Journal of Empirical Finance 17 (2010) 783–802

Contents lists available at ScienceDirect

Journal of Empirical Finance

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

GMM estimation of the number of latent factors: With application to international stock markets

Seung C. Ahn [a], M. Fabricio Perez [b,*]

[a] Department of Economics, W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287, United States
[b] Finance Area of the School of Business and Economics at Wilfrid Laurier University, Waterloo Ontario, Canada N2L 3C5

ARTICLE INFO

Article history:
Received 22 September 2008
Received in revised form 13 April 2010
Accepted 13 April 2010
Available online 22 April 2010

JEL classification:
G1
F2
C1
C3
C4
C5

Keywords:
Factor models
GMM
Number of factors
International stock market
Asset pricing

ABSTRACT

We propose new generalized method of moments (GMM) estimators for the number of latent factors in linear factor models. The estimators are appropriate for data with a large (small) number of cross-sectional observations and a small (large) number of time series observations. The estimation procedure is simple and robust to the configurations of idiosyncratic errors encountered in practice. In addition, the method can be used to evaluate the validity of observable candidate factors. Monte Carlo experiments show that the proposed estimators have good finite-sample properties. Applying the estimators to international stock markets, we find that international stock returns are explained by one strong global factor. This factor is highly correlated with the Fama–French factors from the U.S. stock market. This result can be interpreted as evidence of market integration. We also find two weak factors closely related to markets in Europe and the Americas, respectively.

1. Introduction

Many theories in economics and finance are based on linear factor models. Probably the most documented financial application is the Arbitrage Price Theory (APT, Ross, 1976), in which asset returns are generated by a relatively small number of common factors.[1] Empirical studies based on linear factor models should address two important questions in order to justify the models' theoretical and empirical validity. The first question is how many factors are to be used in the empirical analysis. Determination of the number of factors is critical. If too few factors are used to estimate an APT model, the estimated factor prices are inconsistent. In contrast, if too many factors are used and some of these factors are not correlated with the returns, the risks associated with the "useless" factors can falsely appear priced (Kan and Zhang, 1999). The second question is what observable candidate factors (macroeconomic and/or financial variables) are related to the unobservable true factors. Answering this question makes it possible to provide economic interpretations of the unobservable factors. In this paper we present a new approach to address these two questions. Specifically, we propose a formal procedure to estimate the number of factors and a new test to evaluate the validity of observable candidate factors.

* Corresponding author. Tel.: +1 519 884 0710x2532.
E-mail addresses: miniahn@asu.edu (S.C. Ahn), mperez@wlu.ca (M.F. Perez).

[1] An excellent summary of factor models can be found in Campbell et al. (1997) and Bai (2003).

0927-5398/$ – see front matter
doi:10.1016/j.jempfin.2010.04.005

This paper makes two contributions to the factor models literature. The first contribution is the development of a Generalized Method of Moments (GMM, Hansen, 1982) estimator for the number of latent factors that is appropriate for panel data sets with either a small number of cross-sectional observations (N) or a small number of time series observations (T). Our method is especially relevant to the economic and financial literature since in numerous applications just one dimension of the available data (N or T) is large and the other one is relatively small (e.g., the analysis of portfolio returns, yields on bond indexes, or country common factors). The estimation and test procedures we propose are also computationally simple. All of the procedures are closed-form solutions and thus do not require non-linear optimization. Any software that can estimate multiple equation models can be used.[2]

The method is applicable to data sets with small N and large T as long as the idiosyncratic components of returns are not cross-sectionally correlated (*exact factor structure*). In this case, the errors are allowed to be serially correlated.[3] The method is also relevant for data with small T and large N if the errors are serially uncorrelated. The errors can be cross-sectionally correlated (*approximate factor structure*) in this second case. Additionally, the method is robust to heteroskedasticity in the idiosyncratic components.

If a data set with both large N and T is available, the number of latent factors can be consistently estimated even if the idiosyncratic errors are both cross-sectionally and serially correlated (see Bai and Ng, 2002). Unfortunately, such data sets are not always available. For example, for the studies of bond yields, formation of a small number of portfolios is desirable because many individual bonds are traded infrequently. Additionally, for the studies of international stock markets, the number of countries with developed capital markets and financial information systems is small.[4] Furthermore, several recent studies have shown the advantages of using portfolios over a large set of individual assets. Zhang (2009) shows that systematic factors may not be detected by analyzing individual stock returns and may be better detected by analyzing portfolios of assets. Also, Boivin and Ng (2006) conclude that use of an approximate factor model for the analysis of a large set of cross-sectional observations can produce less reliable forecasts. These studies suggest the importance of portfolio formation in the analysis of factor models. Portfolio analysis reduces both the number of cross-sectional units and the level of cross-sectional correlations. Our methodology is appropriate to deal with such data sets with small N.

The second contribution of this paper is the development of a test to evaluate what observable variables or candidate factors (i.e., macroeconomics or financial variables) are correlated with the true unobservable factors. By analyzing these correlations, researchers are able to give an economic interpretation of the latent factors model. This test can apply to both exact and approximate factor models without additional assumptions (see Ahn et al. (2009) for an exact factor model application, and Eichengreen et al. (2009) for an approximate factor model).

Using GMM to estimate the number of factors is not completely new. Ahn et al. (2007, 2010) show how GMM can be used to estimate the number of latent common factors in panel data models. However, because their method is designed for panel data models with observed regressors, its applicability to latent factor models in financial studies is limited.[5] Our method is more general than the ALS one since it is applicable even if observed regressors correlated with the latent factors are unavailable.

Earlier empirical studies of factor models are based on the maximum likelihood (ML) method of Jöreskog (1967). The ML method requires restrictive distributional assumptions: the idiosyncratic error terms are required to be normal as well as independently and identically distributed over time. Another common method is to construct candidate factors, repeat the estimation and testing of the model for a different number of factors (L), and observe if the tests are sensitive to increasing L (see Lehmann and Modest, 1988; Connor and Korajczyk, 1988). Success of this method depends on the quality of the chosen candidate factors. Estimators of the ranks of matrices are also used (e.g., Gill and Lewbel, 1992; Cragg and Donald, 1996, 1997). A limitation of this approach is that it is computationally burdensome, especially if the number of analyzed response variables is large (Donald et al., 2005).

We use our methodology to analyze comovement in the international stock markets. Most studies of international stock markets, given availability of data, only analyze the stock markets of developed countries. Such data inevitably include a large number of European countries and very few countries from the Americas or Africa. The empirical results from such data are most likely to exacerbate the effect of the regional factors from Europe. To avoid this problem, we use data from a similar number of countries from each continent. Our GMM method is adequate to analyze this data because only a few cross-sectional observations are available from the regions like Africa, South America, or the Middle East. Our empirical results suggest that one strong global factor exists in international stock markets. Interestingly, this factor is correlated with the Fama and French (1993) factors from the U.S. stock market, suggesting a high degree of international market integration. We also find evidence of second and third weak factors, mostly related to the markets in Europe and the Americas, respectively.

[2] A step by step explanation for the use of our method is provided in Appendix B. Gauss codes are also available for download at http://www.public.asu.edu/~mfperez.

[3] Our simulations show that use of our approach leads to accurate estimates of the number of factors regardless of whether idiosyncratic components of the model are normal, leptokurtic, or mesokurtic.

[4] Even in the case of the U.S. stock market, where extensive time series and cross-sectional observations are available, researchers have followed the approaches of Black et al. (1972) and Fama and MacBeth (1973), among many others, to group stocks into portfolios and test asset pricing models using portfolios as base assets.

[5] Specifically, ALS consider the models where the observable explanatory variables are correlated with the loadings of unobservable common factors. ALS method is unable to detect factors whose loadings are uncorrelated with observable variables. For example in an asset pricing model, ALS method would not be able to estimate *all* the risk factors explaining returns. ALS method would capture just factors with loadings that are correlated with observable variables.