Journal of Econometrics 189 (2015) 148–162

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Smooth coefficient estimation of a seemingly unrelated regression

Daniel J. Henderson [a,*], Subal C. Kumbhakar [b,c], Qi Li [d,e], Christopher F. Parmeter [f]

[a] Department of Economics, Finance and Legal Studies, University of Alabama, Tuscaloosa, AL 35487-0224, United States
[b] Department of Economics, State University of New York at Binghamton, United States
[c] University of Stavanger Business School, Stavanger, Norway
[d] Department of Economics, Texas A& M University, United States
[e] ISEM, Capital University of Economics & Business, China
[f] Department of Economics, University of Miami, United States

ARTICLE INFO

Article history:
Received 5 June 2012
Received in revised form
12 March 2015
Accepted 12 July 2015
Available online 29 July 2015

JEL classification:
C14
C39

Keywords:
Semiparametric smooth coefficient model
System estimation
Bandwidth selection
Banking

ABSTRACT

This paper proposes estimation and inference for the semiparametric smooth coefficient seemingly unrelated regression model. We discuss the imposition of cross-equation restrictions which are required by economic theory as well as methods for data driven bandwidth selection. A test of correct functional form for the entire system of equations is also constructed. Asymptotic and finite sample results are given. We illustrate our estimator by applying it to a cost system for US commercial banks. Our results show that most of the banks are operating under increasing returns to scale, but that returns to scale decrease with bank size.

1. Introduction

Nonparametric methods are now quite popular among statisticians, econometricians and applied economists. However, a well known criticism against the use of nonparametric models is the 'curse of dimensionality'. In applied settings this is likely to be troubling as researchers typically have access to a potentially large number of explanatory variables. While one could employ dimension reduction methods such as projection pursuit (Huber, 1985) or engage in significance testing/automatic variable removal (Lavergne and Vuong, 2000; Hall et al., 2007), a common alternative is to use semiparametric methods. While not as flexible as their nonparametric counterparts, semiparametric methods can lessen the curse of dimensionality while not sacrificing too much flexibility for the problem at hand. Additionally, in some settings the use of semiparametric methods can allow easier implementation of an estimator that satisfies certain faculties of the given problem, imposing constraints for example.

Here we use the semiparametric smooth coefficient model (SPSCM) to illustrate this point. The SPSCM has its origins in econometrics dating back to the seminal work of Robinson (1989).[1] Currently it has seen renewed interest, most likely stemming from the fact that it is easily manipulated to mesh with a variety of econometric settings. Das (2005) and Cai et al. (2006) proposed using this estimator in an instrumental variable setting while Cai and Li (2008) proposed using the SPSCM to estimate a dynamic panel regression model. In an applied setting, Mamuneas et al. (2006) used the SPSCM to study the relationship between development and human capital.

In this paper we develop a SPSCM for estimating a seemingly unrelated regression (SUR) model. There are several reasons why we choose to use the SPSCM for a SUR model. First, as mentioned previously, semiparametric methods lessen the curse of dimensionality, which is important in applied settings. Second, in typ-

* Corresponding author. Tel.: +1 205 348 8991; fax: +1 205 348 0186.
E-mail address: djhender@cba.ua.edu (D.J. Henderson).

[1] These methods were made popular in statistics when they were explored by Cleveland et al. (1991) and Hastie and Tibshirani (1993) where they are commonly referred to as varying coefficient models.

http://dx.doi.org/10.1016/j.jeconom.2015.07.002

ical applications of SUR models, cross-equation restrictions[2] are required and can be imposed via matrix arrangement when we use the SPSCM. Lastly, the SPSCM has a 'conditionally' parametric structure, which makes interpretation of results straightforward.

Our discussion of semiparametric estimation in a system adds to the burgeoning literature in this field, which has seen recent interest (Welsh and Yee, 2006; Matzkin, 2008; Jun and Pinkse, 2009). Our semiparametric estimator is straightforward to implement and should prove valuable for modeling systems of equations when the functional form of the unknown responses is not immediately derived from economic theory and cross-equation coefficient restrictions are necessary to impose. In microeconomic theory, cross-equation coefficient restrictions are quite common. For example, in estimating consumer demand functions based on utility maximization with budget constraints (or minimizing cost for a given level of utility), the demand functions share the parameters in the utility function. Similarly, in production theory, conditional input demand functions for cost minimizing firms share the parameters of the production (cost) function. The (unconditional) input demand and output supply functions for profit maximizing firms depend on the production (profit) function parameters. We show how to incorporate these restrictions in a SUR system and that accounting for these cross-equation coefficient restrictions improves the asymptotic efficiency of the semiparametric smooth coefficient estimator.

In a similar vein, Orbe et al. (2003) develop a semiparametric time-varying smooth coefficient estimator for a system of equations (see also, Orbe et al., 2005). In much the same spirit that Li et al. (2002) generalizes the work of Robinson (1989), our work here generalizes that of Orbe et al. (2003) who developed an estimator to impose various restrictions on (potentially) time-varying coefficients. No theoretical properties of the proposed estimator were provided although the method was demonstrated to work well in both practical and simulated settings. Moreover, the focus of Orbe et al. (2003) was on allowing for seasonality and trending for all coefficients in a system of equations, coupled with restrictions on the coefficients. Their estimator requires solving a recursion formula. In our setting, a closed form solution exists that does not become more difficult as the sample size grows. Further, we provide the asymptotic properties of our estimator as well as for our test of correct parametric specification of the SUR model. In addition, we propose a method for data driven bandwidth selection.

Alternatively, models with varying coefficients stemming from a system of equations which depend on unobserved heterogeneity have recently been proposed. Jun (2009) presents a triangular model with varying coefficients that depend upon unobserved heterogeneity as opposed to explanatory variables. Jun's (2009) model stems from a non-separable triangular system that allows for a wide variety of heterogeneity as well as endogeneity.

A further benefit of the semiparametric system estimator within the production paradigm is the inclusion of non-traditional inputs or environmental variables (which will be illustrated in our empirical example). It is common to encounter key variables in an applied production setting which do not fit into a classic input/output analysis, but more than likely impact the production environment of the firm. Our semiparametric model can incorporate these variables directly into the smooth coefficients and, conditional on these variables, we have a consistent notion of the production environment. Another way to view the influence of these variables is that they change the production landscape in a manner that makes the model a standard parametric production model, but holding the levels of these variables fixed.

We apply our cross-sectional cost system method to US commercial banks in 2010. Since bank size is an important factor in the production environment, we use it as an argument for the smooth coefficients. We consider a single equation cost function with a SPSCM as well as a cost system with a SPSCM. Our results suggest that the impact of the production environment, as measured by bank size depends on whether we use the cost share equations or focus exclusively on the single equation cost function. We find increasing returns for most banks, but our results show that returns-to-scale diminish with bank size. When we use the single equation cost function, the increasing returns hold for even very large banks, whereas for the system estimator, we cannot reject constant returns for the largest banks. This finding is potentially important as increasing returns is often used to justify bank mergers and in policy debates on regulations limiting the size of banks (especially after the recent financial crisis).

The remainder of the paper is organized as follows. Section 2 presents the SPSCM estimator for a SUR system and establishes the large sample theory. Section 3 provides a test of correct functional form for the entire SUR system. Section 4 provides finite sample results from a small Monte Carlo setup. Results from our empirical example are given in Section 5. Finally, Section 6 presents some concluding remarks and direction for future research.

2. Semiparametric smooth coefficient systems of equations

The general setup of a varying coefficient regression takes the form

$$y_i = x_i^T \beta(z_i) + u_i, \quad i = 1, \dots, n \tag{1}$$

where y_i is the response variable of unit i, x_i is a $l \times 1$ vector of regressors, the superscript T denotes transpose, z_i is a vector of environmental variables of dimension q and u_i is an additive idiosyncratic error. One can envision the setup in (1) stemming from the translog cost function presented in (11) via a set of environmental variables that characterize the operating environment of the firms. For example, Feng and Serletis (2009) allow the parameters of their translog cost function to varying depending on the size category that each bank falls within. Asaftei and Parmeter (2010) note that the smooth coefficient model can be thought of as linear in parameters for a fixed value of z.

Li et al. (2002) discuss standard local-constant estimation of this model in the multivariate setting, prove its consistency and provide a test of function form while Lee and Ullah (2001) study the local-linear version of this estimator. Other theoretical contributions include Cai et al. (2000a,b) who propose a one-step local maximum likelihood estimator for generalized linear models with varying coefficients, Cai et al. (2000a,b) who study the time-series properties of the varying coefficient model and show how many practical time series models can have smoothly varying coefficients, and Cai (2007) and Cai et al. (2009) who discuss the asymptotic properties of the local linear smooth coefficient model in the presence of non-stationarity. Further, Fan and Huang (2005) detail inference via profile likelihood estimation of varying coefficient models and show that a profile likelihood ratio test provides power gains over existing tests involving varying coefficients while Li and Racine (2010) study the theoretical and practical properties of the varying coefficient estimator case in the mixed discrete–continuous data environment. As should be evident, the single equation varying coefficient regression estimator is well studied and has been shown to have suitable asymptotic properties across a range of models and assumptions.

[2] The cross-equation coefficient restrictions we consider here are required by economic theory and are not debatable. However, the theory is more general and could be used to impose other restrictions of economic interest (e.g., constant and/or unitary returns to scale, separability, monotonicity, etc.).