Identi cation of a Triangular Two Equation System Without Instruments

Arthur Lewbel, Susanne M. Schennach, and Linqi Zhang
Boston College, Brown University, and Boston College

Original September 2020, Revised December 2022

Abstract

We show that a standard linear triangular two equation system can be point identi ed, without the use of instruments or any other side information. We nd that the only case where the model is not point identi ed is when a latent variable that causes endogeneity is normally distributed. In this non-identi ed case, we derive the sharp identi ed set. We apply our results to Acemoglu and Johnson's (2007) model of life expectancy and GDP, obtaining point identi cation and comparable estimates to theirs, without using their (or any other) instrument.

Keywords : Returns to schooling, identi cation, triangular Systems, Kotlarski, deconvolution.
Corresponding Author: Arthur Lewbel, Department of Economics - Maloney 315, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA.i9 althe@(b)-27c.edu,-gle.(co/wb)-28c.edu/aArthu-lLewbe/.

is the firm's labor per unit of capital, and U is unobserved entrepreneurship, which affects both productivity and the chosen level of inputs.

Such models are traditionally identified in econometrics by finding an instrument, i.e., a variable that correlates with Y but not $''_2$, or equivalently, a variable that correlates with V but not U or R. However, such instruments can be difficult to find. For example, Card (1995, 2002) and others propose using measures of access to schooling, such as distance to or cost of colleges in one's area, as wage equation instruments, while others raise objections to

an empirical application where we establish that our identification and estimation strategy is viable even with a very small sample size. Specifically, we estimate the Acemoglu and Johnson (2007) model without using any instruments, and obtain estimates that are very similar to what they found with their instrument.

Instrumental variables estimation of the model has the advantage that it only requires assumptions regarding first and second moments of the covariates, errors, and instruments. In contrast, our assumptions regarding U, V, and R are, implicitly, restrictions on all moments. However, there are a number of mitigating factors. First, some of our results, such as Lemma 1 below, only rely on lower order moments. Second, our main theorem works via convolutions, and so our independence assumptions can be relaxed to subindependence, as defined and described in Schennach (2019), who points out that subindependence is arguably as weak as a conditional mean assumption in terms of the dimensionality of the restrictions imposed. Third, our independence assumption is actually conditional on other covariates, so, e.g., the identification can handle arbitrary heteroskedasticity and dependence of higher moments on regressors. Similarly, if, e.g., U is ability, then identification only requires ability to be conditionally (sub)independent from other unobserved factors, conditional on covariates. Nevertheless, given our required assumptions, these results should be most useful when instruments either don't exist, or might be invalid.

The identification of equations (4) and (5) without instruments has been previously considered by Rigobon (2003), Klein and Vella (2010), and Lewbel (2012), but these results neither nest nor are nested by ours because they *require* that the errors be heteroskedastic, and identification is obtained by imposing varying restrictions on the structure of that heteroskedasticity.[4]

A number of special cases of our results do appear in the literature, but all of them assume $= 0$, and so they omit the most important feature of the model in applications like ours.

[4] Rigobon (2003) and Klein and Vella (2010) impose different parametric restrictions on the error variances, while Lewbel (2012) imposes a nonparametric restriction. For simplicity we assume homoskedastic errors, but by conditioning our identification theorems on X, we could allow for general heteroskedastity as well, at the expense of likely weaker identification and more complicated estimators.

Kotlarski (1967) is the special case of our model where it is known that = 0 and = 1, and in that case Kotlarski's Lemma shows that point identi cation of the distribution of all the latent variables holds even under normality. Similarly, Reiers l (1950) uses a special case of our model where it is known that = 0 and Y plays the role of a measurement of U contaminated by an error V and establishes conditions under which would be identi ed. As noted in Lewbel (2020), with = 0 and Reiers l's identi cation of , one could rewrite Reiers l's model as $Y = U + V$ and $W/ = U + R/$, and then apply Kotlarski's lemma to the joint distribution of Y and $W/$ to identify the distributions of U, V, and R.[5]

Our results, showing necessary and su cient conditions to identify the more general model of equations (4) and (5) with unknown nonzero , turns out to be a di cult extension. In particular, the methods of proof used by Reiers l (1950) and Kotlarski (1967) do not extend to our problem. The proof of our main result instead relies on similar tools as Khatri and Rao (1972) or Rao (1966, 1971) (see also Comon's (1994) reference to Darmois (1953)).

Some limitations of our results should be acknowledged upfront. We assume that the coe cients and are constants. So, e.g., our results do not immediately extend to random coe cients, such as treatment e ects with unobserved heterogeneity, or to nonlinearity in the dependence of W on Y. However, this limitation may be mitigated to some extent by allowing the distributions of the unobservables to be unknown functions of covariates. Another important restriction on our results is that we require U to be a scalar. While this is a common assumption (as in the examples cited earlier), there are other situations where one might expect a vector of unobservable shocks like U to a ect both Y and W, and our identi cation results would then not apply. We provide examples in Supplement D. Finally,

[5]A special case of non-normality is when the components U and V are asymmetric. Lewbel (1997) and Erickson and Whited (2002) exploit asymmetry to construct simple estimators for the Reiers l (1950) model. See also Bierens (1981). Other papers propose estimators for models like equations (4) and (5) with = 0, by assuming that coe cients like are point identi ed using higher moments, but without explicitly characterizing when that is possible. Examples include Bonhomme and Robin (2010), Fruehwirth, Navarro, and Takahashi (2016), and Navarro and Zhou (2017). A related result, showing identi cation of direction of causality in models under nonnormality, is Peters, Janzing, and Scholkopf (2017). Generalizations of Kotlarski's lemma to models with more components (but again still assuming = 0) include Szekely and Rao (2000) and Li and Zheng (2020). A nonlinear extension of Reiers l (1950) is Schennach and Hu (2013).

a limitation for empirical work is that our estimators depend on higher than second moments of the data, and such moments can lead to very imprecise estimates when sample sizes are small.

In section 2, we provide a few simple moments that will often suffice to point identify our model, and can be used to construct a correspondingly simple GMM estimator. In Section 3, we present our general identification results, including constructing more moments like those in Section 2, and showing that, with minimal regularity, the model is point identified as long as both U and V are not normal. In sections 4 and 5 we derive the sharp identified set when either U or V is normal, and derive some inequalities regarding our model relative to ordinary least squares. Section 6 provides a general pro

Assumption 2 *The unobserved real valued random variables U, V, and R are mean zero and mutually independent,*[6] *with unknown distributions.*

Assumption 3 *R has nite variance, and U and V each have nite fourth moments.*

Assumption 4 *The unknown constants and are real valued, nite, and > 0.*

We can assume our data Y and

Proofs are all in Supplement A. The proof of Lemma 1 works by substituting $W \quad Y =$ $U + R$ and $W \quad Y = \quad V + R$ into equations (7) and (8), and then uses the mutual independence of U, V, and R to verify that these equations hold.

Lemma 1 provides two equations in the two unknowns and . If we solve the rst equation for and substitute that into the second, we obtain a quadratic in . The sign restriction that > 0 then determines which root is the correct one for .

We later provide the formal conditions under which these two equations su ce to point identify and . The main condition, derived in Theorem 1 below, is equation (21). Equation (21) shows that the main cases in which equations (7) and (8) by themselves fail to provide point identi cation are when U and V have the exact same distribution, or when both are symmetrically distributed, or if either U or V is normally distributed. We later show that in nitely many additional equations in , , Y and W can be constructed, based on higher moments of Y and W than those used in Lemma 1. These higher moments can help identify and in applications where Lemma 1 does not su ce.

A simple estimator for and can be constructed by rewriting equations (7) and (8) as moment conditions, and applying standard method of moments or GMM. One can immediately check that these equations take the form

$$E(YW \quad _{yw}) = 0, \qquad E(Y^2 \quad _{yy}) = 0 \tag{9}$$

$$E[(W \quad Y)(W \quad (\ + \ )Y)Y] = 0 \tag{10}$$

$$E \ (W \quad Y)(W \quad (\ + \ )Y) \ Y^2 \quad _{yy} \quad 2(\ _{yw} \quad _{yy})(W \quad (\ + \ )Y)Y \ = 0 \tag{11}$$

where $_{yw} = E(YW)$ and $_{yy} = E(Y^2)$. The parameters $_{yw}$ and $_{yy}$ are estimated along with and by putting equations (9), (10), and (11) into any standard GMM estimation routine. One could replace with e^b in these equations to impose the sign restriction that > 0.

Lemma 1 uses up to fourth moments of the data. Based on results derived in the next section, in Supplement B we provide additional equations (using up to fth moments) that can provide overidenti cation of and , or point identi cation in some cases where Lemma 1 does not su ce.

Let $_U^2$, $_V^2$, and $_R^2$ denote the variances of the error components U, V, and R. It may be of economic interest to estimate these variances, to identify how much of the variance of the model errors is due to unobserved ability U versus the idiosyncratic components V and R. From the model we have $E((W \quad Y)Y) = \quad _U^2$, $E(Y^2) = {}_U^2 + {}_V^2$, and $E\left((W \quad Y)^2\right) = {}^2\,{}_U^2 + {}_R^2$, which implies

$$_U^2 = E((W \quad Y)Y) = \;, \qquad _V^2 = E\left(Y^2\right) \quad _U^2, \qquad _R^2 = E\left((W \quad Y)^2\right) \quad {}^2\,{}_U^2 \tag{12}$$

Given estimates of and , we can replace the expectations in equation (12) with sample averages to estimate these variances.

Alternatively, we can estimate these variances jointly with the model parameters by observing that

$$_{yy} = {}_U^2 + {}_V^2, \qquad _{yw} = \quad _U^2 + \quad _U^2 + {}_V^2 \; : \tag{13}$$

So, in equations (9), (10), and (11) we can replace $_{yy}$ and $_{yw}$ with their expressions in equation (13), and apply GMM using those equations along with the additional equation

$$E\left((W \quad Y)^2 \quad {}^2\,{}_U^2 \quad {}_R^2\right) = 0 \tag{14}$$

to simultaneously estimate , , $_U^2$, $_V^2$, and $_R^2$. We can further replace $_U^2$ with $_U^2 = e^{\;U}$ and similarly for $_V^2$ and $_R^2$, to impose the constraint that variances are positive. See Supplement B for details on these moments.

Higher moments of U, V, and R can be estimated analogously. Alternatively, as discussed later, once we have identi ed and estimated and , we can apply Kotlarski's Lemma to recover the entire distributions of U, V, and R.

We can also easily extend this identi cation and associated estimation to allow for covariates. Suppose we have the model

$$Y = b_1^0 X + U + V \tag{15}$$

$$W = \quad Y + b_2^{0} X + \quad U + R \tag{16}$$

where X is exogenous and is therefore uncorrelated with U, V, and R. The reduced form for W is now

$$W = (\quad b_1 + b_2)^{0} X + (\quad + \quad) U + \quad V + R$$

So we can estimate the coe cient vectors b_1 and b_2 along with and by replacing Y and W in equations (9), (10), and (11) with $Y \quad b_1^{0} X$ and $W \quad (\quad b_1 + b_2)^{0} X$, respectively and estimate those moments along with the moments

$$E \quad W \quad (\quad b_1 + b_2)^{0} X \quad X \quad = 0, \qquad E\left((Y \quad b_1^{0} X) X \right) = 0 \tag{17}$$

The complete set of moments for estimating this model via GMM, which we use in our empirical application, is provided in Supplement B.

Although we did not nd this to be the case in our application, when GMM models are substantially overidenti ed (many more moments then parameters) it is sometimes preferable to only use a subset of available moments for estimation. Since our estimator takes the form of standard GMM, in these cases the existing literature on empirical choice of moments in standard GMM estimation might be applied. See, e.g., Andrews and Lu (2001), Caner (2009), and Liao (2013).

For simplicity, these estimators assumed the errors U, V, and R are homoskedastic, and similarly have higher moments that do not depend on X. This could be relaxed to allow higher moments of these errors to depend in unknown ways on X, by letting the assumptions of Lemma 1 hold conditional on X, thereby replacing the unconditional moments of equations (7) and (8) with conditional moments. Corresponding estimators would then, however, be much more complicated, and parameters like the error variances would need to be replaced by nonparametric functions of X.

3 General Point Identi cation

We now provide a more general and systematic analysis of the identi cation of our model, using more information than the low order moments of Lemma 1. We provide four main results. First, we show that it is possible to construct infinitely many moments like those of Lemma 1, which can be used to construct simple GMM estimators, and we give the conditions under which these moments point identify the coefficients and (equivalently, and). Second, we apply Kotlarski's lemma to point identify the distributions of U, V, and R given point identification of and . Third, we demonstrate that, using the entire joint distribution of Y and W (instead of just some moments) the only case where point identification is not possible is when U or V (or both) are normal. Finally, in the not point identified case, we fully characterize the sharp identified set.

We make extensive use of the characteristic function and its logarithm. Knowing the (log) characteristic function of a vector of random variables is equivalent to knowing the joint distribution of those variables (Theorem 3.1.1 in Lukacs (1970)).

Definition 1 *Given two random variables Y and W, let $_{Y;W}(\;;\;)$ $E\; e^{i\;Y+i\;W}$ denote their joint characteristic function. Similarly for a single random variable, let $_{Y}(\;)$ $E\; e^{i\;Y}$. Moreover, let $_{Y;W}(\;;\;)$ $\ln\; _{Y;W}(\;;\;)$ and $_{Y}(\;)$ $\ln\; _{Y}(\;)$ denote log characteristic functions (which are also called cumulant generating functions).*

Definition 2 *Given two random variables Y and W, define the cumulant of order $k;\ell$ (Lukacs (1970), p. 27) as*

$$^{k;\ell}_{Y;W} \quad \left.\frac{\partial^{k+\ell}\; _{Y;W}(\;;\;)}{i^{k+\ell}\partial\; ^{k}\partial\; ^{\ell}}\right|_{=0;\;=0}.$$

Similarly for a single random variable, define the cumulant of order k as

$$^{k}_{Y} \quad \left.\frac{\partial^{k}\; _{Y}(\;)}{i^{k}\partial\; ^{k}}\right|_{=0}.$$

All cumulants can be expressed in terms of standard moments, as obtained by an explicit differentiation of the log characteristic function and by exploiting the characteristic function moment theorem hose) characteristic functions as well as the corresponding cumulants are directly related, e.g., $_{Y}(\ ) = \ _{Y;W}(\ ;0)$, $_{Y}(\ ) = \ _{Y;W}(\ ;0)$ and $^{k}_{Y} = \ ^{k;0}_{Y;W}$.

With these tools in hand, we are ready to state a general identification result based on moment constraints. As in Lemma 1, we start by rewriting the model of equations (4) and (5) in the reduced form of equations (4) and (6), and focus on the parameters and .

Theorem 1 *Let Assumptions 1, 2, and Equations (4) and (6) hold. Assume $1 < \ < \ < 1$ and let*

$$M_p(\ ;\ ) \quad ^{1+p;2}_{Y;W} \quad ^{2} \ ^{3+p}_{Y} \quad (\ +\ ) \ ^{2+p;1}_{Y;W} \quad ^{3+p}_{Y} : \tag{18}$$

Let $q;\bar{q} \in \mathbb{N} \setminus \{0;1;\ldots\}$ with $q < \bar{q}$. If $E\left[|U|^{\bar{q}}\right]$, $E\left[|V|^{\bar{q}}\right]$ and $E\left[|R|^{\bar{q}}\right]$ exist and $^{3+\bar{q}}_{Y} \ ^{2+q;1}_{Y;W} \neq \ ^{3+q}_{Y} \ ^{2+\bar{q};1}_{Y;W}$ (or, equivalently, if $^{3+\bar{q}}_{U} \ ^{3+q}_{V} \neq \ ^{3+\bar{q}}_{V} \ ^{3+q}_{U}$), then the moment constraints

$$M_q(\ ;\ ) = 0 \tag{19}$$

$$M_{\bar{q}}(\ ;\ ) = 0 \tag{20}$$

point identify the parameters of the model as $(\ ;\ ) = (\ _{+};\ )$, where

$$= \frac{F^{3012}}{2F}$$

Corollary 2 *The assumptions of Theorem 1 with* $q = 0$ *and* $\tilde{q} = 1$ *imply that the assump-*

and $M_1(\;;\;) = 0$, requires that $^4_U\,^3_V \neq \;^4_V\,^3_U$, or equivalently

$$E\;U^4 \quad 3\;E\;U^2\;{}^2\;E\;V^3 \qquad E\;V^4 \quad 3\;E\;V^2\;{}^2\;E\;U^3 \neq 0. \qquad (21)$$

The left-hand side of (21) turns out to be proportional to the determinant of the Jacobian of the moment conditions (7) and (8) evaluated at the true value of the parameters:

$$\begin{matrix} E[V^3] & E[U^3] \\ E[V^4] \quad 3(E[V^2])^2 & E[U^4]+3(E[U^2])^2 \end{matrix} \; . \qquad (22)$$

This connection is expected, since having a nonsingular Jacobian at the true parameter values is a necessary condition for point identi cation.

Condition (21) is violated, for instance, if either U or V is normal, or if both U and V are symmetric, or if both U and V have the exact same distribution. If we add the additional moments corresponding to $M_2(\;;\;) = 0$, then point identi cation only requires that at least one of the inequalities $^4_U\,^3_V \neq \;^4_V\,^3_U$, $^5_U\,^3_V \neq \;^5_V\,^3_U$, or $^5_U\,^4_V \neq \;^5_V\,^4_U$, hold. For example, if the second of these holds then Theorem 1 applies with $q = 0$ and $\bar{q} = 2$. If more than one of these inequalities holds, then we are generally overidenti ed.

Once the parameters and have been identi ed, the full distribution of all unobservables can be determined under the following Assumption.[8]

Assumption 5 *The characteristic functions of $U; V$ and R are nonvanishing on the real line.*

Corollary 3 *If Assumptions 1, 2, 5 and Equations (4) and (6) hold, $E[|Y|] < 1$ and if ; are point identi ed, then the distributions of U, V and R are point identi ed from the joint distribution of Y and W through*

$$_V(\;) = \int_0 \frac{E\left[iYe^{i\frac{W\quad Y}{}}\right]}{E\left[e^{i\frac{W\quad Y}{}}\right]}d$$

A more explicit expression for the distributions of these unobserved variables can be obtained by an inverse Fourier transform. For instance, if V admits a density, it is given by

$$f_V(v) = (2\ \ )^{1} \int_{1}^{1} \exp(\ \ _V(\ \ ))\, e^{\ i\ \ v} d \tag{24}$$

and similarly for the other densities. More general distributions (e.g. discrete and/or singular) can be recovered as well, if equation (24) is interpreted in the appropriate measure theoretic sense.

Although Theorem 1 is quite general, it does require the condition $\ _U^{3+\tilde{q}}\ \ _V^{3+q} \neq \ \ _V^{3+\tilde{q}}\ \ _U^{3+q}$ to deliver identi cation, so it is natural to ask whether this is fundamentally necessary. It is in fact possible to formulate an estimation strategy that relaxes this condition. For instance, as discussed above, one could stack the moment conditions of the form (19) and (20) obtained with di erent values of $(q;\tilde{q})$. The resulting moment conditions would only fail to identify $(\ ;\ )$ if the condition $\ _U^{3+\tilde{q}}\ \ _V^{3+q} \neq \ \ _V^{3+\tilde{q}}\ \ _U^{3+q}$ fails simultaneously for all the choices of q and $\tilde{q}$ considered.

An even more general strategy could be to start from the fundamental relationships between the log characteristic functions of the observables and unobservables ($\ _{Y;W}(\ ;\ ) = \ _U(\ \ +\ \ \ ) + \ _V(\ \ +\ \ \ ) + \ _R(\ \ )$) and cast identi cation as an optimization problem that minimizes deviations between the observed quantities (i.e. $\ _{Y;W}(\ ;\ )$) and predicted quantities:

$$(\ ;\ ;\ \ _U;\ \ _V;\ \ _R) \tag{25}$$

$$=\ \arg\min_{(\ ;\ ;}$$

An estimator based on Equation (25) would be obtained replacing $\ _{Y;W}(\ ;\ )$ by its sample analogue and trimming or downweighting the high-frequency tails in the integral.

The question remains, do there exist situations where neither this nor any other estimator can consistently estimate the model, due to lack of point identification? The following theorem fully addresses this question, by showing that there exist cases that are not point identified. However, all such cases are when U or V (or both) are normal.

This differs from, and is simpler than, Reiersøl's (1950) well-known result in linear univariate errors-in-variables models, where the nonidentified cases arise when the model contains normal factors (see below). However, the required methods of proof differ significantly. For instance, the presence of two slope parameters and (instead of one), and the presence of both latent variables U and V in both equations of the model, prevents us from using Reiersøl's proof method, which is based on the fact that two functions of different variables that are equal to each other must be constant. In our case, we have sums of many different functions of different variables on each side of an equality, and possible cancellation between terms that complicates the argument significantly.

Assumption 6 *$E\ |U|^3\ ; E\ |V|^3\ ; E\ |R|^3$ are finite.*

Theorem 4 *Let Assumptions 1, 2, 5, 6 and Equations (4) and (6) hold and assume that $1 < \ < \ < 1$. If neither U nor V are normally distributed, then $\ ;\ $ are uniquely determined by the joint distribution of Y and W by Equation (25).*

Note that U or V normal implies Y has full real line support, so having the support of Y be bounded is a simple sufficient condition for point identification. In the next section, we address what happens when either U or V (or both) are normally distributed.

4 Set Identification

In the case where Theorem 4 does not apply, so that the parameters are not point identified, the objective function of Equation (25) is maximized over a set rather than at a single point. In order to precisely characterize this *identi ed set* exp (^{0} =2) and checking if the result is a valid characteristic function (e.g., by verifying if the inverse Fourier transform is a nonnegative measure). An alternative check for the validity of a given function ( ) to be a valid characteristic function can be based on Bochner's Theorem (Theorem 4.2.2 in Lukacs (1970)): is a characteristic function iff

$$\sum_{i=1}\sum_{j=1} c_i c_j \left({}_i \quad {}_j \right) \geq 0 \text{ for all } c_1, \ldots, c_n \in \mathbb{C} \text{ for all } {}_1, \ldots, {}_n \in \mathbb{R} \text{ for all integer } n \geq 1$$

(Bochner's Theorem also includes the conditions that ( ) be continuous and (0) = 1 but these are automatically satisfied in our context.)

Using Lemma 2, we can decompose the observed $Z = (Y; W)$ into Gaussian (g) and non-Gaussian (n) factors

$$(Y; W) = (Y_g; W_g) + (Y_n; W_n) \tag{27}$$

This decomposition can be accomplished without the knowledge of or . The non-Gaussian or Gaussian nature of the two factors is important in our context, because it is associated with the features that can or cannot be point-identified. This type of decomposition is not a purely theoretical construct; it can be empirically implemented. Independent Component Analysis techniques, which are widely used in signal processing, (see Hyvärinen and Oja (2000) for a review) specifically rely on such decompositions into Gaussian and non-Gaussian components.

Define

$$B_s = \frac{E[W_s Y_s]}{E[Y_s^2]} \tag{28}$$

$$D_s = \frac{E[W_s^2]\,E[Y_s^2] - (E[W_s Y_s])^2}{(E[Y_s^2])^2} \geq 0 \tag{29}$$

where the subscript s is either set to "g", or to "n", or is removed. We can now state our set-identification theorem:

Theorem 5 *Let Assumptions 1, 2 and Equations (4) and (6) hold and assume that $E[Y^2]$, $E[W^2]$, $E[R^2] < \infty$ and that $-\infty < \quad < \quad < \infty$. Then, the following bounds (illustrated in Figure 1) are sharp:*

1. *If both U and V are Gaussian (and* $E[Y^2] > 0$*), then*

$$B_g \qquad (30)$$

$$B_g \quad \frac{D_g}{B_g} \quad B_g: \qquad (31)$$

2. *If V*

This looser bound is also related to the measurement error bounds in Frisch (1934). If one is willing to rely on this relaxed bound, then a simple GMM estimator for the resulting identi ed set could be obtained based on the moment conditions

$$E \quad {}^2 \; {}^2_U + \; {}^2 \; Y^2 \quad {}^2_U \; + \; {}^2_R \quad W^2 \quad = \quad 0 \tag{36}$$

$$E \quad {}^2_U + \quad Y^2 \quad {}^2_U \quad YW \quad = \quad 0 \tag{37}$$

while optimizing over $\;;\;;\; {}^2_U;\; {}^2_R$, subject to the constraints $\; < \;$ (equivalent to $\; > 0$), ${}^2_U \quad 0$ and ${}^2_R \quad 0$. These moment conditions are obtained from Equations (66) and (67) in the proof of Theorem 5, without extracting the Gaussian parts. The bounds of Corollary 6 are also obeyed in the case of point identi ed models, since they are obtained solely from

V/F38 11.9552 Tf 11.761(26 Td [(2)]T31Td [(2)]5W)]TJ
Wa es< 2 28(one)-38048can-pro of Tb(6(the)-327pim]TJ edi)-3v [((ofr326(obtaine)-32T8isditionsbserv1(The)/F38 at73(<)-358( )]TJ/F2115 74(o)27

heterogeneity). In particular, in the returns to schooling context, we would expect both and to be positive (because unobserved ability U should a ect schooling Y and wages W in the same direction, and increased schooling should increase wages). By the above analysis, this in turn means that we would expect $0 <$ B.

However, as noted by Card (2001), most returns to schooling empirical applications yield estimates of , using instrumental variables methods, that are greater than B, which contradicts this inequality and hence also contradicts the model. One possible explanation for this contradiction is that, in the returns to schooling context, Y may also contain signi cant measurement error. Standard attenuation bias under classical measurement error implies that the ordinary least squares coe cient B is biased towards zero relative to , which if $0 <$ would imply $B <$. If the model is correct for returns to education, but in addition Y is mismeasured, then B could be either larger or smaller than , depending on the relative magnitude of the measurement error.

6 Monte Carlo

corresponding to Lemma 1, given by equations (77), (78), and (79) (without covariates, so $\hat{Y} = Y$ and $\hat{W} = W$), as given in Supplement B. The over-identified estimator is GMM using these same equations, plus equations (81) and (82) of Supplement B.

Tables C1 to C4 of the Supplement report results from designs 1 to 4, respectively. Each Table has four panels, corresponding to the two different GMM estimators, each with the two different sample sizes. We report estimates of , , the error component variances ${}^2_U$, ${}^2_V$, and ${}^2_R$, and, when over-identified, ${}_{WW}$. Reported summary statistics of each parameter estimate across the simulations are the mean (MEAN), the standard deviation (SD), the 25% quantile (LQ), the median (MED), the 75% quantile (UQ), the root mean squared error

Well (2007), Lorentzen, McMillan, and Wacziarg (2008), Aghion, Howitt, and Murtin (2010), Cervellati and Sunde (2011), Ecevit (2013), Bloom, Canning, and Fink (2014), and Bloom, Canning, Kotschy, Prettner, and Schünemann (2019).

Based on a neo-classical growth model, AJ estimate a model in the form of equations (1) and (2), where Y

this application, with only 47 countries. Nevertheless, AJ's estimates of are statistically significant.[12]

Now suppose we had not observed predicted mortality, or we are uncertain of its validity as an instrument. We can instead consider applying our GMM estimators. First, consider the distribution of Y. Assuming (measured) life expectancy is bounded away from zero, log life expectancy is bounded, which suffices for point identification since it rules out U or V being normal.[13] We therefore attempt to apply our GMM estimators.

In Table 1, we report two sets of GMM estimates along with AJ's 2SLS results. Columns labeled GMM1, GMM2, and GMM3 are GMM estimates of equations (15) and (16), which do not make use of the predicted mortality instrument in any way. Specifically, these are estimates based on the over-identifying set of moments given by equations (77) to (82) in Supplement B. The last three columns of Table 1 then give GMM estimates that use both our over-identifying set of moments and the additional moment given by AJ's instrument (as discussed at the end of Supplement B).[14]

Panel A in Table 1 reports the main parameter of interest , and also reports b_2, the other covariate coefficients in equation (16). The variables in columns (4) and (7) have been demeaned so there is no constant.[15] Our main takeaway from Panel A of Table 1 is that our estimates of are quite comparable to AJ's. In GMM1 and GMM2, the estimates of are 1.984 and 1.241, virtually the same range as AJ's 2SLS estimates, and are

[12] Our standard errors in columns (1)-(3) of Table 1 differ from those reported by AJ. AJ's estimates are from ivreg in Stata 9. We used ivregress 2sls, which replaced ivreg as of Stata 10. ivreg and ivregress can give different robust standard error estimates, because ivreg uses HC1 (MacKinnon and White 1985) robust standard errors while ivregress 2sls uses HC0 (Huber-White). Also, to reduce the number of coefficients in GMM estimation, we differenced the data while AJ used level data with fixed effects. Since T=2, these are asymptotically equivalent estimators.

[13] More heuristically, if Y is close to normal, then it may be that U or V is close to normal. Y has a skewness of 0.170 and a kurtosis of 1.791, which is reasonably far from normal in terms of the low order moments our GMM estimator is based on. The p-value of a Shapiro-Wilk test of normality of Y is $.02$, rejecting normality, and even lower if one tests the residuals after regressing Y on either of the covariates in X.

[14] These GMM models are estimated in Stata, using the vce(robust) option to compute standard errors.

[15] In Supplement B: Moments for GMM Estimation, it is noted that "For the model without covariates, one can replace b_1 and b_2 with zero in the above expressions, and drop equation (80). Note that in this case Y and W should be demeaned." In columns (4) and (7), we demeaned Y and W so b_1 and b_2 are zeros.

statisically significant. GMM3 gives an estimate of a lower magnitude 0.383, but this estimate is statistically insignificant with a very large standard error, suggesting that our higher moment based estimator is imprecise for this particular combination of covariates and small sample size. The last three columns of Table 1, which combine both our moments and the AJ instrument, give estimates very close to those of AJ, with somewhat smaller standard errors, which is exactly what one would expect to see if both sets of moments are valid and if AJ's instrument is strong. In the bottom row of Table 1 we report Hansen's J-test; we do not reject validity of the joint set of overidentifying restrictions in any of the GMM estimates.

Panels B and C of Table 1 provide the other estimated parameters of the model. Panel C gives the estimated b_1 coefficients from equation (15), while Panel B gives the estimates of and the estimated variances of our error components. appears to be difficult to precisely estimate, with large standard errors.[16] In the specifications where is statistically significant, the variance of U (the source of endogeneity in the model) is much smaller than the variances of the idiosyncratic components V and R, but very precisely estimated with small standard errors.

Later tables have the same format as Table 1, providing additional results. In Table 2, we re-estimate the model using the exactly identified set of moments from Lemma 1. As expected with fewer moments, these estimates are less efficient, and turn out to be quite a bit noisier than those of Table 1. GMM5, with the quality of institutions as the covariate, is still reasonably comparable to AJ with of 1.401, while now both GMM4 and GMM6 are insignificant and more variable. The estimates combining these moments with AJ's instrument behave as before.

We also perform a number of robustness checks in Supplement D, using alternative outcome variables that AJ considered in their Tables 8-9. These additional outcomes are log population, log births, percentage of population under age 20, log GDP, and log GDP per working age population. Some of the alternative outcomes suffer from the issue that U might

[16] In contrast is, like , much more precisely estimated, but apparently the difference is harder to pin down.

also contain measurement error, and in those cases, our identi cation results would not apply. The results of our GMM estimators with other outcomes are generally more erratic than with log per capita GDP. The estimates that combine our moments and the AJ instrument remain comparable to AJ's 2SLS estimates.

We conclude that, in all speci cations where the standard errors were small enough to yield statistically signi cant results, our estimates based on higher moments, without side information, are very close to those obtained by AJ that required an instrument.

8 Conclusions

We have shown that a standard linear triangular structural model is generally point identi ed, without an instrument or other side information that is generally used to identify such models. We illustrate the result with Monte Carlo simulations and in an empirical application. Our application shows that, without using an instrument, GMM estimation of moments based on the model yields estimates close to those that were obtained by previous authors using an instrument. Even when instruments are available, our estimator could be usefully combined with instrument based moments to either increase estimation precision by adding more moments to the model, or to provide overidentifying moments that might be used for speci cation testing.

What makes point identi cation possible is the assumed error structure, which takes the standard form of a scalar common component U in each equation, plus additional scalar idiosyncratic components V and R. One goal for future work could include deriving alternative estimators for the model. These could include estimators that allow U, V, and R to depend nonparametrically on covariates X (e.g., allowing heteroskedasticity of unknown form), and estimators that make direct use of all the information in Theorem 4, perhaps based directly on characteristic functions rather than moments. Other possibilities for further work include extending the model to more equations, allowing the common component U to a ect outcomes nonlinearly, and extending the model to also allow for measurement error in Y. Based on Card (2001), this last extension would likely be needed for returns to education applications.

Acknowledgments

Susanne Schennach acknowledges support from NSF grant SES-1950969. Vincent Starck is gratefully acknowledged for valuable comments.

Disclosures

The authors report there are no competing interests to declare.

References

Acemoglu, D., and S. Johnson, (2007), \Disease and development: The e ect of life expectancy on economic growth," Journal of Political Economy, 115(6), 925-985.

Aghion, P., P. Howitt, P., and F. Murtin, (2010), \The relationship between health and growth: When Lucas meets Nelson-Phelps, " National Bureau of Economic Research.

Andrews, D.W.K. and Lu, B. (2001): "Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models", Journal of Econometrics, 101(1), 123-164.

Angrist, J. and A. Krueger, (1991), \Does Compulsory School Attendance A ect Schooling and Earnings?" Quarterly Journal of Economics, 106, 979-1014.

Angrist, J. and A. Krueger, (2001), \Instrumental Variables and the Search for Identi cation: From Supply and Demand to Natural Experiments," Journal of Economic Perspectives, 15, 69-85.

Bierens, H. J. (1981) Robust Methods and Asymptotic Theory in Nonlinear Econometrics, Springer Verlag

Bloom, D. E., D. Canning, and G. Fink (2014), \Disease and development revisited ," Journal of Political Economy, 122(6), 1355-1366.

Bloom, D. E., D. Canning, R. Kotschy, K. Prettner, and J. J. Schunemann (2019), \Health and Economic Growth: Reconciling the Micro and Macro Evidence," NBER Working Paper No. 26003.

Bonhomme, S. and J. - M. Robin (2010), \Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," The Review of Economic Studies, 77, 491{533.

Caner, M. (2009): "Lasso-type GMM Estimator", Econometric Theory, 25(1), 270-290.

Card, D (1995) \Using Geographic Variation in College Proximity to Estimate the Return to Schooling," In L.N. Christo des, E.K. Grant, and R. Swidinsky, editors, Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp , University of Toronto Press, Toronto.

Card, D. (2001), \Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," Econometrica, 69, 1127{1160.

Carneiro, P. and J. J. Heckman, (2002), \The Evidence on Credit Constraints in Post-Secondary Schooling," The Economic Journal, 112(482), 705{734.

Cervellati, M., and U. Sunde, (2011), \Life expectancy and economic growth: The role of the demographic transition,"Journal of economic growth, 16(2), 99-133.

Comon, P. (1994), \Independent component analysis, a new concept?," Signal processing, 36(3), 287-314.

Darmois, G. (1953), \Analyse generale des liaisons stochastiques: etude particuliere de l'analyse factorielle lineaire," Revue de l'Institut international de statistique, 2-8.

Fruehwirth, J. C., S. Navarro, and Y. Takahashi (2016), \How the Timing of Grade Retention A ects Outcomes: Identi cation and Estimation of Time-Varying Treatment Effects," Journal of Labor Economics 34:4, 979-1021

Erickson, T. and T. M. Whited, (2002), \Two-step GMM estimation of the errors-in-variables model using high-order moments," Econometric Theory, 18(3), 776-799.

Evdokimov, K. and H. White (2012), \Some Extensions of a Lemma of Kotlarski," Econometric Theory, 28(4), 925{932.

Frisch, R. (1934), \Statistical con uence analysis by means of complete regression systems," Vol. 5, Universitetets konomiske Instituut.

Hyvarinen, A. and E. Oja (2000) \Independent component analysis: algorithms and applications," Neural Networks 13, 411{430.

Khatri, C. and Rao, C. R. (1972), \Functional equations and characterization of probability laws through linear functions of random variables," Journal of Multivariate Analysis 2, 162{173.

Klein, R., and F. Vella, (2010), \Estimating a class of triangular simultaneous equations models without exclusion restrictions," Journal of Econometrics 154(42), 154{164.

Kotlarski, I. I. (1967), \On characterizing the gamma and normal distribution," Paci c Journal of Mathematics, 20, 69{76.

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.

Lewbel, A. (2012), \Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models," Journal of Business and Economic Statistics, 30, 67{80.

Lewbel, A. (2020), \Kotlarski With a Factor Loading," Unpublished Manuscript, Boston College.

Li, S. and X. Zheng, (2020), \A Generalization of Lemma 1 in Kotlarski (1967)," Statistics and Probability Letters,165, article 108814.

Li, T. and Q. Vuong (1998), \Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," Journal of Multivariate Analysis, 65, 139{165.

Liao, Z. (2013): "Adaptive GMM Shrinkage Estimation With Consistent Moment Selection", Econometric Theory, 29(5), 857-904.

Lorentzen, P., J. McMillan, and R. Wacziarg, (2008), \Death and development," Journal of Economic Growth, 13(2), 81{124.

Lukacs, E. (1970), \Characteristic Functions," Second edition, Gri n, London.

Maddison, A. (2003), \Development centre studies the world economy historical statistics: Historical statistics," OECD Publishing.

Navarro, S. and J. Zhou, (2017), \Identifying agent's information sets: An application to a lifecycle model of schooling, consumption and labor supply," Review of Economic Dynamics, 25, 58-92.

Peters, J., D. Janzing, and B. Scholkopf, (2017), \Elements of causal inference: foundations and learning algorithms," MIT press.

Rao, C. R. (1966), \Characterisation of the distribution of random variables in linear structural relations," Sankhya: The Indian Journal of Statistics, Series A, 251-260.

Rao, C. R. (1971), \Characterization of probability laws by linear functions," Sankhya: The Indian Journal of Statistics, Series A, 265-270.

Reiers l, O. (1950), \Identi ability of a linear relation between variables which are subject to error," Econometrica, 18, 375-389.

Rigobon, R. (2003), \Identi cation Through Heteroskedasticity," Review of Economics and Statistics 85(4), 777{792.

Schennach, S. M. (2000), \Estimation of nonlinear models with measurement error," Working Paper, University of Chicago.

Schennach, S. M. (2019), \Convolution without independence," Journal of Econometrics, 211(1), 308-318.

Schennach, S. M. and Y. Hu (2013), \Nonparametric Identi cation and Semiparametric Estimation of Classical Measurement Error Models Without Side Information," Journal of the American Statistical Association, 108, 177-186.

GJ Szekely, G. J. and C. R. Rao (2000), \Identi ability of distributions of independent random variables by linear combinations and moments," Sankhya: The Indian Journal of Statistics, Series A, 62(2), 193-202.

Well, D. N. (2007), \Accounting for the E ect Of Health on Economic Growth," The Quarterly Journal of Economics, 122(3), 1265{1306.

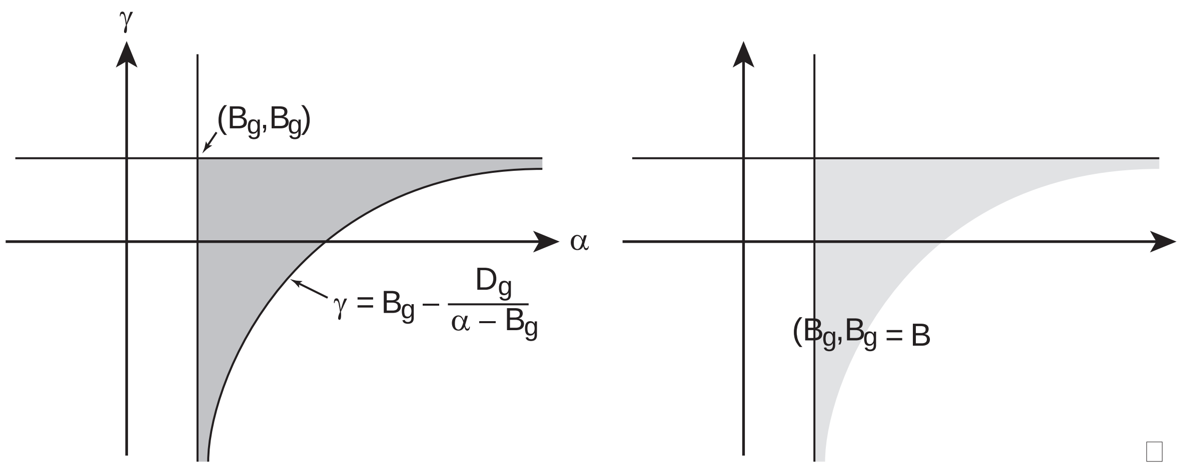

Figure 1: Identi ed set of Theorem 5 for (a) Case 1 and (b) Case 2 (Case 3, analogous to Case 2, is not shown).

Table 1: Over identi ed moments: Base sample 1940 and 1980

	(1) 2SLS1	(2) 2SLS2	(3) 2SLS3	(4) GMM1	(5) GMM2	(6) GMM3	(7) GMM1+AJ	(8) GMM2+AJ	(9) GMM3+AJ
Panel A. Dependent Variable: Growth in GDP per Capita 1940-1980									
Life expectancy	-1.316***	-1.643***	-1.589*	-1.984**	-1.241**	-0.383	-1.341***	-1.642***	-1.573*
	(0.382)	(0.521)	(0.876)	(0.911)	(0.626)	(0.383)	(0.334)	(0.520)	(0.866)
Institutions		-0.0490			-0.0291			-0.0489	
		(0.0418)			(0.0472)			(0.0417)	
Initial (1930) value of dependent variable			-0.0730			0.149			-0.0638
			(0.198)			(0.112)			(0.193)
Constant	1.336***	1.681***	1.990		1.448***	-0.127		1.680***	1.910
	(0.124)	(0.367)	(1.807)		(0.445)	(0.983)		(0.366)	(1.760)
Panel B. and variances									
				2.319***	4.527	54.28*	3.966**	9.523***	1.215
				(0.743)	(11.71)	(115.1)	(2.277)	(5.420)	(1.601)
^2_U				0.0147***	0.00171**	6.68e-07	0.00467***	0.00152***	0.0136***
				(0.0136)	(0.00486)	(0)	(0.00436)	(0.00192)	(0.0155)
^2_V				0.0150***	0.0177***	0.0139***	0.0260***	0.0179***	4.72e-05
				(0.0143)	(0.00552)	(0.00444)	(0.00516)	(0.00412)	(0.0155)
^2_R				0.0547***	0.0943**	0.120***	0.0586***	1.75e-09	0.123***
				(0.0383)	(0.0973)	(0.0263)	(0.0250)	(0)	(0.0319)
$_{ww}$				0.147***	0.143***	0.124***	0.137***	0.143***	0.125***
				(0.0261)	(0.0271)	(0.0206)	(0.0235)	(0.0270)	(0.0207)
Panel C. Dependent Variable: Growth in Life Expectancy 1940-1980									
Institutions		-0.0310***			-0.0496***			-0.0496***	
		(0.00755)			(0.00997)			(0.00996)	
Initial (1930) value of dependent variable			-0.117***			-0.184***			-0.185***
			(0.0310)			(0.0223)			(0.0212)
Constant		0.324***	1.122***		0.579***	1.757***		0.580***	1.760***
		(0.0595)	(0.267)		(0.0523)	(0.173)		(0.0522)	(0.165)
Observations		47	47	47	47	47	47	47	47
Hansen J				0.122	0	0.000493	0.850	0.000109	0.0598
p-val				0.727	1	0.982			

Notes: In all models, the endogenous regressor is the changes in log life expectancy between 1940 and 1980. 2SLS1 is the two-stage least squares regression of growth in GDP per capita on growth in life expectency, using predicted mortality as the instrument. 2SLS2 includes a measure of quality of institutions as exogenous covariate. 2SLS3 adds the initial (1930) value of log GDP per capita. GMM1-GMM3 are the same models as 2SLS1-2SLS3 estimated by GMM estimators based on over identi ed moments. GMM1-GMM3+AJ combine our over identi ed moments and the AJ moment, i.e. $E(IV'') = 0$. The last row reports the p value of the J statistics under the null hypothesis that the overidentifying restrictions are valid.

Table 2: Exactly identi ed moments: Base sample 1940 and 1980

	(1) 2SLS1	(2) 2SLS2	(3) 2SLS3	(4) GMM4	(5) GMM5	(6) GMM6	(7) GMM4+AJ	(8) GMM5+AJ	(9) GMM6+AJ
Panel A. Dependent Variable: Growth in GDP per Capita 1940-1980									
Life expectancy	-1.316***	-1.643***	-1.589*	-3.636	-1.401	-0.340	-1.344***	-1.634***	-1.599*
	(0.382)	(0.521)	(0.876)	(4.184)	(9.818)	(1.022)	(0.348)	(0.520)	(0.871)
Institutions		-0.0490			-0.0370			-0.0478	
		(0.0418)			(0.483)			(0.0416)	
Initial (1930) value of dependent variable			-0.0730			0.156			-0.0706
			(0.198)			(0.195)			(0.194)
Constant	1.336***	1.681***	1.990		1.541	-0.195		1.670***	1.969
	(0.124)	(0.367)	(1.807)		(5.663)	(1.836)		(0.365)	(1.775)
Panel B. and variances									
				3.091	13.61	1.589	2.228	0.805	1.235
				(3.588)	(2,713)	(5.015)	(1.235)	(6.184)	(1.190)
^2_U				0.0278***	0.000712	8.72e-06	0.00804***	0.0177	0.0138***
				(0.0107)	(0.128)	(0.00959)	(0.00714)	(0.136)	(0.0123)
^2_V				0.00253	0.0187	0.0139***	0.0220***	0.00170	0.000140
				(0.00960)	(0.128)	(0.0133)	(0.00718)	(0.136)	(0.0100)
^2_R				0.101***	9.05e-07	0.123***	0.0863***	0.126***	0.123***
				(0.0307)	(28.70)	(0.0306)	(0.0198)	(0.0843)	(0.0317)
Panel C. Dependent Variable: Growth in Life Expectancy 1940-1980									
Institutions		-0.0310***			-0.0496***			-0.0494***	
		(0.00755)			(0.00997)			(0.00996)	
Initial (1930) value of dependent variable			-0.117***			-0.184***			-0.184***
			(0.0310)			(0.0223)			(0.0219)
Constant		0.324***	1.122***		0.579***	1.761***		0.578***	1.758***