Here you can insert the title of your seminar paper

First Student Second Student

Module: Modulename
Course: Coursename

Lecturer: Name of the lecturer

December 2, 2016

Abstract

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Contents

1 Introduction

Following Li and Racine (2007) notation, the model

$$Y = g(X'\beta_0) + \epsilon, \tag{1}$$

where

1. Y is the dependent variable, $X \in \mathbb{R}^q$ is a vector of explanatory variables, β_0 is the q $\times$ 1 vector of unknown parameters;
2. $X'\beta_0$ is a single index because it is a scalar;
3. $E(\epsilon|x) = 0$;
4. $g : \mathbb{R} \to \mathbb{R}$ is not known;

is a single index model.

The above model is referred to as a linear single index model by Ichimura (1993). This model is semiparametric as the functional form of the linear index is specified, while $g(\cdot)$ is left unspecified and the conditional probability of ϵ conditioned on X is not specified except $E(\epsilon|X) = 0$.
Semiparametric single index models arise naturally in binary choice models. We now analyze binary choice models in the semiparametric single index setting proposed by Li and Racine (2007)[1]. More specifically, by accepting the parametric linear index assumption governing choices while not specifying the unknown distribution of the error term, we obtain a semiparametric single index model. Thus, when considering the relationship between a binary dependent variable (Y) and some covariates (X) this relationship can be modelled as follows:

$$Yi = \begin{cases} 1, & \text{if } Y_i^* \overset{def}{=} \alpha + X_i'\beta + u_i > 0 \\ 0, & \text{if } Y_i^* = \alpha + X_i'\beta + u_i \leq 0 \end{cases} \tag{2}$$

where Y^* is a latent variable. Consider Y representing labor force participation, equal to 1 if the individual participates in the labor market and 0 otherwise. The explanatory variables can contain a series of economic factors that might influence labor participation, such as gender, age, marital status and education. Assuming a linear relationship between labor force participation and potential determinants, the empirical analysis focuses on the estimation of β. Parametric methods to estimate β require assumptions on the distribution of the error term u. A common assumption in the parametric framework is $u \sim N(0, \sigma^2)$. With further identification conditions such as $\sigma = 1$, β can be jointly identified (see Madalla (1986)) and we can use maximum likelihood to estimate it. Let $F_u(\cdot)$ denote the true CDF of u. Nevertheless, excluding the assumption of normality of the error term would in general lead to inconsistent estimates. Then we get:

[1]While focusing on binary choice, we acknowledge that Y is not strictly bound to a binary character.

$$
\begin{aligned}
E(Y|x) &= \sum_{y=0,1} yP(y|x) \\
&= 1 \times P(Y = 1|x) + 0 \times P(Y = 0|x) \\
&= P(Y = 1|x) \\
&= P(\alpha + x'\beta + u_i > 0) \\
&= P(u_i > -(\alpha + x'\beta)) \\
&= 1 - P(u_i \leq -(\alpha + x'\beta)) \\
&= 1 - F(-(\alpha + x'\beta))
\end{aligned}
$$

where $F(\cdot)$ is the cummulative distribution function (CDF) of u. If $u \sim N(0, 1)$ is true, then u has a symmetric distribution and $1-F(-(\alpha+x'\beta)) = F(\alpha+x'\beta)$. Then model (2) is commonly referred to as a Probit:

$$E(Y|X) = P(Y = 1|x) = \Phi(\alpha + x'\beta), \tag{3}$$

where Φ is the CDF of a standard normal variable. On the other hand, for u having a symmetric logistic distribution, we obtain:

$$E(Y|X) = P(Y = 1|x) = \frac{1}{1 + e^{\alpha+x'\beta}}, \tag{4}$$

From (3) and (4) we can see that different functional forms for u lead to different functional forms for the conditional probability of $Y = 1$. Therefore, consistent parametric estimation of $E(Y|X) = P(Y = 1|x)$ requires the *correct* distribution specification of u.
Generally speaking, model (1) contains many widely used parametric models that assume that g is known up to a finite-dimensional parameter. If g is the identity function we obtain a linear model. If g is the cumulative normal or logistic distribution, we obtain a binary probit or logit model. Further, assuming $g(X'\beta_0) = E(Y|X = x)$ model (1) represents a tobit model:

$$Y = max(0, g(X'\beta_0) + \epsilon),$$

where ϵ is an unobserved, normally distributed random variable that has mean zero and is independent of X.
In conclusion, model (1) with unknown $g(\cdot)$ leads to a more flexible version of parametric models while retaining many of its desirable features. As shown above, model (1) avoids the problem of error distribution mispecification. Further, the single index model achieves dimension reduction and avoids the curse of dimensionality: g is estimated with the same convergence rate in probability as in the case when $X'\beta_0$ is observable and β_0 achieves the same convergence rate $n^{-\frac{1}{2}}$ as parametric models. Thus, the single index model is as accurate as a parametric model for estimating $X'\beta_0$. On the other hand, the assumptions are weaker than those of a parametric model and stronger than those of a fully nonparametric model. [2]

[2]However, the necessary assumptions for consistent parametric estimation can be relaxed. In particular, the single index model might have weaker assumptions than a fully parameterized model for structural economic models.

Essentially, while the risk of mispecification is reduced relative to parametric models, it also avoids inconveniences of fully nonparametric models such as the curse of dimensionality, difficulty of interpretation, and the lack of extrapolation capability. Moreover, notice that model (1) implies $E(Y|x) = g(x'\beta)$, and thus Y depends on x only through the linear combination $x'\beta_0$, and this relationship is characterized by the link function $g(\cdot)$.

2 Identification

Before estimating β_0 and $g(\cdot)$, restrictions must be imposed to ensure identification of the semiparametric model:

$$E(Y|x) = g(x'\beta_0)$$

For the discussion of the identification strategy we follow Li and Racine (2007).

Proposition 1. *(Identification of a Single Index Model). Identification of β_0 and $g(\cdot)$ in model (1) requires:*

(i) x should not contain a constant and it must contain at least one continuous variable. Furthermore, $\|\beta_0\| = 1$.

(ii) $g(\cdot)$ is differentiable and it is not a constant function on the support of $x'\beta_0$.

(iii) For the discrete components of x, varying the values of the discrete variables will not divide the support of $x'\beta_0$ into disjoint subsets.

Intuition. We start with *(*i) and by emphasizing that the elements of x cannot suffer from multicollinearity, i.e., "β_0 is not identified, if $Pr(x'\alpha = c) = 1$ where α is a constant and c is a scalar. Further, the requirement that x contains at least one continuous variable prevents that x as well as the scalar variable $v = x'\beta_0$ for any vector β have a finite support. Then $E(Y|X = x) = g(x'\beta)$ would impose only a finite number of restrictions on $g(\cdot)$, leading to a infinite number of different choices for $g(\cdot)$ and β_0 that satisfy those restrictions. [3] In addition, *location normalization* and *scale normalization* requirements are necessary. Define the function g^* by the relation $g^*(\gamma + v\delta) = g(v)$ for all v in the support of $x'\beta_0$. Then

$$E(Y|X = x) = g(x'\beta_0) \tag{5}$$

and

$$E(Y|X = x) = g^*(\gamma + x'\beta_0\delta) \tag{6}$$

The models (5) and (6) are observationally equivalent. Thus, β_0 and g are not identified unless restrictions are imposed to uniquely specify γ and δ. The restriction on γ is then a *location normalization* and it can be satisfied by for example requiring x not to include a constant. The restriction on δ is a *scale*

[3]Note however that if $g(\cdot)$ is assumed to be increasing, we can identify bounds on the components of β_0. See Horowitz (2009) for more concrete examples.

normalization. In this case we use the approach that assumes that the vector β_0 has unit length, i.e., $\|\beta_0\| = 1$ using an Euclidean norm. As to what concerns part (ii), $g(\cdot)$ cannot be a constant function, as otherwise β_0 is clearly not identified. What makes the identification of $E(Y|X = x)$ possible is that it remains constant if x changes in a way such that $x'\beta_0$ stays constant. Then $P(x'\beta_0 = c) = 0$, for $x'\beta_0$ continuously distributed and for some constant c. For g differentiable, the set of x for which $x'\beta_0$ is within any specified nonzero distance of c has nonzero probability for c in the interior of the support of $x'\beta_0$ and we identify β_0 by the approximate constancy of $x'\beta_0$. The violation of (iii) might for example lead to the violation of *location normalization* and *scale normalization*. For a concrete example see Horowitz (2009).

3 Section about quotations

In this section, an example for a literal quotation is given.

> *"A persona is a rich picture of an imaginary person who represents your core user group."* (**?**)

Sometimes you might want to make use of the authors name within the text. Before, we used the command `citep{}`, which creates the brackets around author name and year. You can also use the `cite` command like this:

? defined the concept of persona as follows:

> *"A persona is a rich picture of an imaginary person who represents your core user group."* (**?**)

You may notice, that this increases the readability of the text.

According to APA format[4] there are some rules, when and how to include page numbers, when referring to literature.

> *"Include page numbers for any citations in the text of your paper that include direct quotations or refer to a specific part of the work you are referencing. Direct quotations must include a page number as part of the citation. The quoted material should be followed by a citation in parentheses that gives the author's name, the year in which the work was published, and the page number from which the quoted material appears."* (**?**)

Check out the example and recommendations of **?** on `http://www.ehow.com/how_5689799_cite-numbers-apa-format.html`. In LaTeXyou can include the pages very easy. For example:

?, p. 86 stated:

> *"We hope that our preliminary attempts to begin answering the question will convince the reader, not necessarily that our views are correct, but that the question was and is well worth asking"* (**?**, p. 86)

[4] American Psychological Association (APA)

Note that in the first reference, we used `citet[]{}` in order to have brackets just around year and page number; later we used `citep[]{}`.

4 Section about references within the document

If you want to refer to you own chapters, figures, tables or the like, you can make use of the `ref{}` command, for example:

- section 3 on page 6

4.1 Subsection within Foundations

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

4.2 Another subsection within Foundations

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

5 Methodology

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua (**?**). At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

6 Results

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua (**?**, p. 48). At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

7 Conclusion

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

References