

# Erratum regarding "Instrumental variables with unrestricted heterogeneity and continuous treatment"

Authors: Stefan Hoderlein, Hajo Holzmann, Maximilian Kasy, Alexander Meister

Persistent link: <http://hdl.handle.net/2345/bc-ir:104980>

This work is posted on [eScholarship@BC](#),  
Boston College University Libraries.

---

Boston College Working Papers in Economics, 2016

Originally posted on: <http://ideas.repec.org/p/boc/bocoec/896.html>

# Erratum regarding “Instrumental variables with unrestricted heterogeneity and continuous treatment”

Stefan Hoderlein      Hajo Holzmann      Maximilian Kasy  
Alexander Meister

July 18, 2015

Kasy (2014) considers a triangular system of equations characterized by the following assumptions:

**Assumption 1** (Triangular System).

$$\begin{aligned} Y &= g(X, U) \\ X &= h(Z, V) \end{aligned} \tag{1}$$

where  $X, Y, Z$  are random variables taking their values in  $\mathbb{R}$ , the unobservables  $U, V$  have their support in an arbitrary measurable space of unrestricted dimensionality, and

$$Z \perp (U, V). \tag{2}$$

**Assumption 2** (Continuous treatment).

*The treatment  $X$  is continuously distributed in  $\mathbb{R}$  conditional on  $Z$ .*

**Assumption 3** (First stage monotonic in instrument).

*The first stage relationship  $h(z, v)$  is strictly increasing in  $z$  for all  $v$ .*

**Assumption 4** (Continuous instrument).

*The instrument  $Z$  is continuously distributed in  $\mathbb{R}$ , with support  $[z_l, z_u]$ . The first stage relationship  $h$  is continuous in  $z$  for all  $z$  and almost all  $v$ , and  $P(X \leq x | Z = z)$  is continuous in  $z$  for all  $x$ .*

Under these assumptions, the following definitions are introduced:

**Definition 1** (Potential outcomes).

*We denote by*

$$\begin{aligned} Y^x &= g(x, U) \\ X^z &= h(z, V). \end{aligned} \tag{3}$$

Furthermore, we define

$$Z^x = \begin{cases} h^{-1}(x, V) & \text{if } h(z_l, V) \leq x \leq h(z_u, V) \\ -\infty & \text{if } x < h(z_l, V) \\ \infty & \text{if } h(z_u, V) < x. \end{cases} \quad (4)$$

It is claimed in the statement and proof of Kasy (2014), theorem 1, that under these assumptions

$$\begin{aligned} P(Y \leq y | X = x, Z = z) &= P(Y^x \leq y | X = x, Z = z) \\ &= P(Y^x \leq y | Z^x = z, Z = z) \\ &= P(Y^x \leq y | Z^x = z). \end{aligned}$$

This assertion is wrong, the following theorem 1 states a corrected version. For theorem 1 to hold, we need to additionally impose the following regularity conditions.

**Assumption 5** (Regularity conditions).

There exist  $0 < c_l < c_u < \infty$ , such that  $c_l \leq \partial_z h(z, v) \leq c_u$  for all  $z$  and  $v$ . Further,  $V$  can be decomposed as  $V = (V_1, V_2)$ , where  $V_2$  is scalar and absolutely continuous given  $(Z, V_1)$  with bounded conditional density, and  $\partial_{v_2} h(z, v_1, v_2) \geq c > 0$  for all  $z$  and  $v = (v_1, v_2)$ .

**Theorem 1.** Under assumptions 1 through 5,

$$P(Y \leq y | X = x, Z = z) = E[\lambda \cdot \mathbf{1}(Y^x \leq y) | Z^x = z] \quad (5)$$

where

$$\lambda = \frac{E[\partial_z h(z, V) | X^z = x]}{\partial_z h(z, V)} = \frac{[\partial_z h(z, V)]^{-1}}{E[\partial_z h(z, V)]^{-1} | Z^x = z]}. \quad (6)$$

The proof of theorem 1 and of the following two corollaries can be found in section 2 below.

Theorem 1 immediately implies the following two corollaries. Corollary 1 considers the control function approach of Imbens and Newey (2009), using the control function  $V^* = F(X|Z)$ . Corollary 1 provides a representation as a weighted average for the estimand for the average structural function proposed by Imbens and Newey (2009).

Corollary 2 establishes that the claims of Kasy (2014) *do* hold, once the assumption of one-dimensional first stage heterogeneity  $V$  is imposed.

**Corollary 1.** Let  $V^* = F(X|Z)$ , and assume  $\text{supp}(V^* | X = x) = [0, 1]$ . Then

$$\int_0^1 E[Y | X = x, V^* = v^*] dv^* = \int_{-\infty}^{\infty} E[\lambda \cdot Y^x | Z^x = z] dF_{Z^x}(z).$$

**Corollary 2.** If  $\dim(V) = 1$  (ie.,  $V = V_2$ ) and  $h$  is strictly monotonic in  $V$ , then

$$P(Y \leq y | X = x, Z = z) = E[\mathbf{1}(Y^x \leq y) | Z^x = z]. \quad (7)$$

# 1 Discussion

1. The key step which fails in the original derivation of Kasy (2014) is the asserted equality

$$P(Y^x \leq y | X = x, Z = z) = P(Y^x \leq y | Z^x = z, Z = z).$$

The conditioning event on both sides is the same, that is

$$(X = x, Z = z) = (Z^x = z, Z = z).$$

If this conditioning event had a positive probability, as would be the case for discrete random variables, the asserted equality would indeed hold. As we are dealing with the continuous case, however, this event has probability 0. Conditional expectations (probabilities) given events of probability 0 are only well defined relative to a given  $\sigma$ -algebra. Since the  $\sigma$ -algebra generated by the random variables  $(X, Z)$  and the  $\sigma$ -algebra generated by the random variables  $(Z^x, Z)$  are different, equality of conditional distributions need not hold in general.

2. As implied by corollary 2, the assertions of Kasy (2014) *do* hold under the assumption imposed by Imbens and Newey (2009), that first stage heterogeneity  $V$  is one-dimensional and enters  $h$  monotonically. Despite the failure of its central theorem to hold, Kasy (2014) might thus be thought of as providing alternative estimation procedures, based on reweighting rather than based on regression with controls, which are valid under the assumptions of Imbens and Newey (2009).
3. It was shown in Kasy (2011) that control function approaches such as the one of Imbens and Newey (2009) yield the required conditional independence of potential outcomes and treatment only if first-stage heterogeneity is one-dimensional. Theorem 1 implies that in general we cannot recover the distribution of potential outcomes  $Y^x$  from the observed conditional distributions of  $Y$  given  $X$  and  $Z$ , unless the weights  $\lambda$  are constant and equal to 1. This in turn is only the case if there is no heterogeneity in  $\partial_z h(z, V)$  given  $Z^x$ . This suggests that the argument of Kasy (2011) generalizes, and that reweighting approaches can yield (unconditional) independence of potential outcomes and treatment again only if first-stage heterogeneity is one-dimensional.
4. Based on the result of theorem 1 in this note, it appears that it would be impossible to recover structural objects such as the average structural function without imposing restrictions on heterogeneity. Three general strategies have been pursued in the literature which rely on such restrictions:
  - (a) Restricting heterogeneity  $V$  of the first stage relationship to be one-dimensional, as in Imbens and Newey (2009), Florens et al. (2008), or Hoderlein and Sasaki (2015).

- (b) Restricting heterogeneity  $U$  in the structural equation of interest to be one-dimensional, as in Newey and Powell (2003) or Horowitz (2011).
- (c) Restricting heterogeneity in both equations to be finite dimensional in conjunction with additional structure. A special case is for instance studied in Hoderlein et al. (2015), where the model under consideration takes the form

$$\begin{aligned} Y &= B_0 + B_1X + B_2'W, \\ X &= A_0 + A_1'Z + A_2'W, \end{aligned}$$

where the  $d_{B_2} + d_{A_1} + d_{A_2} + 3$  dimensional vector of unobserved random coefficients  $(A, B)$  is independent of  $d_W + d_Z$  dimensional vector of exogenous variables  $(Z, W)$ , but not of the two endogenous variables  $(Y, X)$ . The random coefficients specification is probably the most generic specification to model complex heterogeneity; compared to the model laid out in assumption 1, this model imposes linearity in random coefficients.

In this class of models, Hoderlein et al. (2015) show formally that nonidentification generically prevails. To make models comparable, in the special case of this model where  $A_2 = B_2 = 0$ ,  $Z$  scalar and  $A_1 > 0$ , Hoderlein et al. (2015) show formally that the distribution of  $Y^x$  is not point identified, and hence additional assumptions are required to achieve point identification. One such assumption is the independence of  $A_1$  from  $B$ , which opens up a way for constructive identification of  $f_B$ , and hence  $f_{Y^x}$ . However, for this argument, monotonicity is not required to hold. Hoderlein et al. (2015) show that this argument generalizes to the case of vector valued  $Z$  and exogenous variables  $W$ , provided that one element of the vector  $A_1$  is independent of  $B$ , which is trivially satisfied, if one element of  $A_1$  is not random. These results seem to suggest that introducing nonlinear terms may be possible, but at the expense of introducing stronger independence conditions, or, perhaps, alternative assumptions like monotonicity.

## 2 Proofs

### Proof of theorem 1:

This proof is structured as follows. We first consider the right hand side of equation (5), and show that for random variables  $\phi$  such that  $(\phi, V) \perp Z$ , we get

$$E[\phi|Z^x = z] = \frac{E[\phi \cdot \partial_z h(z, V)|X^z = x]}{E[\partial_z h(z, V)|X^z = x]}. \quad (8)$$

We then turn to the left hand side, and show, for  $\psi$  such that  $(\psi, V) \perp Z$ ,

$$E[\psi|X = x, Z = z] = E[\psi|X^z = x].$$

The claim of the theorem then follows once we consider  $\psi = \mathbf{1}(Y^x \leq y)$  and  $\phi = \lambda \cdot \mathbf{1}(Y^x \leq y)$ .

Consider some non-negative random variable  $\phi$ , defined on the same probability space as  $V$  and  $Z$ , such that  $(\phi, V) \perp Z$  and  $0 < E[|\phi|] < \infty$ . Since

$$\begin{aligned} \partial_z E[\phi \cdot \mathbf{1}(Z^x \leq z)] &= \partial_z \int_{-\infty}^z E[\phi|Z^x = z'] \cdot f_{Z^x}(z') dz' \\ &= E[\phi|Z^x = z] \cdot f_{Z^x}(z) \end{aligned}$$

and

$$\begin{aligned} \partial_z E[\mathbf{1}(Z^x \leq z)] &= \partial_z \int_{-\infty}^z f_{Z^x}(z') dz' \\ &= f_{Z^x}(z), \end{aligned}$$

we can write

$$E[\phi|Z^x = z] = \frac{\partial_z E[\phi \cdot \mathbf{1}(Z^x \leq z)]}{\partial_z E[\mathbf{1}(Z^x \leq z)]}.$$

Next, note that  $Z^x \leq z$  if and only if  $X^z = h(z, V) \geq x$  (this holds by monotonicity of  $h$ ). We get

$$\begin{aligned} f_{Z^x}(z) &= \partial_z E[\mathbf{1}(Z^x \leq z)] \\ &= \partial_z E[\mathbf{1}(X^z \geq x)] \\ &= -\partial_z F_{X^z}(x) \\ &= E[\partial_z h(z, V)|X^z = x] \cdot f_{X^z}(x). \end{aligned}$$

These equalities hold (i) by definition of the pdf  $f_{Z^x}(z)$ , (ii) by the equality  $\mathbf{1}(Z^x \leq z) = \mathbf{1}(X^z \geq x)$  (due to monotonicity of  $h$ ), and (iii) by equation (D1) in Chernozhukov et al. 2015 (see also Hoderlein and Mammen 2007). This last step requires the regularity conditions of assumption 5.

Now consider the probability measure  $P^\phi$ , defined by  $P^\phi = (\phi/E[\phi]) \cdot P$ , that is the probability measure with density  $(\phi/E[\phi])$  relative to  $P$ , and let  $E^\phi$

be the expectation operator with respect to  $P^\phi$ . Applying the same reasoning as before to this new measure yields

$$\begin{aligned}
\frac{1}{E[\phi]} \cdot \partial_z E[\phi \cdot \mathbf{1}(Z^x \leq z)] &= \partial_z E^\phi[\mathbf{1}(Z^x \leq z)] \\
&= \partial_z E^\phi[\mathbf{1}(X^z \geq x)] \\
&= -\partial_z F_{X^z}^\phi(x) \\
&= E^\phi[\partial_z h(z, V) | X^z = x] \cdot f_{X^z}^\phi(x) \\
&= \frac{1}{E[\phi]} \cdot E[\phi \cdot \partial_z h(z, V) | X^z = x] \cdot f_{X^z}(x).
\end{aligned}$$

The claim of equation (8) follows from what we have shown so far. This proves our first assertion, and also implies the equality of the two definitions of  $\lambda$  (set  $\phi = 1/\partial_z(z, V)$ ) given in the statement of the theorem.

Let us now turn to the left hand side of the equality asserted in the theorem. Consider some random variable  $\psi$ , again defined on the same probability space as  $V$  and  $Z$ , such that  $(\psi, V) \perp Z$  and  $E[|\psi|] < \infty$ . Using statistical independence of  $Z$  and  $(\psi, V)$ , we get

$$\begin{aligned}
E[\psi | X = x, Z = z] &= E[\psi | h(z, V) = x, Z = z] \\
&= E[\psi | h(z, V) = x] = E[\psi | X^z = x].
\end{aligned}$$

Setting  $\psi = \mathbf{1}(Y^x \leq y)$  and  $\phi = \lambda \cdot \mathbf{1}(Y^x \leq y)$  concludes the proof.  $\square$

**Proof of corollary 1:**

This follows immediately from

$$\int_0^1 E[Y | X = x, V^* = v^*] dv^* = \int_{-\infty}^{\infty} E[Y | X = x, Z = z] dF_{Z^x}(z),$$

cf. Kasy (2014), proof of theorem 2.  $\square$

**Proof of corollary 2:** Under this condition,  $V$  is pinned down by  $Z^x = z$ , so that  $\lambda \equiv 1$  given  $Z^x = z$ .  $\square$

## References

- Chernozhukov, V., Fernandez-Val, I., Hoderlein, S., Holzmann, H., and Newey, W. (2015). Nonparametric identification in panels using quantiles. *Journal of Econometrics*, (0):-.
- Florens, J. P., Heckman, J. J., Meghir, C., and Vytlacil, E. (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica*, 76(5):pp. 1191–1206.
- Hoderlein, S., Holzmann, H., and Meister, A. (2015). The triangular model with random coefficients. *Working Paper*.
- Hoderlein, S. and Mammen, E. (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica*, 75(5):1513–1518.
- Hoderlein, S. and Sasaki, Y. (2015). Outcome conditioned treatment effects. Technical report, Discussion paper.
- Horowitz, J. L. (2011). Applied Nonparametric Instrumental Variables Estimation. *Econometrica*, 79(2):347–394.
- Imbens, G. W. and Newey, W. (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica*, 77:1481–1512.
- Kasy, M. (2011). Identification in triangular systems using control functions. *Econometric Theory*, 27(03):663–671.
- Kasy, M. (2014). Instrumental variables with unrestricted heterogeneity and continuous treatment. *Review of Economic Studies*, 81(4).
- Newey, W. K. and Powell, J. L. (2003). Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, 71(5):1565–1578.