

Spurious Factor Analysis

Alexei Onatski* and Chen Wang†

*Faculty of Economics, University of Cambridge.

†Department of Statistics and Actuarial Science,
University of Hong Kong.

September 14, 2018

Abstract

This paper draws parallels between the Principal Components Analysis of factorless high-dimensional nonstationary data and the classical spurious regression. We show that a few of the principal components of such data absorbs nearly all the data variation. The corresponding scree plot suggests that the data contains a few factors, which is corroborated by the standard panel information criteria. Furthermore, the Dickey-Fuller tests of the unit root hypothesis applied to the estimated ‘idiosyncratic terms’ often reject, creating an impression that a few factors are responsible for most of the nonstationarity in the data. We warn empirical researchers of these peculiar effects and suggest to always compare the analysis in levels with that in differences.

KEY WORDS: Spurious regression, principal components, factor models, Karhunen-Loève expansion.

1 Introduction

Researchers applying factor analysis to nonstationary macroeconomic panels face a choice: keep the data in levels or first-difference them. If all the nonstationarity is due to the factors, no differencing is necessary. A simple principal components estimator of the factors is consistent (e.g. Bai, 2004). Otherwise, the standard advice is to extract the factors from the first-differenced data, and then, accumulate them to obtain estimates of the factors in levels (e.g. Bai and Ng, 2004).