SEMIPARAMETRIC EFFICIENCY BOUNDS UNDER SHAPE RESTRICTIONS

GAUTAM TRIPATHI

Department of Economics
University of Wisconsin-Madison.

This Version: September 22, 1997.

Abstract. Consider the regression $y = x \beta_0 + f^*(x) + \varepsilon$ with $\varepsilon \sim N(0, \sigma_0^2)$. In this paper we show how to compute the asymptotic lower bounds for efficiently estimating $\beta_0$, when the only information we have about $f^*$ is that it has a certain shape; i.e., we calculate efficiency bounds for $\beta_0$ when the only thing we know about $f^*$ is that it lies in $\mathcal{F}$, where $\mathcal{F}$ is a compact set of functions with certain shape properties. These shape properties are such that $\mathcal{F}$ is a closed convex cone. The efficiency bounds, for $n^{1/2}$ consistent regular estimators of $\beta_0$, are shown to be determined by a projection onto $\text{lin} T(\mathcal{F}, f^*)$, the smallest closed linear space containing the tangent cone to $\mathcal{F}$ at $f^*$. This tangent cone, denoted by $T(\mathcal{F}, f^*)$, seems at first sight to be the natural space to determine the efficiency bounds. However, we prove an “impossibility” result showing that a projection onto $T(\mathcal{F}, f^*)$ may yield bounds that are not attainable by any $n^{1/2}$ consistent, regular estimator of $\beta_0$. This “impossibility” result is used to show that in the class of all $n^{1/2}$ consistent regular estimators of $\beta_0$, concavity of $f^*$ does not help in estimating $\beta_0$ more efficiently, while homogeneity of $f^*$ may lead to dramatic efficiency gains in estimating $\beta_0$.

1. Introduction

The objective of this paper is to determine how finite dimensional parameters can be efficiently estimated for an important class of economic models. In these models

Key words and phrases. Efficiency Bounds, Semiparametric Models, Shape Restrictions.

I wish to thank Professors Ian Domowitz, Rosa Matzkin and Tom Severini for their constant encouragement and advice during the course of this research. Without their help this paper would not have been written. I am also grateful for the comments provided by seminar audiences at Johns Hopkins, Michigan State, Northwestern, Ohio State, Virginia and the University of Wisconsin-Madison. Needless to say, all errors still remain my responsibility. This paper is a highly modified version of Chapters 2 and 3 of my Northwestern PhD thesis, which was supported by an Alfred P. Sloan Dissertation Fellowship in Economics.
some of the functions are known up to a finite dimensional parameter $\beta_0 \in \mathbb{R}^p$, while the other functions are known only to possess some shape properties such as concavity, linear homogeneity and monotonicity. These shape properties are actually the restrictions that economic theory imposes upon the unknown function. We let $\mathcal{F}$ denote the set of shape restricted functions.

Semiparametric estimation of the finite dimensional parameters, when $\mathcal{F}$ is just a set of functions satisfying some smoothness properties, has been studied extensively. However, when dealing with economic data, the functions in $\mathcal{F}$ often have to satisfy other restrictions besides smoothness. Though extensively studied in economic theory, there has only been a limited use of these restrictions in econometric practice notwithstanding their tremendous usefulness. In the models that we will consider, the imposition of shape restrictions usually leads to a reduction in the variance of estimators (see Matzkin (1994) for more on this). Omitting shape restrictions, when economic theory demands otherwise, would typically lead to inefficient estimation procedures thus reducing the power of subsequent statistical analysis. This could have important policy implications since semiparametric models are being increasingly used to answer policy related questions. However, inclusion of shape restrictions complicates estimation because such restrictions generate constraints that are infinite dimensional in nature. To deal with these infinite dimensional constraints, we use in this paper certain techniques borrowed from nonlinear analysis.

In this paper we show how to calculate the minimum asymptotic variance, hereafter called the efficiency bound, that a $n^{1/2}$ consistent regular estimator of the finite dimensional parameter ($\beta_0$) can achieve in a semiparametric model when the unknown function in the model, denoted by $f^*$, is either concave or homogeneous of degree $r$, where $r$ is known; i.e. $f^* \in \mathcal{F}$ where $\mathcal{F}$ is the set of all concave or homogeneous functions of degree $r$. The paper is organized as follows. We develop an argument to show that computing efficiency bounds for an estimator of $\beta_0$ requires a projection onto $T(\mathcal{F}, f^*)$, the tangent cone to $\mathcal{F}$ at the point $f^*$. However, as shown later, when $T(\mathcal{F}, f^*)$ is a proper cone\(^1\) efficiency bounds determined by a projection onto $T(\mathcal{F}, f^*)$ cannot be attained by any regular $n^{1/2}$ consistent estimator of $\beta_0$. We call this an “impossibility” result and, following van der Vaart (1989), conclude that efficiency bounds for regular estimators of $\beta_0$ are determined by a projection onto $\text{lin} T(\mathcal{F}, f^*)$, the smallest closed linear space containing $T(\mathcal{F}, f^*)$, rather than the tangent cone $T(\mathcal{F}, f^*)$ itself. Using this procedure we calculate efficiency bounds for regular estimators of $\beta_0$, when $f^*$ is either concave or homogeneous of degree $r$. The impossibility result is used to show that in the class of all regular $n^{1/2}$ consistent estimators of $\beta_0$, concavity of $f^*$ does not help in estimating $\beta_0$ more efficiently. However, by means of some simple examples, we demonstrate that homogeneity of $f^*$ may lead to large efficiency gains in estimating $\beta_0$.

The extension of efficiency bounds from a purely parametric to the semiparametric case was first proposed by Stein (1956) and subsequently developed in the statistical works cited in Bickel, Klassen, Ritov, and Wellner (1993). Attracted by

\(^1\)Which happens, for instance, when $\mathcal{F}$ is the set of all concave functions and $f^* \in \text{bdry}(\mathcal{F})$.\)
the elegance of the semiparametric approach and its wide applicability to economics, several econometricians mentioned in Newey (1990)'s excellent survey article have also made valuable contributions to this area in recent years. However, most of the research to date has concentrated upon developing efficiency bounds for distribution free models, i.e. models in which the distribution of the error term is unknown (Chamberlain 1986, Cosslett 1987). Where shape restrictions have been involved, they have been imposed on the error distribution (Newey 1988), rather than on the unknown function. But these cases form too narrow a class, since they exclude models with shape restrictions which arise regularly in microeconometrics; e.g. at the firm or the consumer level. In fact, to the best of my knowledge, this attempt is the first of its kind to develop efficiency bounds for models where the shape restrictions are imposed on the unknown functional form rather than on the distribution of the error term. This research, therefore, extends the class of models that econometricians can deal with efficiently. It should be of particular interest to any practitioner in the field, because it provides new insights into incorporating shape restrictions in estimation procedures.

This paper limits itself to the analysis of i.i.d. observations which are generated by a partially linear model. Moreover, we have only concentrated upon the computation of efficiency bounds (under shape restrictions) in this paper. For the construction of estimators that achieve these bounds see Tripathi (1997b). All mathematical definitions have been included in the paper to make it as self contained as possible. Furthermore, to enhance readability, all proofs have been confined to the appendices.

2. The Partially Linear Model

Let us begin by constructing two examples of partially linear models that may typically occur in microeconomics. In these examples we are interested in estimating the finite dimensional parameters when the only information we have about the unknown function is that it belongs to a set of functions whose elements satisfy certain shape properties.

**Example 2.1.** A firm produces two different goods with production functions \( F_1 \) and \( F_2 \). That is, \( y_1 = F_1(x) \), and \( y_2 = F_2(z) \), with \( (x \times z) \in \mathbb{R}^n \times \mathbb{R}^m \). The firm maximizes total profits \( p_1 y_1 - w'_1 x + p_2 y_2 - w'_2 z \). The maximized profit can be written as \( \pi_1(u) + \pi_2(v) \), where \( u = (p_1, w_1) \), and \( v = (p_2, w_2) \). Now suppose that the econometrician has sufficient information about the first good to parameterize the first profit function as \( \pi_1(u) = u' \theta_0 \). Then the observed profit \( \pi_i = u'_i \theta_0 + \pi_2(v_i) + \epsilon_i \), where \( \pi_2 \) is monotone, convex, linearly homogeneous and continuous in its arguments. □

**Example 2.2.** Again, suppose we have a similar but geographically dispersed firms with the same profit function. This could happen if, for instance, these firms had access to similar technology. Now suppose that the observed profit depends not only upon the price vector, but also on a linear index of exogenous variables. That is,
\[ \pi_i = x_i^\prime \theta_0 + \pi^* (p_1, \ldots, p_{k_0}) + \varepsilon_i, \] where the profit function \( \pi^* \) is continuous, monotone, convex, and homogeneous of degree one in its arguments. \( \square \)

We are now ready to begin our study of a general shape restricted semiparametric model by analyzing the partially linear model. But first, some definitions that may be useful for the reader.

**Definition 2.1 (Cone).** Let \( X \) be a vector space over \( \mathbb{R} \). A subset \( C \) of \( X \) is called a cone if and only if for any \( c \in C \), and any \( \lambda \geq 0 \) we have \( \lambda c \in C \). \( \square \)

**Definition 2.2 (Proper Cone).** A cone \( C \) is said to be proper if it is not a linear space. \( \square \)

**Notation 2.1.** Let \( Z \) be a compact subset of \( \mathbb{R}^q \) and, let \( \mathcal{H} \) denote the set of all \( C^2(Z) \) functions with uniformly bounded values, gradients, and Hessians. \( \mathcal{F} \subset \mathcal{H} \) is a closed (w.r.t. \( C^2 \) norm) convex cone in \( \mathcal{H} \), while \( \overline{lin} \mathcal{F} \) is the smallest closed linear space containing \( \mathcal{F} \). \( \ell(\cdot) \) represents the loglikelihood function for a single observation. All vectors are expressed in boldface. For instance, \( \mathbf{x} = (x_1, x_2, \ldots, x_p) \) denotes the value taken by random the variable \( \mathbf{X} = (X_1, X_2, \ldots, X_p) \). The inner product of two vectors \( (\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{2p} \) is given by \( \mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^p a_i b_i \). \( \square \)

For \( i = 1, \ldots, n \), consider the regression

\[ y_i = \mathbf{x}_i \cdot \beta_0 + f^*(\mathbf{z}_i) + \varepsilon_i. \]

**Assumption 2.1.** The data \( \{\mathbf{x}_i, y_i, \mathbf{z}_i\}_{i=1}^n \) are assumed to be realizations of i.i.d. random variables \( (\mathbf{X}, Y, Z) \) which take values in \( \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^q \). Furthermore,

(i) \( \varepsilon \overset{d}{=} \mathcal{N}(0, \sigma_\varepsilon^2) \), where \( \sigma_\varepsilon^2 \) is known;

(ii) \( \beta_0 = (\beta_{10}, \ldots, \beta_{p0}) \in \text{int} (\mathcal{B}) \), where \( \mathcal{B} \) is a compact subset of \( \mathbb{R}^p \);

(iii) \( (\mathbf{x}, \mathbf{z}) \) come from distributions with compact support \( \mathbf{X} \times \mathbf{Z} \in \mathbb{R}^p \times \mathbb{R}^q \), and have a joint density function \( g_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}) \).

(iv) The joint pdf \( g_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}) \) and the marginal pdf \( g_{\mathbf{Z}}(\mathbf{z}) \) are both twice continuously differentiable w.r.t. \( \mathbf{z} \).

(v) \( f^* \in \mathcal{F} \).

(vi) \( \varepsilon \) and \( (\mathbf{x}, \mathbf{z}) \) are independent. \( \square \)

**Remark 2.1.** (i) Following Newey (1990), it can be shown that the efficiency bounds for estimating \( \beta_0 \) are not affected by the knowledge of \( \sigma_\varepsilon^2 \). Therefore, the assumption that \( \sigma_\varepsilon^2 \) is known is w.l.o.g.

(ii) Since \( \mathcal{H} \) is a compact subset of \( C^2(Z) \) w.r.t. the \( C^2 \) norm and \( \mathcal{F} \) is a closed subset of \( \mathcal{H} \), \( \mathcal{F} \) is also compact w.r.t. the \( C^2 \) norm.

(iii) It is easy to see that \( \beta_0 \), without any intercept term, is identified if and only if the elements of the vector \( \mathbf{x} - \mathbb{E} (\mathbf{x}|\mathbf{z}) \) are linearly independent. This assumption is maintained throughout the paper.

(iv) Throughout this paper we have also maintained the assumption that the operations of integration and differentiation can be exchanged. \( \square \)
Since the errors are Gaussian, the loglikelihood function for a single observation is given by \( l(\beta_0, f^*(x, y, z)) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} [y - x \cdot \beta_0 - f^*(z)]^2 + \log g_{x,z}(x, z) \). The score functions \( \frac{\partial l}{\partial \beta} |_{\beta = \beta_0} \) are assumed to satisfy the following assumption.

**Assumption 2.2.** The score functions are elements of the Hilbert space \( L^2(\mathcal{D}) \), where \( \mathcal{D} \) is the probability measure induced by \( (X, Y, Z) \). □

This is very useful since the geometry of Hilbert spaces facilitates the solution of projection problems which we shall soon encounter. Let us begin by looking at the geometry bounds for the case when the parameter of interest is a scalar. Once the groundwork has been laid, it will be extended to the case when the parameter of interest is multi-dimensional. So for each \( n \), let \( \hat{\beta}_n \) be an estimate (based on \( n \) i.i.d. observations) of a real valued parameter \( \beta_0 \). Suppose that for each \( \beta_0 \), \( n^{-1/2} (\hat{\beta}_n - \beta_0) \Rightarrow N(0, v(\beta_0)) \). Then according to Fisher, \( v(\beta_0) \geq \frac{1}{i_{\beta_0}} \), where \( i_{\beta_0} \) is the information contained in a single observation, and \( v(\beta_0) \) is called the asymptotic variance of \( \hat{\beta}_n \). However, in the absence of suitable regularity conditions, this relationship does not necessarily hold as is indicated by the canonical example of a superefficient estimator. Therefore, to ensure that the information inequality holds, we only consider regular estimators as defined below.

**Definition 2.3 (Regular Estimator).** Let \( \beta_0, \delta \in \mathbb{R}^p \), and \( \beta_n = \beta_0 + n^{-1/2} \delta \). Also, let \( \hat{\beta}_n \) be a consistent estimator of \( \beta_0 \). Then the sequence of estimators \( \hat{\beta}_n \) is said to be regular if \( n^{-1/2} (\hat{\beta}_n - \beta_n) \) converges in distribution (under \( \beta_n \)) to a limiting distribution that does not depend upon \( \delta \). □

This notion of regularity is akin to stability; i.e. a regular estimator is stable in the sense that small perturbations of the true parameter do not alter its asymptotic distribution. See, for instance, van der Vaart (1989). We also need to define the following terms.

**Definition 2.4 (Tangent Vector).** Let \( M \) be a subset of a Banach space \( X \). A vector \( z \in X \) is said to be tangent to the set \( M \) at a point \( x_0 \), if there exist an \( \epsilon > 0 \) and a mapping \( t \mapsto r(t) \) of the interval \( (0, \epsilon) \) into \( X \) such that

\[
x_0 + tx + r(t) \in M \quad \forall t \in (0, \epsilon) \quad \text{and} \quad \frac{\| r(t) \|}{t} \to 0 \quad \text{as} \quad t \to 0. \quad \square
\]

**Definition 2.5 (Tangent Cone).** The set of vectors which are tangent to a set \( M \) at the point \( x_0 \), is denoted by \( T(M, x_0) \), and is a closed non-empty cone. This cone is called the tangent cone to \( M \) at \( x_0 \). □

**Definition 2.6 (Tangent Space).** If the tangent cone \( T(M, x_0) \) is a linear space it is called the tangent space to \( M \) at \( x_0 \). □

**Notation 2.2.** Let \( \text{lin} T(M, x_0) \) be the tangent cone to \( M \) at \( x_0 \). Then \( \text{lin} T(M, x_0) \) denotes the smallest closed linear space containing \( T(M, x_0) \). □
An equivalent characterization of tangent vectors and tangent cones is given in Appendix A. This appendix also contains several useful results about tangent cones that will be used subsequently.

We now return to our original problem. So let \( \beta_0 \in \text{int}(\mathcal{B}) \subseteq \mathbb{R} \), and \( f^* \in \mathcal{F} \) be the true value of the nuisance parameter. Assume that the parameter space \( \mathcal{B} \times \overline{\text{lin}} \mathcal{F} \) is parameterized by a smooth curve \( \beta \mapsto (\beta, \eta_\beta) \) such that \( \eta_\beta |_{\beta = \beta_0} = f^* \). Furthermore, let \( t \mapsto \beta_t \) be a smooth curve in \( \mathcal{B} \) through \( \beta_0 \). Then any point in \( \mathcal{B} \times \overline{\text{lin}} \mathcal{F} \) has coordinates \((\beta_t, \eta_{\beta_t})\). With this parameterization, estimating \( t \) is equivalent to estimating \((\beta_t, \eta_{\beta_t})\). The Fisher information for estimating \( t \) is then given by,

\[
\mathbb{E} \left[ \frac{d}{dt} \ell(\beta_t, \eta_{\beta_t}) \right]_{t=0}^2 = \left( \frac{d \beta_t}{dt} \right)_{t=0}^2 \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \right]^2.
\]

Since \( \eta \) is an element of a function space, the derivative \( \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \) denotes the Fréchet derivative of the loglikelihood function with respect to \( \eta \). This derivative operates on the tangent vector \( \frac{d}{dt} \eta_{\beta_0} \). The partial derivative \( \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \), is the score w.r.t. the parameter of interest while \( \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} \) may be interpreted as the score w.r.t. the nuisance parameter. Therefore, \( \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta}(\overline{\text{lin}} \mathcal{F}, f^*) \) is the space spanned by the scores w.r.t. the nuisance parameter.

Let \( \tilde{\tau}_{\beta_0} \) be the Fisher information for estimating \( \beta_0 \). Then since \( \beta \) is parameterized as a function of \( t \), an application of the chain rule gives

\[
\tilde{\tau}_{\beta_0} = \mathbb{E} \left[ \frac{d}{dt} \ell(\beta_t, \eta_{\beta_t}) \right]_{t=0}^2 = \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \right]^2.
\]

Notice that the information \( \tilde{\tau}_{\beta_0} \) depends upon the parameterization \( \eta_\beta \) only through the tangent vector \( \frac{d}{dt} \eta_{\beta_0} \). The tangent vector \( \delta^* \) which gives the least information for estimating \( \beta \) is then

\[
\delta^* = \arg\min_{\delta \in \overline{\text{lin}} \mathcal{T}(\mathcal{F}, f^*)} \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta) \right]^2.
\]

That is

\[
\frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta}(\delta^*) = -\text{proj} \left( \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \overline{\text{lin}} \mathcal{T}(\mathcal{F}, f^*) \right) \right)
\]

and, in a Hilbert space, \( \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta}(\delta^*) \) is characterized by the necessary and sufficient conditions given in Theorem B.2 (see Appendix B). Now let

\[
i_{\beta_0} = \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta^*) \right]^2.
\]

Then \( i_{\beta_0} \) is called the semiparametric information for \( \beta_0 \) and it may be shown that \( i_{\beta_0}^{-1} \) is the lower bound for the asymptotic variance of any regular estimator of \( \beta \).
The direction $\delta^*$ is the projection of the score w.r.t. the parameters of interest onto the space generated by the nuisance parameter scores. Following Severini and Wong (1992), we call $\delta^*$ the least favorable direction for estimating $\beta_0$. A curve $\eta_\beta$ which gives rise to $\delta^*$ as the tangent vector, at $\beta = \beta_0$, is called a least favorable curve. Notice that while Theorem B.1 (in Appendix B) implies that the least favorable direction is unique, no such implication holds for the least favorable curve. For instance, the curves $t \mapsto f^* + t\delta^*$ and $t \mapsto f^* + t(t + 1)\delta^*$ give rise to the same least favorable direction at $t = 0$. Hence, we have the following definition.

**Definition 2.7 (Least Favorable Curve and Direction).** Let $B \times \overline{\text{lin}F}$ be parameterized by a smooth curve $\beta \mapsto (\beta, \eta_\beta)$ such that, $\eta_\beta|_{\beta = \beta_0} = f^*$. Then $\eta_\beta \in \overline{\text{lin}F}$ is said to be a least favorable curve in $\overline{\text{lin}F}$ for estimating $\beta_0$, if $\frac{d}{d\beta} \eta_{\beta_0} \in \overline{\text{lin}T(F, f^*)}$ minimizes

$$
\mathbb{E} \left[ \frac{d\ell(\beta_0, \eta_{\beta_0})}{d\beta} \right]^2 = \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \frac{d}{d\beta} \eta_{\beta_0} \right) \right]^2.
$$

Moreover, $\frac{d}{d\beta} \eta_{\beta_0}$ is called the least favorable direction in $\overline{\text{lin}T(F, f^*)}$ for estimating $\beta_0$. □

**Remark 2.2.** Let $\frac{d}{d\beta} \eta_{\beta_0}$ be the least favorable direction in $\overline{\text{lin}T(F, f^*)}$ for estimating $\beta_0$, where $\eta_\beta$ is a curve through $f^*$. For instance, $\eta_\beta(z) = f^*(z) + (\beta - \beta_0) \frac{d}{d\beta} \eta_{\beta_0}(z)$. Notice, however, that $\eta_\beta$ may not necessarily lie in the cone $F$. But since we can easily show that $\overline{\text{lin}T(F, f^*)} \subseteq \overline{\text{lin}F}$, it is clear that $\eta_\beta$ does lie in $\overline{\text{lin}F}$. Hence, we can always find a curve $\eta_\beta$ in $\overline{\text{lin}F}$ which gives rise to the least favorable direction $\frac{d}{d\beta} \eta_{\beta_0} \in \overline{\text{lin}T(F, f^*)}$. In fact, since $F$ is a convex cone containing $f^*$, we have the following result. □

**Lemma 2.1.** Let $F$ be a convex cone and $f^* \in F$. Then $\overline{\text{lin}T(F, f^*)} = \overline{\text{lin}F}$.

**Proof.** See Appendix A. □

Let us now consider the case when the parameter of interest is multidimensional. So suppose that $\beta_0 \in \text{int}(B) \subseteq \mathbb{R}^p$, where $p > 1$. Assume that $B \times \overline{\text{lin}F}$ is parameterized by a smooth curve $\beta \mapsto (\beta, \eta_\beta)$ such that $\eta|_{\beta = \beta_0} = f^*$. Following Severini (1987), we then have the following definition.

**Definition 2.8 (Least Favorable Surface).** Let $t \mapsto \beta_t$ be a smooth curve in $B$ through $\beta_0$. Then $\eta_t$ is called a least favorable surface in $\overline{\text{lin}F}$ for estimating $\beta_0$, if $\eta_{\beta_t}$ is a least favorable curve in $\overline{\text{lin}F}$ for estimating $t$. That is, $\eta_t$, minimizes $\mathbb{E} \left[ \frac{d\ell(\beta_t, \eta_{\beta_t})}{dt} \right]_{t=0}^2$. □

This definition leads to the following results. Severini (1987) first obtained these results for the case when $F$ was a linear space. The following theorems therefore extend Severini’s results to include the case when the nuisance parameter is restricted to lie in a convex cone. The proofs of these theorems are very similar to those in Severini (1987), and are provided in Appendix C for the sake of completeness.
Theorem 2.1. Let \( \eta_0 \) be a least favorable surface in \( \text{lin}\mathcal{F} \) for estimating \( \beta_0 \). Denote the least favorable direction in \( \text{lin}\mathcal{T}(\mathcal{F}, f^*) \) by \( \delta^* = (\frac{d}{d\beta_0}, \eta_0)_{\beta=\beta_0} \); i.e. \( \delta_i^* = (\frac{d}{d\beta_i}, \eta_0)_{\beta=\beta_0} \) for \( i = 1, \ldots, p \). Then \( \delta^* = (\delta_1^*, \ldots, \delta_p^*) \) satisfies

\[
\mathbb{E} \left[ \frac{\partial \ell(\beta_0, \eta_0)}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_0)}{\partial \eta} (\delta_i^*) \right] \frac{\partial \ell(\beta_0, \eta_0)}{\partial \eta}(\delta) = 0,
\]

for all \( \delta \in \text{lin}\mathcal{T}(\mathcal{F}, f^*) \) and, \( i = 1, \ldots, p \).

Theorem 2.2. Let \( \eta_0 \) be a smooth curve such that \( \eta_{\beta_0} = f^* \). Then \( \eta_0 \) is a least favorable surface in \( \text{lin}\mathcal{F} \) for estimating \( \beta_0 \), if and only if for all \( \delta_i \in \text{lin}\mathcal{T}(\mathcal{F}, f^*) \) and, \( i = 1, \ldots, p \)

\[
\sum_{i=1}^{p} \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \eta_0)}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_0)}{\partial \eta}(\delta_i^*) \right] \frac{\partial \ell(\beta_0, \eta_0)}{\partial \eta}(\delta_i) = 0.
\]

Theorem 2.3. Let,

\[
I = \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \eta_0)}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_0)}{\partial \eta}(\frac{d}{d\beta_0} \eta_0) \right] \left[ \frac{\partial \ell(\beta_0, \eta_0)}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_0)}{\partial \eta}(\frac{d}{d\beta_0} \eta_0) \right]^{-1}
\]

be the information matrix when \( \eta_0 \) is a smooth curve in \( \text{lin}\mathcal{F} \) such that \( \eta_{\beta_0} = f^* \). Then there exists a matrix \( I_{\beta_0} \) such that \( \alpha'(I_{\beta_0} - I)\alpha \leq 0 \) for all \( \alpha \in \mathbb{R}^p \), if and only if \( I_{\beta_0} \) corresponds to the information matrix when \( \eta_0 \) is a least favorable surface in \( \text{lin}\mathcal{F} \) for estimating \( \beta_0 \).

Theorem 2.3 shows that a least favorable surface in \( \text{lin}\mathcal{F} \), denoted by \( \eta_0 \), minimizes Fisher's information \( I_{\beta_0} \) in the usual sense. That is, for any other information matrix \( I \), the matrix difference \( I_{\beta_0} - I \) is always negative semidefinite. It follows from van der Vaart (1989, Theorem 2.5, Page 1491), that \( I_{\beta_0}^{-1} \) is a valid lower bound for the asymptotic variance of regular estimators of \( \beta_0 \). Furthermore, Theorem 2.1 and Theorem 2.3 together imply that to find the least favorable direction in the multiparameter case, we simply find the least favorable direction corresponding to each component of the parameter vector. That is, these results also show that finding the lower efficiency bound for a \( p \) dimensional parameter of interest is equivalent to solving \( p \) individual optimization problems.

Let us now compute the efficiency bounds for \( \beta_0 \) in the partially linear model. As before, let \( (\beta, \eta_0) \) be a smooth curve in \( \mathcal{B} \times \text{lin}\mathcal{F} \), such that \( \eta_{\beta}|_{\beta=\beta_0} = f^* \). The vector \( (\beta, \eta_0) \) is often called a parametric submodel. The word parametric here refers to the fact that since the nonparametric part is now indexed by \( \beta \), the estimation problem is restricted to finite dimensional or parametric space. The term submodel simply means that \( (\beta, \eta_0) \) is just one of the several parameterizations that may be chosen. Notice that the parametric submodel passes through the truth when \( \beta = \beta_0 \). Assuming that the data is generated by this parametric submodel, the loglikelihood function is

\[
(2.1) \quad \ell(\beta, \eta_0|x, y, z) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} [y - x \cdot \beta - \eta_0(z)]^2 + \log g_{x,z}(x, z)
\]
When $\beta = \beta_0$, this likelihood function equals the true likelihood. The score function for $\beta_0$ is

$$S_{\beta_0} = \frac{d\ell(\beta_0, \eta_{\beta_0})}{d\beta} = \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta}(\frac{d}{d\beta} \eta_{\beta_0}).$$

Even though $S_{\beta_0}$ is a $p \times 1$ vector it suffices to look at the componentwise scores, as discussed previously. Now it may be easily seen that

(i) the score w.r.t. $\beta_0$, $\frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} = \epsilon x$;

(ii) while the Fréchet derivative $\frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} = \epsilon$.

Therefore, $S_{\beta_0} = -\epsilon [x_i + \frac{d}{d\beta} \eta_{\beta_0}]$ where, $i = 1, \ldots, p$. Furthermore, the least favorable direction in $\text{lin} T(\mathcal{F}, f^*)$ for estimating $\beta_0$ is given by

$$\delta_i^* = \arg\min_{\xi \in \text{lin} T(\mathcal{F}, f^*)} \mathbb{E} [x_i + \xi]^2 = -\text{proj}(x_i | \text{lin} T(\mathcal{F}, f^*)), \quad (2.2)$$

Hence the efficient score for computing the semiparametric efficiency bounds of $\beta_0$ is given by $\hat{S} = \epsilon \begin{pmatrix} x_1 - \delta_1^* \\ \vdots \\ x_p - \delta_p^* \end{pmatrix}$. The matrix $(\mathbb{E} \hat{S} \hat{S}^t)^{-1}$ then gives the semiparametric efficiency bounds for regular estimators of $\beta_0$. Therefore, to determine the efficiency bounds we simply have to solve the projection problem in Equation (2.2).

**Remark 2.3 (Solving the Projection Problem).** One important point about calculating the projection in Equation (2.2) needs to be made evident. The optimization problem in Equation (2.2) requires that we find a function in $\text{lin} T(\mathcal{F}, f^*)$ that is closest (in $L^2$ norm) to $x_i$. But $\text{lin} T(\mathcal{F}, f^*)$ is not a Hilbert space. So how can we use the Classical Projection Theorem (applicable only to Hilbert spaces) to solve this problem? The answer is simple. First solve this optimization problem over the bigger space $L^2$; i.e. find a $\delta_i^*(z) \in L^2$ that solves Equation (2.2). Secondly, show that $\delta_i^*(z)$ is also an element of $\text{lin} T(\mathcal{F}, f^*)$. This will clearly imply that $\delta_i^*(z)$ is indeed the solution to Equation (2.2). This technique of working in $L^2$ and then showing that the projection also satisfies certain additional restrictions has considerable technical advantages. It allows us to use results valid for Hilbert spaces, instead of working with $C^2$ functions throughout.

3. PRELUDE TO THE IMPOSSIBILITY RESULT

**Notation 3.1.** Let $\mathcal{F}_h \subset \mathcal{H}$ be the set of all homogeneous functions of degree $r$, where $r$ is known. Also, let $\mathcal{F}_c \subset \mathcal{H}$ denote the set of all concave functions.

Until now we had simply assumed that $\mathcal{F}$ was a closed convex cone in $\mathcal{H}$. This certainly includes the case when $\mathcal{F}$ is a linear space, as for instance, when $\mathcal{F} = \mathcal{F}_h$. In this case, $\mathcal{F}_h = T(\mathcal{F}_h, f^*) = \text{lin} T(\mathcal{F}_h, f^*)$ are all linear spaces and calculating the efficiency bounds is relatively straightforward (see Section 5). However, as described in the introduction, typical shape restrictions that one encounters in microeconomics include restrictions like concavity and monotonicity. But the set of all concave
(or monotone) functions is not a linear space but a convex cone. An immediate consequence of the fact that \( F_\epsilon \) is a proper convex cone is that \( T(F_\epsilon, f^*) \) is also a convex cone and not, in general, a linear space. As we shall soon discover, when \( T_1(F_\epsilon, f^*) \) is a proper cone, a projection onto \( T(F_\epsilon, f^*) \) in Equation (2.2) (instead of a projection onto \( \text{lin} T(F_\epsilon, f^*) \)) may lead to a better lower bound for estimating \( \beta_0 \).

However, if we want this lower bound to be attained by a regular \( n^{1/2} \) consistent estimator of \( \beta_0 \), the projection in Equation (2.2) must be taken on \( \text{lin} T(F, f^*) \) and not on \( T(F, f^*) \). This may be seen as follows.

**Assumption 3.1.** For the remainder of this section assume that \( F \subset H \) is a proper convex cone and that \( f^* \in F \). \( \square \)

To simplify exposition, assume for the moment that we are only dealing with a scalar parameter of interest; i.e. \( \beta_0 \in \mathbb{R} \). Now let \( \gamma_\beta \) be a smooth curve in \( \text{lin} \bar{F} \) such that \( \gamma_{\beta_0} = f^* \), and define

\[
i_1 = \min_{\zeta \in \text{lin} T(F, f^*)} \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \gamma_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \gamma_{\beta_0})}{\partial \gamma} (\zeta) \right]^2.
\]

Again, let \( \lambda_\beta \) be a smooth curve in \( F \) such that \( \lambda_{\beta_0} = f^* \), and define

\[
i_2 = \min_{\delta \in \text{lin} T(F, f^*)} \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \lambda} (\delta) \right]^2.
\]

Note that \( i_1 \) (resp. \( i_2 \)) is just the information obtained by projecting the scores w.r.t. the parameters of interest onto \( \text{lin} T(F, f^*) \) (resp. \( T(F, f^*) \)). Since \( T(F, f^*) \subseteq \text{lin} T(F, f^*) \), it is clear that \( i_1 \leq i_2 \). Now if \( i_1 = i_2 \), it really does not matter on which space we take the projection since the lower bound will be the same in both the cases. Therefore, the only interesting case is when \( i_1 < i_2 \). In this case, there is a real possibility that projecting the parametric scores onto the tangent cone \( T(F, f^*) \) will yield a lower bound \( i_2^{-1} \) which is strictly better than the lower bound \( i_1^{-1} \), obtained by a projection onto \( \text{lin} T(F, f^*) \). What this means is that when \( T(F, f^*) \) is a proper cone, we may achieve gains in efficiency by projecting onto \( T(F, f^*) \), instead of \( \text{lin} T(F, f^*) \). However, as it turns out, the promise of this gain in efficiency may just be illusory. We will now construct two parametric examples to show that when the tangent cone is a proper cone, a projection onto \( T(F, f^*) \) will actually lead to lower bounds that are either

(i) too optimistic for the m.l.e. of \( \beta_0 \) (i.e. unattainable by the m.l.e.) or,

(ii) which are actually beaten by the m.l.e. (i.e. are invalid bounds).

These parametric examples provide intuition for the following result in the semi-parametric case: When \( T(F, f^*) \) is a proper cone, efficiency bounds obtained by projecting the scores w.r.t. the parameter of interest onto \( T(F, f^*) \) will not be attainable by any regular \( n^{1/2} \) consistent estimator of \( \beta_0 \). We call this result an "impossibility" result. This impossibility result will be used to show that in the class of all regular \( n^{1/2} \) consistent estimators of \( \beta_0 \), concavity and monotonicity (unlike homogeneity) of \( f^* \) do not lead to any efficiency gains in estimating \( \beta_0 \).
Since it follows from van der Vaart (1989) that projections taken on \( \text{lin} T(F, f^*) \) yield valid lower bounds, the space on which the parametric scores should be projected to obtain efficiency bounds for \( \beta_0 \) is \( \text{lin} T(F, f^*) \), and not \( T(F, f^*) \).

\textbf{Remark 3.1.} In Equation (3.2), let \( \delta^* \in T(F, f^*) \) be the tangent vector that achieves \( i_2 \). Then following Definition 2.7, \( \delta^* \) is called the least favorable direction in \( T(F, f^*) \) for estimating \( \beta_0 \). Now let \( \eta_{\beta} \) be a curve through \( f^* \) such that \( \frac{d}{d \beta} \eta_{\beta_0} = \delta^* \). Again following Definition 2.7, \( \eta_{\beta} \) is called a least favorable curve in \( F \) for estimating \( \beta_0 \). Similarly, for the multidimensional case, we can define a least favorable surface (resp. direction) in \( F \) (resp. \( T(F, f^*) \)) for estimating \( \beta_0 \). In Theorems 2.1 – 2.3 if \( \text{lin} F \) (resp. \( \text{lin} T(F, f^*) \)) is replaced by \( F \) (resp. \( T(F, f^*) \)), the following analogous results are obtained. As Theorems 3.1 – 3.3 deal with proper cones \( F \) (resp. \( T(F, f^*) \)) instead of \( \text{lin} F \) (resp. \( \text{lin} T(F, f^*) \)), their proofs are obtained by using Theorem B.2 in Appendix B instead of the classical projection theorem. Since the proofs of these theorems are very similar to the the proofs of Theorems 2.1 – 2.3, they have been omitted from this paper. □

**Theorem 3.1.** Let \( \eta_{\beta} \) be a least favorable surface in \( F \) for estimating \( \beta_0 \). Denote the least favorable direction in \( T(F, f^*) \) by \( \delta^* = \left( \frac{d}{d \beta} \eta_{\beta_0} \right)_{\beta = \beta_0} \); i.e., \( \delta^*_i = \left( \frac{d}{d \beta_i} \eta_{\beta} \right)_{\beta = \beta_0} \) for \( i = 1, \ldots, p \). Then \( \delta^* = (\delta^*_1, \ldots, \delta^*_p) \) satisfies

(i) \( E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \frac{d}{d \beta} \eta_{\beta_0} \right) \right] \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta^*_i) = 0 \) and,

(ii) \( E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \frac{d}{d \beta} \eta_{\beta_0} \right) \right] \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta) \geq 0 \), for all \( \delta \in T(F, f^*) \).

**Theorem 3.2.** Let \( \eta_{\beta} \) be a smooth curve such that \( \eta_{\beta_0} = f^* \). Then \( \eta_{\beta} \) is a least favorable surface in \( F \) for estimating \( \beta_0 \) if and only if

(i) for all \( \delta_i \in T(F, f^*) \)

\[
\sum_{i=1}^{p} E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \frac{d}{d \beta} \eta_{\beta_0} \right) \right] \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta_i) \geq 0,
\]

(ii) and, if \( \delta_i = (\frac{d}{d \beta_i} \eta_{\beta_0})_{\beta = \beta_0} \)

\[
\sum_{i=1}^{p} E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \frac{d}{d \beta} \eta_{\beta_0} \right) \right] \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \frac{d}{d \beta} \eta_{\beta_0} \right) = 0.
\]

**Theorem 3.3.** Let

\[
J = E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \frac{d}{d \beta} \eta_{\beta_0} \right) \right] \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \left( \frac{d}{d \beta} \eta_{\beta_0} \right) \right]^t
\]

be the information matrix when \( \eta_{\beta} \) is smooth curve in \( F \) such that \( \eta_{\beta_0} = f^* \). Then there exists a matrix \( J_{\beta_0} \) such that \( \alpha' (J_{\beta_0} - J) \alpha \leq 0 \) for all \( \alpha \in \mathbb{R}^p \) if and only if \( J_{\beta_0} \) corresponds to the information matrix when \( \eta_{\beta} \) is a least favorable surface in \( F \) for estimating \( \beta_0 \).
To gain insight into our impossibility result we now impose monotonicity and convexity on nuisance parameters in simple linear regression, and see how the imposition of such a shape restriction affects the efficiency bounds for the parameter of interest.

**Example 3.1 (Imposing Monotonicity).** Consider the following linear regression

\[ y_i = \theta_0 + \lambda_0 z_i + \epsilon_i, \quad i = 1, \ldots, n. \]

Here \( \epsilon \stackrel{\text{i.i.d.}}{\sim} \text{NIID}(0,1) \), and \( z \) is a random variable with positive variance that is independent of \( \epsilon \). The parameter of interest is \( \theta_0 \in \mathbb{R} \), and the nuisance parameter is \( \lambda_0 \in \Lambda = [0, \infty) \); i.e. we want to fit an increasing line to the i.i.d. observations \((y, z)\). Following the procedure in Section 2, it is easy to obtain the results in Table 1.

**Table 1. Lower Bounds for Estimating \( \theta_0 \)**

<table>
<thead>
<tr>
<th>Nuisance Parameter</th>
<th>( \mathcal{T}(\Lambda, \lambda_0) )</th>
<th>Lower Bound for ( \theta_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_0 = 0 )</td>
<td>([0, \infty))</td>
<td>(1 / \mathbb{E} \left[ 1 - Z \min {0, (\mathbb{E} Z)/\mathbb{E} Z^2 } \right]^2)</td>
</tr>
<tr>
<td>( \lambda_0 &gt; 0 )</td>
<td>((-\infty, \infty))</td>
<td>(\mathbb{E} Z^2 / (\text{Var} Z))</td>
</tr>
</tbody>
</table>

Notice that the efficiency bound depends upon the true value of the nuisance parameter \( \lambda_0 \). Let us now see if the m.l.e. of \( \theta_0 \) achieves this bound. So define \( S_{zz} = \sum_{i=1}^{n} (z_i - \bar{z}_n)^2 \), \( \bar{z}_n = \frac{1}{n} \sum_{i=1}^{n} z_i \), \( z' z = \sum_{i=1}^{n} z_i^2 \), and let \( \hat{\theta}_n \) denote the m.l.e. of \( \theta_0 \). Then,

\[ \hat{\theta}_n = \begin{cases} 
\bar{y}_n - \hat{\lambda}_n \bar{z}_n & \text{if } \hat{\lambda}_n \geq 0 \\
\bar{y}_n & \text{if } \hat{\lambda}_n < 0,
\end{cases} \]

with \( \hat{\lambda}_n = \sum_{i=1}^{n} (z_i - \bar{z}_n) y_i / S_{zz} \). Let \( \phi(\cdot) \) denote the p.d.f., and \( \Phi(\cdot) \) the c.d.f. of a standard Normal random variable. Then after some tedious algebra it may be shown that conditional on observing \( z_1, \ldots, z_n \),

\[
\Pr\{n^{1/2} |\hat{\theta}_n - \theta_0| \leq t\} = \int_{u = -\infty}^{t \sqrt{S_{zz} / \mathbb{E} Z^2}} \frac{e^{-u^2/2}}{\sqrt{2\pi}} \Phi(n^{1/2} \lambda_0 \sqrt{z' z / n - \bar{z}_n \sqrt{n / S_{zz} u}}) \, du \\
+ \Phi(-n^{1/2} \lambda_0 \sqrt{S_{zz} / n}) \Phi(t - n^{1/2} \lambda_0 \bar{z}_n).
\]

Now letting \( F(w) = \int_{u = -\infty}^{w} \phi(u) \Phi(-u \sqrt{\mathbb{E} Z / \mathbb{E} Z^2}) \, du \), we can show that

\[
\Pr\{n^{1/2} |\hat{\theta}_n - \theta_0| \leq t\} \rightarrow \begin{cases} 
\frac{1}{2} \Phi(t) + F(t) & \text{if } \lambda_0 = 0 \\
\Phi(t \sqrt{\text{Var} \bar{z} / \mathbb{E} Z^2}) & \text{if } \lambda_0 > 0,
\end{cases}
\]
and that the asymptotic variance

$$\text{AsVar}(n^{1/2}(\hat{\theta}_n - \theta_0)) = \begin{cases} \frac{1}{2}(1 + \frac{\mathbb{E} Z^2}{\text{Var} Z}) - \frac{1}{2\pi} \frac{(\mathbb{E} Z)^2}{\text{Var} Z} & \text{if } \lambda_0 = 0 \\ \frac{\mathbb{E} Z^2}{\text{Var} Z} & \text{if } \lambda_0 > 0. \end{cases}$$

Notice that when $\mathbb{E} Z \neq 0$

$$1 < \frac{1}{2}(1 + \frac{\mathbb{E} Z^2}{\text{Var} Z}) - \frac{1}{2\pi} \frac{(\mathbb{E} Z)^2}{\text{Var} Z} < \frac{\mathbb{E} Z^2}{\text{Var} Z}$$

and, the asymptotic variance is not continuous at $\lambda_0 = 0$. Also, the asymptotic distribution of the m.l.e. is not normal at $\lambda_0 = 0$. All this algebra yields the following results, which are summarized in Table 2.

**Table 2. Imposing Monotonicity in Linear Regression**

<table>
<thead>
<tr>
<th>$\mathbb{E} Z$</th>
<th>$\lambda_0$</th>
<th>Lower Bound for $\theta_0$</th>
<th>$\hat{\theta}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{E} Z = 0$</td>
<td>$\lambda_0 = 0$</td>
<td>Attained by the m.l.e.</td>
<td>Not Regular</td>
</tr>
<tr>
<td>$\mathbb{E} Z &gt; 0$</td>
<td>$\lambda_0 = 0$</td>
<td>Not attained by the m.l.e.</td>
<td>Not Regular</td>
</tr>
<tr>
<td>$\mathbb{E} Z &lt; 0$</td>
<td>$\lambda_0 = 0$</td>
<td>Beaten by the m.l.e.</td>
<td>Not Regular</td>
</tr>
<tr>
<td>$\mathbb{E} Z = 0$</td>
<td>$\lambda_0 &gt; 0$</td>
<td>Attained by the m.l.e.</td>
<td>Regular</td>
</tr>
<tr>
<td>$\mathbb{E} Z &gt; 0$</td>
<td>$\lambda_0 &gt; 0$</td>
<td>Attained by the m.l.e.</td>
<td>Regular</td>
</tr>
<tr>
<td>$\mathbb{E} Z &lt; 0$</td>
<td>$\lambda_0 &gt; 0$</td>
<td>Attained by the m.l.e.</td>
<td>Regular</td>
</tr>
</tbody>
</table>

(i) When $\lambda_0 = 0$, the efficiency bounds are attained only when $\mathbb{E} Z = 0$. When $\mathbb{E} Z \neq 0$, the bound is either not attained (when $\mathbb{E} Z > 0$), or is actually beaten by $\hat{\theta}_n$ (when $\mathbb{E} Z < 0$). However, it may be shown that in all these cases the estimator $\hat{\theta}_n$ is not regular.

(ii) When $\lambda_0 > 0$, not only is $\hat{\theta}_n$ regular but it also attains the lower bound. Also, when $\lambda_0 > 0$ the tangent cone $T(\Lambda, \lambda_0) = (-\infty, \infty)$. But $(-\infty, \infty)$ is also the smallest closed linear space containing $[0, \infty)$. Hence, if we restrict ourselves to the class of regular estimators, the space on which the projection is taken to obtain the efficiency bound should be $\text{lin} T(\Lambda, \lambda_0)$, rather than the tangent cone itself. Projection on this larger space will lead to bounds that can be attained by regular estimators. To do any better, we would have to use an estimator that is not regular. □

**Example 3.2 (Imposing Convexity).** Imposing convexity in linear regression is easily done by substituting $Z = X^2$ in the previous example. Thus the shape restricted regression now becomes, $y_i = \theta_0 + \lambda_0 z_i^2 + \epsilon_i$ under the same conditions as before. The restriction $\lambda_0 \geq 0$ now implies that we are fitting a convex function to the data. We now have the following results which are stronger than those obtained in the previous example.

As before, when $\lambda_0 > 0$, the efficiency bound is attained by $\hat{\theta}_n$. However, when $\lambda_0 = 0$, the efficiency bound is not attained. These results once again show that to obtain efficiency bounds which are attainable by regular estimators, the projection must be taken on $\text{lin} T(\Lambda, \lambda_0)$, rather than the tangent cone $T(\Lambda, \lambda_0)$ itself. □
Table 3. Imposing Convexity in Linear Regression

<table>
<thead>
<tr>
<th>$\lambda_0$</th>
<th>Lower Bound for $\theta_0$</th>
<th>$\theta_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0$</td>
<td>Not attained by the m.l.e.</td>
<td>Not Regular</td>
</tr>
<tr>
<td>$\lambda_0 &gt; 0$</td>
<td>Attained by the m.l.e.</td>
<td>Regular</td>
</tr>
</tbody>
</table>

4. The Impossibility Theorem

As the parametric examples in Section 3 demonstrate, a projection on $T(\mathcal{F}, f^*)$, when $T(\mathcal{F}, f^*)$ is a proper cone, leads to efficiency bounds that are unattainable by the m.l.e. of $\beta_0$. In this section we make this argument rigorous for semiparametric models. We will now show that if efficiency bounds for estimating $\beta_0$ are obtained by projecting the parametric scores onto a proper tangent cone, then no regular $n^{1/2}$-consistent estimator of $\beta_0$ can achieve these bounds. This result will be shown to hold under the following condition.

Condition 4.1 (Condition for Impossibility Result). Let $\gamma_\beta$ be a smooth curve in $\overline{\text{lin}\mathcal{F}}$ such that $\gamma_{\beta_0} = f^*$, and define

$$I_1 = \inf_{\zeta \in \overline{\text{lin}\mathcal{T}(\mathcal{F}, f^*)}} \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \gamma_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \gamma_{\beta_0})}{\partial \gamma}(\zeta) \right] \left[ \frac{\partial \ell(\beta_0, \gamma_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \gamma_{\beta_0})}{\partial \gamma}(\zeta) \right]' .$$

Again, let $\lambda_\beta$ be a smooth curve in $\mathcal{F}$ such that $\lambda_{\beta_0} = f^*$ and define the matrix $I_2$ as follows.

$$I_2 = \inf_{\zeta \in \overline{\text{lin}\mathcal{T}(\mathcal{F}, f^*)}} \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \lambda}(\zeta) \right] \left[ \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \lambda}(\zeta) \right]' .$$

Then $I_1 - I_2$ is negative definite.

The infima in Condition 4.1 are taken w.r.t. the usual order on the space of all $p \times p$ matrices. From Theorem 2.3 and Theorem 3.3, $I_1$ and $I_2$ exist if $\gamma_{\beta_0}$ (resp. $\lambda_{\beta_0}$) is a least favorable curve in $\overline{\text{lin}\mathcal{F}}$ (resp. $\mathcal{F}$). Therefore, a necessary condition for this assumption to hold is that the least favorable directions in $\overline{\text{lin}\mathcal{T}(\mathcal{F}, f^*)}$ and $\times_{i=1}^p T(\mathcal{F}, f^*)$ be different. This condition certainly holds for the parametric examples in the previous section. It may also hold for the semiparametric models under study. To see this, consider the following example.

Example 4.1 (Tangent Cone to Set of Concave Functions). Let $\mathcal{F}$ be the set of all concave functions in $\mathcal{H}$ and, let $f^* \in \mathcal{F}$ be concave but not strictly concave. Since $f^*$ is not strictly concave, there exists a nonempty set $Z_0 \subseteq Z$ on which the Hessian $\nabla^2 f^*$ vanishes, while on $Z - Z_0$ the Hessian matrix $\nabla^2 f^*$ is
negative definite. \(^2\) Now let
\[ \mathcal{W} = \{ f \in \mathcal{H} : f \text{ is concave on } Z_0 \subseteq Z \}. \]

Since the Hessian of \( f \in \mathcal{W} \) is negative semi-definite on \( Z_0, \mathcal{F}_c \subseteq \mathcal{W} \). If \( Z_0 = Z \), \( f^* \) is affine and thus \( \mathcal{W} = \mathcal{F}_c \). A function could be strictly convex and still be in \( \mathcal{W} \) if its Hessian vanishes on \( Z_0 \). For instance, the function \( (x, y) \mapsto x^4 + y^4 \) is strictly convex but its Hessian is zero at \((0, 0)\). Hence if \( f^* \) is chosen such that \( Z_0 = \{(0,0)\} \), then the function \( (x, y) \mapsto x^4 + y^4 \) is an element of \( \mathcal{W} \). This shows that \( \mathcal{W} \) is a closed convex cone and not a linear space, since all strictly concave functions are in \( \mathcal{W} \) while some strictly convex functions are not. Now following the details in Appendix D, we can show the following results.

**Lemma 4.1.** Let \( f^* \in \mathcal{F}_c \) be concave but not strictly concave. Then \( T(\mathcal{F}_c, f^*) = \mathcal{W} \).

**Proof.** See Appendix D. \( \square \)

**Lemma 4.2.** Let \( f^* \in \mathcal{F}_c \). Then \( \text{lin} T(\mathcal{F}_c, f^*) = \mathcal{H} \).

**Proof.** See Appendix D. \( \square \)

When \( f^* \) is concave but not strictly concave, these results show that \( I_2 \) (as defined in Condition 4.1) is achieved by the least favorable direction \( \xi^* \) in \( \mathcal{W} \). Similarly, \( I_1 \) is achieved by the least favorable direction \( \zeta^* \) in \( \text{lin} T(\mathcal{F}_c, f^*) \), which is just a \( C^2 \) function. It also follows from Equation (2.2) that for the partially linear model, \( \zeta^* \) is just the projection of \( x \) onto \( \text{lin} T(\mathcal{F}_c, f^*) \).

**Lemma 4.3.** Let \( \zeta^*(z) = -\text{proj}(x|\text{lin} T(\mathcal{F}_c, f^*)) \). Then \( \zeta^*(z) = -E(x|z) \).

**Proof.** See Appendix D. \( \square \)

From Lemma 4.3 we can see that \( \zeta^*(z) \) is just the usual conditional expectation, subject to some smoothness restrictions. Since, without imposing additional restrictions, this conditional expectation is not concave on \( Z_0 \), we have that \( \zeta^* \not\in \mathcal{W} \). Therefore, the two least favorable directions are different and this implies that \( I_1 - I_2 \) is negative definite. Thus Condition 4.1 holds. We are only interested in the case of \( f^* \in \text{bdry}(\mathcal{F}_c) \) because if \( f^* \in \text{int}(\mathcal{F}_c) \), which happens when \( f^* \) is strictly concave, then \( T(\mathcal{F}_c, f^*) = \text{lin} T(\mathcal{F}_c, f^*) = \mathcal{H} \) and the analysis is straightforward. \( \square \)

We now come to the main result of this section.

**Theorem 4.1 (Impossibility Theorem).** Let \( \mathcal{F} \) be a convex cone in \( \mathcal{H} \), and let \( f^* \in \text{bdry}(\mathcal{F}) \) be such that the tangent cone \( T(\mathcal{F}, f^*) \) is a proper cone. Then under Condition 4.1, no regular \( n^{1/2} \)-consistent estimator of \( \beta_0 \) can achieve the efficiency bounds obtained by projecting the scores w.r.t. the parameter of interest onto the tangent cone \( T(\mathcal{F}, f^*) \).

**Proof.** See Appendix E. \( \square \)

\(^2\)Geometrically, the fact that \( f^* \) is concave but not strictly concave means that \( f^* \in \text{bdry}(\mathcal{F}_c) \).
To see the utility of this impossibility result, consider Example 4.1 once again. Since Condition 4.1 holds, Theorem 4.1 implies that projecting the parametric scores onto $T(F, f^*)$ will lead to efficiency bounds that are unattainable by any regular $n^{1/2}$-consistent estimator of $\beta_0$. To obtain attainable efficiency bounds we have to project onto $\text{lin } T(F, f^*)$ which is just $\mathcal{H}$. Using Lemma 4.2 and Lemma 4.3, the efficient score for estimating $\beta_0$ is given by

$$\bar{S} = \varepsilon \begin{pmatrix} X_1 - \mathbb{E}(X_1 | Z) \\ \vdots \\ X_p - \mathbb{E}(X_p | Z) \end{pmatrix}.$$ 

But this is the same efficient score vector that would be obtained if the only thing we knew about $f^*$ was that it was twice continuously differentiable. Hence, in the class of $n^{1/2}$-consistent regular estimators of $\beta_0$, computing efficiency bounds for $\beta_0$ when $f^*$ is concave is equivalent to computing efficiency bounds for $\beta_0$ when $f^*$ is just a $C^2(Z)$ function. Therefore, concavity of $f^*$ does not help us in estimating $\beta_0$ more efficiently. Even though we have not verified them in this paper, analogous results should hold for the case when $f^*$ is monotone. This follows because while concavity implies a restriction on the second derivative of $f^*$, monotonicity will impose a restriction on the first derivative of $f^*$.

The impossibility theorem says that if $T(F, f^*)$ is a proper cone, a $n^{1/2}$ consistent regular estimator of $\beta_0$ will not achieve the efficiency bounds obtained by projecting the scores w.r.t. the parameters of interest onto $T(F, f^*)$. However, this result leaves open the possibility of a non-regular estimator of $\beta_0$ attaining such bounds. Construction of a non-regular efficient estimator has to be dealt on a case by case basis. The degree of difficulty in constructing non-regular yet efficient estimators obviously depends upon the problem being studied. For instance, let us look at Example 4.1 once again. If we want gains from concavity, we must find a non-regular estimator that achieves the efficiency bounds for Example 4.1. The first step in this case is to obtain the bounds themselves. This requires a projection onto $T(F, f^*) = \mathcal{W}$ which, to the best of my knowledge, does not have a closed form solution. However, it may still be possible to develop an algorithm that allows us to determine a projection iteratively. After obtaining this projection we calculate the lower bound and, finally, construct an estimator with asymptotic variance equal to this bound. From Theorem 4.1 we know that this estimator will not be regular. As far as I know, this is still an open problem and one that merits further research.

5. Efficiency Bounds Under Homogeneity

In this section we will calculate efficiency bounds for estimating $\beta_0$ in the partially linear model, when the only thing that we know about $f^*$ is that it is homogeneous of degree $r$, where $r$ is known; i.e. $f^* \in F_h$, where $F_h \subset \mathcal{H}$ is the space of all homogeneous functions of degree $r$. Since $F_h$ is a linear space, $F_h = T(F_h, f^*) = \text{lin } T(F_h, f^*)$ and, obtaining the efficiency bounds requires a projection onto the space of homogeneous functions. In this section we will also show how to obtain such
a projection. Determining this projection has an additional fringe benefit. It allows us to develop kernel estimators of homogeneous functions. One such estimator is used in Tripathi (1997a) to develop a test for homogeneity of functional form.

However, merely knowing the efficiency bounds is not enough. To be of any use, these bounds must be attainable. That is, we must be able to construct an estimator of $\beta_0$ with asymptotic variance $(E S \hat{S} S')^{-1}$, where $\hat{S}$ is given in Equation (5.1). To keep our presentation concise, we have excluded the estimation part of the problem from this paper. In Tripathi (1997b), following the approach of Severini and Wong (1992), we construct such an estimator and also show that it is efficient; i.e. it attains the efficiency bounds.

**Remark 5.1.** When $f^*$ is homogeneous of degree $r > 0$, $f^*(0) = 0$. In this case even the intercept term in $\beta_0$ is identified. □

So let us determine the efficiency bounds for estimating $\beta_0$, when $f^*$ is homogeneous of degree $r$. Following Equation (3.1), we simply have to project the scores w.r.t. the parameters of interest onto $\text{lin} T(F_h, f^*) = F_h$. Since the score w.r.t. to $\beta_{i0}$ is just $X_i$, we have the following result.

**Lemma 5.1.** The projection of $X_i$ onto $F_h$ is the function

$$
\delta_i^*(z) = -\frac{z_i^r E(X_i \gamma_{i0}^* | \gamma_{i1}^* = \frac{z_1}{\gamma_0}, \ldots, \gamma_{ip-1}^* = \frac{z_{p-1}}{\gamma_0})}{E(Z_i^2 | \gamma_{i1}^* = \frac{z_1}{\gamma_0}, \ldots, \gamma_{ip-1}^* = \frac{z_{p-1}}{\gamma_0})}.
$$

**Proof.** See Appendix F. □

Using Theorem 5.1, the efficient score $\hat{S}$ for estimating $\beta_0$ is

$$
\hat{S} = \varepsilon \left( \begin{array}{c}
\frac{z_1^r E(X_1 \gamma_{10}^* | \gamma_{11}^* = \frac{z_1}{\gamma_0}, \ldots, \gamma_{1p-1}^* = \frac{z_{p-1}}{\gamma_0})}{E(Z_1^2 | \gamma_{11}^* = \frac{z_1}{\gamma_0}, \ldots, \gamma_{1p-1}^* = \frac{z_{p-1}}{\gamma_0})} \\
\vdots \\
\frac{z_p^r E(X_p \gamma_{p0}^* | \gamma_{p1}^* = \frac{z_1}{\gamma_0}, \ldots, \gamma_{p1}^* = \frac{z_{p-1}}{\gamma_0})}{E(Z_p^2 | \gamma_{p1}^* = \frac{z_1}{\gamma_0}, \ldots, \gamma_{ip-1}^* = \frac{z_{p-1}}{\gamma_0})}
\end{array} \right).
$$

As mentioned before, the matrix $(E \hat{S} S')^{-1}$ then gives the semiparametric efficiency bounds for estimating $\beta_0$ when the true function $f^*$ is homogeneous of degree $r$.

A natural question at this point is to inquire about the gain in efficiency obtained by imposing the shape restriction of homogeneity. The following example provides an interesting case in point.

**Example 5.1 (Homogeneity Increases Efficiency).** Let $(\beta, z) \in \mathbb{R}^2 \times \mathbb{R}^2$. Our model is then $y = x_1 \beta_1 + x_2 \beta_2 + f^*(z_1, z_2) + \varepsilon$, with $\varepsilon \overset{d}{=} \mathcal{N}(0, 1)$. To simplify matters even further, let $z_1 = z_2 = z$, and let $x = (x_1, x_2)$ be completely predictable by $z$. Say for instance, $x_1 = z^2$ and $x_2 = z^3$. The model then reduces to

$$
y = z^2 \beta_1 + z^3 \beta_2 + f^*(z, z) + \varepsilon.
$$

Now consider the following two cases.
**Case I:** No shape restrictions on \( f^* \). That is, just assume that \( f^* \in \mathcal{H} \). We claim that in this case, \((\beta_1, \beta_2, f^*)\) is not identified. To see this, let \(g\) be the joint density of \((y, z)\) and define \(S_1 \equiv (\beta_1, \beta_2, f^*)\) and, \(S_2 \equiv (\alpha_1, \alpha_2, h^*)\). If we can now show that there exist structures \(S_1, S_2\) (with \(S_1 \neq S_2\)) such that \(g(y, z; S_1) = g(y, z; S_2)\), then \((\beta_1, \beta_2, f^*)\) is not identified. So let \(S_1 = (1, 0, f^*(z, z))\) and \(S_2 = (1, 1, f^*(z, z) - z^3)\). Clearly \(S_1 \neq S_2\), and since \(g(y, z) = g(y | z)g(z)\) we have

\[
g(y, z; S_1) = (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} (y - z^2 - z^3 \cdot 0 - f^*(z, z))^2 \right\} g(z)
= (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} (y - z^2 - f^*(z, z))^2 \right\} g(z)
\]

and,

\[
g(y, z; S_2) = (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} (y - z^2 - z^3 - f^*(z, z) + z^3)^2 \right\} g(z)
= (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} (y - z^2 - f^*(z, z))^2 \right\} g(z).
\]

Therefore, \(g(y, z; S_1) = g(y, z; S_2)\) and so \((\beta_1, \beta_2, f^*)\) is not identified. It is not difficult to see that this also implies that both \((\beta_1, \beta_2)\) and \(f^*\) are separately not identified. Due to this lack of identification, the lower bound for the variance of any estimator of \((\beta_1, \beta_2)\) is \((\infty, \infty)\).

**Case II:** Now let \( f^* \) be homogeneous of degree 1. Since \( f^* \) is homogeneous of degree 1,

\[
y = z^2 \beta_1 + z^3 \beta_2 + z f^*(1, 1) + \epsilon.
\]

But this clearly shows that in this case \((\beta_1, \beta_2, f^*(1, 1))\) is identified. Therefore, the efficiency bounds for estimators of \(\beta_1\) and \(\beta_2\) are finite. Hence, by imposing a shape restriction on \( f^* \) we can identify the finite dimensional parameters and achieve a dramatic gain in efficiency. \(\square\)

Even though this example has an artificial flavor, it illustrates the potential gains in efficiency that may be obtained by imposing shape restrictions. An interesting exercise here would be to examine what shape restrictions on \( f^* \), besides homogeneity, allow us to identify \( \beta \). Notice that this example also illustrates the strength of homogeneity as a shape restriction. Here is another example demonstrating that large gains in efficiency are possible under homogeneity, even when the parameter of interest is identified.

**Example 5.2 (Another Simple Example).** For \((\beta_0, z) \in \mathbb{R} \times \mathbb{R}\), and \(\epsilon \equiv N(0, 1)\), let \(y = x \beta_0 + f^*(z) + \epsilon\), where \(x, z \equiv \text{UID}(0, 1)\), and \(f^*\) is linearly homogeneous. Since \(f^*\) is homogeneous of degree one, \(y = x \beta_0 + z f^*(1) + \epsilon\) and the asymptotic lower bound for \(\beta_0\) can be shown to be 8.4. However, if homogeneity is not imposed upon \(f^*\) it is easy to see that \(\beta_0\) still remains identified, but the lower bound increases to 12. Therefore, the asymptotic relative efficiency of the estimator under homogeneity w.r.t. the estimator when homogeneity is not imposed is \(\frac{12}{8.4} = 1.428\). Thus the loss in efficiency by not imposing homogeneity, when \(f^*\) is truly homogeneous, is 42.8%. \(\square\)
6. Conclusion

Recent trends clearly indicate the growing popularity of semiparametric techniques in econometrics. As econometricians incorporate restrictions of economic theory in these techniques, they will gain even wider acceptance among applied economists. This paper is a step in this direction, namely, the integration of economic theory with econometric practice. Hopefully, it will be a stepping stone to a general theory of efficient semiparametric estimation under shape restrictions. Such a theory will be obtained when the class of shape restriction is extended to include all popular restrictions imposed by economic theory on unknown functions. However, in this paper we have only concentrated upon the two basic shape restrictions of concavity and homogeneity.

Appendix A. Tangent Cones

In this section we collect some results about tangent cones. These results are available in standard mathematical literature but seem to be scattered all over the place. We begin with a definition from Krabs (1979, Page 154).

Definition A.1 (Tangent Vector and Tangent Cone). Let $E$ be a normed vector space, $A$ a non-empty subset of $E$, and $x_0$ any point of $A$. A vector $h \in E$ is called a tangent vector to $A$ at $x_0$ if there is a sequence $x_n$ of elements of $A$ and a sequence $\lambda_n$ of positive real numbers with $\lim_{n \to \infty} x_n = x_0$ and $\lim_{n \to \infty} \lambda_n(x_n - x_0) = h$. Furthermore, let $T(A, x_0)$ be the set of all tangent vectors to $A$ at $x_0$. Then $T(A, x_0)$ is called the tangent cone to $A$ at $x_0$.

Remark A.1. We now look at some properties of $T(A, x_0)$.

(i) Since $T(A, x_0)$ certainly contains the null vector of $E$, it is not empty.
(ii) In the above definition, $x_0$ is necessarily a point of closure of $A$. Moreover, in general $T(A, x_0)$ is not a convex set.
(iii) Notice that if $A \subseteq B$ and $x_0 \in A \cap B$, then $T(A, x_0) \subseteq T(B, x_0)$.
(iv) We can also show that if $x_0 \in A \cap \text{int}(B)$, then $T(A \cap B, x_0) = T(A, x_0)$.

Lemma A.1. $T(A, x_0)$ is a cone.

Proof. Let $h \in T(A, x_0)$. Therefore, there exists a sequence of real numbers $\lambda_n > 0$ and a sequence of elements $x_n \in A$ with $x_n \to x_0$ such that $h = \lim_{n \to \infty} \lambda_n(x_n - x_0)$.

To show that $T(A, x_0)$ is a cone we have to show that $\alpha h \in T(A, x_0)$ for all $\alpha > 0$. Now, notice that,

$$\alpha h = \alpha \lim_{n \to \infty} \lambda_n(x_n - x_0) = \lim_{n \to \infty} \mu_n(x_n - x_0),$$

where $\mu_n = \alpha \lambda_n$. This shows that there exists a sequence of real numbers $\mu_n > 0$, and a sequence of elements $x_n \in A$ with $x_n \to x_0$ such that $\alpha h = \lim_{n \to \infty} \mu_n(x_n - x_0)$.

i.e. $\alpha h$ is also a tangent vector at $x_0$, which implies that $\alpha h \in T(A, x_0)$. Therefore, $T(A, x_0)$ is a cone.

Lemma A.2. $T(A, x_0)$ is closed.
Proof. See Krabs (1979, Page 154). □

Lemma A.3. Let $A$ be a non-empty convex subset of a vector space $E$. Then, $T(A, x_0)$ contains $A - x_0$ and is convex.

Proof. We first show that $T(A, x_0)$ contains $A - x_0$ if $A$ is convex. So let $h \in A$. Now define the sequence $h_n = x_0 + \frac{1}{n}(h - x_0)$, i.e. $h_n = \frac{1}{n}h + (1 - \frac{1}{n})x_0$. Clearly, $h_n \in A$ since $A$ is convex. Also, $h_n \to x_0$ and $n(h_n - x_0) \to h - x_0$. Therefore, $h - x_0 \in T(A, x_0)$ and, since $h$ was an arbitrary element of $A$, this implies that $A - x_0 \subseteq T(A, x_0)$.

We now show that $T(A, x_0)$ is convex. Let $h_1, h_2 \in T(A, x_0)$. Then there exist sequences $x^1_n, x^2_n \in A$ with $x^1_n \to x_0, x^2_n \to x_0$, and sequences of positive real numbers $\mu^1_n, \mu^2_n$ such that $h_1 = \lim_{n \to \infty} \mu^1_n (x^1_n - x_0)$ and $h_2 = \lim_{n \to \infty} \mu^2_n (x^2_n - x_0)$. Now, let $0 \leq \lambda \leq 1$ and define $h = \lambda h_1 + (1 - \lambda)h_2$. Then, $h = \lim_{n \to \infty} \delta_n (z_n - x_0)$, where,

$$
\delta_n = \lambda \mu^1_n + (1 - \lambda)\mu^2_n \quad \text{and,} \quad z_n = \frac{\lambda \mu^1_n}{\lambda \mu^1_n + (1 - \lambda)\mu^2_n} x^1_n + \frac{(1 - \lambda)\mu^2_n}{\lambda \mu^1_n + (1 - \lambda)\mu^2_n} x^2_n.
$$

Now $\delta_n$ is a sequence of positive real numbers, and $z_n \in A$ since $A$ is convex. So if we can show that $z_n \to x_0$ we would be done, since then $h$ would be an element of $T(A, x_0)$. We show this as follows. Notice that,

$$
\|z_n - x_0\| = \left\| \frac{\lambda \mu^1_n}{\lambda \mu^1_n + (1 - \lambda)\mu^2_n} (x^1_n - x_0) + \frac{(1 - \lambda)\mu^2_n}{\lambda \mu^1_n + (1 - \lambda)\mu^2_n} (x^2_n - x_0) \right\|
\leq \frac{\lambda \mu^1_n}{\lambda \mu^1_n + (1 - \lambda)\mu^2_n} \|x^1_n - x_0\| + \frac{(1 - \lambda)\mu^2_n}{\lambda \mu^1_n + (1 - \lambda)\mu^2_n} \|(x^2_n - x_0)\|
= \frac{\lambda}{\lambda + (1 - \lambda)\mu^2_n} \|x^1_n - x_0\| + \frac{\mu^2_n}{(1 - \lambda) + \lambda \mu^2_n} \|(x^2_n - x_0)\|.
$$

And since both the coefficients are bounded by 1,

$$
\|z_n - x_0\| \leq \|x^1_n - x_0\| + \|x^2_n - x_0\| \to 0,
$$

since $\|x^1_n - x_0\| \to 0$ and, $\|x^2_n - x_0\| \to 0$. Hence we are done. □

Using the properties given above, we get the following characterization of a tangent cone.

Theorem A.1. Let $A$ be a non-empty convex subset of a vector space $E$. Then, $T(A, x_0)$ is the smallest closed convex cone containing $A - x_0$.

Proof. By the previous theorem, when $A$ is a convex set containing $x_0$, $T(A, x_0)$ is a closed convex cone containing $A - x_0$. It only remains to show that it is the smallest closed convex cone containing $A - x_0$. So let $C$ be any closed convex cone containing $A - x_0$, and suppose that $h \in T(A, x_0)$. Then $h = \lim_{n \to \infty} \lambda_n (x_n - x_0)$, where $\lambda_n$ is a sequence of positive reals and $x_n$ is a sequence of elements in $A$ approaching $x_0$.

So define $h_n = \lambda_n (x_n - x_0)$. Clearly, $x_n - x_0 \in A - x_0$. But since $A - x_0 \subseteq C$ we have that $x_n - x_0 \in C$. Now, the fact that $C$ is a cone implies that $h_n = \lambda_n (x_n - x_0) \in C$;
i.e. \( h_n \) is a convergent sequence in \( C \). But since \( C \) is closed, the limit \( h \in C \). This implies that \( T(A, x_0) \subseteq C \). \( \Box \)

Using this theorem, we get the following important result about \( T(A, x_0) \), when \( A \) is itself a cone. This is a result of Aubin and Frankowska (1990, Lemma 4.2.5, Page 143).

**Theorem A.2.** Let \( A \) be a non-empty convex cone in a vector space \( E \), and \( x_0 \in A \). Then, \( T(A, x_0) = A - \mathbb{R}_{++}x_0 \).

**Proof.** \( \implies \) Let \( h \in T(A, x_0) \). Therefore, there exists a sequence of real numbers \( \lambda_n > 0 \) and a sequence of elements \( x_n \in A \) with \( x_n \to x_0 \), such that \( h = \lim_{n \to \infty} \lambda_n (x_n - x_0) \). But \( \lambda_n x_n \in A \) since \( A \) is a cone, and clearly \( \lambda_n x_0 \in \mathbb{R}_{++}x_0 \). This implies that \( \lambda_n x_n - \lambda_n x_0 \in A - \mathbb{R}_{++}x_0 \), which shows that \( h = \lim_{n \to \infty} \lambda_n (x_n - x_0) \in A - \mathbb{R}_{++}x_0 \).

\( \iff \) Let \( x \) be any arbitrary element of \( A \), and \( \lambda > 0 \) any element in \( \mathbb{R}_{++} \). If we could show that \( x - \lambda x_0 \in T(A, x_0) \), we would be done since this would imply that \( A - \mathbb{R}_{++}x_0 \subseteq T(A, x_0) \). Then taking the closure on both sides, and keeping in mind that \( T(A, x_0) \) is closed, we would have \( A - \overline{\mathbb{R}_{++}x_0} \subseteq T(A, x_0) \). So we show that \( x - \lambda x_0 \in T(A, x_0) \).

Since \( \lambda \in \mathbb{R}_{++} \), choose a \( t > 0 \) such that \( \lambda t < 1 \). Then since \( A \) is a convex cone, we have that \( (1 - \lambda t)x_0 + tx \in A \), i.e. \( x_0 + t(x - \lambda x_0) \in A \). This implies that \( t(x - \lambda x_0) \in A - x_0 \). But we know that \( T(A, x_0) \) contains \( A - x_0 \). Hence, we have that \( t(x - \lambda x_0) \in T(A, x_0) \). But since \( t > 0 \) and \( T(A, x_0) \) is a cone, this implies that \( x - \lambda x_0 \in T(A, x_0) \). \( \Box \)

**Corollary A.1.** Let \( A \) be a closed linear subspace of a vector space \( E \), and let \( x_0 \in A \). Then \( T(A, x_0) = A \).

**Proof.** Follows from the previous theorem. \( \Box \)

**Proof.** [Lemma 2.1] We will first show that whenever \( F \) is a cone with \( f^* \in F \), then \( f^* \in T(F, f^*) \). This is important because it may not be true when \( F \) is not a cone. For instance, let \( F = \{1\} \). Then \( T(F, 1) = \{0\} \), but \( 1 \notin T(F, 1) \). So let \( F \) be a cone and, let \( f^* \in F \). Now consider the curve \( \gamma(t) = f^* + tf^* \) for \( t \geq 0 \). Since \( F \) is a cone, \( \gamma(t) \in F \) for all \( t \geq 0 \). Also, \( \frac{d}{dt}\gamma(t)|_{t=0} = f^* \). Hence, \( \gamma(t) \) is a curve in \( F \) with \( f^* \) as the tangent vector; i.e. \( f^* \in T(F, f^*) \). Furthermore, if in addition to being a cone \( F \) is also convex then \( F \subseteq T(F, f^*) \). This may be seen as follows. From Theorem A.1, whenever \( F \) is a convex cone, \( T(F, f^*) \) is the smallest closed convex cone containing \( F - f^* \). That is,

\[
F - f^* \subseteq T(F, f^*) \iff F \subseteq T(F, f^*) + f^*.
\]

But the cone property of \( F \) implies that \( f^* \in T(F, f^*) \) (see above). Hence \( F \subseteq T(F, f^*) \), since \( T(F, f^*) \) is a convex cone.

We now use this result to show that \( \text{lin} F \subseteq \text{lin} T(F, f^*) \). Clearly, this will imply that \( \overline{\text{lin} F} \subseteq \overline{\text{lin} T(F, f^*)} \). So let \( f \in \text{lin} F \). This implies that for some positive integer \( m \), there exist real numbers \( \{\alpha_1, \ldots, \alpha_m\} \) and \( \{f_1, \ldots, f_m\} \in \times_{i=1}^m F \) such
that $f = \sum_{i=1}^{m} \alpha_i f_i$. But as we have just shown, whenever $F$ is a convex cone containing $f^*$, $F \subseteq T(F, f^*)$. This means that each $f_i \in F \subseteq T(F, f^*)$. Therefore, $f = \sum_{i=1}^{m} \alpha_i f_i \in \text{lin}T(F, f^*)$. Hence, $\text{lin}F \subseteq \text{lin}T(F, f^*)$.

Let us now show that $\text{lin}T(F, f^*) \subseteq \text{lin}F$. Again, this will clearly imply that $\text{lin}T(F, f^*) \subseteq \text{lin}F$. So let $f \in \text{lin}T(F, f^*)$. This means that for some $m \in \mathbb{N}$, there exist real numbers $\{\alpha_1, \ldots, \alpha_m\}$ and $\{f_1, \ldots, f_m\} \in \times_{i=1}^{m} T(F, f^*)$ such that $f = \sum_{i=1}^{m} \alpha_i f_i$. Using Theorem A.2, since each $f_i \in T(F, f^*) = F - \mathbb{R}_{++} f^*$, there exist $(g_n, \lambda_n) \in F \times \mathbb{R}_{++}$ such that $f_i = \lim_{n \to \infty} (g_n - \lambda_n f^*)$. But since $g_n \in F \subseteq \text{lin}F$ and $\lambda_n f^* \in F \subseteq \text{lin}F$, we get that $g_n - \lambda_n f^*$ is a convergent sequence in $\text{lin}F$. Hence its limit is an element of $\text{lin}F$, i.e., $f_i \in \text{lin}F$. Therefore, each $f_i$ is an element of $\text{lin}F$. Thus, $f = \sum_{i=1}^{m} \alpha_i f_i \in \text{lin}F$ which implies that $\text{lin}T(F, f^*) \subseteq \text{lin}F$.

The lemma stands proved since we have just shown that $\text{lin}F \subseteq \text{lin}T(F, f^*)$, and that $\text{lin}T(F, f^*) \subseteq \text{lin}F$. □

APPENDIX B. PROJECTION THEOREMS

**Theorem B.1 (Classical Projection Theorem).** Let $H$ be a Hilbert space and $M$ a closed subspace of $H$. Corresponding to any vector $x \in H$, there is a unique vector $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$. Furthermore, a necessary and sufficient condition that $m_0 \in M$ be the unique minimizing vector is that $x - m_0$ be orthogonal to $M$.


**Theorem B.2 (Projection on Convex Cones).** Let $H$ be a Hilbert space and $M$ a closed convex cone in $H$. Corresponding to any vector $x \in H$, there is a unique vector $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$. Furthermore, a necessary and sufficient condition that $m_0 \in M$ be the unique minimizing vector is that $(x - m_0, m_0) = 0$, and that $(x - m_0, m) \leq 0$ for all $m \in M$.

*Proof.* See Barlow, Bartholomew, Bremmer, and Brunck (1972). □

APPENDIX C. PROOFS OF THEOREM 2.1–THEOREMS 2.3

**Proof. [Theorem 2.1]** Clearly, $\beta_t = (\beta_{10}, \beta_{20}, \ldots, \beta_{(t+1)0}, \ldots, \beta_p)$ is a smooth curve in $B$ which passes through $\beta_3$ when $t = 0$. Then since $\eta_3$ is a least favorable surface, $\eta_3$ must be a least favorable curve for estimating $\beta_t$. That is, it must minimize $E \left[ \frac{d}{dt} \ell(\beta_t, \eta_t) \right]_{t=0}^2$. Now,

$$
\frac{d}{dt} \ell(\beta_t, \eta_t) = \sum_{i=1}^{p} \frac{\partial \ell(\beta_t, \eta_t)}{\partial \beta_i} \frac{d \beta_i(t)}{dt} + \frac{\partial \ell(\beta_t, \eta_t)}{\partial \eta} \sum_{i=1}^{p} \left( \frac{d}{d \beta_i} \eta_t \right) \frac{d \beta_i(t)}{dt}
$$

$$
= \frac{\partial \ell(\beta_t, \eta_t)}{\partial \beta_i} + \frac{\partial \ell(\beta_t, \eta_t)}{\partial \eta} \left( \frac{d}{d \beta_i} \eta_t \right), \text{ implying,}
$$

$$
\frac{d}{dt} \ell(\beta_t, \eta_t) \bigg|_{t=0} = \frac{\partial \ell(\beta_0, \eta_0)}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_0)}{\partial \eta} \left( \frac{d}{d \beta_i} \eta_0 \right).
$$
Therefore, minimizing $E \left[ \frac{d}{dt} \ell(\beta_t, \eta_{t_0}) \big| t = 0 \right]^2$ is equivalent to minimizing

$$E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \right] = E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \right]^2,$$

where $\delta_i \in \text{lin} T(f^*, F^*)$ for $i = 1, \ldots, p$. Hence the minimizer $\delta_i^*$ satisfies

$$E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta_i^*) \right] \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta) = 0$$

for all $\delta \in \text{lin} T(f^*, F^*)$. □

Proof. [Theorem 2.2] \( \implies \) So suppose that the result is true. Since it holds for all $(\delta_1, \ldots, \delta_p) \in \text{lin} T(f^*, F^*) \times \ldots \times \text{lin} T(f^*, F^*)$ it also holds for $(\delta_1, 0, \ldots, 0) \in \text{lin} T(f^*, F^*) \times \ldots \times \text{lin} T(f^*, F^*)$, because 0 is always an element of the tangent cone. Therefore, for all $\delta_1 \in \text{lin} T(f^*, F^*)$,

$$E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} \right] \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta_1) = 0.$$

But this implies that $\frac{d}{d\beta_i} \eta_{\beta_0}$ is the least favorable direction. The same holds for $\delta_2, \ldots, \delta_p$, and we get that $\delta^* = (\frac{d}{d\beta_1} \eta_{\beta_0}, \ldots, \frac{d}{d\beta_p} \eta_{\beta_0})$ is the least favorable direction. Therefore, $\eta_{\beta}$ is a least favorable surface since $\delta^* = \frac{d}{d\beta} \eta_{\beta}|_{\beta=\beta_0}$ is the tangent vector to $\eta_{\beta}$ at $f^*$.

$\iff$ Now suppose that $\eta_{\beta}$ is a least favorable curve, and let the least favorable direction be $\delta = (\frac{d}{d\beta_1} \eta_{\beta}|_{\beta=\beta_0}, \ldots, \frac{d}{d\beta_p} \eta_{\beta}|_{\beta=\beta_0})$. Then from Theorem 2.1, $\delta$ satisfies

$$E \left[ \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \beta_i} + \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta_i^*) \right] \frac{\partial \ell(\beta_0, \eta_{\beta_0})}{\partial \eta} (\delta_i) = 0$$

for all $\delta_i \in \text{lin} T(f^*, F^*)$, and $i = 1, \ldots, p$. Summation over $i$ then yields the required result. □

Proof. [Theorem 2.3] \( \implies \) Let $\lambda_{\beta}$ be a smooth curve in $\overline{\text{lin} F}$ such that $\lambda_{\beta_0} = f^*$ and, let $I_{\beta_0}$ be given by

$$E \left[ \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \lambda} \right] \left[ \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \lambda} \right]^2.$$

Assume that $\alpha'(I_{\beta_0} - I) \alpha \leq 0$ for all $\alpha \in \mathbb{R}^p$. We now show that $\lambda_{\beta}$ is a least favorable surface in $\overline{\text{lin} F}$ for estimating $\beta_0$. Since the given condition holds for all $\alpha \in \mathbb{R}^p$ choose $\alpha = e_j$, the $j$th unit vector in $\mathbb{R}^p$. Then $e_j'(I_{\beta_0} - I)e_j \leq 0$ becomes,

$$E \left[ \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \beta_j} + \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \lambda} \right] \left( \frac{d}{d\beta_j} \lambda_{\beta_0} \right)^2 \leq E \left[ \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \beta_j} + \frac{\partial \ell(\beta_0, \lambda_{\beta_0})}{\partial \lambda} \right] \left( \frac{d}{d\beta_j} \lambda_{\beta_0} \right)^2.$$
for \( j = 1, \ldots, p \). This implies that \( \frac{\partial l(\beta_0, \lambda_\beta)}{\partial \lambda} = (\frac{d}{d \lambda}, \lambda_\beta) \) is the projection of \( \frac{\partial l(\beta_0, \lambda_\beta)}{\partial \beta} \) onto \( \frac{\partial l(\beta_0, \lambda_\beta)}{\partial \beta} \). Therefore, \( \frac{d}{d \beta}, \lambda_\beta \) is the least favorable direction in \( \text{lin} T(\mathcal{F}, f^*) \) for estimating \( \beta_j \), implying that \( \lambda_\beta \) is a least favorable curve.

Now suppose that \( \eta_\beta \) is a least favorable surface in \( \text{lin} \mathcal{F} \) for estimating \( \beta_\theta \). Let \( S_{\beta_\theta} = \frac{d l(\beta_0, \eta_\beta)}{d \beta} \) and define \( I_{\beta_\theta} = \mathbb{E} S_{\beta_\theta} S_{\beta_\theta}^t \). Then we show that \( \alpha' (I_{\beta_\theta} - I) \alpha \leq 0 \) for all \( \alpha \in \mathbb{R}^p \), where \( I \) is the Fisher information matrix corresponding to any other \( p \) dimensional parameterization of the nuisance parameter.

So let \( \beta_\theta = \beta_0 + t \alpha \). Clearly, \( \beta_\theta \) is a smooth curve in \( \mathcal{B} \) through \( \beta_0 \), with tangent vector \( \alpha \). Since \( \eta_\beta \) is a least favorable surface, \( \eta_\beta \), must be a least favorable curve for estimating \( \beta \); i.e. \( \eta_\beta \), minimizes \( \mathbb{E} \left[ \frac{d l(\beta_0, \eta_\beta)}{d t} \right]_{t=0}^2 \). But by the chain rule,

\[
\frac{d l(\beta, \eta_\beta)}{d t} \bigg|_{t=0} = \left[ \frac{\partial l(\beta, \eta_\beta)}{\partial \beta} + \frac{\partial l(\beta, \eta_\beta)}{\partial \eta} \left( \frac{d}{d \beta} \eta_\beta \right) \right]' \left[ \frac{d}{d t} \bigg|_{t=0} \right] = S_{\beta_\theta}^t \alpha.
\]

Therefore,

\[
\mathbb{E} \left[ \frac{d l(\beta, \eta_\beta)}{d t} \bigg|_{t=0}^2 \right] = \mathbb{E} \left( S_{\beta_\theta}^t \alpha \right)' \left( S_{\beta_\theta}^t \alpha \right) = \alpha' S_{\beta_\theta} S_{\beta_\theta}^t \alpha = \alpha' I_{\beta_\theta} \alpha,
\]

and we get that \( \eta_\beta \), minimizes \( \alpha' I_{\beta_\theta} \alpha \). Similarly, if \( \lambda_\beta \) is any other smooth curve in \( \text{lin} \mathcal{F} \) through \( f^* \),

\[
\mathbb{E} \left[ \frac{d l(\beta, \lambda_\beta)}{d t} \bigg|_{t=0}^2 \right] = \alpha' I \alpha,
\]

where \( I \) is the now information matrix corresponding to \( \lambda_\beta \). But since \( \eta_\beta \) was a least favorable surface, it minimized \( \alpha' I_{\beta_\theta} \alpha \); i.e. \( \alpha' I_{\beta_\theta} \alpha \leq \alpha' I \alpha \), for all \( \alpha \in \mathbb{R}^p \). \( \square \)

**Appendix D. Proofs of Results in Example 4.1**

**Remark D.1.** In this section, all convergence is w.r.t. the \( C^2 \) norm. \( \square \)

**Proof.** [Proof of Lemma 4.1] Using Theorem A.2, we already know that \( T(\mathcal{F}, f^*) = \mathcal{F} - \mathcal{F} \cap f^* \). So we simply show that \( \mathcal{F} - \mathcal{F} \cap f^* = \mathcal{W} \).

We first show that \( \mathcal{F} - \mathcal{F} \cap f^* \subseteq \mathcal{W} \). So let \( f \in \mathcal{F} - \mathcal{F} \cap f^* \). Then there exists a sequence \( (f_n, \lambda_n) \in \mathbb{R}^+ \times \mathcal{F} \), such that \( f_n - \lambda_n f^* \to f \). So let \( g_n(z) = f_n(z) - \lambda_n f^*(z) \). Notice that on \( Z_0 \), \( \nabla^2 g_n(z) - \lambda_n \nabla^2 f^* = 0 \); i.e. the Hessian of \( g_n - f_n \) vanishes on \( Z_0 \). But this implies that \( g_n - f_n \in \mathcal{W} \), i.e. \( g_n \in \mathcal{W} + f_n \). However, \( f_n \in \mathcal{F} \subseteq \mathcal{W} \) and this implies that \( g_n \in \mathcal{W} \) since \( \mathcal{W} \) is a convex cone. But this means that \( g_n \) is a convergent sequence in the closed cone \( \mathcal{W} \). Therefore, its limit (denoted by \( f \)) also lies in \( \mathcal{W} \); i.e. \( \mathcal{F} - \mathcal{F} \cap f^* \subseteq \mathcal{W} \).

Now for the reverse inequality. Let \( \delta \) be any function in \( \mathcal{W} \), and for \( t > 0 \) define \( \eta_t(z) = f^*(z) + t \delta(z) \). If we can show that \( \eta_t(z) \in \mathcal{F} \) for all \( z \in Z \), we are done because then \( \delta = \eta_t - f^* \in \mathcal{F} - f^* \subseteq \mathcal{F} - \mathcal{F} \cap f^* \), implying that \( \mathcal{W} \subseteq \mathcal{F} - \mathcal{F} \cap f^* \). So let us show that \( \eta_t \in \mathcal{F} \). Now we know that by assumption the Hessian of \( f^* \)

\(^3\)Since \( \lambda_\beta \) and \( \eta_\beta \) are smooth curves in \( \text{lin} \mathcal{F} \) through \( f^* \), the tangent vectors \( \frac{d}{d \beta}, \lambda_\beta \) and \( \frac{d}{d \beta}, \eta_\beta \) are both elements of \( \text{lin} T(\mathcal{F}, f^*) \). (See Lemma 2.1).
is negative definite on $Z - Z_0$ and that $Z$ is compact. Therefore, for any $\alpha \in \mathbb{R}^n$, $z \in Z - Z_0$ and, sufficiently small $t$

$$\alpha'[^{\nabla^2 \eta_t}(z)]\alpha = \alpha'[^{\nabla^2 f^*}(z)]\alpha + t\alpha'[^{\nabla^2 \delta}(z)]\alpha < 0;$$

i.e. for small enough $t$, $\eta_t$ is strictly concave on $Z - Z_0$. Thus we have shown that for all $\alpha \in \mathbb{R}^n$ and sufficiently small $t$

$$\alpha'[^{\nabla^2 \eta_t}(z)]\alpha = \begin{cases} 
\leq 0 & \text{if } z \in Z_0, \\
< 0 & \text{if } z \in Z - Z_0;
\end{cases}$$

i.e. $\eta_t$ is concave on $Z$. And since $\eta_t \in \mathcal{H}$ by construction, we have that $\eta_t \in \mathcal{F}$.

Notice that if $f^*$ was strictly concave, $f^* \in \text{int}(\mathcal{F})$ and, therefore, $T(\mathcal{F}, f^*) = \mathcal{H}$. □

To prove Lemma 4.2 we need the following definition.

**Definition D.1.** Let $\mathcal{D}_W = \{f \in \mathcal{H} : f = f_1 - f_2, \text{ for some } f_1, f_2 \in \mathcal{W}\}$; i.e. $\mathcal{D}_W$ is the set of all functions in $\mathcal{H}$ which can be expressed as the difference of two functions in $\mathcal{W}$. □

It is easy to see from its definition that $\mathcal{D}_W \subseteq \mathcal{H}$ is a linear space containing $\mathcal{W}$. In fact, we will now show that $\mathcal{D}_W$ is the smallest closed linear space containing $\mathcal{W}$.

**Lemma D.1.** Let $\overline{\text{lin}} \mathcal{W}$ denote the smallest closed linear space containing $\mathcal{W}$. Then, $\mathcal{D}_W = \overline{\text{lin}} \mathcal{W}$.

*Proof.* Let $\mathcal{C}$ be the collection of all Banach spaces containing $\mathcal{W}$. Then $\overline{\text{lin}} \mathcal{W} = \cap_{\mathcal{X} \in \mathcal{C}} \mathcal{X}$. Since $\mathcal{H} \in \mathcal{C}$, $\mathcal{C}$ is not empty. We first show that $\mathcal{D}_W \subseteq \overline{\text{lin}} \mathcal{W}$. So let $f$ be any element of $\mathcal{D}_W$. Therefore, $f = f_1 - f_2$ for some $f_1, f_2 \in \mathcal{W}$. Now let $\mathcal{X}$ be an arbitrary Banach space containing $\mathcal{W}$. Notice that since $\mathcal{X}$ is a linear space, $-\mathcal{W} \subseteq \mathcal{X}$. Hence $f_1 \in \mathcal{W} \subseteq \mathcal{X}$ and, $-f_2 \in -\mathcal{W} \subseteq \mathcal{X}$. Therefore, by linearity of $\mathcal{X}$, $f_1 + (-f_2) \in \mathcal{X}$. That is, $\mathcal{D}_W \subseteq \mathcal{X}$, and since $\mathcal{X}$ was an arbitrary element of $\mathcal{C}$, $\mathcal{D}_W \subseteq \cap_{\mathcal{X} \in \mathcal{C}} \mathcal{X} = \overline{\text{lin}} \mathcal{W}$.

The other direction is even easier to show. By Lemma D.2, $\mathcal{D}_W$ is a Banach space containing $\mathcal{W}$. Now, by definition, $\overline{\text{lin}} \mathcal{W}$ is the smallest Banach space containing $\mathcal{W}$. Therefore, $\overline{\text{lin}} \mathcal{W} \subseteq \mathcal{D}_W$. □

We are now ready to prove Lemma 4.2.

*Proof.* [Lemma 4.2] When $f^*$ is strictly concave, i.e. $f^* \in \text{int}(\mathcal{F})$, $T(\mathcal{F}, f^*) = \mathcal{H}$ and there is nothing to prove. When $f^*$ is concave but not strictly concave, the tangent cone $T(\mathcal{F}, f^*) = \mathcal{W}$ and thus $\overline{\text{lin}} T(\mathcal{F}, f^*) = \overline{\text{lin}} \mathcal{W}$. From Lemma D.1 we have that $\overline{\text{lin}} \mathcal{W} = \mathcal{D}_W$. So it only remains to show that $\mathcal{D}_W = \mathcal{H}$. Now, by its very definition, $\mathcal{D}_W \subseteq \mathcal{H}$. But from Corollary D.1 we know that $\mathcal{H} \subseteq \mathcal{D}_W$. Therefore, $\mathcal{D}_W = \mathcal{H}$ and we are done. □

**Lemma D.2.** $\mathcal{D}_W$ is a Banach space containing $\mathcal{W}$.
Proof. It is easy to see that $\mathcal{D}_W \subseteq \mathcal{H}$ is a linear space containing $\mathcal{W}$. Hence it only remains to show that $\mathcal{D}_W$ is complete. So let $f_n$ be a sequence in $\mathcal{D}_W$ that converges in $C^2$ norm to $f$. We show that $f \in \mathcal{D}_W$.

Since the convergence is in the $C^3$ norm, $f \in \mathcal{H}$. So if we can show that $f$ can be written as the difference of two functions in $\mathcal{W}$, we are done. This is shown as follows. For $z \in Z$, define $g_0(z) = -\sum_{i=1}^t z_i^2$. Then since the Hessian of $g_0$ is negative definite on $Z$, $g_0 \in \mathcal{W}$. Now for any $z \in Z$, consider the function $h_\epsilon(z) = \frac{1}{\epsilon}g_0(z) + f(z)$, where $\epsilon > 0$. Therefore, for all $\alpha \in \mathbb{R}^t$ and sufficiently small $\epsilon$,

$$\alpha'[\nabla^2 h_\epsilon(z)]\alpha = \frac{1}{\epsilon} \alpha'[\nabla^2 g_0(z)]\alpha + \alpha'[\nabla^2 f(z)]\alpha < 0.$$ 

This follows from the fact that the Hessian of $g_0$ is negative definite and that $Z$ is compact. Therefore, for sufficiently small $\epsilon$, $h_\epsilon$ is strictly concave i.e. $h_\epsilon \in \mathcal{W}$. Moreover, the fact that $g_0 \in \mathcal{W}$ implies that $\frac{1}{\epsilon}g_0 \in \mathcal{W}$ since $\epsilon > 0$ and $\mathcal{W}$ is a cone. Therefore, $f$ can be written as the difference of two functions in $\mathcal{W}$. Hence, $f \in \mathcal{D}_W$. □

Corollary D.1. $\mathcal{H} \subseteq \mathcal{D}_W$.

Proof. Follows directly from Lemma D.2. □

Proof. [Lemma 4.3] Lemma 4.2 shows that $\zeta^* = -\text{proj}(x|\mathcal{H})$. But from Assumption 2.1(iv) it is easily seen that $E(x|z) \in C^2(Z)$; i.e. $\zeta^*(z) \in \mathcal{H}$. Hence all that remains is to verify the orthogonality condition of the classical projection theorem. But this is straightforward. □

APPENDIX E. PROOF OF THE IMPOSSIBILITY THEOREM

The proof of Theorem 4.1 requires the following definition.

Definition E.1 (LAN Condition). For any $\delta \in \mathbb{R}^p$ define $\beta_0 = \beta_\alpha + n^{-1/2}\delta$. Furthermore, let $L_n(\beta_\alpha, \gamma_{\beta_\alpha}) = \sum_{i=1}^n \ell(\beta_\alpha, \gamma_{\beta_\alpha}; x_i, y_i, z_i)$, where $\gamma_{\beta_\alpha}$ is any curve in $C^2(Z)$ such that $\gamma_{\beta_\alpha}|_{\beta_\alpha = \beta_0} = f^*$. Also let $L_n(\beta_0, \gamma_{\beta_0}) - L_n(\beta_0, f^*)$. Then for sufficiently large $n$, $L_n = n^{-1/2}b^T \frac{d}{d\beta} L_n(\beta_0, \gamma_{\beta_0}) - \frac{1}{2}b^T I_L b + o_p(1)$ where $I_L = E\left[\frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0})\right] \frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0})^T$. □

Throughout this section we assume that the LAN condition holds. Sufficient conditions under which this is true are given in the following lemma.

Lemma E.1 (Pflanzagl). With $\gamma_{\beta}$ as defined above and $i, j = 1, \ldots, p$, assume that for all $\beta \in B$ the loglikelihood $\beta \mapsto \ell(\beta, \gamma_{\beta}; x, y, z)$ satisfies,

(i) $E\left[\frac{d}{d\beta}, \ell(\beta, \gamma_{\beta}; x, y, z)\right] = 0$,

(ii) $E\left[\frac{d^2}{d\beta_i d\beta_j} \ell(\beta, \gamma_{\beta}; x, y, z)\right] + E\left[\frac{d}{d\beta_i} \ell(\beta, \gamma_{\beta}; x, y, z) \frac{d}{d\beta_j} \ell(\beta, \gamma_{\beta}; x, y, z)\right] = 0$. 

Proof. □
(iii) Let \( \mathcal{D} \) be the measure induced by \((x, y, z)\). Then for \( i = 1, \ldots, p \), the functions \( \frac{d}{d\beta_i} \ell(\cdot; \beta, \gamma_\beta) \) are linearly \( \mathcal{D} \)-independent. That is, if

\[
\sum_{k=1}^{m} a_k \frac{d}{d\beta_j} \ell(x_k, y_k, z_k; \beta, \gamma_\beta) = 0
\]

for \( \mathcal{D} \) - a.a. \((x, y, z)\), then \( a_k = 0 \) for all \( k \).

(iv) There exists a neighborhood \( N_{0} \) of \( \beta_0 \), such that for all \((x, y, z)\) the map

\[
\beta \mapsto \frac{d^2}{d\beta d\gamma} \ell(\beta, \gamma_\beta; x, y, z)
\]

is continuous on \( N_0 \), and,

\[
\mathbb{E} \sup_{\gamma_\beta \in N_{0}} \left| \frac{d^2}{d\beta d\gamma} \ell(\beta, \gamma_\beta; x, y, z) \right| < \infty.
\]

Then the LAN condition holds.

Proof. Let \( \tilde{\ell}(\beta; x, y, z) \equiv \ell(\beta, \gamma_\beta; x, y, z) \); i.e. treat \( \ell(\beta, \gamma_\beta; x, y, z) \) just as a parametric function of \( \beta \). Now, using \( \tilde{\ell}(\beta; x, y, z) \), proceed as in Pfanzagl (1994, Page 265). \( \Box \)

Let us now prove the impossibility theorem.

Proof. \textbf{[Theorem 4.1]} We obtain a proof by contradiction. So let Condition 4.1 hold and, suppose that there exists a regular \( n^{1/2} \) consistent estimator for \( \beta_0 \) that achieves the efficiency bounds when the parametric scores are projected onto \( T(\mathcal{F}, f^*) \). Let \( \hat{\beta}_n \) denote this estimator. We will now show that the existence of \( \hat{\beta}_n \) will violate Condition 4.1.

As a consequence of the theorem convolution theorem (Pfanzagl 1994, Page 289), \( \hat{\beta}_n \) is asymptotically linear; i.e. there exists a smooth curve \( \lambda_\beta \) such that \( \lambda_{\beta_0} = f^* \) and

\[
n^{1/2}(\hat{\beta}_n - \beta_0) = n^{-1/2} I_2^{-1} \sum_{i=1}^{n} \frac{d}{d\beta} \ell(\beta_0, \lambda_{\beta_0}; x_i, y_i, z_i) + o_p(1),
\]

where \( I_2 \) is defined in Condition 4.1. Since \( \hat{\beta}_n \) achieves the lower bound \( I_2^{-1} \), obtained by projecting the scores w.r.t. the parameter of interest onto the tangent cone \( T(\mathcal{F}, f^*) \), Theorem 3.3 shows that \( \lambda_\beta \) is a least favorable curve in \( \mathcal{F} \) for estimating \( \beta_0 \); i.e. \( \lambda_\beta \in \mathcal{F} \) and,

\[
I_2 = \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\beta_0}) \right] \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\beta_0}) \right]' .
\]

But as the LAN condition holds for a larger class of functions, it implies that for all \( (\delta, \gamma_\beta) \in \mathbb{R}^p \times \overline{lin} \mathcal{F} \) with \( \gamma_{\beta_0} = f^* \)

\[
\Sigma_n = n^{-1/2} \sum_{i=1}^{n} \delta \frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0}; x_i, y_i, z_i) - \frac{1}{2} \delta' I_L \delta + o_p(1),
\]

where, \( I_L = \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0}) \right] \left[ \frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0}) \right]' \). From Equation (E.1) and Equation (E.2) it should be clear that under \( \beta_0 \),

\[
n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, I_2^{-1}) \quad \text{and,} \quad \Sigma_n \xrightarrow{d} N(-\frac{1}{2} \delta' I_L \delta, \delta' I_L \delta).
\]
Therefore, \( \left( n^{1/2} (\hat{\beta}_n - \beta_0) \right) \frac{d}{\delta_0} \) \( \sim N \left( \left[ \begin{array}{c} 0 \\ -\frac{1}{2} \delta' \Gamma \delta \end{array} \right], \left[ I_2^{-1} \rho \rho' \right] \right) \) from the Cramér-Wold device, where \( \rho = I_2^{-1} \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\beta_0}) \right] \left[ \frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0}) \right]' \delta \). Hence by LeCam's Third Lemma (Bickel, Klassen, Ritov, and Wellner 1993, Page 503),

\[
n^{1/2} (\hat{\beta}_n - \beta_0) \frac{d}{\beta_0} \sim N \left( I_2^{-1} \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\beta_0}) \right] \frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0})' \delta, I_2^{-1} \right).
\]

Let \( I_p \) (resp. \( O_p \)) denote the \( p \times p \) identity (resp. zero) matrix. Then since \( n^{1/2} (\hat{\beta}_n - \beta_0) = \delta \), we have that \( n^{1/2} (\hat{\beta}_n - \beta_0) \frac{d}{\beta_n} \) \( \sim N (\mu, I_2^{-1}) \) with the bias term

\[
\mu = I_2^{-1} \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\beta_0}) \right] \frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0})' \delta - \delta
\]

\[
= \left\{ I_2^{-1} \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\beta_0}) \right] \frac{d}{d\beta} \ell(\beta_0, \gamma_{\beta_0})' - I_p \right\} \delta.
\]

Now \( \hat{\beta}_n \) is a regular estimator if and only if its asymptotic distribution (under \( \beta_n \)) does not depend upon \( \delta \); i.e. \( \hat{\beta}_n \) is regular if and only if \( \mu = 0 \). But since \( \delta \) is arbitrary,
\[ \mu = 0 \Leftrightarrow I_2^{-1} \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\sigma_0}) \right] \left[ \frac{d}{d\beta} \ell(\beta_0, \gamma_{\sigma_0}) \right]' = I_p = O_p \]

(\text{E.3})

\[ \Leftrightarrow \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\sigma_0}) \right] \left[ \frac{d}{d\beta} \ell(\beta_0, \gamma_{\sigma_0}) \right]' = I_2 \]

\[ \Leftrightarrow \mathbb{E} \left\{ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\sigma_0}) \right\} = \mathbb{E} \left[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\sigma_0}) \right]' = O_p \]

(\text{E.4})

\[ \Leftrightarrow \mathbb{E} \left\{ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\sigma_0}) \right\} \left\{ \frac{\partial \ell(\beta_0, \gamma_{\sigma_0})}{\partial \gamma} \left( \frac{d}{d\beta} \gamma_{\sigma_0} \right) - \frac{\partial \ell(\beta_0, \gamma_{\sigma_0})}{\partial \lambda} \left( \frac{d}{d\beta} \lambda_{\sigma_0} \right) \right\}' = O_p \]

(\text{E.5})

\[ \Leftrightarrow \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \lambda_{\sigma_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\sigma_0})}{\partial \lambda} \left( \frac{d}{d\beta} \lambda_{\sigma_0} \right) \right] \frac{\partial \ell(\beta_0, \gamma_{\sigma_0})}{\partial \gamma} \left( \frac{d}{d\beta} \gamma_{\sigma_0} \right)' = O_p \]

(\text{E.6})

Remark E.1. We now explain briefly how the above equations are obtained.

(i) Equation (E.3) simply follows from the definition of \( I_2 \).

(ii) To see how Equation (E.4) is obtained from the previous equation, notice that

\[ \frac{d}{d\beta} \ell(\beta_0, \gamma_{\sigma_0}) = \frac{\partial \ell(\beta_0, \gamma_{\sigma_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \gamma_{\sigma_0})}{\partial \gamma} \left( \frac{d}{d\beta} \gamma_{\sigma_0} \right) \text{ and,} \]

\[ \frac{d}{d\beta} \ell(\beta_0, \lambda_{\sigma_0}) = \frac{\partial \ell(\beta_0, \lambda_{\sigma_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\sigma_0})}{\partial \lambda} \left( \frac{d}{d\beta} \lambda_{\sigma_0} \right). \]

However, \( \frac{\partial \ell(\beta_0, \gamma_{\sigma_0})}{\partial \beta} = \frac{\partial \ell(\beta_0, \lambda_{\sigma_0})}{\partial \beta} \) since the scores with respect to the parameters of interest are equal.
(iii) Now \( \frac{d}{d \beta} \lambda_{\theta_0} \) is the least favorable direction in \( T(\mathcal{F}, f^*) \), and Equation (E.5) follows from the the previous equation since

\[
\mathbb{E} \left[ \frac{\partial \ell(\beta_0, \lambda_{\theta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\theta_0})}{\partial \lambda} \left( \frac{d}{d \beta} \lambda_{\theta_0} \right) \right] \frac{\partial \ell(\beta_0, \lambda_{\theta_0})}{\partial \lambda} \left( \frac{d}{d \beta} \lambda_{\theta_0} \right)' = O_p.
\]

By Theorem B.2, this is one of the necessary and sufficient conditions for projecting \( \frac{\partial \ell(\theta_0, \lambda_{\theta_0})}{\partial \lambda} \) onto \( T(\mathcal{F}, f^*) \).

(iv) Equation (E.6) follows from the previous equation because \( \frac{\partial \ell(\theta_0, \lambda_{\theta_0})}{\partial \lambda} \) is just the restriction of the Fréchet derivative \( \frac{\partial \ell(\theta_0, \lambda_{\theta_0})}{\partial \gamma} \) (defined over \( \text{lin} T(\mathcal{F}, f^*) \)) to the tangent cone \( T(\mathcal{F}, f^*) \). \( \square \)

Therefore, Theorem B.1 and Equation (E.6) imply that \( \frac{d}{d \beta} \lambda_{\theta_0} \in \mathbb{R}_{++} \) is the unique solution to the optimization problem

\[
\inf_{\xi \in \mathbb{R}_{++}} \mathbb{E} \left[ \frac{\partial \ell(\beta_0, \lambda_{\theta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\theta_0})}{\partial \lambda} (\xi) \right] \left[ \frac{\partial \ell(\beta_0, \lambda_{\theta_0})}{\partial \beta} + \frac{\partial \ell(\beta_0, \lambda_{\theta_0})}{\partial \lambda} (\xi) \right]'.
\]

This means that even when we search over a bigger space \( \text{lin} T(\mathcal{F}, f^*) \), the least favorable direction \( \frac{d}{d \beta} \lambda_{\theta_0} \) is found to lie in the strictly smaller space \( T(\mathcal{F}, f^*) \). But this violates Assumption 4.1. Hence we are done. \( \square \)

**Appendix F. Proofs of Results in Section 5**

**Proof. [Lemma 5.1]** The theorem is proved using the "guess and verify" technique. Let \( \lambda > 0 \). Since

\[
\delta^*_r(\lambda z) = \frac{(\lambda z)^r}{\mathbb{E} [X, Z^r | Z_q = \frac{z_q}{z_q}, \ldots, Z_{q-1} = \frac{z_{q-1}}{z_q}, Z_{q-2} = \frac{z_{q-2}}{z_q}]} = \lambda^r \delta^*_r(z),
\]

\( \delta^*_r(\cdot) \) is homogeneous of degree \( r \). To verify the orthogonality conditions, let \( g(\cdot) \) be
a homogeneous function of degree \( r \). Then since \( g(Z) = Z_1 g(Z_1, \ldots, Z_{n-1} Z_q, 1), \)

\[
\mathbb{E} [(X_i - \delta^*_i(Z)) g(Z)] = \mathbb{E} \left[ \left\{ X_i - \frac{Z_1^r \mathbb{E} (X_i, Z_1^r | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q})}{\mathbb{E} (Z_1^r | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q})} \right\} Z_1^r g(Z_1, \ldots, Z_{n-1}, 1) \right]
\]

\[
= \mathbb{E} [X_i g(Z)] - \mathbb{E} \left[ \frac{\mathbb{E} (Z_1^r X_i g(Z_1, \ldots, Z_{n-1}, 1) | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q})}{\mathbb{E} (Z_1^r | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q})} \right]
\]

\[
= \mathbb{E} [X_i g(Z)] - \frac{\mathbb{E} (X_i g(Z) | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q})}{\mathbb{E} (Z_1^r | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q})}
\]

\[
= \mathbb{E} [X_i g(Z)] - \mathbb{E} \left[ \frac{\mathbb{E} (X_i g(Z) | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q})}{\mathbb{E} (Z_1^r | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q})} \right]
\]

Thus \( \delta^*_i(Z) \), a homogeneous function of degree \( r \), solves the projection problem in \( \mathcal{L}^2 \). So if we can now show that \( \delta^*_i(Z) \) is also an element of \( \mathcal{F}_i \), we are done. But to show this we simply have to show that \( \delta^*_i \in \mathcal{C}^2(Z) \); i.e. it suffices to show that the conditional expectations \( \mathbb{E} [X_i, Z_1^r | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q}] \) and \( \mathbb{E} [Z_1^r | \frac{Z_1}{Z_q}, \ldots, \frac{Z_{n-1}}{Z_q}] \) are twice continuously differentiable w.r.t. \( Z \). After a little algebra, this may be shown to follow from Assumption 2.1(iv). \( \square \)

**Remark F.1.** Suppose that in the above lemma we had taken the projection as

\[
\delta^*_i(Z) = -\frac{z_i^r \mathbb{E} (X_i, Z_1^r | Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1})}{\mathbb{E} (Z_1^r | Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1})}.
\]

That this is a valid projection can be seen immediately, since this also satisfies the necessary and sufficient conditions of the classical projection theorem. Then the uniqueness of the projections (postulated by the classical projection theorem) would imply that

\[
\frac{z_i^r \mathbb{E} (X_i, Z_1^r | Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1})}{\mathbb{E} (Z_1^r | Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1})} = -\frac{z_i^r \mathbb{E} (X_i, Z_1^r | Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1})}{\mathbb{E} (Z_1^r | Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1})}.
\]

That this is indeed the case, may be seen as follows.
\[
\frac{z_1^* E \left( X, Z_q^* \left| \frac{Z_1}{Z_q} = \frac{z_1}{z_q}, \ldots, \frac{Z_{q-1}}{Z_q} = \frac{z_{q-1}}{z_q} \right. \right)}{\frac{z_1^* E \left( Z_{q-1}^* \left| \frac{Z_q}{Z_1} = \frac{z_q}{z_1}, \ldots, \frac{Z_{q-1}}{Z_q} = \frac{z_{q-1}}{z_q} \right. \right) = \frac{z_1}{z_q}}}
\]

This is a nice test of the validity of the result obtained in Lemma 5.1. □

References


Department of Economics, University of Wisconsin, Madison, WI - 53706.
E-mail address: gtripath@ssc.wisc.edu