TRIANGULAR SIMULTANEOUS EQUATIONS MODEL UNDER STRUCTURAL MISSPECIFICATION

A Dissertation

submitted to the UCL Faculty Social & Historical Sciences
in fulfillment of the
requirements for the degree of
Master of Philosophy (M.Phil.)

in
The Department of Economics

by
Ilker Kandemir, DEA, BA
November, 2011

I, Ilker Kandemir, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Triangular simultaneous equation models are commonly used in econometric analysis to analyse endogeneity problems caused, among others, by individual choice or market equilibrium. Empirical researchers usually specify the simultaneous equation models in an ad hoc linear form; without testing the validity of such specification. In this paper, approximation properties of a linear fit for structural function in a triangular system of simultaneous equations are explored. I demonstrate that, linear fit can be interpreted as the best linear prediction to the underlying structural function in a weighted mean squared (WMSE) error sense. Furthermore, it is shown that with the endogenous variable being a continuous treatment variable, under misspecification, the pseudo-parameter that defines the returns to treatment intensity is weighted average of the Marginal Treatment Effects (MTE) of Heckman and Vytlacil (2001). Misspecification robust asymptotic variance formulas for estimators of pseudo-true parameters are also derived. The approximation properties are further investigated with Monte-Carlo experiments.

Table of Contents

Abstract 3

List of Tables 5

List of Figures 6

Acknowledgements 7

Chapter 1 Triangular Simultaneous Equations Model under Structural Misspecification 8
- 1.1 Introduction 8
- 1.2 Definition and Characterization of Misspecification 11
- 1.3 Interpretation of Linear Fit under Misspecification 13
  - 1.3.1 Complete Linear Misspecification 13
  - 1.3.2 Structural Linear Misspecification 13
  - 1.3.3 Omitted Variable Bias Formula 18
- 1.4 Relation with Treatment Effects 19
- 1.5 Estimators for Pseudo-Parameters under Misspecification and Their Asymptotic Properties 20
- 1.6 Simulation 23
- 1.7 Conclusion 25
- 1.8 Proofs 27

References 32

List of Tables

Table 1: Estimated Probability Limits of Estimates of Partially Linear Model and Weighted Least Squares Approximation. 24

Table 2: Estimates of Partially Linear Model and Weighted Least Squares Approximation from a Random Replication. 25

List of Figures

Figure 1.1 Kernel Density of Weights from a Random Replication 26

Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I want to thank Simon Lee, Whitney Newey, Adam Rosen and João Santos Silva for valuable discussions. I have furthermore to thank my former advisors Juan Mora and Elena Martinez and my former classmates Brice Corgnet, Ismael Rodriguez, Ufuk Otag and M. Hakan Eratalay who were all continuous resources of encouragement.

I am grateful to my supervisor Lars Nesheim whose help, stimulating suggestions and encouragement helped me in all the time of research for and writing of this thesis.

Several graduate students from UCL Economics supported me in my research work. I want to thank them for all their help, support, interest and valuable hints. Especially I am obliged to express my gratitude to Sami Stouli, Luigi Minale, Marieke Schnabel, Kari V. Salvanes, Kazuki Baba, Michael Amior and Ali Reza Sepahsalari for their support and time. I will remain forever indebted to Jan Stuhler, Italo Lopez Garcia and Lucia Rizzica for being there at difficult times. I also want to thank to Administrative Staff of UCL Economics. In particular, Daniella Fauvrelle, Katie Canada and Tina Fowler were of great help whenever I needed it.

I will be forever grateful to my parents, Huseyin and Sebiha; and my brother Ilkay, for their neverending love and support.

I would like to give my special thanks to my wife Jana whose patient love and endless encouragement enabled me to complete this work in difficult circumstances.

Chapter1 Triangular Simultaneous Equations Model under Structural Misspecification

1.1 Introduction

Despite recent innovations in nonparametric/semiparametric methods, linear modelling remains widely popular in regression analysis, since it is simpler to handle and statistical inference in linear models is well established. Linear models ensure fast convergence rates, thus require smaller sample size compared to nonparametric models. Linear regression, unlike nonparametric regression, permits direct interpretation of estimates in terms of policy-relevant quantities, such as elasticities or derivatives of the response variable. However, a linear specification is often assumed without any kind of economic justification whatsoever; this in turn, leads to misspecification biases. In this paper, I explore the nature of pseudo-parameters in a linearly misspecified structural equation within a triangular system. Specifically, I show that the pseudo-parameters solve a weighted least squares problem with possibly negative weights. An explicit expression for the weight function, as well as a misspecification bias formula for the structural function is provided.

Several interpretations of linear estimates under misspecification have been proposed. Early interpretations include Cramer (1969), Denny and Fuss (1977) who define the OLS (ordinary least squares) estimator for a misspecified linear regression model as a Taylor series approximation to the underlying true conditional expectation, under strict assumptions on the functional form of true conditional expectation. White (1980), under relatively mild conditions, demonstrates that the OLS estimator for the misspecified linear regression model is the best linear prediction to the true conditional expectation in the mean squared error minimizing sense; and derives the asymptotical properties of OLS estimator under misspecification. An analogous result in the case of quantile regression has recently been established by Angrist, Chernozhukov and Fernandez-Val (2006), albeit in a weighted mean squared error minimizing sense.

Since the pioneering work of Haavelmo (1943), simultaneous equation models have been used in empirical analysis to analyse endogeneity problems caused by individual choice, market equilibrium or unobserved heterogeneity. Many applied researchers specify the triangular systems of simultaneous equation models in an ad hoc linear form. Basmann (1957) and Theil (1953) introduced well-known 2SLS estimation of linear coefficients in triangular models. Following Heckman (1979), a "control function" approach was introduced to correct for endogeneity in triangular systems of simultaneous equations. This approach amounts to the inclusion of a first stage residual as a covariate in the structural equation, therefore treating the endogeneity problem as an omitted variables problem. During the last two decades, there has been a move towards the use of nonparametric methods in triangular simultaneous equations models, thanks to methods proposed by Newey and Powell (1989) and Newey, Powell and Vella (1999), Chesher (2003), Florens et al. (2008), Imbens and Newey (2009). These last two papers analyze systems with non-additive errors.

Several interesting questions naturally arise when considering the intersection of triangular systems of simultaneous equations and linear misspecification. How well does the linear form in a triangular system approximates true underlying, potentially nonparametric, structural function? How should one interpret estimates of the pseudo-parameters[1] of the misspecified model? What are the asymptotic properties of the estimators under misspecification? The last question is particularly relevant for empirical researchers who wish to confine themselves to a linear specification because of sample size limitations. In this case, researchers would still like to know the asymptotic variance of their estimates under possible misspecification and nonobservability of the control variable to ensure robust statistical inference.

I first define the notion of structural misspecification in a triangular array of simultaneous

[1] I use the term "pseudo-parameters" to indicate the parameters of the specified model. In contrast, the word "population parameter" refers to the parameters of the underlying data generation process.

equations and discuss two different types of misspecification. I then characterize the approximation properties of the linear fit under misspecification of the structural function. I show that under linear misspecification of the structural function, the linear fit for the structural function approximates the true underlying structural function in weighted minimum squared error sense, as in Angrist et. al (2006); albeit with possibly negative weights. I show that the weights are directly related to the signs and magnitudes of both true specification error and the expectation of the specification error conditional on the control variable. In particular I conclude that, even in models with small specification error, the weights might have large variation across observations depending on the average specification error given the control variable. It must be noted that the analysis here assumes that a consistent estimator of the true control variable is available. In the case where only a contaminated control is available from the first stage (Kim and Petrin, 2011), it is difficult to trace the approximation properties of the linear fit in the second stage.

This weighted mean squared error approximation result is analogous to conclusions in Heckman and Vytlacil (2001, 2005) that unify different evaluation measures via the concept of marginal treatment effects. Specifically, I show that the pseudo-parameter defining the returns to treatment intensity can be written as a weighted average of marginal treatment effects as in Heckman and Vytlacil (2001).

I describe a two-stage estimator for the structural pseudo-parameters, and present its asymptotic properties. I use the results from the literature on asymptotics of semiparametric M-estimators described in Ichimura and Lee (2010) to demonstrate asymptotic normality of the estimators. The paper is concluded with a simulation study.

The next section defines the concept of misspecification in triangular simultaneous equation models. Section 3 explores the properties of the misspecified model. Section 4 relates the misspecified estimator to unifying theorems of Heckman and Vytlacil (2001, 2005). Section 5 gives the asymptotic distributions for estimators of misspecified models. Section 6 analyses the properties of the misspecified model by a Monte Carlo experiment.

1.2 Definition and Characterization of Misspecification

It is natural to ask the question "what does a misspecified triangular model really estimate?". Note that, unlike the single equation case, the answer to the question depends on two distinct factors: the nature of true underlying data generating process and the specific combination of various "elemental" misspecifications that I will discuss below.

The true data generating process is assumed to be the error separable triangular nonparametric simultaneous equations model as in Newey, Powell and Vella (1999) (NPV henceforth)[2],

$$Y = h(X, Z_1) + \varepsilon \tag{1.1}$$

$$X = g(Z) + \nu \tag{1.2}$$

$$E[\varepsilon|X, Z] = E[\varepsilon|X - g(Z)] = E[\varepsilon|\nu] = \lambda(\nu) \tag{1.3}$$

$$E[\nu|Z] = 0 \tag{1.4}$$

$$\eta = \varepsilon - E[\varepsilon|\nu] = \varepsilon - \lambda(\nu) \text{ is such that } E[\eta^2] < \infty \tag{1.5}$$

where Y is the outcome variable, X is the univariate endogenous variable. Z is a $d_z \times 1$ vector of instrumental variables that consists of a subvector Z_1 of size $d_{z1} \times 1$ and a subvector Z_2 of size $d_{z2} \times 1$ with $d_{z1} + d_{z2} = d_{z.}$ $g(Z)$ is a function of instruments Z and ε, ν are the disturbances. This is effectively a nonparametric generalization of the linear triangular simultaneous equations model. Theorem 2.3 of NPV concludes that assuming $\lambda(\nu) = E[\varepsilon|\nu, Z] = E[\varepsilon|\nu]$, $g(Z)$ and $h(X, Z_1)$

2 The instrumental equation could be specified nonseparably as in Matzkin (2003), without affecting our approximation analysis. In that case, the control variable to be included in the estimation of the second stage would be the cumulative distribution function of the endogenous variable conditional on the set of instruments.

differentiable, and (Z, ε) has a density that vanishes at the boundary and $rank(\partial g(Z)/\partial Z_2) = 1$; $h(X, Z_1)$ is identified. Notice that, if $g(Z)$ were linear in Z, then the rank condition would be precisely the identification condition of a linear simultaneous equations system in terms of reduced form coefficients. The identification of $h(X, Z_1)$ is obviously up to an additive constant, however, with a simple location restriction such as $\lambda(\nu^*) = \lambda^*$ for some known ν^*,λ^* then complete identification of $h(X, Z_1)$ is achieved.

Linear misspecification is defined as one or more of the components in the structural equation being specified in a linear form. I focus on the cases, where linear forms are assumed for functions $h(X, Z_1)$ and/or $\lambda(\nu)$:

$$Y = X\beta + Z_1'\gamma + \tilde{\varepsilon} \tag{1.6}$$

$$E[\tilde{\varepsilon}|\nu, Z] = E[\tilde{\varepsilon}|\nu] = \phi(\nu), \tag{1.7}$$

and

$$\phi(\nu) = c + \alpha\nu. \tag{1.8}$$

Notice that misspecification of (1.1) as (1.6) and (1.7) as (1.8) may exist together. Specifically, the linear misspecification of the control function may arise under the false assumption that (ε, ν) is jointly normal or entirely out of convenience. Although White (1980) gives a single equation interpretation about misspecification of this type, it is silent about how well the linear fit for the structural function alone approximates the structural function.

When the control function is specified nonparametrically and (1.1) is misspecified as (1.6); the analysis will be the one for a misspecified partially linear regression model.

1.3 Interpretation of Linear Fit under Misspecification

1.3.1 Complete Linear Misspecification

$Y = h(X, Z_1) + \varepsilon$ is misspecified as

$$Y = X\beta + Z_1'\gamma + \tilde{\varepsilon}$$

and in addition, the control function $E[\tilde{\varepsilon}|\nu, Z] = E[\tilde{\varepsilon}|\nu] = \phi(\nu)$ is misspecified as $c + \alpha\nu$, with the rest of the system correctly specified. The linearity assumption regarding the control function might be motivated by false assumption about the joint normality of the structural error and instrumental error. On the other hand, as I have discussed in the introduction, linear misspecification of the structural function might arise out of convenience or out of sample size limitations.

For any t_0, t_1, t_2, t_3 define structural specification error $\Delta_{h,t_0,t_1}(X, Z_1) = h(X, Z_1) - Xt_0 - Z_1't_1$ and specification error of the control function $\Delta_{\lambda,t_2,t_3}(\nu) = \lambda(\nu) - t_2 - \nu t_3$. Under Complete Linear Misspecification, I can make the following proposition in line with White (1980).

Proposition 1 *Under Complete Linear Misspecification; the fit $X\beta + Z_1'\gamma + c + \alpha\nu$ is a linear projection of the conditional expectation $E[Y|\nu, Z, X] = h(X, Z_1) + \lambda(\nu)$ on X,Z_1 and ν.*

Therefore, one can interpret $X\beta + Z_1'\gamma + c + \alpha\nu$ as the best linear prediction to the conditional expectation $E[Y|\nu, Z, X] = h(X, Z_1) + \lambda(\nu)$; minimizing total specification error $\Delta_{\lambda,t_2,t_3}(\nu) + \Delta_{h,t_0,t_1}(X, Z_1)$ in mean squared error sense along the lines of White (1980).

1.3.2 Structural Linear Misspecification

$Y = h(X, Z_1) + \varepsilon$ is misspecified as

$$Y = X\beta + Z_1'\gamma + \tilde{\varepsilon}$$

with the rest of the system correctly specified

$$X = g(Z) + \nu,\ E[\varepsilon|\nu, Z] = E[\varepsilon|\nu] = \lambda(\nu),\ E[\nu|Z] = 0.$$

The resulting misspecified partially linear regression model consists of pseudo-parameters $(\beta, \gamma')'$

and a function $\phi(\nu)$

$$Y = X\beta + Z_1'\gamma + \phi(\nu) + \omega, \tag{1.9}$$

with error $\omega = \tilde{\varepsilon} - E[\tilde{\varepsilon}|\nu]$. For any t_0, t_1 define the structural specification error

$$\Delta_{h,t_0,t_1}(X, Z_1) = h(X, Z_1) - Xt_0 - Z_1't_{1.} \tag{1.10}$$

The following theorem establishes the population quantity that pseudo-parameters $(\beta, \gamma')'$ minimize; and asserts that it is the integrated conditional variance of the specification error $\Delta_{h,\beta,\gamma}(X, Z_1)$.

Theorem 1 *Under structural linear misspecification; pseudo-parameters β and γ minimize the specification error $\Delta_{h,t_0,t_1}(X,Z_1) = h(X,Z_1) - Xt_0 - Z_1't_1$ in Integrated Conditional Variance (ICV) sense:*

$$(\beta,\gamma')' \,\epsilon\, argmin_{t_0,t_1}\ E[Var[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]] = \int\limits_{\nu\in\nu} Var[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]\sigma(\nu).d\nu \quad (1.11)$$

where $\sigma(\nu)$ is the density of error ν. Also, the control function at the misspecified partially linear model 1.9 is related to the true control function (1.3) in DGP with $\phi(\nu) = \lambda(\nu) + E[\Delta_{h,\beta,\gamma}(X,Z_1)|\nu]$.

The last part of the above theorem states that the specification error of the control function is $\Delta_{\lambda,\phi}(\nu) = \lambda(\nu) - \phi(\nu) = -E[\Delta_{h,\beta,\gamma}(X,Z_1)|\nu]$. This means that the control function $\phi(\nu)$ in 1.9 absorbs the specification error $\Delta_{h,\beta,\gamma}(X,Z_1)$ after conditioning it on control variate. Keeping that in mind I may state following lemma;

Lemma 1 *Under structural linear misspecification, the pseudo-parameters $(\beta,\gamma')'$ minimize the distance between mean squared specification error (MSSE) of the structural function and MSSE of the control function;*

$$(\beta,\gamma')' \,\epsilon\, argmin_{t_0,t_1} E[\Delta^2_{h,t_0,t_1}(X,Z_1)] - E[\Delta^2_{\lambda,f_{(t_0,t_1)}}(\nu)] \quad (1.12)$$

where the structural specification error is absorbed by the control function, that is, for any t_0, t_1, $f_{(t_0,t_1)}(\nu) = E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu] + \lambda(\nu)$ so the specification error of the control function is $\Delta_{\lambda,f_{(t_0,t_1)}}(\nu) = \lambda(\nu) - f_{(t_0,t_1)}(\nu)$ where $f_{(\beta,\gamma)}(\nu) = \phi(\nu)$.

Note that above Lemma 1 directly follows from Theorem 1 simply because the structural specification error is absorbed into the specification error of the control function conditional on ν, that is, $-E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu] = \lambda(\nu) - f_{(t_0,t_1)}(\nu) = \Delta_{\lambda,f_{(t_0,t_1)}}(\nu)$. A few words can be said about Lemma 1, which is in contrast with Proposition 1. Under complete linear misspecification, $(\beta,\gamma',\alpha,c)'$ minimizes the total specification error, $\Delta_{\lambda,f_{(t_0,t_1)}}(\nu) + \Delta_{h,t_0,t_1}(X,Z_1)$, in Mean Squared Error (MSE) sense. Under structural linear misspecification, $(\beta,\gamma',\phi(.))'$ minimizes the difference between mean squared specification error of the structural function and the control function. In a way, the optimization 1.12 aims to equalize expected squared structural error and expected squared control function error, while minimizing total specification error $\Delta_{\lambda,f_{(t_0,t_1)}}(\nu) + \Delta_{h,t_0,t_1}(X,Z_1)$ in MSE sense. In empirical studies, it might worth being aware of this particular character of the misspecification. Note the minimized spread in Theorem 1 can also be interpreted as the recentered expectation of the squared specification error, similar in spirit to White (1980).

It is possible to express the above problem as a weighted minimization.

Theorem 2 *Under structural linear misspecification, the true pseudo-parameter vector* $(\beta, \gamma')'$ *is the solution to a weighted MSE problem*

$$(\beta, \gamma')' \in \arg\min_{t_0, t_1} E[\Delta^2_{h,t_0,t_1}(X, Z_1)\mathcal{W}(Z, X)]$$

where the weights

$$w(Z, X; \beta, \gamma) = \begin{cases} 0 & \Delta_{h,\beta,\gamma}(X, Z_1) = 0 \\ 1 - \frac{E(\Delta_{h,\beta,\gamma}(X,Z_1)|\nu)}{\Delta_{h,\beta,\gamma}(X,Z_1)} & \textit{otherwise} \end{cases}. \tag{1.13}$$

This result contrasts with that of White (1980) for single equation mean regression and that of Angrist et al. (2006) for single equation quantile regression models. Although I establish a weighted approximation property like the latter do, I do so with possibly negative weights. The weights given above are negative whenever the true specification error unexplained by conditioning on the control variable, that is, π at

$$\Delta_{h,\beta,\gamma}(X, Z_1) = E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu) + \pi$$

is such that

$$1 > \frac{\pi}{\Delta_{h,\beta,\gamma}(X, Z_1)}$$

It is easy to see that the weight is directly related to the sign and the magnitude of the specification error and the average specification error conditional on the control variable. Contrary to initial perception, the weight does not need to be increasing as the portion of the specification error explained by the control variable decreases: When $E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)$ and $\Delta_{h,\beta,\gamma}(X, Z_1)$ have opposite signs; the weight will indeed decrease as the specification error explained by the control variable, $E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)$, decreases.

1.3.3 Omitted Variable Bias Formula

The above specification is useful to write down an omitted variables-like formula for parameters of the true structural function when it is a polynomial. Assume that the true structural function is

$$h(X, Z_1) = X\theta_0 + Z_1'\theta_1 + \sum_{l=2}^{L} X^l \theta_l \tag{1.14}$$

When $\sum_{l=2}^{L} X^l \theta_l$ is omitted from the model, using Theorem 2, I can show that the resulting parameters $(\beta, \gamma')'$ in the misspecified model

$$Y = X\beta + Z_1'\gamma + \phi(\nu) + \omega$$

are related to the parameters of the true structural function in 1.14 as follows:

$$\begin{aligned}
\begin{pmatrix}\beta \\ \gamma\end{pmatrix} &= E\left[\mathcal{W}_i(Z,X)\begin{pmatrix} X_i^2 & X_i Z_{1i}' \\ X_i Z_{1i} & Z_{1i}Z_{1i}'\end{pmatrix}\right]^{-1} E\left[\mathcal{W}_i(Z,X)\begin{pmatrix} X_i(X_i\theta_0 + Z_{1i}'\theta_1 + \sum_{l=2}^{L} X^l\theta_l) \\ Z_{1i}(X_i\theta_0 + Z_{1i}'\theta_1 + \sum_{l=2}^{L} X^l\theta_l)\end{pmatrix}\right] \\
&= E\left[\mathcal{W}_i(Z,X)\begin{pmatrix} X_i^2 & X_i Z_{1i}' \\ X_i Z_{1i} & Z_{1i}Z_{1i}'\end{pmatrix}\right]^{-1} E\left[\mathcal{W}_i(Z,X)\begin{pmatrix} X_i^2\theta_0 + X_i Z_{1i}'\theta_1 + X_i\sum_{l=2}^{L} X^l\theta_l \\ Z_{1i}X_i\theta_0 + Z_{1i}Z_{1i}'\theta_1 + Z_{1i}\sum_{l=2}^{L} X^l\theta_l\end{pmatrix}\right] \\
&= E\left[\mathcal{W}_i(Z,X)\begin{pmatrix} X_i^2 & X_i Z_{1i}' \\ X_i Z_{1i} & Z_{1i}Z_{1i}'\end{pmatrix}\right]^{-1} \\
&\quad\times\left(E\left[\mathcal{W}_i(Z,X)\begin{pmatrix} X_i^2 & X_i Z_{1i}' \\ X_i Z_{1i} & Z_{1i}Z_{1i}'\end{pmatrix}\begin{pmatrix}\theta_0 \\ \theta_1\end{pmatrix}\right] + E\left[\mathcal{W}_i(Z,X)\begin{pmatrix} X_i\sum_{l=2}^{L} X^l\theta_l \\ Z_{1i}\sum_{l=2}^{L} X^l\theta_l\end{pmatrix}\right]\right) \\
&= \begin{pmatrix}\theta_0 \\ \theta_1\end{pmatrix} + E\left[\mathcal{W}_i(Z,X)\begin{pmatrix} X_i^2 & X_i Z_{1i}' \\ X_i Z_{1i} & Z_{1i}Z_{1i}'\end{pmatrix}\right]^{-1} E\left[\sum_{l=2}^{L}\mathcal{W}_i(Z,X)\begin{pmatrix} X_i \\ Z_{1i}\end{pmatrix} X^l\theta_l\right].
\end{aligned}$$

Denoting $\widetilde{X}_i = (X_i, Z_{1i}')'$, $\theta_P = (\theta_2, ..., \theta_L)'$ and $X_i^P = (X_i^2, X_i^3, ..., X_i^L)'$

$$= \begin{pmatrix}\theta_0 \\ \theta_1\end{pmatrix} + E\left[\mathcal{W}_i\widetilde{X}_i\widetilde{X}_i'\right]^{-1} E\left[\mathcal{W}_i\widetilde{X}_i(X_i^P)'\right]\theta_P$$

where

$$\mathcal{W}_i(Z,X) = 1 - \frac{E(\Delta_{\theta_0,\theta_1,\theta_P,\beta,\gamma}(X,Z_1)|\nu)}{\Delta_{\theta_0,\theta_1,\theta_P,\beta,\gamma}(X,Z_1)} = 1 - \frac{E(X|\nu_i)(\theta_0-\beta) + E(Z_1'|\nu_i)(\theta_1-\gamma) + E((X_i^P)'|\nu_i)\theta_P}{X_i(\theta_0-\beta) + Z_{1i}'(\theta_1-\gamma) + (X_i^P)'\theta_P}$$

with

$$\nu_i = X_i - g(Z_{1i}, Z_{2i}) = X_i - g(Z_i).$$

$E\left[\mathcal{W}_i \widetilde{X}_i \widetilde{X}_i'\right]^{-1} E\left[\mathcal{W}_i \widetilde{X}_i (X_i^P)'\right] \theta_P$ is the bias resulting from the linear misspecification.

1.4 Relation with Treatment Effects

It is possible to interpret the pseudo-parameter β as an aggregate measure of treatment effects where the treatment intensity X is a continuous endogenous regressor. When the true DGP is as in NPV, the Marginal Treatment Effect (MTE) of Heckman and Vytlacil (2001) adapted to the continous treatment case (Florens et al. 2008) becomes

$$\begin{aligned} MTE(X_i, Z_{1i}, \nu_i) &= E\left(\frac{\partial}{\partial X}\left(h(X, Z_1) + \lambda(\nu)\right) | X = X_i, Z_1 = Z_{1i}, \nu = \nu_i\right) \quad (1.15) \\ &= E\left(h_1(X, Z_1) | X = X_i, Z_1 = Z_{1i}, \nu = \nu_i\right) = h_1(X_i, Z_{1i}) \end{aligned}$$

where $h_1(X_i, Z_{1i})$ denotes the partial derivative of function $h(.)$ with respect to its first argument. We know from the proof of Theorem 4 that the pseudo-parameter vector $(\beta, \gamma')'$ can be written as follows

$$\begin{aligned} (\beta, \gamma')' &= E[\widetilde{X}_i \widetilde{X}_i{}' \mathcal{W}(Z_i, X_i)]^{-1} E[\widetilde{X}_i h(X_i, Z_{1i}) \mathcal{W}(Z_i, X_i)] \\ &= E[\widetilde{X}_i \widetilde{X}_i{}' \mathcal{W}_i]^{-1} E[\widetilde{X}_i h(\widetilde{X}_i) \mathcal{W}_i] \end{aligned}$$

where $\widetilde{X}_i = (X_i, Z_{1i}')'$ as before. Note that; $(\beta, \gamma')'$ above can be described as the parameters in the parsimonious instrumental variable model

$$\mathcal{W} h(X, Z_1) = \mathcal{W} \widetilde{X}'(\beta, \gamma')' + \zeta \text{ with } E(\zeta \mid \widetilde{X}) = 0 \text{ by construction} \quad (1.16)$$

with $\widetilde{X}$ being the vector of instruments. The mean independence condition implies,

$$\begin{aligned} E\left[\mathcal{W} h(X, Z_1) - \mathcal{W} \widetilde{X}'(\beta, \gamma')' | \widetilde{X}\right] &= 0 \text{ taking expectation on both sides,} \\ E\left[\mathcal{W} h(X, Z_1)\right] &= E\left[\mathcal{W} \widetilde{X}'\right] (\beta, \gamma')' . \end{aligned}$$

Assuming $\mu_{\mathcal{W}} = E\left[\mathcal{W}\right] < \infty$ and differentiating the above condition w.r.t. treatment intensity X_i, I obtain

$$\begin{aligned} E\left[\mathcal{W}h_1(X,Z_1) - \mathcal{W}\beta\right] &= 0 \\ E\left[\left(\frac{\mathcal{W}}{\mu_{\mathcal{W}}}\right)h_1(X,Z_1)\right] &= \beta \end{aligned}$$

Using the MTE as a building block, I show that the pseudo-parameter β, which defines the returns to the treatment intensity X_i, is weighted average of the marginal treatment effects $h_1(X,Z_1)$ where weights are given by $\left(\frac{\mathcal{W}}{\mu_{\mathcal{W}}}\right)$. Having thus proven it, we state this result at below Theorem.

Theorem 3 *Assume that the weight function 1.13 has a finite mean $\mu_{\mathcal{W}} = E\left[\mathcal{W}\right]$.Then under structural linear misspecification, the pseudo-parameter that defines the returns to treatment intensity is a weighted average of marginal treatment effects with weights $\left(\frac{\mathcal{W}}{\mu_{\mathcal{W}}}\right)$,*

$$\beta = E\left[\left(\frac{\mathcal{W}}{\mu_{\mathcal{W}}}\right)h_1(X,Z_1)\right].$$

1.5 Estimators for Pseudo-Parameters under Misspecification and Their Asymptotic Properties

We now describe estimators of pseudo-parameters under structural linear misspecification and derive asymptotic properties.

We start with the first step estimation. Suppose the function $g(Z)$ in

$$X = g(Z) + \nu$$

is estimated using series estimators, as in NPV or Carneiro and Lee (2009). Let $\{B_k(Z) : k = 1, 2, ...\kappa\}$ denote tensor product B-spline basis for real valued smooth functions defined on the compact support $[-1,1]^{d_z}$ such that a linear combination of $\{B_k(Z) : k = 1, 2, ...\kappa\}$ can approximate $g(Z)$ as the number of approximating functions, κ goes to infinity. Z here can be rescaled to lie within the rectangle $[-1,1]^{d_z}$.

Let $B_\kappa(Z_i)$ denote the tensor product B-spline basis $\{B_k(Z_i) : k = 1, 2, ...\kappa\}$. The estimator of

$g(Z)$, $\widehat{g}(Z)$ is obtained by series estimation

$$\widehat{g}(Z_i) = B_\kappa(Z_i)'\widehat{\delta}_{n\kappa}$$

where

$$\widehat{\delta}_{n\kappa} = \left(\sum_{i=1}^{n} B_\kappa(Z_i)B_\kappa(Z_i)'\right)^{-1}\left(\sum_{i=1}^{n} B_\kappa(Z_i)X_i\right)$$

consequently, the estimator of ν_i is

$$\widehat{\nu}_i = X_i - \widehat{g}(Z_i)$$

The convergence properties of $\widehat{g}(Z_i)$ follows from standard results in the series estimation literature. (Newey, 1997)

Remember that my misspecified model is partially linear;

$$Y_i = X_i\beta + Z_{1i}'\gamma + \phi(\widehat{\nu}_i) + \omega_i \tag{1.17}$$

The pseudo-parameters $(\beta, \gamma')'$ and the function $\phi(.)$ in 1.9 can be consistently estimated with the estimator given in Robinson (1988)[3]. Denote $A_i = 1(\widehat{\nu}_i \in \mathcal{V})$ where $\mathcal{V}$ is a compact set on which continuously distributed ν has a density f_ν bounded away from zero. Then Robinson's estimator for pseudoparameters $(\beta, \gamma')'$ is given by;

$$\begin{aligned}(\widehat{\beta}, \widehat{\gamma}')' &= \left[\sum_{i=1}^{n} A_i\left(\widetilde{X}_i - E(\widetilde{X}|\widehat{\nu}_i, A_i)\right)\left(\widetilde{X}_i' - E(\widetilde{X}'|\widehat{\nu}_i, A_i)\right)\right]^{-1} \\ &\quad \times\left[\sum_{i=1}^{n} A_i\left(\widetilde{X}_i - E(\widetilde{X}|\widehat{\nu}_i, A_i)\right)(Y_i - E(Y|\widehat{\nu}_i, A_i))\right]\end{aligned}$$

A trimming function of the type $A_i = 1(\widehat{\nu}_i \in \mathcal{V})$ is considered here to avoid the noise at the tails of the distribution of $\widehat{\nu}$.

Assumption 1 The data, $\{(Y_i, X_i, Z_i) : i = 1, ..., n\}$ is independent and identically distributed, and the $\nu ar(X|Z)$ is bounded.

Assumption 2 The support of Z lies within the rectangle $[-1, 1]^{d_z}$ on which Z has an absolutely continuous probability density which is bounded above by a positive constant and bounded away from zero.

[3] Alternatively, one could employ Speckman's (1988) estimator.

Assumption 3 The function $g(Z)$ is $r_g - times$ continuously differentiable on $[-1,1]^{d_z}$ with $r_g > 2d_z$.

Assumption 1-3 are standard in the series estimation literature. See Li and Racine (2007, pp.450-451) for details.

Assumption 4 Continuously distributed ν has a density σ_ν that is bounded away from zero on a compact set ν.

The following theorem is standard in the series estimation literature.

Theorem 4 *Suppose assumptions 1-4 hold. Then*

$$\max_{i=1,.....n} |\widehat{\nu}_i - \nu_i| = \max_{i=1,.....n} |\widehat{g}(Z_i) - g(Z_i)| = O_p\left(\frac{\kappa}{n^{1/2}} + \kappa^{-\left(\frac{2r_g - d_z}{2d_z}\right)}\right)$$

Denote $\widetilde{X}_i = (X_i, Z'_{1i})'$ as before.

Assumption 5 The conditional expectation $E(\widetilde{X}|\nu_i, A_i)$ is twice continuously differentiable with repect to ν_i and its kernel estimator $\widehat{E}_h(\widetilde{X}|\nu_i, A_i)$ is consistent uniformly over $\nu \in \mathcal{V}$. Furthermore, assume that

$$\sup_{\nu \in \nu} \left| \widehat{E}_h(\widetilde{X}|\nu_i, A_i) - E(\widetilde{X}|\nu_i, A_i) \right| = o_p(n^{-1/4})$$

Assumption 6 $\kappa^4/n \to 0$ and $\kappa^{(2r_g/d_z)}/n \to \infty$

Note that Assumption 6 is satisfied with $\kappa \propto n^a$ with $(d_z/2r_g) < a < (1/4)$.

Define

$$\Sigma_0 = E\left[\left(\widetilde{X}_i - E(\widetilde{X}|\nu_i, A_i)\right)\left(\widetilde{X}'_i - E(\widetilde{X}'|\nu_i, A_i)\right) \times A_i\right]$$
$$\Phi(Z_i) = E\left[-A_i \times \left(\widetilde{X}_i - E(\widetilde{X}|\nu_i, A_i)\right)\left(\left.\frac{\partial \phi(\nu)}{\partial \nu}\right|_{\nu=\nu_i}\right)\middle| Z = Z_i\right]$$

and

$$\begin{aligned} \Omega_0 &= E\left[A_i \times \omega_i^2 \times \left(\widetilde{X}_i - E(\widetilde{X}|\nu_i, A_i)\right)\left(\widetilde{X}'_i - E(\widetilde{X}'|\nu_i, A_i)\right)\right] \\ &\quad + E\left[\nu_i^2 \times \Phi(Z_i) \times \Phi(Z_i)'\right] \end{aligned}$$

Assumption 7 Σ_0 is positive definite, $\Phi(Z_i)$ is continuously differentiable with respect to its argument, $E\left[\Phi(Z_i) \times \Phi(Z_i)'\right]$ is nonsingular and Ω_0 is finite.

The following theorem gives the asymptotic distribution of the estimators of the pseudoparameters.

Theorem 5 *Under Structural Linear Misspecification; the estimators $(\widehat{\beta}, \widehat{\gamma}')'$ for pseudo-parameters $(\beta, \gamma')'$ are asymptotically normal as*

$$\sqrt{n}\left(\begin{pmatrix}\widehat{\beta}\\ \widehat{\gamma}\end{pmatrix} - \begin{pmatrix}\beta\\ \gamma\end{pmatrix}\right) \to_d N(0, \Sigma_0^{-1}\Omega_0\Sigma_0^{-1})$$

The term $E\left[\nu_i^2 \times \Phi(Z_i) \times \Phi(Z_i)'\right]$ arises since ν_i is not observed and must be estimated.

1.6 Simulation

In order to illustrate the nature of the structural misspecification; I present the results of a Monte Carlo experiment here. The true DGP assumed is;

$$\begin{aligned} Y &= 3X + X^2 + 2Z_1 + \varepsilon \\ X &= \sin Z_1 + \log Z_2 + \nu \end{aligned}$$

where Z_2 is the square root of a folded standard normally distributed random variable, $Z_2 = \sqrt{|N|}$ and $Z_1 = N^3U/10$ where N refers to a standard normally distributed random variable and U a random variable drawn from uniform $[0, 1]$ distribution. Likewise, the unobservable

$$\varepsilon = 0.25\nu^2 + \eta$$

where ν and η are normalized independently distributed random variables. They are both constructed from interaction of a set of independent standard normal and uniform $[0, 1]$ distributions, i.e. $\nu = NU$ and $\eta = NU$. We use MATLAB software to perform the simulation. I choose this slightly complex specification of DGP in order to illustrate that the previous results are robust to the specification of DGP.

We first want to estimate the probability limit of the estimates obtained from weighted minimization problem at Theorem 4, and want to see whether this estimated limit coincides with the true-pseudo parameters as predicted by Theorem 4. Since it is not possible to obtain true pseudo-parameters algebraically, I use the estimate of probability limit of Robinson's estimators (1988) of the model 1.9. Since I want to increase the accuracy of the estimates of these probability limits, when I estimate the model 1.9 I use true ν at this stage. The estimates of the probability limits are obtained by simply averaging the estimates of the model across replications. The conditional expectations involved in Robinson's estimation (1988) are replaced by Nadaraya-Watson estimators using rule of thumb bandwidths given in Li and Racine (2007, pp. 66).

Secondly, I need to contruct the weight function as accurately as possible. Using the estimate of the probability limit of estimators of model 1.9, I construct the weight function 1.13. However, the weight function 1.13 involves a conditional expectation which can not be obtained algebraically. We therefore replace the conditional expectation $E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)$ with its Nadaraya-Watson estimate as well.

We set the sample size $n = 10000$ and the number of replications to be 1000. We first obtain $(\widehat{\beta}^r_{PL}, \widehat{\gamma}^r_{PL})$ for $r = 1, ..., 1000$ from the partially linear model 1.9 using the true ν. The average of $(\widehat{\beta}^r_{PL}, \widehat{\gamma}^r_{PL})$'s across replications, $(\overline{\beta}_{PL}, \overline{\gamma}_{PL})$ should, by consistency of the estimates; be close to the true pseudo parameters. We then estimate the weights at 1.13 using near-true values for pseudo-parameters $(\overline{\beta}_{PL}, \overline{\gamma}_{PL})$ I obtained previously and obtain estimates from the weighted linear regression $(\widehat{\beta}^r_W, \widehat{\gamma}^r_W)$. Denote the estimate of the probability limit of $(\widehat{\beta}^r_W, \widehat{\gamma}^r_W)$ with $(\overline{\beta}_W, \overline{\gamma}_W)$.

	$(\overline{\beta}_{PL}, \overline{\gamma}_{PL})$	$(\overline{\beta}_W, \overline{\gamma}_W)$
β	1.0335	0.9909
γ	2.6359	2.6950

Table 1: Probability Limits of Estimators of Partially Linear Model and Weighted Linear Fit

The estimated probability limit of the estimates obtained from weighted minimization, $(\overline{\beta}_W, \overline{\gamma}_W)$, is, as predicted, fairly close to the "true" pseudo-parameter values, $(\overline{\beta}_{PL}, \overline{\gamma}_{PL})$. The gap between $(\overline{\beta}_W, \overline{\gamma}_W)$ and $(\overline{\beta}_{PL}, \overline{\gamma}_{PL})$, despite the large sample size and large number of replications, is possibly due to the fact that I replace the conditional expectations in the above procedure with their Nadaraya-Watson estimates.

To explore the effects of two stage estimation, I estimate a misspecified partially linear model;

$$Y = \beta X + \gamma Z_1 + \phi(\widehat{\nu}) + \zeta$$

where $\widehat{\nu}$ is the residual from the first step estimation. We estimate $\widehat{g}(Z_1, Z_2)$ in

$$X = g(Z_1, Z_2) + \nu$$

by series estimation described in Section 1.5. As basis functions $B_\kappa(Z_i)$ I use the tensor product of cubic B-splines with 4 knots placed equiprobably. Then I estimate the partially linear model using the estimator described in Section 1.5 with a trimming function that eliminates the observations with estimated density of $\widehat{\nu}$ below the 5th percentile of the density estimates. Table 2 gives the estimates from a random replication.

Estimates of the Partially Linear Model			
Parameter	**Estimate**	**Standard Error**	**Corrected Standard Error**
β	1.2163	0.0147	0.0755
γ	2.4498	0.0105	0.0120
Estimates of the Weighted Least Squares Approximation			
Parameter	**Estimate**	**Standard Error**	
β	1.0406	0.0269	
γ	2.6287	0.0366	

Table 2

Looking at Table 2, the corrected standard errors are higher due to the two-stage nature of the estimation. The kernel density of the weights from the same replication is given at Figure 1.1.

Figure 1.1 reveals that the distribution is highly centered. As I have pointed out earlier in Section 1.4, in order to interpret the pseudo-parameter β as a weighted average of marginal treatment effects, above distribution needs to have finite mean. This seems very likely to be true in this particular example, although further numerical analysis is required to be certain.

1.7 Conclusion

In this work the nature of linear misspecification of structural function in a triangular

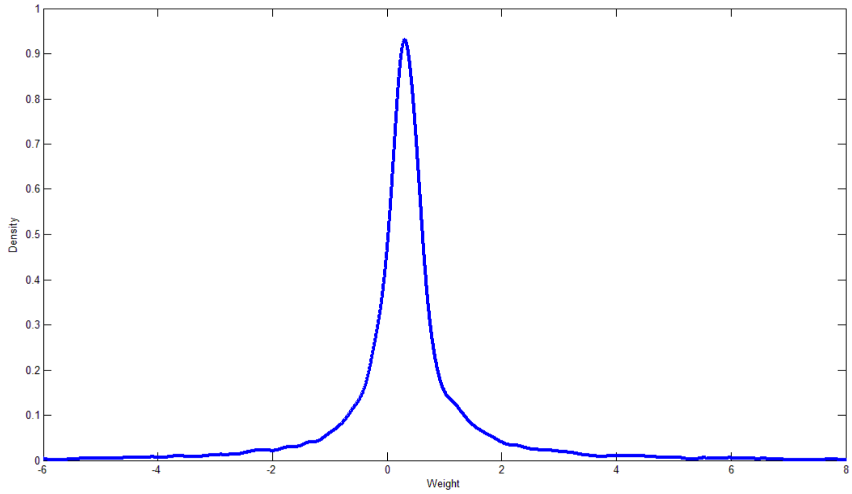

Figure 1.1: Kernel Density of Weights from a Random Replication

simultaneous equations system is explored. It is found that the linear fit approximates the structural function in a weighted mean squared error sense, possibly with negative weights. The pseudo-parameters of the misspecified model are also found to be the average marginal treatment effects, when the weights have finite mean. Building on the weighted approximation result, an omitted variable formula is derived for cases where the true structural function is a polynomial.

There are a few routes left to explore. The nature of total linear misspecification of the triangular system, that is, when both the instrumental equation and the structural equation are linearly misspecified, is left for future work. Future research should also explore the nature of linear instrumental variable estimators when the true DGP is a fully nonparametric triangular simultaneous equations system.

1.8 Proofs

Proof. [Proposition 1] Note that from the usual OLS problem I can write

$$
\begin{aligned}
(c,\alpha,\beta,\gamma) &\in \arg\min_{t_0,t_1,t_2,t_3} E[Y - Xt_0 - Z_1't_1 - t_2 - t_3\nu]^2 \\
&= \arg\min_{t_0,t_1,t_2,t_3} E[h(X,Z_1) + \lambda(\nu) + \eta - Xt_0 - Z_1't_1 - t_2 - t_3\nu]^2 \\
&= \arg\min_{t_0,t_1,t_2,t_3} E[\Delta_{h,t_0,t_1}(X,Z_1) + \Delta_{\lambda,t_2,t_3}(\nu) + \eta]^2 \\
&= \arg\min_{t_0,t_1,t_2,t_3} [E[\Delta_{h,t_0,t_1}(X,Z_1) + \Delta_{\lambda,t_2,t_3}(\nu)]^2 \\
&\quad + E\eta^2 + 2E[\eta(\Delta_{h,t_0,t_1}(X,Z_1) + \Delta_{\lambda,t_2,t_3}(\nu))]] \\
&= \arg\min_{t_0,t_1,t_2,t_3} E[\Delta_{h,t_0,t_1}(X,Z_1) + \Delta_{\lambda,t_2,t_3}(\nu)]^2
\end{aligned}
$$

The last two terms cancel to zero since η is mean independent of any measurable function of X,Z,ν; and $E\eta^2$ is a constant with respect to the minimization. QED. ■

Proof. [Theorem 1] Note that pseudoparametes minimize

$$
\begin{aligned}
(\beta,\gamma) &\in \arg\min_{t_0,t_1} E\left[(Y - E[Y|\nu]) - (X - E[X|\nu])'t_0 - (Z_1 - E[Z_1|\nu])'t_1\right]^2 \\
&= \arg\min_{t_0,t_1} E(h(X,Z_1) - Xt_0 - Z_1't_1 + \varepsilon - \lambda(\nu) \\
&\quad - E[h(X,Z_1)|\nu] - E[X|\nu]t_0 - E[Z_1|\nu]'t_1]]^2 \\
&= \arg\min_{t_0,t_1} E(\Delta_{h,t_0,t_1}(X,Z_1) + \varepsilon - \lambda(\nu) - E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]]^2 \\
&= \arg\min_{t_0,t_1} E(\Delta_{h,t_0,t_1}(X,Z_1) + \eta - E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]]^2 \\
&= \arg\min_{t_0,t_1} E[\Delta_{h,t_0,t_1}(X,Z_1) - E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]]^2 \\
&\quad + E[\eta^2] + 2E[\eta(\Delta_{h,t_0,t_1}(X,Z_1) - E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu])] \qquad (1.18)
\end{aligned}
$$

where $\eta = \varepsilon - E[\varepsilon|\nu,Z] = \varepsilon - \lambda(\nu)$ with $E[\eta|\nu,Z] = 0$. Therefore, η is mean orthogonal to any function of ν and Z, thus X. The last two terms cancel to zero

$$\begin{aligned}(\beta,\gamma) &\in \arg\min_{b,\gamma} E[\Delta_{h,t_0,t_1}(X,Z_1) - E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]]^2 \\ &= \arg\min_{t_0,t_1} E[E[(\Delta_{h,t_0,t_1}(X,Z_1) - E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu])^2|\nu]] \text{ (by law of iterated expectations)} \\ &= \arg\min_{t_0,t_1} E[Var[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]] \qquad (1.19)\end{aligned}$$

As for $\phi(\nu)$, note that

$$\begin{aligned}\phi(\nu) &= E[\widetilde{\varepsilon}|\nu] = E[Y - X\beta - Z_1'\gamma|\nu] \\ &= E[h(X,Z_1) + \varepsilon - X\beta - Z_1'\gamma|\nu] \\ &= E[\Delta_{h,\beta,\gamma}(X,Z_1) + \varepsilon|\nu] \\ &= E[\Delta_{h,\beta,\gamma}(X,Z_1)|\nu] + E[\varepsilon|\nu] \\ &= E[\Delta_{h,\beta,\gamma}(X,Z_1)|\nu] + \lambda(\nu).\end{aligned}$$

QED. ■

Proof. [Lemma 1] From Theorem 2

$$(\beta,\gamma)'\epsilon \arg\min_{t_0,t_1} E[Var[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]] = E[E[\Delta^2_{h,t_0,t_1}(X,Z_1)|\nu]] - E[[E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]]^2].$$

Noting $E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]] = -\Delta_{\lambda,f_{(t_0,t_1)}}(\nu)$ where $f_{(t_0,t_1)}(\nu) = E[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]] + \lambda(\nu)$, therefore the control

$$\begin{aligned}&(\beta,\gamma)' \,\epsilon\, argmin_{t_0,t_1}\; E[\Delta^2_{h,t_0,t_1}(X,Z_1)] - E[[-\Delta_{\lambda,f(t_0,t_1)}(\nu)]^2] \\ =\;& E[\Delta^2_{h,t_0,t_1}(X,Z_1)] - E[\Delta^2_{\lambda,f(t_0,t_1)}(\nu)]\,.\end{aligned}$$

QED. ■

Proof. [Theorem 2] Directly from Theorem 2, $(\beta,\gamma)'$ solve

$$\min_{t_0,t_1}\; E[Var[\Delta_{h,t_0,t_1}(X,Z_1)|\nu]]$$

$$= \min_{t_0,t_1} E[E(\Delta^2_{h,t_0,t_1}(X,Z_1)|\nu) - E(\Delta_{h,t_0,t_1}(X,Z_1)|\nu)^2]$$

$$= \min_{t_0,t_1}[E[\Delta^2_{h,t_0,t_1}(X,Z_1) - E(\Delta_{h,t_0,t_1}(X,Z_1)|\nu)^2|\Delta_{h,t_0,t_1}(X,Z_1)=0]\Pr(\Delta_{h,t_0,t_1}(X,Z_1)=0)$$

$$+E[\Delta^2_{h,t_0,t_1}(X,Z_1) - E(\Delta_{h,t_0,t_1}(X,Z_1)|\nu)^2|\Delta_{h,t_0,t_1}(X,Z_1)\neq 0]\Pr(\Delta_{h,t_0,t_1}(X,Z_1)\neq 0)].$$

Since X and Z have continuous support, $\Pr(\Delta_{h,t_0,t_1}(X, Z_1) = 0) = 0$ for any t_0, t_1.Then,

$$\min_{t_0,t_1} E[Var[\Delta_{h,t_0,t_1}(X, Z_1)|\nu]]$$

$$= \min_{t_0,t_1} E[\Delta^2_{h,t_0,t_1}(X, Z_1) - E(\Delta_{h,t_0,t_1}(X, Z_1)|\nu)^2 | \Delta_{h,t_0,t_1}(X, Z_1) \neq 0].$$

By first order condition,

$$0 = -2E[\widetilde{X}\Delta_{h,t_0,t_1}(X, Z_1) - E(\Delta_{h,t_0,t_1}(X, Z_1)|\nu)E(\widetilde{X}|\nu)|\Delta_{h,t_0,t_1}(X, Z_1) \neq 0],$$

this equality must hold at the true β, γ;

$$0 = -2E[\widetilde{X}\Delta_{h,\beta,\gamma}(X, Z_1) - E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)E(\widetilde{X}|\nu)|\Delta_{h,\beta,\gamma}(X, Z_1) \neq 0].$$

Define $\zeta = \widetilde{X} - E(\widetilde{X}|\nu)$, which is by construction mean orthogonal to $E(\widetilde{X}|\nu)$. Then

$$E[E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)E(\widetilde{X}|\nu)|\Delta_{h,\beta,\gamma}(X, Z_1) \neq 0]$$

$$= E[(E(\widetilde{X}|\nu) + \zeta)E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)|\Delta_{h,\beta,\gamma}(X, Z_1) \neq 0]$$

$$= E[\widetilde{X}E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)|\Delta_{h,\beta,\gamma}(X, Z_1) \neq 0].$$

Plugging this into the the first order condition;

$$\begin{aligned} 0 &= -2E[\widetilde{X}\Delta_{h,\beta,\gamma}(X, Z_1) - \widetilde{X}E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)|\Delta_{h,\beta,\gamma}(X, Z_1) \neq 0] \\ &= -2E\left[\widetilde{X}\Delta_{h,\beta,\gamma}(X, Z_1) - \widetilde{X}\Delta_{h,\beta,\gamma}(X, Z_1)\left(\frac{E(\Delta_{h,\beta,\gamma}(X, Z_1)|\nu)}{\Delta_{h,\beta,\gamma}(X, Z_1)}\right)|\Delta_{h,\beta,\gamma}(X, Z_1) \neq 0\right], \end{aligned}$$

which is the first order condition of the minimization problem

$$\min_{t_0,t_1} E[\Delta_{h,t_0,t_1}(X, Z_1)w(Z, X; \beta, \gamma)$$

with

$$w(Z, X; \beta, \gamma) = \begin{cases} 0 & \Delta_{h,\beta,\gamma}(X, Z_1) = 0 \\ 1 - \frac{E(\Delta_{h,\beta,\gamma}(X,Z_1)|\nu)}{\Delta_{h,\beta,\gamma}(X,Z_1)} & \text{otherwise} \end{cases}.$$

QED. ■

Proof. [Theorem 4] Noting that $|\widehat{\nu}_i - \nu_i| = |\widehat{g}(Z_i) - g(Z_i)|$, the proof follows from Theorem 7 of Newey (1997) given in Corollary 15.1 of Li and Racine (2007). QED. ■

Proof. [Theorem 5] Our proof and notation rely heavily on the proof of Theorem 3 found in Carneiro and Lee (2009, hereafter CL). In principle, the theorem is proven by verifying the assumptions of the general Theorem 3.2 in Ichimura and Lee (2010, hereafter IL). Like CL, I make the derivations implicit in the trimming function $A_i = 1(\widehat{\nu}_i \in \mathcal{V})$. Then the estimator $\widehat{\theta}$ can be written as an M-estimator with;

$$m[(y,x,z),t,f(.)] = \frac{1}{2}[y - f_1(x - f_3(z)) - \{\widetilde{x} - f_2(x - f_3(z))\}'t]^2$$

where $f = f_1, f_2, f_3$ are the nonparametric components. The true function $f_0 = f_{10}, f_{20}, f_{30}$ is given by $f_{10}(.) = E[Y|\nu = .]$, $f_{20} = E[\widetilde{X}|\nu = .]$ and $f_{30} = E[X|Z = .] = g(.)$

Assumption 3.1(a) of IL is not necessary in my case because my estimator minimizes a convex objective function. Assumptions 3.1(b) is satisfied with the assumption that Σ_0 is positive definite[4]. The consistency of the estimator follows from Assumption 5, thus Assumption 3.1(c) of IL is satisfied. Assumptions 3.2 and 3.3 are trivial given the form of my objective function. Following our Theorem 4, Assumption 6 implies that;

$$\max_{i=1,....,n} |\widehat{g}(Z_i) - g(Z_i)| = o_p\left(n^{-1/4}\right)$$

Using above and Assumption 5, Assumption 3.4 of IL is satisfied. Given the form of m, it is straightforward to verify Assumption 3.5 of IL[5]. Assumption 3.6 of IL is the critical assumption which characterizes the effect of first stage estimation. Following the notation of IL;

$$\begin{aligned}
D_{f_1}m^*(t, f_0(.))[h_1(.)] &= -E[\{(Y - E[Y|\nu]) - (\widetilde{X} - E[\widetilde{X}|\nu])'t\}h_1(.)]; \\
D_{f_2}m^*(t, f_0(.))[h_2(.)] &= E[\{(Y - E[Y|\nu]) - (\widetilde{X} - E[\widetilde{X}|\nu])'t\}h_2(.)'t]; \\
D_{f_3}m^*(t, f_0(.))[h_3(.)] &= -E[\{(Y - E[Y|\nu]) - (\widetilde{X} - E[\widetilde{X}|\nu])'t\} \\
&\quad \times\{-\left.\frac{\partial f_{10}(\nu)}{\partial \nu}\right|_{\nu=\nu} + \left.\frac{\partial f_{20}(\nu)}{\partial \nu}\right|_{\nu=\nu} t\}h_3(.)].
\end{aligned}$$

4 see Section 4 of Robinson (1988)

5 see Proposition 3.1, Example 4.1 and 4.2 of IL on verification of this assumption.

Then;

$$\begin{aligned} \frac{\partial}{\partial t} D_{f_1} m^*(t, f_0(.))[h_1(.)]|_{t=\theta} &= E[(\widetilde{X} - E[\widetilde{X}|\nu])' h_1(.)] \\ &= E[(\widetilde{X} - E[\widetilde{X}|\nu])'(\widehat{E}[Y|\nu] - E[Y|\nu])] \\ &= 0 \end{aligned}$$

the last equality follows from the fact that the residual $(\widetilde{X} - E[\widetilde{X}|\nu])$ is mean orthogonal to any measurable function of ν. Similarly;

$$\frac{\partial}{\partial t} D_{f_2} m^*(t, f_0(.))[h_2(.)]|_{t=\theta} = 0$$

$$\frac{\partial}{\partial t} D_{f_3} m^*(t, f_0(.))[h_3(.)]|_{t=\theta}$$

$$= E\left[-(\widetilde{X} - E[\widetilde{X}|\nu]) \left.\frac{\partial \phi(\nu)}{\partial \nu}\right|_{\nu=\nu} h_3(.)\right]$$

$$= E\left[-(\widetilde{X} - E[\widetilde{X}|\nu]) \left.\frac{\partial \phi(\nu)}{\partial \nu}\right|_{\nu=\nu} (\widehat{g}(Z) - g(Z))\right]$$

Therefore, only the third term affects the asymptotic distribution. Its limiting behavior at $\widehat{g}(Z) - g(Z)$ can be described by the argument at Section 4 of Newey (1997), because it is a linear functional of $\widehat{g}(Z) - g(Z)$. Assumption 7 of Newey (1997) is satisfied with $\Phi(Z)$ given previously (see Assumption 7 at Section 1.5). Then the proof follows from Theorem 3.2 of IL with the restriction $n^{1/2}\kappa^{-(r_g/d_z)} \to 0$ given at Assumption 6. QED. ■

References

Angrist, J., Chernozhukov, V. and Fernández-Val, I., (2006), "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure", *Econometrica* 74, 539-563

Basmann, R. L., (1957), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation", *Econometrica* 25, 77-83

Carneiro, P., and Lee, S., (2009), "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality", *Journal of Econometrics* 149, 191-208

Chesher, A., (2003), "Identification in Nonseparable Models", *Econometrica* 71, 1405-1441

Cramer, J., (1969), Empirical Econometrics, *London: North Holland Publishing Co.*

Denny, M., and Fuss, M., (1977), "The Use of Approximation Analysis to Test for Separability and the Existence of Consistent Aggregates," *American Economic Review* 67, 404-418.

Fan, J., and Gijbels, I., (1996), Local Polynomial Modelling and Its Applications, Chapman-Hill

Florens, J. P., Heckman, J. J., Meghir, C., and Vytlacil, E. J., "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects", *Econometrica* 76, 1191-1206

Haavelmo, T., (1943), "The Statistical Implications of a System of Simultaneous Equations", *Econometrica* 11, 1-12

Heckman, J. J., (1979), "Sample Selection Bias as a Specification Error", *Econometrica* 47, 153-161

Heckman, J. J., and Vytlacil, E., (2001), "Policy-Relevant Treatment Effects", *The American Economic Review* 91, 107-111

Heckman, J. J., and Vytlacil, E., (2005), "Structural Equations, Treatment Effects and Econometric Policy Evaluation" *Econometrica* 73, 669-73

Ichimura, H., and Lee, S., (2010) "Characterization of the Asymptotic Distribution of semiparametric M-estimators", *Journal of Econometrics* 159, 252-266

Imbens, G. W., and Newey, W. K., (2009), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity", *Econometrica* 77, 1481-1512

Kim, K., and Petrin, A., (2011), "Revisiting the Classic Control Function Approach, with Implications for Parametric and Non-Parametric Regressions", mimeo, University of Minnesota Department of Economics

Li, Q, and Racine, J. S., (2007), Nonparametric Econometrics: Theory and Practice, Princeton University Press

Matzkin, R. L., (2003), "Nonparametric Estimation of Nonadditive Random Functions", *Econometrica* 71, 1339-1375

Newey, W. K., (1994), "The asymptotic variance of semiparametric estimators", *Econometrica* 62, 1349-1382.

Newey, W. K., (1997), "Convergence rates and asymptotic normality for series estimators", *Journal of Econometrics* 79, 147-168.

Newey, W. K., and Powell, J. L., (1989), "Nonparametric Instrumental Variables Estimation," Working Paper, MIT Department of Economics.

Newey, W. K., Powell, J. L., and Vella, F., (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica* 67, 565-603

Robinson, P. M., (1988), "Root-N-Consistent Semiparametric Regression", *Econometrica* 56, 931-954

Speckman, P., (1988), "Kernel Smoothing in Partial Linear Models", *Journal of The Royal Statistical Society. Series B.* 50, 413-436

Theil, H., (1953), "Estimation and Simultaneous Correlation in Complete Equation Systems", *mimeo*, The Hague: Centraal Planbureau

White, H., (1980), "Using Least Squares to Approximate Unknown Regression Functions", *International Economic Review* 21, 149-170