Kotlarski with a Factor Loading

Arthur Lewbel[*]   Boston College

Original March 2020, revised December 2020

Abstract

This note extends the Kotlarski (1967) Lemma to show exactly what is identified when we allow for an unknown factor loading on the common unobserved factor. That is, this note completely characterizes identification of the model Y = cV + U and X = V + W, where the joint distribution of Y and X is known, while the constant c and the mutually independent random variables V, U, and W are unobserved. Potential applications include measurement error models and panel data factor models.

Suppose we observe, or can identify from data, the joint distribution of two random variables Y and X. Suppose that

$$
\begin{aligned}
Y &= cV + W \\
X &= V + U
\end{aligned}
$$

where U, V, and W are unobserved, mutually independent real valued random variables, and c is a nonzero constant. If we knew that $c = 1$, we could apply Kotlarski's (1967) Lemma to show that the distributions of U, V, and W are point identified. Since its introduction into the econometrics literature by Li and Vuong (1998), Kotlarski's Lemma has been widely applied. Prominent examples include Bonhomme and Robin (2010) and Cunha, Heckman, and Schennach (2010).

This note extends Kotlarski's Lemma to the case of unknown c. Much of this extension is a one line proof, combining Kotlarski (1967) with a theorem of Reiersøl (1950), and the rest exploits bounds like those of Frisch (1934). Given the age of these references, it's a bit surprising that the results below hadn't been previously made explicit.

The main result is that the constant c and the distributions of U, V, and W are all point identified unless V is normal, and in addition, either U or W contains a normal factor. When c is not point identified, the sharp identified set for c is an interval.

This extension of Kotlarski is useful because many economic models have error structures of the above type, where V is an unobserved factor that affects outcomes Y and X, and c is a factor loading. Here are two examples of such models:

Example 1: Measurement Error in Linear Regression: Suppose X is the observed, mismeasured version of some unobserved true variable X^*, so $X = X^* + U$ where U is measurement error. Consider the regression model $Y = a + bX^* + e$. The goal would be to identify the coefficients a and b and the distributions of the true X^*, the measurement error U, and the model error e. Letting $V = X^*$, $W = a + e$, $a = E(W)$, and $b = c$, we then get that this model is equivalent to our framework of $X = V + U$ and $Y = cV + W$.

Example 2: Fixed T Panel Data Models: Consider the panel data model $Y_{it} = g_t(Z_{it}) + \lambda_t V_i + \varepsilon_{it}$, with unknown functions g_t, unobserved individual specific effects V_i, unknown constant factor loadings λ_t, and idiosyncratic errors ε_{it}. For the moment, assume $g_t(Z_{it})$ is just a constant α_t. The goal is then to identify α_t, the distribution of the individual specific effects V_i, each factor loading λ_t, and the distributions of the idiosyncratic effects ε_{it} in each period t. Impose the free normalization $\lambda_1 = 1$, and consider any other time period s. If we define $U_i = \alpha_1 + \varepsilon_{i1}$, $W_i = \alpha_s + \varepsilon_{is}$, and $\lambda_s = c$, we get that $\alpha_t + \lambda_t V_i + \varepsilon_{it}$ equals $V + U$ in period $t = 1$ and equals $cV + W$ in period $t = s$. We may then apply Theorem 1 below, followed by $\alpha_1 = E(U_i)$, $\varepsilon_{i1} = U_i - E(U_i)$, $\alpha_s = E(W_i)$, and $\varepsilon_{is} = W_i - E(W_i)$. This then gives α_t, λ_t, and the distributions of V_i and of ε_{it} for every observed time period t. Finally, if instead of constants α_t we had functions $g_t(Z_{it})$ for observed covariates Z_{it}, we could repeat the analysis conditioning on Z_{i1} and Z_{is}, and replace the constructions of α_{i1} and α_{is} with conditional means, conditioning on Z_{i1} and Z_{is}, respectively. Evdokimov (2010) uses the Kotlarski Lemma to identify a similar panel structure without factor loadings.

A number of special cases of this paper's result already appear in the literature, referring either directly to the system $Y = cV + W$ and $X = V + U$, or to the equivalent measurement error model of example 1 above. For example, the factor loading c (or equivalently, the mismeasured regressor coefficient b) is known to be point identified when V is asymmetrically distributed. This fact is used by estimators proposed in Lewbel (1997) and Erickson and Whited (2002). Other papers assume more generally that the factor loading is point identified by higher moments, but do not explicitly characterize when that is possible (or what is identifiable in that case). Examples include Bonhomme and Robin (2010), Fruehwirth, Navarro, and Takahashi (2016), and Navarro and Zhou (2017). Generalizations of Kotlarski's lemma to models with more components include Székely and Rao (2000), Li and Zheng (2019), and Lewbel, Schennach, and Zhang (2020).[1]

ASSUMPTION A1: We observe the joint distribution of two real valued, nondegenerate random variables Y and X.

ASSUMPTION A2: $X = V + U$ and $Y = cV + W$, where U, V, and W are mutually

---

[1]Lewbel, Schennach, and Zhang (2020), which is a later working paper to the present paper, considers a more general model where Y also depends on X.

independent, real valued, nondegenerate random variables with unknown distributions. The unknown constant c is real valued, finite, and nonzero.

ASSUMPTION A3: Either the median or the mean of U, V, or W is zero. The characteristic functions of U, V, and W do not vanish.

Assumption A3 is mainly used for identification of the distributions of U, V, and W, not for the identification of c. Kotlarski's Lemma requires some location normalization, as in Assumption A3. Evdokimov and White (2012) provide alternative conditions under which Kotlarski's Lemma holds even when the characteristic functions of U, V, and/or W can have zeros.

Kotlarski's Lemma assumes $c = 1$. We assume $c \neq 0$ because, if $c = 0$ then trivially we can only identify the distributions of W and of $V + U$. Moreover, we can immediately tell if $c = 0$, because in that case the distributions of X and Y will be independent.

For any random variables R and S, let $\sigma_R^2 = var(R)$ if this variance exists, and let $\sigma_{RS} = cov(R, S)$ if this covariance exists. Also let $\psi_R(t) = \ln E\left[\exp(itR)\right]$, the log characteristic function (also known as the cumulant generating function) of R, and similarly $\psi_{R,S}(t_1, t_2) = \ln E\left[\exp(it_1 R + it_2 S)\right]$.

We begin with a tiny Lemma:

LEMMA 1: Let Assumptions A1, A2, and A3 hold. If the constant c is point identified, then the distributions of U, V, and W are all also point identified.

Proof of Lemma 1: When c is point identified, we have $Y/c = V + W/c$. We can then apply Kotlarski's Lemma to the observed joint distribution of Y/c and X to identify the distributions of U, V, and W/c, and hence also identify the distribution of W.

Next we have a theorem from model B of Reiersøl (1950).

THEOREM 1: Let Assumptions A1, A2, and A3 hold. If V is not normally distributed, then the constant c is point identified from the equation

$$\psi_{Y,X}(t_1, t_2) - \psi_Y(t_1) - \psi_X(t_2) = \psi_V(t_1 c + t_2) - \psi_V(t_1 c) - \psi_V(t_2)$$

otherwise, V is normally distributed and

$$\psi_{Y,X}(t_1, t_2) - \psi_Y(t_1) - \psi_X(t_2) = -\sigma_V^2 c t_1 t_2$$

The proof of Theorem 1 is given in Reiersøl (1950). Note that the left side of the equations in Theorem 1 are identified from the joint distribution of Y and X. That the equations in Theorem 1 hold can be immediately verified by the definition of the log characteristic function and the independence of U, V, and W. The deep result in Reiersøl (1950) is proving that c is point identified from these equations whenever V is non-normal, even though the distribution of V, and hence the log characteristic function ψ_V, is unknown. When V is normal, these

equations give a bound on c (it must lie between zero and the coefficient of $-t_1 t_2$), but this bound is tightened below.

Lemma 1 and Theorem 1 together show how to tell if V is normal or not, and show that Kotlarski's Lemma extends to point identification with an unknown factor loading c as long as V is non-normal.

Now consider the case where V is normal. For this case, we need some more notation. For a random variable R, define R's "largest normal factor" to be the variable $\widetilde{R}$ having the maximum variance such that $R = \widetilde{R} + \overline{R}$, where $\widetilde{R}$ and $\overline{R}$ are independently distributed and $\widetilde{R}$ is normally distributed. Without loss of generality, assume $\widetilde{R}$ has mean zero. Call $\overline{R}$ the non-normal factor. If no normal $\widetilde{R}$ exists, then R does not have a normal factor, and in this case we can let $\widetilde{R} = 0$ and $\overline{R} = R$. If R is normal then $\widetilde{R} = R - E(R)$ and $\overline{R} = E(R)$. See Schennach and Hu (2013) and Lewbel, Schennach, and Zhang (2020) for a similar use of normal factors. Reiersøl (1950) calls a normal factor a normal divisor.

Given a random variable R, the variance of $\widetilde{R}$ can be determined by

$$\sigma_{\widetilde{R}}^2 = \sup \left\{ \sigma^2 \in \mathbb{R}^+ : \psi_R(t) + t^2 \sigma^2 / 2 \text{ is a log characteristic function} \right\}$$

If $\sigma_{\widetilde{R}}^2 = 0$ then R does not have a normal factor, otherwise, $\sigma_{\widetilde{R}}^2$ given by this expression is the variance of the largest normal factor $\widetilde{R}$. This follows immediately from the definition of a characteristic function, since a positive $\sigma_{\widetilde{R}}^2$ means by construction that R equals the convolution of two independent random variables, one of which has the log characteristic function of a mean zero normal.[2] This means that if R has a known distribution, and hence a known characteristic function, we can determine if it has a normal factor or not, and we can point identify the distributions of $\widetilde{R}$ and $\overline{R}$.

THEOREM 2: Let Assumptions A1, A2, and A3 hold. Assume V is normally distributed. Then $\sigma_{\widetilde{X}\widetilde{Y}}$, $\sigma_{\widetilde{X}}^2$, and $\sigma_{\widetilde{Y}}^2$ are identified. If $\sigma_{\widetilde{X}\widetilde{Y}}/\sigma_{\widetilde{X}}^2 = \sigma_{\widetilde{Y}}^2/\sigma_{\widetilde{X}\widetilde{Y}}$ then c is point identified by $c = \sigma_{\widetilde{X}\widetilde{Y}}/\sigma_{\widetilde{X}}^2 = \sigma_{\widetilde{Y}}^2/\sigma_{\widetilde{X}\widetilde{Y}}$ and in this case neither W nor U have a normal factor. Otherwise, c is interval identified by $c \in \left[\sigma_{\widetilde{X}\widetilde{Y}}/\sigma_{\widetilde{X}}^2, \sigma_{\widetilde{Y}}^2/\sigma_{\widetilde{X}\widetilde{Y}} \right]$, and for each value of c in this interval, there is a corresponding, identified unique distribution for U, V, and W. This interval bound on c is sharp.

The fact that c is point identified when neither W nor U have a normal factor also appears in Reiersøl (1950). The identified sets in Theorem 2 are new, but are closely related to the Frisch (1934) bounds on mismeasured linear regressions. Taken together, Lemma 1, Theorem 1, and Theorem 2 completely characterize the identification of our model.

Proof of Theorem 2: Separating Y and X into their normal and non-normal factors, we have $Y = \widetilde{Y} + \overline{Y}$ and $X = \widetilde{X} + \overline{X}$. Similarly, Separating W and U into normal and non-normal factors, we also have $Y = cV + \widetilde{W} + \overline{W}$ and $X = V + \widetilde{U} + \overline{U}$. When V is normal, this implies $\widetilde{Y} = cV + \widetilde{W}$, $\overline{Y} = \overline{W}$, $\widetilde{X} = V + \widetilde{U}$ and $\overline{X} = \overline{U}$. This in turn means that, with V

---

[2]An explicit mathematical expression for "being a characteristic function" and hence defining $\sigma_{\widetilde{R}}^2$ can be obtained from Bochner's Theorem, e.g., Theorem 4.2.2 in Lukacs (1970).

normal, $\overline{X}$ and $\overline{Y}$ are independent of each other and of the joint distribution of $\widetilde{Y}$ and $\widetilde{X}$. Since the marginal distributions of $\overline{Y}$ and $\overline{X}$ are identified, we can identify the left side of

$$\psi_{Y,X}(t_1, t_2) - \psi_{\overline{Y}}(t_1) - \psi_{\overline{X}}(t_2) = \psi_{\widetilde{Y},\widetilde{X}}(t_1, t_2)$$

And therefore the joint normal distribution of the mean zero variables $\widetilde{Y}$ and $\widetilde{X}$ is identified. In particular, this means that $\sigma^2_{\widetilde{Y}}$, $\sigma^2_{\widetilde{X}}$, and $\sigma_{\widetilde{X}\widetilde{Y}}$ are identified.

The remaining step now borrows heavily from the Frisch (1934) bounds on mismeasured linear regression. From the identified second moments of $\widetilde{Y}$ and $\widetilde{X}$, we have $\sigma^2_{\widetilde{Y}} = c^2\sigma^2_V + \sigma^2_{\widetilde{W}}$, $\sigma^2_{\widetilde{X}} = \sigma^2_V + \sigma^2_{\widetilde{U}}$, and $\sigma_{\widetilde{X}\widetilde{Y}} = c\sigma^2_V$, which provides three equations in the four unknown constants $\sigma^2_{\widetilde{U}}$, $\sigma^2_{\widetilde{W}}$, σ^2_V, and c. The only constraints on these parameter values are that $c \neq 0$, $\sigma^2_{\widetilde{U}}$ and $\sigma^2_{\widetilde{W}}$ must be non-negative (either can be zero if the corresponding normal factor doesn't exist), and σ^2_V must be positive. These being the only constraints is what makes the corresponding bounds be sharp. The equation $\sigma_{\widetilde{X}\widetilde{Y}} = c\sigma^2_V$ means that the sign of c equals the sign of $\sigma_{\widetilde{X}\widetilde{Y}}$ to ensure $\sigma^2_V > 0$. Then $\sigma^2_{\widetilde{U}} \geq 0$ requires $\sigma^2_{\widetilde{X}} - \sigma_{\widetilde{X}\widetilde{Y}}/c \geq 0$ and $\sigma^2_{\widetilde{W}} \geq 0$ requires $\sigma^2_{\widetilde{Y}} - c\sigma_{\widetilde{X}\widetilde{Y}} \geq 0$. Therefore, either $\sigma_{\widetilde{X}\widetilde{Y}} > 0$ and $\sigma_{\widetilde{X}\widetilde{Y}}/\sigma^2_{\widetilde{X}} \leq c \leq \sigma^2_{\widetilde{Y}}/\sigma_{\widetilde{X}\widetilde{Y}}$, or $\sigma_{\widetilde{X}\widetilde{Y}} < 0$ and $\sigma^2_{\widetilde{Y}}/\sigma_{\widetilde{X}\widetilde{Y}} \leq c \leq \sigma_{\widetilde{X}\widetilde{Y}}/\sigma^2_{\widetilde{X}}$. Either way c lies in the interval between $\sigma_{\widetilde{X}\widetilde{Y}}/\sigma^2_{\widetilde{X}}$ and $\sigma^2_{\widetilde{Y}}/\sigma_{\widetilde{X}\widetilde{Y}}$, and this interval does not include zero.

Finally, given any c that lies in this interval, there's an identified corresponding unique distribution for $\widetilde{U}$, V, and $\widetilde{W}$ that satisfies the assumptions, given by $V \sim N\left(0, \sigma_{\widetilde{X}\widetilde{Y}}/c\right)$, $\widetilde{U} \sim N\left(0, \sigma^2_{\widetilde{X}} - \sigma_{\widetilde{X}\widetilde{Y}}/c\right)$, and $\widetilde{W} \sim N\left(0, \sigma^2_{\widetilde{Y}} - c\sigma_{\widetilde{X}\widetilde{Y}}\right)$, and corresponding identified unique distributions for U and W given by $U = \widetilde{U} + \overline{X}$ and $W = \widetilde{W} + \overline{Y}$.

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.

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

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitiveand Noncognitive Skill Formation," Econometrica, 78, 883–931.

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

Evdokimov, K. (2010), "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Unpublished Manuscript.

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

Frisch, R. (1934), "Statistical confluence analysis by means of complete regression systems," Vol. 5, Universitetets Økonomiske Instituut.

Kotlarski, I. I. (1967), "On characterizing the gamma and normal distribution," Pacific 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., S. M. Schennach, and L. Zhang (2020), "Identification of a Triangular Two Equation System Without Instruments," Centre for microdata methods and practice working paper CWP41/20.

Li, S. and X. Zheng, (2019), "A Generalization of Lemma 1 in Kotlarski (1967)," Rice University unpublished manuscript.

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

Lukacs, E. (1970), "Characteristic Functions", Second edition, Griffin, London.

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.

Reiersøl, O. (1950), "Identifiability of a linear relation between variables which are subject to error," Econometrica, 18, 375-389.

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

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