbf{n}_2}^c. \end{cases} \quad (3.7)$$

The rest of the identification arguments is the same as the duopoly case. Hence, we obtain the following result.

**Theorem 3.2.** *In the entry model defined by Equations (2.1) and (2.2), the density  $f_\beta$  is point identified if Assumptions 2.1-2.4, 3.2, and 3.3 hold.*

### 3.3 Recovering the joint density of $\Delta_1, \Delta_2$

We note that the unnormalized coefficients satisfy

$$\theta_j^* \equiv \beta_j^* + \tilde{\Delta}_j \mathbf{n}_j \text{ for } j = 1, 2. \quad (3.8)$$

This relationship suggests that the density of the strategic interaction effects can be partially identified generally through Makarov bounds (see Fan and Park (2010) and Gautier and Hoderlein (2012)) and can be fully recovered from  $f_\theta$  and  $f_\beta$  under an additional independence assumption.

**Assumption 3.4.**  $\tilde{\Delta} \perp \beta^*$

If this assumption holds, the unnormalized coefficient  $\theta^*$  is the convolution of  $\beta^*$  and the vector  $(\tilde{\Delta}_1 \mathbf{n}_1, \tilde{\Delta}_2 \mathbf{n}_2)'$ . In the following, we let  $\Delta \equiv (\tilde{\Delta}_1 / \|\theta_1^*\|, \tilde{\Delta}_2 / \|\theta_2^*\|)'$  denote the normalized interaction effects and let  $f_\Delta$  denote its density. We note that the scale of the interaction effects is not identified because the entry observations are only informative about the normalized coefficients. Assumption 3.4 gives an integral equation that ties the three densities  $(f_\beta, f_\theta, f_\Delta)$ .

We may then use a deconvolution technique to disentangle the distribution of the interaction effects from  $f_\theta$  and  $f_\beta$ .<sup>4</sup>

The following theorem characterizes the integral equation and gives a sufficient condition for point identification of  $f_\Delta$ .

**Theorem 3.3.** *Suppose the conditions of Theorems 3.1 and 3.2 hold. Suppose further that Assumption 3.4 holds. Then,  $f_\beta, f_\theta$ , and  $f_\Delta$  satisfy  $f_\theta = \mathcal{K}f_\Delta$ , where  $\mathcal{K} : L^2([-1, 0]^2) \rightarrow L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  is an operator defined by:*

$$\mathcal{K}h(t_1, t_2) = \int_{(-1, 0)^2} K(t_1 - w_1 \mathbf{n}_1, t_2 - w_2 \mathbf{n}_2) h(w_1, w_2) dw_1 dw_2 \quad (3.9)$$

where

$$K(u_1, u_2) = f_\beta \left( \frac{u_1}{\|u_1\|}, \frac{u_2}{\|u_2\|} \right) \|u_1\|^{-k_1-1} \|u_2\|^{-k_2-1}. \quad (3.10)$$

Moreover, if  $\Psi_K(s_1, s_2) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} K(u_1, u_2) e^{is'_1 u_1 + is'_2 u_2} d\sigma(u_1, u_2) \neq 0$  almost everywhere in  $\mathbb{R}^{k_1+1} \times \mathbb{R}^{k_2+1}$ ,  $f_\Delta$  is identified.

The regularity condition imposed on  $K$  is an analog of the condition in Devroye (1989) and Carrasco and Florens (2010).

**Remark 3.2.** *To construct a convenient estimator, we have assumed full independence of  $\Delta$  from other coefficients  $\beta$ , but this is stronger than necessary for identification. In fact, it suffices to have  $(\tilde{u}_1, \tilde{u}_2) \perp \tilde{\Delta}$ , but an estimator based on this weaker condition requires marginalization of  $f_\beta$  and  $f_\theta$  to obtain the distributions of  $(\tilde{u}_1/\|\beta_1^*\|, \tilde{u}_2/\|\beta_2^*\|)$  and  $((\tilde{u}_1 + \tilde{\Delta}_1)/\|\theta_1^*\|, (\tilde{u}_2 + \tilde{\Delta}_2)/\|\theta_2^*\|)$  which can be done numerically in practice (See GK, 2013). The estimator based on the full independence condition does not require this extra step.*

## 4 Estimation

This section establishes that the identification principle put forward in this paper can be used directly to construct a sample counterparts estimator. We specify assumptions to construct such an estimator and analyze its large sample behavior.

---

<sup>4</sup>Deconvolution problems are common in both statistics and econometrics; see Carroll and Hall (1988), Devroye (1989), Hu and Ridder (2007), and Carrasco and Florens (2010) among others.



## 4.1 Overview

Throughout, we let  $f_Z, f_{Z_1}, f_{Z_2}, f_{Z_1|Z_2}, f_{Z_2|Z_1}$  denote the joint, marginal, and conditional densities of  $Z_1$  and  $Z_2$ . We construct estimators of  $f_\theta$  and  $f_\beta$  using developments in GK (2013). Although we do not pursue here, construction of an alternative estimator may also be possible. For instance, Gautier and Le Pennec (2011) develop an adaptive estimator for the density of random coefficients in binary choice models using the recent theory of needlets.

Below, we take  $f_\theta$  as an example. First, we rewrite  $R_{(1,1)}^{--}$  as

$$R_{(1,1)}^{--}(z_1, z_2) = \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} \left\{ E \left[ \frac{(4W+1)}{f_Z(Z_1, Z_2)} q_{2p_1+1, 2p_2+1, k_1, k_2}(z_1, z_2, Z_1, Z_2) \right] \right. \\ \left. - E \left[ \frac{q_{2p_1+1, k_1}(z_1, Z_1)}{f_{Z_1}(Z_1)} \right] E \left[ \frac{2W q_{2p_2+1, k_2}(z_2, Z_2)}{f_{Z_2|Z_1}(Z_2|c_1)} \middle| Z_1 = c_1 \right] \right. \\ \left. - E \left[ \frac{q_{2p_2+1, k_2}(z_2, Z_2)}{f_{Z_2}(Z_2)} \right] E \left[ \frac{2W q_{2p_1+1, k_1}(z_1, Z_1)}{f_{Z_1|Z_2}(Z_1|c_2)} \middle| Z_2 = c_2 \right] \right\}, \quad (4.1)$$

where  $W = Y_1 Y_2$ , and  $q_{n_1, n_2, k_1, k_2}, q_{n_1, k_1}$ , and  $q_{n_2, k_2}$  are all known functions that will be defined shortly. We then construct a sample counterpart estimator  $\hat{R}_{(1,1)}^{--}$  by replacing expectations with sample averages and unknown densities with their nonparametric estimators.

In the second step, we invert the operator  $\mathcal{T}$  to obtain  $\hat{f}_\theta^{--} = \mathcal{T}^{-1} \hat{R}_{(1,1)}^{--}$ . We also obtain estimators  $\hat{f}_{\theta_1}, \hat{f}_{\theta_2}$  of marginal densities, using GK (2013). In the final step, we estimate  $f_\theta$  by  $\hat{f}_\theta \equiv 4\hat{f}_\theta^{--}(t_1, t_2) 1\{\hat{f}_\theta^{--}(t_1, t_2) > 0, \hat{f}_{\theta_1}^{--}(t_1) > 0, \hat{f}_{\theta_2}^{--}(t_2) > 0\}$ . An estimator for  $f_\beta$  can be constructed in the same way. Based on the estimators of  $f_\theta$  and  $f_\beta$ , we take another deconvolution step to estimate  $f_\Delta$ .

## 4.2 Condensed harmonic expansion

As a main building block, we use the condensed harmonic expansion in  $L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  to derive (4.1). The motivation for using this expansion is as follows. First, any function  $f \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  can be represented as the sum of its projections to orthogonal subspaces  $H^{n_1, k_1+1} \otimes H^{n_2, k_2+1}$ , where  $H^{n, d}$  is the space of functions, called spherical harmonics of degree  $n$  and dimension  $d$ .<sup>5</sup>  $\mathcal{T}^{-1}$  applied to any function in  $H^{n_1, k_1+1} \otimes H^{n_2, k_2+1}$  is then a simple multiplication by a known constant. This allows us to reduce the computational cost of our estimator.

Formally, the condensed harmonic expansion of  $f \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  is defined by

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} Q_{n_1, n_2, k_1, k_2} f.$$

<sup>5</sup>Details on the spherical harmonics and related objects are provided in Appendix B. See also GK.

Here, the map  $Q_{n_1, n_2, k_1, k_2}$  defined by

$$(Q_{n_1, n_2, k_1, k_2} f)(z_1, z_2) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) f(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2), \quad (4.2)$$

projects  $f \in: L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  to the subspace  $H^{n_1, k_1+1} \otimes H^{n_2, k_2+1}$ . The kernel  $q_{n_1, n_2, k_1, k_2}$  of this map can be written as  $q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) = q_{n_1, k_1}(z_1, \tilde{z}_1) \times q_{n_2, k_2}(z_2, \tilde{z}_2)$  with:  $q_{n_j, k_j}(z_j, \tilde{z}_j) = h(n_j, k_j + 1) L_{n_j}^{k_j+1}(z_j' \tilde{z}_j) / |\mathbb{S}^{k_j}|$ , where  $|\mathbb{S}^{k_j}|$  is the surface area of the sphere, and the constant  $h(n, d)$  and the polynomial  $L_n^d$  are defined in Appendix B.

### 4.3 A sample counterpart estimator

The following theorem shows  $R_{(1,1)}^{--}$  has a representation that suggests a simple sample counterpart estimator.

**Theorem 4.1.** *Suppose Assumptions 2.1, 3.1-3.2 hold. Then, Eq. (4.1) holds.*

Theorem (4.1) suggests estimating  $R_{(1,1)}^{--}$  by

$$\begin{aligned} \hat{R}_{(1,1)}^{--}(z_1, z_2) \equiv & \frac{1}{N} \sum_{i=1}^N \frac{4W_i + 1}{\hat{f}_Z(Z_{1i}, Z_{2i})} \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} q_{2p_1+1, k_1}(z_1, Z_{1i}) q_{2p_2+1, k_2}(z_2, Z_{2i}) \\ & - \frac{1}{N} \sum_{i=1}^N \frac{\sum_{p_1=0}^{\infty} q_{2p_1+1, k_1}(z_1, Z_{1i})}{\hat{f}_{Z_1}(Z_{1i})} \hat{E}_N \left[ \frac{2W_i \sum_{p_2=0}^{\infty} q_{2p_2+1, k_2}(z_2, Z_{2i})}{\hat{f}_{Z_2|Z_1}(Z_{2i}|c_1)} \Big| Z_1 = c_1 \right] \\ & - \frac{1}{N} \sum_{i=1}^N \frac{\sum_{p_2=0}^{\infty} q_{2p_2+1, k_2}(z_2, Z_{2i})}{\hat{f}_{Z_2}(Z_{2i})} \hat{E}_N \left[ \frac{2W_i \sum_{p_1=0}^{\infty} q_{2p_1+1, k_1}(z_1, Z_{1i})}{\hat{f}_{Z_1|Z_2}(Z_{1i}|c_2)} \Big| Z_2 = c_2 \right], \end{aligned} \quad (4.3)$$

where  $\hat{f}_Z, \hat{f}_{Z_j}, \hat{f}_{Z_j|Z_{-j}}$  and  $\hat{E}_N$  are suitable estimators of their population counterparts.

We note here that recovering  $f_{\theta}^{--}$  from  $R_{(1,1)}^{--}$  is an ill-posed inverse problem. To see this, we note that  $q_{n_1, n_2, k_1, k_2}(\cdot, \cdot, \tilde{z}_1, \tilde{z}_2)$  belongs to  $H^{n_1, k_1+1} \otimes H^{n_2, k_2+1}$ , which in turn implies by Proposition 2.4 in GK (2013),

$$\mathcal{T}^{-1} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) = \lambda(n_1, k_1 + 1)^{-1} \lambda(n_2, k_2 + 1)^{-1} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2),$$

where  $\lambda(n, d)$  is an eigenvalue of  $\mathcal{H}_{\mathbb{S}^{d+1}}$ , which tends to 0 as  $n$  increases. Hence, as is standard in the literature, we regularize the inverse. Specifically, for each  $j = 1, 2$ , let  $T_j \in \mathbb{N}$  and  $K_{T_j}(z_j, \tilde{z}_j)$  be the *smoothed projection kernel* defined by

$$K_{T_j}(z_j, \tilde{z}_j) \equiv \sum_{n_j=0}^{T_j} \chi_j(n_j, T_j) q_{n_j, k_j}(z_j, \tilde{z}_j), \quad (4.4)$$

where  $\chi_j(n_j, T_j)$  tends to 0 as  $n_j$  increases. Similarly, let  $K_{T_j}^-$  be the odd part of  $K_{T_j}$  defined by

$$K_{T_j}^-(z_j, \tilde{z}_j) \equiv \sum_{p_j=0}^{T_j-1} \chi_j(2p_j + 1, 2T_j) q_{2p_j+1, k_j}(z_j, \tilde{z}_j).$$

Truncating the sum and introducing the coefficients  $\chi_j$  regularize the inverse when  $\mathcal{T}^{-1}$  is applied to the smoothed projection kernel.

Unknown densities are also estimated using the smoothed projection kernel. Specifically, the joint and marginal distributions are estimated by

$$\hat{f}_Z(z_1, z_2) \equiv \frac{1}{N} \sum_{i=1}^N K_{T_1}(z_1, Z_{1i}) K_{T_2}(z_2, Z_{2i}), \quad \text{and} \quad \hat{f}_{Z_j}(z_j) \equiv \frac{1}{N} \sum_{i=1}^N K_{T_j}(z_j, Z_{ji}) \quad \text{for } j = 1, 2. \quad (4.5)$$

The estimator of  $f_{Z_1|Z_2}$  is then given by  $\hat{f}_{Z_1|Z_2}(z_1|z_2) \equiv \hat{f}_Z(z_1, z_2)/\hat{f}_{Z_2}(z_2)$  and similarly for  $f_{Z_2|Z_1}$ . Since the unknown densities are in the denominators in Eq. (4.1), we use trimmed estimators to handle the random denominator problem. For this, let  $a_N = (\ln N)^{-r}$  for some positive  $r > 0$  and define  $\hat{f}_Z^a \equiv \hat{f}_Z \vee a_N$ ,  $\hat{f}_{Z_j}^a \equiv \hat{f}_{Z_j} \vee a_N$ , and  $\hat{f}_{Z_j|Z_{-j}}^a \equiv \hat{f}_{Z_j|Z_{-j}} \vee a_N$ . We further define the estimator of the conditional mean  $E[V|Z_j]$  of a random variable  $V$  by

$$\hat{E}_N[V|Z_j = c] \equiv \frac{1}{N} \sum_{i=1}^N \frac{V_i K_{T_j}(c, Z_{ji})}{\hat{f}_{Z_j}^a(Z_{ji})}. \quad (4.6)$$

These estimators can be viewed as variants of the projection estimator studied in Hendriks (1990).

Summarizing, our estimator of  $f_{\theta}^{-}$  is defined by

$$\begin{aligned} \hat{f}_{\theta}^{-}(t_1, t_2) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{4W_i + 1}{\hat{f}_Z^a(Z_{1i}, Z_{2i})} \mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(\cdot, Z_{1i}))(t_1) \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2) \\ &\quad - \frac{1}{N} \sum_{i=1}^N \frac{\mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(\cdot, Z_{1i}))(t_1)}{\hat{f}_{Z_1}^a(Z_{1i})} \frac{1}{N} \sum_{i=1}^N \frac{2W_i \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2) K_{T_1}(c_1, Z_{1i})}{\hat{f}_{Z_2}^a(Z_{2i}|c_1) \hat{f}_{Z_1}^a(Z_{1i})} \\ &\quad - \frac{1}{N} \sum_{i=1}^N \frac{\mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2)}{\hat{f}_{Z_2}^a(Z_{2i})} \frac{1}{N} \sum_{i=1}^N \frac{2W_i \mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(\cdot, Z_{1i}))(t_1) K_{T_2}(c_2, Z_{2i})}{\hat{f}_{Z_1}^a(Z_{1i}|c_2) \hat{f}_{Z_2}^a(Z_{2i})}. \end{aligned} \quad (4.7)$$

Based on (3.5), our estimator of  $f_{\theta}$  is now defined pointwise by

$$\hat{f}_{\theta}(t_1, t_2) \equiv 4\hat{f}_{\theta}^{-}(t_1, t_2) \mathbf{1}\{\hat{f}_{\theta}^{-}(t_1, t_2) > 0, \hat{f}_{\theta_1}^{-}(t_1) > 0, \hat{f}_{\theta_2}^{-}(t_2) > 0\}. \quad (4.8)$$

An estimator  $\hat{f}_{\beta}$  of  $f_{\beta}$  can be constructed in the same manner.

## 4.4 Asymptotic properties of $\hat{f}_\theta$

In order to investigate asymptotic properties of our estimator, we need additional regularity conditions. First, we assume that the densities of interest belong to a suitable smoothness class. For this, let  $s \geq 0$ . For each  $f \in L^2(\mathbb{S}^\ell)$ , define the Sobolev norm by

$$\|f\|_{\mathbb{W}_2^s}^2 \equiv \sum_{n=0}^{\infty} (1 + \zeta_{n,\ell})^s \|Q_{n,\ell} f\|_{L^2}^2,$$

where  $\zeta_{n,\ell} \equiv n(n + \ell - 2)$ . We define the Sobolev space by  $\mathbb{W}_2^s(\mathbb{S}^\ell) \equiv \{f : \|f\|_{\mathbb{W}_2^s} < \infty\}$ .

**Assumption 4.1.** *There exist  $s_1, s_2 \geq 0$  such that  $f_\theta^{--} \in \mathbb{W}_2^{s_1}(\mathbb{S}^{k_1}) \otimes \mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})$ .*

Our assumptions on the smoothed projection kernel and the densities of the covariates are analogous to those made in GK (2013) and collected in the appendix (Assumptions C.1 and C.2). For each  $j = 1, 2$ , let  $\rho_j \equiv (2k_j + 1)/s_j$ . The following theorem establishes the convergence rate of our density estimator in the  $L^2$ -norm.

**Theorem 4.2.** *Suppose the conditions of Theorem 3.1 hold. Suppose further that Assumptions C.1 and C.2 hold. If  $T_j$  satisfies*

$$T_j \asymp \left( \frac{N}{(\ln N)^{2r}} \right)^{\frac{1}{s_j(\rho_1 + \rho_2 + 2)}}, \quad j = 1, 2, \quad (4.9)$$

then,

$$\|\hat{f}_\theta - f_\theta\|_{L^2} = O_p \left( \left( \frac{N}{(\ln N)^{2r}} \right)^{\frac{-1}{\rho_1 + \rho_2 + 2}} \right). \quad (4.10)$$

## 4.5 Estimation of $f_\Delta$

The joint density of the strategic interaction terms  $f_\Delta$  is of particular interest. Theorem 3.3 shows that this density is related to  $f_\beta$  and  $f_\theta$  by the following formula

$$f_\theta \left( \frac{(t_1, b_1)}{\|(t_1, b_1)\|}, \frac{(t_2, b_2)}{\|(t_2, b_2)\|} \right) = \int_{-1}^0 \int_{-1}^0 f_\beta \left( \frac{(t_1 - r_1, b_1)}{\|(t_1 - r_1, b_1)\|}, \frac{(t_2 - r_2, b_2)}{\|(t_2 - r_2, b_2)\|} \right) \|(t_1 - r_1, b_1)\|^{-k_1 - 1} \|(t_2 - r_2, b_2)\|^{-k_2 - 1} f_\Delta(r_1, r_2) dr_1 dr_2. \quad (4.11)$$

Here  $t_j, r_j \in \mathbb{R}$  and  $(t_j, b_j) \in \mathbb{R}^{k_j + 1}$  is a value of  $\beta_j^*$ . Note that this relation holds for every fixed value  $b = (b_1, b_2)$  of  $\tilde{\beta}$ . Hence, we can integrate the last equation over  $b$ . This will reduce the curse of dimensionality for the estimator we want to construct. To state the integrated form of

the last equation we define the functions

$$\begin{aligned} K(t_1, t_2) &\equiv \int_{\mathbb{R}^{k_1}} \int_{\mathbb{R}^{k_2}} f_\beta \left( \frac{(t_1, b_1)}{\|(t_1, b_1)\|}, \frac{(t_2, b_2)}{\|(t_2, b_2)\|} \right) \|(t_1, b_1)\|^{-k_1-1} \|(t_2, b_2)\|^{-k_2-1} db_1 db_2 \\ g_\theta(t_1, t_2) &\equiv \int_{\mathbb{R}^{k_1}} \int_{\mathbb{R}^{k_2}} f_\theta \left( \frac{(t_1, b_1)}{\|(t_1, b_1)\|}, \frac{(t_2, b_2)}{\|(t_2, b_2)\|} \right) db_1 db_2. \end{aligned} \quad (4.12)$$

This enables us to derive from (4.11) the following integral equation which is the tensor product of two convolutions

$$\begin{aligned} g_\theta(t_1, t_2) &= \int_{-1}^0 \int_{-1}^0 K(t_1 - r_1, t_2 - r_2) f_\Delta(r_1, r_2) dr_1 dr_2 \\ &= (K(* \otimes *) f_\Delta)(t_1, t_2). \end{aligned} \quad (4.13)$$

Thereby the estimation of  $f_\Delta$  requires the solution of this special deconvolution problem. The eigenvectors of a usual convolution operator are the Fourier basis functions. Hence, the eigenvectors of the tensor product of two convolution operators is the tensor product basis of two Fourier bases. For the domain  $[-1, 1]^2$  it is

$$\varphi_{m_1, m_2}(t_1, t_2) = \frac{1}{2} \exp(\pi i(m_1 t_1 + m_2 t_2)) \quad m_1, m_2 \in \mathbb{Z}.$$

Hence,

$$f_\Delta(t_1, t_2) = \sum_{(m_1, m_2) \in \mathbb{Z}^2} \frac{\langle g_\theta, \varphi_{m_1, m_2} \rangle_{L^2}}{\langle K, \varphi_{m_1, m_2} \rangle_{L^2}} \varphi_{m_1, m_2}(t_1, t_2).$$

Our estimator for  $f_\Delta$  uses the estimators  $\hat{f}_\beta$  and  $\hat{f}_\theta$  introduced in the last section and is defined by

$$\hat{K}(t_1, t_2) \equiv \int_{\mathbb{R}^{k_1}} \int_{\mathbb{R}^{k_2}} \hat{f}_\beta \left( \frac{(t_1, b_1)}{\|(t_1, b_1)\|}, \frac{(t_2, b_2)}{\|(t_2, b_2)\|} \right) \|(t_1, b_1)\|^{-k_1-1} \|(t_2, b_2)\|^{-k_2-1} db_1 db_2 \quad (4.14)$$

$$\hat{g}_\theta(t_1, t_2) \equiv \int_{\mathbb{R}^{k_1}} \int_{\mathbb{R}^{k_2}} \hat{f}_\theta \left( \frac{(t_1, b_1)}{\|(t_1, b_1)\|}, \frac{(t_2, b_2)}{\|(t_2, b_2)\|} \right) db_1 db_2 \quad (4.15)$$

$$\hat{f}_\Delta(t_1, t_2) \equiv \sum_{(m_1, m_2) \in \mathbb{Z}^2} w_{m_1, m_2} \frac{\langle \hat{g}_\theta, \varphi_{m_1, m_2} \rangle_{L^2}}{\langle \hat{K}, \varphi_{m_1, m_2} \rangle_{L^2}} \varphi_{m_1, m_2}(t_1, t_2). \quad (4.16)$$

Here  $w_{m_1, m_2} \in [0, 1]$  are weights that are smoothing the solution to overcome the ill-posedness of the deconvolution. The weights can be chosen in accordance to some smoothed projection kernel, in accordance to Tikhonov regularization with regularization parameter  $\alpha_n$  depending

on the sample size  $n$

$$w_{m_1, m_2} = \left( 1 + \alpha_n \frac{\langle \hat{g}_\theta, \varphi_{m_1, m_2} \rangle_{L^2}}{\langle \hat{K}, \varphi_{m_1, m_2} \rangle_{L^2}} \right)^{-1},$$

or adaptively by asymptotically minimizing the mean integrated square error

$$w_{m_1, m_2} = \max \left\{ 0, 1 - \frac{\langle \hat{K}, \varphi_{m_1, m_2} \rangle_{L^2}^2}{n \langle \hat{g}_\theta, \varphi_{m_1, m_2} \rangle_{L^2}^2} \right\}.$$

In addition  $w_{m_1, m_2}$  is usually set to 0 if  $|m_1|$  or  $|m_2|$  exceed some thresholds.

## 5 Extensions

In this section, we discuss various extensions of our identification results.

### 5.1 The case of strategic complements

We now investigate identification when the sign of the interaction effects is known to be positive. In this case, each of the two the monopoly outcomes  $(0, 1)$  (or  $(1, 0)$ ) realizes as a unique pure strategy Nash equilibrium. Under Assumptions 2.1, 2.3-2.4, the conditional entry probabilities are given by:

$$r_{(0,1)}(z) = \mathcal{U}_1(f_{\beta_1, \theta_2})(z) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} 1\{z'_1 b_1 \leq 0\} 1\{z'_2 t_2 > 0\} f_{\beta_1, \theta_2}(b_1, t_2) d\sigma(b_1, t_2) \quad (5.1)$$

$$r_{(1,0)}(z) = \mathcal{U}_2(f_{\theta_1, \beta_2})(z) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} 1\{z'_1 t_1 > 0\} 1\{z'_2 b_2 \leq 0\} f_{\theta_1, \beta_2}(t_1, b_2) d\sigma(t_1, b_2). \quad (5.2)$$

Letting  $\mathcal{M}_{-1, k_j}$  a map that multiplies a function on  $\mathbb{S}^{k_j}$  by -1 pointwise, the maps  $\mathcal{U}_1, \mathcal{U}_2$  can be equivalently written as:

$$\mathcal{U}_1 = (\mathcal{H}_{\mathbb{S}^{k_1}} \circ \mathcal{M}_{-1, k_1}) \otimes \mathcal{H}_{\mathbb{S}^{k_2}} \quad \text{and} \quad \mathcal{U}_2 = \mathcal{H}_{\mathbb{S}^{k_1}} \otimes (\mathcal{H}_{\mathbb{S}^{k_2}} \circ \mathcal{M}_{-1, k_2}).$$

Since the maps  $\mathcal{M}_{-1, k_j}$ ,  $j = 1, 2$  do not affect the null space, this implies that  $\mathcal{N}(\mathcal{U}_1) = \mathcal{N}(\mathcal{U}_2) = \mathcal{N}(\mathcal{T})$ . Therefore, our previous identification argument applies. Under Assumptions 3.1-3.3, we may uniquely extend the conditional entry probabilities to define  $R_{(0,1)}$  and  $R_{(1,0)}$  on  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ . Further, Assumptions 3.1 and 3.3 imply that there exist  $(e_1, c_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}$  such that  $\text{supp}(f_{\beta_1, \theta_2}) \subseteq H_{-e_1} \times H_{c_2}$ . Similarly, there exist  $(c_1, e_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}$  such that  $\text{supp}(f_{\theta_1, \beta_2}) \subseteq H_{c_1} \times H_{-e_2}$ . These conditions ensure that  $f_{\beta_1, \theta_2}$  and  $f_{\theta_1, \beta_2}$  are determined by their component-wise odd parts. These functions can be recovered by applying the inverse of the operators to

$R_{(0,1)}^-$  and  $R_{(1,0)}^-$ . Therefore, in the case of strategic complements, the densities  $f_{\beta_1, \theta_2}$  and  $f_{\theta_1, \beta_2}$  are point identified under Assumptions 2.1, 2.3-3.3.

Identification of the marginal densities  $f_{\Delta_1}, f_{\Delta_2}$  of the interaction effects are possible. For each  $j$ , the three marginal densities  $(f_{\theta_j}, f_{\beta_j}, f_{\Delta_j})$  can be shown to be related through the integral equation:

$$f_{\theta_j}(t_j) = \mathcal{K}_j f_{\Delta_j}(t_j) = \int_{(-1,0)} K_j(t_j - w_j \mathbf{n}_1) f_{\Delta_j}(w_j) d\mu(w_j), \quad (5.3)$$

where  $K_j(u_j) = f_{\beta_j}(\frac{u_j}{\|u_j\|}) \|u_j\|^{-k_j-1}$ . Provided that the inverse Fourier transform of  $K_j$  is non-zero *a.e.*, we may then identify the marginal distributions of the interaction effects by deconvolution. A crucial difference from the competitive case is that we may not identify the joint distribution. This is because the conditional entry probability of  $(0, 1)$  (or  $(1, 0)$ ) is informative about only one of the interaction effects. Still, as we will see in Section 5.8, the marginal density is useful for studying various structural objects including the average effect of the other player's entry. Further, functionals of the joint density can be partially identified. For example, we may obtain bounds on a measure of dependence between  $\Delta_1$  and  $\Delta_2$  using the Fréchet-Hoeffding bounds. Results on these bounds are well known. See, for example, Heckman, Smith, and Clements (1997) and Fan and Zhu (2009).

## 5.2 Interaction effects with point masses

For the identification result and for the estimator we assumed that the  $\tilde{\Delta}$  has a Lebesgue density. However, in some markets, the opponent's action may not affect a player's payoff. In such a case,  $\tilde{\Delta}_j$  is degenerated at 0. Nevertheless, the distribution of  $\tilde{\Delta}$  is identified by our model. Motivated by this example, we generalize our results in a way which allows the probability measure of  $\tilde{\Delta}_j$  to be any Radon measure. As above, we distinguish the cases of strategic substitutes and strategic complements. We present only the case of strategic substitutes. The other case can be studied as in the previous section.

To generalize the identification result, Assumption 2.2 (ii) is replaced by the assumption that  $f_{\tilde{\Delta}} \in \mathcal{D}'(\mathbb{R}^2)$  is a distribution, i.e. a generalized function. Here  $\mathcal{D}'(\mathbb{R}^2)$  is the dual space of all infinitely smooth functions with compact support  $C_c^\infty(\mathbb{R}^2)$ . It contains all Radon measures. This new assumption is not in conflict with Assumption 2.3. If for example  $f_{\tilde{\Delta}}$  has compact support and  $f_{\tilde{u}} \in L^2(\mathbb{R}^2)$ , then  $\tilde{\Delta} + \tilde{u}$  has a  $L^2$  density, since  $f_{\tilde{\Delta}} * f_{\tilde{u}} \in L^2(\mathbb{R}^2)$ . Hence,  $\theta$  can have a  $L^2$  density as well.

Therefore, the identification analysis of  $f_\theta$  and  $f_\beta$  presented above need not to be changed. Only the identification result for  $f_\Delta$  has to be generalized as  $f_\Delta$  is now a distribution with support in  $[-1, 0]$ . Hence,  $f_\Delta$  is a distribution with compact support, i.e.  $f_\Delta \in \mathcal{E}'(\mathbb{R}^2)$ . This

makes the generalization straightforward, because it implies that the convolution of  $f_\Delta$  with any  $L^2$  function is again in  $L^2$  and that the convolution theorem holds. The operator  $\mathcal{K}$  in Theorem 3.3 is a convolution operator with the convolution kernel  $K \in L^2(\mathbb{R}^{k_1+1} \times \mathbb{R}^{k_2+1})$ . So the extension of the operator to  $\mathcal{K} : \mathcal{E}'(\mathbb{R}^2) \rightarrow L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  is well defined. Under Assumption 3.4 the first assertion of Theorem 3.3 that  $f_\theta = \mathcal{K}f_\Delta$  is still valid for  $f_\Delta \in \mathcal{E}'(\mathbb{R}^2)$ . Furthermore, as the convolution theorem can be applied, the second assertion of Theorem 3.3 is true as well. I.e.  $f_\Delta$  is identified in  $\mathcal{E}'(\mathbb{R}^2)$  if the Fourier transform of the convolution kernel is nonzero  $\mathcal{F}(K) \neq 0$  almost everywhere.

For the numerical implementation of the deconvolution our main interest is in distributions, which have a small number of point masses at some points and are continuously distributed elsewhere in  $[-1, 0]$ . So we assume  $f_\Delta$  has the form

$$f_\Delta(w) = g_\Delta(w) + \sum_{m=1}^M d_m \delta_{x_m}(w),$$

with  $g_\Delta \in L^1([-1, 0])$  non negative,  $\delta_{x_m}$  is a Dirac delta at  $x_m \in [-1, 0]$ ,  $d_m \geq 0$ , and  $\int g_\Delta(x)dw + \sum d_m = 1$ . Let us denote by  $S_M$  the set of all these distributions with at most  $M$  point masses.

The class of distribution we consider now does not admit a representation by Fourier series as  $f_\Delta$  in Section 4.5. Therefore, we propose an other estimator for the deconvolution problem which uses Tikhonov regularization to overcome the ill-posedness of the deconvolution.

$$\hat{f}_\Delta := \underset{f \in S_M}{\operatorname{argmin}} (\|f_\theta - \mathcal{K}f\|_{L^2} + \alpha \mathcal{R}(f)) \quad (5.4)$$

Note that the approximate solution  $\hat{f}_\Delta \in S_M$  is by definition a probability distribution. Since  $f_\theta \in L^2$  and  $\mathcal{K}f \in L^2$ , it is quite natural to evaluate the data misfit (the first term on the r.h.s.) by the  $L^2$ -norm. Other convex distance measures like the Kullback-Leibler divergence are possible as well. The regularization functional  $\mathcal{R}$  is supposed to be convex and  $\alpha \geq 0$  is a regularization parameter that has to be chosen carefully. An appropriate choice for the regularization functional is  $\mathcal{R}(f_\Delta) = \|g_\Delta\|_{L^2} + \sum_{m=1}^M d_m$ . Alternatively,  $g_\Delta$  can be regularized by a Sobolev norm or by maximum entropy.

The minimization problem (5.4) is convex and has therefore a unique solution under mild assumptions. This solution can be calculated by convex optimization algorithms like the semi-smooth Newton method or sequentially quadratic programming among others. Convergence rates and parameter choice strategies for  $\alpha$  in algorithms with similar regularization functionals can be found in Eggermont (1993), Burger and Osher (2004), Resmerita (2005), and Grasmair, Haltmeier, Scherzer (2008).



### 5.3 Games with more than 2 players

So far, our analysis has focused on the case with two players. Our identification analysis on  $f_\beta$  and  $f_\theta$  can be extended to the case with  $J$  players where  $J \geq 3$ . In the case of strategic substitutes with more than two players, the no entry outcome  $(0, \dots, 0)$  and “full entry” outcome  $(1, \dots, 1)$  still arise as unique equilibria. These give the following two integral equations that involve  $J$ -fold tensor products of hemispherical transforms.

$$r_{(1, \dots, 1)}(z) = \int_{\mathbb{S}^{k_1} \times \dots \times \mathbb{S}^{k_J}} 1\{z'_1 t_1 > 0\} \cdots 1\{z'_J t_J > 0\} f_\theta(t_1, \dots, t_J) d\sigma(t_1, \dots, t_J) \quad (5.5)$$

$$r_{(0, \dots, 0)}(z) = \int_{\mathbb{S}^{k_1} \times \dots \times \mathbb{S}^{k_J}} 1\{z'_1 b_1 \leq 0\} \cdots 1\{z'_J b_J \leq 0\} f_\beta(b_1, \dots, b_J) d\sigma(b_1, \dots, b_J). \quad (5.6)$$

Inverting them yields identification of random coefficients except the interaction effects provided that we have a sign restriction for each player. With  $J$  players, however, the interaction effects become quite high-dimensional. This raises a challenge for identification. We expect that our identification strategy, which recovers  $f_\Delta$  through deconvolution of  $f_\theta$  and  $f_\beta$  does not extend readily to this general case, however, identification of  $f_\Delta$  may be possible under additional symmetry restrictions e.g. the existence of a potential function, see Fox and Lazzati, (2012) for details.

### 5.4 Semiparametric specification

The full random coefficient specification is appealing but requires strong identifying assumptions. In particular, all instruments need to be continuously distributed with full supports. In this section, we consider a semiparametric specification, which allows us to relax this assumption.

Below, we classify instruments into three categories. For each  $j$ , let  $\tilde{Z}_j^{FD} : \Omega \rightarrow \mathbb{R}^{k_j^{FD}}$  be instruments with potentially limited supports. Here, we allow  $\tilde{Z}_1^{FD}$  and  $\tilde{Z}_2^{FD}$  to be discrete. We also allow them to have variables in common. It is, however, assumed that their coefficients  $\tilde{\beta}_j^{FD}$  are non-random. Similarly, let  $\tilde{Z}_j^{FC} : \Omega \rightarrow \mathbb{R}^{k_j^{FC}}$  be instruments with full supports whose coefficients  $\tilde{\beta}_j^{FC}$  are non-random. Further, let  $\tilde{Z}_j^R : \Omega \rightarrow \mathbb{R}^{k_j^R}$  denote instruments whose coefficients  $\tilde{\beta}_j^R$  are random. We assume that  $(\tilde{Z}_1^R, \tilde{Z}_2^R)$  are continuously distributed with full supports.

Our semiparametric model is then given by

$$\begin{aligned}
Y_1^* &= \tilde{Z}_1^{R'} \tilde{\beta}_1^R + \tilde{Z}_1^{FC'} \tilde{\beta}_1^{FC} + \tilde{Z}_1^{FD'} \tilde{\beta}_1^{FD} + Y_2 \tilde{\Delta}_1 + \tilde{u}_1 \\
Y_2^* &= \tilde{Z}_2^{R'} \tilde{\beta}_2^R + \tilde{Z}_2^{FC'} \tilde{\beta}_2^{FC} + \tilde{Z}_2^{FD'} \tilde{\beta}_2^{FD} + Y_1 \tilde{\Delta}_2 + \tilde{u}_2, \\
Y_j &= \begin{cases} 1 & \text{if } Y_j^* > 0 \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, 2.
\end{aligned} \tag{5.7}$$

Again, we normalize the coefficients and variables. For  $j$  and  $\tilde{z}_j^{FD}$ , let

$$\tilde{\gamma}_j(\tilde{z}_j^{FD}) \equiv \tilde{u}_j + \tilde{z}_j^{FD'} \tilde{\beta}_j^{FD} \tag{5.8}$$

$$\tilde{\delta}_j(\tilde{z}_j^{FD}) \equiv \tilde{u}_j + \tilde{\Delta}_j + \tilde{z}_j^{FD'} \tilde{\beta}_j^{FD}. \tag{5.9}$$

Further, for each  $j$  and  $\tilde{z}_j^{FD} \in \mathbb{R}^{k_j^{FD}}$ , let  $W_j^* \equiv (1, \tilde{Z}_j^{R'}, \tilde{Z}_j^{FC'})'$ ,  $\beta_j^*(\tilde{z}_j^{FD}) \equiv (\tilde{\gamma}_j(\tilde{z}_j^{FD}), \tilde{\beta}_j^{R'}, \tilde{\beta}_j^{FC'})'$ , and  $\theta_j^*(\tilde{z}_j^{FD}) \equiv (\tilde{\delta}_j(\tilde{z}_j^{FD}), \tilde{\beta}_j^{R'}, \tilde{\beta}_j^{FC'})'$ . For each  $j$ , we then use  $W_j$ ,  $\beta_j(\tilde{z}_j^{FD})$ , and  $\theta_j(\tilde{z}_j^{FD})$  to denote their normalized versions.

We make the following assumptions.

**Assumption 5.1.**  $(\beta_1^*(\tilde{Z}_1^{FD}), \beta_2^*(\tilde{Z}_2^{FD}), \tilde{\Delta}_1, \tilde{\Delta}_2) \perp W | \tilde{Z}^{FD}$ .

**Assumption 5.2.** *There exists  $(c_1, c_2) : \text{supp}(f_{\tilde{Z}^{FD}}) \rightarrow \mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$  such that  $(c_1(\tilde{Z}^{FD}), c_2(\tilde{Z}^{FD})) \in \text{supp}(f_{W_1, W_2 | \tilde{Z}^{FD}})$  and  $\text{supp}(f_{\theta(\tilde{Z}^{FD}) | \tilde{Z}^{FD}}) \subseteq H_{c_1(\tilde{Z}^{FD})} \times H_{c_2(\tilde{Z}^{FD})}$  almost surely.*

**Assumption 5.3.** *There exists  $(e_1, e_2) : \text{supp}(f_{\tilde{Z}^{FD}}) \rightarrow \mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$  such that  $(-e_1(\tilde{Z}^{FD}), -e_2(\tilde{Z}^{FD})) \in \text{supp}(f_{W_1, W_2 | \tilde{Z}^{FD}})$  and  $\text{supp}(f_{\beta(\tilde{Z}^{FD}) | \tilde{Z}^{FD}}) \subseteq H_{-e_1(\tilde{Z}^{FD})} \times H_{-e_2(\tilde{Z}^{FD})}$  almost surely.*

**Assumption 5.4.** *The support of  $f_{W_1, W_2 | (\tilde{Z}_1^{FD}, \tilde{Z}_2^{FD})}$  is  $H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}$  almost surely, where  $H_{\mathbf{n}_j} \subset \mathbb{S}^{k_j^R + k_j^{FC}}$  is the hemisphere as in Assumption 3.2.*

The identification strategy is the same as before. Therefore, we just briefly sketch the argument. Let  $f_{\beta(\tilde{Z}^{FD}) | \tilde{Z}^{FD}}$  be the density of  $\beta(\tilde{Z}^{FD})$  conditional on  $\tilde{Z}^{FD}$ . For any  $(w_1, w_2)$ , Assumption 5.1 allows us to write

$$r_{(1,1)}(w_1, w_2) = (\mathcal{T} f_{\theta(\tilde{Z}^{FD}) | \tilde{Z}^{FD}})(w_1, w_2) \tag{5.10}$$

Assumption 5.2 ensures that  $f_{\theta(\tilde{Z}^{FD}) | \tilde{Z}^{FD}}$  is determined by its component-wise odd part, and the odd part of the marginals. Assumption 5.4 ensures an extension of  $r_{(1,1)}$  to  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$  exists. Then, by inverting  $\mathcal{T}$ , we may identify  $f_{\theta(\tilde{Z}^{FD}) | \tilde{Z}^{FD}}$ . A similar argument also applies to identification of  $f_{\beta(\tilde{Z}^{FD}) | \tilde{Z}^{FD}}$ .

For identification of  $f_\Delta$ , we note that the following relationship holds:

$$\theta_j^*(\tilde{Z}_j^{FD}) = \beta_j^*(\tilde{Z}_j^{FD}) + \tilde{\Delta}_j \mathbf{n}_j. \quad (5.11)$$

This implies that  $f_{\theta(\tilde{Z}^{FD})|\tilde{Z}^{FD}}$ ,  $f_{\beta(\tilde{Z}^{FD})|\tilde{Z}^{FD}}$ , and  $f_{\Delta|\tilde{Z}^{FD}}$  satisfy a convolution relationship under the following assumption.

**Assumption 5.5.**  $\tilde{\Delta} \perp \beta^*(\tilde{Z}^{FD})|\tilde{Z}^{FD}$ .

Together with a regularity condition on the Fourier transform of  $f_{\beta(\tilde{Z}^{FD})|\tilde{Z}^{FD}}$ , Assumption 5.5 allows us to recover the conditional distribution  $f_{\Delta|\tilde{Z}^{FD}}$  by deconvolution. Since  $Z^{FD}$  is observable, one may estimate  $f_{\tilde{Z}^{FD}}$ . Then,  $f_\Delta$  can be recovered by integrating  $f_{\Delta|\tilde{Z}^{FD}} \times f_{\tilde{Z}^{FD}}$  over the support of  $\tilde{Z}^{FD}$ .

The knowledge of  $f_{\beta(\tilde{Z}^{FD})|\tilde{Z}^{FD}}$  also allows us to recover the joint distribution of normalized random coefficients:  $(\tilde{\beta}_1^R/\|\beta_1^*(\tilde{z}_1^{FD})\|, \tilde{\beta}_2^R/\|\beta_2^*(\tilde{z}_2^{FD})\|)$  conditional on  $\tilde{Z}^{FD}$ . Marginalizing this density using  $f_{\tilde{Z}^{FD}}$  gives the joint density of the normalized random coefficients.

We also note that the fixed coefficients are identified up to scale. For example,  $f_{\beta_1(\tilde{Z}_1^{FD})|\tilde{Z}_1^{FD}}$  being identified implies that one knows

$$E \left[ \frac{\tilde{\gamma}_1(\tilde{Z}_1^{FD})}{\|\beta_1^*(\tilde{Z}_1^{FD})\|} \middle| \tilde{Z}^{FD} = \tilde{z}^{FD} \right] = E \left[ \frac{\tilde{u}_1}{\|\beta_1^*(\tilde{Z}_1^{FD})\|} \right] + E \left[ \frac{1}{\|\beta_1^*(\tilde{Z}_1^{FD})\|} \right] \tilde{z}^{FD} \tilde{\beta}_1^{FD}.$$

With enough variation of  $\tilde{Z}^{FD}$ , we may identify  $\tilde{\beta}_1^{FD}$  up to scale. Similarly, the knowledge of  $f_{\beta_1(\tilde{Z}_1^{FD})|\tilde{Z}_1^{FD} = \tilde{z}^{FD}}$  also identifies  $\tilde{\beta}^{FC}$  up to scale.

## 5.5 Discrete explanatory variables with random coefficients

The previous approach allows for discrete explanatory variables, but presupposes that the coefficient on these variables is fixed. However, often times discrete explanatory variables are believed to have a heterogeneous impact, e.g., throughout the treatment effects literature. Because of this leading case, we focus in what follows on a binary explanatory variable, wlog the first for the first player, denoted  $\tilde{Z}_{11}$ . This allows us to study identification using developments in Gautier and Hoderlein (2012). Separating  $\tilde{Z}_1 = (\tilde{Z}_{11}, \tilde{Z}'_{-11})'$ , and analogously for the coefficients, we obtain

$$\begin{aligned} Y_1^* &= \tilde{Z}_{11} \tilde{\beta}_{11} + \tilde{Z}'_{-11} \tilde{\beta}_{-11} + Y_2 \tilde{\Delta}_1 + \tilde{u}_1 \\ Y_2^* &= \tilde{Z}'_2 \tilde{\beta}_2 + Y_1 \tilde{\Delta}_2 + \tilde{u}_2, \\ Y_j &= \begin{cases} 1 & \text{if } Y_j^* > 0 \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, 2, \end{aligned} \quad (5.12)$$

Next, if we condition the choice probabilities on the events  $\tilde{Z}_{11} = 1$ , and  $\tilde{Z}_{11} = 0$ , we obtain four conditional choice probabilities (instead of two), that allow us to recover the marginal densities of  $(\tilde{\beta}_{-11}, \tilde{u}_1)$ ,  $(\tilde{\beta}_{-11}, \tilde{\Delta}_1 + \tilde{u}_1)$ ,  $(\tilde{\beta}_{-11}, \tilde{\beta}_{11} + \tilde{u}_1)$ , and  $(\tilde{\beta}_{-11}, \tilde{\Delta}_1 + \tilde{\beta}_{11} + \tilde{u}_1)$ . Much as before with the density of the interaction effects, we can invoke (conditional) independence conditions, to recover the density of  $f_{\tilde{\beta}_{11}}$ . In fact, analogous conditional independence conditions are amply sufficient for identification, as there are several ways to recover  $f_{\tilde{\beta}_{11}}$ . The same is true for Makarov-type bounds that may be obtained, if one is reluctant to invoke these independence assumptions, see, e.g., Gautier and Hoderlein (2012), Section 3.3.

## 5.6 Interaction effects with observable components and multidimensional unobservable heterogeneity

In what follows, we assume that non-constant variables that affect the interaction effects are also included in the instrument  $\tilde{Z}$  and denote them by  $\tilde{X} = (\tilde{X}'_1, \tilde{X}'_2)' \in \mathbb{R}^{l_1} \times \mathbb{R}^{l_2}$ . We also reorder  $\tilde{Z}$  so that for each  $j$ , the first  $l_j + 1$  components of  $\tilde{Z}_j$  are given by  $(1, \tilde{X}_j)$ . The reduced form model then becomes:

$$Y_1^* = \tilde{Z}'_1 \tilde{\beta}_1 + Y_2(\tilde{\Delta}_1 + \tilde{X}'_1 \tilde{\eta}_1) + \tilde{u}_1, \quad (5.13)$$

$$Y_2^* = \tilde{Z}'_2 \tilde{\beta}_2 + Y_1(\tilde{\Delta}_2 + \tilde{X}'_2 \tilde{\eta}_2) + \tilde{u}_2, \quad (5.14)$$

$$Y_j = \begin{cases} 1 & \text{if } Y_j^* > 0 \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, 2, \quad (5.15)$$

where  $\tilde{\eta}_1 : \Omega \rightarrow \mathbb{R}^{l_1}$  and  $\tilde{\eta}_2 : \Omega \rightarrow \mathbb{R}^{l_2}$  are random coefficients. We then let  $\theta_j^* \equiv (\tilde{\Delta}_j + \tilde{u}_j, \tilde{\beta}_1 + \tilde{\eta}_1, \dots, \tilde{\beta}_{l_j} + \tilde{\eta}_{l_j}, \tilde{\beta}_{l_j+1}, \dots, \tilde{\beta}_{k_j})$ . We make the following assumption, which replaces Assumption 2.2.

**Assumption 5.6.** *For each  $\tilde{x}_j \in \text{supp}(\tilde{X}_j)$ ,  $\tilde{\Delta}_j + \tilde{x}'_j \tilde{\eta}_j \leq 0$  with probability 1.*

Under this assumption, we may recover  $f_\beta$  from the conditional probability of the no entry outcomes. Similarly,  $f_\theta$  can be recovered from the probability of the duopoly outcome. Define the scaled coefficients  $\gamma_j \equiv (\tilde{\Delta}_j / \|\theta_j^*\|, \tilde{\eta}_j / \|\theta_j^*\|)'$ .  $f_\gamma$  is partially identified generally and point identified if  $\gamma \perp \tilde{X}$  and  $\gamma \perp \beta^*$  and additional regularity conditions hold. Specifically, under independence,  $f_\beta, f_\theta$ , and  $f_\gamma$  satisfy  $f_\theta = \mathcal{L}f_\gamma$ , where  $\mathcal{L}$  is an integral operator defined by:

$$\mathcal{L}h(t_1, t_2) = \int_{(-1,0) \times (-1,1)^{l_2} \times (-1,0) \times (-1,1)^{l_1}} L(t_1 - v_1 \mathbf{m}_1, t_2 - v_2 \mathbf{m}_2) h(v_1, v_2) d\mu(v_1, v_2),$$

where

$$L(u_1, u_2) = f_\beta \left( \frac{u_1}{\|u_1\|}, \frac{u_2}{\|u_2\|} \right) \|u_1\|^{-k_1-1} \|u_2\|^{-k_2-1}.$$

Therefore, if  $\Psi_L(s_1, s_2) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} L(u_1, u_2) e^{is'_1 u_1 + is'_2 u_2} d\sigma(u_1, u_2) \neq 0, a.e.$ , then  $f_\gamma$  is identified.

## 5.7 Endogenous explanatory variables

To discuss endogenous explanatory variables, it is worthwhile to return to the reduced form of the baseline model, but we let one of the explanatory variables, for simplicity the first denoted  $\tilde{Z}_{11}$ , be correlated with the vector  $\tilde{\beta}_1$ . Separating  $\tilde{Z}_1 = (\tilde{Z}_{11}, \tilde{Z}'_{-11})'$ , and analogously for the coefficients, we obtain

$$\begin{aligned} Y_1^* &= \tilde{Z}_{11} \tilde{\beta}_{11} + \tilde{Z}'_{-11} \tilde{\beta}_{-11} + Y_2 \tilde{\Delta}_1 + \tilde{u}_1 \\ Y_2^* &= \tilde{Z}'_2 \tilde{\beta}_2 + Y_1 \tilde{\Delta}_2 + \tilde{u}_2, \\ Y_j &= \begin{cases} 1 & \text{if } Y_j^* > 0 \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, 2, \end{aligned} \tag{5.16}$$

If we have access to an excluded instrumental variable  $S$ , which, together with  $\tilde{Z}'_{-11}, \tilde{Z}'_2$  is fully independent of  $(\tilde{U}, \tilde{\beta}, \tilde{\Delta})$ , but which is related to the endogenous variable via a nonseparable equation

$$\tilde{Z}_{11} = \varphi(S, \tilde{Z}'_{-11}, \tilde{Z}'_2, V),$$

where  $\varphi$  is strictly monotonic in the last argument  $V$ . If we strengthen the independence condition to  $(S, \tilde{Z}'_{-11}, \tilde{Z}'_2)$  fully independent of  $(\tilde{U}, \tilde{\beta}, \tilde{\Delta}, V)$ , then this allows the construction of a control function in the sense of Imbens and Newey (2009). This implies that  $\tilde{Z}$  is independent of  $(\tilde{U}, \tilde{\beta}, \tilde{\Delta})$  conditional on  $V$ , and therefore allows to do the entire analysis performed above, if we condition in addition on  $V = v$  for every  $v \in V$ . This procedure allows to recover the conditional density  $f_{\tilde{U}, \tilde{\beta}, \tilde{\Delta} | V}$  and by integrating out  $V$ , allows to recover  $f_{\tilde{U}, \tilde{\beta}, \tilde{\Delta}}$ . See Hoderlein and Sherman (2012) for a related procedure in the binary choice random coefficients model. This assumption could be relaxed to allow for random coefficients in the selection equation, as in Gautier and Hoderlein (2012), but we leave the details for future research.

## 5.8 Recovering structural objects

While the distribution of random coefficients is of interest in itself, and allows to determine means, variances or other functionals of the distribution, often time the counterfactual choice probabilities are at the center of interest. For instance, given an estimator for the density  $f_{\theta_1}$ ,

we can estimate

$$P_c [Y_1 = 1|Z_1 = z, Y_2 = 1] = \int 1\{z't_1 > 0\} f_{\theta_1}(t_1) d\sigma_{k_1}(t_1)$$

where the subscript  $c$  denotes counterfactuals. From this quantity, one may recover all derivatives, respectively discrete differences, e.g.,

$$P_c [Y_1 = 1|Z_1 = z', Y_2 = 1] - P_c [Y_1 = 1|Z_1 = z, Y_2 = 1].$$

Another interesting object would be the probability of a specific action profile being a pure strategy Nash equilibrium (NE).<sup>6</sup> For example, as in Aradillas-Lopez (2012), we may estimate

$$\begin{aligned} P_c((1, 0) \text{ is a NE} | (Z_1, Z_2) = (z_1, z_2)) \\ = \int 1\{z'b_1 > 0\} 1\{z'_2 t_2 \leq 0\} f_{\beta_1, \theta_2}(b_1, t_2) d\sigma(b_1, t_2). \end{aligned} \quad (5.17)$$

This ensures that one may estimate related objects. For example, the aggregate propensity of the equilibrium selection mechanism to select the action profile  $(1, 0)$  is given by the ratio of the actual entry probability  $r_{(1,0)}(z_1, z_2)$  and  $P_c((1, 0) \text{ is a NE} | (Z_1, Z_2) = (z_1, z_2))$ .

## 5.9 Common explanatory variables

Thus far, we assumed that the explanatory variables  $\tilde{Z}_1$  and  $\tilde{Z}_2$  do not have elements in common. However, the behavior of firms who are acting on the same market will at least partially depend on the same environment, and one may hence want to choose explanatory variables that are common to both players. To illustrate the limitations in the case of common explanatory variables, we show that the operators  $\mathcal{T}$  and  $\mathcal{S}$  degenerate if  $Z \equiv Z_1 = Z_2 \in \mathbb{S}^k$ . Afterward, we discuss two additional sets of assumptions which allow to overcome these limitations: one set that involves individual specific covariates, and which is otherwise not restrictive, and one where all variables are common. We only present the case of strategic substitutes in which the interaction coefficients are non-positive.

Let us assume the function  $R_{(1,1)}$  is known on  $\mathbb{S}^k$  and that  $Z \equiv Z_1 = Z_2 \in \mathbb{S}^k$ . As before, the situation of a duopoly is described by the equation

$$R_{(1,1)}(z, z) = \iint_{\mathbb{S}^k \times \mathbb{S}^k} 1\{z't_1 > 0\} 1\{z't_2 > 0\} f_{\theta}(t_1, t_2) d\sigma(t_1, t_2).$$

This can be written as an operator equation  $R_{(1,1)}(z, z) = (\mathcal{T}_c f_{\theta})(z)$  where  $\mathcal{T}_c : L^2(\mathbb{S}^k \times \mathbb{S}^k) \rightarrow$

---

<sup>6</sup>We are indebted to Andres Aradillas-Lopez for this point.

$L^2(\mathbb{S}^k)$ . It is instructive to characterize the null space in the one dimensional case  $k = 1$ .

**Theorem 5.1.** *Let  $\lambda_n := \lambda(n, 1)$  be the eigenvalues of the hemispherical transform  $\mathcal{H}_{\mathbb{S}^1}$  to the Fourier basis  $\varphi_n(t) = (2\pi)^{-1} \exp(-int)$ . The null space of  $\mathcal{T}_c : L^2(\mathbb{S}^1 \times \mathbb{S}^1) \rightarrow L^2(\mathbb{S}^1)$  is*

$$\begin{aligned} \mathcal{N}(\mathcal{T}_c) = \text{span}(\{ & \lambda_{n_1} \lambda_{n_2} \varphi_{n_1} \varphi_{n_2} - \lambda_{m_1} \lambda_{m_2} \varphi_{m_1} \varphi_{m_2} \mid n_1 + n_2 = m_1 + m_2 \} \\ & \cup \{ \varphi_0 \varphi_n - \varphi_n \varphi_0 \mid n \text{ is even} \} \\ & \cup \{ \varphi_{n_1} \varphi_{n_2} \mid n_1 \text{ or } n_2 \text{ is even} \}). \end{aligned}$$

This theorem is a direct consequence of Theorem D.1. The last part of the null space contains everything but the component-wise odd part and the odd part of the marginals. The second part contains the difference between the odd part of the marginals  $f_{\theta_1}^- - f_{\theta_2}^-$ . Therefore, we only get information about the sum of the odd part of the marginals  $f_{\theta_1}^- + f_{\theta_2}^-$ . This is also true for higher dimensions as is shown in the Appendix. Finally, the first part of the null space contains much of the dependence structure of  $f_{\theta_1}^-$  and  $f_{\theta_2}^-$ . Obviously, no useful information about the dependence structure can be recovered. By an analogous argument, the same holds true for  $f_{\beta_1}^- + f_{\beta_2}^-$  if  $R_{(0,0)}(z, z)$  is known in  $\mathbb{S}^k$ . This illustrates that the information provided by the data in the common covariates case is not sufficient to recover the joint or marginal distribution of random parameters.

Indeed, these objects do not even provide enough information for recovering the distribution of the interaction effects. To give an example of an assumption that allows identification of  $f_{\Delta_1}$  and  $f_{\Delta_2}$  in the case when all covariates are common, we consider the following common coefficient assumption:

**Assumption 5.7.**  $\Delta_1 = \Delta_2$  almost surely.

With this assumption,  $f_{\Delta_1} = f_{\Delta_2}$  is related to  $f_{\theta_1} + f_{\theta_2}$  and  $f_{\beta_1} + f_{\beta_2}$  by a special convolution similar to Theorem 3

$$(f_{\theta_1} + f_{\theta_2})(t) = \int_{-1}^0 (f_{\beta_1} + f_{\beta_2})((t - w\mathbf{n}_1)/\|t - w\mathbf{n}_1\|) \|t - w\mathbf{n}_1\|^{k-1} f_{\Delta_1}(w) dw.$$

For every  $t \in \mathbb{S}^k$  this is a one dimensional deconvolution problem. It gives the same solution  $f_{\Delta_1}$  for every  $t$ , if it is identified. Hence,  $f_{\Delta_1}$  is identified if for every  $n \in \mathbb{Z}$  there is a  $t \in \mathbb{S}^k$ , such that the Fourier coefficient

$$\int_{-1}^0 (f_{\beta_1} + f_{\beta_2})((t - w\mathbf{n}_1)/\|t - w\mathbf{n}_1\|) \|t - w\mathbf{n}_1\|^{k-1} e^{-i2\pi n w} dw \neq 0$$

does not vanish.

The situation turns out to be much more benign in the case where some instruments coincide for both players, and some do not. This case is arguably the most common in applications, and can be shown to yield point identification under additional independence assumption. Let us denote the common variables by  $\tilde{Z}_c$  and player specific variables by  $\tilde{Z}_1$  and  $\tilde{Z}_2$ . The coefficient vectors  $\tilde{\beta}_1$  and  $\tilde{\beta}_2$  can be separated accordingly into coefficients  $\tilde{\beta}_{c,1}$  and  $\tilde{\beta}_{c,2}$  corresponding to the common variables and coefficients  $\tilde{\beta}_{*,1}$  and  $\tilde{\beta}_{*,2}$  corresponding to  $\tilde{Z}_1$  and  $\tilde{Z}_2$ . Hence, we can write  $\tilde{\beta}'_i = (\tilde{\beta}'_{c,i}, \tilde{\beta}'_{*,i})$ . We will analyze this case only under the assumption that the coefficients for the common variables are independent of the coefficients for the specific variables for each player.

**Assumption 5.8.**  $\tilde{\beta}_{0i}$  is independent of  $\tilde{\beta}_{*i}$  for  $i = 1, 2$

One consequence of this assumption is that the term  $z'_c \tilde{\beta}_{c,i}$  can be treated like an intercept for each player. Hence, we can integrate it into the player specific intercept  $\tilde{u}_i$  of our model by setting  $\tilde{u}_{c,i} \equiv \tilde{u}_i + \tilde{\beta}'_{c,i} Z_c$ . So, for every value  $z_c$  of  $\tilde{Z}_c$  the Equations (2.1) and (2.1) of the model can be rewritten as

$$\begin{aligned} Y_1^* &= (z'_c, \tilde{Z}'_1) (\tilde{\beta}'_{c,1}, \tilde{\beta}'_{*,1})' + Y_2 \tilde{\Delta}_1 + \tilde{u}_1 = \tilde{Z}'_1 \tilde{\beta}_{*,1} + Y_2 \tilde{\Delta}_1 + (\tilde{u}_{c,1} | z_c), \\ Y_2^* &= (z'_c, \tilde{Z}'_2) (\tilde{\beta}'_{c,2}, \tilde{\beta}'_{*,2})' + Y_1 \tilde{\Delta}_2 + \tilde{u}_2 = \tilde{Z}'_2 \tilde{\beta}_{*,2} + Y_1 \tilde{\Delta}_2 + (\tilde{u}_{c,2} | z_c). \end{aligned}$$

Where  $(\tilde{u}_{c,1} | z_c)$  denotes the new intercept conditioned on  $z_c$ . Treating the common variables and its coefficients as an intercept transforms the problem formally into a problem with only specific variables. For every  $z_c$  we can apply the method presented in Chapters 3 and 4 to estimate the densities of  $(\theta | z_c)$  and  $(\beta | z_c)$ . Marginalizing the density  $f_{\beta | z_c}(t_1, t_2 | z_c)$  to the first components of the vectors  $t_1$  and  $t_2$  gives the densities of  $(\tilde{u}_{c,1} / \|\beta_1^*\| | z_c)$  and  $(\tilde{u}_{c,2} / \|\beta_2^*\| | z_c)$ . This allows to recover the densities of the scaled coefficients  $\tilde{\beta}_{c,1} / \|\beta_1^*\|$  and  $\tilde{\beta}_{c,2} / \|\beta_2^*\|$  by inverting a Radon transform. See HKM (2010). It is, however, not possible to recover the joint of  $\tilde{\beta}_{c,1} / \|\beta_1^*\|$  and  $\tilde{\beta}_{c,2} / \|\beta_2^*\|$  with this method because both coefficients can be observed only for one common explanatory variable.

Furthermore, the joint densities of the scaled strategic interaction terms  $\Delta_1$  and  $\Delta_2$  can be computed by deconvolving the densities of  $(\theta | z_c)$  and  $(\beta | z_c)$ . Under Assumption 3.4, the interaction terms do not depend on  $z_c$ . Hence, it is as well possible to compute first the densities of  $\theta$  and  $\beta$  with the unconditioned intercepts  $u_{c,i}$ , and then the deconvolution.

## 6 A numerical study

We illustrate our estimation procedure through a numerical study. In this experiment, we let  $Z_j^* = (1, \tilde{Z}_j^{(1)}, \tilde{Z}_j^{(2)})$  for  $j = 1, 2$ , where  $(\tilde{Z}_j^{(1)}, \tilde{Z}_j^{(2)})'$  follows the standard bivariate normal



distribution. Similarly, for each  $j$ , we generate  $(\tilde{u}_j, \tilde{\beta}_j^{(1)})$  as a standard bivariate normal vector. We then let  $\tilde{\beta}_j^{(2)} = 1$  for  $j = 1, 2$ . In this setting, Assumptions 3.1 and 3.3 are satisfied with  $c_j = (0, 0, 1)$  and  $e_j = (0, 0, -1)$ . The interaction effects are generated as  $(\tilde{\Delta}_1, \tilde{\Delta}_2) = (-\exp(V_1), -\exp(V_2))$ , where  $(V_1, V_2)$  is a bivariate normal vector with mean  $\mu_\Delta$  and covariance matrix  $\Sigma_\Delta$ . We consider two specifications. Specification 1 sets  $\mu_\Delta = (-0.7, -0.7)'$  and  $\Sigma_\Delta$  to the identity matrix. Specification 2 is the same as Specification 1 except that we introduce a positive correlation between  $\Delta_1$  and  $\Delta_2$  by setting the off-diagonal components of  $\Sigma_\Delta$  to 0.9. The entry outcomes are then generated according to (2.1)-(2.3). The sample size is  $n = 1000$ .

Our estimator of  $f_\theta$  is implemented using the smoothed projection kernel with

$$\chi_j(n, T) = (1 - (\zeta_{n, k_j+1} / \zeta_{T, k_j+1} + 1)^{s/2})^l, \quad (6.1)$$

where we use the tuning parameters  $s = 1$ ,  $l = 9$ , and  $T_N = 11$ .<sup>7</sup> The trimming parameter is  $r = 4$ . For the nonparametric estimators of unknown densities, we use the projection estimators defined in (4.5) and (4.6) with the smoothed projection kernel with  $s = 2$ ,  $l = 3$ , and  $T_N = 5$ .

Figure 1 and 2 show the joint density of  $(\theta_1^{(1)}, \theta_1^{(2)})$  and that of  $(\theta_1^{(1)}, \theta_2^{(1)})$  respectively and their estimates under Specification 1. These plots are produced by marginalizing the joint density  $f_\theta$  by numerical integration. Marginalization is carried out so that the resulting density is evaluated on a one-dimensional unit sphere.<sup>8</sup> For example, in Figure 1, the red curve represents the true density  $f_{\theta_1^{(1)}, \theta_1^{(2)}}$ . This density is defined on  $\mathbb{S}^1$ , which is depicted as a dashed circle in the figure. Each evaluation point  $(t_1^{(1)}, t_1^{(2)}) \in \mathbb{S}^1$  of the density is then a point on this unit circle. For each  $(t_1^{(1)}, t_1^{(2)}) \in \mathbb{S}^1$ , the red curve's distance (or height) from the unit circle represents the value of the density:  $f_{\theta_1^{(1)}, \theta_1^{(2)}}(t_1^{(1)}, t_1^{(2)})$ . In other words, its distance from the origin gives  $1 + f_{\theta_1^{(1)}, \theta_1^{(2)}}(t_1^{(1)}, t_1^{(2)})$ . Similarly, the blue curve represents our estimate  $\hat{f}_{\theta_1^{(1)}, \theta_1^{(2)}}$  whose distance from the unit circle corresponds to  $\hat{f}_{\theta_1^{(1)}, \theta_1^{(2)}}(t_1^{(1)}, t_1^{(2)})$ . Overall, our estimator captures the shape of the true density well. This is still true when the two interaction effects are correlated. Figure 3 shows the joint density of  $(\theta_1^{(1)}, \theta_2^{(1)})$  and its estimate under Specification 2. The shape of the true density is also captured by the estimator in this case.

## 7 Conclusion and Outlook

This paper studies nonparametric identification of the joint distribution of random coefficients in static games of complete information. We give conditions under which the joint distribution of random coefficients except those on the interaction effects is identified. Moreover, we provide

<sup>7</sup>The smoothed projection kernel with  $\chi_j$  in (6.1) is called the Riesz kernel. See Ditzian (1998) and GK (2013) for details.

<sup>8</sup>See GK (2013) Section 5.1 for details on marginalization of densities defined on spheres.

additional conditions that allow to point identify the joint density of the interaction effects. We also discuss various ways to extend our main identification result. We further show that our constructive identification strategy allows us to construct sample counterpart estimators. We analyze their asymptotic properties, and illustrate their finite sample behavior in a numerical study.

We have focused on nonparametric identification of the density of random coefficients from uniquely predicted outcomes. An interesting direction would be to study possible efficiency gains by considering simultaneously the two integral equality restrictions obtained from the no entry and duopoly outcomes and additional integral inequality restrictions, which can be obtained from the monopoly outcomes. We pursue this in another paper that studies a setting, in which the density of random coefficients are partially identified by integral equality and inequality restrictions.

Another interesting direction would be to apply the developed estimation procedure to empirical examples in which heterogeneity plays an important role. Such examples include airline markets, households' labor supply decisions, and bilateral trade agreements.

## References

- [1] Amemiya, T. (1974): "Multivariate Regression and Simultaneous Equation Models when the Dependent Variables Are Truncated Normal". *Econometrica*, **42**, 999–1012.
- [2] Aradillas-Lopez, A. (2010): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information". *Journal of Econometrics*, **157**, 409–431.
- [3] Aradillas-Lopez, A. (2012): "Inference in Ordered Response Games with Complete Information". Working Paper.
- [4] Bajari, P. and H. Hong and S.P. Ryan (2010): "Identification and Estimation of a Discrete Game of Complete Information" . *Econometrica*, **78**, 1529–1568.
- [5] Beran, R., A. Feuerverger, and P. Hall (1996): "On Nonparametric Estimation of Intercept and Slope in Random Coefficients Regression". *Annals of Statistics*, **2**, 2569–2592.
- [6] Beresteanu, A. and I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions . *Econometrica*, **79**, 1785–1821.
- [7] Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry". *Econometrica*, **60**, 889–917.

- [8] Berry, S. T. and P. A. Haile (2009): “Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers”. Working paper.
- [9] Berry, S. T. and P. A. Haile (2011): “Identification in a Class of Nonparametric Simultaneous Equations Models”. Working Paper.
- [10] Berry, S. T. and E. Tamer (2006): “Identification in Models of Oligopoly Entry”. *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume 2*.
- [11] Bjorn, P.A. and Q. H. Vuong (1985): “Simultaneous Equations Models for Dummy Endogenous Variables: a Game Theoretic Formulation with an Application to Labor Force Participation”. Working paper.
- [12] Blundell, W. R. and J. L. Powell (2004): “Endogeneity in Semiparametric Binary Response Models”. *Review of Economic Studies*, **71**, 655–679.
- [13] Bresnahan, T. F. and P.C. Reiss (1990): “Entry in Monopoly Market”. *Review of Economic Studies*, **57**, 531–553.
- [14] Bresnahan, T. F. and P.C. Reiss (1991): “Empirical Models of Discrete Games”. *Journal of Econometrics*, **48**, 57–81.
- [15] Burger, M. and Osher, S. (2004): “Convergence rates of convex variational regularization”. *Inverse Problems*, **20**, 1411–1421.
- [16] Carrasco, M. and J.P. Florens (2010): “A Spectral Method for Deconvolving a Density”. *Econometric Theory*, **27**, 546–581.
- [17] Carroll, R.J. and P. Hall (1988): “Optimal Rates of Convergence for Deconvolving a Density”. *Journal of the American Statistical Association*, **83**, 1184–1186.
- [18] Chesher, A. and A. M. Rosen “Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification”. CEMMAP Working Paper
- [19] Ciliberto, F. and E. Tamer (2009): “Market structure and multiple equilibria in airline markets”. *Econometrica*, **77**, 1791–1828.
- [20] de Paula, A., and X. Tang (2012): “Inference of Signs of Interaction Effects in Simultaneous Games With Incomplete Information”. *Econometrica*, **80**, 143–172.
- [21] Devroye, L. (1989): “Consistent Deconvolution in Density Estimation”. *Canadian Journal of Statistics*, **17**, 235–239.

- [22] Ditzian, Z. (1998): “Fractional Derivatives and Best Approximation”. *Acta Mathematica Hungarica*, **81**, 323–348.
- [23] Eggermont, P. (1993): “Maximum Entropy Regularization for Fredholm Integral Equations of the First Kind”. *SIAM Journal on Mathematical Analysis*, **24**, 1557–1576.
- [24] Fan, Y., and S. S. Park (2010): “Sharp Bounds on the Distribution of Treatment Effects and Their Statistical Inference”. *Econometric Theory*, **26**, 931–951.
- [25] Fan, Y. and D. Zhu (2009): “Partial Identification and Confidence Sets for Functionals of the Joint Distribution of Potential Outcomes”. Working paper.
- [26] Folland, G.B. (1999): *Real Analysis: Modern Techniques and Their Applications*. Wiley, New York.
- [27] Fox, J.T. and N. Lazzati (2012): “Identification of Discrete Games and Choice Models for Bundles”. Working paper.
- [28] Fox, J.T., S.P. Ryan, P. Bajari, and K. Kim (2012): “The random coefficients logit model is identified”. *Journal of Econometrics*, **166**, 204–212.
- [29] Gautier, E., and S. Hoderlein (2012): “Estimating the Distribution of Treatment Effects”. CeMMAP Working Paper.
- [30] Gautier, E., and Y. Kitamura (2013): “Nonparametric Estimation in Random Coefficients Binary Choice Models”. *Econometrica*, forthcoming.
- [31] Gautier, E., and E. Le Pennec (2011): “Adaptive Estimation in the Nonparametric Random Coefficients Binary Choice Model by Needlet Thresholding, Working Paper
- [32] Grasmair, M and Haltmeier, M. and Scherzer, O. (2008): “Sparse regularization with  $l^q$  penalty term”. *Inverse Problems*, **24**, 055020.
- [33] Groemer, H. (1996): *Geometric Applications of Fourier Series and Spherical Harmonics*. Cambridge University Press., Cambridge.
- [34] Heckman, J. J. (1978): “Dummy Endogenous Variables in a Simultaneous Equation System”. *Econometrica*, **46**, 931–959.
- [35] Heckman, J. J., J. Smith, and N. Clements (1997): “Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts”. *Review of Economic Studies*, **64**, 487–535.

- [36] Hendriks, H. (1990): “Nonparametric Estimation of a Probability Density on a Riemannian Manifold Using Fourier Expansions”. *The Annals of Statistics*, **18**, 832–849.
- [37] Hoderlein, S, J. Klemelä, and E. Mammen (2010): “Analyzing the Random Coefficient Model Nonparametrically”. *Econometric Theory*, **26**, 804–837.
- [38] Hoderlein, S. and R. Sherman (2012): Identification and estimation in a correlated random coefficients binary response model, CeMMAP Working Paper.
- [39] Hu, Y. and G. Ridder (2010): “On deconvolution as a first stage nonparametric estimator”. *Econometric Reviews*, **29**, 365–396.
- [40] Ichimura, H., and T. Thompson (1998): “Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution”. *Journal of Econometrics*, **86**, 269–295.
- [41] Imbens, G. and W. Newey (2009): “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity”. *Econometrica*, **77**, 1481–1512.
- [42] Kline, B. (2011): “Identification of Complete Information Games”. Working Paper.
- [43] Kline, B. and Tamer, E. (2012): “Bounds for Best Response Functions in Binary Games”. *Journal of Econometrics*, **166**, 92–105.
- [44] Lewbel, A., and X. Tang (2012): “Identification and Estimation of Games with Incomplete Information using Excluded Regressors”. Working Paper
- [45] Lukacs, E. (1970): *Characteristic Functions*. Statistical Monographs and Courses. Griffin.
- [46] Masten, M. (2012): “Random Coefficients on Endogenous Variables in Simultaneous Equations Models”. Working Paper
- [47] Matzkin, R. L. (2008): “Identification in Nonparametric Simultaneous Equations Models”. *Econometrica*, **76**, 94578.
- [48] Matzkin, R. L. (2012): “Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity”. *Journal of Econometrics*, **166**, 10615.
- [49] Natterer, F. (1986): *The Mathematics of Computerized Tomography*. Wiley, Chichester.
- [50] Reed, M. and B. Simon (1980): *Methods of modern mathematical physics. I* Academic Press Inc., New York.

- [51] Reiss, P. C. and P. Spiller (1989): “Competition and Entry in Small Airline Markets”. *Journal of Law and Economics*, **32**, 179–202.
- [52] Resmerita, E. (2005): “Regularization of ill-posed problems in Banach spaces: convergence rates”. *Inverse Problems*, **21**, 1303–1314.
- [53] Rubin, B. (1999): “Inversion and characterization of the hemispherical transform”. *J. Anal. Math.*, **77**, 105–128.
- [54] Tamer, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria”. *Review of Economic Studies*, **70**, 147–165.

## SUPPLEMENTAL APPENDIX

In this supplemental appendix, we include the proofs of results stated in the main text. The contents of the supplemental appendix are organized as follows. Appendix A contains the proof of Theorems 3.1, 3.2, and 3.3 and required auxiliary results. Appendix B gives a brief review of Fourier series on spheres and the proof of auxiliary lemmas useful for constructing our estimator in Section 4. Appendix C contains regularity conditions required by Theorem 4.2, auxiliary lemmas, and the proof of Theorem 4.2. Appendix D contains the results of the numerical study.

### APPENDIX A: Proof of Theorems 3.1, 3.2, and 3.3

**Lemma A.1.** *Let  $k_1, k_2 \in \mathbb{N}$ . Let  $f$  be a non-negative function on  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ . Suppose that  $\text{supp}(f) \subseteq H_{v_1} \times H_{v_2}$  for some  $(v_1, v_2) \in \mathbb{S}^{\ell_1} \times \mathbb{S}^{\ell_2}$ . Then,*

$$f(z_1, z_2) = \begin{cases} 4f^{--}(z_1, z_2) & (z_1, z_2) \in H_{v_1} \times H_{v_2} \\ 0 & \text{otherwise.} \end{cases}$$

*Proof.* Let  $(z_1, z_2) \in H_{v_1} \times H_{v_2}$ . Then,  $f(-z_1, z_2) = f(z_1, -z_2) = f(-z_1, -z_2) = 0$ , because  $\text{supp}(f) \subseteq H_{v_1} \times H_{v_2}$ . Therefore, by (2.6),  $f^{--}(z_1, z_2) = f(z_1, z_2)/4$  on  $H_{v_1} \times H_{v_2}$  and vanishes elsewhere.  $\square$

**Lemma A.2.** *Suppose Assumptions 2.3 and 3.1 hold. Then,*

$$f_\theta(t_1, t_2) = 4f_\theta^{--}(t_1, t_2)1\{f_\theta^{--}(t_1, t_2) > 0, f_{\theta_1}^-(t_1) > 0, f_{\theta_2}^-(t_2) > 0\}, \quad \text{for all } (t_1, t_2) \in \mathbb{S}^{k_1} \times \mathbb{S}^{k_2}.$$

*Proof of Lemma A.2.* By Assumption 3.1 and Lemma A.1,

$$f_\theta(t_1, t_2) = \begin{cases} 4f_\theta^{--}(t_1, t_2) & (t_1, t_2) \in \text{supp}(f_\theta) \subset H_{c_1} \times H_{c_2} \\ 0 & \text{otherwise.} \end{cases}$$

For the conclusion of the Lemma, it then suffices to show that  $(t_1, t_2) \in \text{supp}(f_\theta)$  if and only if  $f_\theta^{--}(t_1, t_2) > 0$ ,  $f_{\theta_1}^-(t_1) > 0$ , and  $f_{\theta_2}^-(t_2) > 0$ . Toward this end, define

$$\begin{aligned} \text{supp}(f_\theta)^{-+} &\equiv \{(t_1, t_2) \in \mathbb{S}^{k_1} \times \mathbb{S}^{k_2} : (-t_1, t_2) \in \text{supp}(f_\theta)\} \\ \text{supp}(f_\theta)^{+-} &\equiv \{(t_1, t_2) \in \mathbb{S}^{k_1} \times \mathbb{S}^{k_2} : (t_1, -t_2) \in \text{supp}(f_\theta)\} \\ \text{supp}(f_\theta)^{--} &\equiv \{(t_1, t_2) \in \mathbb{S}^{k_1} \times \mathbb{S}^{k_2} : (-t_1, -t_2) \in \text{supp}(f_\theta)\}. \end{aligned}$$

We examine the signs of  $f_\theta^{--}$ ,  $f_{\theta_1}^-$ , and  $f_{\theta_2}^-$  on a partition of  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ , which consists of the five disjoint subsets:  $\text{supp}(f_\theta)$ ,  $\text{supp}(f_\theta)^{-+}$ ,  $\text{supp}(f_\theta)^{+-}$ ,  $\text{supp}(f_\theta)^{--}$ , and the rest. Specifically, Assumption 3.1 and

(2.6) imply

$$f_{\theta}^{-}(t_1, t_2) = \begin{cases} f_{\theta}(t_1, t_2)/4 & (t_1, t_2) \in \text{supp}(f_{\theta}) \\ -f_{\theta}(-t_1, t_2)/4 & (t_1, t_2) \in \text{supp}(f_{\theta})^{-+} \\ -f_{\theta}(t_1, -t_2)/4 & (t_1, t_2) \in \text{supp}(f_{\theta})^{+-} \\ f_{\theta}(-t_1, -t_2)/4 & (t_1, t_2) \in \text{supp}(f_{\theta})^{--} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.1})$$

Similarly,  $f_{\theta_j}^{-}(t_j)$  being given by  $(f_{\theta_j}(t_j) - f_{\theta_j}(-t_j))/2$  implies

$$f_{\theta_1}^{-}(t_1) = \begin{cases} f_{\theta_1}(t_1)/2 & (t_1, t_2) \in \text{supp}(f_{\theta}) \\ -f_{\theta_1}(-t_1)/2 & (t_1, t_2) \in \text{supp}(f_{\theta})^{-+} \\ f_{\theta_1}(t_1)/2 & (t_1, t_2) \in \text{supp}(f_{\theta})^{+-} \\ -f_{\theta_1}(-t_1)/2 & (t_1, t_2) \in \text{supp}(f_{\theta})^{--} \\ 0 & \text{otherwise} \end{cases}, \text{ and } f_{\theta_2}^{-}(t_2) = \begin{cases} f_{\theta_2}(t_2)/2 & (t_1, t_2) \in \text{supp}(f_{\theta}) \\ f_{\theta_2}(t_2)/2 & (t_1, t_2) \in \text{supp}(f_{\theta})^{-+} \\ -f_{\theta_2}(-t_2)/2 & (t_1, t_2) \in \text{supp}(f_{\theta})^{+-} \\ -f_{\theta_2}(-t_2)/2 & (t_1, t_2) \in \text{supp}(f_{\theta})^{--} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.2})$$

By (A.1)-(A.2), it holds that  $(t_1, t_2) \in \text{supp}(f_{\theta})$  if and only if  $f_{\theta}^{-}(t_1, t_2) > 0$ ,  $f_{\theta_1}^{-}(t_1) > 0$ , and  $f_{\theta_2}^{-}(t_2) > 0$ . This establishes the claim of the lemma.  $\square$

**Lemma A.3.** *Suppose Assumptions 2.1-2.4, and 3.2 hold.*

(i) *If additionally Assumption 3.1 holds, then there exists a unique extension  $R_{(1,1)}$  of  $r_{(1,1)}$  defined on  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ , which is given by (3.6).*

(ii) *If additionally Assumption 3.3 holds, then there exists a unique extension  $R_{(0,0)}$  of  $r_{(0,0)}$  defined on  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ , which is given by (3.7).*

Proof of Lemma A.3. For (i), we make use of Assumption 3.2 to reconstruct  $r_{(1,1)}$  on the rest of  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ . First, under Assumptions 2.1-2.4,  $f_{\theta}$  exists and satisfies equation (2.4). Equation (2.4) and the definition of the hemispherical transform then imply

$$R_{(1,1)}(-z_1, z_2) + r_{(1,1)}(z_1, z_2) = \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} \mathbf{1}\{z_2' t_2 > 0\} f_{\theta}(t_1, t_2) d\sigma(t_1, t_2) = \mathcal{H}_{\mathbb{S}^{k_2}} f_{\theta_2}(z_2).$$

Rewriting this yields

$$R_{(1,1)}(-z_1, z_2) = \mathcal{H}_{\mathbb{S}^{k_2}} f_{\theta_2}(z_2) - r_{(1,1)}(z_1, z_2), \quad (\text{A.3})$$

and by symmetry,

$$R_{(1,1)}(z_1, -z_2) = \mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}(z_1) - r_{(1,1)}(z_1, z_2). \quad (\text{A.4})$$



Furthermore, the formula  $\mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}(-z_1) = 1 - \mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}(z_1)$  yields

$$R_{(1,1)}(-z_1, -z_2) = 1 - (\mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}(z_1, z_2) + \mathcal{H}_{\mathbb{S}^{k_2}} f_{\theta_2}(z_1, z_2)) + r_{(1,1)}(z_1, z_2). \quad (\text{A.5})$$

The last three equations define a unique extension of  $r$  on  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ . But to make use of this extension, we have to identify the hemispherical transforms of the marginal distributions in advance. As already argued in Section 2.2 this is possible under Assumption 3.1. Since  $(c_1, c_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}$ ,

$$\mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}(z_1) = r_{(1,1)}(z_1, c_2) \quad \text{and} \quad \mathcal{H}_{\mathbb{S}^{k_2}} f_{\theta_2}(z_2) = r_{(1,1)}(c_1, z_2)$$

are well-defined on  $H_{\mathbf{n}_1}$  and  $H_{\mathbf{n}_2}$  respectively. We can extend these functions to  $\mathbb{S}^{k_1}$  and to  $\mathbb{S}^{k_2}$  by using  $\mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}(-z_1) = 1 - \mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}(z_1)$  and similarly  $\mathcal{H}_{\mathbb{S}^{k_2}} f_{\theta_2}(-z_2) = 1 - \mathcal{H}_{\mathbb{S}^{k_2}} f_{\theta_2}(z_2)$ . Finally, we can write the extension of  $r_{(1,1)}$  for arbitrary  $(z_1, z_2) \in \mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$  as:

$$R_{(1,1)}(z_1, z_2) = \begin{cases} r_{(1,1)}(z_1, z_2) & \text{for } z_1 \in H_{\mathbf{n}_1} \text{ and } z_2 \in H_{\mathbf{n}_2} \\ r_{(1,1)}(c_1, z_2) - r_{(1,1)}(-z_1, z_2) & \text{for } z_1 \notin H_{\mathbf{n}_1} \text{ and } z_2 \in H_{\mathbf{n}_2} \\ r_{(1,1)}(z_1, c_2) - r_{(1,1)}(z_1, -z_2) & \text{for } z_1 \in H_{\mathbf{n}_1} \text{ and } z_2 \notin H_{\mathbf{n}_2} \\ 1 - (r_{(1,1)}(-z_1, c_2) + r_{(1,1)}(c_1, -z_2)) + r_{(1,1)}(-z_1, -z_2) & \text{for } z_1 \notin H_{\mathbf{n}_1} \text{ and } z_2 \notin H_{\mathbf{n}_2}. \end{cases}$$

The proof of (ii) is similar. Hence it is omitted.  $\square$

**Lemma A.4.** *Suppose Assumptions 2.1-2.4, and 3.1-3.2 hold. Then,*

$$\mathcal{T}f_{\theta}^{--}(z_1, z_2) = R_{(1,1)}^{--}(z_1, z_2), \quad \mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}^-(z_1) = R_{(1,1)}^-(z_1, c_2), \quad \text{and} \quad \mathcal{H}_{\mathbb{S}^{k_2}} f_{\theta_2}^-(z_2) = R_{(1,1)}^-(c_1, z_2). \quad (\text{A.6})$$

Proof of Lemma A.4. First, by Assumptions 2.1-2.4, and 3.1-3.2 and Lemma A.3 (i),  $R_{(1,1)}$  exists. By (2.6), it is then straightforward to show that

$$R_{(1,1)}^{--}(z_1, z_2) = \begin{cases} r_{(1,1)}(z_1, z_2) - \frac{1}{2}r_{(1,1)}(c_1, z_2) - \frac{1}{2}r_{(1,1)}(z_1, c_2) + \frac{1}{4} & z_1 \in H_{\mathbf{n}_1}, z_2 \in H_{\mathbf{n}_2} \\ -r_{(1,1)}(-z_1, z_2) + \frac{1}{2}r_{(1,1)}(c_1, z_2) + \frac{1}{2}r_{(1,1)}(-z_1, c_2) - \frac{1}{4} & z_1 \notin H_{\mathbf{n}_1}, z_2 \in H_{\mathbf{n}_2} \\ -r_{(1,1)}(z_1, -z_2) + \frac{1}{2}r_{(1,1)}(c_1, -z_2) + \frac{1}{2}r_{(1,1)}(z_1, c_2) - \frac{1}{4} & z_1 \in H_{\mathbf{n}_1}, z_2 \notin H_{\mathbf{n}_2} \\ r_{(1,1)}(-z_1, -z_2) - \frac{1}{2}r_{(1,1)}(c_1, -z_2) - \frac{1}{2}r_{(1,1)}(-z_1, c_2) + \frac{1}{4} & z_1 \notin H_{\mathbf{n}_1}, z_2 \notin H_{\mathbf{n}_2}. \end{cases}$$

Let  $(z_1, z_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}$ . Then, again by (2.6),

$$\begin{aligned}
\mathcal{T}f_\theta^{--}(z_1, z_2) &= \frac{1}{4} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} 1\{z'_1 t_1 > 0\} 1\{z'_2 t_2 > 0\} \\
&\quad \times (f_\theta(t_1, t_2) - f_\theta(-t_1, t_2) - f_\theta(t_1, -t_2) + f_\theta(-t_1, -t_2)) d\sigma(t_1, t_2) \\
&= \frac{1}{4} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} 1\{z'_1 t_1 > 0\} 1\{z'_2 t_2 > 0\} (f_\theta(t_1, t_2) - f_\theta(-t_1, t_2)) d\sigma(t_1, t_2) \\
&\quad - \frac{1}{4} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} 1\{z'_1 t_1 > 0\} 1\{z'_2 t_2 \leq 0\} (f_\theta(t_1, t_2) - f_\theta(-t_1, t_2)) d\sigma(t_1, t_2) \\
&= \frac{1}{4} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} 1\{z'_1 t_1 > 0\} (2 \times 1\{z'_2 t_2 > 0\} - 1) (f_\theta(t_1, t_2) - f_\theta(-t_1, t_2)) d\sigma(t_1, t_2) \\
&= \frac{1}{4} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} (2 \times 1\{z'_1 t_1 > 0\} - 1) (2 \times 1\{z'_2 t_2 > 0\} - 1) (f_\theta(t_1, t_2)) d\sigma(t_1, t_2) \\
&= r_{(1,1)}(z_1, z_2) - \frac{1}{2} r_{(1,1)}(c_1, z_2) - \frac{1}{2} r_{(1,1)}(z_1, c_2) + \frac{1}{4} = R_{(1,1)}^{--}(z_1, z_2).
\end{aligned}$$

Similar calculations for the cases  $(z_1, z_2) \in H_{\mathbf{n}_1}^c \times H_{\mathbf{n}_2}$ ,  $(z_1, z_2) \in H_{\mathbf{n}_1} \times H_{\mathbf{n}_2}^c$ , and  $(z_1, z_2) \in H_{\mathbf{n}_1}^c \times H_{\mathbf{n}_2}^c$  can be carried out to show that  $\mathcal{T}f_\theta^{--} = R_{(1,1)}^{--}$  on each of these subsets of  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ . Hence,  $\mathcal{T}f_\theta^{--}(z_1, z_2) = R_{(1,1)}^{--}(z_1, z_2)$  for all  $(z_1, z_2) \in \mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ .

Next, using  $1\{-c'_2 \theta_2 > 0\} = 0, P - a.s.$ , it is straightforward to show that

$$R_{(1,1)}(z_1, c_2) = \begin{cases} r(z_1, c_2) & z_1 \in H_{\mathbf{n}_1} \\ 1 - r(z_1, c_2) & z_1 \notin H_{\mathbf{n}_1}. \end{cases}$$

By equations (3.2)-(3.4) in GK, the claim  $\mathcal{H}_{\mathbb{S}^{k_1}} f_{\theta_1}^-(z_1) = R_{(1,1)}^-(z_1, c_2)$  then follows.  $\mathcal{H}_{\mathbb{S}^{k_2}} f_{\theta_2}^-(z_2) = R_{(1,1)}^-(c_1, z_2)$  can be established similarly.  $\square$

**Proof of Theorem 3.1.** By Assumptions 2.1-2.4, 3.1-3.2, and Lemma A.3,  $R_{(1,1)}$  exists. Since  $r_{(1,1)}$  is identified,  $R_{(1,1)}$  is also identified. By Assumptions 2.1-2.4, and 3.1-3.2, and Lemma A.4, the functions  $f_\theta^{--}, f_{\theta_1}^-, f_{\theta_2}^-$  are in  $\mathcal{N}(\mathcal{T})^\perp$ , and hence can be obtained by inverting the operator equations in (A.6). Therefore,  $f_\theta^{--}, f_{\theta_1}^-, f_{\theta_2}^-$  are identified. By Assumptions 2.3, 3.1 and Lemma A.2,  $f_\theta$  is then identified. This establishes the claim of the theorem.  $\square$

**Proof of Theorem 3.2.** The proof of Theorem 3.2 is very similar to that of Theorem 3.1. Thus, it is omitted for brevity.  $\square$

**Proof of Theorem 3.3.** Let  $\mu$  denote Lebesgue measure on  $\mathbb{R}^{k_1+1} \times \mathbb{R}^{k_2+1}$ . Let  $f_{\beta^*}, f_{\theta^*}$  be the probability density functions of  $\beta^*$  and  $\theta^*$  with respect to  $\mu$ , respectively. For any Borel subset  $E$  of  $\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ , let

$$E_\infty \equiv \{(x_1, x_2) \in \mathbb{R}^{k_1+1} \times \mathbb{R}^{k_2+1} : (x_1/\|x_1\|, x_2/\|x_2\|) \in E\} \tag{A.7}$$

Then, for each Borel set  $E \subseteq \mathbb{S}^{k_1} \times \mathbb{S}^{k_2}$ , Theorem 2.49 in Folland (1999) implies

$$\begin{aligned} \int_E f_\theta(t_1, t_2) d\sigma(t_1, t_2) &= \int_{E_\infty} f_{\theta^*}(t_1^*, t_2^*) d\mu(t_1^*, t_2^*) \\ &= \int_0^\infty \int_0^\infty \int_E f_{\theta^*}(t_1 r_1, t_2 r_2) r_1^{k_1} r_2^{k_2} d\sigma(t_1, t_2) dr_1 dr_2 \end{aligned} \quad (\text{A.8})$$

Note that the integrals in (A.8) are with respect to  $\sigma$ -finite measures and that the integrand is jointly measurable. Further, the integrands are non-negative, and the integrals are finite. By interchanging integrals using Tonelli's theorem and noting that  $E$  was arbitrary:

$$f_\theta(t_1, t_2) = \int_0^\infty \int_0^\infty f_{\theta^*}(t_1 r_1, t_2 r_2) r_1^{k_1} r_2^{k_2} dr_1, dr_2 \quad (\text{A.9})$$

for  $\sigma$ -almost all  $(t_1, t_2)$ . Similarly, for almost all  $(b_1, b_2)$ ,

$$f_\beta(b_1, b_2) = \int_0^\infty \int_0^\infty f_{\beta^*}(b_1 r_1, b_2 r_2) r_1^{k_1} r_2^{k_2} dr_1, dr_2. \quad (\text{A.10})$$

By Assumption 3.4 and  $\theta_j^* = \beta_j^* + \tilde{\Delta}_j \mathbf{n}_j$  for  $j = 1, 2$ ,  $\theta_j^*$  is the convolution of  $\beta_j^*$  and  $\tilde{\Delta}_j \mathbf{n}_j$ . Therefore, one may write

$$f_{\theta^*}(t_1^*, t_2^*) = \int_{-\infty}^0 \int_{-\infty}^0 f_{\beta^*}(t_1^* - w_1^* \mathbf{n}_1, t_2^* - w_2^* \mathbf{n}_2) f_{\tilde{\Delta}}(w_1^*, w_2^*) dw_1^* dw_2^*. \quad (\text{A.11})$$

$$= \int_{-1}^0 \int_{-1}^0 f_{\beta^*}((t_1^*/\|t_1^*\| - w_1 \mathbf{n}_1) \|t_1^*\|, (t_2^*/\|t_2^*\| - w_2 \mathbf{n}_2) \|t_2^*\|) f_{\Delta}(w_1, w_2) dw_1 dw_2, \quad (\text{A.12})$$

where  $w_j^* = w_j \|t_j^*\|$ , and  $f_{\tilde{\Delta}}(w_1^*, w_2^*) = f_{\Delta}(w_1^*/\|t_1^*\|, w_2^*/\|t_2^*\|) \frac{1}{\|t_1^*\|} \frac{1}{\|t_2^*\|}$ . By letting  $t_j = t_j^*/\|t_j^*\|$  and  $r_j = \|t_j^*\|$ , we may then write

$$f_{\theta^*}(t_1 r_1, t_2 r_2) = \int_{-1}^0 \int_{-1}^0 f_{\beta^*}((t_1 - w_1 \mathbf{n}_1) r_1, (t_2 - w_2 \mathbf{n}_2) r_2) f_{\Delta}(w_1, w_2) dw_1 dw_2. \quad (\text{A.13})$$

By (A.9), (A.10), (A.13), and the change of variables with  $s_j = r_j \times \|t_j - w_j \mathbf{n}_j'\|$ ,  $j = 1, 2$ , we then

obtain:

$$f_\theta(t_1, t_2) = \int_0^\infty \int_0^\infty \int_{-1}^0 \int_{-1}^0 f_{\beta^*}((t_1 - w_1 \mathbf{n}_1)r_1, (t_2 - w_2 \mathbf{n}_2)r_2) f_\Delta(w_1, w_2) dw_1 dw_2 r_1^{k_1} r_2^{k_2} dr_1, dr_2 \quad (\text{A.14})$$

$$= \int_{-1}^0 \int_{-1}^0 \int_0^\infty \int_0^\infty f_{\beta^*}((t_1 - w_1 \mathbf{n}_1)r_1, (t_2 - w_2 \mathbf{n}_2)r_2) r_1^{k_1} r_2^{k_2} dr_1, dr_2 f_\Delta(w_1, w_2) dw_1 dw_2 \quad (\text{A.15})$$

$$= \int_{-1}^0 \int_{-1}^0 \int_0^\infty \int_0^\infty f_{\beta^*} \left( \frac{t_1 - w_1 \mathbf{n}_1}{\|t_1 - w_1 \mathbf{n}_1\|} s_1, \frac{t_2 - w_2 \mathbf{n}_2}{\|t_2 - w_2 \mathbf{n}_2\|} s_2 \right) s_1^{k_1} s_2^{k_2} ds_1, ds_2 \quad (\text{A.16})$$

$$\times \|t_1 - w_1 \mathbf{n}_1\|^{-k_1-1} \|t_2 - w_2 \mathbf{n}_2\|^{-k_2-1} f_\Delta(w_1, w_2) dw_1 dw_2 \quad (\text{A.17})$$

$$= \int_{-1}^0 \int_{-1}^0 f_\beta \left( \frac{t_1 - w_1 \mathbf{n}_1}{\|t_1 - w_1 \mathbf{n}_1\|}, \frac{t_2 - w_2 \mathbf{n}_2}{\|t_2 - w_2 \mathbf{n}_2\|} \right) \|t_1 - w_1 \mathbf{n}_1\|^{-k_1-1} \|t_2 - w_2 \mathbf{n}_2\|^{-k_2-1} f_\Delta(w_1, w_2) dw_1 dw_2, \quad (\text{A.18})$$

where the second equality follows from Tonelli's theorem, which is applicable because the integrand is jointly measurable, non-negative, and the integral equals  $f_\theta(t_1, t_2) < \infty$ . This establishes the first claim of the theorem.

The proof of the second claim is similar to that in Devroye (1981) and Carrasco and Florens (2010). Let  $\mathcal{D}$  be the set of probability density functions in  $L^2([-1, 0]^2)$ , i.e.  $\{\varphi \in L^2([-1, 0]^2) : \varphi \geq 0, \text{ and } \int \varphi(w_1, w_2) dw_1 dw_2 = 1\}$ . Let  $\mathcal{K}|_{\mathcal{D}}$  be the restriction of  $\mathcal{K}$  to  $\mathcal{D}$ . Below, we show that  $\mathcal{N}(\mathcal{K}|_{\mathcal{D}}) = \{0\}$ .

Define the map  $\mathcal{F} : L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}) \rightarrow L^2(\mathbb{R}^{k_1+1} \times \mathbb{R}^{k_2+1})$  pointwise by

$$\mathcal{F}f(x_1, x_2) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} f(u_1, u_2) e^{i(x'_1 u_1 + x'_2 u_2)} d\sigma(u_1, u_2). \quad (\text{A.19})$$

Note that  $(t_1, t_2, w_1, w_2) \mapsto K(t_1 - w_1 \mathbf{n}_1, t_2 - w_2 \mathbf{n}_2) f_\Delta(w_1, w_2) e^{i(x'_1 t_1 + x'_2 t_2)}$  is jointly measurable and  $|K(t_1 - w_1 \mathbf{n}_1, t_2 - w_2 \mathbf{n}_2) f_\Delta(w_1, w_2) e^{i(x'_1 t_1 + x'_2 t_2)}| \leq |K(t_1 - w_1 \mathbf{n}_1, t_2 - w_2 \mathbf{n}_2) f_\Delta(w_1, w_2)|$ . Further,

$$\begin{aligned} & \int_{-1}^0 \int_{-1}^0 \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} |K(t_1 - w_1 \mathbf{n}_1, t_2 - w_2 \mathbf{n}_2) f_\Delta(w_1, w_2)| d\sigma(t_1, t_2) dw_1 dw_2 \\ &= \|K\|_{L^1(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \|f_\Delta\|_{L^1([-1, 0]^2)} < \infty. \end{aligned}$$

By Tonelli's theorem and the change of variables with  $v_j = t_j - w_j \mathbf{n}_j$  and  $dv_j = dt_j$ , we obtain:

$$\mathcal{F}(\mathcal{K}|_{\mathcal{D}} f_\Delta)(x_1, x_2) = \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} \int_{-1}^0 \int_{-1}^0 K(t_1 - w_1 \mathbf{n}_1, t_2 - w_2 \mathbf{n}_2) f_\Delta(w_1, w_2) dw_1 dw_2 e^{i(x'_1 t_1 + x'_2 t_2)} d\sigma(t_1, t_2) \quad (\text{A.20})$$

$$= \int_{-1}^0 \int_{-1}^0 \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} K(v_1, v_2) e^{i(x'_1 (v_1 + w_1 \mathbf{n}_1) + x'_2 (v_2 + w_2 \mathbf{n}_2))} d\sigma(v_1, v_2) f_\Delta(w_1, w_2) dw_1 dw_2 \quad (\text{A.21})$$

$$= \Psi_K(x_1, x_2) \Psi_{f_\Delta}(x_1^{(1)}, x_2^{(1)}), \quad (\text{A.22})$$

where  $\Psi_{f_\Delta}$  is the characteristic function of  $f_\Delta$ . Suppose that there is another  $f' \in \mathcal{D}$  such that  $\mathcal{K}|_{\mathcal{D}}(f_\Delta - f') = 0$ . Then, by (A.22) and linearity of  $\mathcal{K}|_{\mathcal{D}}$ , this holds if and only if

$$\Psi_K(x_1, x_2) \{ \Psi_{f_\Delta}(x_1^{(1)}, x_2^{(1)}) - \Psi_{f'}(x_1^{(1)}, x_2^{(1)}) \} = 0, \quad (\text{A.23})$$

for all  $(x_1, x_2) \in \mathbb{R}^{k_1+1} \times \mathbb{R}^{k_2+1}$ . Let  $\mathcal{S} \equiv \{(x_1^{(1)}, x_2^{(1)}) \in \mathbb{R}^2 : \Psi_K(x_1, x_2) = 0\}$ . By (A.23), it must hold that  $\Psi_{f_\Delta}(e_1, e_2) = \Psi_{f'}(e_1, e_2)$  for all  $(e_1, e_2) \notin \mathcal{S}$ . By hypothesis,  $\mathcal{S}$  has measure 0. Therefore, for any  $\epsilon > 0$ , there is a sequence of open sets  $I_n \subset \mathbb{R}^2$  with positive Lebesgue measure such that  $\mathcal{S} \subset \bigcup_n I_n$  and  $\sum_n \text{Leb}(I_n) < \epsilon$ . Then, for any  $(x_1, x_2) \in \mathcal{S}$  and  $\epsilon > 0$ , we may find an open neighborhood  $I_\epsilon$  of  $(x_1, x_2)$  such that  $\Psi_{f_\Delta}(s_{1,\epsilon}, s_{2,\epsilon}) = \Psi_{f'}(s_{1,\epsilon}, s_{2,\epsilon})$  for any  $(s_{1,\epsilon}, s_{2,\epsilon}) \in I_\epsilon \setminus \{(x_1, x_2)\}$ . Now by letting  $\epsilon \downarrow 0$  and noting that the  $\Psi_{f_\Delta}$  and  $\Psi_{f'}$  are continuous (See Theorem 2.1.2 in Lukacs, 1960), it follows that they must also coincide at all points in  $\mathcal{S}$ . Since the characteristic functions determine their densities, it must hold that  $f_\Delta = f'$ . Thus, the second claim follows.  $\square$

## APPENDIX B: Proof of Theorem 4.1

This appendix introduces technical tools and auxiliary lemmas used to prove Theorem 4.1. Please see Groemer (1996) for more details.

Let  $\mathbb{R}^d$  be a Euclidean space. Let  $\Delta$  be the Laplace operator defined by  $\Delta \equiv \partial^2/\partial x_1^2 + \dots + \partial^2/\partial x_d^2$ . A *harmonic polynomial* is a polynomial defined on  $\mathbb{R}^d$  such that  $\Delta f = 0$ . A *spherical harmonic* of order  $n$  and dimension  $d$  is the restriction of a  $n$ -th order harmonic polynomial to  $\mathbb{S}^{d-1}$ .

Let  $H^{n,d}$  be the space of all spherical harmonics of order  $n$  and dimension  $d$ , and let  $h(n, d)$  denote the dimension of  $H^{n,d}$ , which is by Theorem 3.1.4 in Groemer (1996):

$$h(n, d) = \frac{2n + d - 2}{n + d - 2} \binom{n + d - 2}{d - 2}.$$

For each  $n_1, n_2 \geq 0$  and  $k_1, k_2 \geq 2$ , let  $\{\varphi_{n_1, l}\}_{l=1}^{h(n_1, k_1+1)}$  and  $\{\psi_{n_2, m}\}_{m=1}^{h(n_2, k_2+1)}$  be orthonormal bases of  $H^{n_1, k_1+1}$  and  $H^{n_2, k_2+1}$  respectively. Let  $\{\Phi_j, j = 1, \dots\}$  be a sequence such that for each  $n_1 \geq 0$  it contains  $h(n_1, k_1 + 1)$  orthonormal spherical harmonics  $\varphi_{n_1, l}, l = 1, \dots, h(n_1, k_1 + 1)$ . Similarly, let  $\{\Psi_j, j = 1, \dots\}$  be a sequence such that for each  $n_2 \geq 0$  it contains  $h(n_2, k_2 + 1)$  orthonormal spherical harmonics  $\psi_{n_2, m}, m = 1, \dots, h(n_2, k_2 + 1)$ . These sequences are called the *standard sequences* and form orthonormal bases of  $L^2(\mathbb{S}^{k_1})$  and  $L^2(\mathbb{S}^{k_2})$  respectively. We note that  $\{\Phi_j \Psi_k, j = 1, \dots, k = 1, \dots\}$  then forms an orthonormal basis in  $L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$ . (See for example Theorem II.10 in Reed and Simon, 1980.)

The *harmonic expansion* of any  $f \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  is given by

$$f(z_1, z_2) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \left( \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} f(\tilde{z}_1, \tilde{z}_2) \Phi_j(\tilde{z}_1) \Psi_k(\tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \right) \Phi_j(z_1) \Psi_k(z_2). \quad (\text{B.1})$$

Now, for each  $n_1, n_2 \geq 0$ ,  $z_1, \tilde{z}_1 \in \mathbb{S}^{k_1}$ , and  $z_2, \tilde{z}_2 \in \mathbb{S}^{k_2}$ , define

$$q_{n_1, k_1}(z_1, \tilde{z}_1) \equiv \sum_{l=1}^{h(n_1, k_1+1)} \varphi_{n_1, l}(z_1) \varphi_{n_1, l}(\tilde{z}_1) \quad (\text{B.2})$$

$$q_{n_2, k_2}(z_2, \tilde{z}_2) \equiv \sum_{m=1}^{h(n_2, k_2+1)} \psi_{n_2, m}(z_2) \psi_{n_2, m}(\tilde{z}_2) \quad (\text{B.3})$$

$$q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) \equiv \sum_{l=1}^{h(n_1, k_1+1)} \sum_{m=1}^{h(n_2, k_2+1)} \varphi_{n_1, l}(z_1) \psi_{n_2, m}(z_2) \varphi_{n_1, l}(\tilde{z}_1) \psi_{n_2, m}(\tilde{z}_2). \quad (\text{B.4})$$

Let  $Q_{n_1, n_2, k_1, k_2} : L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}) \rightarrow H^{n_1, k_1+1} \otimes H^{n_2, k_2+1}$  be the projection map defined by

$$(Q_{n_1, n_2, k_1, k_2} f)(z_1, z_2) \equiv \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) f(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2). \quad (\text{B.5})$$

An alternative way to write (B.1) is

$$f(z_1, z_2) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} (Q_{n_1, n_2, k_1, k_2} f)(z_1, z_2). \quad (\text{B.6})$$

As in the main text, we call this expansion the *condensed harmonic expansion*.

For all  $n \geq 0$ , the *Legendre polynomial*  $L_n^d$  of order  $n$  and dimension  $d$  is a polynomial on  $\mathbb{R}$ , which is uniquely defined by the following recurrence formula:

$$(n+d-2)L_{n+1}^d(x) - (2n+d-2)xL_n^d(x) + nL_{n-1}^d(x) = 0, \quad (\text{B.7})$$

where  $L_{-1}^d(x) = 0$  and  $L_0^d(x) = 1$  for all  $x$ . For  $d \geq 3$ , the *Gegenbauer polynomial*  $C_n^{\nu(d)}$  of order  $n$  and dimension  $\nu(d) = (d-2)/d$  is defined by  $C_n^{\nu(d)} \equiv \binom{n+d-3}{d-3} L_n^d$ .

**Lemma B.1.** *For every  $n_1, n_2 \geq 0$ ,  $z_1, \tilde{z}_1 \in \mathbb{S}^{k_1}$  and  $z_2, \tilde{z}_2 \in \mathbb{S}^{k_2}$ ,*

$$q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) = \frac{h(n_1, k_1+1)h(n_2, k_2+1)}{|\mathbb{S}^{k_1}| |\mathbb{S}^{k_2}|} L_{n_1}^{k_1+1}(z_1' \tilde{z}_1) L_{n_2}^{k_2+1}(z_2' \tilde{z}_2) \quad (\text{B.8})$$

$$= \frac{h(n_1, k_1+1)h(n_2, k_2+1)}{|\mathbb{S}^{k_1}| |\mathbb{S}^{k_2}|} \frac{C_{n_1}^{\nu(k_1+1)}(z_1' \tilde{z}_1) C_{n_2}^{\nu(k_2+1)}(z_2' \tilde{z}_2)}{C_{n_1}^{\nu(k_1+1)}(1) C_{n_2}^{\nu(k_2+1)}(1)}, \quad (\text{B.9})$$

where  $L_n^d$  are Legendre polynomials of order  $n$  and dimension  $d$ , and  $C_n^{\nu}$  are Gegenbauer polynomials of order  $n$  and  $\nu(d)$ .

Proof of Lemma B.1. One may rewrite  $q_{n_1, n_2, k_1, k_2}$  as

$$\begin{aligned} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) &= \sum_{l=1}^{h(n_1, k_1+1)} \varphi_{n_1, l}(z_1) \varphi_{n_1, l}(\tilde{z}_1) \sum_{m=1}^{h(n_2, k_2+1)} \psi_{n_2, m}(z_2) \psi_{n_2, m}(\tilde{z}_2) \\ &\equiv q_{n_1, k_1}(z_1, \tilde{z}_1) \times q_{n_2, k_2}(z_2, \tilde{z}_2). \end{aligned}$$

Now, Eq. (B.8) follows from Theorem 3.3.3 in Groemer (1996). Eq. (B.9) follows from Theorem 2.1 in GK.  $\square$

**Lemma B.2.** *For every  $n_1 > 0, n_2 > 0$ ,  $z_1 \in \mathbb{S}^{k_1}$  and  $z_2 \in \mathbb{S}^{k_2}$ ,*

$$\begin{aligned} \int_{\mathbb{S}^{k_1}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) d\sigma_{k_1}(\tilde{z}_1) &= 0, \quad \forall \tilde{z}_2 \in \mathbb{S}^{k_2} \\ \int_{\mathbb{S}^{k_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) d\sigma_{k_2}(\tilde{z}_2) &= 0, \quad \forall \tilde{z}_1 \in \mathbb{S}^{k_1}. \end{aligned}$$

Proof of Lemma B.2. For each  $z_1 \in \mathbb{S}^{k_1}$ ,  $\tilde{z}_1 \mapsto L_{n_1}^{k_1+1}(z_1' \tilde{z}_1)$  is a  $k_1 + 1$ -dimensional spherical harmonic of order  $n_1 > 0$  by Theorem 3.3.3 in Groemer (1996). Thus, by Corollary 3.2.2 in Groemer (1996).

$$\int_{\mathbb{S}^{k_1}} L_{n_1}^{k_1+1}(z_1' \tilde{z}_1) d\sigma_{k_1}(\tilde{z}_1) = 0. \quad (\text{B.10})$$

By (B.8) and (B.10), the first claim follows. The second claim can be proved in the same manner.  $\square$

**Lemma B.3.** *For every  $n_1 > 0, n_2 > 0$ ,  $z_1, \tilde{z}_1 \in \mathbb{S}^{k_1}$ , and  $z_2, \tilde{z}_2 \in \mathbb{S}^{k_2}$ ,*

$$\begin{aligned} q_{n_1, n_2, k_1, k_2}(z_1, z_2, -\tilde{z}_1, \tilde{z}_2) &= (-1)^{n_1} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) \\ q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, -\tilde{z}_2) &= (-1)^{n_2} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2). \end{aligned}$$

Proof of Lemma B.3. We first note that for each  $t$ ,  $C_n^\nu(-t) = (-1)^n C_n^\nu(t)$  (See p.32 in GK). This, together with (B.9), establishes the claim of the lemma.  $\square$

Proof of Theorem 4.1. The proof of this theorem is similar to that of Theorem 4.1 in GK. We note that  $R_{(1,1)}$  is bounded and hence square integrable with respect to  $\sigma$ . Hence, it has the condensed harmonic expansion:

$$R_{(1,1)}(z_1, z_2) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} (Q_{n_1, n_2, k_1, k_2} R_{(1,1)})(z_1, z_2). \quad (\text{B.11})$$

By Lemma B.3 and Eq. (2.6), the condensed harmonic expansion of  $R_{(1,1)}^{--}$  is then given by

$$R_{(1,1)}^{--} = \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} (Q_{2p_1+1, 2p_2+1, k_1, k_2} R_{(1,1)}). \quad (\text{B.12})$$

By Eq. (3.6), we may write  $Q_{n_1, n_2, k_1, k_2} R_{(1,1)}$  as

$$(Q_{n_1, n_2, k_1, k_2} R_{(1,1)})(z_1, z_2) = \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) r_{(1,1)}(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \quad (\text{B.13})$$

$$+ \int_{H_{n_1}^c \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) \{r_{(1,1)}(c_1, \tilde{z}_2) - r_{(1,1)}(-\tilde{z}_1, \tilde{z}_2)\} d\sigma(\tilde{z}_1, \tilde{z}_2) \quad (\text{B.14})$$

$$+ \int_{H_{n_1} \times H_{n_2}^c} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) \{r_{(1,1)}(\tilde{z}_1, c_2) - r_{(1,1)}(\tilde{z}_1, -\tilde{z}_2)\} d\sigma(\tilde{z}_1, \tilde{z}_2) \quad (\text{B.15})$$

$$+ \int_{H_{n_1}^c \times H_{n_2}^c} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) \{1 - r_{(1,1)}(-\tilde{z}_1, c_2) - r_{(1,1)}(c_1, -\tilde{z}_2) + r_{(1,1)}(-\tilde{z}_1, -\tilde{z}_2)\} d\sigma(\tilde{z}_1, \tilde{z}_2). \quad (\text{B.16})$$

By Eqs. (B.13)-(B.16), and Lemma B.3,

$$(Q_{n_1, n_2, k_1, k_2} R_{(1,1)})(z_1, z_2) = 4 \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) r_{(1,1)}(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \quad (\text{B.17})$$

$$- 2 \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) r_{(1,1)}(c_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \quad (\text{B.18})$$

$$- 2 \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) r_{(1,1)}(\tilde{z}_1, c_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \quad (\text{B.19})$$

$$+ \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2). \quad (\text{B.20})$$

Combining Eq. (B.17) and (B.20) yields

$$\begin{aligned} & \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) \{4r_{(1,1)}(\tilde{z}_1, \tilde{z}_2) + 1\} d\sigma(\tilde{z}_1, \tilde{z}_2) \\ &= \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) E \left[ \frac{4W + 1}{f_Z(\tilde{z}_1, \tilde{z}_2)} \middle| Z_1 = \tilde{z}_1, Z_2 = \tilde{z}_2 \right] f_Z(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \\ &= E \left[ \frac{4W + 1}{f_Z(Z_1, Z_2)} q_{n_1, n_2, k_1, k_2}(z_1, z_2, Z_1, Z_2) \right], \end{aligned} \quad (\text{B.21})$$

where the last equality follows from the law of iterated expectations. Eq. (B.18) can be written as

$$\begin{aligned} & -2 \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) r_{(1,1)}(c_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \\ &= - \int_{H_{n_1}} \frac{q_{n_1, k_1}(z_1, \tilde{z}_1)}{f_{Z_1}(\tilde{z}_1)} f_{Z_1}(\tilde{z}_1) d\sigma_{k_1}(\tilde{z}_1) \int_{H_{n_2}} \frac{q_{n_2, k_2}(z_2, \tilde{z}_2)}{f_{Z_2|Z_1}(\tilde{z}_2|c_1)} E[2W|Z_1 = c_1, Z_2 = \tilde{z}_2] f_{Z_2|Z_1}(\tilde{z}_2|c_1) d\sigma_{k_2}(\tilde{z}_2) \\ &= -E \left[ \frac{q_{n_1, k_1}(z_1, Z_1)}{f_{Z_1}(Z_1)} \right] \times E \left[ \frac{2W q_{n_2, k_2}(z_2, Z_2)}{f_{Z_2|Z_1}(Z_2|c_1)} \middle| Z_1 = c_1 \right]. \end{aligned} \quad (\text{B.22})$$



Similarly, Eq. (B.19) can be written as

$$\begin{aligned}
-2 \int_{H_{n_1} \times H_{n_2}} q_{n_1, n_2, k_1, k_2}(z_1, z_2, \tilde{z}_1, \tilde{z}_2) r_{(1,1)}(\tilde{z}_1, c_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \\
= -E \left[ \frac{q_{n_2, k_2}(z_2, Z_2)}{f_{Z_2}(Z_2)} \right] \times E \left[ \frac{2W q_{n_1, k_1}(z_1, Z_1)}{f_{Z_1|Z_2}(Z_1|c_2)} \Big| Z_2 = c_2 \right].
\end{aligned} \tag{B.23}$$

By (B.17)-(B.23), we obtain

$$(Q_{n_1, n_2, k_1, k_2} R_{(1,1)})(z_1, z_2) = E \left[ \frac{4W + 1}{f_Z(Z_1, Z_2)} q_{n_1, n_2, k_1, k_2}(z_1, z_2, Z_1, Z_2) \right] \tag{B.24}$$

$$- E \left[ \frac{q_{n_1, k_1}(z_1, Z_1)}{f_{Z_1}(Z_1)} \right] E \left[ \frac{2W q_{n_2, k_2}(z_2, Z_2)}{f_{Z_2|Z_1}(Z_2|c_1)} \Big| Z_1 = c_1 \right] \tag{B.25}$$

$$- E \left[ \frac{q_{n_2, k_2}(z_2, Z_2)}{f_{Z_2}(Z_2)} \right] E \left[ \frac{2W q_{n_1, k_1}(z_1, Z_1)}{f_{Z_1|Z_2}(Z_1|c_2)} \Big| Z_2 = c_2 \right]. \tag{B.26}$$

Therefore, (B.12) and (B.24)-(B.26) establish the claim of the Theorem.  $\square$

## APPENDIX C: Proof of Theorem 4.2

Throughout, we let  $\bar{c}$  denote a generic positive constant that may be different in different appearances. In what follows, we write  $\hat{f}_\theta^{--}$  as  $\hat{f}_\theta^{--} = \hat{A} - \hat{B}\hat{C} - \hat{D}\hat{E}$ , where

$$\begin{aligned}
\hat{A}(t_1, t_2) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{4W_i + 1}{\hat{f}_Z^a(Z_{1i}, Z_{2i})} \mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(\cdot, Z_{1i}))(t_1) \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2) \\
\hat{B}(t_1) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{\mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(\cdot, Z_{1i}))(t_1)}{\hat{f}_{Z_1}^a(Z_{1i})} \\
\hat{C}(t_2) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{2W_i \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2) K_{T_1}(c_1, Z_{1i})}{\hat{f}_{Z_2}^a(Z_{2i}|c_1) \hat{f}_{Z_1}^a(Z_{1i})} \\
\hat{D}(t_2) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{\mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2)}{\hat{f}_{Z_2}^a(Z_{2i})} \\
\hat{E}(t_1) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{2W_i \mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(\cdot, Z_{1i}))(t_1) K_{T_2}(c_2, Z_{2i})}{\hat{f}_{Z_1}^a(Z_{1i}|c_2) \hat{f}_{Z_2}^a(Z_{2i})}.
\end{aligned}$$

Further, we let their population counterparts be defined by

$$\begin{aligned}
A(t_1, t_2) &\equiv E \left[ \frac{4W + 1}{f_Z(Z_1, Z_2)} \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} (\mathcal{H}_{\mathbb{S}^{k_1}} \otimes \mathcal{H}_{\mathbb{S}^{k_2}})^{-1} q_{2p_1+1, 2p_2+1, k_1, k_2}(t_1, t_2, Z_1, Z_2) \right] \\
B(t_1) &\equiv E \left[ \frac{\sum_{p_1=0}^{\infty} \mathcal{H}_{\mathbb{S}^{k_1}}^{-1} q_{2p_1+1, k_1}(t_1, Z_1)}{f_{Z_1}(Z_1)} \right] \\
C(t_2) &\equiv E \left[ \frac{2W \sum_{p_2=0}^{\infty} \mathcal{H}_{\mathbb{S}^{k_2}}^{-1} q_{2p_2+1, k_2}(t_2, Z_2)}{f_{Z_2|Z_1}(Z_2|c_1)} \middle| Z_1 = c_1 \right] \\
D(t_2) &\equiv E \left[ \frac{\sum_{p_1=0}^{\infty} \mathcal{H}_{\mathbb{S}^{k_2}}^{-1} q_{2p_2+1, k_2}(t_2, Z_2)}{f_{Z_2}(Z_2)} \right] \\
E(t_1) &\equiv E \left[ \frac{2W \sum_{p_1=0}^{\infty} \mathcal{H}_{\mathbb{S}^{k_1}}^{-1} q_{2p_1+1, k_1}(t_1, Z_1)}{f_{Z_1|Z_2}(Z_1|c_2)} \middle| Z_2 = c_2 \right].
\end{aligned}$$

The assumptions on the smoothed projection kernel are the same as GK's. For easy reference, we give them below.

**Assumption C.1.** (i) For each  $j$  and  $\|K_{T_j}(z_j, \cdot)\|_{L^1}$  is uniformly bounded in  $T_j$ .

(ii) For each  $j$ , there exist constants  $\bar{c} > 0$  and  $\alpha_j > 0$ , such that for all  $z_j, \tilde{z}_j, \tilde{z}'_j \in \mathbb{S}^{k_j}$ ,

$$|K_{T_j}(z_j, \tilde{z}_j) - K_{T_j}(z_j, \tilde{z}'_j)| \leq \bar{c} \|\tilde{z}_j - \tilde{z}'_j\|^{\alpha_j}.$$

(iii) For each  $j$  and  $s_j > 0$ , there exists a constant  $\bar{c} > 0$  such that

$$\left\| f(\cdot) - \int_{\mathbb{S}^{k_j}} K_{T_j}(\cdot, \tilde{z}_j) f(\tilde{z}_j) d\sigma_{k_j}(\tilde{z}_j) \right\|_{L^2} \leq \bar{c} \|f\|_{\mathbb{W}_2^{s_j} T_j^{-s_j}}, \quad \text{for all } f \in \mathbb{W}_2^{s_j}(\mathbb{S}^{k_j}).$$

(iv)  $\chi_j(\cdot, T_j)$  takes values in  $[0, 1]$  and is such that there exists  $\bar{c} > 0$  such that for all  $0 \leq n \leq \lfloor T_j/2 \rfloor$ ,  $\chi_j(n, T_j) \geq \bar{c}$ .

Our assumptions the joint, conditional, and marginal densities of  $Z$  and their estimators are the following.

**Assumption C.2.** (i)  $f_Z$  is bounded on  $H_1 \times H_2$ .

(ii) There exists a constant  $r > 0$  such that

$$\begin{aligned}\sigma\left(\{(z_1, z_2) : 0 < f_Z(z_1, z_2) < (\ln N)^{-r}\}\right) &= o\left(\left(\frac{N}{(\ln N)^{2r}}\right)^{-\frac{\rho_1+\rho_2}{2(\rho_1+\rho_2+2)}}\right), \\ \sigma_{k_j}\left(\{z_j : 0 < f_{Z_j}(z_j) < (\ln N)^{-r}\}\right) &= o\left(\left(\frac{N}{(\ln N)^{2r}}\right)^{-\frac{\rho_j}{2(\rho_1+\rho_2+2)}}\right), \text{ for } j = 1, 2, \\ \int_{\bar{S}_{jN}} f_{Z_j}(z_j)^{-1} d\sigma_{k_j}(z_j) &= o\left(\left(\frac{N}{(\ln N)^{2r}}\right)^{-\frac{\rho_j}{2(\rho_1+\rho_2+2)}}\right), \text{ for } j = 1, 2,\end{aligned}$$

where  $\bar{S}_{jN} \equiv \{z_j : 0 < f_{Z_j}(z_j) < (\ln N)^{-r}\}$ .

(iii) Further,

$$\begin{aligned}\left(\frac{N}{(\ln N)^{2r}}\right)^{\frac{(k_1+1)/s_1+(k_2+1)/s_2}{2(\rho_1+\rho_2+2)}} (\ln N)^r \max_{i=1, \dots, N} \left| \frac{f_Z^a(Z_{1,i}, Z_{2,i})}{\hat{f}_Z^a(Z_{1,i}, Z_{2,i})} - 1 \right| &= o_p(1) \\ \left(\frac{N}{(\ln N)^{2r}}\right)^{\frac{(k_j+1)/s_j}{2(\rho_1+\rho_2+2)}} (\ln N)^r \max_{i=1, \dots, N} \left| \frac{f_{Z_j}^a(Z_{j,i})}{\hat{f}_{Z_j}^a(Z_{j,i})} - 1 \right| &= o_p(1) \\ \left(\frac{N}{(\ln N)^{2r}}\right)^{\frac{(k_j+1)/s_j}{2(\rho_1+\rho_2+2)}} (\ln N)^{2r} \max_{i=1, \dots, N} \left| \frac{f_{Z_j|Z_{-j}}^a(Z_{ji}|c_{-j}) f_{Z_{-j}}^a(Z_{-ji})}{\hat{f}_{Z_j|Z_{-j}}^a(Z_{ji}|c_{-j}) \hat{f}_{Z_{-j}}^a(Z_{-ji})} - 1 \right| &= o_p(1).\end{aligned}$$

The conditions above are similar to those assumed in GK and hold for a reasonable class of distributions.

**Lemma C.1.** *Let  $k_1, k_2 \geq 2$ . Suppose Assumptions 4.1 and C.1 hold. Then,*

$$\left\| f(\cdot) - \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} K_{T_1}(\cdot, \tilde{z}_1) K_{T_2}(\cdot, \tilde{z}_2) f(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \right\|_{L^2} \leq \bar{c}(T_1^{-s_1} \vee T_2^{-s_2}) \|f\|_{\mathbb{W}_2^{s_1} \otimes \mathbb{W}_2^{s_2}},$$

for all  $f \in \mathbb{W}_2^{s_1}(\mathbb{S}^{k_1}) \times \mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})$ .

Proof of Lemma C.1. By the triangle inequality,

$$\begin{aligned}& \left\| f(\cdot, \cdot) - \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} K_{T_1}(\cdot, \tilde{z}_1) K_{T_2}(\cdot, \tilde{z}_2) f(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2 \\ & \leq \left\| f(\cdot, \cdot) - \int_{\mathbb{S}^{k_1}} K_{T_1}(\cdot, \tilde{z}_1) f(\tilde{z}_1, \cdot) d\sigma_{k_1}(\tilde{z}_1) \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2 \\ & \quad + \left\| \int_{\mathbb{S}^{k_1}} K_{T_1}(\cdot, \tilde{z}_1) f(\tilde{z}_1, \cdot) d\sigma_{k_1}(\tilde{z}_1) - \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} K_{T_1}(z_1, \tilde{z}_1) K_{T_2}(z_2, \tilde{z}_2) f(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2.\end{aligned}$$

Note that

$$\begin{aligned}
& \left\| f(\cdot, \cdot) - \int_{\mathbb{S}^{k_1}} K_{T_1}(\cdot, \tilde{z}_1) f(\tilde{z}_1, \cdot) d\sigma_{k_1}(\tilde{z}_1) \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2 \\
&= \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} \left( f(z_1, z_2) - \int_{\mathbb{S}^{k_1}} K_{T_1}(z_1, \tilde{z}_1) f(\tilde{z}_1, z_2) d\sigma_{k_1}(\tilde{z}_1) \right)^2 d\sigma(z_1, z_2) \\
&= \int_{\mathbb{S}^{k_2}} \left\| f(\cdot, z_2) - \int_{\mathbb{S}^{k_1}} K_{T_1}(\cdot, \tilde{z}_1) f(\tilde{z}_1, z_2) d\sigma_{k_1}(\tilde{z}_1) \right\|_{L^2(\mathbb{S}^{k_1})}^2 d\sigma_{k_2}(z_2) \\
&\leq (\bar{c}T_1^{-s_1})^2 \int_{\mathbb{S}^{k_2}} \|f(\cdot, z_2)\|_{\mathbb{W}_2^{s_1}}^2 d\sigma_{k_2}(z_2) \\
&\leq (\bar{c}T_1^{-s_1})^2 \|f\|_{\mathbb{W}_2^{s_1} \otimes \mathbb{W}_2^{s_2}}^2,
\end{aligned}$$

where the last inequality follows from the following result.

$$\begin{aligned}
\int_{\mathbb{S}^{k_2}} \|f(\cdot, z_2)\|_{\mathbb{W}_2^{s_1}}^2 d\sigma_{k_2}(z_2) &= \int_{\mathbb{S}^{k_2}} \sum_{n_1=0}^{\infty} (1 + \zeta_{n_1, k_1})^{s_1} \|Q_{n_1, k_1} f(\cdot, z_2)\|_{L^2(\mathbb{S}^{k_1})}^2 d\sigma_{k_2}(z_2) \\
&= \sum_{n_1=0}^{\infty} (1 + \zeta_{n_1, k_1})^{s_1} \int_{\mathbb{S}^{k_2}} \|Q_{n_1, k_1} f(\cdot, z_2)\|_{L^2(\mathbb{S}^{k_1})}^2 d\sigma_{k_2}(z_2) \\
&= \sum_{n_1=0}^{\infty} (1 + \zeta_{n_1, k_1})^{s_1} \int_{\mathbb{S}^{k_2}} \left\| \sum_{n_2=0}^{\infty} Q_{n_2, k_2} Q_{n_1, k_1} f(\cdot, z_2) \right\|_{L^2(\mathbb{S}^{k_1})}^2 d\sigma_{k_2}(z_2) \\
&= \sum_{n_1=0}^{\infty} (1 + \zeta_{n_1, k_1})^{s_1} \int_{\mathbb{S}^{k_1}} \left\| \sum_{n_2=0}^{\infty} Q_{n_2, k_2} Q_{n_1, k_1} f(\cdot, z_2) \right\|_{L^2(\mathbb{S}^{k_2})}^2 d\sigma_{k_1}(z_1) \\
&= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} (1 + \zeta_{n_1, k_1})^{s_1} \int_{\mathbb{S}^{k_1}} \|Q_{n_2, k_2} Q_{n_1, k_1} f(\cdot, z_2)\|_{L^2(\mathbb{S}^{k_2})}^2 d\sigma_{k_1}(z_1) \\
&= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} (1 + \zeta_{n_1, k_1})^{s_1} \|Q_{n_1, k_1, n_2, k_2} f\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2 \leq \|f\|_{\mathbb{W}_2^{s_1} \otimes \mathbb{W}_2^{s_2}}^2,
\end{aligned}$$

where the second equality follows from the monotone convergence theorem, the third equality follows from  $Q_{n_1, k_1} f(\cdot, z_2) \in L^2(\mathbb{S}^{k_2})$ , the fourth equality follows from Tonelli's theorem, the fifth equality follows from Theorem 3.2.10 in Groemer (1996) and the monotone convergence theorem, and the last inequality follows from the fact that  $(1 + \zeta_{n_2, k_2}) \geq 1$  for all  $n_2 \geq 0$ .

Similarly,

$$\begin{aligned}
& \left\| \int_{\mathbb{S}^{k_1}} K_{T_1}(\cdot, \tilde{z}_1) f(\tilde{z}_1, \cdot) d\sigma_{k_1}(\tilde{z}_1) - \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} K_{T_1}(\cdot, \tilde{z}_1) K_{T_2}(\cdot, \tilde{z}_2) f(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2 \\
&= \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} \left( \int_{\mathbb{S}^{k_1}} K_{T_1}(z_1, \tilde{z}_1) f(\tilde{z}_1, z_2) d\sigma_{k_1}(\tilde{z}_1) - \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} K_{T_1}(z_1, \tilde{z}_1) K_{T_2}(z_2, \tilde{z}_2) f(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \right)^2 d\sigma(z_1, z_2) \\
&= \int_{\mathbb{S}^{k_1}} \left\| \int_{\mathbb{S}^{k_1}} K_{T_1}(z_1, \tilde{z}_1) f(\tilde{z}_1, \cdot) d\sigma_{k_1}(\tilde{z}_1) - \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} K_{T_1}(z_1, \tilde{z}_1) K_{T_2}(\cdot, \tilde{z}_2) f(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \right\|_{L^2(\mathbb{S}^{k_2})}^2 d\sigma_{k_1}(z_1) \\
&\leq (\bar{c} T_2^{-s_2}) \int_{\mathbb{S}^{k_1}} \left\| \int_{\mathbb{S}^{k_1}} K_{T_1}(z_1, \tilde{z}_1) f(\tilde{z}_1, \cdot) d\sigma_{k_1}(\tilde{z}_1) \right\|_{\mathbb{W}_2^{s_2}}^2 d\sigma_{k_1}(z_1) \\
&\leq (\bar{c} T_2^{-s_2}) \left\| \int_{\mathbb{S}^{k_1}} K_{T_1}(\cdot, \tilde{z}_1) f(\tilde{z}_1, \cdot) d\sigma_{k_1}(\tilde{z}_1) \right\|_{\mathbb{W}_2^{s_1} \otimes \mathbb{W}_2^{s_2}}^2 \\
&\leq (\bar{c} T_2^{-s_2}) \|f\|_{\mathbb{W}_2^{s_1} \otimes \mathbb{W}_2^{s_2}}^2,
\end{aligned}$$

where the last inequality follows from the following result.

$$\begin{aligned}
\int_{\mathbb{S}^{k_1}} K_{T_1}(z_1, \tilde{z}_1) f(\tilde{z}_1, z_2) d\sigma_{k_1}(\tilde{z}_1) &= \sum_{n_1=0}^{T_1} \chi(n_1, T_1) Q_{n_1, k_1} f(z_1, z_2) \\
&= \sum_{n_1=0}^{T_1} \sum_{n_2=0}^{\infty} \chi(n_1, T_1) Q_{n_1, k_1, n_2, k_2} f(z_1, z_2).
\end{aligned}$$

Hence,

$$\begin{aligned}
& \left\| \int_{\mathbb{S}^{k_1}} K_{T_1}(\cdot, \tilde{z}_1) f(\tilde{z}_1, \cdot) d\sigma_{k_1}(\tilde{z}_1) \right\|_{\mathbb{W}_2^{s_1} \otimes \mathbb{W}_2^{s_2}}^2 \\
&= \sum_{n_1=0}^{T_1} \sum_{n_2=0}^{\infty} \chi(n_1, T_1) (1 + \zeta_{n_1, k_1})^{s_1} (1 + \zeta_{n_2, k_2})^{s_2} \|Q_{n_1, k_1, n_2, k_2} f\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2 \leq \|f\|_{\mathbb{W}_2^{s_1} \otimes \mathbb{W}_2^{s_2}}^2,
\end{aligned}$$

where the last inequality holds because  $\chi(n_1, T_1) \leq 1$  for all  $n_1$  by Assumption C.1.  $\square$

**Lemma C.2.** *Let  $\hat{f}_\theta^-$  be given as in (4.8) and let  $q \geq 1$ . Then,  $\|f_\theta - \hat{f}_\theta\|_{L^q(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^q \leq \bar{c} \|f_\theta^- - \hat{f}_\theta^-\|_{L^q(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^q$  for some  $\bar{c} > 0$ .*

Proof of Lemma C.2. The proof is very similar to that of Theorem 5.1 in GK (See p.39). Thus, it is omitted.  $\square$

The following lemma is an analog of Theorem 3.2 in GK.

**Lemma C.3.** *For every  $k_1, k_2 \geq 2$ , there exists a positive constant  $B(k_1, k_2)$  such that for any function  $s \in L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$  of the form  $s = s_1 s_2$  with  $s_1 \in \bigoplus_{p_1=0}^{T_1} H^{2p_1+1, k_1}$  and  $s_2 \in \bigoplus_{p_2=0}^{T_2} H^{2p_2+1, k_2}$ , it holds that*

$$\|(\mathcal{H}_{\mathbb{S}^{k_1}} \otimes \mathcal{H}_{\mathbb{S}^{k_2}})^{-1} s\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \leq B(k_1, k_2) T_1^{(k_1+1)/2} T_2^{(k_2+1)/2} \|s\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}.$$

Proof of Lemma C.3. As in the proof of Theorem 3.2 in GK, we first write  $(\mathcal{H}_{\mathbb{S}^{k_1}} \otimes \mathcal{H}_{\mathbb{S}^{k_2}})^{-1}$  as a combinations of unbounded operators with non-positive eigenvalues. Define

$$\begin{aligned}\mathcal{P}_1 s &\equiv \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} \frac{1}{\lambda_1(4p_1+3, k_1)\lambda_2(4p_2+1, k_2)} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} q_{4p_1+3, 4p_2+1, k_1, k_2}(\cdot, \cdot, \tilde{z}_1, \tilde{z}_2) s(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \\ \mathcal{P}_2 s &\equiv \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} \frac{1}{\lambda_1(4p_1+1, k_1)\lambda_2(4p_2+3, k_2)} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} q_{4p_1+1, 4p_2+3, k_1, k_2}(\cdot, \cdot, \tilde{z}_1, \tilde{z}_2) s(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \\ \mathcal{P}_3 s &\equiv \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} \frac{1}{\lambda_1(4p_1+1, k_1)\lambda_2(4p_2+1, k_2)} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} q_{4p_1+1, 4p_2+1, k_1, k_2}(\cdot, \cdot, \tilde{z}_1, \tilde{z}_2) s(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2) \\ \mathcal{P}_4 s &\equiv \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} \frac{1}{\lambda_1(4p_1+3, k_1)\lambda_2(4p_2+3, k_2)} \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} q_{4p_1+3, 4p_2+3, k_1, k_2}(\cdot, \cdot, \tilde{z}_1, \tilde{z}_2) s(\tilde{z}_1, \tilde{z}_2) d\sigma(\tilde{z}_1, \tilde{z}_2).\end{aligned}$$

We then write  $(\mathcal{H}_{\mathbb{S}^{k_1}} \otimes \mathcal{H}_{\mathbb{S}^{k_2}})^{-1} = \mathcal{P}_1 + \mathcal{P}_2 - \mathcal{P}_3 - \mathcal{P}_4$ . By Theorem 3.2 in Ditzian (1998) and the triangle inequality, there exists a constant  $B(k_1, k_2)$  such that for all  $s \in (\bigoplus_{p_1=0}^{T_1} H^{2p_1+1, k_1}) \otimes (\bigoplus_{p_2=0}^{T_2} H^{2p_2+1, k_2})$ ,

$$\|(\mathcal{H}_{\mathbb{S}^{k_1}} \otimes \mathcal{H}_{\mathbb{S}^{k_2}})^{-1} s\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \leq \frac{B(k_1, k_2)}{\lambda(2T_1+1, k_1)\lambda(2T_2+1, k_2)} \|s\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}.$$

The claim of the lemma then follows from Eq. (9.11) in GK.  $\square$

**Lemma C.4.** *Suppose Assumption 4.1 holds. Then,  $A \in \mathbb{W}_2^{s_1}(\mathbb{S}^{k_1}) \otimes \mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})$ ,  $B, D \in \mathbb{W}_2^{s_1}(\mathbb{S}^{k_1})$ , and  $C, E \in \mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})$ .*

Proof of Lemma C.4. Note that  $A, BC$ , and  $DE$  are in  $L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})$ . Since  $f_{\theta}^{--} = A + BC + DE$ , the Sobolev norm of  $f_{\theta}^{--}$  being finite implies that the same must be true for  $A, BC$ , and  $DE$ . Hence,  $A \in \mathbb{W}_2^{s_1}(\mathbb{S}^{k_1}) \otimes \mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})$ . Since  $B$  only depends on  $t_1$  but not on  $t_2$ , and  $C$  only depends on  $t_2$  but not on  $t_1$ ,  $\|BC\|_{\mathbb{W}_2^{s_1}(\mathbb{S}^{k_1}) \otimes \mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})} = \|B\|_{\mathbb{W}_2^{s_1}(\mathbb{S}^{k_1})} \|C\|_{\mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})} < \infty$ . A similar result holds for  $DE$ . Therefore,  $B, D \in \mathbb{W}_2^{s_1}(\mathbb{S}^{k_1})$ , and  $C, E \in \mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})$ .  $\square$

Proof of Theorem 4.2. First, by Lemma C.2,  $\|f_{\theta} - \hat{f}_{\theta}\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2 \leq \bar{c} \|f_{\theta}^{--} - \hat{f}_{\theta}^{--}\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2$  for some  $\bar{c} > 0$ . Hence, it suffices to derive  $\|f_{\theta}^{--} - \hat{f}_{\theta}^{--}\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2$ . Notice that

$$\|f_{\theta}^{--} - \hat{f}_{\theta}^{--}\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \leq \|\hat{A} - A\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} + \|\hat{B}\|_{L^2(\mathbb{S}^{k_1})} \|\hat{C} - C\|_{L^2(\mathbb{S}^{k_2})} \quad (\text{C.1})$$

$$+ \|B - \hat{B}\|_{L^2(\mathbb{S}^{k_1})} \|C\|_{L^2(\mathbb{S}^{k_2})} + \|\hat{D}\|_{L^2(\mathbb{S}^{k_2})} \|\hat{E} - E\|_{L^2(\mathbb{S}^{k_1})} + \|D - \hat{D}\|_{L^2(\mathbb{S}^{k_2})} \|E\|_{L^2(\mathbb{S}^{k_1})}. \quad (\text{C.2})$$

Note that  $L^2$ -norms of  $B, C, D$ , and  $E$  are finite by Lemma C.4. Further, as we show below,  $\|\hat{B}\|_{L^2(\mathbb{S}^{k_1})} = O_p(1)$  and  $\|\hat{D}\|_{L^2(\mathbb{S}^{k_2})} = O_p(1)$  since they converge to  $\|B\|_{L^2(\mathbb{S}^{k_1})} < \infty$  and  $\|D\|_{L^2(\mathbb{S}^{k_2})} < \infty$  respectively.

We now work with  $\|\hat{A} - A\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}$  in (C.1). First, we define two hypothetical estimators:  $\hat{A}^{IT}$

the ‘‘infeasible trimmed estimator’’ and  $\hat{A}^I$  the ‘‘infeasible estimator’’ defined by

$$\begin{aligned}\hat{A}^{IT}(t_1, t_2) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{4W_i + 1}{f_Z^a(Z_{1i}, Z_{2i})} \mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(\cdot, Z_{1i}))(t_1) \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2) \\ \hat{A}^I(t_1, t_2) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{4W_i + 1}{f_Z(Z_{1i}, Z_{2i})} \mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(\cdot, Z_{1i}))(t_1) \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2)\end{aligned}$$

Now write

$$\hat{A} - A = (\hat{A} - \hat{A}^{IT}) - (\hat{A}^{IT} - E[\hat{A}^{IT}]) - (E[\hat{A}^{IT}] - E[\hat{A}^I]) - (E[\hat{A}^I] - A). \quad (\text{C.3})$$

The first term is the stochastic component due to plug-in; the second term is the stochastic component of the infeasible trimmed estimator; the third term is trimming bias; and the fourth term is approximation bias.

For  $\hat{A} - \hat{A}^{IT}$ , by Lemma C.3 and the triangle inequality, it follows that:

$$\|\hat{A} - \hat{A}^{IT}\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \quad (\text{C.4})$$

$$= \left\| (\mathcal{H}_1 \otimes \mathcal{H}_2)^{-1} \frac{1}{N} \sum_{i=1}^N \frac{4W_i + 1}{f_Z^a(Z_{1i}, Z_{2i})} K_{T_1}^-(z_1, Z_{1i}) K_{T_2}^-(z_2, Z_{2i}) \left( \frac{f_Z^a(Z_{1i}, Z_{2i})}{\hat{f}_Z^a(Z_{1i}, Z_{2i})} - 1 \right) \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \quad (\text{C.5})$$

$$\leq \bar{c} T_1^{(k_1+1)/2} T_2^{(k_2+1)/2} (\ln N)^r \left\| \frac{1}{N} \sum_{i=1}^N K_{T_1}^-(z_1, Z_{1i}) K_{T_2}^-(z_2, Z_{2i}) \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \max_{i=1, \dots, N} \left| \frac{f_Z^a(Z_{1i}, Z_{2i})}{\hat{f}_Z^a(Z_{1i}, Z_{2i})} - 1 \right|. \quad (\text{C.6})$$

We now decompose the  $L^2$ -norm in (C.6) as  $\|H_1\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} + \|H_2\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}$ , where

$$H_1(z_1, z_2) \equiv \frac{1}{N} \sum_{i=1}^N |K_{T_1}^-(z_1, Z_{1i}) K_{T_2}^-(z_2, Z_{2i})| - E|K_{T_1}^-(z_1, Z_{1i}) K_{T_2}^-(z_2, Z_{2i})|$$

$$H_2(z_1, z_2) \equiv E|K_{T_1}^-(z_1, Z_{1i}) K_{T_2}^-(z_2, Z_{2i})|.$$

Now, we calculate the convergence rate of  $H_1$ . First, consider  $E[\|H_1\|_{L^2}^2]$ . Note that

$$E[\|H_1\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}^2] = \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} E[H_1(z_1, z_2)^2] d\sigma(z_1, z_2).$$

Further,

$$\begin{aligned}E[H_1(z_1, z_2)^2] &\leq \frac{1}{N} E \left( |K_{T_1}^-(z_1, Z_{1i}) K_{T_2}^-(z_2, Z_{2i})|^2 \right) \\ &\leq \frac{\bar{c}}{N} \|K_{T_1}^-(z_1, \cdot)\|_{L^2(\mathbb{S}^{k_1})}^2 \|K_{T_2}^-(z_2, \cdot)\|_{L^2(\mathbb{S}^{k_2})}^2 \leq \frac{\bar{c}}{N} T_1^{k_1} T_2^{k_2},\end{aligned}$$

for some constant  $\bar{c} > 0$ , where the second inequality follows from Assumption C.2, and the third

inequality follows from (9.15) and Lemma 9.2 in GK. Hence, it holds that

$$T_1^{\frac{k_1+1}{2}} T_2^{\frac{k_2+1}{2}} (\ln N)^r \|H_1\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} = O_p(N^{-1/2} (\ln N)^r T_1^{\frac{2k_1+1}{2}} T_2^{\frac{2k_2+1}{2}}). \quad (\text{C.7})$$

Regarding  $H_2$ , by Assumptions C.1 (i) and C.2,

$$\|H_2\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \leq \bar{c} \|K_{T_1}^-(\cdot, \cdot) K_{T_2}^-(\cdot, \cdot)\|_{L^1(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \|L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})\| \quad (\text{C.8})$$

$$\leq \bar{c} \sigma(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})^{1/2} \|K_{T_1}^-(\cdot, \cdot) K_{T_2}^-(\cdot, \cdot)\|_{L^1(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} = O(1), \quad (\text{C.9})$$

where we used the fact that  $(z_1, z_2) \mapsto \|K_{T_1}^-(z_1, \cdot) K_{T_2}^-(z_2, \cdot)\|_{L^1(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}$  is a constant map. By (C.6), (C.7), and (C.9), it follows that

$$\|\hat{A} - \hat{A}^{IT}\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \leq \bar{c} \left\{ O_p(N^{-1/2} (\ln N)^r T_1^{\frac{2k_1+1}{2}} T_2^{\frac{2k_2+1}{2}}) + O_p((\ln N)^r T_1^{\frac{k_1+1}{2}} T_2^{\frac{k_2+1}{2}}) \right\} \quad (\text{C.10})$$

$$\times \max_{i=1, \dots, N} \left| \frac{f_Z^a(Z_{1,i}, Z_{2,i})}{\hat{f}_Z^a(Z_{1,i}, Z_{2,i})} - 1 \right|. \quad (\text{C.11})$$

We now turn to  $\|\hat{A}^{IT} - E[\hat{A}^{IT}]\|_{L^2}$ . By Lemma C.3, it follows that

$$\|\hat{A}^{IT} - E[\hat{A}^{IT}]\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \leq \bar{c} T_1^{(k_1+1)/2} T_2^{(k_2+1)/2} (\ln N)^r \quad (\text{C.12})$$

$$\times \left\| \frac{1}{N} \sum_{i=1}^N K_{T_1}^-(z_1, Z_{1i}) K_{T_2}^-(z_2, Z_{2i}) - E[K_{T_1}^-(z_1, Z_{1i}) K_{T_2}^-(z_2, Z_{2i})] \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \quad (\text{C.13})$$

$$= O_p(N^{-1/2} (\ln N)^r T_1^{\frac{2k_1+1}{2}} T_2^{\frac{2k_2+1}{2}}), \quad (\text{C.14})$$

where the last inequality follows from (C.7).

Now we consider the trimming bias:  $\|E[\hat{A}^{IT}] - E[\hat{A}^I]\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}$ . Let  $S_N \equiv \{(z_1, z_2) : 0 < f_Z(z_1, z_2) < (\ln N)^{-r}\}$  and note that  $|f_Z(z_1, z_2)(\ln N)^r - 1| \leq 1$  on this set. We may then write

$$\begin{aligned} E[\hat{A}^{IT}] - E[\hat{A}^I] &= \int_{S_N} E[4W_i + 1 | Z_1 = z_1, Z_2 = z_2] \\ &\quad \times \mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(z_1, \cdot))(t_1) \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(z_2, \cdot))(t_2) (f_Z(z_1, z_2)(\ln N)^r - 1) d\sigma(z_1, z_2). \end{aligned}$$

By Proposition 2.2 in GK (applied twice), Lemma C.3, and by the fact that  $|f_Z(z_1, z_2)(\ln N)^r - 1| \leq 1$  on  $S_N$ , it follows that

$$\begin{aligned} &\|E[\hat{A}^{IT}] - E[\hat{A}^I]\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \\ &\leq 5 \|\mathcal{H}_{\mathbb{S}^{k_1}}^{-1}(K_{T_1}^-(z_1, \cdot))(t_1)\|_{L^2(\mathbb{S}^{k_1})} \|\mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(z_2, \cdot))(t_2)\|_{L^2(\mathbb{S}^{k_2})} \sigma(S_N) \leq \bar{c} T_1^{\frac{k_1+1}{2}} T_2^{\frac{k_2+1}{2}} \sigma(S_N), \end{aligned} \quad (\text{C.15})$$

for some  $\bar{c} > 0$ . We note that, under the choice of  $T_1$  and  $T_2$  in (4.9),  $\|E[\hat{A}^{IT}] - E[\hat{A}^I]\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}$  vanishes faster than other terms by Assumption C.2 (ii).



For the approximation bias, we note that  $A \in \mathbb{W}_2^{s_1}(\mathbb{S}^{k_1}) \otimes \mathbb{W}_2^{s_2}(\mathbb{S}^{k_2})$  by Assumption 4.1 and Lemma C.4. Hence, Lemma C.1 ensures

$$\begin{aligned} & \|E[\hat{A}^I] - A\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \\ &= \left\| \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} A(\tilde{t}_1, \tilde{t}_2) K_{T_1}^-(t_1, \tilde{t}_1) K_{T_1}^-(t_2, \tilde{t}_2) d\sigma(\tilde{t}_1, \tilde{t}_2) - A(t_1, t_2) \right\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} \leq \bar{c}(T^{-s_1} \vee T^{-s_2}), \end{aligned} \quad (\text{C.16})$$

for some  $\bar{c} > 0$  by Assumption C.1 (iii). Now, by (C.11), (C.14), (C.15), and (C.16), we have

$$\|\hat{A} - A\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})} = \left\{ O_p(N^{-1/2}(\ln N)^r T_1^{\frac{2k_1+1}{2}} T_2^{\frac{2k_2+1}{2}}) + O_p((\ln N)^r T_1^{\frac{k_1+1}{2}} T_2^{\frac{k_2+1}{2}}) \right\} \quad (\text{C.17})$$

$$\times \max_{i=1, \dots, N} \left| \frac{f_Z^a(Z_{1,i}, Z_{2,i})}{\hat{f}_Z^a(Z_{1,i}, Z_{2,i})} - 1 \right| \quad (\text{C.18})$$

$$+ O_p(N^{-1/2}(\ln N)^r T_1^{\frac{2k_1+1}{2}} T_2^{\frac{2k_2+1}{2}}) \quad (\text{C.19})$$

$$+ O(T_1^{\frac{k_1+1}{2}} T_2^{\frac{k_2+1}{2}} \sigma(S_N)) + O(T_1^{-s_1} \vee T_2^{-s_2}). \quad (\text{C.20})$$

A similar argument ensures that the stochastic orders of  $\|\hat{B} - B\|_{L^2(\mathbb{S}^{k_1})}$  and  $\|\hat{D} - D\|_{L^2(\mathbb{S}^{k_2})}$  are

$$\left\{ O_p(N^{-1/2}(\ln N)^r T_j^{\frac{2k_j+1}{2}}) + O_p((\ln N)^r T_j^{\frac{k_j+1}{2}}) \right\} \times \max_{i=1, \dots, N} \left| \frac{f_{Z_j}^a(Z_{j,i})}{\hat{f}_{Z_j}^a(Z_{j,i})} - 1 \right| \quad (\text{C.21})$$

$$+ O_p(N^{-1/2}(\ln N)^r T_j^{\frac{2k_j+1}{2}}) + O(T_j^{\frac{k_j+1}{2}} \sigma(S_{jN})) + O(T_j^{-s_j}), \quad (\text{C.22})$$

where  $j = 1$  for  $\|\hat{B} - B\|_{L^2(\mathbb{S}^{k_1})}$  and  $j = 2$  for  $\|\hat{D} - D\|_{L^2(\mathbb{S}^{k_2})}$ . We also note that these results follow directly from Theorem 5.1 in GK.

Now, for  $\hat{C}$ , we again define infeasible estimators as follows.

$$\begin{aligned} \hat{C}^{IT}(t_2) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{2W_i \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2) K_{T_1}(c_1, Z_{1i})}{f_{Z_2}^a(Z_{2i}|c_1) f_{Z_1}^a(Z_{1i})} \\ \hat{C}^I(t_2) &\equiv \frac{1}{N} \sum_{i=1}^N \frac{2W_i \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, Z_{2i}))(t_2) K_{T_1}(c_1, Z_{1i})}{f_{Z_2}(Z_{2i}|c_1) f_{Z_1}(Z_{1i})}. \end{aligned}$$

Similar to the analysis of  $\|\hat{A} - \hat{A}^{IT}\|_{L^2(\mathbb{S}^{k_1} \times \mathbb{S}^{k_2})}$ , the stochastic component of the plug-in estimator

obeys

$$\|\hat{C} - \hat{C}^{IT}\|_{L^2(\mathbb{S}^{k_2})} \leq B(k_2, 2)T_2^{\frac{k_2+1}{2}} (\ln N)^{2r} \left\| \frac{1}{N} \sum_{i=1}^N K_{T_2}^-(\cdot, Z_{2i})(t_2) K_{T_1}(c_1, Z_{1i}) \right\|_{L^2(\mathbb{S}^{k_2})} \quad (\text{C.23})$$

$$\times \max_{i=1, \dots, N} \left| \frac{f_{Z_2}^a(Z_{2i}|c_1) f_{Z_1}^a(Z_{1i})}{\hat{f}_{Z_2}^a(Z_{2i}|c_1) \hat{f}_{Z_1}^a(Z_{1i})} - 1 \right| \quad (\text{C.24})$$

$$= \{O_p(N^{-1/2}(\ln N)^{2r} T_2^{\frac{k_2+1}{2}}) + O_p((\ln N)^{2r} T_2^{\frac{k_2+1}{2}})\} \times \max_{i=1, \dots, N} \left| \frac{f_{Z_2}^a(Z_{2i}|c_1) f_{Z_1}^a(Z_{1i})}{\hat{f}_{Z_2}^a(Z_{2i}|c_1) \hat{f}_{Z_1}^a(Z_{1i})} - 1 \right|, \quad (\text{C.25})$$

where we used Lemma C.3 the fact that  $\|K_{T_1}(c_1, \cdot)\|_\infty < \infty$ , which follows from (4.4), (B.8), and  $t \mapsto L_{n_1}^{k_1}(t)$  being bounded on  $[0, 1]$ . Similarly, again by applying Lemma C.3, it follows that

$$\|\hat{C}^{IT} - E[\hat{C}^{IT}]\|_{L^2(\mathbb{S}^{k_2})} = O_p(N^{-1/2}(\ln N)^{2r} T_2^{\frac{k_2+1}{2}}). \quad (\text{C.26})$$

Below, we let  $S_{1N} \equiv \{(z_1, z_2) : 0 < f_{Z_1}(z_1) < (\ln N)^{-r}\}$  and  $U_{2N} \equiv \{(z_1, z_2) : 0 < f_{Z_2}(z_2|c_1) < (\ln N)^{-r}\}$ . For the trimming bias, we may then write

$$E[\hat{C}^{IT}] - E[\hat{C}^I] = \int_{\mathbb{S}^{k_1} \times \mathbb{S}^{k_2}} E[2W_i | Z_1 = z_1, Z_2 = z_2] \mathcal{H}_{\mathbb{S}^{k_2}}^{-1}(K_{T_2}^-(\cdot, z_2))(t_2) K_{T_1}(c_1, z_1) \quad (\text{C.27})$$

$$\times \{f_{Z_2}(z_2|c_1) f_{Z_1}(z_1) (\ln N)^{2r} - 1\} 1\{(z_1, z_2) \in S_{1N} \cap U_{2N}\} \quad (\text{C.28})$$

$$+ (f_{Z_2}(z_2|c_1) (\ln N)^r - 1) 1\{(z_1, z_2) \in U_{2N} \setminus S_{1N}\} \quad (\text{C.29})$$

$$+ (f_{Z_1}(z_1) (\ln N)^r - 1) 1\{(z_1, z_2) \in S_{1N} \setminus U_{2N}\} \frac{f_{Z_2}(z_2|z_1)}{f_{Z_2}(z_2|c_1)} d\sigma(z_1, z_2). \quad (\text{C.30})$$

By Proposition 2.2 in GK (applied twice) and Lemma C.3, it then follows that

$$\|E[\hat{C}^{IT}] - E[\hat{C}^I]\|_{L^2(\mathbb{S}^{k_2})} \quad (\text{C.31})$$

$$\leq 5B(k_2, 2)T_2^{\frac{k_2+1}{2}} \|K_{T_2}^-(\cdot, z_2)\|_{L^2(\mathbb{S}^{k_2})} \|K_{T_1}(c_1, \cdot)\|_{L^2(\mathbb{S}^{k_2})} \times \int_{S_{1N} \cup U_{2N}} \frac{f_{Z_2}(z_2|z_1)}{f_{Z_2}(z_2|c_1)} d\sigma(z_1, z_2) \quad (\text{C.32})$$

$$\leq \bar{c}T_2^{\frac{k_2+1}{2}} |f_{Z_1}(c_1)| \int_{\bar{S}_{1N}} f_{Z_1}(z_1)^{-1} d\sigma_{k_1}(z_1), \quad (\text{C.33})$$

where  $\bar{S}_{1N}$  is the projection of  $S_{1N}$  to  $\mathbb{S}^{k_1}$ . Finally, by Assumption 4.1, we have  $C \in \mathbb{W}_2^{s_1}(\mathbb{S}^{k_2})$ . Hence, by Assumption C.1 (iii),

$$\|E[\hat{C}^I] - C\|_{L^2(\mathbb{S}^{k_2})} = \left\| \int_{\mathbb{S}^{k_2}} C(\tilde{t}_2) K_{T_2}^-(t_2, \tilde{t}_2) d\sigma_{k_2}(\tilde{t}_2) - C(t_2) \right\|_{L^2(\mathbb{S}^{k_2})} \leq \bar{c}T^{-s_2}. \quad (\text{C.34})$$

By (C.25), (C.26), (C.30), and (C.34), we have

$$\|\hat{C} - C\|_{L^2(\mathbb{S}^{k_2})} = \{O_p(N^{-1/2}(\ln N)^{2r}T_2^{\frac{k_2+1}{2}}) + O_p((\ln N)^{2r}T_2^{\frac{k_2+1}{2}})\} \times \max_{i=1, \dots, N} \left| \frac{f_{Z_2}^a(Z_{2i}|c_1)f_{Z_1}^a(Z_{1i})}{\hat{f}_{Z_2}^a(Z_{2i}|c_1)\hat{f}_{Z_1}^a(Z_{1i})} - 1 \right| \quad (\text{C.35})$$

$$+ O_p(N^{-1/2}(\ln N)^{2r}T_2^{\frac{k_2+1}{2}}) + O(T_2^{\frac{k_2+1}{2}})|f_{Z_1}(c_1)| \int_{S_{1N}} f_{Z_1}(z_1)^{-1} d\sigma_{k_1}(z_1) + O(T^{-s_2}). \quad (\text{C.36})$$

Similarly,

$$\|\hat{E} - E\|_{L^2(\mathbb{S}^{k_1})} = \{O_p(N^{-1/2}(\ln N)^{2r}T_1^{\frac{k_1+1}{2}}) + O_p((\ln N)^{2r}T_1^{\frac{k_1+1}{2}})\} \times \max_{i=1, \dots, N} \left| \frac{f_{Z_1}^a(Z_{1i}|c_2)f_{Z_2}^a(Z_{2i})}{\hat{f}_{Z_1}^a(Z_{1i}|c_2)\hat{f}_{Z_2}^a(Z_{2i})} - 1 \right| \quad (\text{C.37})$$

$$+ O_p(N^{-1/2}(\ln N)^{2r}T_1^{\frac{k_1+1}{2}}) + O(T_1^{\frac{k_1+1}{2}})|f_{Z_2}(c_2)| \int_{S_{2N}} f_{Z_2}(z_2)^{-1} d\sigma_{k_2}(z_2) + O(T^{-s_1}), \quad (\text{C.38})$$

where  $\bar{S}_{2N} \equiv \{z_2 \in \mathbb{S}^{k_2} : 0 < f_{Z_2}(z_2) < (\ln N)^{-r}\}$ .

Given these results, we choose  $T_1$  and  $T_2$  so that we may balance the variance, which is of the order  $O_p(N^{-1/2}(\ln N)^r(T_1^{\frac{2k_1+1}{2}} + \ln N + 1)(T_2^{\frac{2k_2+1}{2}} + \ln N + 1))$  and the bias, which is of the order  $O(T_1^{-s_1} \vee T_2^{-s_2})$ . This leads to the choice of  $T_j$  in (4.9). Under this choice, the convergence rate in (4.10) follows.  $\square$

#### APPENDIX D: Proof of Theorem 5.1

**Theorem D.1.** *Let  $\lambda_n := \lambda(n, 1)$  be the eigenvalues of the hemispherical transform  $\mathcal{H}_{\mathbb{S}^1}$  to the Fourier basis  $\varphi_n(t) = (2\pi)^{-1} \exp(-int)$ . A singular system  $(\Phi_n, \varphi_n, \sigma_n)$  for non-zero singular values of  $\mathcal{T}_c$  is given by the singular values*

$$\sigma_n \equiv \frac{1}{2\pi} \left( \sum_{n_1+n_2=n} \lambda_{n_1}^2 \lambda_{n_2}^2 \right)^{\frac{1}{2}} \quad n \in \mathbb{Z},$$

the following functions in  $L^2(\mathbb{S}^1 \times \mathbb{S}^1)$

$$\Phi_n \equiv \frac{1}{2\pi\sigma_n} \sum_{n_1+n_2=n} \lambda_{n_1} \lambda_{n_2} \varphi_{n_1} \varphi_{n_2},$$

and the Fourier basis  $\varphi_n \in L^2(\mathbb{S}^1)$ . I.e.,  $\mathcal{T}_c \Phi_n = \sigma_n \varphi_n$ .

Proof of Theorem D.1. Note that all  $\Phi_n$  and  $\varphi_n$  have norm 1. We show that  $\Phi_n$  are the eigenfunctions

of  $\mathcal{T}_c^* \mathcal{T}_c$  and  $\sigma_n^2$  the corresponding eigenvalues. Let us start with the observation

$$\mathcal{T}_c(\varphi_{n_1} \varphi_{n_2}) = \frac{1}{2\pi} \lambda_{n_1} \lambda_{n_2} \varphi_{n_1+n_2}$$

which allows to characterize  $\mathcal{T}_c^*$  by

$$\frac{1}{2\pi} \lambda_{n_1} \lambda_{n_2} = \langle \mathcal{T}_c(\varphi_{n_1} \varphi_{n_2}), \varphi_{n_1+n_2} \rangle_{L^2(\mathbb{S}^1)} = \langle \varphi_{n_1} \varphi_{n_2}, \mathcal{T}_c^* \varphi_{n_1+n_2} \rangle_{L^2(\mathbb{S}^1 \times \mathbb{S}^1)}.$$

Hence,

$$\mathcal{T}_c^* \varphi_n = \frac{1}{2\pi} \sum_{n_1+n_2=n} \lambda_{n_1} \lambda_{n_2} \varphi_{n_1} \varphi_{n_2}.$$

This sum exists in  $L^2(\mathbb{S}^1 \times \mathbb{S}^1)$  because the  $\lambda_n$  are square-summable. Now we can compute

$$\begin{aligned} \mathcal{T}_c^* \mathcal{T}_c(\Phi_n) &= \mathcal{T}_c^* \mathcal{T}_c \left( \frac{1}{2\pi \sigma_n} \sum_{n_1+n_2=n} \lambda_{n_1} \lambda_{n_2} \varphi_{n_1} \varphi_{n_2} \right) \\ &= \mathcal{T}_c^* \left( \frac{1}{4\pi^2 \sigma_n} \sum_{n_1+n_2=n} \lambda_{n_1}^2 \lambda_{n_2}^2 \varphi_n \right) \\ &= \frac{1}{8\pi^3 \sigma_n} \left( \sum_{n_1+n_2=n} \lambda_{n_1}^2 \lambda_{n_2}^2 \right) \sum_{\tilde{n}_1+\tilde{n}_2=n} \lambda_{\tilde{n}_1} \lambda_{\tilde{n}_2} \varphi_{\tilde{n}_1} \varphi_{\tilde{n}_2} \\ &= \sigma_n^2 \left( \frac{1}{2\pi \sigma_n} \sum_{\tilde{n}_1+\tilde{n}_2=n} \lambda_{\tilde{n}_1} \lambda_{\tilde{n}_2} \varphi_{\tilde{n}_1} \varphi_{\tilde{n}_2} \right) \\ &= \sigma_n^2 \Phi_n. \end{aligned}$$

Hence, all  $\Phi_n$  are eigenfunctions to the eigenvalue  $\sigma_n^2$ . In addition, the first step of the last computation shows that  $\mathcal{T}_c(\Phi_n) = \sigma_n \varphi_n$ . Thus,  $\mathcal{T}_c$  is a bijection between  $\text{span}\{\Phi_n | n \in \mathbb{Z}\}$  and  $L^2(\mathbb{S}^1)$ . So, there can be no further eigenfunction which is not in the null space of  $\mathcal{T}_c^* \mathcal{T}_c$ . This completes the proof.  $\square$

**Corollary D.1.** *If  $\varphi_n$  is an odd function on  $\mathbb{S}^1$ , i.e.  $n$  is an odd number, then  $\sigma_n = 2^{-1/2} \lambda_n$  and  $\Phi_n(z_1, z_2) = (2\sqrt{2}\pi)^{-1}(\varphi_n(z_1) + \varphi_n(z_2))$*

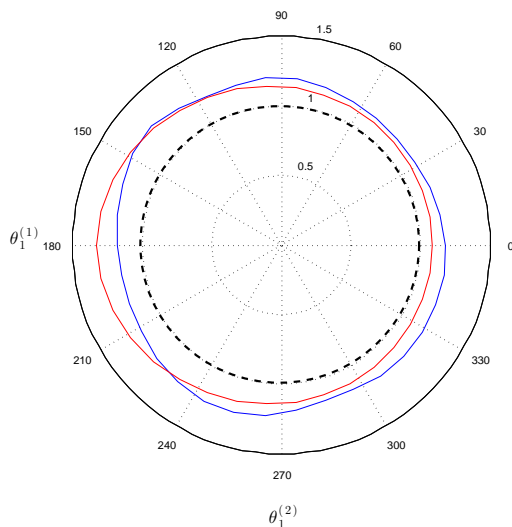
Proof of Corollary D.1. If  $\varphi_n$  is odd, then  $\varphi_n = 2\pi \varphi_{n_1} \varphi_{n_2}$  holds only if one of the functions is odd and the other one is even. The eigenvalues  $\lambda_n$  vanish for even functions that are non constant. Hence, only  $\varphi_n(z_1) \varphi_0(z_2)$  and  $\varphi_0(z_1) \varphi_n(z_2)$  contribute to the sums in the definitions of  $\sigma_n$  and  $\Phi_n$ . This shows the assertion.  $\square$

**Corollary D.2.** *If  $\varphi_n$  is an odd spherical harmonic on  $\mathbb{S}^k$  and  $\lambda_n$  the corresponding eigen value of the hemispherical transform, then  $\langle \mathcal{T}_c f_\theta, \varphi_n \rangle = \lambda_n / 2 (a_{0,n} + a_{n,0})$ .*

Proof of Corollary D.2. The argument to proof this corollary is the same as for Corollary D.1. Hence, it is omitted.  $\square$

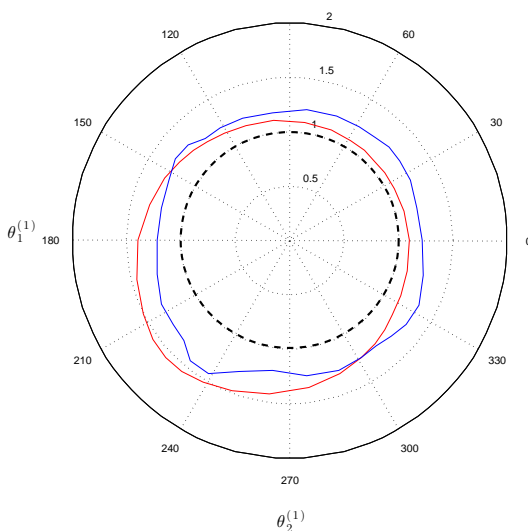
APPENDIX E: Figures

Figure 1:  $f_{\theta_1^{(1)}, \theta_1^{(2)}}$  and  $\hat{f}_{\theta_1^{(1)}, \theta_1^{(2)}}$  under Specification 1.



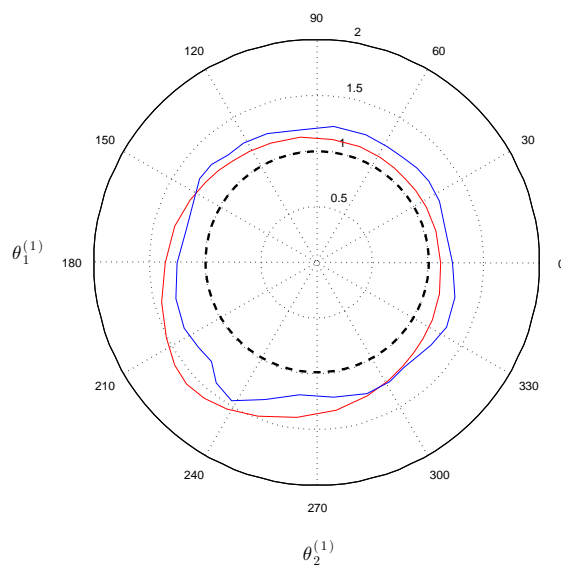
Note: For each  $(t_1^{(1)}, t_1^{(2)}) \in \mathbb{S}^1$ , the red curve's distance from the unit circle (dashed circle) gives  $f_{\theta_1^{(1)}, \theta_1^{(2)}}(t_1^{(1)}, t_1^{(2)})$ . Similarly, the blue curve's distance from the unit circle gives  $\hat{f}_{\theta_1^{(1)}, \theta_1^{(2)}}(t_1^{(1)}, t_1^{(2)})$ .

Figure 2:  $f_{\theta_1^{(1)}, \theta_2^{(1)}}$  and  $\hat{f}_{\theta_1^{(1)}, \theta_2^{(1)}}$  under Specification 1



Note: For each  $(t_1^{(1)}, t_2^{(1)}) \in \mathbb{S}^1$ , the red curve's distance from the unit circle (dashed circle) gives  $f_{\theta_1^{(1)}, \theta_2^{(1)}}(t_1^{(1)}, t_2^{(1)})$ . Similarly, the blue curve's distance from the unit circle gives  $\hat{f}_{\theta_1^{(1)}, \theta_2^{(1)}}(t_1^{(1)}, t_2^{(1)})$ .

Figure 3:  $f_{\theta_1^{(1)}, \theta_2^{(1)}}$  and  $\hat{f}_{\theta_1^{(1)}, \theta_2^{(1)}}$  under Specification 2



Note: For each  $(t_1^{(1)}, t_2^{(1)}) \in \mathbb{S}^1$ , the red curve's distance from the unit circle (dashed circle) gives  $f_{\theta_1^{(1)}, \theta_2^{(1)}}(t_1^{(1)}, t_2^{(1)})$ . Similarly, the blue curve's distance from the unit circle gives  $\hat{f}_{\theta_1^{(1)}, \theta_2^{(1)}}(t_1^{(1)}, t_2^{(1)})$ .