USC Dornsife
Center for Economic and Social Research

USC Schaeffer
Leonard D. Schaeffer Center for Health Policy & Economics

Kotlarski's lemma for dyadic models

Grigory Franguridi, Hyungsik Roger Moon

Paper No: 2026-002

CESR-SCHAEFFER WORKING PAPER SERIES

Kotlarski's lemma for dyadic models*

Grigory Franguridi[1] and Hyungsik Roger Moon[2]

[1]Center for Economic and Social Research, University of Southern California

[2]Department of Economics, University of Southern California

January 16, 2026

Abstract

We show how to identify the distributions of the latent components in the two-way dyadic model for bipartite networks $y_{i,\ell} = \alpha_i + \eta_\ell + \varepsilon_{i,\ell}$. This is achieved by a repeated application of the extension of the classical lemma of Kotlarski (1967) in Evdokimov and White (2012). We provide two separate sets of assumptions under which all the latent distributions are identified. Both rely on some of the latent components being identically distributed.

JEL Classification: C23

Keywords: Kotlarski lemma, deconvolution, dyadic data, two-way error component, bipartite network

1 Introduction

Identifying latent variables from observed data is a central problem in econometrics. One of the important tools for addressing this problem is a lemma of Kotlarski (1967) and its variants, which provide conditions under which the distribution (characteristic function) of latent variables is identified.

Kotlarski's lemma has been used for identification, estimation, and inference in a variety of economic settings, such as measurement error models (Li and Vuong, 1998; Li, 2002; Schennach, 2004; Kurisu and Otsu, 2022), auctions (Li et al., 2000; Krasnokutskaya, 2011; Grundl and Zhu, 2019; Andreyanov and Caoui, 2022), and models of earning dynamics (Bonhomme and Robin, 2010; Botosaru and Sasaki, 2018; Hu et al., 2019). More recently, Kotlarski's lemma has been used for robust inference in Kato et al. (2021). Finally, generalizations of Kotlarski's lemma exist to the cases of multiple error components or unknown factor loadings (Székely and Rao, 2000; Li and Zheng, 2020; Lewbel, 2022;

*We thank an anonymous referee for helpful comments and suggestions.

Lewbel et al., 2024). For a more complete overview of the variants and uses of Kotlarski's lemma, see, e.g., Schennach (2016).

The classical Kotlarski lemma used in most applications assumes repeated measurement with a common latent variable and independent errors, for example, $y_{i,\ell} = c + \alpha_i + \varepsilon_{i,\ell}$, where $\ell = a, b$ for each i. In this note, we show how to use the classical Kotlarski lemma to identify the standard two-way dyadic model for bipartite networks, $y_{i,\ell} = c + \alpha_i + \eta_\ell + \varepsilon_{i,\ell}$. We discuss two cases: the partially connected bipartite network and the fully connected bipartite network.

Our main theorem relies on the version of Kotlarski's lemma in Evdokimov and White (2012), and hence does not assume that the characteristic functions (CF) of the error components have no zeros, which would rule out many distributions of interest, including all continuous distributions with compact support and many discrete distributions. Instead, the CFs are allowed to have zeros, as long as they do not overlap with zeros of their first derivatives.

2 Classical lemma by Kotlarski

In this section, we briefly discuss the classical lemma by Kotlarski (1967) and its extension in Evdokimov and White (2012). Suppose we observe two repeated noisy measurements X_1, X_2 of a variable M,

$$X_1 = M + U_1,$$
$$X_2 = M + U_2,$$

where U_1, U_2 are noise variables. Assume that M, U_1, U_2 are jointly independent and $\mathbb{E}[U_1] = 0$. The goal is to identify the distributions of M, U_1, U_2.

Kotlarski's lemma states that, if the CFs ϕ_M, ϕ_{U_1}, and ϕ_{U_2} of M,U_1, and U_2, respectively, are nonvanishing everywhere, then these CFs (and hence the distributions) can be recovered from the CF of the observables X_1, X_2.

The assumption of everywhere nonvanishing CFs rules out many interesting distributions, such as any nondegenerate distribution with compact support and any discrete distribution with finite support. Evdokimov and White (2012) provide an extension of Kotlarski's lemma that relaxes this assumption. Specifically, it only requires that the real zeros of ϕ_{U_1} and its derivative ϕ'_{U_1} are disjoint and that the zeros of ϕ_{U_2} form a set of isolated points, see Assumption A and Lemma 1(b) in Evdokimov and White (2012).

3 Application to dyadic models

3.1 Case of partially connected network

Consider a bipartite network with two sets of nodes, $\{1, 2\}$ and $\{a, b\}$. Assume that the node pairs $(1, a), (1, b)$, and $(2, a)$ are linked, so that the associated variables $y_{1,a}, y_{1,b}$ and $y_{2,a}$ are observed. Notice that we do not assume that nodes 2 and b are linked so that $y_{2,b}$ may be unobserved. Consider the dyadic model

$$
\begin{aligned}
y_{1,a} &= \alpha_1 + \eta_a + \varepsilon_{1,a}, \\
y_{1,b} &= \alpha_1 + \eta_b + \varepsilon_{1,b}, \\
y_{2,a} &= \alpha_2 + \eta_a + \varepsilon_{2,a},
\end{aligned}
$$

where $\alpha_1, \alpha_2, \eta_a, \eta_b$ are unobserved random effects and $\varepsilon_{1,a}, \varepsilon_{1,b}$, and $\varepsilon_{2,a}$ are idiosyncratic errors. We assume that all the latent variables are jointly independent. We are interested in identifying the distributions of all the latent components α_1, α_2, η_a, η_b, $\varepsilon_{1,a}$, $\varepsilon_{1,b}$, and $\varepsilon_{2,a}$ from the distributions of $y_{1,a}$, $y_{1,b}$, and $y_{2,a}$.

Our identification strategy consists of two parts. First, we identify the distributions of α_1, η_a, and $\varepsilon_{1,a}$. Then, we provide two sets of restrictions under which the remaining distributions are identified: (i) the equality of the distributions of $\varepsilon_{1,a}$, $\varepsilon_{1,b}$, and $\varepsilon_{2,b}$ (Assumption 4) and (ii) the equality of the distributions of α_1 and α_2 and those of η_a and η_b (Assumption 5).

Let us now provide the intuition on how the distributions of α_1, η_a, and $\varepsilon_{1,a}$ can be identified by a repeated application of Kotlarski's lemma. First, for a pair $(1, a), (1, b)$, write

$$
\begin{aligned}
y_{1,a} &= \alpha_1 + \eta_a + \varepsilon_{1,a} = M + U_1, \\
y_{1,b} &= \alpha_1 + \eta_b + \varepsilon_{1,b} = M + U_2,
\end{aligned}
$$

where $M = \alpha_1$, $U_1 = \eta_a + \varepsilon_{1,a}$, and $U_2 = \eta_b + \varepsilon_{1,b}$. By Kotlarski's lemma, the CF ϕ_{α_1} is identified. Similarly, for a pair $(1, a), (2, a)$, write

$$
\begin{aligned}
y_{1,a} &= \alpha_1 + \eta_a + \varepsilon_{1,a} = \tilde{M} + \tilde{U}_1, \\
y_{2,a} &= \alpha_2 + \eta_a + \varepsilon_{2,a} = \tilde{M} + \tilde{U}_2,
\end{aligned}
$$

where $\tilde{M} = \eta_a$, $\tilde{U}_1 = \alpha_1 + \varepsilon_{1,a}$, and $\tilde{U}_2 = \alpha_2 + \varepsilon_{2,a}$. By Kotlarski's lemma, the CF ϕ_{η_a} is identified. Joint independence of α_1, η_a, and $\varepsilon_{1,a}$ implies

$$\phi_{\varepsilon_{1,a}}(t) = \frac{\phi_{y_{1,a}}(t)}{\phi_{\alpha_1}(t)\phi_{\eta_a}(t)},$$

identifying the distribution of $\varepsilon_{1,a}$, and hence the distributions of $\varepsilon_{1,b}$ and $\varepsilon_{2,b}$. Finally, joint independence of α_2, η_a, and $\varepsilon_{2,a}$ implies

$$\phi_{\alpha_2}(t) = \frac{\phi_{y_{2,a}}(t)}{\phi_{\eta_a}(t)\phi_{\varepsilon_{2,a}}(t)}, \tag{1}$$

and joint independence of α_1, η_b, and $\varepsilon_{1,b}$ implies

$$\phi_{\eta_b}(t) = \frac{\phi_{y_{1,b}}(t)}{\phi_{\alpha_1}(t)\phi_{\varepsilon_{1,b}}(t)}, \tag{2}$$

identifying the distributions of the remaining components α_2 and η_b.

We now state the assumptions needed to make the intuition above rigorous. For a random variable ζ, denote by $\mathcal{Z}_\zeta$ the set of zeros of its CF ϕ_ζ and denote by $\mathcal{Z}'_\zeta$ the set of zeros of the derivative ϕ'_ζ of its CF.

Assumption 1. *(i) $\alpha_1, \alpha_2, \eta_a, \eta_b, \varepsilon_{1,a}, \varepsilon_{1,b}, \varepsilon_{2,a}$ are integrable with zero means.*

(ii) $\alpha_1, \alpha_2, \eta_a, \eta_b, \varepsilon_{1,a}, \varepsilon_{1,b}, \varepsilon_{2,a}$ are jointly independent.

Assumption 2. *(i) The sets $\mathcal{Z}_{\alpha_1}, \mathcal{Z}_{\eta_a}, \mathcal{Z}_{\varepsilon_{1,a}}$ are pairwise disjoint.*

(ii) The sets $\mathcal{Z}_{\alpha_1}$ and $\mathcal{Z}'_{\alpha_1}$ are disjoint.

(iii) The sets $\mathcal{Z}_{\eta_a}$ and $\mathcal{Z}'_{\eta_a}$ are disjoint.

(iv) The sets $\mathcal{Z}_{\varepsilon_{1,a}}$ and $\mathcal{Z}'_{\varepsilon_{1,a}}$ are disjoint.

(v) The sets $\mathcal{Z}_{\alpha_2}, \mathcal{Z}_{\eta_b}, \mathcal{Z}_{\varepsilon_{1,b}}, \mathcal{Z}_{\varepsilon_{2,a}}$ consist of isolated points.

In the identification strategy described above, we apply Kotlarski's lemma in the case when the error terms $U_1, U_2, \tilde{U}_1, \tilde{U}_2$ consist of two latent components. The lemma relies on assumptions about these error terms, which are not primitives of our dyadic model. Instead, we want to restrict the latent components in a way that would guarantee that the assumptions on the error terms hold. The following abstract result shows how this can be achieved.

Lemma 1. *Let $A = B + C$, where B is independent of C. Assume that*

1. *$\mathcal{Z}_B \cap \mathcal{Z}_C = \varnothing$,*
2. *$\mathcal{Z}_B \cap \mathcal{Z}'_B = \varnothing$,*

3. $\mathcal{Z}_C \cap \mathcal{Z}'_C = \varnothing$.

Then $\mathcal{Z}_A \cap \mathcal{Z}'_A = \varnothing$.

Proof. Take any $t \in \mathcal{Z}_A$. Then either (i) $t \in \mathcal{Z}_B$ or (ii) $t \in \mathcal{Z}_C$. In the case (i), we have $\phi'_A(t) = \phi'_B(t)\phi_C(t) + \phi_B(t)\phi'_C(t) = \phi'_B(t)\phi_C(t)$. By condition 1, $t \notin \mathcal{Z}_C$, and by condition 2, $t \notin \mathcal{Z}'_B$. Therefore, $\phi'_A(t) \neq 0$ and so $t \notin \mathcal{Z}'_A$. Case (ii) is analogous. □

We are now ready to state our adaptation of the main theorem of Evdokimov and White (2012) to dyadic data.

Theorem 1. *Under Assumptions 1 and 2,* ϕ_{α_1} *and* ϕ_{η_a} *are identified and*

$$\phi_{\varepsilon_{1,a}}(s) = \frac{\phi_{y_{1,a}}(s)}{\phi_{\alpha_1}(s)\phi_{\eta_a}(s)}, \quad s \notin \mathcal{Z}_{\alpha_1} \cup \mathcal{Z}_{\eta_a}.$$

Proof. We use the notations $M, U_1, U_2, \tilde{M}, \tilde{U}_1, \tilde{U}_2$ from the discussion above. Assumptions 2(i), 2(iii), 2(iv) and Lemma 1 imply that the zeros of ϕ_{U_1} and ϕ'_{U_1} are disjoint. Assumption 2(v) implies that the zeros of ϕ_{U_2} are isolated. Combining with Assumption 1 proves Assumption A in Evdokimov and White (2012). Invoking their Lemma 1(b) establishes identification of ϕ_{α_1}.

Similarly, Assumptions 2(i), 2(ii), 2(iv) and Lemma 1 imply that the zeros of $\phi_{\tilde{U}_1}$ and $\phi'_{\tilde{U}_1}$ are disjoint. Assumption 2(v) implies that the zeros of $\phi_{\tilde{U}_2}$ are isolated. Combining with Assumption 1 proves Assumption A in Evdokimov and White (2012). Invoking their Lemma 1(b) establishes identification of ϕ_{η_a}.

Finally, independence implies $\phi_{y_{1,a}}(t) = \phi_{\alpha_1}(t)\phi_{\eta_a}(t)\phi_{\varepsilon_{1,a}}(t)$, completing the proof. □

Theorem 1 establishes the distributional identification of α_1 and η_a and also shows that $\phi_{\varepsilon_{1,a}}$ is identified at all points of the real line except for zeros of ϕ_{α_1} or ϕ_{η_a}. We now provide two sets of assumptions under which *all* the latent distributions are identified (which we call *total identification*). Both sets of assumptions include the following condition.

Assumption 3. *The sets* $\mathcal{Z}_{\alpha_1}$ *and* $\mathcal{Z}_{\eta_a}$ *consist of isolated points.*

Let us show that total identification holds under the distributional homogeneity of the three error terms $\varepsilon_{1,a}, \varepsilon_{1,b}, \varepsilon_{2,a}$.

Assumption 4. $\varepsilon_{1,a} \stackrel{d}{=} \varepsilon_{1,b} \stackrel{d}{=} \varepsilon_{2,a}$.

Corollary 1. *Under Assumptions 1, 2, 3, and 4, the distributions of all the latent variables* α_1, α_2, η_a, η_b, $\varepsilon_{1,a}$, $\varepsilon_{1,b}$, *and* $\varepsilon_{2,b}$ *are identified.*

Proof. By Assumption 3, $\mathcal{Z}_{\alpha_1} \cup \mathcal{Z}_{\eta_a}$ consists of isolated points, and hence the formula for $\phi_{\varepsilon_{1,a}}(s)$ in Theorem 1 can be extended to all $s \in \mathbb{R}$ by continuity. This identifies the distribution of $\phi_{\varepsilon_{1,a}}$, and, in view of Assumption 4, the distributions of $\varepsilon_{1,b}$ and $\varepsilon_{2,a}$. The formulas (1) and (2) then identify the distributions of α_2 and η_b. □

Next, we show that total identification holds under the distributional homogeneity of the random effects, α_1, α_2 and η_a, η_b.

Assumption 5. $\alpha_1 \overset{d}{=} \alpha_2$ *and* $\eta_a \overset{d}{=} \eta_b$.

Corollary 2. *Under Assumptions 1, 2, 3, and 5, the distributions of all the latent variables* α_1, α_2, η_a, η_b, $\varepsilon_{1,a}$, $\varepsilon_{1,b}$, *and* $\varepsilon_{2,b}$ *are identified.*

Proof. By Theorem 1, ϕ_{α_1} and ϕ_{η_a} are identified everywhere. By Assumption 5, $\phi_{\alpha_1} = \phi_{\alpha_2} := \phi_\alpha$ and $\phi_{\eta_a} = \phi_{\eta_b} := \phi_\eta$. Finally, joint independence yields

$$\phi_{\varepsilon_{1,b}} = \frac{\phi_{y_{1,b}}(s)}{\phi_\alpha(s)\phi_\eta(s)}, \quad s \notin \mathcal{Z}_\alpha \cup \mathcal{Z}_\eta,$$
$$\phi_{\varepsilon_{2,a}} = \frac{\phi_{y_{2,a}}(s)}{\phi_\alpha(s)\phi_\eta(s)}, \quad s \notin \mathcal{Z}_\alpha \cup \mathcal{Z}_\eta.$$

By Assumption 3, $\mathcal{Z}_\alpha \cup \mathcal{Z}_\eta$ consists of isolated points, and hence the formulas above can be extended by continuity. This identifies the distributions of $\varepsilon_{1,b}$ and $\varepsilon_{2,a}$. □

3.2 Case of fully connected network

When the bipartite network on the node sets $\{1, 2\}$ and $\{a, b\}$ is fully connected, i.e., when all the node pairs $(1, a), (1, b), (2, a)$, and $(2, b)$ are linked, the distributions of all the latent variables α_1, α_2, η_a, η_b, $\varepsilon_{1,a}$, $\varepsilon_{1,b}$, $\varepsilon_{2,a}$, and $\varepsilon_{2,b}$ are identified without any assumptions on the distribution homogeneity (cf. Assumptions 4 and 5). To see that, notice that applying Kotlarski's lemma to the pair $y_{i,a}, y_{i,b}$ identifies the distribution of α_i, $i = 1, 2$. Then applying the lemma to the pair $y_{1,c}, y_{2,c}$ identifies the distribution of η_c, $c = a, b$. Finally, the distribution of $\varepsilon_{i,c}$ is identified via

$$\phi_{\varepsilon_{i,c}}(t) = \frac{\phi_{y_{i,c}}(t)}{\phi_{\alpha_i}(t)\phi_{\eta_c}(t)}, \quad i = 1, 2, \ \ c = a, b.$$

Formulating rigorous conditions under which this identification strategy is valid can be done along the lines of Section 3.1.

4 Conclusion

We show how the classical lemma of Kotlarski can be employed to identify distributions of latent components in dyadic models for bipartite networks with two-way random effects. When the bipartite graph is partially linked, we provide two sets of assumptions under which all the latent distributions are identified. The first set of assumptions restricts the errors to be identically distributed. The second set of assumptions restricts the random effects to be identically distributed. Our identification result can be used to develop an estimation procedure for this class of models. This is beyond the scope of the current paper, and we defer this agenda to future work.

References

Andreyanov, P. and E. H. Caoui (2022): "Secret reserve prices by uninformed sellers," *Quantitative Economics*, 13, 1203–1256.

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.

Botosaru, I. and Y. Sasaki (2018): "Nonparametric heteroskedasticity in persistent panel processes: An application to earnings dynamics," *Journal of Econometrics*, 203, 283–296.

Evdokimov, K. and H. White (2012): "Some extensions of a lemma of Kotlarski," *Econometric Theory*, 28, 925–932.

Grundl, S. and Y. Zhu (2019): "Identification and estimation of risk aversion in first-price auctions with unobserved auction heterogeneity," *Journal of Econometrics*, 210, 363–378.

Hu, Y., R. Moffitt, and Y. Sasaki (2019): "Semiparametric estimation of the canonical permanent-transitory model of earnings dynamics," *Quantitative Economics*, 10, 1495–1536.

Kato, K., Y. Sasaki, and T. Ura (2021): "Robust inference in deconvolution," *Quantitative Economics*, 12, 109–142.

Kotlarski, I. (1967): "On characterizing the gamma and the normal distribution," *Pacific Journal of Mathematics*, 20, 69–76.

Krasnokutskaya, E. (2011): "Identification and estimation of auction models with unobserved heterogeneity," *The Review of Economic Studies*, 78, 293–327.

Kurisu, D. and T. Otsu (2022): "On the uniform convergence of deconvolution estimators from repeated measurements," *Econometric Theory*, 38, 172–193.

Lewbel, A. (2022): "Kotlarski with a factor loading," *Journal of Econometrics*, 229, 176–179.

Lewbel, A., S. M. Schennach, and L. Zhang (2024): "Identification of a triangular two equation system without instruments," *Journal of Business & Economic Statistics*, 42, 14–25.

Li, S. and X. Zheng (2020): "A generalization of Lemma 1 in Kotlarski (1967)," *Statistics & Probability Letters*, 165, 108814.

Li, T. (2002): "Robust and consistent estimation of nonlinear errors-in-variables models," *Journal of Econometrics*, 110, 1–26.

Li, T., I. Perrigne, and Q. Vuong (2000): "Conditionally independent private information in OCS wildcat auctions," *Journal of econometrics*, 98, 129–161.

Li, T. and Q. Vuong (1998): "Nonparametric estimation of the measurement error model using multiple indicators," *Journal of Multivariate Analysis*, 65, 139–165.

Schennach, S. M. (2004): "Nonparametric regression in the presence of measurement error," *Econometric Theory*, 20, 1046–1093.

——— (2016): "Recent advances in the measurement error literature," *Annual Review of Economics*, 8, 341–377.

Székely, G. and C. Rao (2000): "Identifiability of distributions of independent random variables by linear combinations and moments," *Sankhyā: The Indian Journal of Statistics, Series A*, 193–202.