econometrics

MDPI

Article

When It Counts—Econometric Identification of the Basic Factor Model Based on GLT Structures

Sylvia Frühwirth-Schnatter [1], Darjus Hosszejni [1,*] and Hedibert Freitas Lopes [2,3]

1 Department of Finance, Accounting, and Statistics, WU Vienna University of Economics and Business, 1020 Vienna, Austria; sfruehwi@wu.ac.at
2 School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85281, USA
3 Insper Institute of Education and Research, São Paulo 04546-042, Brazil
* Correspondence: darjus.hosszejni@wu.ac.at

Abstract: Despite the popularity of factor models with simple loading matrices, little attention has been given to formally address the identifiability of these models beyond standard rotation-based identification such as the positive lower triangular (PLT) constraint. To fill this gap, we review the advantages of variance identification in simple factor analysis and introduce the generalized lower triangular (GLT) structures. We show that the GLT assumption is an improvement over PLT without compromise: GLT is also unique but, unlike PLT, a non-restrictive assumption. Furthermore, we provide a simple counting rule for variance identification under GLT structures, and we demonstrate that within this model class, the unknown number of common factors can be recovered in an exploratory factor analysis. Our methodology is illustrated for simulated data in the context of post-processing posterior draws in sparse Bayesian factor analysis.

Keywords: identifiability; sparsity; rank deficiency; rotational invariance; variance identification

JEL Classification: C11; C38; C63

check for updates

Citation: Frühwirth-Schnatter, Sylvia, Darjus Hosszejni, and Hedibert Freitas Lopes. 2023. When It Counts—Econometric Identification of the Basic Factor Model Based on GLT Structures. *Econometrics* 11: 26. https://doi.org/10.3390/econometrics11040026

Academic Editor: Martin Wagner

Received: 24 February 2023
Revised: 29 October 2023
Accepted: 1 November 2023
Published: 20 November 2023

1. Introduction

Ever since the pioneering work of Thurstone (1935, 1947), factor analysis has been a popular method to model the covariance matrix $\boldsymbol{\Omega}$ of correlated, multivariate observations $\mathbf{y}_t$ of dimension m (see, e.g., Anderson (2003) for a comprehensive review). Assuming r uncorrelated factors in a basic factor model, for instance, yields the representation $\boldsymbol{\Omega} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma}_0$, with a $m \times r$ factor loading matrix $\boldsymbol{\Lambda}$ and a diagonal matrix $\boldsymbol{\Sigma}_0$. This considerable reduction in the number of parameters compared to the $m(m+1)/2$ parameters of an unconstrained covariance matrix is the main motivation for applying factor models to covariance estimation, especially if m is large (see, among many others, Fan et al. (2008)). In addition, shrinkage estimation has been shown to lead to very efficient covariance estimation (see, for example, Kastner (2019) in Bayesian factor analysis and Ledoit and Wolf (2020) in a non-Bayesian context).

In numerous applications, factor analysis reaches beyond covariance modeling (see, among many others, Forni et al. (2009) in the context of structural factor models). From the very beginning, the goal of factor analysis has been to extract the underlying loading matrix $\boldsymbol{\Lambda}$ to understand the driving forces behind the observed correlation between the measurements (see, e.g., Owen and Wang (2016) for a recent review). However, also in this setting, the only source of information is the observed covariance of the data, making the decomposition of the covariance matrix $\boldsymbol{\Omega}$ into the cross-covariance matrix $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$ and the variance $\boldsymbol{\Sigma}_0$ of the idiosyncratic errors more challenging than estimating only $\boldsymbol{\Omega}$ itself.

A huge amount of literature, dating back to Koopmans and Reiersøl (1950) and Reiersøl (1950), has addressed this problem of identification which can be resolved only by imposing additional structure on the factor model. Anderson and Rubin (1956) consider two kinds of