Journal of Econometrics 211 (2019) 308–318

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Convolution without independence

Susanne M. Schennach *

Brown University, United States

ARTICLE INFO

Article history:
Available online 12 December 2018

JEL classification:
C02
C12

Keywords:
Subindependence
Measurement error
Error-in-variables
Deconvolution
Characteristic function

ABSTRACT

Widely used convolution and deconvolution techniques traditionally rely on independence assumptions, often criticized as being strong. We observe that the convolution theorem actually holds under a weaker assumption, known as subindependence. We show that this notion is arguably as weak as a conditional mean assumption. We report various simple characterizations of subindependence and devise constructive methods to generate subindependent random variables. We extend subindependence to multivariate settings and propose the new concepts of conditional and mean subindependence, relevant to measurement error problems. We finally introduce three tests of subindependence based on characteristic functions, generalized method of moments and randomization, respectively.

1. Introduction

Convolutions and deconvolutions play a central role in the identification and estimation of measurement error models (Fan, 1991; Fan and Truong, 1993; Li, 2002; Li and Vuong, 1998; Wang and Hsiao, 2011; Taupin, 2001; Hu and Ridder, 2012, 2010; Bonhomme and Robin, 2010; Carrasco and Florens, 2011; Wilhelm, 2015; Schennach, 2004, 2007, 2008, 2013, 2016) and, more generally, in any problem involving sums of independent random variables. The use of convolution techniques in this context yields very computationally and conceptually convenient methods. However, the requirement that the variables (e.g. the true quantity of interest and its measurement error) be independent is often criticized as being too strong (Bound et al., 2001; Hu and Schennach, 2008). In this note, we observe that independence is, in fact, not necessary for the convolution theorem to hold. Instead, a much weaker notion, known as subindependence, is the appropriate necessary and sufficient condition.

Although the concept of subindependence and its relation to convolutions is known (e.g., Hamedani and Volkmer (2009), Ebrahimi et al. (2010), Hamedani (2013), and references therein), it has received surprisingly little attention. This paper contributes to this literature (i) by motivating the usefulness of this concept by showing that it is as weak as a conditional mean assumption in a well-defined sense, (ii) by providing a number of equivalent characterizations, (iii) by introducing generalizations of this concept in multivariate settings, (iv) by devising a simple and general method to generate pairs of subindependent random variables and (v) by introducing three simple tests of subindependence: One based on characteristic functions, one expressed as a generalized method of moment and one based on a permutation test. All proofs can be found in the appendix.

* Correspondence to: Brown University, Box B, Department of Economics, Providence, RI 02912, United States.
E-mail address: smschenn@alum.mit.edu.

https://doi.org/10.1016/j.jeconom.2018.12.018
0304-4076/

2. Scalar variables

Let X and Y denote two scalar real-valued random variables and let $Z = X + Y$. Let the characteristic functions (c.f.) of some random variable X be denoted by $\phi_X(\xi) \equiv E\left[e^{i\xi X}\right]$ and let the joint c.f. of two variables X and Y be denoted by $\phi_{XY}(\xi,\gamma) = E\left[e^{i\xi X}e^{i\gamma Y}\right]$. We denote the density of a random variable X (with respect to the Lebesgue measure) by f_X while its cdf is denoted by F_X, and similarly for joint densities and cdf.

The convolution theorem (Loève, 1977; Lukacs, 1970) states that, under independence of X and Y, we have the convenient factorization $\phi_Z(\xi) = \phi_{X+Y}(\xi) = E\left[e^{i\chi X+\gamma Y}\right] = E\left[e^{i\chi X}\right]E\left[e^{i\gamma Y}\right] = \phi_X(\xi)\phi_Y(\xi)$. Such a result does not actually require full independence, because full independence is equivalent to the following assumption (by Theorem 16-B in Loève (1977)):

Definition 1. Two random variables X and Y are *independent* (denoted $X \perp\!\!\!\perp Y$) iff $E\left[e^{i(\chi X+\gamma Y)}\right] = E\left[e^{i\chi X}\right]E\left[e^{i\gamma Y}\right]$ for any $\chi \in \mathbb{R}$ and $\gamma \in \mathbb{R}$.

Note that independence requires the factorization to hold for any χ and any γ when the convolution theorem only needs the factorization to hold for $\chi = \gamma$. This observation leads to the following weaker assumption:

Definition 2. Two random variables X and Y are *subindependent* (denoted $X \top\!\!\!\top Y$) iff $E\left[e^{i\xi(X+Y)}\right] = E\left[e^{i\xi X}\right]E\left[e^{i\xi Y}\right]$ for all $\xi \in \mathbb{R}$.

Note that the number of restrictions imposed by subindependence is considerably less than for independence: Only a one-dimensional subset of the domain of the joint c.f. of X and Y is constrained, instead of its whole two-dimensional domain. For comparison, this imposes as few constraints on the joint c.f. as a conditional mean assumption $E[Y \mid X = x] = 0$,[1] which can also be expressed as a constraint on the c.f. on a one-dimensional subset (see Proposition 2 in Schennach (2014)):

Definition 3. Two random variables X and Y satisfy a *conditional mean* restriction (denoted $E[Y \mid X] = 0$) iff $[\partial\phi_{XY}(\chi,\gamma)/\partial\gamma]_{\gamma=0} = E\left[Ye^{i\chi X}\right] = 0$ for all $\chi \in \mathbb{R}$.

Informally, one could interpret these observations as follows. If one were to select a generating process for X, Y at "random", the chances that it satisfies subindependence are of the same order as the chances that it satisfies a conditional mean assumption, while the chances of satisfying independence are considerably smaller. Another interpretation is that, in a case where the convolution theorem does not hold, the error made in using it anyway is related to the "distance" to the nearest generating process satisfying subindependence, which is much "closer" than the nearest model satisfying independence. A third interpretation would be that a model that assumes subindependence is as robust to deviations from this assumption as a model that relies on a conditional mean assumption, while being far more robust to deviations than a model assuming independence.

One could argue that, regardless of favorable dimensionality arguments, the notion of subindependence is less intuitive than the one of conditional mean, because the former is apparently tied to a Fourier representation. To address this concern, we now state simple ways to characterize and generate subindependent pairs of random variables that do not involve Fourier transforms. First, here is an equivalent real-space characterization of subindependence:

Lemma 1. *Two scalar real-valued random variables X and Y are subindependent iff*

$$F_{X+Y}(z) = \iint 1(x+y \le z)\, dF_X(x)\, dF_Y(y). \tag{1}$$

Note that the left-hand side of (1) is the distribution of the sum $X + Y$ (accounting for possible dependence between X and Y), while the right-hand side is the convolution of the marginal distributions of X and Y, expressed in a form that allows for general probability measures. The two expressions are obviously equal under independence, but this lemma shows that it holds under the weaker conditions of subindependence. Property (1) is given the name *summable uncorrelated marginals* by Ebrahimi et al. (2010). As the equivalence of this notion to subindependence may not be obvious to many readers, we provide a proof of this fact in the Appendix, for completeness.

For densities (instead of general probability measures), one can give a lengthier, but more transparent, characterization of subindependence:

Lemma 2 (*Adapted from Lemma 2 in Ebrahimi et al. (2010)*)**.** *Two scalar real-valued random variables X and Y with continuous joint density $f_{XY}(x,y)$ (and marginals $f_X(x)$ and $f_Y(y)$, respectively) are subindependent iff*

$$\Delta f_{XY}(x,y) \equiv f_{XY}(x,y) - f_X(x)f_Y(y) \tag{2}$$

satisfies

$$\int_{-\infty}^{\infty} \Delta f_{XY}(x,y)\, dx = 0 \text{ for any } y \in \mathbb{R} \tag{3}$$

[1] For almost every $x \in \mathbb{R}$, and assuming that $E[|Y|\,|X = x] < \infty$ and $E[|Y|] < \infty$.

$$\int_{-\infty}^{\infty} \Delta f_{XY}(x, y)\, dy = 0 \text{ for any } x \in \mathbb{R} \tag{4}$$

$$\int_{-\infty}^{\infty} \Delta f_{XY}(x, z - x)\, dx = 0 \text{ for any } z \in \mathbb{R}. \tag{5}$$

Although it is straightforward to find functions $\Delta f_{XY}(x, y)$ satisfying (3) and (4), it is more difficult to do so while at the same time satisfying (5). For this reason, we provide a simple construction to generate pairs of subindependent random variables.

Theorem 4. *Let X and Y be scalar real-valued random variables with marginal density $f_X(x)$ and $f_Y(y)$, respectively, and satisfying $E[|X|] < \infty$ and $E[|Y|] < \infty$. Any continuous joint density $f_{XY}(x, y)$ such that X and Y are subindependent can be written in the form*

$$f_{XY}(x, y) = f_X(x) f_Y(y) + \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right) p(x, y) \tag{6}$$

for some function $p : \mathbb{R}^2 \mapsto \mathbb{R}$ such that

$$\int_{-\infty}^{\infty} p(x, u)\, dx = 0 \text{ and } \int_{-\infty}^{\infty} p(u, y)\, dy = 0 \tag{7}$$

for any $u \in \mathbb{R}$ and

$$\lim_{|x| \to \infty} p(x, z - x) = 0 \tag{8}$$

for any $z \in \mathbb{R}$.

Remark 1. This theorem does not guarantee that, for any choice of $p(x, y)$, the resulting $f_{XY}(x, y)$ is a valid probability density. However, it does guarantee that if one considered every possible $p(x, y)$ satisfying the restrictions and such that (6) is a well-defined density, one would have covered all possible joint densities that satisfy subindependence. There are essentially two ways in which $f_{XY}(x, y)$ could fail to be a valid density: (i) if $p(x, y)$ is not sufficiently differentiable, which is easy to avoid and (ii) if the resulting function $f_{XY}(x, y)$ reaches negative values, in which case one can merely rescale $p(x, y)$ so that $f_{XY}(x, y) \geq 0$ everywhere.

Drawing random variables from a density of the form (6), can be straightforwardly accomplished: For instance, one can draw trial pairs X and Y of independent random variables from the densities $f_X(x)$ and $f_Y(y)$, respectively, and then accept those trial pairs with a probability $q(x, y)$, where

$$q(x, y) = \alpha \left(1 + \frac{\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right) p(x, y)}{f_X(x) f_Y(y)}\right) \tag{9}$$

and where α is a constant chosen such that $q(x, y) \leq 1$ for all $(x, y) \in \mathbb{R}^2$. This approach offers the advantage that it can be entirely expressed in terms of univariate random number generators.

We can also use Theorem 4 to construct simple examples that provide graphical intuition into the concept of subindependence.

Example 1. Taking $p(x, y) = xye^{-\left(x^2+y^2\right)/2}$ yields $f_{XY}(x, y) = f_X(x) f_Y(y) + \Delta f_{XY}(x, y)$ with $\Delta f_{XY}(x, y) = \left(y - x + xy^2 - yx^2\right) e^{-\frac{1}{2}x^2 - \frac{1}{2}y^2}$.

The deviation $\Delta f_{XY}(x, y)$ is shown in Fig. 1 and illustrates perhaps the simplest general shape of a deviation from independence that will preserve subindependence. One can also easily construct an example (illustrated in Fig. 2) where independence is violated but subindependence and conditional mean $E[Y \mid X] = 0$ hold. This is useful to see that subindependence is not incompatible with the natural conditional mean assumption.

Example 2. Taking $p(x, y) = \left(y^2 - 1\right) xe^{-\left(x^2+y^2\right)/2}$ yields $f_{XY}(x, y) = f_X(x) f_Y(y) + \Delta f_{XY}(x, y)$ with $\Delta f_{XY}(x, y) = \left(-1 + x^2 + y^2 - 3xy + xy^3 - x^2y^2\right) e^{-\frac{1}{2}x^2 - \frac{1}{2}y^2}$. Note that $\int y f_{XY}(x, y)\, dy = 0$ in this case (if $\int y f_Y(y)\, dy = 0$).

We conclude this section by providing a few convenient results for transformed variables. It is well-known that nonlinear transformations preserve independence, i.e. $X \perp\!\!\!\perp Y \implies f(X) \perp\!\!\!\perp g(Y)$ for two measurable functions f and g. For subindependence, this only holds for specific linear transformations. There is also some interesting relation between independence and subindependence: a family of subindependence restrictions implies independence. These results are summarized below.

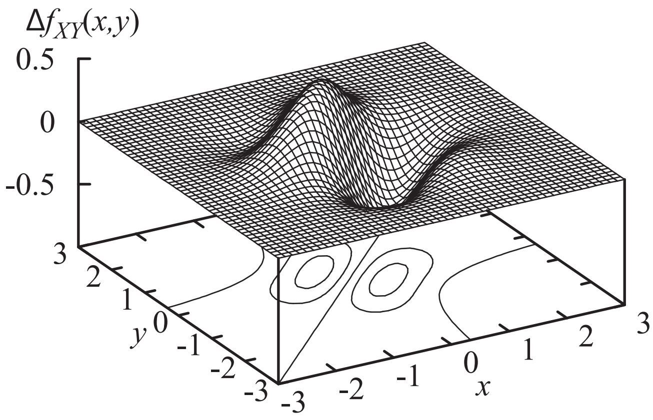

Fig. 1. Function $\Delta f_{XY}(x, y)$ from Example 1.

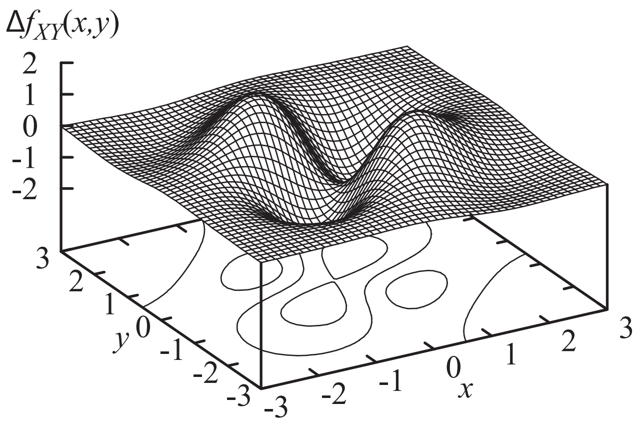

Fig. 2. Function $\Delta f_{XY}(x, y)$ from Example 2.

Theorem 5. *Let X and Y be scalar real-valued random variables. Then,*

1. $X \top\!\!\!\top Y \Rightarrow (a + cX) \top\!\!\!\top (b + cY)$ *for any* $a, b, c \in \mathbb{R}$. *(However,* $X \top\!\!\!\top Y \nRightarrow (a + cX) \top\!\!\!\top (b + dY)$ *for any* $a, b, c, d \in \mathbb{R}$ *if* $c \neq d$*).*
2. $X \top\!\!\!\top aY$ *for any* $a \in \mathbb{R} \Rightarrow X \perp\!\!\!\perp Y$.

3. Multivariate extensions

The concept of subindependence in multivariate settings leads to a broader array of related concepts, which we now discuss.

Definition 6. Conformable random vectors $X_1, \ldots, X_n$ are said to *mutually subindependent* (denoted $\top\!\!\!\top (X_1, \ldots, X_n)$) if either (i) $E\left[\exp\left(i\xi' \sum_{k=1}^{n} X_k\right)\right] = \prod_{k=1}^{n} E\left[e^{i\xi' X_k}\right]$ or (ii) $F_{X_1+\cdots+X_n}(s) = \int \cdots \int 1\left(\sum_{k=1}^{n} x_k \leq s\right) dF_{X_1}(x_1) \cdots dF_{X_n}(x_n)$ (with $\leq$ holding for all elements).

Theorem 7. *Let* X_1, X_2, X_3 *be random vectors of the same dimension. The following hold:*

1. $X_1 \top\!\!\!\top X_2$ *and* $(X_1 + X_2) \top\!\!\!\top X_3 \Rightarrow \top\!\!\!\top (X_1, X_2, X_3)$.
2. $X_1 \top\!\!\!\top X_2$ *and* $X_2 \top\!\!\!\top X_3$ *and* $X_3 \top\!\!\!\top X_1 \nRightarrow \top\!\!\!\top (X_1, X_2, X_3)$.

These results parallel analogue implications for independence. The second result is interesting in the sense that one would have expected that it should hold, as subindependence looks like a fundamentally pairwise property. The concept of conditional independence also admits a subindependence analogue that is useful in measurement error problems.

Definition 8. Let X, Y and Z be random vectors (where X and Y have the same dimension). Then, X and Y are said to be *subindependent conditional* on Z (denoted $X \top\!\!\!\top Y \mid Z$) iff either (i) $E\left[e^{i\xi'(X+Y)} \mid Z\right] = E\left[e^{i\xi'X} \mid Z\right] E\left[e^{i\xi'Y} \mid Z\right]$ for any $\xi \in \mathbb{R}$ or, equivalently, (ii) $F_{X+Y|Z}(s \mid z) = \int\int 1(x+y \le s)\, dF_{X|Z}(x \mid z)\, dF_{Y|Z}(y \mid z)$.

The following result is useful to deconvolve a joint distribution involving other variables not measured with error.

Theorem 9. *If $X \top\!\!\!\top Y \mid Z$ and $Y \perp\!\!\!\perp Z$, then*

$$E\left[e^{i\zeta Z} e^{i\xi(X+Y)}\right] = E\left[e^{i\zeta Z} e^{i\xi X}\right] E\left[e^{i\xi Y}\right] \tag{10}$$

for any $\xi \in \mathbb{R}$ and

$$F_{X+Y|Z}(w|z) = \iint 1(x+y \le w)\, dF_{X|Z}(x|z)\, dF_Y(y). \tag{11}$$

If densities exist, then

$$f_{X+Y,Z}(w,z) = \int f_{XZ}(w-y,z) f_Y(y)\, dy. \tag{12}$$

Corollary 10. *If $X \top\!\!\!\top Y \mid Z$ and $Y \perp\!\!\!\perp Z \implies E\left[g(Z)\, e^{i\xi(X+Y)}\right] = E\left[g(Z)\, e^{i\xi X}\right] E\left[e^{i\xi Y}\right]$ for any $\xi \in \mathbb{R}$ and for any measurable function g.*

We can also introduce an analogue of the concept of mean independence $E[Y|X] = E[Y]$ in a subindependence context. Observe that mean independence of $g(Y)$ from X implies that $E\left[g(Y)\, e^{i\xi X}\right] = E[g(Y)]\, E\left[e^{i\xi X}\right]$ for any given measurable function g. This suggests the following definition:

Definition 11. Random variable Y is said to be *mean subindependent* from X iff $E\left[Ye^{i\xi(Y+X)}\right] = E\left[Ye^{i\xi Y}\right] E\left[e^{i\xi X}\right]$ for any $\xi \in \mathbb{R}$.

Note that had we defined Y *mean subindependent* from X as $E\left[Ye^{i\xi X}\right] = E[Y]\, E\left[e^{i\xi X}\right]$, it would actually have coincided with the usual concept of mean independence. Also, if we had defined it as $E\left[Ye^{i\xi Y}|X\right] = E\left[Ye^{i\xi Y}\right]$, then it would have reduced to the usual notion of independence. The need for factorizations of the form $E\left[Ye^{i\xi(Y+X)}\right] = E\left[Ye^{i\xi Y}\right] E\left[e^{i\xi X}\right]$ often arises in the derivation of results similar to Kotlarski's identity (Kotlarski, 1967).

We conclude this section with a simple sufficient condition for subindependence of sums.

Theorem 12. *If $(X_1, Y_1) \perp\!\!\!\perp (X_2, Y_2)$ and $X_1 \top\!\!\!\top Y_1$ and $X_2 \top\!\!\!\top Y_2 \implies (X_1+X_2) \top\!\!\!\top (Y_1+Y_2)$.*

4. Testing

It is useful to develop empirical tests of the assumption of subindependence. Of course, as would be the case for testing independence, the methods proposed here assume that all variables are observed. In measurement error contexts, this would typically be verified with validation data.

The most obvious approach, conceptually, is to directly test Definition 2 in terms of characteristic functions. One can proceed in analogy with characteristic function based tests of independence and consider test statistics of the form

$$\hat{T} = \int \left|\hat{\phi}_X(\xi)\, \hat{\phi}_Y(\xi) - \hat{\phi}_{X+Y}(\xi)\right|^2 \omega(\xi)\, d\xi \tag{13}$$

where $\hat{\phi}_X(\xi) \equiv \frac{1}{n}\sum_{i=1}^{n} e^{i\xi X_i}$ for any random variable X (and an associated iid sample $(X_1, \ldots, X_n)$ drawn from the same marginal distribution) and where $\omega(\xi)$ is a user-specified weighting function satisfying:

Assumption 1. $\omega(\xi)$ is strictly positive for all $\xi \in \mathbb{R}$ and $\omega(\xi) \le C(1+|\xi|)^{-r}$ for some $r > 2$ and some $C > 0$.

We also need a few basic assumptions.

Assumption 2. $E[|X_i|] < \infty$, $E[|Y_i|] < \infty$.

Assumption 3. X_i and Y_i take values in $\mathbb{R}$ and (X_i, Y_i) forms an iid sequence.

Theorem 13. *Under Assumptions 1–3, if $X \;\top\!\!\!\top\; Y$, we have $\sqrt{n}\hat{T} \xrightarrow{d} N\left(0, E\left[(\psi(X,Y))^2\right]\right)$ where*

$$\begin{aligned}\psi(x,y) = {} & 2\int \left(\phi_X(-\xi)\,\phi_Y(-\xi) - \phi_{X+Y}(-\xi)\right)\phi_Y(\xi)\,\omega(\xi)\,e^{i\xi x}d\xi + \\ & 2\int \left(\phi_X(-\xi)\,\phi_Y(-\xi) - \phi_{X+Y}(-\xi)\right)\phi_X(\xi)\,\omega(\xi)\,e^{i\xi y}d\xi + \\ & 2\int \left(\phi_X(-\xi)\,\phi_Y(-\xi) - \phi_{X+Y}(-\xi)\right)\omega(\xi)\,e^{i\xi(x+y)}d\xi.\end{aligned} \tag{14}$$

A consistent asymptotic variance estimator can be obtained by replacing, in (14), the characteristic function $\phi_X(\xi)$, etc., by their empirical counterparts $\hat{\phi}_X(\xi) = \frac{1}{n}\sum_{i=1}^{n} e^{i\xi X_i}$, etc.

This test offers the advantage that it has power against any deviations from subindependence. The user-specified weighting function $\omega(\xi)$ is necessary to ensure that the test statistic has a finite asymptotic variance.

If one wishes to avoid using Fourier transforms, a continuum of restrictions or complex numbers, one can formulate a polynomial test of subindependence, based on the following result.

Theorem 14. *If $E\left[|X|^l\right] < \infty$ and $E\left[|Y|^l\right] < \infty$ for $l = 1, \ldots, k$ and $X \;\top\!\!\!\top\; Y$, then*

$$E\left[(X+Y)^k\right] = \sum_{l=0}^{k}\binom{k}{l} E\left[X^l\right] E\left[Y^{k-l}\right]. \tag{15}$$

Conversely, if the moment generating function of (X, Y) exists in an open neighborhood of the origin and if (15) holds for any $k \in \mathbb{N}$, then $X \;\top\!\!\!\top\; Y$.

Drawbacks of this formulation are (i) the strong moment existence requirements and (ii) the difficulty in practically implementing a test that has power against all deviations from subindependence, since this would require considering all $k \in \mathbb{N}$. On the other hand, a finite dimensional version of this test (with moments up to K) is very simple to implement via the Generalized Method of Moment (GMM) (Hansen, 1982) with the moment vector:

$$E\left[(X+Y)^k\right] - \sum_{l=0}^{k}\binom{k}{l} m_l E\left[Y^{k-l}\right] = 0 \text{ for } k = 1, \ldots, K \tag{16}$$

$$E\left[X^l\right] - m_l = 0 \text{ for } l = 1, \ldots, K, \tag{17}$$

where we rephrased the moment conditions in an equivalent form by introducing the parameter vector m_l to obtain moment conditions that are linear in the generating process, as required by standard GMM.

One can also exploit the equivalent definition provided by Lemma 1 to devise a resampling/randomization test of subindependence. Given an iid sample of two variables (X_i, Y_i) with $i = 1, \ldots, n$ one can proceed as follows:

Theorem 15. *Under Assumption 3, a valid statistical test of subindependence of X and Y can be obtained as follows from a sample $(X_i, Y_i)_{i=1,\ldots,n}$:*

1. *For each $i = 1, \ldots, n$, draw the r_i at random uniformly from $\{1, \ldots, n\}$ (with replacement), independently from all X_i and Y_i.*
2. *Let $Z_i = X_i + Y_i$ and $W_i = X_i + Y_{r_i}$.*
3. *Test equality of the distributions of Z_i and W_i using a two-sample Kolmogorov–Smirnov test (Conover, 1971).*

The principle underlying this test is that the distribution of Z_i is the true distribution of the sum of X_i and Y_i (accounting for possible dependence between them) while the distribution of W_i is the convolution of the ones of X_i and Y_i, since any dependence has been essentially removed through the randomization of Y_{r_i} via r_i and by relying on the iid assumption on Y_i.

Which test is preferable depends on the situation. The test based on Theorem 14 (for finite k) is perhaps the simplest to implement but it is blind to some deviations from subindependence and thus tends to be dominated by the other two tests in terms of power. The tests based on Theorems 13 and 15 are both able, asymptotically, to detect any types of deviations from the subindependence but one test does not uniformly dominate the other. The test based on Theorem 13 will tend to perform better (in terms of power) for smooth distributions (since the weighting function $\omega(\xi)$ effectively performs smoothing) while the test based on Theorem 15 will tend to perform better for nonsmooth ones. As is often the case in nonparametric tests, it is difficult to establish some form of power optimality property for any of the proposed tests.

Of course, in practice, one often does not have access to the true, correctly measured, variables that would be necessary to perform the above direct tests in measurement error models. In such cases, indirect tests could be devised by exploiting, for instance, the fact that two different choices of valid instruments or of repeated measurements should yield estimates that are not statistically significantly different. Such test, however, jointly tests all the assumptions of the estimation method and not just subindependence.

5. Conclusion

This paper's aim is not to try to argue that economic models should be stated in terms of subindependence, which would admittedly be an unnatural assumption. Rather, we are arguing that inferences made under the assumption of independence are robust to large deviations from independence that maintain subindependence. This considerably expands the scope of validity of the wide range of methods developed under independence, because subindependence is arguably just as weak an assumption as conditional mean. Indeed, both conditions, when phrased in terms of c.f., impose constraints on a subset of its domain that is of the same dimension. We provide explicit examples that illustrate that deviations from independence that maintain subindependence have quite simple and plausible shapes. We also propose extensions of the subindependence concept to multivariate settings and introduce simple tests of subindependence.

Acknowledgments

The author acknowledges support from United States National Science Foundation (NSF) grants SES-0752699, SES-1156347, SES-1357401 and SES-1659334.

Appendix. Proofs

Proof of Lemma 1. To show the equivalence of Eq. (1) and Definition 2, we observe that the equality of the cdf $F_{X+Y}(z)$ and $\tilde{F}_Z(z) \equiv \iint 1(x+y \leq z)\, dF_X(x)\, dF_Y(y)$ is equivalent to the equality of the corresponding probability measures $dF_{X+Y}(z)$ and $d\tilde{F}_Z(z)$. By the well-known uniqueness of c.f. (Loève (1977), Lukacs (1970)), this is equivalent to the equality between the corresponding Fourier transforms for all real ξ:

$$\int e^{i\xi z} dF_{X+Y}(z) = \int e^{i\xi z} d\tilde{F}_Z(z)\,. \tag{18}$$

The left-hand side of (18) is obviously

$$\int e^{i\xi z} dF_{X+Y}(z) = E\left[e^{i\xi(X+Y)}\right] \tag{19}$$

by construction. Evaluating the right-hand-side yields:

$$\begin{aligned}\int e^{i\xi z} d\tilde{F}_Z(z) &= \int e^{i\xi z} d\left(\int\int \mu_{x+x}(z)\, dF_X(x)\, dF_Y(y)\right)\\ &= \iiint e^{i\xi z} d\mu_{x+y}(z)\, dF_X(x)\, dF_Y(y)\end{aligned} \tag{20}$$

where $\mu_{z_0}(z) \equiv 1(z_0 \leq z)$ and where the second equality follows from Fubini's theorem for finite measures (see Chapter 5 in Bhattacharya and Rao (2010)). Since $d\mu_{x+y}(z)$ represents a unit point mass at $z = x + y$, we have

$$\begin{aligned}\int e^{i\xi z} d\tilde{F}_Z(z) &= \iint e^{i\xi(x+y)} dF_X(x)\, dF_Y(y)\\ &= \iint e^{i\xi x} e^{i\xi y} dF_X(x)\, dF_Y(y)\\ &= \int e^{i\xi x} dF_X(x) \int e^{i\xi y} dF_Y(y)\\ &= E\left[e^{i\xi x}\right] E\left[e^{i\xi z}\right]\end{aligned} \tag{21}$$

where we have again used Fubini's theorem for finite (complex) measures. Equating (19) and (21) for any $\xi \in \mathbb{R}$ yields Definition 2. ■

Proof of Theorem 4. Subindependence of X and Y requires that $\phi_{XY}(\chi, \chi) = \phi_X(\chi)\phi_Y(\chi) = \phi_{XY}(\chi, 0)\phi_{XY}(0, \chi)$. Therefore, $\phi_{XY}(\chi, \gamma)$ can be written as:

$$\phi_{XY}(\chi, \gamma) = \phi_X(\chi)\phi_Y(\gamma) + \Delta\phi_{XY}(\chi, \gamma) \tag{22}$$

where $\Delta\phi_{XY}(\chi, \gamma) = 0$ if either $\chi = 0, \gamma = 0$ or $\chi = \gamma$. Since $\Delta\phi_{XY}(\chi, \gamma)$ is a difference of c.f., which are always bounded, it is also bounded. Since $E[|X|]$ and $E[|Y|]$ are finite, $\phi_X(\chi)$, $\phi_Y(\gamma)$ and $\phi_{XY}(\chi, \gamma)$ are everywhere continuously differentiable and, therefore, so is $\Delta\phi_{XY}(\chi, \gamma)$. In particular, this implies that near the line $\chi = \gamma$ (where $\Delta\phi_{XY}(\chi, \gamma)$ vanishes), $\Delta\phi_{XY}(\chi, \gamma)$ behaves linearly and the ratio $\frac{\Delta\phi_{XY}(\chi,\gamma)}{i(\chi-\gamma)}$ does not diverge ($i = \sqrt{-1}$ is a constant introduced for convenience). Away from this line, $(\chi - \gamma)$ is nonzero, so the ratio does not diverge either and $\Delta\phi_{XY}(\chi, \gamma)$ can be written in the form:

$$\Delta\phi_{XY}(\chi, \gamma) = i(\chi - \gamma)\psi(\chi, \gamma) \tag{23}$$

where $\psi(\chi, \gamma)$ is finite at each $(\chi, \gamma) \in \mathbb{R}^2$ and such that $\psi(\chi, 0) = 0$ and $\psi(0, \gamma) = 0$. Since $(\chi - \gamma)$ is nonzero along the lines $\chi = 0$ and $\gamma = 0$ (except at $\chi = \gamma = 0$) the constraints $\Delta\phi_{XY}(\chi, 0) = 0$ and $\Delta\phi_{XY}(0, \gamma) = 0$ translate directly into the constraint that $\psi(\chi, 0) = 0$ and $\psi(0, \gamma) = 0$. The inverse Fourier transform of these constraints yields (7). (The value at $\psi(0, 0)$ is irrelevant since it is finite and multiplied by $(\chi - \gamma) = 0$.) Since $\Delta\phi_{XY}(\chi, \gamma)$ and $\psi(\xi, \gamma)$ are bounded, they are a special case of tempered distributions and admit an inverse Fourier transform, given by Lighthill (1962):

$$\Delta f_{XY}(x, y) = \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right) p(x, y) \tag{24}$$

where $\Delta f_{XY}(x, y)$ and $p(x, y)$ denote the inverse Fourier transforms of $\Delta\phi_{XY}(\chi, \gamma)$ and $\psi(x, \gamma)$, respectively. Combining (24) with the inverse Fourier transform of (22) yields (6). We can further restrict the behavior of $p(x, y)$ by invoking Eq. (5) from Lemma 2 with $\Delta f_{XY}(x, y)$ given by (24):

$$\int_{-\infty}^{\infty} \left[\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right) p(x, y)\right]_{y=z-x} dx = 0 \tag{25}$$

for any $z \in \mathbb{R}$. Letting superscripts denote orders of derivatives with respect to each argument, we can write this expression in terms of a total differential and integrate it:

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} \left(p^{(10)}(x, z - x) - p^{(01)}(x, z - x)\right) dx = \int_{-\infty}^{\infty} \frac{d}{dx} p(x, z - x)\, dx \\ &= \lim_{x\to\infty} p(x, z - x) - \lim_{x\to-\infty} p(x, z - x) \end{aligned} \tag{26}$$

Hence $\lim_{x\to\infty} p(x, z - x) = \lim_{x\to-\infty} p(x, z - x)$. These limits must also be zero by the continuity of $\psi(\chi, \gamma)$, thus implying (8). ∎

Proof of Theorem 5. To show the first implication we observe that, for any constant $a, b, c \in \mathbb{R}$ and letting $\tilde{\xi} \equiv c\xi$, we have

$$\begin{aligned} E\left[e^{i\xi(a+cX+b+cY)}\right] &= e^{i\xi a} e^{i\xi b} E\left[e^{i\xi c(X+Y)}\right] \\ &= e^{i\xi a} e^{i\xi b} E\left[e^{i\tilde{\xi}(X+Y)}\right] = e^{i\xi a} e^{i\xi b} E\left[e^{i\tilde{\xi} X}\right] E\left[e^{i\tilde{\xi} Y}\right] \\ &= e^{i\xi a} e^{i\xi b} E\left[e^{i\xi cX}\right] E\left[e^{i\xi cY}\right] = E\left[e^{i\xi(a+cX)}\right] E\left[e^{i\xi(b+cY)}\right] \end{aligned} \tag{27}$$

where we have used the fact that $X \top\!\!\top Y$ to write $E\left[e^{i\tilde{\xi}(X+Y)}\right] = E\left[e^{i\tilde{\xi}X}\right] E\left[e^{i\tilde{\xi}Y}\right]$. Note that this result relies crucially on X and Y having the same prefactor c, hence the implication fails to hold if the prefactors are different, i.e., one cannot generally conclude that $(a + cX) \top\!\!\top (b + dY)$ if $c \neq d$.

To show the second implication, we note that $E\left[e^{i\xi X + i\gamma Y}\right] = E\left[e^{i\xi(X+aY)}\right]$ for $a = \gamma/\xi$, next, by $X \top\!\!\top aY$ for any $a \in \mathbb{R}$, we have

$$E\left[e^{i\xi(X+aY)}\right] = E\left[e^{i\xi X}\right] E\left[e^{i\xi aY}\right] = E\left[e^{i\xi X}\right] E\left[e^{i\gamma Y}\right]. \quad ∎ \tag{28}$$

Proof of Theorem 7. To show the first implication we first write $E\left[e^{i\xi X_1} e^{i\xi X_2} e^{i\xi X_3}\right] = E\left[e^{i\xi(X_1+X_2)} e^{i\xi X_3}\right] = E\left[e^{i\xi(X_1+X_2)}\right] E\left[e^{i\xi X_3}\right]$ since $(X_1 + X_2) \top\!\!\top X_3$. Then $E\left[e^{i\xi(X_1+X_2)}\right] = E\left[e^{i\xi X_1}\right] E\left[e^{i\xi X_2}\right]$ by $X_1 \top\!\!\top X_2$ and the result follows.

To shows the failure of the second implication, consider the characteristic function $\phi_{X_1X_2X_3}(\xi_1, \xi_2, \xi_3) = E\left[e^{i\xi_1 X_1 + i\xi_2 X_2 + i\xi_3 X_3}\right]$ and note that $X_1 \top\!\!\top X_2$ and $X_2 \top\!\!\top X_3$ imply constraints on $\phi_{X_1X_2X_3}(\xi_1, \xi_2, \xi_3)$ over the set $\{(\xi, \xi, 0) : \xi \in \mathbb{R}\} \cup \{(0, \xi, \xi) : \xi \in \mathbb{R}\}$ while $\top\!\!\top (X_1, X_2, X_3)$ implies constraints on $\phi_{X_1X_2X_3}(\xi_1, \xi_2, \xi_3)$ over the set $\{(\xi, \xi, \xi) : \xi \in \mathbb{R}\}$. Since these set are always different, the implication cannot hold. ∎

Proof of Theorem 9. To show the first result, we use, in turn, (i) iterated expectations, (ii) $X \top\!\!\top Y|Z$, (iii) $Y \perp\!\!\!\perp Z$ (iv) iterated expectations again:

$$\begin{aligned} E\left[e^{i\zeta Z} e^{i\xi(X+Y)}\right] &= E\left[e^{i\zeta Z} E\left[e^{i\xi(X+Y)}|Z\right]\right] = E\left[e^{i\zeta Z} E\left[e^{i\xi X}|Z\right] E\left[e^{i\xi Y}|Z\right]\right] \\ &= E\left[e^{i\zeta Z} E\left[e^{i\xi X}|Z\right] E\left[e^{i\xi Y}\right]\right] = E\left[e^{i\xi Y}\right] E\left[e^{i\zeta Z} E\left[e^{i\xi X}|Z\right]\right] \\ &= E\left[e^{i\xi Y}\right] E\left[e^{i\zeta Z} e^{i\xi X}\right] \end{aligned} \tag{29}$$

Similarly, in terms of cdf, using in turn (i) $X \top\!\!\top Y|Z$ and (ii) $Y \perp\!\!\!\perp Z$, we have:

$$\begin{aligned} F_{X+Y|Z}(w|z) &= \iint 1(x + y \leq w)\, dF_{X,Y|Z}(x, y|z) \\ &= \iint 1(x + y \leq w)\, dF_{X|Z}(x|z)\, dF_{Y|Z}(y|z) \\ &= \iint 1(x + y \leq w)\, dF_{X|Z}(x|z)\, dF_{Y}(y). \end{aligned} \tag{30}$$

If densities exist, the corresponding steps become:

$$\begin{aligned} f_{X+Y,Z}(w,z) &= f_{X+Y,Z}(w|z) f_Z(z) = \int f_{X,Y|Z}(w-y,y|z)\,dy\, f_Z(z) \\ &= \int f_{X|Z}(w-y|z) f_{Y|Z}(y|z)\,dy\, f_Z(z) = \int f_{X|Z}(w-y|z) f_Y(y)\,dy\, f_Z(z) \\ &= \int f_{X,Z}(w-y,z) f_Y(y)\,dy \quad \blacksquare \end{aligned} \tag{31}$$

Proof of Corollary 10. As in the proof of Theorem 9, we have:

$$\begin{aligned} E\left[g(Z)e^{i\xi(X+Y)}\right] &= E\left[g(Z)E\left[e^{i\xi(X+Y)}|Z\right]\right] = E\left[g(Z)E\left[e^{i\xi X}|Z\right]E\left[e^{i\xi Y}|Z\right]\right] \\ &= E\left[g(Z)E\left[e^{i\xi X}|Z\right]E\left[e^{i\xi Y}\right]\right] = E\left[g(Z)E\left[e^{i\xi X}|Z\right]\right]E\left[e^{i\xi Y}\right] \\ &= E\left[g(Z)e^{i\xi X}\right]E\left[e^{i\xi Y}\right]. \quad \blacksquare \end{aligned} \tag{32}$$

Proof of Theorem 12. We can write:

$$\begin{aligned} E\left[e^{i\xi(X_1+X_2+Y_1+Y_2)}\right] &= E\left[e^{i\xi(X_1+Y_1)}\right]E\left[e^{i\xi(X_2+Y_2)}\right] \\ &= E\left[e^{i\xi X_1}\right]E\left[e^{i\xi Y_1}\right]E\left[e^{i\xi X_2}\right]E\left[e^{i\xi Y_2}\right] \\ &= E\left[e^{i\xi X_1}\right]E\left[e^{i\xi X_2}\right]E\left[e^{i\xi Y_1}\right]E\left[e^{i\xi Y_2}\right] \\ &= E\left[e^{i\xi(X_1+X_2)}\right]E\left[e^{i\xi(Y_1+Y_2)}\right], \end{aligned} \tag{33}$$

which implies that $(X_1, Y_1) \perp\!\!\!\perp (X_2, Y_2)$. $\blacksquare$

Proof of Theorem 13. The calculation of the limiting distribution can be performed by calculating the influence function and showing negligibility of suitable remainder terms. Let us define $\delta\phi_X(\xi) \equiv \hat{\phi}_X(\xi) - \phi_X(\xi) = \frac{1}{n}\sum_{j=1}^n e^{i\xi X_j} - E\left[e^{i\xi X_j}\right]$ and thus write $\hat{\phi}_X(\xi) = \phi_X(\xi) + \delta\phi_X(\xi)$, and similarly for all other quantities. An expansion of[2]

$$\hat{T} = \int \left(\hat{\phi}_X(-\xi)\hat{\phi}_Y(-\xi) - \hat{\phi}_{X+Y}(-\xi)\right)\left(\hat{\phi}_X(\xi)\hat{\phi}_Y(\xi) - \hat{\phi}_{X+Y}(\xi)\right)\omega(\xi)\,d\xi \tag{34}$$

in $\delta\phi_X(\xi)$, $\delta\phi_Y(\xi)$ and $\delta\phi_{X+Y}(\xi)$ to linear order yields:

$$\begin{aligned} \hat{T} &= 2\mathrm{Re}\int (\phi_X(-\xi)\phi_Y(-\xi) - \phi_{X+Y}(-\xi))\,\phi_Y(\xi)\,\omega(\xi)\,\delta\phi_X(\xi)\,d\xi \\ &\quad + 2\mathrm{Re}\int (\phi_X(-\xi)\phi_Y(-\xi) - \phi_{X+Y}(-\xi))\,\phi_X(\xi)\,\omega(\xi)\,\delta\phi_Y(\xi)\,d\xi + \\ &\quad 2\mathrm{Re}\int (\phi_X(-\xi)\phi_Y(-\xi) - \phi_{X+Y}(-\xi))\,\omega(\xi)\,\delta\phi_{X+Y}(\xi)\,d\xi + R \end{aligned} \tag{35}$$

where R includes all higher order terms (and will be bounded below). After substituting in the fact that $\delta\phi_X(\xi) = \frac{1}{n}\sum_{j=1}^n e^{i\xi X_j} - E\left[e^{i\xi X_j}\right]$ (and similarly for $\delta\phi_Y(\xi)$ and $\delta\phi_{X+Y}(\xi)$) and evaluating the integral over ξ we obtain:

$$\hat{T} = \frac{1}{n}\sum_{j=1}^{n} \psi\left(X_j, Y_j\right) - E\left[\psi\left(X_j, Y_j\right)\right] + R. \tag{36}$$

for $\psi(x, y)$ given in the statement of the theorem. Provided R is negligible, asymptotic normality and the form of the asymptotic variance then follow from Assumption 3 and the Lindeberg–Lévy central limit theorem.

To bound R we observe that it consists of a finite linear combination of terms of the form

$$R_m = \int (\phi_*(\xi))^k (\delta\phi_*(\xi))^l \omega(\xi)\, e^{i\xi *} d\xi \tag{37}$$

where $(\phi_*)^k$ denotes a product of k (potentially different) terms $\phi_*(\xi)$ each taken from $\{\phi_X(\xi), \phi_Y(\xi), \phi_{X+Y}(\xi)\}$ and $\delta\phi_*(\xi)$ denotes a product of l (potentially different) terms $\delta\phi_*(\xi)$ each taken from $\{\delta\phi_X(\xi), \delta\phi_Y(\xi), \delta\phi_{X+Y}(\xi)\}$ and $e^{i\xi *}$ denotes one of $\left\{e^{i\xi x}, e^{i\xi y}, e^{i\xi(x+y)}\right\}$. Note that $2 \le l \le 4$ and $0 \le k \le 3$.

Since $\phi_*(\xi)$ is a characteristic function, it satisfies $|\phi_*(\xi)| \le 1$. We also have that $|\delta\phi_*(\xi)| \le 2$ since it is a difference of two characteristic functions (even empirical characteristic function are bounded by 1). We use Lemma 6 in Schennach (2004) (with $a_j = 1$ and $z_j = X_j$) to establish that, for any $A, \kappa > 0$, $\sup_{|\xi| \le An^{\kappa}} |\delta\phi_X(\xi)| = O_p\left(n^{-1/2+\varepsilon}\right)$ for arbitrarily small

2 Note that the complex conjugate of $\hat{\phi}_X(\xi)$ is equal to $\hat{\phi}_X(-\xi)$.

$\varepsilon > 0$ (and similarly for $\delta\phi_Y(\xi)$, $\delta\phi_{X+Y}(\xi)$). By Assumption 1, $\omega(\xi) \leq C(1+|\xi|)^{-r} \leq C$ for $r = 2+\delta$ with $\delta > 0$. Breaking up the integral in (37) into two domains, and exploiting the above bounds, we can write:

$$
\begin{aligned}
|R_m| &\leq \int_{|\xi|\leq An^{\kappa}} |\phi_*(\xi)|^k |\delta\phi_*(\xi)|^l \omega(\xi) e^{i\xi*} d\xi + \int_{|\xi|\geq An^{\kappa}} |\phi_*(\xi)|^k |\delta\phi_*(\xi)|^l \omega(\xi) e^{i\xi*} d\xi \\
&\leq \int_{|\xi|\leq An^{\kappa}} 1^k O_p\left(n^{-1/2+\varepsilon}\right)^l C d\xi + \int_{|\xi|\geq An^{\kappa}} 1^k 2^l C(1+|\xi|)^{-r} d\xi \\
&= O_p(n^{(-1/2+\varepsilon)l}) C 2An^{\kappa} + C' 2^l 2(1+An^{\kappa})^{-r+1} \\
&\leq O_p\left(n^{(-1/2+\varepsilon)l+\kappa}\right) + O_p\left(n^{\kappa(-r+1)}\right) \\
&\leq O_p\left(n^{(-1/2+\varepsilon)2+\kappa}\right) + O_p\left(n^{\kappa(-2-\delta+1)}\right) \\
&= O_p\left(n^{(-1/2+\varepsilon)2+\kappa}\right) + O_p\left(n^{\kappa(-1-\delta)}\right).
\end{aligned} \tag{38}
$$

Both terms can be made $o\left(n^{-1/2}\right)$ by picking κ such that $(1+\delta)^{-1}/2 < \kappa < 1/2$ and a positive $\varepsilon < 1/4 - \kappa/2$ (note that a $\varepsilon > 0$ is possible since $\kappa < 1/2$). Hence $R = o_p\left(n^{-1/2}\right)$ and the linear terms (that are asymptotically normal) dominate.

To show consistency of the asymptotic variance estimate obtained by replacing characteristic functions by their empirical counterparts, we observe that the difference, arising from estimating the first term in Equation (14),

$$
\begin{aligned}
\tilde{R}(x,y) &= \int \left(\hat{\phi}_X(-\xi)\hat{\phi}_Y(-\xi) - \hat{\phi}_{X+Y}(-\xi)\right)\hat{\phi}_Y(\xi)\,\omega(\xi)\, e^{i\xi x} d\xi \\
&\quad - \int \left(\phi_X(-\xi)\phi_Y(-\xi) - \phi_{X+Y}(-\xi)\right)\phi_Y(\xi)\,\omega(\xi)\, e^{i\xi x} d\xi
\end{aligned} \tag{39}
$$

can also be written as a linear combination of terms of the form of R_m in Eq. (37) with $1 \leq l \leq 4$ and $0 \leq k \leq 3$ and we can show through a similar reasoning that $\sup_{x,y\in\mathbb{R}} \left|\tilde{R}(x,y)\right| = o_p(1)$. (Note that the order is reduced from $o_p\left(n^{-1/2}\right)$ to $o_p(1)$, relative to the earlier derivation, because now l can take the value 1. Also note that the bound is uniform in x and y because all terms involving x or y are complex exponential with unit modulus.) Similar arguments hold for the two other terms of Equation (14). ■

Proof of Theorem 14. To show that subindependence implies Eq. (15), we write:

$$
\begin{aligned}
&\frac{d^k}{d\xi^k}\left(E\left[e^{i\xi X}\right]E\left[e^{i\xi Y}\right] - E\left[e^{i\xi(X+Y)}\right]\right) \\
&= \sum_{l=0}^{k}\binom{k}{l} E\left[(iX)^l e^{i\xi X}\right]E\left[(iY)^{k-l}e^{i\xi Y}\right] - E\left[(i(X+Y))^k e^{i\xi(X+Y)}\right].
\end{aligned} \tag{40}
$$

Evaluating this at $\xi = 0$ yields:

$$
\begin{aligned}
&\sum_{l=0}^{k}\binom{k}{l} E\left[(iX)^l\right]E\left[(iY)^{k-l}\right] - E\left[(i(X+Y))^k\right] \\
&= i^k\sum_{l=0}^{k}\binom{k}{l} E\left[X^l\right]E\left[Y^{k-l}\right] - i^k E\left[(X+Y)^k\right].
\end{aligned} \tag{41}
$$

To show the converse implication, we use the well-known fact that, if the joint moment generating function of (X, Y) exists in an open neighborhood of the origin, the knowledge of all derivatives of their joint characteristic function $\phi_{XY}(\xi,\gamma)$ at the origin uniquely determines $\phi_{XY}(\xi,\gamma)$ for all $(\xi,\gamma) \in \mathbb{R}^2$ (through a convergent Taylor expansion and an analytic continuation scheme). Then, consider the function $\tau(\xi) = \phi_{XY}(\xi,0)\phi_{XY}(0,\xi) - \phi_{XY}(\xi,\xi)$. By the same reasoning, the knowledge of all derivatives of $\tau(\xi)$ at the origin uniquely determines its value at all $\xi \in \mathbb{R}$. But the kth derivative of $\tau(\xi)$ at the origin is simply given by Eq. (15), as was shown above. Thus, if Eq. (15) holds for all k, $\left[d\tau^k(\xi)/d\xi^k\right]_{\xi=0} = 0$ for all k and a Taylor expansion, implies that $\tau(\xi)$ must vanish for all $\xi \in \mathbb{R}$. Subindependence thus holds. ■

Proof of Theorem 15. By construction, the distribution of Z_i is the true distribution of the sum of X_i and Y_i, i.e. the left-hand side of Eq. (1). The distribution of $W_i = X_i + Y_{r_i}$ can be shown to be the convolution of the marginal distributions of X_i and of Y_i (i.e. the right-hand side of (1)) up to a negligible remainder. If $r_i \neq i$, then clearly X_i and Y_{r_i} are independent by Assumption 3. There is a probability n^{-1} that $r_i = i$ in which case X_i and Y_i are not independent, however this only "misplaces" of the order of one point of the empirical cdf and thus introduces an error of order n^{-1} in the empirical cdf, which is negligible relative to the statistical noise of the Kolmogorov–Smirnov statistic. Since X_i and Y_{r_i} are independent up to a negligible remainder, the distribution of $W_i = X_i + Y_{r_i}$ is the convolution of the marginal of X_i and of Y_{r_i} up to a negligible error, as desired.

The two-sample Kolmogorov–Smirnov test requires the two samples $(Z_i)_{i=1}^n$ and $(W_i)_{i=1}^n$ to be independent. This is not the case here because these variables have X_i in common. This potential problem can be avoided by simply drawing at random another permutation vector s_i, with each element drawn uniformly from $\{1, \ldots, n\}$ (**without** replacement) and defining $\tilde{W}_i = X_{s_i} + Y_{s_{r_i}}$. The new $\tilde{W}_i$ satisfies the required independence from $(Z_i)_{i=1}^n$ (up to a n^{-1} remainder). However, the two-sample Kolmogorov–Smirnov statistic is invariant under permutation of the sample, so the statistics obtained with $\left(Z_i, \tilde{W}_i\right)$ and with (Z_i, W_i) are numerically identical. This additional randomization trick is thus unnecessary and (Z_i, W_i) can be directly used for testing. ■

References

Bhattacharya, R.N., Rao, R.R., 2010. Normal Approximation and Asymptotic Expansions. SIAM, Philadelphia.
Bonhomme, S., Robin, J.-M., 2010. Generalized non-parametric deconvolution with an application to earnings dynamics. Rev. Econom. Stud. 77, 491–533.
Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. V. Elsevier Science.
Carrasco, M., Florens, J.-P., 2011. Spectral method for deconvolving a density. Econometric Theory 27, 546–581.
Conover, W.J., 1971. Practical Nonparametric Statistics. Wiley, New York.
Ebrahimi, N., Hamedani, G., Soofi, E.S., Volkmer, H., 2010. A class of models for uncorrelated random variables. J. Multivariate Anal. 101, 1859–1871.
Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. 19 (3), 1257–1272.
Fan, J., Truong, Y.K., 1993. Nonparametric regression with errors in variables. Ann. Statist. 21 (4), 1900–1925.
Hamedani, G., 2013. Sub-independence: An expository perspective. Commun. Statist. Theory-Methods 42, 3608–3613.
Hamedani, G.G., Volkmer, H.W., 2009. Comment on 'conditional moments and independence' by A. De Paula. Amer. Statist. 63, 295–295.
Hansen, L.P., 1982. Large sample properties of generalized method of moment estimators. Econometrica 50, 1029–1054.
Hu, Y., Ridder, G., 2010. On deconvolution as a first stage nonparametric estimator. Econometric Rev. 29, 1–32.
Hu, Y., Ridder, G., 2012. Estimation of nonlinear models with measurement error using marginal information. J. Appl. Econometrics 27, 347–385.
Hu, Y., Schennach, S.M., 2008. Instrumental variable treatment of nonclassical measurement error models. Econometrica 76, 195–216.
Kotlarski, I., 1967. On characterizing the gamma and the normal distribution. Pac. J. Math. 20, 69–76.
Li, T., 2002. Robust and consistent estimation of nonlinear errors-in-variables models. J. Econometrics 110, 1–26.
Li, T., Vuong, Q., 1998. Nonparametric estimation of the measurement error model using multiple indicators. J. Multivariate Anal. 65, 139–165.
Lighthill, M.J., 1962. Introduction to Fourier Analysis and Generalized Function. Cambridge University Press, London.
Loève, M., 1977. Probability Theory I. Springer, New York.
Lukacs, E., 1970. Characteristic Functions, second ed. Griffin, London.
Schennach, S.M., 2004. Estimation of nonlinear models with measurement error. Econometrica 72, 33–75.
Schennach, S.M., 2007. Instrumental variable estimation of nonlinear errors-in-variables models. Econometrica 75, 201–239.
Schennach, S.M., 2008. Quantile regression with mismeasured covariates. Econometric Theory 24, 1010–1043.
Schennach, S.M., 2013. Measurement error in nonlinear models — A review. In: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), Advances in Economics and Econometrics, Vol. 3. Cambridge University Press, pp. 296–337, chap. 8.
Schennach, S.M., 2014. Entropic latent variable integration via simulation. Econometrica 82, 345–386.
Schennach, S.M., 2016. Recent advances in the measurement error literature. Annu. Rev. Econ. 8, 341–377.
Taupin, M.-L., 2001. Semi-parametric estimation in the nonlinear structural errors-in-variables model. Ann. Statist. 29, 66–93.
Wang, L., Hsiao, C., 2011. Method of moments estimation and identifiability of nonlinear semiparametric errors-in-variables models. J. Econometrics 165, 30–44.
Wilhelm, D., 2015. Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error, cemmap working paper CWP34/15.