Semi-Instrumental Variables: A Test for Instrument Admissibility

Tianjiao Chu
tchu@andrew.cmu.edu

Richard Scheines
Philosophy Department
Carnegie Mellon University
Pittsburgh, PA 15213
scheines@cmu.edu

Peter Spirtes
ps7z@andrew.cmu.edu

Abstract

In a causal graphical model, an instrument for a variable X and its effect Y is a random variable that is a cause of X and independent of all the causes of Y except X. (Pearl (1995), Spirtes et al (2000)). Instrumental variables can be used to estimate how the distribution of an effect will respond to a manipulation of its causes, even in the presence of unmeasured common causes (confounders). In typical instrumental variable estimation, instruments are chosen based on domain knowledge. There is currently no statistical test for validating a variable as an instrument. In this paper, we introduce the concept of semi-instrument, which generalizes the concept of instrument. We show that in the framework of additive models, under certain conditions, we can test whether a variable is semi-instrumental. Moreover, adding some distribution assumptions, we can test whether two semi-instruments are instrumental. We give algorithms to estimate the p-value that a random variable is semi-instrumental, and the p-value that two semi-instruments are both instrumental. These algorithms can be used to test the experts' choice of instruments, or to identify instruments automatically.

Key Words: Causality, Instrumental Variables

1 INTRODUCTION

One of the big advantages of knowing the causal relation among the variables we are interested in is that we can predict the effect of intervention. For example, if we know that a random variable X is the cause of Y, then intervening X will change the value of Y in a certain way, but intervening Y should not change the value of X.[1] By regressing Y on X, we can estimate the effect of the intervening of X on Y, up to a constant.

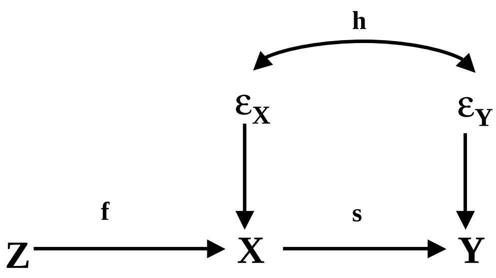

Figure 1: One Instrumental Variable

However, this advantage becomes problematic if we know, or even suspect, that there are some unmeasured common causes of X and Y, in which case we cannot estimate the direct effect of X on Y.

Consider the the causal structure illustrated in figure 1. Suppose that Z is not observable for the moment. Among X, Y, ϵ_X, and ϵ_Y, only X and Y are observed variables. The functional relations among them are given as:

$$X = f(Z) + \epsilon_X \tag{1}$$

[1] Here by intervention we mean the manipulation of the value of one random variable, with the assumption that this manipulation can affect other variables only through the manipulated variable. We also assume no feedback, i.e., no directed cycles in a causal graph.

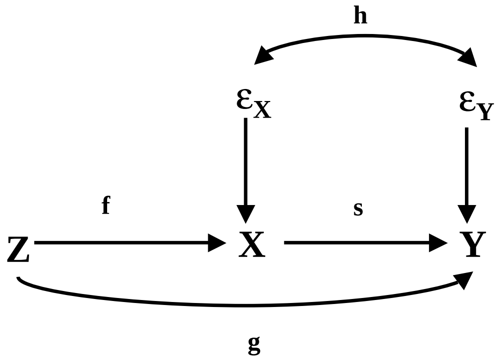

Figure 2: One Common Cause

$$Y = s(X) + \epsilon_Y \quad (2)$$
$$\mathrm{E}[\epsilon_Y|\epsilon_X] = h(\epsilon_X) \quad (3)$$

Note that ϵ_Y and X are dependent, that is, $\mathrm{E}[\epsilon_Y|X]$ will be a non-constant function of X, say, $\lambda(X)$. The regression of Y on X will give:

$\mathrm{E}[Y|X] = s(X) + \mathrm{E}[\epsilon_Y|X] = s(X) + \lambda(X)$

Because we have no way to estimate $\lambda(X)$, we also have no way to identify $s(X)$, even up to a constant.

However, with the help of variable Z, which is an instrumental variable for X and Y, we can estimate ϵ_X by regress X on Z to get an estimate of $\epsilon_X = Z - \mathrm{E}[X|Z]$. Then, we can regress Y on X and ϵ_X to get:

$\mathrm{E}[Y|X, \epsilon_X] = s(X) + \mathrm{E}[\epsilon_Y|\epsilon_X, X]$

$= s(X) + \mathrm{E}[\epsilon_Y|\epsilon_X] = s(X) + h(\epsilon_X)$.

An additive regression method will give the estimate of $s(X)$ and $h(\epsilon_X)$ simultaneously.[2]

2 PRIOR WORK

There have been intensive studies of the use of instruments in econometric literature. Much work has been focusing on the study of the efficiency of an instrument (see Nelson et al (1990), Shea (1997), and Zivot et al (1998).). There are studies of whether a variable, say, Z_1, is an instrument for X and Y, provided we already have an instrument, say, Z_2, for X and Y (see Wu (1973), Newey (1985), and Magdalinos (1994)), but how to find the first instrument is still an open problem. There is no test available to find out whether a variable is an instrument, without knowing that some other variable is an instrument. [3]

In practice, people usually resort to domain knowledge to determine whether a variable is an instrument. However, in the context of automated knowledge discovery, it is desirable to have an algorithm that can determine whether a variable is an instrument automatically. Moreover, as some studies show, it is often very difficult to find instruments, and a wrong choice may significantly affect the final result (see e.g. Bartels (1991), Heckman (1995), and Bound et al (1995).). Therefore, even in the case where we do have some domain knowledge to help us identify instruments, some kind of testing procedure is still useful in that it can serve as a double check. It is the goal of this paper to find out whether, under certain reasonable conditions, we can test whether a variable is an instrument.

3 APPROACH

Within the framework of linear models, the fact that a variable is an instrument imposes no constraint on the joint marginal distribution of the observed variables. Consider the causal structure illustrated in figure 2, where only Z, X, and Y are observable. Assume for now that all the functional relations are linear. In particular, we assume that:

$$X = fZ + \epsilon_X \quad (4)$$
$$Y = gZ + sX + \epsilon_Y \quad (5)$$

where f, g, and s are non-zero constants.

Now let $g' = 0$, $s' = \frac{g}{f} + s$, and $\epsilon'_Y = \epsilon_Y + (s - s')\epsilon_X$. Let $Z' = Z$, $X' = X$, and $Y' = s'X' + \epsilon'_Y$. Clearly $Y' = Y$, hence (Z, X, Y) has the same distribution as (Z', X', Y'). However, we notice that in the first case, Z is an instrument for X and Y, while in the second case, Z' is not an instrument, but a common cause of X' and Y'. Actually, we can set g' to be any value, and adjust s' and ϵ'_Y accordingly so that $X' = X$ and $Y' = Y$. This means that we cannot identify the value of g, and hence cannot tell whether Z is an instrument.

[2]For the proof of the identifiability of $s(X)$, up to a constant, please see Newey et al (1999).

[3]Pearl (1995) shows that, for discrete data, Z being an instrument for X and Y imposes some constraint, which is called by him as *instrumental inequality*, on the joint distribution of Z, X, and Y. This inequality is not sufficient for Z to be an instrument, and it cannot be extended to the continuous models. The paper does not give an instrument testing procedure for the discrete data.

The above analysis suggests two possible approaches for instrument testing. First, we may extend the space of structural models, and see whether the instrument assumption does impose some constraint in this larger space. For example, we can extend the model space to the set of additive nonparametric models. Another approach is to get more candidates for instrumental variables, and see whether the instrument assumption imposes some constraint on the joint distribution of these candidate variables. In this section we shall try a combination of both approaches.

3.1 SEMI-INSTRUMENTAL VARIABLES FOR THE ADDITIVE NONPARAMETRIC MODELS

Here by an additive nonparametric model we mean that, for each endogenous variable, its expectation given its parents is a linear combination of univariate functions of its parents, plus some error term with unknown distribution. We further assume that all exogenous variables are independent, and that they are independent of all the error terms. But we do allow dependence between two error terms. Later in this paper we shall use additive model and additive nonparametric model interchangeably.

Figure 1 and figure 2 give two very simple additive models. We can take them as two alternative hypotheses about the underlying model that generates a sample with observable variables Z, X, and Y. The problem of testing whether Z is an instrument for X and Y is equivalent to the problem of determining which figure represents the correct model.

Note that a linear model is a special case of the additive model. Thus, given that we cannot determine whether a random variable is an instrument in a linear model, we could not, in general, determine whether it is an instrument in an additive model. Therefore, we should look for some other conditions, preferably a little bit weaker than those required by an instrument, so that a random variable satisfies these conditions can be identified in an additive model.

One such candidate set of conditions is what we call the semi-instrumental conditions. A random variable Z is a *semi-instrument* for X and Y if the following conditions are satisfied:

1. *The joint distribution of Z, X, and Y can be represented by an additive model that preserves all the causal relations among Z, X, and Y;*

2. *In that model, Z, X, and Y are the only three observed variables;*

3. *In that model, Z is an exogenous variable;*

4. *In that model, Z is the cause of X, and X is a cause of Y;*

5. *In that model, if Z is also a cause of Y, then the direct effect of Z on Y is a linear function of the direct effect of Z on X.*

It is easy to see that the semi-instrumental assumption is weaker than the instrumental assumption: All instruments are semi-instruments, but not all semi-instruments are instruments. For example, in figure 3, Z is a semi-instrument, but not an instrument. Moreover, in general, in a linear model, an exogenous Z that is a common cause of X and Y is a semi-instrument for X and Y, but not an instrument.

To test whether a random variable Z is a semi-instrument for X and Y, theoretically we should check whether all the five conditions are satisfied. However, it turns out that not all these five conditions are testable. For example, the third condition, i.e., Z is an exogenous variable, cannot be tested if we only have the joint distribution of X, Y, and Z. From the joint distribution only, there is no way to tell whether Z and the error term associated with X, say, ϵ_X, are dependent.

On the other hand, we do have many cases where it is reasonable to assume that the first four conditions are satisfied. For example, it is typical that when testing whether Z is an instrument for X and Y, we are dealing with an additive model described by conditions 1 to 4, and our question is whether Z is also a direct cause of Y.

Assuming that the first four conditions are satisfied, under certain smoothness conditions, to test whether the fifth condition is also satisfied is equivalent to testing whether $\mathrm{E}[Y|X, \epsilon_X]$ is a linear combination of a univariate function of X and a univariate function of ϵ_X:

Theorem 1 *Consider the additive model given by figure 2. Suppose that (X, ϵ_X) has a joint density, and that all the functions, i.e., f, g, s, and h, are differentiable.* [4] *Then $E[Y|X, \epsilon_X]$ is a linear combination of a univariate function of X and a univariate function of ϵ_X if and only if $g(Z) = af(Z) + b$ for some pair of real numbers (a, b).*

For the proof, please see the appendix.

[4] These two conditions are much stronger than what we need. Actually, we only need to assume the boundary of the support of the (X, ϵ_X) has no positive probability, (see proof of Theorem 2.3 in Newey et al (1999)), and that all the functions are absolutely continuous. We choose these two conditions because they are more intuitive.

The above theorem suggests an algorit
will call the *semi-instrument testing alg*
whether Z is a semi-instrument for a sa
ated from an additive model given by fi

1. Regress X on Z to estimate $\epsilon_X =$
2. Regress Y on X and Z with a surf
let $\hat{Y}_s$ and $MSE_s(S)$ be the estima
the mean square error respectively.
X and ϵ_X with an additive smooth
$MSE_a(S)$ be the estimation of Y
square error respectively.
3. Generate N bootstrap samples
adding a permutation of $MSE_s(S)$
4. Apply the first two steps to the N
estimate the distribution of MSE_a
5. $P(MSE_a - MSE_s \leq MSE_a(S) -$
the p-value that Z is a semi-instrum
for X and Y.

Note that if we know that Z is a semi-instrument, i.e., $\mathrm{E}[Y|X, \epsilon_X]$ is additive in X and ϵ_X, we could estimate the second order derivative of the direct effect of X on Y, or the first order derivative up to a constant. This is not interesting in a linear model, because we know that the second order derivative must be 0, and the first order derivate must be a constant. However, it does provide us with some insight in an additive model. The basic algorithm for computing the first order derivative, up to a constant, is:

1. Regress X on Z to estimate $\epsilon_X = X - \mathrm{E}[X|Z]$
2. Estimate the partial derivative of $\mathrm{E}[Y|X, \epsilon_X]$ with respect to ϵ_X by nonparametric regression.

The above algorithm is based on the fact that, assuming the model specified by equations 4 and 5, and figure 2, assuming further that Z is a semi-instrument, i.e., $g(Z) = af(Z) + b$, then:

$$\frac{\partial}{\partial X}\mathrm{E}[Y|X, \epsilon_X] = s'(X) + a$$

3.2 TWO SEMI-INSTRUMENTS

If the test for whether a random variable, say, Z_1, is a semi-instrument for X and Y gives a negative result, there is not much left to do with Z_1. However, if the test says that Z_1 is a semi-instrument, we will face another problem: Is Z_1 an instrument?

We have pointed out that this question cannot be answered if we only have the joint distribution of Z_1, X, and Y. However, it turns out, with some further assumption, that if there is a second semi-instrument, say, Z_2, we might be able to determine whether Z_1 and Z_2 are both instruments. (If not both of them are instrument, we would not be able to tell whether both are non-instrumental, or only one is non-instrumental.)

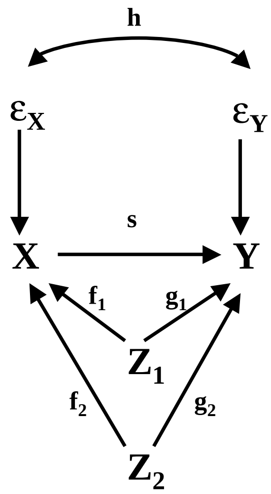

Figure 3: Two Common Causes

Suppose a random variable Z is a semi-instrument for X and Y, then the direct effect of Z on Y, say, $g(Z)$, is a linear function of the direct effect of Z on X, say, $f(Z)$. That is, there is a pair of real numbers (a, b) such that $g(Z) = af(Z) + b$. We shall call a the *linear coefficient of the semi-instrument Z*. The following theorem gives the condition when two semi-instruments are both instruments almost surely.

Theorem 2 *Let Z_1 and Z_2 be two independent random variables that are both semi-instruments for X and Y. Suppose the linear coefficients of Z_1 and Z_2 are independent, and each has a distribution function that has one and only one point of discontinuity 0. If Z_1 and Z_2 have the same linear coefficients, then with probability 1, Z_1 and Z_2 are both instruments.*

For the proof, please see the appendix.

Assume that the sample S was generated from the causal structure illustrated in figure 3, and that both Z_1 and Z_2 are semi-instruments for X and Y. We can use the following algorithm, called the *double instruments testing algorithm*, to test whether Z_1 and Z_2 have the same linear coefficient:

1. Create a new variable $Z = f_1(Z_1) + f_2(Z_2)$, where $f_1(Z_1) = E[X|Z_1]$, $f_2(Z_2) = \mathrm{E}[X|Z_2]$.

2. Estimate the p-value that Z is a semi-instrument.

This algorithm is based on the following observation:

Assuming that g_1 and g_2 are differentiable, and that Z_1 and Z_2 have a joint density. Then $g_1(Z_1)+g_2(Z_2) = a(f_1(Z_1) + f_2(Z_2)) + b$ for some (a, b) iff there are (b_1, b_2) such that $g_1(Z_1) = af_1(Z_1) + b_1$ and $g_2(Z_2) = af_2(Z_2) + b_2$. [5]

4 SIMULATION

We have done some simulation studies to estimate the performance of the two algorithms for semi-instrument testing and double instrument testing. It turns out that we need to find some better regression methods to make these tests powerful. The biggest problem is that the additive regression methods available to us are not suitable when the predictors are dependent. Let S be a sample generated from the additive model given by figure 2. To test whether $g(Z) = af(Z)+b$ for some (a, b), we need to compare $MSE_a(S)$, the mean square error of using additive regression method to regress Y on X and ϵ_X, with $MSE_s(S)$, the mean square error of using surface smoother method to regress Y on X and ϵ_X. In a simulation study, we use the Splus function `gam` as the additive regression method. It turns out that because of the dependence between ϵ_X and X, the performance of `gam` is significantly worse than that of the Splus function `loess`, a surface smoother we used in our simulation study. [6] That is, the test statistic, i.e., $MSE_a(S) - MSE_s(S)$, will be very large. What is worse is that, because we use $\hat{Y}_a$, the estimation of Y obtained by `gam`, to create bootstrap samples $S_1, \cdots, S_N$, the performance of `gam` will not be too bad compared to that of `loess` for these bootstrap samples. That is, $MSE_a(S_i) - MSE_s(S_i)$ for $i = 1, \cdots, N$ will be much smaller than $MSE_a(S) - MSE_s(S)$. This implies that no matter whether the null hypothesis is true, i.e., Z is a semi-instrument, the p-value of test statistic always be extremely small!

We note that Gozalo et al (2001) gives some algorithms for nonparametric additivity testing. However, it turns out that their algorithms also require the predictors being independent. [7] Of course, in a linear model, if the predictors are dependent, we can use ridge regression to get a better estimation (see Fennessey et al (1980)). However, it is not clear whether a similar solution exists for nonparametric additive regression.

Despite the lack of good additive regression method, with some further condition, we could still make the double instruments testing algorithm work. Consider the model given by figure 3. If besides assuming Z_1 and Z_2 are semi-instruments, we further assume that $s(X) = cX + d$, i.e., the direct effect of X on Y is a linear function in X, then we will be able to test whether $Z = f_1(Z_1) + f_2(Z_2)$ is a semi-instrument for X and Y.

Note that because $s(X)$ is linear, we have:

$$\begin{aligned}&\mathrm{E}[Y|Z_1, Z_2, \epsilon_X]\\ &= \mathrm{E}[cX + d + \epsilon_Y + g_1(Z_1) + g_2(Z_2)|Z_1, Z_2, \epsilon_X]\\ &= cf_1(Z_1)+g_1(Z_1)+cf_2(Z_2)+g_2(Z_2)+c\,\epsilon_X+h(\epsilon_X)+d.\end{aligned}$$

That is, $\mathrm{E}[Y|Z_1, Z_2, \epsilon_X]$ is additive in Z_1, Z_2, and ϵ_X, where Z_1, Z_2, and ϵ_X are independent.

Let $\mathrm{E}[g_1(Z_1) + g_2(Z_2)|f_1(Z_1) + f_2(Z_2) = Z] = t(Z)$. We also have:

$$\begin{aligned}&\mathrm{E}[Y|Z, \epsilon_X] = \mathrm{E}[cX + d|Z, \epsilon_X] + \mathrm{E}[\epsilon_Y|Z, \epsilon_X] + t(Z)\\ &= cZ + t(Z) + c\,\epsilon_X + h(\epsilon_X) + d.\end{aligned}$$

That is, $\mathrm{E}[Y|Z, \epsilon_X]$ is additive in Z and ϵ_X, where Z and ϵ_X are independent.

Now if Z is a semi-instrument, i.e., $g_1(Z_1) + g_2(Z_2) = aZ + b$, we will have:

$$\mathrm{E}[Y|Z, \epsilon_X] = (c + a)Z + c\,\epsilon_X + h(\epsilon_X) + b + d.$$

Therefore, to test whether Z is a semi-instrument is equivalent to test whether Y is linear in Z!

The following algorithm, which is a modified double instruments testing algorithm, can be used to test whether two semi-instruments Z_1 and Z_2 have the same linear coefficients, assuming the direct effect of X on Y is linear in X:

1. Use additive regression to regress X on Z_1 and Z_2 to get $\mathrm{E}[X|Z_1, Z_2]$.

2. Let $Z = \mathrm{E}[X|Z_1, Z_2]$. Regress Y on Z using linear regression, let $\hat{Y}_l$ be the estimation of Y.

3. Regress Y on Z_1 and Z_2 using additive regression,

[5] The proof of this observation is similar to that of Theorem 1.

[6] The `gam` function implements the back-fitting algorithm proposed in Hastie et al (1990), which is not very stable when the predictors are dependent.

[7] To illustrate their approach, consider a model where $Y = f(X_1, X_2) + \epsilon$. To test whether $f(X_1, X_2) = f_1(X_1) + f_2(X_2)$ for some f_1 and f_2, They use the difference between $\mathrm{E}[Y|X_1, X_2]$ and $\mathrm{E}[Y|X_1] + \mathrm{E}[Y|X_2]$ as the testing statistic, and reject the additivity assumption if these two estimations differ significantly. It is easy to see that in order for $\mathrm{E}[Y|X_1] + \mathrm{E}[Y|X_2]$ to be a consistent estimator of $\mathrm{E}[Y|X_1, X_2]$, we need not only the additive assumption, but also the assumption that $\mathrm{E}[f_i(X_i)|X_j] = 0$, where $i, j \in \{1, 2\}, i \neq j$.

let $\hat{Y}_a$ be the estimation of Y.

4. The test statistic is the sum of square of $\hat{Y}_a - \hat{Y}_l$. If its value is small, then Z_1 and Z_2 have the same linear coefficient.

To test the above algorithm, we generate two samples in the following way:

For sample 1, we have:

$X = Z_1^2 + Z_2^2 + \epsilon_X$

$Y = X + \epsilon_Y$

where $\mathrm{E}[\epsilon_Y | \epsilon_X] = \epsilon_X^2$.

For sample 2, we use the same Z_1, Z_2, ϵ_X, X, and ϵ_Y, but we let $Y = X + Z_2^2 + \epsilon_Y$. [8]

Using the above algorithm, for a sample size of 200, we the sum of square of $\hat{Y}_a - \hat{Y}_l$ is only 1.695891 for the first sample, and 791.5779 for the second sample. This suggests that our algorithm is very powerful.

5 DISCUSSION AND FUTURE WORK

5.1 THREE ASSUMPTIONS

The semi-instrument testing algorithm assumes that the first four semi-instrumental conditions are satisfied. While in general we cannot test whether a random variable satisfies all the first four semi-instrumental conditions, it is interesting to know whether we can test for one of them, especially the third semi-instrumental condition: Z is an exogenous variable. The answer is: in principle, this assumption can be tested by the method of instrument, if we have an instrument for Z and X. However, it is easy to see that this will lead to an infinite regression.

Another assumption key to the double instruments testing algorithm is: The prior probability that Z is instrumental is positive, while the prior probability is 0 for a semi-instrument to have linear coefficient a if $a \neq 0$. This raises a question: Why does the value 0 have a special status in the range of possible values of the linear coefficients of a semi-instrument? Here we want to give an argument for the plausibility of this assumption: If we take the set of possible causal structures among X, Y, Z_1 and Z_2 as a discrete sample space, it is reasonable to assign a positive prior probability to one element in the space, i.e., the structure where both Z_1 and Z_2 are instruments, which means that both Z_1 and Z_2 have linear coefficients 0. On the other hand, if a semi-instrument is not an instrument, there is no specific reason to believe that its linear coefficient should take any specific non-zero value.

We make a third assumption in an effort to modify the double instruments testing algorithm so that it has sufficient power: We assume that the direct effect of X on Y is a linear function of X. We notice that this is a rather strong assumption for an additive model. Moreover, because currently we do not have a suitable additive regression method for the semi-instrument testing, we also have to assume, without any testing, that Z_1 and Z_2 are semi-instruments. Nevertheless, the modified double instruments testing algorithm is still useful: It provides a double instruments test for linear models.

5.2 FUTURE WORK

To make the semi-instrument testing powerful, we will continue to look for some additive regression method that is suitable for the case where the predictors are independent. Alternative, we may also try to find some new way of generating bootstrap samples so that the performance of currently available regression methods will not get improved for the bootstrap samples. We will also need to get the asymptotic distribution of the test statistic for the modified double instruments testing.

[8] The distribution of these variables are: Z_1 and Z_2 both are uniform between 0 and 4. There is also a latent variable T that is uniform between 0 and 2. ϵ_X is the sum of T and a normal noise with variance 0.25, ϵ_Y is the sum of T^2 and a normal noise with variance 0.25.

Reference

[1] Bartels, L. (1991), "Instrumental and 'Quasi-Instrumental' Variables," in *American Journal of Political Science, 35*, pp. 777-800.

[2] Bound, J., Jaeger, D., Baker, R. (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," in *Journal of the American Statistical Association, Vol. 90*, pp. 443-450.

[3] Fennessey, J., d'Amico, R. (1980), "Collinearity, Ridge Regression, and Investigator Judgment," in *Sociological Methods and Research 8*, pp. 309-340.

[4] Gozalo, P., Linton, O. (2001) "Testing Additivity in Generalized Nonparametric Regression Models with Estimated Parameters," forthcoming in *Journal of Econometrics.*

[5] Hastie, T., Tibshirani, R. (1990), *Generalized Additive Models.* New York : Chapman and Hall.

[6] Heckman, J. (1995), "Instrumental Variables: A Cautionary Tale," *NBER Technical Working Paper No. 185.*

[7] Magdalinos, M. (1994), "Testing Instrument Admissibility: Some Refined Asymptotic Results", in *Econometrica Vol. 62*, pp. 373-403

[8] Nelson, C., Startz, R. (1990), "The Distribution of the Instrumental Variables Estimator and Its t-Ratio when the Instrument is a Poor One," in *Journal of Business, vol. 63*, S125-S140.

[9] Newey, W. (1985), "Generalized Method of Moments Specification Testing," in *Journal of Econometrics, Vol. 29*, pp.229-256.

[10] Newey, W., Powell, J., Vella, F. (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models," in *Econometrica Vol. 67*, pp.565-603.

[11] Pearl, J. (1995) "On the Testability of Causal Models with Latent and Instrumental Variables," in P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence 11*, pp.435-443. San Francisco, CA: Morgan Kaufmann

[12] Shea, J. (1997), "Instrument Relevance in Multivariate Linear Models: A Simple Measure," in *Review of Economics and Statistics Vol. 79*, pp: 348-352.

[13] Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd ed. New York, N.Y.: MIT Press.

[14] Wu, D. (1973), "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," in *Econometrica Vol. 41*, pp.733-750.

[15] Zivot, E., Startz, R., Nelson, C., (1998), "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," in *International Economic Review, Vol. 39*, pp. 1119-44.

Appendix

Proof for Theorem 1.

Because (X, ϵ_X) has a joint density, and that f, g, s, and h are differentiable, it immediately follows that $\mathrm{E}[Y|X = x, \epsilon_X = u]$ is differentiable with respect to X and ϵ_X with probability one.

Also note that $\mathrm{E}[Y|X = x, \epsilon_X = u]$

$= \mathrm{E}[s(X) + g(Z)) + \epsilon_Y|X = x, \epsilon_X = u]$

$= s(x) + \mathrm{E}[g(Z)|f(Z) = x - u] + h(u)$.

It is easy to see that if $g(Z) = af(Z)+b$, i.e., the direct effect of Z on Y is a linear function of the direct effect of Z on X, then with probability one, $\mathrm{E}[Y|X, \epsilon_X] = s(X) + aX + h(\epsilon_X) - a\epsilon_X + b$, i.e., $\mathrm{E}[Y|X, \epsilon_X]$ is a linear combination of a univariate function of X and a univariate function of ϵ_X.

To show the converse, suppose:

$\mathrm{E}[Y|X = x, \epsilon_X = u] = s_1(x) + h_1(u)$.

Let $g_0(x - u) = \mathrm{E}[g(Z)|f(Z) = x - u]$, we have:

$$\frac{\partial g_0(x-u)}{\partial x} = g_0'(x-u) = \frac{d}{dx}(s_1(x) - s(x))$$

$\Longrightarrow$ $g_0'(x - u)$ is constant in u

$\Longrightarrow$ g_0' is a constant

$\Longrightarrow$ g_0 is a linear function

$\Longrightarrow$ $g(Z) = af(Z) + b$, where a, b are constants.

Proof for Theorem 2.

Let L_1 be the linear coefficient of Z_1, L_2 the linear coefficient of Z_2, and μ_{L_1} the distribution of L_1. Then:

$$P(L_1 = L_2 = 0|L_1 = L_2) = \frac{P(L_1 = L_2 = 0)}{P(L_1 = L_2)} = 1, \text{ for:}$$

$$P(L_1 = L_2, L_1 \neq 0) = \int_{\mathbb{R}\setminus\{0\}} P(L_2 = l_1) d\mu_{L_1}(l_1) = 0$$