

Nonparametric identification of the classical errors-in-variables model without side information.

S. M. Schennach*
Department of Economics
University of Chicago
1126 East 59th Street
Chicago IL 60637
smschenn@uchicago.edu

Yingyao Hu
Department of Economics
Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218
yhu@jhu.edu

A. Lewbel
Department of Economics
Boston College
140 Commonwealth Avenue
Chestnut Hill, MA 02467
lewbel@bc.edu

July 16, 2007

Abstract

This note establishes that the fully nonparametric classical errors-in-variables model is identifiable from data on the regressor and the dependent variable alone, unless the specification is a member of a very specific parametric family. This family includes the linear specification with normally distributed variables as a special case. This result relies on standard primitive regularity conditions taking the form of smoothness and monotonicity of the regression function and nonvanishing characteristic functions of the disturbances.

*Corresponding author. S. M. Schennach acknowledges support from the National Science Foundation via grant SES-0452089.

1 Introduction

The identification of regression models in which both the dependent and independent variables are measured with error has received considerable attention over the last few decades. This so-called classical nonlinear errors-in-variables model takes the following form.

Model 1 *Let $y, x, x^*, \Delta x, \Delta y$ be scalar real-valued random variables such that*

$$\begin{aligned}y &= g(x^*) + \Delta y \\x &= x^* + \Delta x.\end{aligned}$$

where only x and y are observed while all remaining variables are not and where $x^, \Delta x, \Delta y$, are mutually independent, $E[\Delta x] = 0$ and $E[\Delta y] = 0$.*

A well-known result is that when $g(x^*)$ is linear while $x^*, \Delta x$ and Δy are normal, the model is not identified, although the regression coefficients can often be consistently bounded (Klepper and Leamer (1984)).¹ This negative result for what is perhaps the most natural regression model has long guided the search for solutions to the errors-in-variables problem towards approaches that rely on additional information (beyond x and y), such as instruments, repeated measurements, validation data, known measurement error distribution, etc (e.g., Hausman, Newey, Ichimura, and Powell (1991), Newey (2001), Schennach (2004a), Schennach (2004b), Schennach (2007), Hu and Schennach (2006), Hu and Ridder (2004), among many others).

Nevertheless, since the seminal work of Geary (1942), a large number of authors (e.g. Reiersol (1950), Kendall and Stuart (1979), Pal (1980), Cragg (1997), Lewbel (1997), Erickson and Whited (2002), Dagenais and Dagenais (1997), Erickson and Whited (2000), Bonhomme and Robin (2006), and the many references therein) have suggested alternative methods to identify a linear regression with nonnormally distributed regressors based on the idea that higher order moments of x and y then provide additional information that can be exploited. However, the question of characterizing the set of identifiable models in fully nonparametric settings while exploiting the joint distribution of all the observable variables remains wide open.

¹Chesher (1998) suggests some settings where a polynomial regression is not identified based on the knowledge of *some* of the moments of the observed data.

We demonstrate that the answer to this question turns out to be surprisingly simple, although proving so is not. Under fairly simple and natural regularity conditions, a specification of the form $g(x^*) = a + b \ln(e^{cx^*} + d)$ is the *only* functional form that is not guaranteed to be identifiable. Even with this specification, the distributions of all the variables must have very specific forms in order to evade identifiability of the model. As expected, this parametric family includes the well-known linear case (with $d = 0$) with normally distributed variables. Given that this very specific unidentified parametric functional form is arguably the exception rather than the rule, our identification result should have a wide applicability.

2 Identification result

We need a few basic regularity conditions.

Assumption 1 $E[e^{i\xi\Delta x}]$ and $E[e^{i\gamma\Delta y}]$ do not vanish for any $\xi, \gamma \in \mathbb{R}$, where $i = \sqrt{-1}$.

The type of assumption regarding the so-called characteristic function has a long history in the deconvolution literature (see Schennach (2004a) and the references therein). Without it, the measurement error effectively masks information regarding the true variables that cannot be recovered.² The only commonly encountered distributions with a vanishing characteristic function are the uniform and the triangular distributions.

Assumption 2 *The distribution of x^* admits a finite density $f_{x^*}(x^*)$ with respect to the Lebesgue measure.*

This assumption rules out pathological case such as fractal-like distributions. It also rules out discrete distributions.³

Assumption 3 *The regression function $g(x^*)$ has a continuous, finite and nonvanishing first derivative at each point⁴ in the interior of the support of x^* .*

²Although our approach could probably be extended to the case of characteristic functions vanishing at isolated points in \mathbb{R} along the lines of Hu and Ridder (2004).

³An extension of our result to purely discrete distributions is straightforward, although such a result would not be very useful in the context of classical measurement error.

⁴It need not be *uniformly* bounded above and below.

This is a smoothness and monotonicity constraint. Without it, it is difficult to rule out extremely complex and pathological joint distributions of x and y (including, once again, fractal-like distributions). In particular, one could imagine an extremely rapidly oscillating $g(x^*)$, where nearly undetectable changes in x^* yield changes in y that are almost observationally indistinguishable from genuine errors in y . Even relaxing this assumption to include less pathological functional forms that oscillate a finite number of times is difficult, due to the overlap between the measurement error distributions in regions where the regression function is not one-to-one and due to the appearance of divergences in some of the densities entering the model. Many recent nonparametric identification results also rely on monotonicity assumptions, as discussed, for instance, in the Handbook of Econometrics chapter by Matzkin (2007).

Our main result can then be stated as follows, after we define the following convenient concept.

Definition 1 *We say that a random variable r is decomposable with F factor if r can be written as the sum of two independent random variables (which may be degenerate), one of which has the distribution F .*

Theorem 1 *Let Assumptions 1, 2 and 3 hold.*

1. *If $g(x^*)$ is **not** of the form*

$$g(x^*) = a + b \ln(e^{cx^*} + d) \tag{1}$$

for some constants $a, b, c, d \in \mathbb{R}$ then $f_{x^}(x^*)$ and $g(x^*)$ (over the support of $f_{x^*}(x^*)$) in Model 1 are **identified**.*

2. *If $g(x^*)$ **is** of the form (1) with⁵ $d > 0$, then neither $f_{x^*}(x^*)$ nor $g(x^*)$ in Model 1 are identified iff x^* has a density of the form*

$$f_{x^*}(x^*) = A \exp(-Be^{Cx^*} + CDx^*) (e^{Cx^*} + E)^{-F} \tag{2}$$

with⁶ $C \in \mathbb{R}$, $A, B, D, E, F \in [0, \infty[$ and Δx and Δy are decomposable with a type I extreme value factor.⁷

⁵A case where $d < 0$ can be converted into a case with $d > 0$ by permuting the roles of x and y .

⁶The constants A, B, C, D, E, F depend on a, b, c, d , although this dependence is omitted here for simplicity. Constants yielding a valid density can be found for any a, b, c, d (with $d > 0$).

⁷A type I extreme value distribution has a density of the general form $f(u) = K_1 \exp(K_2 \exp(K_3 u) + K_4 u)$. Here, the constant K_1, K_2, K_3, K_4 are such that $f(u)$ integrates to 1 and has zero mean and may depend on a, b, c, d , although this dependence is omitted here for simplicity.

3. If $g(x^*)$ is linear (i.e. of the form (1) with $d = 0$), then neither $f_{x^*}(x^*)$ nor $g(x^*)$ in Model 1 are identified iff x^* is normally distributed and either Δx or Δy is decomposable with a normal factor.

The phrasing of Cases 2 and 3 should make it clear that the conclusion of the theorem remains unchanged if one focuses on identifying $g(x^*)$ only and not $f_{x^*}(x^*)$, because the observationally equivalent models ruling identifiability out have different regression functions in all of the unidentified cases.

The proof of this result (provided in the Appendix) proceeds in four steps:

1. We reduce the identification problem of a model with errors along x and y into the equivalent problem of finding two observationally models, one having errors only along the x axis and one having errors only along the y axis.
2. We rule out a number of pathological cases in which the error distributions do not admit densities with respect to the Lebesgue measure by showing that such occurrences would actually imply identification of the model (in essence, any nonsmooth point gives away the shape of the regression function).
3. We derive necessary conditions for lack of identification that take the form of differential equations involving all densities. This establishes that the large class of models where these equations do not hold are identified.
4. Cases that do satisfy the differential equations are then systematically checked to see if they yield valid densities for all variables, thus pointing towards the only cases that are actually not identified and securing necessary and sufficient conditions for identifiability.

It is somewhat unexpected that in a fully nonparametric setting, the nonidentified family of regression functions would still be parametric with such a low dimension (only 4 adjustable parameters). It is also surprising that, even in the presumably difficult case of normally distributed regressors, most nonlinear specifications are actually identified. While our findings regarding linear regressions (Case 3) coincide with Reiersol (1950), the functional forms in the other nonidentified models (Case 2) are hardly trivial and would have been difficult to find without a systematic approach such as ours.

Theorem 1 can be extended in various useful directions. For instance, perfectly observed covariates w can be included simply by conditioning all densities (and expectations) on these covariates. We then establish identification of $f_{x^*|w}(x^*|w)$ and $g(x^*, w) \equiv E[y|x^*, w]$ and therefore of $f_{x^*,w}(x^*, w) = f_{x^*|w}(x^*|w) f_w(w)$. The above results do not yet establish identification of the measurement error distributions, but this can be trivially achieved by deconvolution techniques (once $g(x^*)$ and $f_{x^*}(x^*)$ have been determined) under the additional assumption that $E[e^{i\xi x^*}]$ and $E[e^{i\gamma g(x^*)}]$ do not vanish.

3 Conclusion

This note answers the long-standing question of the identifiability of the nonparametric classical errors-in-variables model with a rather encouraging result, namely, that only a specific 4-parameter parametric family of regression functions may exhibit lack of identifiability. Our identification result is agnostic regarding the type of estimator to be used in practice. One could use higher-order moment equalities, characteristic function equalities, or nonparametric sieve-type likelihoods. Finding the most convenient and statistically powerful method remains a nontrivial and important avenue of future research. It would also be useful to investigate whether these results extend to the case of nonclassical measurement error (i.e. relaxing some of the independence assumptions), where the dimensionality of the unknown distributions is greater or equal to the dimensionality of the observable distributions.

A Proof of Theorem 1

Let \mathcal{S}_u denote the support of the random variable u and let $f_u(u)$ denote its density (and similarly for the multivariate case).

Consider an alternative observationally equivalent model defined as:

Model 2 *Similar to Model 1 with x^* , Δx , Δy , $g(\cdot)$ replaced, respectively, by \tilde{x}^* , $\Delta\tilde{x}$, $\Delta\tilde{y}$, $\tilde{g}(\cdot)$.*

It is clear that any assumptions (including regularity conditions) made regarding Model 1 must hold for this alternative model as well.

We first reduce the identification problem to a simpler but equivalent problem involving only one error term. Consider the following two models:

Model 3 Let $\bar{x}, \bar{y}, x^*, \Delta\bar{x}$ be scalar real-valued random variables such that

$$\begin{aligned}\bar{y} &= g(x^*) \\ \bar{x} &= x^* + \Delta\bar{x}\end{aligned}$$

where \bar{x} and \bar{y} are observable (and may differ from x, y in Model 1), where the unobservable x^* and $g(x^*)$ are as in Model 1, and $\Delta\bar{x}$ is independent from x^* , $E[\Delta\bar{x}] = 0$ and the distribution of $\Delta\bar{x}$ is a factor⁸ of the distribution of Δx in Model 1.

Model 4 Let $\bar{x}, \bar{y}, \tilde{x}^*, \Delta\bar{y}$ be scalar real-valued random variables such that

$$\begin{aligned}\bar{y} &= \tilde{g}(\tilde{x}^*) + \Delta\bar{y} \\ \bar{x} &= \tilde{x}^*\end{aligned}$$

where the observables \bar{x} and \bar{y} are as in Model 3, where the unobservable \tilde{x}^* and $\tilde{g}(\tilde{x}^*)$ are as in Model 2 and where $\Delta\bar{y}$ is independent from \tilde{x}^* , $E[\Delta\bar{y}] = 0$ and the distribution of $\Delta\bar{y}$ is a factor of the distribution of Δy in Model 2.

Note that, given the above definitions, $\Delta x = \Delta\bar{x} + \Delta\tilde{x}$. This assumes, without loss of generality, that the distribution of $\Delta\tilde{x}$ is a factor of the distribution of Δx (otherwise, one can just permute the role of Models 1 and 2, which interchanges the role of tilded and non tilded symbols).

Lemma 1 Under Assumptions 1-3, there exist two distinct observationally equivalent Models 1 and 2 iff there exist two distinct observationally equivalent models of the form of Models 3 and 4. Moreover, when two such models exist, the distributions of $\bar{x}, \bar{y}, \Delta\bar{x}$ and $\Delta\bar{y}$ all admit a density with respect to the Lebesgue measure and are supported on all of \mathbb{R} .

Proof. (1) The joint characteristic function of x and y , defined as $E[e^{i\xi x} e^{i\gamma y}]$, conveys the same information as the joint distribution of x and y . Under Model 1,

$$E[e^{i\xi x} e^{i\gamma y}] = E[e^{i\xi x^*} e^{i\gamma g(x^*)} e^{i\xi \Delta x} e^{i\gamma \Delta y}].$$

The independence conditions stated in Model 1 then imply that

$$E[e^{i\xi x} e^{i\gamma y}] = E[e^{i\xi x^*} e^{i\gamma g(x^*)}] E[e^{i\xi \Delta x}] E[e^{i\gamma \Delta y}]. \quad (3)$$

⁸A distribution F is said to be a *factor* of a distribution H if there exists a distribution G (which may be degenerate) such that the random variable $h = f + g$ has distribution H , where f, g are independent random variables drawn from F, G respectively.

We seek an alternative observationally equivalent model (Model 2, denoted with \sim) also satisfying:

$$E [e^{i\xi x} e^{i\gamma y}] = E [e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{g}(\tilde{x}^*)}] E [e^{i\xi \Delta \tilde{x}}] E [e^{i\gamma \Delta \tilde{y}}]. \quad (4)$$

Define

$$\alpha(\xi) \equiv \frac{E [e^{i\xi \Delta x}]}{E [e^{i\xi \Delta \tilde{x}}]} \quad \beta(\gamma) \equiv \frac{E [e^{i\gamma \Delta \tilde{y}}]}{E [e^{i\gamma \Delta y}]}$$

and note that $\alpha(\xi)$ and $\beta(\gamma)$ are everywhere continuous, nonvanishing and finite.⁹ Also, $\alpha(0) = 1$, $\alpha'(0) = 0$ and $\beta(0) = 1$, $\beta'(0) = 0$. Rearranging, we obtain

$$\begin{aligned} E [e^{i\xi \Delta \tilde{x}}] &= \frac{E [e^{i\xi \Delta x}]}{\alpha(\xi)} \\ E [e^{i\gamma \Delta \tilde{y}}] &= \beta(\gamma) E [e^{i\gamma \Delta y}]. \end{aligned}$$

Substituting these expressions into (4), yields

$$E [e^{i\xi x} e^{i\gamma y}] = \left(E [e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{g}(\tilde{x}^*)}] \frac{\beta(\gamma)}{\alpha(\xi)} \right) E [e^{i\xi \Delta x}] E [e^{i\gamma \Delta y}].$$

But, by (3), this is also equal to $E [e^{i\xi x^*} e^{i\gamma g(x^*)}] E [e^{i\xi \Delta x}] E [e^{i\gamma \Delta y}]$ and therefore

$$E [e^{i\xi x^*} e^{i\gamma g(x^*)}] E [e^{i\xi \Delta x}] E [e^{i\gamma \Delta y}] = \left(E [e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{g}(\tilde{x}^*)}] \frac{\beta(\gamma)}{\alpha(\xi)} \right) E [e^{i\xi \Delta x}] E [e^{i\gamma \Delta y}].$$

Since $E [e^{i\xi \Delta x}]$, $E [e^{i\gamma \Delta y}]$ and $\alpha(\xi)$ are finite and nonvanishing, we can multiply each side by $\alpha(\xi) / (E [e^{i\xi \Delta x}] E [e^{i\gamma \Delta y}])$ to yield:

$$E [e^{i\xi x^*} e^{i\gamma g(x^*)}] \alpha(\xi) = E [e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{g}(\tilde{x}^*)}] \beta(\gamma) \quad (5)$$

or

$$E [e^{i\xi x^*} e^{i\gamma g(x^*)}] \alpha(\xi) E [e^{i\gamma 0}] = E [e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{g}(\tilde{x}^*)}] E [e^{i\xi 0}] \beta(\gamma).$$

In other words, Models 1 and 2 are observationally equivalent iff there exists a model with errors only in the regressor (Model 3, where $\alpha(\xi)$ is the characteristic function of $\Delta \tilde{x}$) that is observationally equivalent to a model with errors in the dependent variable (Model 4, where $\beta(\gamma)$ is the characteristic function of $\Delta \tilde{y}$). This completes the first part the proof.

⁹That is, finite at each point, though not necessarily uniformly bounded.

(2) It remains to be shown that we can indeed limit ourselves to $\alpha(\xi)$ and $\beta(\gamma)$ that are valid characteristic functions and, more specifically, to characteristic functions of densities supported on \mathbb{R} . Define $y^* \equiv g(x^*)$ and $h(y^*) \equiv g^{-1}(y^*)$ and note that y^* admits a density. $f_{y^*}(y^*) = f_{x^*}(h(y^*)) / g'(h(y^*))$ since $g'(x^*) \neq 0$ by assumption. We can then rewrite Equation (5) as

$$E [e^{i\xi h(y^*)} e^{i\gamma y^*}] \alpha(\xi) = E [e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{g}(\tilde{x}^*)}] \beta(\gamma). \quad (6)$$

We now calculate the inverse Fourier transform (FT) of each side using the convolution theorem. To this effect, we calculate the FT of each term individually. Since we can write

$$\begin{aligned} E [e^{i\xi h(y^*)} e^{i\gamma y^*}] &= \int \int \delta(x^* - h(y^*)) f_{y^*}(y^*) e^{i\xi x^*} e^{i\gamma y^*} dx^* dy^* \\ E [e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{g}(\tilde{x}^*)}] &= \int \int \delta(\tilde{y}^* - \tilde{g}(\tilde{x}^*)) \tilde{f}_{\tilde{x}^*}(\tilde{x}^*) e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{y}^*} d\tilde{x}^* d\tilde{y}^*, \end{aligned}$$

the inverse FT of $E [e^{i\xi h(y^*)} e^{i\gamma y^*}]$ and $E [e^{i\xi \tilde{x}^*} e^{i\gamma \tilde{g}(\tilde{x}^*)}]$ are, respectively, $\delta(x^* - h(y^*)) f_{y^*}(y^*)$ and $\delta(\tilde{y}^* - \tilde{g}(\tilde{x}^*)) \tilde{f}_{\tilde{x}^*}(\tilde{x}^*)$, where $\delta(\cdot)$ denotes a delta function.

Let $\mathcal{W}_{\Delta\bar{x}}$ denote the set where the inverse FT of $\alpha(\xi)$ is well-defined and finite¹⁰ and let $f_{\Delta\bar{x}}(\Delta\bar{x})$ denote this inverse FT for $\Delta\bar{x} \in \mathcal{W}_{\Delta\bar{x}}$. Similarly define $\mathcal{W}_{\Delta\bar{y}}$ and $\tilde{f}_{\Delta\bar{y}}(\Delta\bar{y})$ for $\beta(\gamma)$. Note that the sets $\mathcal{W}_{\Delta\bar{x}}$ and $\mathcal{W}_{\Delta\bar{y}}$ cannot be empty since it would then be impossible for $\alpha(\xi)$ and $\beta(\gamma)$ to be finite everywhere.¹¹ By (6) and the convolution theorem, we have

$$\int \delta(x^* - h(\bar{y})) f_{\bar{y}}(\bar{y}) f_{\Delta\bar{x}}(\bar{x} - x^*) dx^* = \int \delta(\tilde{y}^* - \tilde{g}(\bar{x})) f_{\bar{x}}(\bar{x}) \tilde{f}_{\Delta\bar{y}}(\bar{y} - \tilde{y}^*) d\tilde{y}^*$$

where we have used the equivalence $y^* = \bar{y}$ (under Model 3) and $\bar{x} = \tilde{x}^*$ (under Model 4). Using the properties of the delta function $\delta(\cdot)$,

$$f_{\bar{y}}(\bar{y}) f_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) = f_{\bar{x}}(\bar{x}) \tilde{f}_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \quad (7)$$

an equality which holds for (\bar{x}, \bar{y}) such that $\bar{x} - h(\bar{y}) \in \mathcal{W}_{\Delta\bar{x}}$ and $\bar{y} - \tilde{g}(\bar{x}) \in \mathcal{W}_{\Delta\bar{y}}$.

Suppose that at some point (\bar{x}_0, \bar{y}_0) in the interior of the support of (\bar{x}, \bar{y}) , we have that $\tilde{f}_{\Delta\bar{y}}(\bar{y}_0 - \tilde{g}(\bar{x}_0))$ changes sign, becomes zero, infinite or undefined. Then the same behavior must necessarily occur in $f_{\Delta\bar{x}}(\bar{x}_0 - h(\bar{y}_0))$ at the same point (\bar{x}_0, \bar{y}_0)

¹⁰That is, for a given $\Delta\bar{x}$, $\lim_{t \rightarrow \infty} \int_{-t}^t \alpha(\xi) e^{i\xi \Delta\bar{x}} d\xi$ exists in \mathbb{C} .

¹¹If $\mathcal{W}_{\Delta\bar{x}}$ is empty, $\tilde{f}_{\Delta\bar{x}}(\Delta\bar{x})$ would be undefined or infinite for all points in \mathbb{R} , hence its Fourier transform $\alpha(\xi)$ could not exist.

because multiplication by a bounded positive number (here, $f_{\bar{y}}(\bar{y}_0)$ and $f_{\bar{x}}(\bar{x}_0)$ are finite by assumption) does not affect whether a quantity is well-defined, positive, nonzero or finite. Furthermore, the same behavior would occur along the whole curve (\bar{x}, \bar{y}) giving the same value of $v \equiv \bar{y}_0 - \tilde{g}(\bar{x}_0) = \bar{y} - \tilde{g}(\bar{x})$ or the same value of $u \equiv \bar{x}_0 - h(\bar{y}_0) = \bar{x} - h(\bar{y})$. If the curves

$$\mathcal{V}_v = \{(\tilde{x}^*, \tilde{g}(\tilde{x}^*) + v) : \tilde{x}^* \in \mathcal{S}_{\tilde{x}^*}\} \text{ and } \mathcal{U}_u = \{(h(y^*) + u, y^*) : y^* \in \mathcal{S}_{y^*}\} \quad (8)$$

did not coincide, then it would be possible to recursively construct the following sequence of sets

$$\begin{aligned} \mathcal{V}^0 &\equiv \mathcal{V}_v \\ \mathcal{U}^0 &\equiv \mathcal{U}_u \\ \mathcal{V}^{n+1} &= \bigcup_{v:\mathcal{V}_v \cap \mathcal{U}^n \neq \emptyset} \mathcal{V}_v \\ \mathcal{U}^{n+1} &= \bigcup_{u:\mathcal{U}_u \cap \mathcal{V}^{n+1} \neq \emptyset} \mathcal{U}_u \end{aligned}$$

that is such that $\mathcal{V}^n \rightarrow \mathcal{S}_{\bar{x}\bar{y}}$ and $\mathcal{U}^n \rightarrow \mathcal{S}_{\bar{x}\bar{y}}$. This implies that $f_{\Delta\bar{x}}$ and $\tilde{f}_{\Delta\bar{y}}$ are either everywhere zero, everywhere changing sign, everywhere infinite or everywhere undefined. None of these situations are possible, since the FT of $f_{\Delta\bar{x}}$ and $\tilde{f}_{\Delta\bar{y}}$, respectively, $\alpha(\xi)$ and $\beta(\gamma)$, are everywhere well-defined and nonzero.

Hence the curves in (8) would have to coincide. We can reparametrize the right-hand side curve, letting $y^* = g(x^*)$, to yield $\{(x^* + u, g(x^*)) : x^* \in \mathcal{S}_{x^*}\}$ and we must then have the equality.

$$(\tilde{x}^*, \tilde{g}(\tilde{x}^*) + v) = (x^* + u, g(x^*))$$

implying that

$$\tilde{g}(x^* + u) + v = g(x^*),$$

i.e., $\tilde{g}(\cdot)$ and $g(\cdot)$ are just horizontally and vertically shifted versions of each other. But any nonzero shift would imply that either one of the models is violating one of the zero mean assumptions on the disturbances.¹² Hence, for any pair of valid models 3 and 4, we must have $\tilde{g}(x^*) = g(x^*)$. The density of x^* can then be determined (up to

¹²The only exception in the linear specification, where two nonzero shifts along each axes may cancel each other. But in this case, the shifted curve is identical to the original one.

a multiplicative constant determined by the normalization of unit total probability) from the density $f_{\bar{x}\bar{y}}(\bar{x}, \bar{y})$ along the line $\bar{y} = g(\bar{x}) + u$ for some $u \in \mathcal{W}_{\Delta\bar{y}}$.

This means that if there are any points where $\tilde{f}_{\Delta\bar{y}}$ or $f_{\Delta\bar{x}}$ are ill-defined, change sign, become zero or are infinite, then Model 3 and 4 are such that $\tilde{g}(x^*) = g(x^*)$ and $\tilde{f}_{\bar{x}^*}(x^*) = f_{x^*}(x^*)$. So any pair of *distinct* but *observationally equivalent* models must be such that $\tilde{f}_{\Delta\bar{y}}$ and $f_{\Delta\bar{x}}$ are well-defined densities with respect to the Lebesgue measure that are nonzero, finite and never change sign (and are positive, since $\alpha(0) = 1$ and $\beta(0) = 1$). Since $\tilde{f}_{\Delta\bar{y}}$ and $f_{\Delta\bar{x}}$ are supported on \mathbb{R} , so must $f_{\bar{x}}$ and $f_{\bar{y}}$, in light of Equation (7). \blacksquare

Now, continuing the proof of Theorem 1: Under Model 3, the joint density of \bar{x} and \bar{y} can be written as:

$$f_{\bar{x}\bar{y}}(\bar{x}, \bar{y}) = f_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) f_{\bar{y}}(\bar{y}) \quad (9)$$

where $h(y) \equiv g^{-1}(y)$ (which exists by Assumption 3), while under Model 4, we have

$$f_{\bar{x}\bar{y}}(\bar{x}, \bar{y}) = \tilde{f}_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) f_{\bar{x}}(\bar{x}) \quad (10)$$

where the \sim on the densities emphasizes the quantities that differ under the alternative model.

Since the two models must be observationally equivalent, we equate (9) and (10):

$$f_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) f_{\bar{y}}(\bar{y}) = \tilde{f}_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) f_{\bar{x}}(\bar{x}). \quad (11)$$

After rearranging (11) and taking logs, we obtain:

$$\ln \tilde{f}_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) - \ln f_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) = \ln f_{\bar{y}}(\bar{y}) - \ln f_{\bar{x}}(\bar{x}), \quad (12)$$

where these densities are always positive (by Lemma 1), so that the $\ln(\cdot)$ are always well-defined.

We will find necessary conditions for Equation (12) to hold, in order to narrow down the search for possible solutions that would provide distinct but observationally equivalent models. Next, we will need to check that these solutions actually lead to proper densities (i.e. with finite area) for all variables in order to obtain necessary and sufficient condition for identifiability.

We use the following Lemma:

Lemma 2 *A twice-continuously differentiable function $c(x, y)$ is such that $\partial^2 c(x, y) / \partial x \partial y = 0 \forall x, y$ iff it can be written as $c(x, y) = a(x) + b(y)$.*

Proof. We may write

$$c(x, y) = c(0, 0) + \int_0^x \frac{\partial c(u, 0)}{\partial x} du + \int_0^y \frac{\partial c(x, v)}{\partial y} dv$$

where

$$\frac{\partial c(x, v)}{\partial y} = \frac{\partial c(0, v)}{\partial y} + \int_0^x \frac{\partial^2 c(u, v)}{\partial x \partial y} du = \frac{\partial c(0, y)}{\partial y} + 0$$

if $\partial^2 c(x, y) / \partial x \partial y = 0$. Hence,

$$c(x, y) = \underbrace{c(0, 0) + \int_0^x \frac{\partial c(u, 0)}{\partial x} du}_{a(x)} + \underbrace{\int_0^y \frac{\partial c(0, v)}{\partial y} dv}_{b(y)}.$$

Conversely,

$$\frac{\partial^2 c(x, y)}{\partial x \partial y} = \frac{\partial^2 a(x)}{\partial x \partial y} + \frac{\partial^2 b(y)}{\partial x \partial y} = 0.$$

■

Note that differentiability of $g(x^*)$, combined with $g'(x^*) \neq 0$ implies that $h(\bar{y}) \equiv g^{-1}(\bar{y})$ is differentiable.

Let F denote the logarithms of the corresponding lowercase density and rewrite Equation (12) as

$$\tilde{F}_{\Delta \bar{y}}(\bar{y} - \tilde{g}(\bar{x})) - F_{\Delta \bar{x}}(\bar{x} - h(\bar{y})) = F_{\bar{y}}(\bar{y}) - F_{\bar{x}}(\bar{x}).$$

By Lemma 2, we must then have

$$\begin{aligned} \frac{\partial^2}{\partial \bar{x} \partial \bar{y}} \tilde{F}_{\Delta \bar{y}}(\bar{y} - \tilde{g}(\bar{x})) - \frac{\partial^2}{\partial \bar{x} \partial \bar{y}} F_{\Delta \bar{x}}(\bar{x} - h(\bar{y})) &= 0 \\ \tilde{F}_{\Delta \bar{y}}''(\bar{y} - \tilde{g}(\bar{x})) \tilde{g}'(\bar{x}) - F_{\Delta \bar{x}}''(\bar{x} - h(\bar{y})) h'(\bar{y}) &= 0 \end{aligned} \quad (13)$$

In the above, we have assumed differentiability of $\tilde{F}_{\Delta \bar{y}}$ and $\tilde{F}_{\Delta \bar{x}}$, but if this fails to hold, we can show that the model is actually identified: The functions $\tilde{g}'(\bar{x})$ and $h'(\bar{y})$ are bounded, continuous and nonzero by Assumption 3. Hence, the points (\bar{x}, \bar{y}) where $\tilde{F}_{\Delta \bar{y}}(\bar{y} - \tilde{g}(\bar{x}))$ and $\tilde{F}_{\Delta \bar{x}}(\bar{x} - h(\bar{y}))$ and not twice continuously differentiable must coincide. By the same reasoning as in the second part of the proof of Lemma

1, the alternative model would have to be identical to the true model.¹³ We can therefore rule out insufficient continuous differentiability for the purpose of finding models that are not identified. To proceed, we need the following Lemma.

Lemma 3 *Let Assumptions 1-3 hold, $h(\cdot) \equiv g^{-1}(\cdot)$ and let $g(\cdot)$ and $\tilde{g}(\cdot)$ be as defined in Models 3 and 4, respectively. These models are assumed to be distinct. If two functions $a(\cdot)$ and $b(\cdot)$ are such that $a(\bar{y} - \tilde{g}(\bar{x})) = b(\bar{x} - h(\bar{y})) \forall (\bar{x}, \bar{y}) \in \mathbb{R}^2$, then $a(\cdot)$ and $b(\cdot)$ are constant functions over \mathbb{R} . Similarly if $a(\bar{y} - \tilde{g}(\bar{x})) = 0 \Leftrightarrow b(\bar{x} - h(\bar{y})) = 0 \forall (\bar{x}, \bar{y}) \in \mathbb{R}^2$, then $a(\cdot)$ and $b(\cdot)$ are zero over \mathbb{R} if either one vanishes at a single point.*

Proof. Note that, by Lemma 1, $\{(\bar{y} - \tilde{g}(\bar{x}), \bar{x} - h(\bar{y})) : \forall (\bar{x}, \bar{y}) \in \mathbb{R}^2\} = \mathbb{R}^2$. It is therefore possible to vary \bar{x} and \bar{y} so that $\Delta\bar{y} = \bar{y} - \tilde{g}(\bar{x})$ remains constant while $\Delta\bar{x} = \bar{x} - h(\bar{y})$ varies or vice-versa. Hence, it is possible to vary (\bar{x}, \bar{y}) in such a way such that $\Delta\bar{x}$ varies but $\Delta\bar{y}$ remains constant. Having $a(\Delta\bar{y})$ constant implies that $b(\Delta\bar{x})$ also is, even though its argument is varying. This shows that $b(\Delta\bar{x})$ is constant along a one-dimensional slice of constant $\Delta\bar{y}$. Then, varying (\bar{x}, \bar{y}) so that the argument of the $b(\Delta\bar{x})$ is constant, we can show that the $a(\Delta\bar{y})$ is constant along a one-dimensional slice of constant $\Delta\bar{x}$. Repeating the process we can show that $a(\Delta\bar{y})$ and $b(\Delta\bar{x})$ are constant for all $(\Delta\bar{x}, \Delta\bar{y}) \in \mathbb{R}^2$ and therefore for all $(\bar{x}, \bar{y}) \in \mathbb{R}^2$. A similar argument demonstrates the second conclusion of the Lemma. \blacksquare

Continuing with the proof of Theorem 1, we can rearrange Equation (13) to yield

$$\tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) = \frac{h'(\bar{y})}{\tilde{g}'(\bar{x})} F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y})), \quad (14)$$

where the ratio $h'(\bar{y})/\tilde{g}'(\bar{x})$ is nonzero and finite by assumption. Hence if $F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))$ is zero, then so is $\tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}))$ and vice versa. If either of those two functions vanishes at a point, by Lemma 3, they must vanish everywhere. It would follow that $\tilde{F}_{\Delta\bar{y}}(\Delta\bar{y})$ and $F_{\Delta\bar{x}}(\Delta\bar{x})$ would be linear and that the corresponding densities $\tilde{f}_{\Delta\bar{y}}(\Delta\bar{y})$ and $f_{\Delta\bar{x}}(\Delta\bar{x})$ would be exponential over \mathbb{R} , which is an improper density. It follows that our presumption that either $F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))$ or $\tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}))$ vanish at some point is incorrect.

¹³Note that even if a function is nowhere differentiable to some given order, the singularities cannot be fully translation-invariant. Informally, if a derivative is “ $+\infty$ ” at every point, then the function would be infinite everywhere, a situation already ruled out in Lemma 1. Divergence in the derivatives must change sign to maintain the density finite. These changes in derivative sign could be exploited to gain identification as in Lemma 1.

Hence we may assume that $F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))$ and $\tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}))$ do not vanish. Since these functions are continuous, this means they never change sign. Also note that, by assumption, $h'(\bar{y})$ and $\tilde{g}'(\bar{x})$ never change sign or vanish either. We can thus, without loss of generality, rewrite Equation (13) as:

$$\frac{\left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right|}{\left| F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) \right|} = \frac{|h'(\bar{y})|}{|\tilde{g}'(\bar{x})|} \quad (15)$$

or

$$\ln \left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right| - \ln |F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))| = \ln |h'(\bar{y})| - \ln |\tilde{g}'(\bar{x})|$$

Again, since the right-hand side is a difference of functions of \bar{y} and \bar{x} , respectively, we must have¹⁴ (by Lemma 2)

$$\begin{aligned} \frac{\partial^2}{\partial x \partial y} \ln \left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right| - \frac{\partial^2}{\partial x \partial y} \ln |F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))| &= 0 \\ \left(\ln \left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right| \right)'' \tilde{g}'(\bar{x}) - \left(\ln |F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))| \right)'' h'(\bar{y}) &= 0 \end{aligned}$$

By the same argument as before, if $\left(\ln \left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right| \right)'' = 0$ or $\left(\ln |F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))| \right)'' = 0$ at a point then they must vanish everywhere, a situation covered in Case 2 below. (We can also re-use the argument that lack of sufficient continuous differentiability implies identification, hence we can assume sufficient continuous differentiability.)

Case 1 If $\left(\ln \left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right| \right)''$ and $\left(\ln |F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))| \right)''$ do not vanish, we may write

$$\frac{\left| \left(\ln \left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right| \right)'' \right|}{\left| \left(\ln |F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))| \right)'' \right|} = \frac{|h'(\bar{y})|}{|\tilde{g}'(\bar{x})|}$$

combined with Equation (15) this implies:

$$\begin{aligned} \frac{\left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right|}{\left| F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) \right|} &= \frac{\left| \left(\ln \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right)'' \right|}{\left| \left(\ln F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) \right)'' \right|} \\ \frac{\left| \left(\ln \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right)'' \right|}{\left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right|} &= \frac{\left| \left(\ln F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) \right)'' \right|}{\left| F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) \right|} \end{aligned} \quad (16)$$

¹⁴The notation $\left(\ln \left| \tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) \right| \right)''$ stands for $\left(\ln \left| \tilde{F}''_{\Delta\bar{y}}(u) \right| \right)'' \big|_{u=\bar{y}-\tilde{g}(\bar{x})}$.

By Lemma 3, each side of this equality must equal a constant, say A . Note that this equality is only a necessary condition for lack of identifiability. For instance, it does not ensure that $\left| \frac{\tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}))}{F''(u)} \right| / |F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))|$ can actually be written as a ratio of a function of \bar{y} and a function of \bar{x} , as required by Equation (15). This will need to be subsequently checked.

We now find densities such that the left-hand (or right-hand) side of Equation (16) is constant. Letting $u = \bar{y} - \tilde{g}(\bar{x})$ and $F(\cdot) \equiv \tilde{F}_{\Delta\bar{y}}(\cdot)$ (or similarly, $u = \bar{x} - h(\bar{y})$ and $F(\cdot) \equiv F_{\Delta\bar{x}}(\cdot)$), we must have that

$$\begin{aligned} \frac{(\ln |F''(u)|)''}{F''(u)} &= \pm A \\ (\ln |F''(u)|)'' &= \pm A F''(u) \\ (\ln |F''(u)|)' &= \pm A F'(u) + B \\ \ln |F''(u)| &= \pm A F(u) + Bu + C \\ F''(u) &= \pm \exp(\pm A F(u) + Bu + C) \\ F''(u) &= -\exp(A F(u) + Bu + C) \end{aligned}$$

where A, B, C are some constants and where one of the “ \pm ” has been incorporated into the constant A and the other has been set to “ $-$ ”, because the “ $+$ ” solution does not lead to a proper density.

Lemma 4 *The solution $F(u)$ to*

$$F''(u) = -\exp(A F(u) + Bu + C) \tag{17}$$

is:

$$F(u) = -\frac{B}{A}u - \frac{C}{A} + \frac{1}{A} \ln \left(\frac{2D^2}{A} \rho(D(u - u_0)) \right) \tag{18}$$

where

$$\rho(v) = 1 - \tanh^2(v) = 4(\exp(v) + \exp(-v))^{-2}$$

and where A, B, C, D, u_0 are constants.

Proof. This solution can be verified by substitution into the differential equation and noting that any initial conditions in $F(0)$ and $F'(0)$ can be accommodated by adjusting the constants D, u_0 . ■

The density corresponding to $F(u)$ is

$$f(u) = C_1 \exp\left(-\frac{B}{A}u\right) (\rho(D(u - u_0)))^{1/A}$$

where C_1 is such that the density integrates to 1. To check that this is a valid solution, we first calculate what the implied forms of $\tilde{g}(\bar{x})$ and $h(\bar{y})$ are. From Equation (14), we know that

$$\frac{\tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}))}{F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))} = \frac{h'(\bar{y})}{\tilde{g}'(\bar{x})} \quad (19)$$

where we can find an expression for $F''_{\Delta\bar{x}}(\cdot)$ and $\tilde{F}''_{\Delta\bar{y}}(\cdot)$, generically denoted $F''(\cdot)$ using Equations (17) and (18):

$$\begin{aligned} F''(u) &= \exp\left(A\left(-\frac{B}{A}u - \frac{C}{A} + \frac{1}{A}\ln\left(\frac{2D^2}{A}\rho(D(u - u_0))\right)\right) + Bu + C\right) \\ &= \frac{2D^2}{A}\rho(D(u - u_0)). \end{aligned}$$

The constants D and u_0 may differ for $F''_{\Delta\bar{x}}(\cdot)$ and $\tilde{F}''_{\Delta\bar{y}}(\cdot)$ and we distinguish them by subscripts $\Delta\bar{x}$ or $\Delta\bar{y}$. The constant A is the same, however. Next, we calculate the ratio:

$$\begin{aligned} \frac{\tilde{F}''_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}))}{F''_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))} &= \frac{\frac{2D_{\Delta\bar{y}}^2}{A}\rho(D_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}) - u_{0\Delta\bar{y}}))}{\frac{2D_{\Delta\bar{x}}^2}{A}\rho(D_{\Delta\bar{x}}(\bar{x} - h(\bar{y}) - u_{0\Delta\bar{x}}))} \\ &= \frac{D_{\Delta\bar{y}}^2 (\exp(D_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}) - u_{0\Delta\bar{y}})) + \exp(-D_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}) - u_{0\Delta\bar{y}})))^{-2}}{D_{\Delta\bar{x}}^2 (\exp(D_{\Delta\bar{x}}(\bar{x} - h(\bar{y}) - u_{0\Delta\bar{x}})) + \exp(-D_{\Delta\bar{x}}(\bar{x} - h(\bar{y}) - u_{0\Delta\bar{x}})))^{-2}} \\ &= \frac{D_{\Delta\bar{x}}^{-2} (2 + \exp(2D_{\Delta\bar{x}}(\bar{x} - h(\bar{y}) - u_{0\Delta\bar{x}})) + \exp(-2D_{\Delta\bar{x}}(\bar{x} - h(\bar{y}) - u_{0\Delta\bar{x}})))}{D_{\Delta\bar{y}}^{-2} (2 + \exp(2D_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}) - u_{0\Delta\bar{y}})) + \exp(-2D_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}) - u_{0\Delta\bar{y}})))} \end{aligned}$$

and note that it cannot be written as a ratio of a function of \bar{y} and a function of \bar{x} (unless $\tilde{g}(\bar{x})$ or $h(\bar{y})$ are constant, a situation ruled out by Assumption 3). Hence Equation (15) cannot possibly hold and this solution is not valid. Hence, except possibly when $(\ln F''(u))'' = 0$, there exists no pair of observationally equivalent models of the forms of Model 3 and 4.

Case 2 We now consider the (so far excluded) case where $(\ln F''(u))'' = 0$ for $F =$

$F_{\Delta\bar{x}}$ and $\tilde{F}_{\Delta\bar{y}}$. We have

$$\begin{aligned} (\ln |F''(u)|)'' &= 0 \\ |F''(u)| &= \exp(Au + B) \end{aligned} \quad (20)$$

$$\begin{aligned} F''(u) &= \pm \exp(Au + B) \\ F'(u) &= \pm A^{-1} \exp(Au + B) + C \\ F(u) &= -A^{-2} \exp(Au + B) + Cu + D \end{aligned} \quad (21)$$

for some adjustable constants A, B, C, D with $A \neq 0$ (the case $A = 0$ is covered in case 3 below). We have selected the negative branch of the “ \pm ” of since it is the only one yielding a proper density. The density corresponding to (21) is of the form

$$f(u) = \exp(-A^{-2} \exp(Au + B) + Cu + D) \quad (22)$$

where the constants A, B, C, D are selected so as to satisfy the normalization constraint and the zero mean assumption. In the sequel, we will distinguish the constants A, B, C, D by subscripts $\Delta\bar{x}, \Delta\bar{y}$ corresponding to the densities of $\Delta\bar{x}$ and $\Delta\bar{y}$, respectively. We first determine $h(\bar{y})$ and $g(\bar{x})$ through relationship (15):

$$\begin{aligned} \frac{|h'(\bar{y})|}{|\tilde{g}'(\bar{x})|} &= \frac{|\tilde{F}_{\Delta\bar{y}}''(\bar{y} - \tilde{g}(\bar{x}))|}{|F_{\Delta\bar{x}}''(\bar{x} - h(\bar{y}))|} = \frac{\exp(A_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) + B_{\Delta\bar{y}})}{\exp(A_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) + B_{\Delta\bar{x}})} \\ &= \frac{\exp(A_{\Delta\bar{x}}h(\bar{y}) + A_{\Delta\bar{y}}\bar{y} + B_{\Delta\bar{y}})}{\exp(A_{\Delta\bar{y}}\tilde{g}(\bar{x}) + A_{\Delta\bar{x}}\bar{x} + B_{\Delta\bar{x}})} \end{aligned}$$

Rearranging, we must have

$$\frac{|h'(\bar{y})|}{\exp(A_{\Delta\bar{x}}h(\bar{y}) + A_{\Delta\bar{y}}\bar{y} + B_{\Delta\bar{y}})} = \frac{|\tilde{g}'(\bar{x})|}{\exp(A_{\Delta\bar{y}}\tilde{g}(\bar{x}) + A_{\Delta\bar{x}}\bar{x} + B_{\Delta\bar{x}})}$$

and each side must be equal to the same constant (say, $-A_{hg}$) since they depend on different variables. The solution to the differential equation

$$h'(\bar{y}) = \pm A_{hg} \exp(A_{\Delta\bar{x}}h(\bar{y}) + A_{\Delta\bar{y}}\bar{y} + B_{\Delta\bar{y}}) \quad (23)$$

is

$$h(\bar{y}) = -\frac{B_{\Delta\bar{y}}}{A_{\Delta\bar{x}}} - \frac{1}{A_{\Delta\bar{x}}} \ln \left(\pm \frac{A_{\Delta\bar{x}}A_{hg}}{A_{\Delta\bar{y}}} (e^{A_{\Delta\bar{y}}\bar{y}} + C_{1\Delta\bar{y}}A_{\Delta\bar{y}}) \right), \quad (24)$$

where $C_{1\Delta\bar{y}}$ is a constant. (This can be shown by substitution of (24) into (23) and by noting that any initial condition $h(0)$ can be accommodated by adjusting $C_{1\Delta\bar{y}}$.)

Similarly,

$$\tilde{g}'(\bar{x}) = \pm A_{hg} \exp(A_{\Delta\bar{y}}\tilde{g}(\bar{x}) + A_{\Delta\bar{x}}\bar{x} + B_{\Delta\bar{x}})$$

and

$$\tilde{g}(\bar{x}) = -\frac{B_{\Delta\bar{x}}}{A_{\Delta\bar{y}}} - \frac{1}{A_{\Delta\bar{y}}} \ln \left(\pm \frac{A_{\Delta\bar{y}} A_{hg}}{A_{\Delta\bar{x}}} (e^{A_{\Delta\bar{x}}\bar{x}} + C_{1\Delta\bar{x}} A_{\Delta\bar{x}}) \right) \quad (25)$$

where $C_{1\Delta\bar{x}}$ is a constant. From Equations (11), (22) (24) and (25), we have

$$\begin{aligned} \frac{f_{\bar{y}}(\bar{y})}{f_{\bar{x}}(\bar{x})} &= \frac{\tilde{f}_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}))}{f_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))} = \frac{\exp(A_{\Delta\bar{y}}^{-2} \exp(A_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) + B_{\Delta\bar{y}}) + C_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) + D_{\Delta\bar{y}})}{\exp(A_{\Delta\bar{x}}^{-2} \exp(A_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) + B_{\Delta\bar{x}}) + C_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) + D_{\Delta\bar{x}})} \\ &= \frac{\exp\left(A_{\Delta\bar{y}}^{-2} \exp(A_{\Delta\bar{y}}\bar{y} + B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \left(\frac{\pm A_{\Delta\bar{y}} A_{hg}}{A_{\Delta\bar{x}}}\right) (e^{A_{\Delta\bar{x}}\bar{x}} + C_{1\Delta\bar{x}} A_{\Delta\bar{x}})\right) + C_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) + D_{\Delta\bar{y}}}{\exp\left(A_{\Delta\bar{x}}^{-2} \exp(A_{\Delta\bar{x}}\bar{x} + B_{\Delta\bar{y}} + B_{\Delta\bar{x}}) \left(\frac{\pm A_{\Delta\bar{x}} A_{hg}}{A_{\Delta\bar{y}}}\right) (e^{A_{\Delta\bar{y}}\bar{y}} + C_{1\Delta\bar{y}} A_{\Delta\bar{y}})\right) + C_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) + D_{\Delta\bar{x}}} \\ &= \frac{\exp\left(\exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{hg}}{A_{\Delta\bar{y}} A_{\Delta\bar{x}}} \exp(A_{\Delta\bar{y}}\bar{y}) \exp(A_{\Delta\bar{x}}\bar{x})\right)}{\exp\left(\exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{hg}}{A_{\Delta\bar{x}} A_{\Delta\bar{y}}} \exp(A_{\Delta\bar{y}}\bar{y}) \exp(A_{\Delta\bar{x}}\bar{x})\right)} \times \\ &\quad \times \frac{\exp\left(A_{\Delta\bar{y}}^{-2} \exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{\Delta\bar{y}} A_{hg}}{A_{\Delta\bar{x}}} \exp(A_{\Delta\bar{y}}\bar{y}) (C_{1\Delta\bar{x}} A_{\Delta\bar{x}}) + C_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x})) + D_{\Delta\bar{y}}\right)}{\exp\left(A_{\Delta\bar{x}}^{-2} \exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{\Delta\bar{x}} A_{hg}}{A_{\Delta\bar{y}}} \exp(A_{\Delta\bar{x}}\bar{x}) (C_{1\Delta\bar{y}} A_{\Delta\bar{y}}) + C_{\Delta\bar{x}}(\bar{x} - h(\bar{y})) + D_{\Delta\bar{x}}\right)} \\ &= \frac{\exp\left(\exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{hg} C_{1\Delta\bar{x}}}{A_{\Delta\bar{y}}} \exp(A_{\Delta\bar{y}}\bar{y}) + C_{\Delta\bar{y}}\bar{y} + D_{\Delta\bar{y}}\right) \exp(C_{\Delta\bar{x}} h(\bar{y}))}{\exp\left(\exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{hg} C_{1\Delta\bar{y}}}{A_{\Delta\bar{x}}} \exp(A_{\Delta\bar{x}}\bar{x}) + C_{\Delta\bar{x}}\bar{x} + D_{\Delta\bar{x}}\right) \exp(C_{\Delta\bar{y}} \tilde{g}(\bar{x}))} \\ &= \frac{\exp\left(\exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{hg} C_{1\Delta\bar{x}}}{A_{\Delta\bar{y}}} \exp(A_{\Delta\bar{y}}\bar{y}) + C_{\Delta\bar{y}}\bar{y} + D_{\Delta\bar{y}}\right)}{\exp\left(\exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{hg} C_{1\Delta\bar{y}}}{A_{\Delta\bar{x}}} \exp(A_{\Delta\bar{x}}\bar{x}) + C_{\Delta\bar{x}}\bar{x} + D_{\Delta\bar{x}}\right)} \times \\ &\quad \times \frac{\exp\left(-\frac{C_{\Delta\bar{x}} B_{\Delta\bar{y}}}{A_{\Delta\bar{x}}}\right) \left(\frac{\pm A_{\Delta\bar{x}} A_{hg}}{A_{\Delta\bar{y}}}\right)^{-\frac{C_{\Delta\bar{x}}}{A_{\Delta\bar{x}}}} (e^{A_{\Delta\bar{y}}\bar{y}} + C_{1\Delta\bar{y}} A_{\Delta\bar{y}})^{-\frac{C_{\Delta\bar{x}}}{A_{\Delta\bar{x}}}}}{\exp\left(-\frac{C_{\Delta\bar{y}} B_{\Delta\bar{x}}}{A_{\Delta\bar{y}}}\right) \left(\frac{\pm A_{\Delta\bar{y}} A_{hg}}{A_{\Delta\bar{x}}}\right)^{-\frac{C_{\Delta\bar{y}}}{A_{\Delta\bar{y}}}} (e^{A_{\Delta\bar{x}}\bar{x}} + C_{1\Delta\bar{x}} A_{\Delta\bar{x}})^{-\frac{C_{\Delta\bar{y}}}{A_{\Delta\bar{y}}}}}, \end{aligned}$$

implying that

$$\begin{aligned} f_{\bar{y}}(\bar{y}) &= A_{n\Delta\bar{y}} \exp\left(\exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{hg} C_{1\Delta\bar{x}}}{A_{\Delta\bar{y}}} \exp(A_{\Delta\bar{y}}\bar{y}) + C_{\Delta\bar{y}}\bar{y}\right) (e^{A_{\Delta\bar{y}}\bar{y}} + C_{1\Delta\bar{y}} A_{\Delta\bar{y}})^{-\frac{C_{\Delta\bar{x}}}{A_{\Delta\bar{x}}}} \\ f_{\bar{x}}(\bar{x}) &= A_{n\Delta\bar{x}} \exp\left(\exp(B_{\Delta\bar{x}} + B_{\Delta\bar{y}}) \frac{\pm A_{hg} C_{1\Delta\bar{y}}}{A_{\Delta\bar{x}}} \exp(A_{\Delta\bar{x}}\bar{x}) + C_{\Delta\bar{x}}\bar{x}\right) (e^{A_{\Delta\bar{x}}\bar{x}} + C_{1\Delta\bar{x}} A_{\Delta\bar{x}})^{-\frac{C_{\Delta\bar{y}}}{A_{\Delta\bar{y}}}}. \end{aligned}$$

where the constants $A_{n\Delta\bar{y}}$ and $A_{n\Delta\bar{x}}$ incorporate any prefactor that would have cancelled in the ratio $f_{\bar{y}}(\bar{y})/f_{\bar{x}}(\bar{x})$ as well as the constants $\exp(D_{\Delta\bar{y}}) \exp(-C_{\Delta\bar{x}} B_{\Delta\bar{y}}/A_{\Delta\bar{x}}) (\pm A_{\Delta\bar{x}} A_{hg}/A_{\Delta\bar{y}})^{-\frac{C_{\Delta\bar{x}}}{A_{\Delta\bar{x}}}}$ and $\exp(D_{\Delta\bar{x}}) \exp(-C_{\Delta\bar{y}} B_{\Delta\bar{x}}/A_{\Delta\bar{y}}) (\pm A_{\Delta\bar{y}} A_{hg}/A_{\Delta\bar{x}})^{-\frac{C_{\Delta\bar{y}}}{A_{\Delta\bar{y}}}}$, respectively. The constants $A_{n\Delta\bar{y}}$ and $A_{n\Delta\bar{x}}$ are determined by the fact that these

densities must integrate to 1. It can be readily, albeit tediously, verified that it is possible to set the signs of all constants so as to obtain valid densities for all variables. Hence, we have found one special case where Model 1 is not identified. This is case 2 in the statement of Theorem 1.

Case 3 In the special case where $A = 0$ in Equation (20) (not included in Case 2), we let $B_2 = \exp(B)$ and write, for $F = F_{\Delta x}, \tilde{F}_{\Delta y}$:

$$\begin{aligned} F''(u) &= B_2 \\ F(u) &= B_2 u^2 + Cu + D \end{aligned}$$

for some constants B_2, C, D (that differ for $F_{\Delta x}$ and $\tilde{F}_{\Delta y}$) to conclude that $f(u)$ is a normal and therefore that $\Delta\bar{x}$ and $\Delta\bar{y}$ are normally distributed. Since under Model 3 the distribution of $\Delta\bar{x}$ is a factor of the distribution of Δx and under model 4 the distribution of $\Delta\bar{y}$ is a factor of the distribution of Δy , we conclude that either Δx must have a normal factor or Δy must have a normal factor. Next,

$$\frac{|h'(\bar{y})|}{|\tilde{g}'(\bar{x})|} = \frac{|\tilde{F}_{\Delta\bar{y}}''(\bar{y} - \tilde{g}(\bar{x}))|}{|F_{\Delta\bar{x}}''(\bar{x} - h(\bar{y}))|} = B_3$$

where B_3 is the ratio of the constants B_2 obtained for $F_{\Delta x}$ and $\tilde{F}_{\Delta y}$. Rearranging, we obtain

$$|h'(\bar{y})| = B_3 |\tilde{g}'(\bar{x})|$$

and it follows that $h'(\bar{y})$ and $\tilde{g}'(\bar{x})$ must be constant, i.e., that $h(\bar{y})$ and $\tilde{g}(\bar{x})$ are linear. From $\frac{f_{\bar{y}}(\bar{y})}{f_{\bar{x}}(\bar{x})} = \frac{f_{\Delta\bar{y}}(\bar{y} - \tilde{g}(\bar{x}))}{f_{\Delta\bar{x}}(\bar{x} - h(\bar{y}))}$, we can show that $f_{\bar{y}}(\bar{y})$ and $f_{\bar{x}}(\bar{x})$ must also be normal. Either Model 3 or 4 then implies that x^* must be normal. So we recover the more familiar unidentified case 3 in the statement of Theorem 1.

References

- BONHOMME, S., AND J.-M. ROBIN (2006): “Using High-Order Moments to Estimate Linear Independent Factor Models,” Working Paper, University College London.
- CHESHER, A. (1998): “Polynomial Regression with Normal Covariate Measurement Error,” Discussion Paper 98/448, University of Bristol.
- CRAGG, J. C. (1997): “Using higher moments to estimate the simple errors-in-variables model,” *Rand Journal of Economics*, 28, S71–S91.
- DAGENAIS, M. G., AND D. L. DAGENAIS (1997): “Higher Moment Estimators for Linear Regression Models with Errors in Variables,” *Journal of Econometrics*, 76, 193–221.
- ERICKSON, T., AND T. M. WHITED (2000): “Measurement Error and the Relationship between Investment and “q”,” *Journal of Political Economy*, 108, 1027–1057.
- (2002): “Two-step GMM estimation of the errors-in-variables model using high-order moments,” *Econometric Theory*, 18, 776–799.
- GEARY, R. C. (1942): “Inherent Relations Between Random Variables,” *Proceedings of the Royal Irish Academy*, 47A, 63–76.
- HAUSMAN, J., W. NEWEY, H. ICHIMURA, AND J. POWELL (1991): “Measurement Errors in Polynomial Regression Models,” *Journal of Econometrics*, 50, 273–295.
- HU, Y., AND G. RIDDER (2004): “Estimation of Nonlinear Models with Measurement Error Using Marginal Information,” Working Paper, University of Southern California, Department of Economics.
- HU, Y., AND S. M. SCHENNACH (2006): “Identification and estimation of nonclassical nonlinear errors-in-variables models with continuous distributions,” Working Paper, University of Chicago.
- KENDALL, M. G., AND A. STUART (1979): *The Advanced Theory of Statistics*. Macmillan, New York, 4th edition edn.
- KLEPPER, S., AND E. E. LEAMER (1984): “Consistent Sets of Estimates for Regressions with Errors in all Variables,” *Econometrica*, 52, 163–183.

- LEWBEL, A. (1997): “Constructing Instruments for Regressions with Measurement Error when no Additional Data are Available, with an Application to Patents and R&D,” *Econometrica*, 65(5), 1201–1213.
- NEWKEY, W. (2001): “Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models,” *Review of Economics and Statistics*, 83, 616–627.
- PAL, M. (1980): “Consistent moment estimators of regression-coefficients in the presence of errors in variables,” *Journal of Econometrics*, 14, 349–364.
- REIERSOL, O. (1950): “Identifiability of a Linear Relation between Variables Which Are Subject to Error,” *Econometrica*, 18, 375–389.
- SCHENNACH, S. M. (2004a): “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72, 33–75.
- (2004b): “Nonparametric Estimation in the Presence of Measurement Error,” *Econometric Theory*, 20, 1046–1093.
- (2007): “Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models,” *Econometrica*, 75, 201–239.