A Stochastic Frontier Model with Endogenous Treatment Status and Mediator

Yi-Ting Chen, Yu-Chin Hsu, Hung-Jen Wang

IEAS Working Paper No. 14-A006
December, 2014

Institute of Economics
Academia Sinica
Taipei 115, TAIWAN
http://www.sinica.edu.tw/econ/

Copyright © 2014 (Yi-Ting Chen, Yu-Chin Hsu, Hung-Jen Wang)
A Stochastic Frontier Model with Endogenous Treatment Status and Mediator

Yi-Ting Chen
Institute of Economics
Academia Sinica

Yu-Chin Hsu
Institute of Economics
Academia Sinica

Hung-Jen Wang
Department of Economics
National Taiwan University, and
Institute of Economics
Academia Sinica

This version: December 15, 2014

*We thank the editors, referees, and participants of 2014 Asia-Pacific Productivity Conference in Brisbane for very helpful comments and suggestions.
Abstract

Government policies are frequently used throughout the world to promote productivity. While some of the policies are designed to work through technology enhancement, others are meant to exert the influence through efficiency improvement. For a given technology and efficiency channel, another issue to consider in designing and evaluating policies is whether a mediator is required/effective in achieving the final outcome. To help better understanding and evaluating the policies, we propose a new stochastic frontier model with treatment effect and mediators. The model allows us to decompose the total program (treatment) effect into the technology and efficiency components, and it also allows us to investigate whether the effect takes place directly from the program or indirectly through a mediator. Both of the treatment status and the mediator are allowed to be endogenous in the fully-parametric model. We illustrate the empirical application of the model using the data of India to study the effects of large dams on the local districts’ agricultural production.

Keywords: stochastic frontier models, treatment effect, mediation analysis

1 Introduction

In developing as well as developed countries, government policies are often instituted for the purpose of increasing productivity at the industry as well as the firm levels. The increase in productivity may be achieved either by promoting better technology or by improving production efficiency, and there are ample examples of policies that are designed specifically for either (or both) of the channels. For a given channel, another issue to consider in

\footnote{The Argentine federal government sets up the Argentinean Technological Fund to provide programs and funding for helping private firms to gain competitiveness through technological innovation (Chudnovsky et al. 2007). The Mexican federal government instituted the Program for Training the Industrial Workforce between the year 2001 and 2006, with the goal of increasing workforce efficiency for the small and medium enterprises (SMEs; Lopez-Acevedo and Tinajero-Bravo 2011). Starting in 1991, Chile government set up the National Productivity and Technological Development Fund to provide support for the development of new products and production processes, technology transfer and acquisition, and the investment in investigating}
designing and evaluating policies is whether a mediator is required/effective in achieving the final outcome. For instance, R&D subsidy programs may help to elevate technology by way of building advanced labs. Building dams may improve agricultural production efficiency (as well as changing the method of production) through irrigation systems. In these examples, new labs and irrigation systems are mediators in the channels, and the policy may exert all or part of the effect on final outcome through the mediator in the channel. Mechanisms of mediators can therefore be included as part of the policy or be used in evaluating the policy.

Given that such policies are widely adopted by government throughout the world, it is important to have methods for evaluating the policy effectiveness. Does the policy increase productivity at all? Through which channel –technology advancement or efficiency improvement– does the effect take place? Are there effective mediators in the channel? Finally, does the policy work in the way it is designed for?

To help answer the above questions, we propose in this paper a new stochastic frontier model with treatment effect and mediators. The model merges the stochastic frontier analysis with the program evaluation methods, which is well-developed in the literature, and the mediation analysis which is currently growing and gaining traction in the literature. In the framework of stochastic frontier analysis, we show that the model is able to disentangle the frontier (viz., technology) and the (in)efficiency components of the program effect, and it may further decompose the total program effect into a direct and an indirect (viz., mediator) ones for each of the frontier and inefficiency components. The model may thus potentially useful technologies (Benavente et al. 2007). In Taiwan, the SME Administration in the Ministry of Economic Affairs provides various management consulting programs for the SMEs with the goals of improving efficiency and increasing profitability.

\(^2\)This is equivalent of decomposing the total, direct, and indirect effects of a program into their frontier and inefficiency components. As in the literature on program evaluation and mediation analysis, we can only measure these effects and their components in the group of “compliers” because of the endogenous treatment status.
help policy makers to understand which channel, technology or efficiency, is important in conducing the policy effect on productivity, and whether the policy exerts the effect directly on the outcome or indirectly through the mediator.

A feature of the proposed model is that it allows for the treatment status to be endogenous. This is important because program participation is often voluntary-based. Following Abadie (2003), we deal with the parameter identification issue in the presence of the endogeneity by using a binary instrument variable (IV). Furthermore, to make the model even more general, we also include an endogenous binary mediator in the model in addition to the endogenous treatment status. We apply a recent method by Frölich and Huber (2014), which extended the treatment effect model to the context of mediation analysis, for parameter identification. We show that, after integrating the stochastic frontier model into a parametric structure model of output, mediator and treatment status, the model parameters can be identified in a two-step approach and estimated by a two-stage weighted nonlinear least square (WNLS) method. As a consequence, we can easily establish the asymptotic properties of the WNLS estimator (WNLSE) by a traditional two-stage estimation theory. The remaining parameters of the entire parametric structure model can also be easily identified and estimated. A Monte Carlo analysis is provided to show the performance of WNLSE in finite samples. Attention is paid to include endogenous treatment effect and mediation effect in the data generating process in a structural way. Results show that the WNLSE performs reasonably well.

An empirical example is provided by applying the model to estimate the effects of large dams on the local agricultural production in India. As discussed in Duflo and Pande (2007), the decision of dam construction may be endogenous because it may directly related to poverty and production outcome of the local. We follow Duflo and Pande and use river gradients to construct the instrumental variable for the dams construction decision. The mediator variable is constructed from the ratio of irrigated area to total cultivated area since irrigation is one of the major advantages of dams on agricultural production. Results
show that production efficiency was significantly improved in the local after dams were built, and the improvement took place directly from dams rather than indirectly through the mediator. The total output, however, shows insignificant increases, which is consistent with the result of Duflo and Pande (2007).

Stochastic frontier models have long been used in policy analysis. Examples are abundant, such as the study of deregulation on the efficiency gains of electricity generation (Kleit and Terrell 2001), the effect of government subsidized credit on participating financial institutions’ productivity (Wang et al. 2007), and a cross country comparison of the efficiency of health care delivery (Greene 2005). Most of the studies, however, were not in a treatment effect framework and the endogenous treatment issue was often ignored. A few exceptions include Crespo-Cebada et al. (2013) in which the authors studied the impact of school ownership on the technical efficiency of schools. In the analysis student performance is the outcome variable, and the endogeneity arises because families with different socio-economic status may self-select into different types of schools. A propensity score matching approach was proposed to solve the problem.

The treatment effect model with endogenous treatment status and a binary IV is introduced by Imbens and Angrist (1994). Abadie (2003) and Frölich (2007) generalized the IV framework to allow for covariates in the model which is also adopted by Hong and Nekipelov (2010), and Donald, Hsu and Lieli (2014a, 2014b). Among these, Abadie (2003) and Hong and Nekipelov (2010) are closest to our current study where they considered parameters defined by moment conditions, but their models do not contain mediators and do not allow for the mediation analysis. For treatment effect models with exogenous treatment status, please see Imbens and Wooldridge (2009) for a review.

Our model is also related to the mediation analysis which is relatively new in the literature. Our model is similar to that of Frölich and Huber (2014) which has a nonparametric structure model and the mediator can be continuous or discrete. In comparison, we adopt a parametric structure model to account for the fact that the stochastic fron-
tier analysis usually relies on a parametric specification to separate the frontier function from the inefficiency component. This framework allows us to define conditional effects on covariates, and is also easier for establishing asymptotic distributions for our estimators. Meanwhile, we focus on a binary mediator for simplicity. The use of a continuous mediator would complicate our model and estimation method. We leave it for a future study. Flores and Flores-Lagunes (2009) and Huber (2013) also conducted mediation analysis but they were in the framework of exogenous treatment status. The mediation analysis was also conducted in various fields; see, for example, Rosenbaum (1984), Robins and Greenland (1992), Pearl (2001), Rubin (2004), Petersen et al. (2006), MacKinnon (2008), and Imai et al. (2010). See also Huber (2013) and Frölich and Huber (2014) for more discussions.

The rest of the paper is organized as the follows. Section 2 lays out the basic model and the assumptions therein, which is followed by the identification and estimation methods in Section 3 and 4. A Monte Carlo analysis is conducted in Section 5 to show the finite sample performance of the estimator, and an empirical example is provided in Section 6. Section 7 concludes the paper.

2 Model and Effects

Let $Y$ be the observable output of an individual or a firm, and $X$ be a column vector of observable covariates. In the literature, the stochastic frontier model for $Y|X$ has been widely applied to various empirical studies. Conventionally, this model is of the form:

$$Y = h(X, \beta^h) + v - u,$$

where $v$ is a pure random error with $E[v|X] = 0$, $u \geq 0$ is a non-negative random variable which represents the production inefficiency with the conditional mean:

$$E[u|X] = g(X, \beta^g);$$

(2.1)

(2.2)
$h(\cdot)$ is a frontier function with the parameter vector $\beta^h$, and $g(\cdot)$ is a non-negative (conditional mean) inefficiency function with the parameter vector $\beta^g$. These two functions are of distinctive parametric forms which allow $\beta^h$ and $\beta^g$ to be separately identified if the functions share common covariates. In this study, we extend this traditional model to a new stochastic frontier model for program evaluation and mediation analysis with endogenous treatment and mediator.

### 2.1 A New Stochastic Frontier Model

Corresponding to $(Y, X^\top)^\top$, where “$X^\top$” is the transpose of $X$, we let $D$ be a binary random variable which represents the treatment status about the program being evaluated; $D$ has the outcome $d = 1$ for treatment and $d = 0$ for no treatment. We also let $M$ be a binary random variable, with the outcomes $\{0, 1\}$, which represents a post-treatment mediator between $D$ and $Y$. Both $D$ and $M$ are endogenous. We are interested in the program-evaluation problem: evaluating the effectiveness of $D$ on $Y$ and the mediation analysis: analyzing whether and how $D$ directly influences $Y$ and indirectly affects $Y$ through the channel of $M$. Since both tasks involve counterfactual assessments, we conduct the potential-outcome setting of $M$ and $Y$ to facilitate the analyses; see, e.g., Huber (2013) and Frölich and Huber (2014) for the setting.

In this setting, the mediator $M$ is considered as a function of $D$, denoted by $M(D)$, and the output $Y$ is set as a function of $D$ and $M$, denoted by $Y(D, M)$. The random variables $M(D)$ and $Y(D, M)$ are, respectively, a potential mediator and a potential output that provide an assessment about how $M$ would potentially vary with $D$ and how $Y$ would potentially change with $D$ and $M$, respectively. Given the event $D = d$, $M(d)$ is the same

---

3The theoretical framework of the model allows the frontier function and the inefficiency function to have the same or different covariates, and we use a general $X$ in these two functions in the theoretical discussion for notational simplicity. For empirical applications, tests or economic justifications may be provided to determine which function a covariate should enter.
as the actual $M$, and $M(1-d)$ is a counterfactual $M$ about what $M$ would happen if the outcome of $D$ was $1-d$; $Y(d, M(d))$ is the same as the actual $Y$, and $Y(d, M(1-d))$, $Y(1-d, M(d))$ and $Y(1-d, M(1-d))$ are, respectively, the counterfactual $Y$'s about what $Y$ would happen when $(D, M) = (d, M(1-d))$, $(1-d, M(d))$ and $(1-d, M(1-d))$. By construction, $Y$ and $Y(d, M(d))$ have the relationship:

$$Y = D \times Y(1, M(1)) + (1-D) \times Y(0, M(0)).$$

(2.3)

Let $Z_1$ be an exogenous binary IV with the outcome $z_1 \in \{0,1\}$, and consider $D$ a function of $Z_1$, denoted by $D(Z_1)$. The random variable $D(Z_1)$ is a potential treatment status. Given the event $Z_1 = z_1$, $D(z_1)$ is the same as the actual $D$, and $D(1-z_1)$ is a counterfactual $D$. It is customary to think of $Z_1$ as a variable that indicates whether an “exogenous incentive” to obtain treatment is present or as a variable signaling “intention to treat.” Like (2.3), $D$ and $D(z_1)$ have the relationship:

$$D = Z_1 \times D(1) + (1-Z_1) \times D(0).$$

As in Imbens and Angrist (1994), we use the four outcomes of $(D(1), D(0))$ to divide the whole population into four types of mutually disjoint subpopulations:

$$(D(1), D(0)) = \begin{cases} 
(1,1), \text{ always takers,} \\
(1,0), \text{ compliers (hereafter, denoted by } \mathcal{C}), \\
(0,1), \text{ defiers,} \\
(0,0), \text{ never takers.} 
\end{cases}$$

(2.4)

For program evaluation, it is known that, by focusing on the subpopulation of $\mathcal{C}$, we can identify the conditional local average treatment effect (CLATE):

$$CLATE(x) \equiv E[Y(1,M(1))|X=x,\mathcal{C}] - E[Y(0,M(0))|X=x,\mathcal{C}]$$

(2.5)

and the local average treatment effect (LATE):

$$LATE = E[CLATE(X)|\mathcal{C}] = E[Y(1,M(1))|\mathcal{C}] - E[Y(0,M(0))|\mathcal{C}]$$
without relying the assumption of unconfoundedness when $D$ is endogenous; see, e.g., Abadie (2003, Theorem 3.1).

For mediation analysis, we need to decompose the CLATE into a conditional direct LATE (CDLATE) and a conditional indirect LATE (CILATE). This needs not only $E[Y(0, M(0))|X, C]$ and $E[Y(1, M(1))|X, C]$ but also $E[Y(0, M(1))|X, C]$ and $E[Y(1, M(0))|X, C]$ to facilitate the decomposition of (2.5):

$$CLATE(x) = CDLATE(x) + CILATE(x),$$

where

$$CDLATE(x) \equiv E[Y(1, M(1))|X = x, C] - E[Y(0, M(1))|X = x, C]$$

(2.7)

and

$$CILATE(x) \equiv E[Y(0, M(1))|X = x, C] - E[Y(0, M(0))|X = x, C];$$

(2.8)

see Frölich and Huber (2014) for such a decomposition of LATE. The CDLATE of $D$ on $Y$ is defined by holding the mediator $M$ at $M(1)$, and the CILATE of $D$ on $Y$ is generated through the change of $M$ from $M(0)$ to $M(1)$ by holding the treatment status $D$ at $D = 0$. It is known that the decomposition is path-dependent; see Fortin, Lemieux, and Firpo (2011) for a review. One can alternatively define the CDLATE by fixing $M$ at $M(0)$ instead and define the CILATE by holding the treatment status at $D = 1$.

For program evaluation and mediation analysis in the stochastic frontier context, we propose the following potential-output model for the subpopulation of $C$: for $d, d' \in \{0, 1\}$,

$$Y(d, M(d')) = \tilde{h}(d, M(d'), X) + v(d, M(d')) - u(d, M(d'), X),$$

(2.9)
where \( \tilde{h}(d, M(d'), X) \) is a potential frontier function with the outcomes:

\[
\tilde{h}(d, M(d'), X) = \begin{cases} 
    h(X, \beta_{d1}^h), & (d, M(d')) = (1, 1), \\
    h(X, \beta_{d0}^h), & (d, M(d')) = (1, 0), \\
    h(X, \beta_{01}^h), & (d, M(d')) = (0, 1), \\
    h(X, \beta_{00}^h), & (d, M(d')) = (0, 0),
\end{cases}
\tag{2.10}
\]

\( v(d, M(d')) \) is a pure potential random error with \( E[v(d, M(d'))|X, C] = 0, u(d, M(d'), X) \) is a non-negative potential production inefficiency such that

\[
u(d, M(d')) = \tilde{g}(d, M(d'), X) + \tilde{u}(d, M(d')) \]

\( E[\tilde{u}(d, M(d'))|X, C] = 0 \), and \( \tilde{g}(d, M(d'), X) \) is a non-negative potential inefficiency function with the outcomes:

\[
\tilde{g}(d, M(d'), X) = \begin{cases} 
    g(X, \beta_{d1}^g), & (d, M(d')) = (1, 1), \\
    g(X, \beta_{d0}^g), & (d, M(d')) = (1, 0), \\
    g(X, \beta_{01}^g), & (d, M(d')) = (0, 1), \\
    g(X, \beta_{00}^g), & (d, M(d')) = (0, 0); 
\end{cases}
\tag{2.11}
\]

this model has the parameter vector: \( \beta_d \equiv (\beta_{d1}^h, \beta_{d0}^h, \beta_{d1}^g, \beta_{d0}^g) \top \) for \( d \in \{0, 1\} \).

To see the conditional expectation \( E[Y(d, M(d'))|X, C] \) implied by this new model, we use the law of total probability to show that, for \( d \in \{0, 1\} \),

\[
E[\tilde{h}(d, M(d'), X)|X, C] = P(M(d') = 1|X, C)\tilde{h}(d, 1, X) + P(M(d') = 0|X, C)\tilde{h}(d, 0, X) \\
= E[M(d')|X, C]h(X, \beta_{d1}^h) + (1 - E[M(d')|X, C]) h(X, \beta_{d0}^h)
\]

and

\[
E[\tilde{g}(d, M(d'), X)|X, C] = P(M(d') = 1|X, C)\tilde{g}(d, 1, X) + P(M(d') = 0|X, C)\tilde{g}(d, 0, X) \\
= E[M(d')|X, C]g(X, \beta_{d1}^g) + (1 - E[M(d')|X, C]) g(X, \beta_{d0}^g).
\]

As will be shown in (3.6), under suitable assumptions, we can obtain the following potential-mediator model: for \( d' \in \{0, 1\} \),

\[
E[M(d')|X, C] = m_{d'}(X, \alpha_m),
\tag{2.12}
\]
where $m_{d'}(\cdot)$ is a non-negative weighting function in $(0,1)$ with a parameter vector $\alpha_m$. By combining the potential-output model with the potential-mediator model, we can further obtain the following model for $E[Y(d, M(d'))|X, C]$: for $d, d' \in \{0, 1\}$,

$$E[Y(d, M(d'))|X, C] = h_{d'}(X, \alpha_m, \beta_{d1}^h, \beta_{d0}^h) - g_{d'}(X, \alpha_m, \beta_{d1}^g, \beta_{d0}^g),$$  

(2.13)

where

$$h_{d'}(X, \alpha_m, \beta_{d1}^h, \beta_{d0}^h) \equiv m_{d'}(X, \alpha_m)h(X, \beta_{d1}^h) + (1 - m_{d'}(X, \alpha_m))h(X, \beta_{d0}^h)$$  

(2.14)

and

$$g_{d'}(X, \alpha_m, \beta_{d1}^g, \beta_{d0}^g) \equiv m_{d'}(X, \alpha_m)g(X, \beta_{d1}^g) + (1 - m_{d'}(X, \alpha_m))g(X, \beta_{d0}^g).$$  

(2.15)

This model is of the parameter vector: $(\alpha_m^\top, \beta_{d1}^h, \beta_{d0}^h)^\top$ for $d \in \{0, 1\}$.

By comparing (2.13) with the specification of $E[Y|X]$ implied by the traditional stochastic frontier model in (2.1):

$$E[Y|X] = h(X, \beta^h) - g(X, \beta^g),$$

we can observe that the proposed model is substantially different from the traditional model with two important features. Firstly, the proposed model comprises a system of stochastic frontier models for the latent variables: $Y(0, M(0))$, $Y(0, M(1))$, $Y(1, M(0))$ and $Y(1, M(1))$ and a pair of models for $M(0)$ and $M(1)$, and its conditioning set is also latent because of the inclusion of $C$. Thus, unlike traditional stochastic frontier models, it needs a new parameter identification method. Secondly, the proposed model allows the parameters of the frontier function $h(\cdot)$ and the inefficiency function $g(\cdot)$ to vary with $(d, M(d'))$. Therefore, unlike traditional stochastic frontier models, it connects the program evaluation and mediation analysis to the stochastic frontier context.

In addition, although the CLATE in (2.5) and its decomposition into CDLATE and CILATE in (2.6) may be identified and estimated in a nonparametric way as considered
by the literature of program evaluation and mediation analysis, the proposed model allows us to further decompose these treatment effects, direct effects and indirect effects into their production-frontier components and production-inefficiency components, as discussed below. Such an extension is undoubtedly important for stochastic frontier analysis, and it is facilitated by the use of the parametric specification in (2.13).

2.2 Conditional Local Average Treatment Effects

Suppose that, for \(d, d' \in \{0, 1\}\), (2.13) is correctly specified for \(E[Y(d, M(d'))|X, C]\), and the parameter vector \((\alpha_m^\top, \beta^\top_d)\) is identifiable as will be discussed in the next section. Let \(x\) denote a value of \(X\), which represents a particular type of observable features of individuals.

For program evaluation in the stochastic frontier context, according to (2.13), (2.14) and (2.15), we can decompose (2.5) into:

\[
CLATE(x) = CLATE_h(x) - CLATE_g(x),
\]

with the CLATE on the production frontier:

\[
CLATE_h(x) \equiv h_1(x, \alpha_m, \beta_{11}^h, \beta_{10}^h) - h_0(x, \alpha_m, \beta_{01}^h, \beta_{00}^h)
\]

\[
= m_1(x, \alpha_m)h(x, \beta_{11}^h) + (1 - m_1(x, \alpha_m)) h(x, \beta_{10}^h)
\]

\[
- \left( m_0(x, \alpha_m)h(x, \beta_{01}^h) + (1 - m_0(x, \alpha_m)) h(x, \beta_{00}^h) \right)
\]

and the CLATE on the production inefficiency:

\[
CLATE_g(x) \equiv g_1(x, \alpha_m, \beta_{11}^g, \beta_{10}^g) - g_0(x, \alpha_m, \beta_{01}^g, \beta_{00}^g)
\]

\[
= m_1(x, \alpha_m)g(x, \beta_{11}^g) + (1 - m_1(x, \alpha_m)) g(x, \beta_{10}^g)
\]

\[
- \left( m_0(x, \alpha_m)g(x, \beta_{01}^g) + (1 - m_0(x, \alpha_m)) g(x, \beta_{00}^g) \right).
\]

For mediation analysis in this context, we can further decompose (2.7) and (2.8) into:

\[
CDLATE(x) = CDLATE_h(x) - CDLATE_g(x)
\]

(2.17)
\[ \text{CILATE}(x) = \text{CILATE}_h(x) - \text{CILATE}_g(x), \]  

(2.18)

respectively, with the CDLATE on the production frontier:

\[ \text{CDLATE}_h(x) \equiv h_1(x, \alpha_m, \beta_{11}^h, \beta_{10}^h) - h_0(x, \alpha_m, \beta_{01}^h, \beta_{00}^h) = m_1(x, \alpha_m) \left( h(x, \beta_{11}^h) - h(x, \beta_{01}^h) \right) + (1 - m_1(x, \alpha_m)) \left( h(x, \beta_{10}^h) - h(x, \beta_{00}^h) \right), \]

the CDLATE on the production inefficiency:

\[ \text{CDLATE}_g(x) \equiv g_1(x, \alpha_m, \beta_{11}^g, \beta_{10}^g) - g_0(x, \alpha_m, \beta_{01}^g, \beta_{00}^g) = m_1(x, \alpha_m) \left( g(x, \beta_{11}^g) - g(x, \beta_{01}^g) \right) + (1 - m_1(x, \alpha_m)) \left( g(x, \beta_{10}^g) - g(x, \beta_{00}^g) \right), \]

the CILATE on the production frontier:

\[ \text{CILATE}_h(x) \equiv h_1(x, \alpha_m, \beta_{01}^h, \beta_{00}^h) - h_0(x, \alpha_m, \beta_{01}^h, \beta_{00}^h) = (m_1(x, \alpha_m) - m_0(x, \alpha_m)) \left( h(x, \beta_{01}^h) - h(x, \beta_{00}^h) \right), \]

and the CILATE on the production inefficiency:

\[ \text{CILATE}_g(x) \equiv g_1(x, \alpha_m, \beta_{01}^g, \beta_{00}^g) - g_0(x, \alpha_m, \beta_{01}^g, \beta_{00}^g) = (m_1(x, \alpha_m) - m_0(x, \alpha_m)) \left( g(x, \beta_{01}^g) - g(x, \beta_{00}^g) \right). \]

Studying these conditional effects is useful for exploring whether and how the treatment influences the outputs of the individuals in \( \mathcal{C} \) with the features characterized by \( X = x \).

Corresponding to these decompositions, we can establish a set of parameter restrictions for program evaluation:

- \( H_p^o \) : \( \beta_{11}^h = \beta_{10}^h = \beta_{01}^h = \beta_{00}^h \) and \( \beta_{11}^g = \beta_{10}^g = \beta_{01}^g = \beta_{00}^g \) for no \( \text{CLATE}(x), \forall x \);
- \( H_{p^h} \) : \( \beta_{11}^h = \beta_{10}^h = \beta_{01}^h = \beta_{00}^h \) for no \( \text{CLATE}_h(x), \forall x \);
- \( H_{p^g} \) : \( \beta_{11}^g = \beta_{10}^g = \beta_{01}^g = \beta_{00}^g \) for no \( \text{CLATE}_g(x), \forall x \).
In addition, as will be seen in (3.6), the difference between the functions \( m_1(\cdot) \) and \( m_0(\cdot) \) is fully determined by a parameter \( \alpha_d \), which represents the partial effect of \( D \) on \( M \), such that \( m_1(\cdot) = m_0(\cdot) \) if \( \alpha_d = 0 \); otherwise, \( m_1(\cdot) \neq m_0(\cdot) \). Correspondingly, we can also establish another set of parameter restrictions for mediation analysis:

- \( H^d_0 : (\beta^{h}_{11}, \beta^{h}_{10}, \beta^{g}_{11}, \beta^{g}_{10}) = (\beta^{h}_{01}, \beta^{h}_{00}, \beta^{g}_{01}, \beta^{g}_{00}) \) for no \( CDLATE(x) \), \( \forall x \);
- \( H^{dh}_0 : (\beta^{h}_{01}, \beta^{h}_{00}) = (\beta^{h}_{11}, \beta^{h}_{10}) \) for no \( CDLATE_h(x) \), \( \forall x \);
- \( H^{dg}_0 : (\beta^{g}_{01}, \beta^{g}_{00}) = (\beta^{g}_{11}, \beta^{g}_{10}) \) for no \( CDLATE_g(x) \), \( \forall x \);
- \( H^i_0 : (\beta^{h}_{01}, \beta^{g}_{01}) = (\beta^{h}_{00}, \beta^{g}_{00}) \) or \( \alpha_d = 0 \) for no \( CILATE(x) \), \( \forall x \);
- \( H^{ih}_0 : \beta^{h}_{01} = \beta^{h}_{00} \) or \( \alpha_d = 0 \) for no \( CILATE_h(x) \), \( \forall x \);
- \( H^{ig}_0 : \beta^{g}_{01} = \beta^{g}_{00} \) or \( \alpha_d = 0 \) for no \( CILATE_g(x) \), \( \forall x \).

Note that the hypotheses: \( H^i_0, H^{ih}_0 \) and \( H^{ig}_0 \) can be simplified as the parameter restrictions:

\[
(\beta^{h}_{01}, \beta^{g}_{01}) = (\beta^{h}_{00}, \beta^{g}_{00}), \quad \beta^{h}_{01} = \beta^{h}_{00} \quad \text{and} \quad \beta^{g}_{01} = \beta^{g}_{00},
\]

respectively, when \( \alpha_d \neq 0 \). By testing these hypotheses and comparing the testing results, we may assess the significance of the effectiveness of \( D \) on \( Y \) and disentangle the underlying channels of the influences of \( D \) on \( Y \) for the subpopulation of \( C \) conditional on \( X = x \).

### 2.3 Local Average Treatment Effects

In addition, we may also extend (2.6) to a decomposition of LATE:

\[
LATE = DLATE + ILATE,
\]

with the direct LATE (DLATE):

\[
DLATE = E[CDLATE(X)|C] = E[Y(1,M(1))|C] - E[Y(0,M(1))|C]
\]

and the indirect LATE (ILATE):

\[
ILATE = E[CILATE(X)|C] = E[Y(0,M(1))|C] - E[Y(0,M(0))|C],
\]

13
and use (2.13) to further show the decompositions:

$$LATE = LATE_h - LATE_g,$$

(2.19)

$$DLATE = DLATE_h - DLATE_g$$

(2.20)

and

$$ILATE = ILATE_h - ILATE_g,$$

(2.21)

where

$$LATE_h \equiv E[CLATE_h(X)|C], \quad LATE_g \equiv E[CLATE_g(X)|C],$$

$$DLATE_h \equiv E[CDLATE_h(X)|C], \quad DLATE_g \equiv E[CDLATE_g(X)|C],$$

$$ILATE_h \equiv E[CILATE_h(X)|C], \quad \text{and} \quad ILATE_g \equiv E[CILATE_g(X)|C].$$

Analyzing these effects is useful for implementing the program evaluation and mediation analysis in the subpopulation of $C$ without focusing on a particular type of individual features. Similar to the hypotheses: $H^p_o, H^{ph}_o, H^{pg}_o, H^d_o, H^{dh}_o, H^{dg}_o, H^i_o, H^{ih}_o$ and $H^{ig}_o$, we also define a set of moment restrictions for program evaluation:

- $\tilde{H}^p_o : LATE = 0$,
- $\tilde{H}^{ph}_o : LATE_h = 0$,
- $\tilde{H}^{pg}_o : LATE_g = 0$,

and another set of moment restrictions for mediation analysis:

- $\tilde{H}^d_o : DLATE = 0$,
- $\tilde{H}^{dh}_o : DLATE_h = 0$,
• $\tilde{H}_{\alpha}^{g} : DLATE_{g} = 0$,
• $\tilde{H}_{\alpha}^{i} : ILATE = 0$,
• $\tilde{H}_{\alpha}^{ih} : ILATE_{h} = 0$,
• $\tilde{H}_{\alpha}^{ig} : ILATE_{g} = 0$.

Testing these hypotheses allows us to evaluate the effectiveness of $D$ on $Y$ and disentangle its underlying channels for the subpopulation of $C$ without conditional on $X = x$.

In Table 1, we summarize the CLATE and LATE components. This table also shows that the proposed method connects the program evaluation (the mediation analysis) to the stochastic frontier analysis by further decomposing the CLATE and LATE (the associated direct and indirect effects) on the output into their frontier and inefficiency components. This extension allows researchers to investigate whether and how the program influences the output in a detailed way.

3 Identification

In this section, we demonstrate that the parameter vectors $\alpha_{m}$ and $\beta_{d}$, for $d \in \{0, 1\}$, are identifiable from the distribution of $W \equiv (Y, M, D, Z_{1}, Z_{2}, X^{\top})^{\top}$, where $Z_{2}$ is a continuous IV for the endogenous $M$, and show that the CLATE and LATE components are also identifiable from the distribution of $W$.

3.1 Identification of Parameters

Let $B_{d}$ be the parameter space of $\beta_{d}$ for $d \in \{0, 1\}$. Corresponding to $\beta_{d} = (\beta_{d1}^{h\top}, \beta_{d0}^{h\top}, \beta_{d1}^{g\top}, \beta_{d0}^{g\top})^{\top}$, we let $b_{d} = (b_{d1}^{h\top}, b_{d0}^{h\top}, b_{d1}^{g\top}, b_{d0}^{g\top})^{\top}$ be an arbitrary vector in $B_{d}$. From (2.13), (2.14) and (2.15), we can observe that $\beta_{d}$ is separable from $\alpha_{m}$. Thus, we let the identification of $\beta_{d}$ be conditional on that of $\alpha_{m}$, and make the following assumption:
Assumption 3.1 Assume that

(a) $\alpha_m$ is identifiable,

(b) \((2.13)\) holds for some $\beta_d \in B_d$ and for $d \in \{0, 1\}$, and

(c) $E\left[\left( \left( h_{d'}(X, \alpha_m, \beta_{d'_1}^{d'}, \beta_{d'_0}^{d'}) - g_{d'}(X, \alpha_m, \beta_{d'_1}^{g}, \beta_{d'_0}^{g}) \right) \right)^2 \right] \geq 0$, for $d, d' \in \{0, 1\}$ and for all $b_d \in B_d$ such that $b_d \neq \beta_d$.

This assumption is standard for the parameter identification of a nonlinear regression. As will be discussed later, Assumption 3.1(a) can be shown to be valid under certain conditions. Assumption 3.1(b) requires \((2.13)\) to be correctly specified for $E[Y(d, M(d'))|X, C]$, and Assumption 3.1(c) means that, given $\alpha_m$, the parameter vector $\beta_d$ is unique. Assumption 3.1 implies that, for $d \in \{0, 1\}$,

$$
\beta_d \equiv \arg\min_{b_d \in B} \sum_{d' = 0, 1} E \left[ \left( Y(d, M(d')) - h_{d'}(X, \alpha_m, b_{d'_1}^{d'}, b_{d'_0}^{d'}) + g_{d'}(X, \alpha_m, b_{d'_1}^{g}, b_{d'_0}^{g}) \right)^2 \right] | C.
$$

\((3.1)\)

$E[Y(d, M(1))|X, C]$ and $E[Y(d, M(0))|X, C]$ share the same parameters, so the objective function is a sum of two sub-problems. Given $\alpha_m$, we would be able to identify $\beta_d$ based on \((3.1)\), if the potential outputs $Y(d, M(d'))$’s, the potential mediator $M(d')$’s underlying the $Y(d, M(d'))$ and the $m_{d'}(X, \alpha_m)$’s of \((2.14)\) and \((2.15)\), and the potential treatment status $D(z_1)$’s underlying the definition of $C$ were observed. However, we can only observe $W$, rather than these latent variables, in practice. Thus, \((3.1)\) is indeed not enough for the identification of $\beta_d$.

To circumvent this problem, we need a structure model for $(Y, M, D)$, which is applicable to recovering the potential-outcome models for $Y(d, M(d'))$, $M(d')$ and $D(z_1)$, for $d, d', z_1 \in \{0, 1\}$, from the observable $W$ and a set of conditions for $(Z_1, Z_2)$ that suitably accounts for the endogeneity of $D$ and $M$. Recently, Frölich and Huber (2014) proposed such an identification method in the nonparametric context. In this study, we apply their
method to our parametric context for the identification of $\beta_d$. To facilitate this application, we let $1(\cdot)$ be the indicator function which is equal to one when the associated event is true (otherwise, zero), and make the following assumption:

**Assumption 3.2 (Parametric Structure Model)** Assume that:

(a) A parametric structure model for $(Y, M, D)$:

\[
Y = \tilde{h}(D, M, X) - \tilde{g}(D, M, X) + U_Y, \\
M = 1(\alpha_d D + \alpha_{z2} Z_2 + X^\top \alpha_x + U_M \geq 0), \\
D = 1(\gamma_z Z_1 + X^\top \gamma_x + U_D \geq 0),
\]

where $\tilde{h}(D, M, X)$ and $\tilde{g}(D, M, X)$ follow (2.10) and (2.11); $U_Y$, $U_M$ and $U_D$ are error terms; $(\alpha_d, \alpha_{z2}, \alpha_x^\top)$ and $(\gamma_z, \gamma_x^\top)$ are parameter vectors.

(b) $U_Y \equiv v(D, M) + \tilde{u}(D, M)$ and $E[U_Y|X, C] = 0.$

**Assumption 3.3 (Identification of Equation (3.1))** Assume that

(a) (IV independence)

\[
(Z_1, Z_2) \perp (U_Y, U_M)|U_D, X, \\
Z_1 \perp (U_Y, U_M, U_D)|Z_2, X,
\]

where $\perp$ denotes statistical independence.

(b) (Conditional independence of $Z_1$ and $Z_2$) $Z_1 \perp Z_2|X.$

(c) (Weak monotonicity of treatment choice) $\gamma_z > 0.$

(d) (Monotonicity of mediator in the IV and the observable)

(d1) $\alpha_{z2} > 0,$

(d2) $U_M \perp (U_D, Z_2, X)$ and $U_M$ is continuously distributed with the cumulative distribution function (CDF) $F_{U_M}(\cdot)$ that is strictly increasing in the support of $U_M.$

17
(e) (Common support of M)

\[ 0 < P(Z_1 = 1 | M, U_M, X, C) < 1 \quad \text{almost surely.} \]

**Assumption 3.4 (Parametric models for } Z_1, Z_2 \text{ and } D\)** Assume that

(a) (Model for } Z_1\) \[ Z_1 = 1(X^\top \alpha Z_1 + U Z_1 \geq 0) \text{ with } U Z_1 \perp X, \text{ and the CDF of } U Z_1 \text{ is } F_{U Z_1}(\cdot). \]

(b) (Model for } Z_2\) \( Z_2 \) is independent of \((U_D, X)\) and has a continuous CDF \( F_{Z_2}(\cdot, \alpha_f) \) with the parameter vector \( \alpha_f \).

(c) (Model for } D\) \( U_D \) is independent of \((Z_1, X)\) with the CDF \( F_{U_D}(\cdot) \).

Assumption 3.2 constitutes a parametric structure model for \((Y, M, D)\) which comprises the potential-outcome models for \( Y(d, M(d')), M(d') \) and \( D(z_1) \). Assumption 3.2 and Assumption 3.3 are sufficient for the set of nonparametric identification conditions of the Theorem 4 of Frölich and Huber (2014) that we will use to identify our Equation (3.1). Assumptions 3.3 (a) and (b) are conditions needed so \( Z_1 \) and \( Z_2 \) are valid IV’s. The sign restriction, \( \gamma_z > 0 \), in Assumption 3.3(c) combined with the model \( D = 1(\gamma_z Z_1 + X^\top \gamma_x + U_D \geq 0) \) has two implications. First, \( \gamma_z > 0 \) implies that

\[ D(1) = 1(\gamma_z + X^\top \gamma_x + U_D \geq 0) \geq 1(X^\top \gamma_x + U_D \geq 0) = D(0) \]

so there are no defiers in the population. Second, it implies that

\[ \Delta \equiv E(D(1)) - E(D(0)) > 0 \quad (3.2) \]

so that the population of compliers are of positive measure. These conditions are standard in the treatment effect model with endogenous treatment status. Assumptions 3.3 (e) is a common support condition which is also standard in the treatment effect literature. If this
does not hold, we can redefine the population of interest so that this condition would hold on this restricted subpopulation. Assumptions 3.3 (e) also implies that $0 < P(Z_1 = 1 | X) < 1$ almost surely which is also a common support condition in the literature. Note that $0 < P(Z_1 = 1 | X) < 1$ almost surely is needed for use to identify

$$E \left[ \left( Y(d, M(d')) - h_{d'}(X, \alpha_m, b_{d1}, b_{d0}) + g_{d'}(X, \alpha_m, b_{d1}, b_{d0}) \right)^2 \right]$$

for $(d, d') = (1, 1)$ and $(d, d') = (0, 0)$. If we are only interested in $(d, d') = (1, 1)$ case and $(d, d') = (0, 0)$ case as in Abadie (2003), then these conditions are sufficient. In other words, Assumptions 3.3 (d) is introduced for mediation analysis. The sign condition, $\alpha_{z2} > 0$, implies the monotonicity of $M$ in $D$ and in $Z_2$ which are needed in the proof. Therefore, we are able to extend Abadie (2003) to allow for mediation analysis in a treatment effect model.

Assumption 3.4 is introduced for two reasons. First, it implies the parametric models of $Z_1$, $Z_2$ and $D$, and this in turn implies that the weight functions $\omega(d, d', \alpha)$ introduced below have parametric forms. Second, parametric models of the variables allow us to simplify the asymptotic theory which would have been much more complicated if we did not impose these conditions. We will add more discussions on this assumption after we introduce the identification results in Theorem 3.2.

For further discussions, we define $Q_M(s) \equiv 1 - F_{U_M}(-s)$, $Q_D(s) \equiv 1 - F_{U_D}(-s)$, $Q_{Z_1}(s) \equiv 1 - F_{U_{Z_1}}(-s)$ and $f_{Z_2}(s, \alpha_f) \equiv \partial F_{Z_2}(s, \alpha_f) / \partial s$, for $s \in R$.

From Assumption 3.2, we can obtain a submodel for $Y | D, M, C, X$ which reduces to the potential-output model for $Y(d, M(d'))$ in (2.9) when $(D, M) = (d, M(d'))$. Assumption 3.2 also implies the potential mediator:

$$M(d') = 1(\alpha_d d' + \alpha_{z2} Z_2 + X^\top \alpha_x + U_M \geq 0). \quad (3.3)$$

and the potential treatment status:

$$D(z_1) = 1(\gamma_z z_1 + X^\top \gamma_x + U_D \geq 0). \quad (3.4)$$

19
From Assumption 3.2 and Assumption 3.3 (d2), we obtain a single-index model for $M|D, Z_2, X$:

$$E[M|D, Z_2, X] = P(U_M \geq -(\alpha_d D + \alpha_{z_2} Z_2 + X^\top \alpha_x)|D, Z_2, X)$$

$$= Q_M(\alpha_d D + \alpha_{z_2} Z_2 + X^\top \alpha_x).$$

(3.5)

In the Appendix, we further prove the result:

**Theorem 3.1** Suppose that Assumption 3.2 (a), Assumption 3.3 (d2) and Assumption 3.4 (b) hold. For $d' \in \{0, 1\}$,

$$E[M(d')|X, C] = m_{d'}(X, \alpha_m) = \int_R Q_M(\alpha_d d' + \alpha_{z_2} z_2 + X^\top \alpha_x)f_{Z_2}(z_2, \alpha_f)dz_2,$$

(3.6)

with the parameter vector $\alpha_m \equiv (\alpha_d, \alpha_{z_2}, \alpha_x^\top, \alpha_f^\top)^\top$.

This shows that, under suitable conditions, the potential-mediator model for $E[M(d')|X, C]$ in (2.12) is a linear mixture of the $E[M|D, Z_2 = z_2, X]$’s in (3.5) with different $z_2$’s, weighted by the associated $f_{Z_2}(z_2, \alpha_f)$’s. Correspondingly, $\alpha_m$ is composed of $(\alpha_d, \alpha_{z_2}, \alpha_x^\top)$ and $\alpha_f$ that are, respectively, the parameter vectors of the model in (3.5) and $f_{Z_2}(\cdot, \alpha_f)$, and are, respectively, identifiable from the distribution of $(M, D, Z_2, X^\top)^\top$ and that of $(Z_2, X^\top)^\top$. Importantly, this illustrates that $\alpha_m$ is also identifiable, and shows the validity of Assumption 3.1 (a).

Assumption 3.4 (a) implies a parametric propensity-score model:

$$E[Z_1|X] = Q_{Z_1}(X^\top \alpha_{z_1}),$$

(3.7)

which allows us to present the identification result and implement the statistical inference in a simpler way, and to identify $\alpha_{z_1}$ from the distribution of $(Z_1, X^\top)^\top$.

Denote $\alpha \equiv (\alpha_m^\top, \alpha_{z_1}^\top)^\top$ and recall that $\Delta$ is defined in (3.2). In the Appendix, we prove the following result by matching Assumptions 3.2 and 3.3 with the assumptions of Theorem 4 of Frölich and Huber (2014):
Theorem 3.2 Suppose that Assumptions 3.2, 3.3, 3.4(a) and 3.4(b) hold. For \( d, d' \in \{0, 1\} \),

\[
E \left[ \left( Y(d, M(d')) - h_{d'}(X, \alpha_m, b_{d1}^h, b_{d0}^h) + g_{d'}(X, \alpha_m, b_{d1}^g, b_{d0}^g) \right)^2 \right] \bigg| \mathcal{C} \\
= E \left[ w(d, d', \alpha) \left( Y - h_{d'}(X, \alpha_m, b_{d1}^h, b_{d0}^h) + g_{d'}(X, \alpha_m, b_{d1}^g, b_{d0}^g) \right)^2 \right] / \Delta,
\]

(3.8)

where

\[
w(d, d', \alpha) \equiv \begin{cases} \\
D\lambda(Z_1, X, \alpha_{z1}), & (d, d') = (1, 1), \\
D\varphi_1(Z_2, X, \alpha)\lambda(Z_1, X, \alpha_{z1}), & (d, d') = (1, 0), \\
(D - 1)\varphi_2(Z_2, X, \alpha)\lambda(Z_1, X, \alpha_{z1}), & (d, d') = (0, 1), \\
(D - 1)\lambda(Z_1, X, \alpha_{z1}), & (d, d') = (0, 0), \\
\end{cases}
\]

(3.9)

with

\[
\lambda(Z_1, X, \alpha_{z1}) \equiv (Z_1 - Q_{Z_1}(X^\top \alpha_{z1}))/\left( Q_{Z_1}(X^\top \alpha_{z1})(1 - Q_{Z_1}(X^\top \alpha_{z1})) \right),
\]

\[
\varphi_1(Z_2, X, \alpha) \equiv f_{Z_2}(Z_2 + \alpha_d/\alpha_{z2}, \alpha_f) / f_{Z_2}(Z_2, \alpha_f),
\]

and

\[
\varphi_2(Z_2, X, \alpha) \equiv f_{Z_2}(Z_2 - \alpha_d/\alpha_{z2}, \alpha_f) / f_{Z_2}(Z_2, \alpha_f).
\]

This allows us to represent the minimizer of the mean squared error of potential outputs in (3.1) as the minimizer of the weighted mean squared error of actual outputs:

\[
\beta_d \equiv \arg \min_{b_d \in B} \sum_{d' = 0, 1} E \left[ w(d, d', \alpha) \left( Y - h_{d'}(X, \alpha_m, b_{d1}^h, b_{d0}^h) + g_{d'}(X, \alpha_m, b_{d1}^g, b_{d0}^g) \right)^2 \right],
\]

(3.10)

for \( d \in \{0, 1\} \), where the weights: \( w(1, 1, \alpha) \), \( w(1, 0, \alpha) \), \( w(0, 1, \alpha) \) and \( w(0, 0, \alpha) \) are functions of \((D, Z_1, Z_2, X^\top)\). Importantly, unlike (3.1) which is a conditional expectation of the latent \( Y(d, M(d')) \) on the subpopulation of \( \mathcal{C} \), (3.10) is an unconditional expectation of the observable \( Y \) which is determined by the distribution of \( W \). Therefore, given the
identification of $\alpha$, we can identify the parameter vector $\beta_d$ according to (3.10) based on Assumptions 3.1 and 3.2.

The above results demonstrate that the parameter vectors: $\alpha$ and $\beta_d$, for $d \in \{0, 1\}$, are identifiable from the distribution of $W$. Given (3.4), Assumption 3.4(c) implies that

$$E[D|Z_1, X] = Q_D(\gamma_z Z_1 + X^\top \gamma_x).$$

(3.11)

Thus, the parameter vector $(\gamma_z, \gamma_x^\top)^\top$ is also identifiable from the distribution of $(D, Z_1, X^\top)^\top$.

In principle, we may replace the role of $w(d, d', \alpha)$ in the identification of $\beta_d$ with the nonparametric weight of Frölich and Huber (2014); see the proof of Theorem 3.2. This replacement allows us to be free of Assumption 3.4 but requires us to estimate $E[Z_1|X]$ by a nonparametric estimator, such as the series Logit estimator as in Hirano, Imbens and Ridders (2003) and Donald, Hsu and Lieli (2014a), the local polynomial regression estimator as in Ichimura and Linton (2005) and Donald, Hsu and Lieli (2014b) or the kernel estimator as in Abrevaya, Hsu and Lieli (2014). However, the asymptotics of the resulting estimator for $\beta_d$ and the associated assumptions are much more complicated. For ease of the asymptotic analysis and to be consistent with the parametric context conducted in this study, we impose a parametric structure on the propensity score. Similar argument applies to the model of $Z_2$.

### 3.2 Identification of CLATE and LATE

Note the components of the CLATE, CDLATE and CILATE decompositions in (2.16), (2.17) and (2.18) are determined by $m_d(x, \alpha_m)$, $h(x, \beta_{dj}^h)$ and $g(x, \beta_{dj}^g)$ for $d, d', j \in \{0, 1\}$, and the parameter restrictions in the program-evaluation hypotheses: $H_o^p$, $H_o^{ph}$ and $H_o^{pg}$ and the mediation-analysis hypotheses: $H_o^d$, $H_o^{dh}$, $H_o^{dg}$, $H_o^i$, $H_o^{ih}$ and $H_o^{ig}$ are determined by $\beta_0$, $\beta_1$ and $\alpha_d$. Thus, these conditional effects are also identified when the parameter vectors: $\alpha_m$, $\beta_0$ and $\beta_1$ are identified.

To discuss the identification of the unconditional effects, we let $f_{X|C}(\cdot)$ be the conditional
density function of \(X|C\), and \(f_X(\cdot)\) be the unconditional density function of \(X\). Note that \(\text{LATE} \) and \(\text{CLATE}(x)\) have the relationship:

\[
\text{LATE} = \int \text{CLATE}(x)f_X|C(x)dx = \int \left(\text{CLATE}(x) \frac{P(C|X=x)}{P(C)}\right)f_X(x)dx
\]

\[
= E \left[ \frac{\Delta(X)}{\Delta} \times \text{CLATE}(X) \right],
\]

where

\[
\Delta(X) \equiv P(C|X) = P(D(1) = 1, D(0) = 0|X).
\]

This relationship also holds for the components of the LATE decompositions in (2.19), (2.20) and (2.21).

By the law of total probability, we have the result:

\[
P(D(1) = 1|X) = P(D(1) = 1, D(0) = 1|X) + P(D(1) = 1, D(0) = 0|X),
\]

\[
P(D(0) = 1|X) = P(D(1) = 1, D(0) = 1|X) + P(D(1) = 0, D(0) = 1|X).
\]

Note that the sign restriction: \(\gamma_z > 0\) in Assumption 3.3(c) implies no defiers. Thus, given (3.13), we can use the restriction: \(P(D(1) = 0, D(0) = 1|X) = 0\) to simplify (3.14) as:

\[
\Delta(X) = P(D(1) = 1|X) - P(D(0) = 1|X) = E[D(1)|X] - E[D(0)|X].
\]

In addition, (3.11) implies a parametric model for the potential treatment status:

\[
E[D(z_1)|X] = Q_D(\gamma_z z_1 + X^\top \gamma_x)
\]

and hence the following model for \(\Delta(X)\) and \(\Delta\):

\[
\Delta^*(X, \gamma) \equiv Q_D(\gamma_z + X^\top \gamma_x) - Q_D(X^\top \gamma_x)
\]

and

\[
\Delta^*(\gamma) \equiv E[Q_D(\gamma_z + X^\top \gamma_x)] - E[Q_D(X^\top \gamma_x)],
\]

23
where \( \gamma \equiv (\gamma_z, \gamma_x^\top) \).

Accordingly, we can easily transform the components of the CLATE decompositions to their LATE counterparts in (2.19), (2.20) and (2.21):

\[
\text{LATE}_h = H_{11}(\nu^h_1) - H_{00}(\nu^h_0), \quad \text{LATE}_g = G_{11}(\nu^g_1) - G_{00}(\nu^g_0),
\]

\[
\text{DLATE}_h = H_{11}(\nu^h_1) - H_{01}(\nu^h_0), \quad \text{DLATE}_g = G_{11}(\nu^g_1) - G_{01}(\nu^g_0),
\]

\[
\text{ILATE}_h = H_{01}(\nu^h_0) - H_{00}(\nu^h_0), \quad \text{ILATE}_g = G_{01}(\nu^g_0) - G_{00}(\nu^g_0),
\]

where \( \nu^h_d \equiv (\alpha_m^\top, \beta_{d1}^h, \beta_{d0}^h, \gamma^\top) \), \( \nu^g_d \equiv (\alpha_m^\top, \beta_{d1}^g, \beta_{d0}^g, \gamma^\top) \),

\[
H_{dd'}(\nu^h_d) \equiv E\left[ \frac{\Delta^*(X, \gamma)}{\Delta^*(\gamma)} \times h_{dd'}(X, \alpha_m, \beta_{d1}^h, \beta_{d0}^h) \right]
\]

(3.19)

and

\[
G_{dd'}(\nu^g_d) \equiv E\left[ \frac{\Delta^*(X, \gamma)}{\Delta^*(\gamma)} \times g_{dd'}(X, \alpha_m, \beta_{d1}^g, \beta_{d0}^g) \right],
\]

(3.20)

for \( d, d' \in \{0, 1\} \). Thus, given the identification \((\alpha_m^\top, \beta_d^\top, \gamma^\top)\), these unconditional effects are also identified.

### 4 Estimation and Asymptotics

Denote \( W_i \equiv (Y_i, M_i, D_i, Z_{1i}, Z_{2i}, X_i^\top) \), and let \( \{W_i\}_{i=1}^n \) be a random sample of \( W \) with the sample size \( n \). The identification results in Section 3.1 suggest that we may estimate the parameter vectors: \( \alpha, \beta_d \) and \( \gamma \) in a two-step way. In the first step, we estimate the parameter vectors: \( \alpha \) and \( \gamma \). In the second step, given the estimator of \( \alpha \), we estimate \( \beta_d \) by the WNLSE based on (3.10) for \( d \in \{0, 1\} \). Accordingly, we can further estimate the CLATE and LATE components. In this section, we first discuss this two-stage estimation method, and then demonstrate the asymptotic properties of the estimators and provide a general test for statistical inference.
4.1 Estimation

First-Step Estimation on $\alpha$

Recall that $\alpha \equiv (\alpha_m^\top, \alpha_{z1}^\top)^\top$, where $\alpha_m \equiv (\alpha_d, \alpha_{z2}, \alpha_x^\top, \alpha_f^\top)^\top$ is the parameter vector of the potential-mediator model in (3.6), and $\alpha_{z1}$ is the parameter vector of the propensity-score model in (3.7). The vector $\alpha_m$ comprises the parameter vector $\alpha_q \equiv (\alpha_d, \alpha_{z2}, \alpha_x^\top)^\top$ of the mediator model in (3.5) and the parameter vector $\alpha_f$ of the conditional density function $f_{Z_2}(\cdot, \alpha_f)$. In addition, recall that $\gamma \equiv (\gamma_z, \gamma_x^\top)^\top$ is the parameter vector of the treatment model in (3.11). Given the fact that (3.5), (3.7) and (3.11) are binary response models and $f_{Z_2}(\cdot, \alpha_f)$ is also a parametric model, we can easily estimate $\alpha_q$, $\alpha_f$, $\alpha_{z1}$ and $\gamma$ by the maximum likelihood (ML) method.

Specifically, let $A_q$, $A_f$, $A_{z1}$ and $R$ be, respectively, the parameter spaces of $\alpha_q$, $\alpha_f$, $\alpha_{z1}$ and $\gamma$, and $a_q \equiv (a_d, a_{z2}, a_x^\top)^\top$, $a_f$, $a_{z1}$ and $r$ be, respectively, arbitrary vectors in $A_q$, $A_f$, $A_{z1}$ and $R$. The ML estimators (MLEs) for $\alpha_q$, $\alpha_f$, $\alpha_{z1}$ and $\gamma$ are defined as:

$$\hat{\alpha}_q \equiv (\hat{\alpha}_d, \hat{\alpha}_{z2}, \hat{\alpha}_x^\top)^\top = \text{arg max}_{a_q \in A_q} \frac{1}{n} \sum_{i=1}^{n} \ell_{q,i}(a_q),$$

$$\hat{\alpha}_f = \text{arg max}_{a_f \in A_f} \frac{1}{n} \sum_{i=1}^{n} \ell_{f,i}(a_f),$$

$$\hat{\alpha}_{z1} = \text{arg max}_{a_{z1} \in A_{z1}} \frac{1}{n} \sum_{i=1}^{n} \ell_{z1,i}(a_{z1}),$$

and

$$\hat{\gamma} \equiv (\hat{\gamma}_z, \hat{\gamma}_x^\top)^\top = \text{arg max}_{r \in R} \frac{1}{n} \sum_{i=1}^{n} \ell_{\gamma,i}(r),$$

based on the log-probability or log-density functions:

$$\ell_{q,i}(a_q) \equiv \log \left( Q_M(a_d D_i + a_{z2} Z_{2i} + X_i^\top a_x)^M_i (1 - Q_M(a_d D_i + a_{z2} Z_{2i} + X_i^\top a_x))^{1-M_i} \right),$$

$$\ell_{f,i}(a_f) \equiv \log f_{Z_2}(Z_{2i}, a_f),$$

25
\[ \ell_{z_1,i}(a_{z_1}) \equiv \log(Q_{Z_1}(X_i^\top a_{z_1})^{Z_{i1}}(1 - Q_{Z_1}(X_i^\top a_{z_1}))^{1 - Z_{i1}}) \]

and

\[ \ell_{\gamma,i}(r) \equiv \log(Q_{D}(X_i^\top \gamma)^{Z_{i1}}(1 - Q_{D}(X_i^\top \gamma))^{1 - Z_{i1}}) \].

Alternatively, one may also estimate \( \alpha_q, \alpha_f, \alpha_z \) and \( \gamma \) by other suitable moment-based methods. Given \( \hat{\alpha}_q, \hat{\alpha}_f \) and \( \hat{\alpha}_z \), we can estimate \( \alpha \) by

\[ \hat{\alpha} = (\hat{\alpha}_{q1}^\top, \hat{\alpha}_{f1}^\top, \hat{\alpha}_{z1}^\top)^\top. \]

Second-Step Estimation on \( \beta \)

Let \( w_i(d,d',\alpha) \) be the \( w(d,d',\alpha) \) in (3.9) evaluated at \( W = W_i \). Given the estimator \( \hat{\alpha} \), we define the second-step WNLS as:

\[ \hat{\beta}_d = \left( \hat{\beta}_h^{d1}, \hat{\beta}_h^{0}, \hat{\beta}_g^{d1}, \hat{\beta}_g^{0} \right)^\top = \min_{b_d \in B_d} \frac{1}{n} \sum_{i=1}^{n} \rho_{d,i}(\hat{\alpha}, b_d), \]

where

\[ \rho_{d,i}(\alpha, b_d) \equiv \sum_{d'=0,1} w_i(d,d',\alpha) \left( Y_i - h_{d'}(X_i, \alpha_m, b_{d1}^h, b_{d0}^h) + g_{d'}(X_i, \alpha_m, b_{d1}^g, b_{d0}^g) \right)^2, \]

for \( d \in \{0,1\} \); recall that \( b_d \equiv (b_{d1}^h, b_{d0}^h, b_{d1}^g, b_{d0}^g)^\top \) is an arbitrary vector in \( B_d \). Note that (4.1) amounts to the sample analogue of (3.10) conditional on the first-step estimator \( \hat{\alpha} \).

Note that

\[ \frac{1}{n} \sum_{i=1}^{n} \nabla_{b_d} \rho_{d,i}(\hat{\alpha}, \hat{\beta}_d) = 0, \]

we may also interpret \( \hat{\beta}_d \) as the method-of-moment estimator for \( \beta_d \):

\[ \hat{\beta}_d = \arg \min_{b_d \in B_d} \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla_{b_d} \rho_{d,i}(\hat{\alpha}, b_d) \right] \;
\left[ \frac{1}{n} \sum_{i=1}^{n} \nabla_{b_d} \rho_{d,i}(\hat{\alpha}, b_d) \right] \] .

We actually solve \( \hat{\beta}_d \) by (4.4). The main reason is that the problem in (4.1) is not strictly concave in \( \beta \) because the weight \( \omega(d,d',\hat{\alpha}) \) can be negative.
Estimation on CLATE and LATE

Given \( \hat{\alpha}, \hat{\beta}_d \) and \( \hat{\gamma} \), we can immediately estimate the components of the CLATE, CD-LATE and CILATE decompositions in (2.16), (2.17) and (2.18) by replacing the roles of \( m'(x, \alpha_m), h(x, \beta_{d_j}^h) \) and \( g(x, \beta_{d_j}^g) \) in these components with their estimators: \( m'(x, \hat{\alpha}_m), h(x, \hat{\beta}_{d_j}^h) \) and \( g(x, \hat{\beta}_{d_j}^g) \), for \( d, d', j = \{0, 1\} \). Similarly, we can base the tests for \( H_{po}, H_{ph}, H_{pg}, H_{do}, H_{dh}, H_{dg}, H_{io}, H_{ih} \) and \( H_{ig} \) on the estimators: \( \hat{\beta}_{d_j}^h, \hat{\beta}_{d_j}^g \) and \( \hat{\alpha}_d \), for \( d, j \in \{0, 1\} \).

In addition, we can also estimate the components of the LATE, DLATE and ILATE decompositions in (2.19), (2.20) and (2.21) by replacing the roles of \( H^h_{dd'}, H^g_{dd'} \) in these components with their estimators:

\[
\hat{H}_{dd'}^h(\hat{\nu}_{d'}^h) \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta^*(X_i, \hat{\gamma})}{\Delta^*(\hat{\gamma})} \times h'_{d'}(X_i, \hat{\alpha}_m, \hat{\beta}_{d1}^h, \hat{\beta}_{d0}^h) \tag{4.5}
\]

and

\[
\hat{G}_{dd'}^g(\hat{\nu}_{d'}^g) \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta^*(X_i, \hat{\gamma})}{\Delta^*(\hat{\gamma})} \times g'_{d'}(X_i, \hat{\alpha}_m, \hat{\beta}_{d1}^g, \hat{\beta}_{d0}^g), \tag{4.6}
\]

where \( \hat{\nu}_{d'}^h \equiv (\hat{\alpha}_m^\top, \hat{\beta}_{d1}^h, \hat{\beta}_{d0}^h, \hat{\gamma}^\top) \) and \( \hat{\nu}_{d'}^g \equiv (\hat{\alpha}_m^\top, \hat{\beta}_{d1}^g, \hat{\beta}_{d0}^g, \hat{\gamma}^\top) \), for \( d, d' \in \{0, 1\} \). Meanwhile, we can base the tests for \( H_{po}, H_{ph}, H_{pg}, H_{do}, H_{dh}, H_{dg}, H_{io}, H_{ih} \) and \( H_{ig} \) on the sample moments: \( \hat{H}_{dd'}(\hat{\nu}_{d'}^h) \) and \( \hat{G}_{dd'}(\hat{\nu}_{d'}^g) \), for \( d, d' \in \{0, 1\} \).

4.2 Asymptotics

Because the first-step estimators: \( \hat{\alpha}_q, \hat{\alpha}_f, \hat{\alpha}_{z1} \) and \( \hat{\gamma} \) are, respectively, the MLEs for \( \alpha_q, \alpha_f, \alpha_{z1} \) and \( \gamma \), it is standard to argue the consistency and the asymptotic normality of these estimators by the ML theory under Assumptions 3.2 and 3.4 and suitable regularity conditions. Since the ML theory is well-known, we do not repeat the regularity conditions for the consistency and asymptotic normality of the MLE; see, e.g., White (1994) and Newey and McFadden (1994) for the ML theory and conditions. Given the consistency of \( \hat{\alpha}_q, \hat{\alpha}_f, \hat{\alpha}_{z1} \) and \( \hat{\gamma} \), by taking the mean-value expansions of the estimation equations of...
these estimators under suitable conditions and using the information matrix equality, we can show that $\hat{\alpha}_q$, $\hat{\alpha}_f$, $\hat{\alpha}_{z_1}$ and $\hat{\gamma}$ have the influence-function representations:

$$\sqrt{n}(\hat{\alpha}_q - \alpha_q) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{q,i}(\alpha_q) + o_p(1), \quad (4.7)$$

$$\sqrt{n}(\hat{\alpha}_f - \alpha_f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{f,i}(\alpha_f) + o_p(1), \quad (4.8)$$

$$\sqrt{n}(\hat{\alpha}_{z_1} - \alpha_{z_1}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{z_1,i}(\alpha_{z_1}) + o_p(1), \quad (4.9)$$

and

$$\sqrt{n}(\hat{\gamma} - \gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\gamma,i}(\gamma) + o_p(1), \quad (4.10)$$

where $\psi_{q,i}(\alpha_q)$, $\psi_{f,i}(\alpha_f)$, $\psi_{z_1,i}(\alpha_{z_1})$ and $\psi_{\gamma,i}(\gamma)$ are instantaneous transformations of $W_i$ such that

$$\psi_{q,i}(\alpha_q) \equiv E[\nabla_{a_q} \ell_{q,i}(\alpha_q)\nabla_{a_q}^\top \ell_{q,i}(\alpha_q)]^{-1}\nabla_{a_q} \ell_{q,i}(\alpha_q),$$

$$\psi_{f,i}(\alpha_f) \equiv E[\nabla_{a_f} \ell_{f,i}(\alpha_f)\nabla_{a_f}^\top \ell_{f,i}(\alpha_f)]^{-1}\nabla_{a_f} \ell_{f,i}(\alpha_f),$$

$$\psi_{z_1,i}(\alpha_{z_1}) \equiv E[\nabla_{a_{z_1}} \ell_{z_1,i}(\alpha_{z_1})\nabla_{a_{z_1}}^\top \ell_{z_1,i}(\alpha_{z_1})]^{-1}\nabla_{a_{z_1}} \ell_{z_1,i}(\alpha_{z_1})$$

and

$$\psi_{\gamma,i}(\gamma) \equiv E[\nabla_{r} \ell_{\gamma,i}(\gamma)\nabla_{r}^\top \ell_{\gamma,i}(\gamma)]^{-1}\nabla_{r} \ell_{\gamma,i}(\gamma).$$

From (4.7), (4.8) and (4.9), we can also write that

$$\sqrt{n}(\hat{\alpha} - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\alpha,i}(\alpha) + o_p(1), \quad (4.11)$$

where

$$\psi_{\alpha,i}(\alpha) \equiv (\psi_{q,i}(\alpha_q)^\top, \psi_{f,i}(\alpha_f)^\top, \psi_{z_1,i}(\alpha_{z_1})^\top)^\top.$$
Similarly, because the estimator $\hat{\beta}_d$ can be interpreted as a second-step moment estimator for $\beta$, it is also standard to argue its consistency and asymptotic normality under suitable regularity conditions by the two-stage estimation theory. This estimation theory is also quite standard, so the regularity conditions for the consistency and asymptotic normality of the two-stage estimator are not reported here for sake of brevity; see, e.g., Newey and McFadden (1994) and Wooldridge (2001) for the theory and conditions. Given the consistency of $\hat{\beta}_d$, by taking the mean-value expansions of the estimation equation in (4.3) under suitable conditions and using (4.11), we can show that $\hat{\beta}_d$ has the influence-function representation:

$$\sqrt{n}(\hat{\beta}_d - \beta_d) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\beta_d,i}(\alpha) + o_p(1),$$  \hspace{1cm} (4.12)

where

$$\psi_{\beta_d,i}(\beta_d) \equiv -E[\nabla_{b_d} \rho_{d,i}(\alpha, \beta_d)]^{-1} (\nabla_{b_d} \rho_{d,i}(\alpha, \beta_d) + E[\nabla_a (\nabla_{b_d} \rho_{d,i}(\alpha, \beta_d))] \psi_{\alpha,i}(\alpha)),$$

with $a \equiv (a_{q_d}^T, a_f^T, a_{z_1}^T)^T$, for $d \in \{0, 1\}$.

Denote $\beta \equiv (\beta_0^T, \beta_1^T)^T$, $\hat{\beta} \equiv (\hat{\beta}_0^T, \hat{\beta}_1^T)^T$, $\psi_{\beta,i}(\beta) \equiv (\psi_{\beta_0,i}(\beta_0)^T, \psi_{\beta_1,i}(\beta_1)^T)^T$, $\theta \equiv (\alpha^T, \beta^T, \gamma^T)^T$ and $\hat{\theta} \equiv (\hat{\alpha}^T, \hat{\beta}^T, \hat{\gamma}^T)^T$. From (4.10), (4.11) and (4.12), we can see that

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\theta) + o_p(1), \text{ with } \psi_i(\theta) \equiv \begin{bmatrix} \psi_{\alpha,i}(\alpha) \\ \psi_{\beta,i}(\beta) \\ \psi_{\gamma,i}(\gamma) \end{bmatrix}.$$ \hspace{1cm} (4.13)

Because $\psi_i(\theta)$ is an instantaneous transformation of $W_i$, $\{\psi_i(\theta)\}_{i=1}^n$ is an independently and identically distributed (IID) sequence if $\{W_i\}_{i=1}^n$ is an IID sequence. Given this IIDness and the condition that the elements of $\psi_i(\theta)$ have finite second moments, we may first use a central limit theorem and the Cramér-Wold device to show the asymptotic normality of $n^{-1/2} \sum_{i=1}^n \psi_i(\alpha, \beta, \gamma)$ and then use (4.13) to further show the result:

$$\sqrt{n}(\hat{\theta} - \theta) \overset{d}{\to} N(0, \Sigma),$$ \hspace{1cm} (4.14)
with the asymptotic covariance matrix \( \Sigma \equiv E \left[ \psi_i(\theta) \psi_i(\theta)^T \right] \).

The result in (4.14) is applicable to establishing a generalized test for the CLATE hypotheses: \( H_{po}^i, H_{ph}^i, H_{pg}^i, H_{d}^i, H_{dh}^i, H_{o}^i, H_{th}^o \) and \( H_{og}^i \) discussed in Section 2.2. For ease of exposition, suppose that \( \alpha_d \) is known or assumed to be non-zero, so \( H_{po}^i, H_{th}^o \) and \( H_{og}^i \) only include the restrictions on \( \beta_{01}^h, \beta_{00}^h, \beta_{01}^g \) and \( \beta_{00}^g \). Let \( \Sigma_{\beta} \) be the asymptotic covariance matrix of \( n^{1/2} (\hat{\beta} - \beta) \) implied by (4.14), and \( \delta \) be a finite-dimensional linear transformation of \( \beta \) such that \( \delta = S \delta \beta \) for some \( S \delta \) which is a deterministic matrix of full row rank and dependent on the choice of \( \delta \). These hypotheses are particular examples of the hypothesis: \( \delta = 0 \) with different \( \delta \)'s. Denote \( \hat{\delta} = S \delta \hat{\beta} \). According to (4.14), we have the result:

\[
\sqrt{n}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_{\delta}),
\]

(4.15)

with \( \Sigma_{\delta} = S \delta \Sigma_{\beta} S_{\delta}^T \). Let \( \hat{\Sigma}_{\delta} \) be a positive-definite matrix which is consistent for \( \Sigma_{\delta} \). Given (4.15), we can define the Wald test statistic:

\[
W = n \hat{\delta}^T \hat{\Sigma}_{\delta}^{-1} \hat{\delta}
\]

(4.16)

that has the asymptotic null distribution: \( W \xrightarrow{d} \chi^2(k) \), with \( k \equiv \text{dim}(\delta) \). This generalized test is applicable to the CLATE hypotheses by matching \( \delta \) with the \( \beta \) restrictions of the hypothesis being tested.

By a slight modification, this test is also applicable to examining the LATE hypotheses: \( \tilde{H}_{po}^i, \tilde{H}_{th}^o, \tilde{H}_{og}^i, \tilde{H}_{d}^i, \tilde{H}_{dh}^i, \tilde{H}_{o}^i, \tilde{H}_{th}^o \) and \( \tilde{H}_{og}^i \) discussed in Section 2.3. In this scenario, we need to replace \( \delta \) and \( \hat{\delta} \) by \( \delta = E[\xi_i(\theta)] \) and \( \hat{\delta} = n^{-1} \sum_{i=1}^n \xi_i(\hat{\theta}) \), where \( \xi_i(\cdot) \) is dependent on the hypothesis being tested. For example, \( \delta = H_{11}(\nu_{11}^h) - H_{01}(\nu_{01}^h), \hat{\delta} = \hat{H}_{11}(\hat{\nu}_{11}^h) - \hat{H}_{01}(\hat{\nu}_{01}^h) \) and

\[
\xi_i(\theta) = \left( \frac{QD(\gamma_z + X_i^T \gamma_x) - QD(X_i^T \gamma_x)}{E[QD(\gamma_z + X_i^T \gamma_x)] - E[QD(X_i^T \gamma_x)]} \right) \left( h_1(X_i, \alpha_m, \beta_{11}^h, \beta_{10}^h) - h_1(X_i, \alpha_m, \beta_{01}^h, \beta_{00}^h) \right)
\]

when the hypothesis being tested is \( \tilde{H}_{o}^{dh} \). Under suitable conditions, we may also use the mean-value expansion to show the influence-function representation of \( \hat{\delta} \):

\[
\sqrt{n}(\hat{\delta} - \delta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\delta,i}(\theta) + o_p(1),
\]

(4.17)
where
\[ \psi_{\delta,i}(\theta) \equiv \xi_i(\theta) + E[\nabla_{\theta^T} \xi_i(\theta)] \psi_i(\theta). \]

Similar to the derivation of (4.14), we may use (4.17) and the fact that \( \psi_{\delta,i}(\theta) \) is an instantaneous transformation of \( W_i \) to show that, under certain regularity conditions, (4.15) holds for \( \Sigma_{\delta} = E[\psi_{\delta,i}(\theta)^2] \) with \( k = 1 \). By this modification, the generalized test in (4.16) becomes applicable to testing the LATE hypotheses.

Because \( \Sigma_{\delta} \) is quite complicated in applications, it is useful to consider \( \hat{\Sigma}_{\delta} \) a bootstrap estimator, rather than a plug-in estimator, for \( \Sigma_{\delta} \) in the viewpoint of practitioners. In this study, we use the following procedure to compute the bootstrap estimator \( \hat{\Sigma}_{\delta} \):

1. Generate a bootstrap sample: \( \{W_{i(j)}^*\}_{i=1}^n \) by randomly drawing \( \{W_i\}_{i=1}^n \) with replacement, where \( j = 1, 2, \ldots, B \), and \( B \) denotes the number of replications.

2. Generate the bootstrap estimator vector: \( (\hat{\alpha}_{j}^*, \hat{\beta}_{j}^*, \hat{\gamma}_{j}^*)^T \) by using \( \{W_{i(j)}^*\}_{i=1}^n \) in place of the role of \( \{W_i\}_{i=1}^n \) in the computation of \( (\hat{\alpha}^T, \hat{\beta}^T, \hat{\gamma}^T)^T \).

3. Generate the bootstrap statistic: \( \hat{\delta}_j^* \) by using \( \{W_{i(j)}^*\}_{i=1}^n \) and \( (\hat{\alpha}_{j}^*, \hat{\beta}_{j}^*, \hat{\gamma}_{j}^*)^T \) in place of the role of \( \{W_i\}_{i=1}^n \) and \( (\hat{\alpha}^T, \hat{\beta}^T, \hat{\gamma}^T)^T \) in the computation of \( \hat{\delta} \).

4. Generate the sample of bootstrap statistic \( \{\hat{\delta}_j^*\}_{j=1}^B \) by repeating the above two steps for \( j = 1, 2, \ldots, B \).

5. Compute the bootstrap asymptotic covariance matrix estimator for \( \Sigma_{\delta} \):
\[ \hat{\Sigma}_{\delta} = \frac{1}{B} \sum_{j=1}^B \left( \sqrt{n} \left( \hat{\delta}_j^* - \frac{1}{B} \sum_{j=1}^B \hat{\delta}_j^* \right) \right)^T \left( \sqrt{n} \left( \hat{\delta}_j^* - \frac{1}{B} \sum_{j=1}^B \hat{\delta}_j^* \right) \right)^T. \]

By using this bootstrap method, we can implement the CLATE tests and the LATE tests in a simple way.
5 Monte Carlo Simulation

In this section, we conduct a Monte Carlo simulation to assess the finite-sample performance of the identification and estimation methods. This simulation is based on the following data generating process (DGP) for \( W \):

- \( X = (1, X_1, X_2)^\top \), \( Z_1 = I(U_{Z_1} \geq 0) \) and \((U_{Z_1}, Z_2, X_1, X_2)^\top \sim N(0, I_4)\); \( I_4 \) stands for the \( 4 \times 4 \) identity matrix.

- \((\tilde{U}_Y, U_M, U_D)^\top \) is independent of \((U_{Z_1}, Z_2, X_1, X_2)^\top \), and has the distribution:

\[
\begin{bmatrix}
\tilde{U}_Y \\
U_M \\
U_D
\end{bmatrix} \sim N
\begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix}, \begin{bmatrix}
1 & \rho_{ym} & \rho_{yd} \\
\rho_{ym} & 1 & 0 \\
\rho_{yd} & 0 & 1
\end{bmatrix}
\]

\[(5.1)\]

where \( \phi(\cdot) \) and \( \Phi(\cdot) \) are, respectively, the PDF and the CDF of \( N(0, 1) \).

- \((Y, M, D)\) is generated from the process:

\[
\begin{align*}
Y &= \tilde{h}(D, M, X) - \tilde{g}(D, M, X) + U_Y, \\
M &= 1(\alpha_d D + \alpha_{z_2} Z_2 + U_M \geq 0), \\
D &= 1(\gamma_z Z_1 + \gamma_x X_1 + U_D \geq 0),
\end{align*}
\]

\[(5.3)\]

where

\[
\tilde{h}(D, M, X) = h(X, \beta_{dj}) = \beta_{dj(0)}^h + \beta_{dj(1)}^h X_1
\]

and

\[
\tilde{g}(D, M, X) = g(X, \beta_{dj}) = \sqrt{\frac{2}{\pi}} \exp(\beta_{dj(2)}^g X_2),
\]

\[(5.4)\]

if \((D, M) = (d, j)\), for \( d, j \in \{0, 1\}\).
This DGP is encompassed by the parametric structure model for \((Y, M, D)\) in Assumptions 3.2. Given (5.1) and (5.3), \(M\) and \(D\) are endogenous when \(\rho_{ym}\) and \(\rho_{yd}\) are non-zero; meanwhile, the \(U_Y\) in (5.2) amounts to \(U_Y = \sigma_y(\tilde{U}_Y - E[\tilde{U}_Y|X,C])\) and hence satisfies the condition: \(E[U_Y|X,C] = 0\) in Assumptions 3.2(b); see the Appendix for the derivation of \(E[\tilde{U}_Y|X,C]\). The parameter \(\sigma_y\) in (5.2) is introduced to investigate how the degrees of data noise may affect the performance of our method. The functional form of \(g(X, \beta^g_{dj})\) in (5.4) is non-negative. It is motivated by a popular specification for the production inefficiency term in the tradition context (2.1), in which the conditional mean of \(u|X\) in (2.2) is often set to \(g(X, \beta^g) = \sqrt{2/\pi} \exp(\beta^g \top X)\) under the assumption that \(u|X\) has a conditional half-normal distribution.

In the simulation, we consider the following parameter setting for this DGP:

- \(\rho_{ym} = \rho_{yd} = \rho = 0.1\) or 0.7;
- \(\sigma_y = 0.1, 0.5\) or 1;
- \((\alpha_d, \alpha_z, \gamma_{z1}, \gamma_{x1}) = (0.1, 1, 1, 0.1)\) and
  \[
  (\beta^h_{dj(0)}, \beta^h_{dj(1)}, \beta^g_{dj(2)}) = \begin{cases} 
  (0.3, 0.3, 0.15), & \text{if } (d, j) = (1, 1), \\
  (0.6, 0.6, 0.30), & \text{if } (d, j) = (1, 0), \\
  (0.9, 0.9, 0.45), & \text{if } (d, j) = (0, 1), \\
  (1.2, 1.2, 0.60), & \text{if } (d, j) = (0, 0). 
  \end{cases}
  \]

The correlation coefficient \(\rho\) measures the strength of endogeneity, and the standard deviation \(\sigma_y\) measures the degrees of data noise. In the simulation, we let the model for \((Y, M, D)\) be correctly specified for (5.3). For simplicity, we focus on the submodel for \(Y\), and estimate \(\beta_d = (\beta^h_{d1(0)}, \beta^h_{d1(1)}, \beta^h_{d0(0)}, \beta^h_{d0(1)}, \beta^g_{d1(2)}, \beta^g_{d0(2)})\top\), for \(d \in \{0, 1\}\), by the WNLSE in (4.4) with the replacement of \(\hat{\alpha}\) by \(\alpha\). The sample size is \(n = 250, 1,000\) or 10,000, and the number of replications is 1,000. This simulation design allows us to investigate the estimator’s performance in samples of different sample sizes and degrees of endogeneity and
data noise by using different \((n, \rho, \sigma_y)\)'s. We report the sample mean, bias, and the mean square error (MSE) of the estimates and they are calculated from the 1,000 replications. Results are reported in Table 2 to Table 4.

Table 2 presents the simulation results with \(\sigma_y = 0.1\). The table is further divided into two parts, with the upper panel reporting cases of \(\rho = 0.1\) (weak endogeneity) and the lower panel \(\rho = 0.7\) (strong endogeneity). Given the small value of \(\sigma_y\), the models are in general easy to estimate, and the bias and the MSE are small even for small samples \((n = 250)\). The largest MSE in this case is 0.038. We notice that \(\hat{\beta}^{\rho}\) tends to have larger bias compared to \(\hat{\beta}^{h}\), and we surmise that this result is due to the nonlinearity of the \(g(\cdot)\) function. On the other hand, the size of the MSE appears to come mainly from the variance of the estimate, instead of the bias. This is indicated by the fact that MSE equals the sum of the variance and the square of bias.

Different values of \(\rho\) do not appear to exert strong impacts on the estimation, as can be seen by comparing the upper and the lower panels of the table. Although the MSE appears to be somewhat larger with a larger degree of endogeneity of the model \((\rho = 0.7)\), the difference margin is small.

Table 3 and 4 present the simulation results with \(\sigma_y = 0.5\) and 1, respectively. Larger values of \(\sigma_y\) make the data more noisy, which has adverse effects on the nonlinear model’s estimation. For small samples \((n = 250)\), the bias and the MSE from cases of \(\sigma_y = 1\) is about 10 to 15 times larger than those in the cases of \(\sigma_y = 0.1\). A back-of-the-envelope calculation indicates that the increase in MSE mainly comes from increases in the variance of the estimate, which is not surprising. The increases in bias and MSE, nevertheless, quickly subside as the sample size becomes bigger. With \(n = 10,000\), both of the bias and the MSE are numerically not much different in the cases of \(\sigma_y = 0.1\) and \(\sigma_y = 1\).

\footnote{Our empirical example would most likely to resemble the cases with \(n = 250\) and \(\sigma_y = 0.1\). The actual sample size of the empirical example is 247, and by choosing \(\sigma_y = 0.1\), we are able to match closely the \(R^2\) statistics from an OLS regression of \(Y\) on \(X\).}
In Table 3 and 4, the bias and the MSE are reduced as the sample size increases, which is the evidence of the estimator's consistency. Although there are more cases in Table 2 where the reductions in the bias and the MSE are not obvious, we note that the two statistics are quite small in this table to begin with, and sampling errors may be important in explaining the differences in these cases.

6 Empirical Illustration

We apply the stochastic frontier model of treatment effects with mediator to study the effects of dams on India’s agricultural production in the period of 1981 to 1987 at the district level. The treatment variable ($D$) is an indicator of the building of dams, and the mediator variable ($M$) is constructed from the ratio of irrigated area to total cultivated area. The empirical goal is thus to examine whether having dams would impact the technology and efficiency of local agricultural production, and whether the impact is conducted through the mediator or not.

As motivated by Duflo and Pande (2007), India has one of the largest numbers of dams in the world, and over 90 percent of the dams were built for the purpose of agricultural irrigation. Proposals of dam constructions were initiated by the state government, which were then examined and selected by a federal committee. The construction remained largely the state government’s responsibility although outside funding may be available. One of the most important funding sources is World Bank which extended loans for dam constructions on the justification of agricultural growth and rural poverty alleviation. Given that India had been for many years the largest receiver of World Bank lending for irrigation, an evaluation of the policy effectiveness is important.

In the analysis, the dam construction is the treatment and the subsequent agricultural production is the outcome. An endogeneity issue arises in this context because dam construction decisions may be directly related to poverty and production outcomes of the
local. Duflo and Pande (2007) showed that this endogeneity problem may be amended by using the data of river gradient as an instrument for the geographic suitability of dam construction.

The dataset is based on the one used by Duflo and Pande (2007), although their empirical question is different from ours. In their paper, Duflo and Pande are interested in the issue of whether the dam construction would benefit those who live downstream from a dam while hurting those who live in the vicinity of and upstream from a dam. The issue has been debated for some time and Duflo and Pande (2007) is the first to provide rigorous econometrics analysis using the district-level data. Duflo and Pande estimate the panel data using IV to control for the endogenous decision of dam constructions. Their results confirm the differential effects.

In comparison, our focus is on the effects of dams on the agricultural production in the district where the dam is located. Our empirical methods are different, too. We conduct the analysis in the framework of treatment effects with mediators, and the effects on technology and efficiency are also analyzed in the context of stochastic frontier models. None of the mediator effect nor the technology and efficiency channels are discussed in Duflo and Pande (2007). On the other hand, we adopt Duflo and Pande’s (2007) identification strategy to deal with the issue of endogenous decision of building dams.

6.1 Data and Model Specification

The dataset contains information of agricultural production, geographic characteristics, and the number of dams at the district level from 1976 to 1987 in India. Owing to the data limitation, we focus our attention on the effect of dams built between 1976 to 1980 (the treatment period) on the agricultural production from 1981 to 1987 (the evaluation period). Because our estimator requires each observation to be independent, the requirement would be better served by converting the panel data to a cross-section structure. We did this by taking the sample average of the variables over time for each given district. After
deleting observations with missing values on key variables, the dataset contains a total of 247 observations. Variables used in the model are explained below. Relevant and additional information of the variables can be found in Duflo and Pande (2006) and the Data Appendix of Duflo and Pande (2007).

The output variable $Y_i$ is the log of agricultural production ($production_i$) of the $i$th district with $i = 1, 2, \ldots, n$ and $n = 247$. The following vector comprises the key covariates being used for establishing the stochastic frontier model:

$$X_{i}^{hg} = (1, \text{fertilizer}_i, \text{land}_i, \text{rain}_i, elevation_i, pre\_dam_i, \text{rain2}_i),$$

where $\text{fertilizer}_i$ is the log of fertilizer used in district $i$, $\text{land}_i$ is the log of the gross cultivated area, $\text{rain}_i$ is the rainfall variable which is measured as the fractional deviation of the district’s rainfall from the district mean, $elevation_i$ is the log of mean elevation of district $i$, $pre\_dam_i$ is the number of dams that existed in the district $i$ at the beginning of the treatment period, and $\text{rain2}_i$ is calculated as the square of $\text{rain}_i$.

The stochastic production frontier model is assumed to have a linear frontier function and a half-normal random variable of technical inefficiency. This assumption leads to a scaled exponential function of $g(\cdot)$ in addition to a linear function of $h(\cdot)$:

$$h(X_i, \beta^h_{d,j}) = X_i^{h\top} \beta^h_{d,j},$$
$$g(X_i, \beta^g_{d,j}) = \sqrt{\frac{2}{\pi}} \exp(X_i^{g\top} \beta^g_{d,j}),$$

where $X_i^h$ and $X_i^g$ are subvectors of $X_i^{hg}$, for $d, j \in \{0, 1\}$. As mentioned before, $h(\cdot)$ and $g(\cdot)$

---

5 For a random variable $u$ that follows a half-normal distribution, $u \sim N^+(0, \sigma_u^2)$, it is common in the stochastic frontier literature to parameterize $\sigma_u$ by a vector $X$ in order to investigate the relationship between the variables and the inefficiency: $\sigma_u = \exp(X_i^{g\top} \beta^g_{d,j})$. The conditional mean of $u$ is therefore given by $E(u|X_i^g) = \sqrt{\frac{2}{\pi}} \sigma_u = \sqrt{\frac{2}{\pi}} \exp(X_i^{g\top} \beta^g_{d,j})$. Therefore, the specification of $g(\cdot)$ can be seen as a result of the parameterization of the distribution. 

---

37
$g(\cdot)$ are allowed to include different covariates. In this study, we choose

\[ X^h_i = (1, \text{fertilizer}_i, \text{land}_i, \text{rain}_i, \text{elevation}_i, \text{pre}_\text{dam}_i)^\top, \]  \hspace{1cm} (6.3)

\[ X^g_i = (1, \text{rain}_2_i)^\top; \hspace{1cm} (6.4) \]
correspondingly, $\beta^h_{dj}$ and $\beta^g_{dj}$ are composed of the following parameters:

\[ \beta^h_{dj} = \left( \beta^h_{dj(0)}, \beta^h_{dj(f)}, \beta^h_{dj(L)}, \beta^h_{dj(r)}, \beta^h_{dj(e)}, \beta^h_{dj(p)} \right)^\top, \]  \hspace{1cm} (6.5)

\[ \beta^g_{dj} = \left( \beta^g_{dj(0)}, \beta^g_{dj(r2)} \right)^\top. \hspace{1cm} (6.6) \]

To explain this choice of $X^h_i$ and $X^g_i$, note that $\text{fertilizer}_i$, $\text{land}_i$ and $\text{rain}_i$ are included in the frontier function $h(\cdot)$ because they are direct inputs to agricultural production. We also add two other environmental variables to $h(\cdot)$ to control for the production environment: $\text{elevation}_i$ and $\text{pre}_\text{dam}_i$. Geographical properties are often important to agricultural production and elevation is one of them. For instance, precipitation and temperature may change with the elevation, which in turn affects agricultural yields. The variable $\text{elevation}_i$ is included to control for the effect. The variable $\text{pre}_\text{dam}_i$ is included as a control of initial conditions for the program being evaluated. Given the technology, inputs, and geographical conditions, climate variability and uncertainty may hamper the extent to which the production potential may be realized. In other words, the variability and uncertainty affect the production efficiency. We therefore include $\text{rain}_2_i$ in the inefficiency function $g(\cdot)$. This variable measures the volatility of rainfall in the district $i$.

The treatment status variable, $D_i$, is determined by whether the district $i$ has a larger-than-medium number of dams built during the treatment period. If the number is above the medium of the sample, we assign $D_i = 1$ and $D_i = 0$ for otherwise. The sample average of new dams built in the period is 2.50 with a standard deviation equal to 4.57, and the medium is 0.\footnote{We found the results are sensitive to how the $D_i$ is defined. For instance, changing the threshold to the mean figure of the new dams built, the model would not converge, indicating a model mis-specification.}
To deal with the endogeneity problem of dam-building decision, we construct a binary IV $Z_{1i}$ for the treatment status based on the district’s river gradients. Duflo and Pande (2007) also use the information of river gradients to construct an IV to control for the endogeneity of dam placement. As explained by Duflo and Pande, engineering consideration suggests that the likelihood of dam constructions increases if the river gradient is either gentle (1.5-3 percent) or very steep (more than 6 percent, mostly for hydropower dams). We construct a variable of geographic suitability of dam construction using the fraction of the district in which gentle-gradient or steep-gradient rivers (defined above) flow through. The binary instrument is then obtained by assigning a value of 1 to $Z_{1i}$ if the fraction is above the sample average and a value of 0 for otherwise. The results are almost the same if we change the criterion from the sample average to the medium of the sample.

The binary mediator variable $M_i$ is measured from the ratio of a district’s irrigated area to its total cultivated area. Given that enabling irrigation is one of the benefits of building dams, we presume that a major mediator through which the effect of dams impacts the agricultural production is the irrigated land. Districts with larger portion of irrigated land may better capture the beneficial effects of dams. We compute the average ratio of irrigated land for each district in the evaluation period, and we assign a value of 1 to $M_i$ if the district $i$’s ratio is larger than the median of the sample. The value of $M_i$ is equal to 0 otherwise. We then select the log of the total river length in district $i$, $Z_{2i}$, as the IV for $M_i$. Others being equal, wide river networks (and thus longer total river length) make irrigation easier, and the network itself is not directly related to agricultural production. Therefore, the river length variable may be a suitable IV for $M_i$ in this context.

Table 5 reports the summary statistics of main variables used in the analysis. As it shows, districts that built more dams in the treatment period (i.e., districts with $D = 1$) appear to have larger cultivated areas ($land$) although they were likely to situate in higher elevations ($elevation$). Rainfalls and fertilizers were also less in these districts, and the
agricultural production was of smaller amount. Although the picture seems to paint a familiar picture of dams being built on poor mountainous regions, our analysis in this section shall help to shed light on the impact of dams in more details.

6.2 Estimation Procedure

In this empirical study, we set \( Q_M(\cdot) \), \( Q_D(\cdot) \) and \( Q_{Z_1}(\cdot) \) as \( \Phi(\cdot) \), that is the CDF of \( N(0,1) \), for simplicity. Thus, the binary response models for \( M|D, Z_2, X \), for \( D|Z_1, X \) and for \( Z_1|X \), discussed in Section 3, are specified as probit models.

The submodel for \( M|D, Z_2, X \) in (3.5), is the following probit model:

\[
E[M_i|D_i, Z_{2i}, X_i] = \Phi(\alpha_d D_i + \alpha_{z_2} Z_{2i} + X_i^\top \alpha_x),
\]

(6.7)

where \( X_i \) is a list of explanatory variables similar to those in Equation (2) of Duflo and Pande (2007). The equation was used by the authors to predict the number of dams built in a district in a particular year. The list of variables include the \( X^{hg} \) in (6.1), India state dummies, the predicted number of dams per state in the 1970, and the interactions of that predicted number with variables such as district elevations, district slope, and the size of the district. We obtain \( \hat{\alpha}_q = (\hat{\alpha}_d, \hat{\alpha}_{z_2}, \hat{\alpha}_x)^\top \) from the ML estimation of this model, which is later used to compute \( m_d'(\cdot) \) in (3.6).

To compute \( m_d'(\cdot) \), we also need the PDF of \( Z_{2i} \); that is, \( f_{Z_2}(\cdot, \alpha_f) \). By examining the histogram of \( Z_{2i} \), we observe that \( Z_{2i} \) is likely to have a bimodal distribution which may be described by a mixture distribution of two normals. We thus make the following specification:

\[
f_{Z_2}(z_2, \alpha_f) = \frac{\alpha_{f(p)}}{\alpha_{f(s1)}} \phi \left( \frac{z_2 - \alpha_{f(m1)}}{\alpha_{f(s1)}} \right) + \frac{1 - \alpha_{f(p)}}{\alpha_{f(s2)}} \phi \left( \frac{z_2 - \alpha_{f(m2)}}{\alpha_{f(s2)}} \right),
\]

(6.8)

where \((\alpha_{f(m1)}, \alpha_{f(s1)})\) and \((\alpha_{f(m2)}, \alpha_{f(s2)})\) are the mean and the standard deviation parameters of the two normal distributions, respectively, and \( \alpha_{f(p)} \) and \((1 - \alpha_{f(p)})\) are the weights of the distributions. We obtain the estimator \( \hat{\alpha}_f = (\hat{\alpha}_{f(m1)}, \hat{\alpha}_{f(s1)}, \hat{\alpha}_{f(m2)}, \hat{\alpha}_{f(s2)}, \hat{\alpha}_{f(p)})^\top \)
by the ML method. With \( \hat{\alpha}_m = (\hat{\alpha}^T \beta_r, \hat{\beta}_f)^T \) at hand, the weighting function \( m_{d'}(\cdot) \) in (3.6) is then evaluated at \( \hat{\alpha}_m \) and computed by a numerical integration.

In addition, we also specify the propensity-score model in (3.7) as a probit model:

\[
E[Z_i; X_i] = \Phi(X_i^T \alpha_{z_i}),
\]

(6.9)

and estimate \( \alpha_{z_i} \) by the MLE \( \hat{\alpha}_{z_i} \). Given the first-step estimator \( \hat{\alpha} = (\hat{\alpha}^T \beta_r, \hat{\beta}_f, \hat{\alpha}_{z_i})^T \), we estimate the weight parameter \( w(d, d', \alpha) \) in (3.8) by \( w(d, d', \hat{\alpha}) \) and estimate the parameter vectors of (6.2) with \( j = 0,1 \), \( \beta_d = (\beta_{h1, d}^T, \beta_{h0, d}^T, \beta_{g1, d}^T, \beta_{g0, d}^T)^T \), by the WNLSE \( \hat{\beta}_d = (\hat{\beta}_{h1, d}^T, \hat{\beta}_{h0, d}^T, \hat{\beta}_{g1, d}^T, \hat{\beta}_{g0, d}^T)^T \) in (4.4), for \( d \in \{0,1\} \).

### 6.3 Estimation Results

Table 6 shows the WNLSEs for the \( \beta_{d_j}^h \)'s and the \( \beta_{d_j}^g \)'s, for \( d, j \in \{0,1\} \). Results of \( (d, M(d')) = (1, M(d')) \) are for the treated group \( (D_i = 1) \), and those of \( (d, M(d')) = (0, M(d')) \) are for the untreated group \( (D_i = 0) \).

For both of the treated and the untreated groups, fertilizer is only significant when the mediator effect is taken into account in the model (\( \hat{\beta}_{11(f)}^h \) and \( \hat{\beta}_{01(f)}^h \)). In other words, fertilizer and irrigation may be complements in agricultural production. For the size of cultivated land, \( \hat{\beta}_{11(L)}^h, \hat{\beta}_{10(L)}^h, \alpha_{z_i}^2 - \hat{\beta}_{11(L)}^h, \alpha_{z_i}^2 - \hat{\beta}_{10(L)}^h \) are significant, but \( \hat{\beta}_{01(L)}^h \) is insignificant. The latter result indicates that if the irrigation-based mediator effect is the concern, increasing the cultivated area while building no dam would not help raising agricultural production.

For the rainfall variable, \( \hat{\beta}_{11(r)}^h \) is significantly positive, indicating that by building dams, rainfall can be effectively used in the production through irrigated lands. However, \( \hat{\beta}_{10(r)}^h \) and \( \hat{\beta}_{01(r)}^h \) are both significantly negative. This result may imply that without dams or the irrigation systems to effectively control water supply, an additional increase in the rainfall variable may not contribute to agricultural production.

7 Other estimated auxiliary parameters that may of interests include (partial list):

\( (\hat{\alpha}_d, \hat{\alpha}_{z_2}, \hat{\alpha}_{f(m1)}, \hat{\alpha}_{f(s1)}, \hat{\alpha}_{f(m2)}, \hat{\alpha}_{f(s2)}, \hat{\alpha}_{f(p)}, \hat{\alpha}_{z_1.f}, \hat{\alpha}_{z_1.L}, \hat{\alpha}_{z_1.r}, \hat{\alpha}_{z_1.e}, \hat{\alpha}_{z_1.p}, \hat{\alpha}_{z_1.r2}) \)

= (0.732, 1.286, -0.034, 0.572, -3.739, 0.503, 0.539, -0.120, 0.265, 1.746, 0.585, 14.185, 1.278).
rainfall could be harmful. Intriguingly, $\beta_{00(r)}$ is significantly positive. One possibility is that without dams and irrigation systems, rainfall is the only source of water supply and hence is positive to the production in existing cultivated areas.

Among the environmental variables, the effect of elevation appears to be insignificant in the model. On the other hand, the variable $\text{pre}_\text{dam}_i$, which is included to control for initial conditions of district $i$ before the treatment period, are significant in all of the cases. Since the variable may approximate the local’s initial conditions on the economic prosperity, history of agricultural cultivation, the vintage of agricultural infrastructures, etc., the sign of the coefficients could be ambiguous.

For the rainfall volatility variable in the inefficiency function, the coefficients $\hat{\beta}_{11(r2)}^g$, $\hat{\beta}_{10(r2)}^g$, and $\hat{\beta}_{00(r2)}^g$ are all positive and significant while $\hat{\beta}_{01(r2)}^g$ is positive but insignificant. The result indicates that, in most of the situations, the volatility hampers the production potential to be fully realized and thus cause the inefficiency to increase.

To formally evaluate the effects of dams and its mediation channel, we conduct hypothesis testings on CLATE and LATE, as shown in Section 2.2 and 2.3:

$$CLATE(x) = E[Y(1,M(1))|X=x,C] - E[Y(0,M(0))|X=x,C],$$

$$LATE = E[Y(1,M(1))|C] - E[Y(0,M(0))|C].$$

The $x$ is a value of $X$ which represents a particular type of district in this example. A reject of the null hypothesis of $\text{CLATE}(x) = 0$ for all $x$’s, therefore, indicates that potential outputs with and without treatments are not equal for at least some of the districts in the subpopulation of compliers. LATE, on the other hand, is a test of the effect over the averaged compliers. As discussed in Section 3.2, we base the LATE components on the $\Delta^*(X, \gamma)$ in (3.17) and the $\Delta^*(\gamma)$ in (3.18) that are built on the treatment-status model for $D|Z_1, X$ in (3.11). In this empirical study, we also specify this model as the following probit model:

$$E[D_i|Z_{1i}, X_i] = \Phi(\gamma_z Z_{1i} + X_i^{hg^T} \gamma_x),$$

(6.10)
and estimate $\gamma = (\gamma_z, \gamma_x^\top)^\top$ by the MLE $\hat{\gamma} = (\hat{\gamma}_z, \hat{\gamma}_x^\top)^\top$. Accordingly, we base the LATE component estimators on the $\Delta^*(X_i, \hat{\gamma})$’s and $\Delta^*(\hat{\gamma})$ as described in Section 4.1.

For both CLATE and LATE, the output effect can be broken down into the frontier effect and the inefficiency effect. The output, frontier, and inefficiency effects can each be further decomposed into the direct effect and the indirect (mediator) effect in the mediation analysis; see Table 4. Thus, there is a total of nine testable hypotheses/effects for each of the CLATE and LATE which have been discussed in Sections 2.2 and 2.3.

Table 7 presents results on the tests of CLATE. The first row reports the tests of the null hypotheses that building dams (the treatment) has no effect on output ($H_0^o : CLATE(x) = 0, \forall x$), no effect on the frontier function ($H_0^{oh} : CLATE_h(x) = 0, \forall x$), and no effect on the inefficiency ($H_0^{og} : CLATE_g(x) = 0, \forall x$), respectively. The second and the third rows report results on the mediation analysis, i.e., the tests of direct and indirect effects of building dams on the output, frontier function, and inefficiency. We report the $\chi^2_k$ statistics and the associated $p$-value of the Wald test, where the degrees of freedom, $k$, is equal to the number of restrictions of the test.

Results show that all of the nine null hypotheses of no effect are convincingly rejected. The implication is that, at least for some districts among the compliers, building dams (the treatment) has significantly impacted the agricultural production, and the impact both works directly on the production and indirectly through the irrigated lands (the mediator). As well, the impact changes both of the production frontier function and the inefficiency function.

Table 8 presents estimated effects of LATE, and the content is arranged in a way similar

---

8We note that $\hat{\alpha}_d = 0.732$. As discussed on page 13, $H_i^o$, $H_i^{oh}$, and $H_i^{og}$ are tested based on parameter restrictions on $\beta_{dj}^h$ and $\beta_{dj}^g$.

9For instance, there are 2 coefficients in the $g(\cdot)$ function, and so the null hypothesis $H_{p^g}$ would contain 6 coefficient restrictions across the four vectors of $\beta_{dj}^g$, for $d, j \in (0, 1)$. The $\chi^2$ statistic thus has the degree of freedom equal to 6.
Results indicate that although all of the three effects on output (LATE, DLATE, ILATE) are positive, they are statistically insignificant. In fact, among all of the estimated effects, only two of the effects on the inefficiency (LATE$_g$ and DLATE$_g$) are statistically significant and they are with negative signs. The negative effects indicate reductions in production inefficiency (improvements in production efficiency) after building dams. Therefore, for the mean complier, building dams has significantly improved the local’s agricultural production efficiency, and the mediation analysis further shows that the improvement comes directly from dams rather than indirectly through the mediator. However, we notice that the LATE on the frontier function is generally negative (though insignificant), which may have offset the benefit of efficiency improvements. As explained by McCully (2001) and Singh (2002) and cited in Duflo and Pande (2007), areas around and upstream to the dams may suffer large losses because “flooding reduces agricultural and forest land, and increased salinity and waterlogging reduces the productivity of land in the vicinity of the reservoir.” (Duflo and Pande 2007, p.602)

Putting everything together, we infer that building dams may have significant effects for some types of districts with certain $x$ among the compliers, but the average effects on the compliers (when $X$ is integrated out) are mostly insignificant albeit positive. The later result is consistent with that of Duflo and Pande (2007) whose regression analysis also shows an insignificant increase in agricultural production in the district where the dam is located. Our stochastic frontier analysis, nevertheless, sheds further light into the result: building dams may actually have improved the local’s production efficiency directly (as oppose to indirectly through the mediator), but the positive impact from efficiency improvement is offset by the negative impact from the deteriorated frontier function of production.
7 Conclusion

In this paper we propose an endogenous treatment effect model with a mediator for stochastic frontier analysis. In the context of a stochastic frontier analysis, the model would estimate a program’s total productivity effect and decompose it into the technology and efficiency components. The decomposition allows policy makers to identify channels through which the policy takes effects on productivity. The mediation analysis, which is also part of the model, further allows us to test whether the aforementioned effects take places directly from the program or indirectly via a mediator. Results of the tests may help policy makers to design better mechanisms for the policy, and the mediator itself may also serve as an indicator in program evaluation.

Our model is fully parametric. The endogeneity problem of the model is dealt with using a binary instrumental variable in a way similar to Abadie (2003), and the mediation analysis is conducted by extending the method of Frölich and Huber (2014) from their non-parametric context to our parametric framework.

We illustrate the application of the model using the data of India to estimate the effects of building large dams on the local’s agricultural productions. The endogeneity issue arises here because the decision to build dams in an area is likely to relate to the local’s production outcome. Our results show that dams directly improved the local’s production efficiency, but the overall effect on output was insignificant. This result is consistent with Duflo and Pande (2007) who found that a dam would significantly increase the agricultural production in districts downstream to the dam, but its production effect on the local district is rather limited.

There are two limitations to the current model. Firstly, the estimation strategy relies on the presence of exogenous inefficiency determinants to identify the inefficiency function. This may not really be an adverse restriction to the model given that the exogenous inefficiency determinants are often the focus of a stochastic frontier analysis. Secondly, the model has only one mediator. Accommodating multiple mediators in the model would be
an interesting extension in the future.

Appendix

Proof of Theorem 3.1

From (3.3) and (3.4), we can see that, conditional on \( Z_2 \) and \( X \), the potential mediator \( M(d') \) is determined by \( U_M \); conditional on \( X \), the potential treatment status \( D(z_1) \) and hence the subpopulation type \( (D(1), D(0)) \) are determined by \( U_D \). In addition, Assumption 3.3 (d2) requires that \( U_M \perp (U_D, Z_2, X) \), and hence \( U_M \perp (D(1), D(0)) \mid Z_2, X \).

Therefore,

\[
E[M(d') \mid Z_2, X, C] = E[M(d') \mid Z_2, X] = Q_M(\alpha d' + \alpha z_2 Z_2 + X^\top \alpha_x), \tag{A1}
\]

where the second equality is due to (3.5) with \( D = d' \). Let \( f_{Z_2 \mid X, C}(\cdot) \) be the conditional probability density function (PDF) of \( Z_2 \mid X, C \). By the law of total probability, we have

\[
E[M(d') \mid X, C] = \int_R E[M(d') \mid Z_2 = z_2, X, C] f_{Z_2 \mid X, C}(z_2) dz_2,
\]

where the second equality is obtained by evaluating (A1) at \( Z_2 = z_2 \); the last equality is also due to Assumption 3.4 (b) which implies \( Z_2 \perp (D(1), D(0), X) \). The result in (3.6) is a combination of (2.12) and (A2).

Proof of Theorem 3.2

Note that the model that Frölich and Huber (2014) consider is

\[
Y = \kappa(D, M, X, U_Y),
\]

\[
M = 1(\zeta(D, Z_2, X, U_M) \geq 0),
\]

\[
D = 1(\chi(Z_1, X, U_D) \geq 0),
\]

46
where $\kappa(\cdot), \zeta(\cdot)$ and $\chi(\cdot)$ are measurable functions. Theorem 4 of Frölich and Huber (2014) is proved to be true under the following assumption:

(i) IV independence:

\[
(Z_1, Z_2) \perp (U_Y, U_M)|T, X,
\]

\[
Z_1 \perp (U_Y, U_M, T)|Z_2, X;
\]

$T$ is the subpopulation type: always takers, compliers, defiers or never takers.

(ii) Conditional independence of IVs:

\[
Z_1 \perp Z_2|X.
\]

(iii) Monotonicity: $P(D(1) \geq D(0)|X) = 1$;

Existence of compliers: $E(D(1)) > E(D(0))$.

(iv) Monotonicity of $M$ in $Z_2$ and $U_M$:

$U_M$ is a continuous random variable with a distribution function $F_{U_M|X=x,c}(v)$ which is strictly increasing in the support of $U_M$ for almost all $x$’s.

$\zeta(D, Z_2, X, U_M)$ is (normalized to be) strictly monotonic in $Z_2$ and in $U_M$.

(v) Common support of $M$:

\[
0 < E[Z_1|M, U_M, X, C] < 1.
\]

We first claim that our Assumptions 3.2 and 3.3 are sufficient for these conditions. First, the types $T$ are determined by $U_D$ conditional on $X$. Therefore, our Assumption 3.3 (a) is sufficient for (i). Assumption 3.3 (b) is identical to (ii). As mentioned before, the sign restriction, $\gamma_z > 0$, in Assumption 3.3 (c) combined with the model $D = 1(\gamma_z Z_1 + X^T \gamma_x + U_D \geq 0)$ imply that $D(1) \geq D(0)$. The fact that $\gamma_z > 0$ also implies that $P(D(1) = 1|Z_1 = 1, X) > P(D(0) = 1|Z_1 = 0, X)$ a.s. in $X$ and this is equivalent to $P(D(1) = 1|X) > P(D(0) = 1|X)$ a.s. in $X$ which implies that the mass of compliers is positive a.s. in $X$. This further implies that the mass of compliers in the population is strict positive. Therefore, (iii) is satisfied. Our model for $M$ is $M = 1(\alpha_d D + \alpha_{z2} Z_2 +$
$X^\top \alpha_x + U_M \geq 0$ so $\zeta(D, Z_2, X, U_M) = \alpha_d D + \alpha_{z_2} Z_2 + X^\top \alpha_x + U_M$. The sign condition, $\alpha_{z_2} > 0$, and the model implies that $\alpha_d D + \alpha_{z_2} Z_2 + X^\top \alpha_x + U_M$ is strictly increasing in $Z_2$ and $U_M$. Therefore, Assumption 3.3 (d) combined with Assumption 3.2 is sufficient for (iv). Assumption 3.3 (e) is identical to (v). Then when we combine the proof of Theorem 3.1 of Abadie (2003) and that of Theorem 4 of Frölich and Huber (2014), with this linear index model of $M$, we can show that, for $d, d' \in \{0, 1\}$,

$$
E\left[\left(Y(d, M(d')) - h_{d'}(X, \alpha_m, b^h_{d1}, b^h_{d0}) + g_{d'}(X, \alpha_m, b^g_{d1}, b^g_{d0})\right)^2|C\right]
= E\left[\tilde{w}(d, d')\left(Y - h_{d'}(X, \alpha_m, b^h_{d1}, b^h_{d0}) + g_{d'}(X, \alpha_m, b^g_{d1}, b^g_{d0})\right)^2\right]/\Delta,
$$

where

$$
\tilde{w}(d, d') = \begin{cases} 
\frac{(Z_1 - E[Z_1|X])}{E[Z_1|X](1 - E[Z_1|X])}, & (d, d') = (1, 1), \\
\frac{(Z_1 - E[Z_1|X])}{E[Z_1|X](1 - E[Z_1|X])} \frac{f_{Z_2 | X, C}(Z_2 + \alpha_d/\alpha_{z_2})}{f_{Z_2 | X, C}(Z_2)}, & (d, d') = (1, 0), \\
(D - 1)\frac{(Z_1 - E[Z_1|X])}{E[Z_1|X](1 - E[Z_1|X])} \frac{f_{Z_2 | X, C}(Z_2 + \alpha_d/\alpha_{z_2})}{f_{Z_2 | X, C}(Z_2)}, & (d, d') = (0, 1), \\
(D - 1)\frac{(Z_1 - E[Z_1|X])}{E[Z_1|X](1 - E[Z_1|X])}, & (d, d') = (0, 0).
\end{cases}
$$

For the implication of a single-index model of $M$ on the weight functions $\tilde{w}(1, 0)$ and $\tilde{w}(0, 1)$, please see the end of Section 3.3 of Frölich and Huber (2014). Assumption 3.4 (a) implies that $E[Z_1|X] = Q_{Z_1}(X^\top \alpha_z)$ and Assumption 3.4 (b) implies that $f_{Z_2 | X, C}(Z_2) = f_{Z_2}(Z_2, \alpha_f)$. Then under our assumptions, we have $\tilde{w}(d, d') = w(d, d', \alpha)$ for $d, d' \in \{0, 1\}$. This completes our proof.

**Derivation of $E[\tilde{U}_Y|X, C]$**

We derive the conditional mean of $E[\tilde{U}_Y|X, C]$. For compliers, we have

$$
D(1) = 1(\gamma_2 z_1 + \gamma_2 x_1 + U_D \geq 0),
$$

$$
D(0) = 1(\gamma_2 x_1 + U_D \geq 0).
$$
Conditional on $X$, the event of being a compiler ($D(1) = 1, D(0) = 0$) is equivalent to $(-\gamma z_1 - \gamma x_1, X_1 \leq U_D \leq -\gamma x_1, X_1)$. Therefore,

$$E[\tilde{U}_Y|X, C] = E[\tilde{U}_Y| - \gamma z_1 - \gamma x_1, X_1 \leq U_D \leq -\gamma x_1, X_1]$$

$$= E[\rho_ydU_D + \epsilon| - \gamma z_1 - \gamma x_1, X_1 \leq U_D \leq -\gamma x_1, X_1]$$

$$= E[\rho_ydU_D| - \gamma z_1 - \gamma x_1, X_1 \leq U_D \leq -\gamma x_1, X_1]$$

$$= \rho_yd \frac{E[U_D \cdot 1(-\gamma z_1 - \gamma x_1, X_1 \leq U_D)] - E[U_D \cdot 1(-\gamma x_1, X_1 \leq U_D)]}{P(-\gamma z_1 - \gamma x_1, X_1 \leq U_D) - P(-\gamma x_1, X_1 \leq U_D)}$$

$$= \rho_yd \left( \frac{\phi(\gamma z_1 + \gamma x_1, X_1) - \phi(\gamma x_1, X_1)}{\Phi(\gamma z_1 + \gamma x_1, X_1) - \Phi(\gamma x_1, X_1)} \right)$$

where the second equality follows from the fact that, given (5.1), $\tilde{U}_Y$ can be rewritten as $\rho_ydU_D + \epsilon$ where $\epsilon$ is normally distributed with mean 0 and is independent of $U_D$. Third equality holds because the $\epsilon$ is independent of $U_D$ and its mean is 0. The fourth equality follows by definitions. The fifth equality holds by the fact that $E[U_D \cdot 1(-a \leq U_D)] = \phi(a)$ and $P(U_D \geq -a) = P(U_D \leq a)$ by the symmetry of $U_D$. Then the last equality follows.

See also Cameron and Trivedi (2005, Proposition 6.1) for a related discussion.

References


Table 1: CLATE and LATE components

<table>
<thead>
<tr>
<th></th>
<th>CLATE (x)</th>
<th>CLATE_h (x)</th>
<th>CLATE_g (x)</th>
<th>LATE</th>
<th>LATE_h</th>
<th>LATE_g</th>
</tr>
</thead>
<tbody>
<tr>
<td>program evaluation</td>
<td>CLATE (x)</td>
<td>CLATE_h (x)</td>
<td>CLATE_g (x)</td>
<td>LATE</td>
<td>LATE_h</td>
<td>LATE_g</td>
</tr>
<tr>
<td>mediation analysis</td>
<td>DCLATE (x)</td>
<td>DCLATE_h (x)</td>
<td>CLATE_g (x)</td>
<td>DLATE</td>
<td>DCLATE_h</td>
<td>DCLATE_g</td>
</tr>
<tr>
<td>analysis</td>
<td>ICLATE (x)</td>
<td>ICLATE_h (x)</td>
<td>CLATE_g (x)</td>
<td>ILATE</td>
<td>ICLATE_h</td>
<td>ICLATE_g</td>
</tr>
<tr>
<td>ρ = 0.1</td>
<td>N=250</td>
<td>N=1,000</td>
<td>N=10,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>---------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σₐᵧ = 0.1</td>
<td>mean</td>
<td>bias</td>
<td>MSE</td>
<td>mean</td>
<td>bias</td>
<td>MSE</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₁(0)</td>
<td>0.299</td>
<td>-0.001</td>
<td>0.022</td>
<td>0.300</td>
<td>-0.0004</td>
<td>0.002</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₁(1)</td>
<td>0.306</td>
<td>0.006</td>
<td>0.038</td>
<td>0.298</td>
<td>-0.002</td>
<td>0.004</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₁(2)</td>
<td>0.149</td>
<td>-0.001</td>
<td>0.012</td>
<td>0.149</td>
<td>-0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₀(0)</td>
<td>0.599</td>
<td>-0.001</td>
<td>0.020</td>
<td>0.600</td>
<td>-0.0003</td>
<td>0.002</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₀(1)</td>
<td>0.606</td>
<td>0.006</td>
<td>0.034</td>
<td>0.599</td>
<td>-0.001</td>
<td>0.004</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₀(2)</td>
<td>0.275</td>
<td>-0.025</td>
<td>0.015</td>
<td>0.296</td>
<td>-0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₁(0)</td>
<td>1.201</td>
<td>0.001</td>
<td>0.007</td>
<td>1.195</td>
<td>-0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₁(1)</td>
<td>1.198</td>
<td>-0.002</td>
<td>0.025</td>
<td>1.195</td>
<td>-0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₁(2)</td>
<td>0.555</td>
<td>-0.045</td>
<td>0.019</td>
<td>0.579</td>
<td>-0.021</td>
<td>0.005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ρ = 0.7</th>
<th>N=250</th>
<th>N=1,000</th>
<th>N=10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>σₐᵧ = 0.1</td>
<td>mean</td>
<td>bias</td>
<td>MSE</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₁(0)</td>
<td>0.300</td>
<td>0</td>
<td>0.024</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₁(1)</td>
<td>0.305</td>
<td>0.005</td>
<td>0.044</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₁(2)</td>
<td>0.155</td>
<td>0.005</td>
<td>0.014</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₀(0)</td>
<td>0.601</td>
<td>0.001</td>
<td>0.022</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₀(1)</td>
<td>0.605</td>
<td>0.005</td>
<td>0.040</td>
</tr>
<tr>
<td>β⁽ʰ⁾₁₀(2)</td>
<td>0.282</td>
<td>-0.018</td>
<td>0.015</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₁(0)</td>
<td>0.897</td>
<td>-0.003</td>
<td>0.008</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₁(1)</td>
<td>0.894</td>
<td>-0.006</td>
<td>0.020</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₁(2)</td>
<td>0.458</td>
<td>0.008</td>
<td>0.011</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₀(0)</td>
<td>1.197</td>
<td>-0.003</td>
<td>0.008</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₀(1)</td>
<td>1.194</td>
<td>-0.006</td>
<td>0.020</td>
</tr>
<tr>
<td>β⁽ʰ⁾₀₀(2)</td>
<td>0.558</td>
<td>-0.042</td>
<td>0.019</td>
</tr>
</tbody>
</table>

True values are: (β⁽ʰ⁾₁₁(0), β⁽ʰ⁾₁₁(1), β⁽ʰ⁾₁₁(2)) = (0.3, 0.3, 0.15), (β⁽ʰ⁾₁₀(0), β⁽ʰ⁾₁₀(1), β⁽ʰ⁾₁₀(2)) = (0.6, 0.6, 0.30), (β⁽ʰ⁾₀₁(0), β⁽ʰ⁾₀₁(1), β⁽ʰ⁾₀₁(2)) = (0.9, 0.9, 0.45), (β⁽ʰ⁾₀₀(0), β⁽ʰ⁾₀₀(1), β⁽ʰ⁾₀₀(2)) = (1.2, 1.2, 0.60).
### Table 3: Simulation Results of $\sigma_y = 0.5$

<table>
<thead>
<tr>
<th>$\rho = 0.1$</th>
<th>$\sigma_y = 0.5$</th>
<th></th>
<th></th>
<th></th>
<th>$\rho = 0.7$</th>
<th>$\sigma_y = 0.5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$N=250$</td>
<td>mean</td>
<td>bias</td>
<td>MSE</td>
<td></td>
<td>$N=1,000$</td>
<td>mean</td>
<td>bias</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^h_{11(0)}$</td>
<td>0.317</td>
<td>0.017</td>
<td>0.111</td>
<td></td>
<td>$\hat{\beta}^h_{11(0)}$</td>
<td>0.296</td>
<td>-0.004</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^h_{11(1)}$</td>
<td>0.319</td>
<td>0.019</td>
<td>0.155</td>
<td></td>
<td>$\hat{\beta}^h_{11(1)}$</td>
<td>0.298</td>
<td>-0.002</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^g_{11(2)}$</td>
<td>0.152</td>
<td>0.002</td>
<td>0.043</td>
<td></td>
<td>$\hat{\beta}^g_{11(2)}$</td>
<td>0.151</td>
<td>0.001</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^h_{10(0)}$</td>
<td>0.616</td>
<td>0.016</td>
<td>0.098</td>
<td></td>
<td>$\hat{\beta}^h_{10(0)}$</td>
<td>0.596</td>
<td>-0.004</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^h_{10(1)}$</td>
<td>0.617</td>
<td>0.017</td>
<td>0.140</td>
<td></td>
<td>$\hat{\beta}^h_{10(1)}$</td>
<td>0.598</td>
<td>-0.002</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^g_{10(2)}$</td>
<td>0.226</td>
<td>-0.074</td>
<td>0.051</td>
<td></td>
<td>$\hat{\beta}^g_{10(2)}$</td>
<td>0.281</td>
<td>-0.019</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^g_{01(0)}$</td>
<td>0.906</td>
<td>0.006</td>
<td>0.030</td>
<td></td>
<td>$\hat{\beta}^g_{01(0)}$</td>
<td>0.897</td>
<td>-0.003</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^g_{01(1)}$</td>
<td>0.913</td>
<td>0.013</td>
<td>0.078</td>
<td></td>
<td>$\hat{\beta}^g_{01(1)}$</td>
<td>0.892</td>
<td>-0.008</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^g_{01(2)}$</td>
<td>0.463</td>
<td>0.013</td>
<td>0.030</td>
<td></td>
<td>$\hat{\beta}^g_{01(2)}$</td>
<td>0.469</td>
<td>0.019</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^h_{00(0)}$</td>
<td>1.208</td>
<td>0.008</td>
<td>0.031</td>
<td></td>
<td>$\hat{\beta}^h_{00(0)}$</td>
<td>1.197</td>
<td>-0.003</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^h_{00(1)}$</td>
<td>1.215</td>
<td>0.015</td>
<td>0.076</td>
<td></td>
<td>$\hat{\beta}^h_{00(1)}$</td>
<td>1.191</td>
<td>-0.009</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^g_{00(2)}$</td>
<td>0.529</td>
<td>-0.071</td>
<td>0.042</td>
<td></td>
<td>$\hat{\beta}^g_{00(2)}$</td>
<td>0.569</td>
<td>-0.031</td>
<td>0.009</td>
<td></td>
</tr>
</tbody>
</table>

**True values are:**

$\hat{\beta}^h_{11(0)}, \hat{\beta}^h_{11(1)}, \hat{\beta}^g_{11(2)} = (0.3, 0.3, 0.15),$ $\hat{\beta}^h_{10(0)}, \hat{\beta}^h_{10(1)}, \hat{\beta}^g_{10(2)} = (0.6, 0.6, 0.30),$ $\hat{\beta}^g_{01(0)}, \hat{\beta}^h_{01(1)}, \hat{\beta}^g_{01(2)} = (0.9, 0.9, 0.45),$ $\hat{\beta}^h_{00(0)}, \hat{\beta}^h_{00(1)}, \hat{\beta}^g_{00(2)} = (1.2, 1.2, 0.60).$
<table>
<thead>
<tr>
<th>( \rho = 0.1 )</th>
<th>( n = 250 )</th>
<th>( n = 1,000 )</th>
<th>( n = 10,000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_y = 1 )</td>
<td>mean</td>
<td>bias</td>
<td>MSE</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{11(0)} )</td>
<td>0.357</td>
<td>0.057</td>
<td>0.349</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{11(1)} )</td>
<td>0.341</td>
<td>0.041</td>
<td>0.539</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{g}}_{11(2)} )</td>
<td>0.132</td>
<td>-0.018</td>
<td>0.093</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{10(0)} )</td>
<td>0.654</td>
<td>0.054</td>
<td>0.312</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{10(1)} )</td>
<td>0.639</td>
<td>0.039</td>
<td>0.484</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{g}}_{10(2)} )</td>
<td>0.161</td>
<td>-0.139</td>
<td>0.118</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{01(0)} )</td>
<td>1.213</td>
<td>0.013</td>
<td>0.104</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{01(1)} )</td>
<td>1.234</td>
<td>0.034</td>
<td>0.385</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{g}}_{01(2)} )</td>
<td>0.477</td>
<td>-0.123</td>
<td>0.098</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \rho = 0.7 )</th>
<th>( n = 250 )</th>
<th>( n = 1,000 )</th>
<th>( n = 10,000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_y = 1 )</td>
<td>mean</td>
<td>bias</td>
<td>MSE</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{11(0)} )</td>
<td>0.299</td>
<td>-0.001</td>
<td>0.571</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{11(1)} )</td>
<td>0.367</td>
<td>0.067</td>
<td>0.945</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{g}}_{11(2)} )</td>
<td>0.118</td>
<td>-0.032</td>
<td>0.095</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{10(0)} )</td>
<td>0.600</td>
<td>-0.0004</td>
<td>0.511</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{10(1)} )</td>
<td>0.659</td>
<td>0.059</td>
<td>0.855</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{g}}_{10(2)} )</td>
<td>0.149</td>
<td>-0.151</td>
<td>0.120</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{01(0)} )</td>
<td>0.955</td>
<td>0.055</td>
<td>0.131</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{01(1)} )</td>
<td>0.890</td>
<td>-0.010</td>
<td>0.220</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{g}}_{01(2)} )</td>
<td>0.468</td>
<td>0.018</td>
<td>0.063</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{00(0)} )</td>
<td>1.257</td>
<td>0.057</td>
<td>0.133</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{h}}_{00(1)} )</td>
<td>1.186</td>
<td>-0.014</td>
<td>0.214</td>
</tr>
<tr>
<td>( \hat{\beta}^{\mathit{g}}_{00(2)} )</td>
<td>0.501</td>
<td>-0.099</td>
<td>0.081</td>
</tr>
</tbody>
</table>

True values are: \((\hat{\beta}^{h}_{11(0)}, \hat{\beta}^{h}_{11(1)}, \hat{\beta}^{g}_{11(2)}) = (0.3, 0.3, 0.15), (\hat{\beta}^{h}_{10(0)}, \hat{\beta}^{h}_{10(1)}, \hat{\beta}^{g}_{10(2)}) = (0.6, 0.6, 0.30), (\hat{\beta}^{h}_{01(0)}, \hat{\beta}^{h}_{01(1)}, \hat{\beta}^{g}_{01(2)}) = (0.9, 0.9, 0.45), (\hat{\beta}^{h}_{00(0)}, \hat{\beta}^{h}_{00(1)}, \hat{\beta}^{g}_{00(2)}) = (1.2, 1.2, 0.60)).
Table 5: Summary Statistics of Main Variables

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>D=0</th>
<th>D=1</th>
<th>diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>production</td>
<td>9.727</td>
<td>9.902</td>
<td>9.554</td>
<td>0.475***</td>
</tr>
<tr>
<td></td>
<td>(0.736)</td>
<td>(0.822)</td>
<td>(0.594)</td>
<td>(0.101)</td>
</tr>
<tr>
<td>fertilizer</td>
<td>16.832</td>
<td>17.003</td>
<td>16.663</td>
<td>0.536***</td>
</tr>
<tr>
<td></td>
<td>(1.096)</td>
<td>(1.167)</td>
<td>(0.996)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>land</td>
<td>6.240</td>
<td>6.175</td>
<td>6.304</td>
<td>-0.100</td>
</tr>
<tr>
<td></td>
<td>(0.510)</td>
<td>(0.552)</td>
<td>(0.458)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>rain</td>
<td>-0.017</td>
<td>0.019</td>
<td>-0.053</td>
<td>0.060***</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.102)</td>
<td>(0.076)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>elevation</td>
<td>5.574</td>
<td>5.301</td>
<td>5.846</td>
<td>-0.564***</td>
</tr>
<tr>
<td></td>
<td>(0.666)</td>
<td>(0.645)</td>
<td>(0.571)</td>
<td>(0.073)</td>
</tr>
<tr>
<td>pre_dam</td>
<td>0.055</td>
<td>0.015</td>
<td>0.095</td>
<td>-0.082***</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.029)</td>
<td>(0.103)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>rain2</td>
<td>0.095</td>
<td>0.106</td>
<td>0.085</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td>(0.126)</td>
<td>(0.143)</td>
<td>(0.018)</td>
</tr>
<tr>
<td># of obs.</td>
<td>247</td>
<td>123</td>
<td>124</td>
<td></td>
</tr>
</tbody>
</table>

Note: In the 2nd to the 4th columns, the figures are the mean and the standard deviations (in the parentheses) of the respective variables. The 5th column reports the differences between the variables in $D = 0$ and $D = 1$, and *** represents the statistical significance from two-sample t-tests on the differences. Three *** indicates that the difference is at the 1% level. The original panel dataset was converted into a cross-sectional structural by taking the sample average of the variables over time for each given district. The number of observations is thus the number of districts in the sample.
Table 6: Estimation Results

<table>
<thead>
<tr>
<th></th>
<th>$(d, M(d')) = (1, 1)$</th>
<th></th>
<th>$(d, M(d')) = (0, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$ function</td>
<td>est.</td>
<td>std. err.</td>
<td>$h$ function</td>
</tr>
<tr>
<td>$\beta_{11}^{h}(0)$</td>
<td>0.946***</td>
<td>0.024</td>
<td>$\beta_{01}^{h}(0)$</td>
</tr>
<tr>
<td>$\beta_{11}^{h}(f)$</td>
<td>0.155*</td>
<td>0.093</td>
<td>$\beta_{11}^{h}(f)$</td>
</tr>
<tr>
<td>$\beta_{11}^{h}(L)$</td>
<td>0.900***</td>
<td>0.183</td>
<td>$\beta_{01}^{h}(L)$</td>
</tr>
<tr>
<td>$\beta_{11}^{h}(r)$</td>
<td>3.718***</td>
<td>0.029</td>
<td>$\beta_{01}^{h}(r)$</td>
</tr>
<tr>
<td>$\beta_{11}^{h}(e)$</td>
<td>0.170</td>
<td>0.129</td>
<td>$\beta_{01}^{h}(e)$</td>
</tr>
<tr>
<td>$\beta_{11}^{h}(p)$</td>
<td>-1.398***</td>
<td>0.029</td>
<td>$\beta_{01}^{h}(p)$</td>
</tr>
<tr>
<td>$g$ function</td>
<td>est.</td>
<td>std. err.</td>
<td>$g$ function</td>
</tr>
<tr>
<td>$\beta_{11}^{g}(0)$</td>
<td>-2.627***</td>
<td>0.026</td>
<td>$\beta_{01}^{g}(0)$</td>
</tr>
<tr>
<td>$\beta_{11}^{g}(r2)$</td>
<td>3.611***</td>
<td>0.020</td>
<td>$\beta_{01}^{g}(r2)$</td>
</tr>
<tr>
<td># of obs.</td>
<td>124</td>
<td></td>
<td># of obs.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$(d, M(d')) = (1, 0)$</th>
<th></th>
<th>$(d, M(d')) = (0, 0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$ function</td>
<td>est.</td>
<td>std. err.</td>
<td>$h$ function</td>
</tr>
<tr>
<td>$\beta_{10}^{h}(0)$</td>
<td>8.690***</td>
<td>0.111</td>
<td>$\beta_{00}^{h}(0)$</td>
</tr>
<tr>
<td>$\beta_{10}^{h}(f)$</td>
<td>-0.032</td>
<td>0.157</td>
<td>$\beta_{00}^{h}(f)$</td>
</tr>
<tr>
<td>$\beta_{10}^{h}(L)$</td>
<td>0.610*</td>
<td>0.329</td>
<td>$\beta_{00}^{h}(L)$</td>
</tr>
<tr>
<td>$\beta_{10}^{h}(r)$</td>
<td>-4.359***</td>
<td>0.090</td>
<td>$\beta_{00}^{h}(r)$</td>
</tr>
<tr>
<td>$\beta_{10}^{h}(e)$</td>
<td>-0.424</td>
<td>0.306</td>
<td>$\beta_{00}^{h}(e)$</td>
</tr>
<tr>
<td>$\beta_{10}^{h}(p)$</td>
<td>0.235*</td>
<td>0.141</td>
<td>$\beta_{00}^{h}(p)$</td>
</tr>
<tr>
<td>$g$ function</td>
<td>est.</td>
<td>std. err.</td>
<td>$g$ function</td>
</tr>
<tr>
<td>$\beta_{10}^{g}(0)$</td>
<td>-2.576***</td>
<td>0.175</td>
<td>$\beta_{00}^{g}(0)$</td>
</tr>
<tr>
<td>$\beta_{10}^{g}(r2)$</td>
<td>6.157***</td>
<td>0.190</td>
<td>$\beta_{00}^{g}(r2)$</td>
</tr>
<tr>
<td># of obs.</td>
<td>124</td>
<td></td>
<td># of obs.</td>
</tr>
</tbody>
</table>

Note: Significance: ***: 1% level, **: 5% level; *: 10% level. The standard errors are bootstrapped from 1,000 re-samples.
Table 7: Hypothesis Testings on Conditional Local Average Treatment Effects (CLATE)

<table>
<thead>
<tr>
<th>Program</th>
<th>Output</th>
<th>Frontier</th>
<th>Inefficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evaluation</strong></td>
<td>$H^p_0 : CLATE(x) = 0, \forall x$</td>
<td>$H^{ph}_0 : CLATE_h(x) = 0, \forall x$</td>
<td>$H^{pg}_0 : CLATE_g(x) = 0, \forall x$</td>
</tr>
<tr>
<td>$W = 23094.88$</td>
<td>$W = 17622.50$</td>
<td>$W = 2788.94$</td>
<td></td>
</tr>
<tr>
<td>$k = 24$</td>
<td>$k = 18$</td>
<td>$k = 6$</td>
<td></td>
</tr>
<tr>
<td>$p$-value $= 0.00$</td>
<td>$p$-value $= 0.00$</td>
<td>$p$-value $= 0.00$</td>
<td></td>
</tr>
</tbody>
</table>

| Mediation | $H^{dp}_0 : DCLATE(x) = 0, \forall x$ | $H^{dh}_0 : DCLATE_h(x) = 0, \forall x$ | $H^{dg}_0 : DCLATE_g(x) = 0, \forall x$ |
| Analysis | $W = 5179.11$ | $W = 4192.94$ | $W = 946.91$ |
| $k = 16$ | $k = 12$ | $k = 4$ |
| $p$-value $= 0.00$ | $p$-value $= 0.00$ | $p$-value $= 0.00$ |

| $H^{dp}_0 : ICLATE(x) = 0, \forall x$ | $H^{ih}_0 : ICLATE_h(x) = 0, \forall x$ | $H^{ig}_0 : ICLATE_g(x) = 0, \forall x$ |
| $W = 3061.42$ | $W = 945.55$ | $W = 1787.02$ |
| $k = 8$ | $k = 6$ | $k = 2$ |
| $p$-value $= 0.00$ | $p$-value $= 0.00$ | $p$-value $= 0.00$ |

Note: The set of parameter restrictions of each of the tests are on page 12. The $W$ statistic has a $\chi^2$ distribution with the degree of freedom equal to $k$ which is the number of parameter restrictions of the test.

60
<table>
<thead>
<tr>
<th>Program</th>
<th>output</th>
<th>frontier</th>
<th>inefficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>evaluation</td>
<td>$\tilde{H}_P^o : \text{LATE}=0$</td>
<td>$\tilde{H}_P^{oh} : \text{LATE}_h=0$</td>
<td>$\tilde{H}_P^{og} : \text{LATE}_g=0$</td>
</tr>
<tr>
<td></td>
<td>0.117</td>
<td>-0.416</td>
<td>-0.533***</td>
</tr>
<tr>
<td></td>
<td>(0.472)</td>
<td>(0.504)</td>
<td>(0.150)</td>
</tr>
<tr>
<td>mediation</td>
<td>$\tilde{H}_d^o : \text{DLATE}=0$</td>
<td>$\tilde{H}_d^{oh} : \text{DLATE}_h=0$</td>
<td>$\tilde{H}_d^{og} : \text{DLATE}_g=0$</td>
</tr>
<tr>
<td>analysis</td>
<td>0.095</td>
<td>-0.461</td>
<td>-0.556***</td>
</tr>
<tr>
<td></td>
<td>(0.455)</td>
<td>(0.485)</td>
<td>(0.142)</td>
</tr>
<tr>
<td></td>
<td>0.022</td>
<td>0.045</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.073)</td>
<td>(0.032)</td>
</tr>
</tbody>
</table>

Note: The figures are the estimated effects and the standard errors (in parentheses). The significance level is based on the $t$-statistic version of $W$. Significance: ***: 1% level, **: 5% level; *: 10% level. The standard errors are bootstrapped from 1,000 re-samples.
<table>
<thead>
<tr>
<th>Number</th>
<th>Author(s)</th>
<th>Title</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-A006</td>
<td>Yi-Ting Chen, Yu-Chin Hsu, Hung-Jen Wang</td>
<td>A Stochastic Frontier Model with Endogenous Treatment and Status and Mediator</td>
<td>12/14</td>
</tr>
<tr>
<td>14-A005</td>
<td>黃登興</td>
<td>台灣薪資停滯現象解析—全球化貿易與投資夥伴</td>
<td>03/14</td>
</tr>
<tr>
<td>14-A004</td>
<td>Wei-Ming Lee, Yu-Chin Hsu, Chung-Ming Kuan</td>
<td>Robust Hypothesis Tests for M-Estimators with Possibly Non-differentiable Estimating Functions</td>
<td>03/14</td>
</tr>
<tr>
<td>14-A003</td>
<td>Shin-Kun Peng, Raymond Riezman, Ping Wang</td>
<td>Intermediate Goods Trade, Technology Choice and Productivity</td>
<td>02/14</td>
</tr>
<tr>
<td>14-A002</td>
<td>Wei-Ming Lee, Chung-Ming Kuan, Yu-Chin Hsu</td>
<td>Testing Over - Identifying Restrictions without Consistent Estimation of the Asymptotic Covariance Matrix</td>
<td>02/14</td>
</tr>
<tr>
<td>14-A001</td>
<td>Stephen G. Donald, Yu-Chin Hsu, Robert P. Lieli</td>
<td>Inverse Probability Weighted Estimation of Local Average Treatment Effects: Higher Order MSE Expansion</td>
<td>02/14</td>
</tr>
<tr>
<td>13-A011</td>
<td>Been-Lon Chen, Shian-Yu Liao</td>
<td>Capital, Credit Constraints and the Comovement between Consumer Durables and Nondurables</td>
<td>10/13</td>
</tr>
<tr>
<td>13-A009</td>
<td>Chien-Yu Huang, Juin-Jen Chang, Lei Ji</td>
<td>Cash-In-Advance Constraint on R&amp;D in a Schumpeterian Growth Model with an Endogenous Market Structure</td>
<td>09/13</td>
</tr>
<tr>
<td>13-A008</td>
<td>Been-Lon Chen, Yu-Shan Hsu, Kazuo Mino</td>
<td>Welfare Implications and Equilibrium Indeterminacy in a Two-sector Growth Model with Consumption Externalities</td>
<td>08/13</td>
</tr>
<tr>
<td>13-A007</td>
<td>Hui-ting Hsieh, Ching-chong Lai</td>
<td>A Macroeconomic Model of Imperfect Competition with Patent Licensing</td>
<td>08/13</td>
</tr>
<tr>
<td>13-A006</td>
<td>Yukihiro Nishimura</td>
<td>Emergence of Asymmetric Solutions in the Abatement Game</td>
<td>07/13</td>
</tr>
<tr>
<td>Number</td>
<td>Author(s)</td>
<td>Title</td>
<td>Date</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>13-A005</td>
<td>Chia-Hui Lu, Been-Lon Chen</td>
<td>Optimal Capital Taxation in A Neoclassical Growth Model</td>
<td>05/13</td>
</tr>
<tr>
<td>13-A004</td>
<td>Yu-Chin Hsu, Xiaoxia Shi</td>
<td>Model Selection Tests for Conditional Moment Inequality Models</td>
<td>04/13</td>
</tr>
<tr>
<td>13-A003</td>
<td>Yu-Chin Hsu</td>
<td>Consistent Tests for Conditional Treatment Effects</td>
<td>04/13</td>
</tr>
<tr>
<td>13-A002</td>
<td>Christine Amsler, Peter Schmidt, Wen-Jen Tsay</td>
<td>A Post-Truncation Parameterization of Truncated Normal Technical Inefficiency</td>
<td>04/13</td>
</tr>
<tr>
<td>13-A001</td>
<td>Yu-Chin Hsu, Chung-Ming Kuan, Meng-Feng Yen</td>
<td>A Generalized Stepwise Procedure with Improved Power for Multiple Inequalities Testing</td>
<td>01/13</td>
</tr>
<tr>
<td>12-A017</td>
<td>Stephen G. Donald, Yu-Chin Hsu, Robert P. Lieli</td>
<td>Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT</td>
<td>12/12</td>
</tr>
<tr>
<td>12-A016</td>
<td>Stephen G. Donald, Yu-Chin Hsu</td>
<td>Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models</td>
<td>12/12</td>
</tr>
<tr>
<td>12-A015</td>
<td>Stephen G. Donald, Yu-Chin Hsu</td>
<td>Improving the Power of Tests of Stochastic Dominance</td>
<td>12/12</td>
</tr>
<tr>
<td>12-A014</td>
<td>Costas Azariadis, Been-Lon Chen, Chia-Hui Lu, Yin-Chi Wang</td>
<td>A Two-sector Model of Endogenous Growth with Leisure Externalities</td>
<td>12/12</td>
</tr>
<tr>
<td>12-A012</td>
<td>江豐富</td>
<td>台灣公私部門薪資差異問題之研究</td>
<td>10/12</td>
</tr>
<tr>
<td>12-A011</td>
<td>林忠正</td>
<td>市場經濟—紙市單車票：需求面模型</td>
<td>10/12</td>
</tr>
<tr>
<td>12-A010</td>
<td>林忠正</td>
<td>為何店門前顧客排隊的餐廳不漲價？</td>
<td>10/12</td>
</tr>
<tr>
<td>12-A009</td>
<td>林忠正</td>
<td>「京華風雲」—開幕促銷</td>
<td>10/12</td>
</tr>
<tr>
<td>12-A008</td>
<td>林忠正</td>
<td>初探總額預算下同業制約精神對醫療衡量行為的影響</td>
<td>09/12</td>
</tr>
<tr>
<td>12-A007</td>
<td>林忠正</td>
<td>為什麼您不補習很難？專業化與市場化的補習勞務</td>
<td>09/12</td>
</tr>
<tr>
<td>12-A005</td>
<td>黃登興, 楊子菡, 孫英智</td>
<td>全球大型企業的經營特徵 2005～2011：產業，國籍，與中國企業之優勢與劣勢</td>
<td>06/12</td>
</tr>
</tbody>
</table>