ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Ex-post risk premia estimation and asset pricing tests using large cross sections: The regression-calibration approach[☆]

Soohun Kim[a,*], Georgios Skoulakis[b]

[a] Finance Area, Scheller College of Business, Georgia Institute of Technology, Atlanta, GA 30308, United States

[b] Finance Department, Sauder School of Business, University of British Columbia, Vancouver, BC, Canada V6T 1Z2

ARTICLE INFO

Article history:
Received 6 February 2016
Received in revised form 27 September 2017
Accepted 31 January 2018
Available online xxxx

Keywords:
Large cross section
N-consistent ex-post risk premia estimator
Asset pricing tests

ABSTRACT

We propose a modification of the two-pass cross-sectional regression approach for estimating ex-post risk premia in linear asset pricing models, suitable for the case of large cross sections and short time series. Employing the regression-calibration method, we provide a beta correction method, which deals with the error-in-variables problem, based on which we construct an N-consistent estimator of ex-post risk premia and develop associated novel asset pricing tests. Empirically, we reject the implications of the CAPM and the Fama–French three-factor and five-factor models but also offer new evidence on the relevance of the HML factor for pricing large cross sections of individual stocks.

1. Introduction

A common theme of asset pricing models is that differences in average returns across assets should be attributable to differences in exposures to systematic risk. There is a plethora of proposed models in the literature that differ in the types of systematic risk they identify as relevant. Typically, in these models, systematic risk is captured by a small number of pervasive factors and the average return on an asset is a linear function of the factor betas. There is a long line of research, starting with Black et al. (1972) and Fama and MacBeth (1973), on the empirical evaluation of such models.

In this paper, we develop a framework for estimating and evaluating asset pricing factor models using large cross sections of individual stock return data over short time horizons. When researchers are interested in testing an asset pricing model, they have to specify the cross section of test assets. One approach, introduced by Black et al. (1972) and Fama and MacBeth (1973) and since followed by many others, is to form a small number of portfolios and use them as test assets. In fact, following the seminal work by Fama and French (1992), it has become standard practice to sort stocks according to some characteristic, such as size or book-to-market, in order to form portfolios that are subsequently used as test assets. However, a number of papers in the extant literature argue that we should be cautious with such practice. Kan (2004) shows that the explanatory power of an asset pricing model at the individual firm level can be grossly exaggerated or nullified when sorted portfolios are used as test assets, depending on the choice of the sorting variable. Grauer and Janmaat (2004) show that, under certain conditions, the pricing errors of individual stocks can disappear in portfolios. In addition, Liang (2000) argues that, when the sorting variable used in the portfolio construction is measured with error, the estimation of the asset pricing relation, using portfolios as test assets, might suffer from serious biases. In general, the method used to form the test portfolios could affect the inference results in undesirable ways. As Roll (1977) points out, in the process of forming portfolios, important mispricing in individual stocks can be averaged out within portfolios, making it harder to reject the wrong model. Lo and MacKinlay (1990) are concerned about the exact opposite error: if stocks are grouped into portfolios with respect to attributes already observed to be related to average returns, the correct model may be rejected too often when tested at the portfolio level. More recently, Lewellen et al. (2010) and Daniel and Titman (2012) document that the

☆ The paper builds on and extends a preliminary draft titled "Revisiting the Errors in Variables Problem in Studying the Cross Section of Stock Returns" by Ravi Jagannathan, Soohun Kim, and Georgios Skoulakis. An early version of the paper was a chapter in Soohun Kim's dissertation at Northwestern University. We are extremely grateful to Ravi Jagannathan for encouraging us to work on this topic and suggesting to focus on the ex-post risk premia in the context of large cross sections. We are also thankful to Lorenzo Garlappi, Eric Ghysels, Gerard Hoberg, Mark Loewenstein, Pete Kyle, Vojislav Maksimovic, Olya Rachtchinine, Andrea Vedolin (European Finance Association Conference discussant), Jinming Xue, seminar participants at Arizona State University, Baruch College, KAIST, the University of Maryland, the University of British Columbia, the University of New South Wales, the University of North Carolina at Chapel Hill, Texas A&M University, and conference participants at the European Finance Association 2014 Conference and the Society for Financial Econometrics 2015 Conference for helpful feedback. Any errors are our responsibility. Comments and suggestions are welcome at soohun.kim@scheller.gatech.edu and georgios.skoulakis@sauder.ubc.ca .

* Corresponding author.
E-mail addresses: soohun.kim@scheller.gatech.edu (S. Kim), georgios.skoulakis@sauder.ubc.ca (G. Skoulakis).

https://doi.org/10.1016/j.jeconom.2018.01.007

performance of seemingly successful asset pricing models crucially depends on the choice of test assets. The findings of these papers provide motivation for developing asset pricing tests using individual stock data, as originally suggested by Litzenberger and Ramaswamy (1979).

Our method is a variant of the two-pass cross-sectional regression (henceforth CSR) method, which, being simple and intuitively appealing, is a popular approach in the literature.[1] The two-pass CSR method is subject to the error-in-variables (EIV) problem due to the fact that estimated betas, instead of the true betas, are used in the second pass. While the two-pass CSR risk premia estimator is consistent, as the time-series sample size T tends to infinity and the cross section size N is fixed, the traditional Fama–MacBeth standard errors are not consistent and a suitable asymptotic bias correction is needed. The associated econometric theory, that deals with the aforementioned EIV problem, was originally developed by Shanken (1992) and subsequently refined by Jagannathan and Wang (1998), among others. However, when T is fixed and N increases to infinity, the EIV problem, due to beta estimation error, is more severe in the sense that the ex-post risk premia estimator itself is inconsistent. In this paper, we employ the so-called regression calibration method to provide a suitable correction to the beta estimator yielding an N-consistent ex-post risk premia estimator. We further obtain its asymptotic distribution and provide an N-consistent estimator of its asymptotic variance–covariance matrix, which we employ to construct novel statistics for testing the ex-post risk premia implications of asset pricing models with traded factors. Furthermore, in the spirit of the GRS test of Gibbons et al. (1989), we develop a statistic, valid for general factors, to test whether a large number of stocks are fairly priced simultaneously.

The main focus of the extant methodological literature on the estimation and evaluation of asset pricing models is the case in which T is large while N is small, which is relevant when portfolios, as opposed to individual stocks, are used as test assets.[2] A few recent papers are devoted to the analysis of linear asset pricing factor models when the number of test assets N is large.

Ang et al. (2010) argue that using individual stock data, as opposed to forming portfolios, results in risk premia estimators with smaller variance. Their analysis, however, is justified only when T tends to infinity in the sense that, in their setting, the estimators are T-consistent but not N-consistent. Furthermore, they do not address the issue of bias in the risk premia estimates which turns out to be significant when N is large and T is small as our analysis illustrates. Extending the classical Gibbons et al. (1989) test, Pesaran and Yamagata (2012) propose a number of tests for the zero alpha null hypothesis, while they are not concerned with the implications of the asset pricing model regarding factor risk premia. The feasible versions of their tests are justified when both N and T tend to infinity jointly at suitable rates. Their simulation as well as empirical evidence focuses on the S&P 500 universe of stocks. Fan et al. (2015) develop a novel approach, based on joint N and T asymptotics, aimed to enhance the power of tests on the zero alpha null hypothesis using large cross-sectional data sets. Gagliardini et al. (2016) generalize the two-pass cross-sectional methodology to the case of a conditional factor model incorporating firm characteristics and unbalanced panels. Their asymptotic theory, based on N and T jointly increasing to infinity at suitable rates, facilitates studying time-varying risk premia. Chordia et al. (2015), building on Shanken (1992), use bias-corrected risk premia estimates in a context with individual stocks incorporating firm characteristics. Their focus is the relative contribution of betas and characteristics in explaining cross-sectional differences in expected returns. Jegadeesh et al. (2015) use an instrumental variable approach to deal with the EIV problem in the risk premia estimation using individual stocks, where the instruments are betas estimated over separate time periods. They focus on ex-ante implications of asset pricing models and resort to the original Fama–MacBeth approach for computing standard errors and test statistics, as traditionally used in the large T case.

We contribute to the existing literature by developing a two-pass CSR approach in order to estimate ex-post risk premia and, for the first time, construct associated asset pricing tests, when the number of assets N tends to infinity while the time-series length T is fixed. Recall that the second step of the two-pass procedure is a regression of returns on estimated betas. In the context of the standard linear regression model, it is well known that OLS estimators are consistent as long as a suitable orthogonality condition between the regression shocks and the regressors is satisfied. When N is fixed and T tends to infinity, this condition is satisfied and the two-pass CSR risk premia estimator is T-consistent. The EIV problem due to beta estimation error, however, manifests itself in the computation of standard errors. In contrast, in our context with T fixed and N increasing to infinity, the orthogonality condition is *not* satisfied and, hence, the two-pass CSR estimator is *not* N-consistent, as explained in Section 6 in Shanken (1992). At the heart of our approach is the beta estimate correction that we achieve by employing the regression-calibration approach. Using the corrected betas in the second pass yields N-consistent estimators of the risk premia. We further show that the risk premia estimator asymptotically follows a normal distribution and obtain the asymptotic variance–covariance matrix. Finally, incorporating a cluster structure for idiosyncratic shock correlations, we provide an N-consistent estimator of the asymptotic variance–covariance matrix which we use to develop statistics for testing the ex-post risk premia implications of asset pricing models with traded factors. Furthermore, in the spirit of the GRS test of Gibbons et al. (1989), we develop a test statistic, suitable for our small T-large N context and valid for general factors, based on measures of mispricing at the individual stock level.

This paper is not the first attempt to provide an N-consistent risk premia estimator. Litzenberger and Ramaswamy (1979), Shanken (1992), and Jagannathan et al. (2010) provide related estimators under different sets of assumptions.[3] While these papers offer N-consistent risk premia estimators, they do not develop the associated sampling theory. Our paper, to the best of our knowledge, is the first paper in the literature to fill this gap. It turns out that the estimator developed in Jagannathan et al. (2010), adapted to our framework, is equivalent to our risk premia estimator. However, our estimation scheme is fundamentally different in that it is based on the regression-calibration approach and achieves the desired orthogonality condition in the second-pass regression by using EIV-corrected betas. As a result, our risk premia estimator has a convenient OLS form that we exploit to establish its asymptotic

[1] Alternative approaches for estimating and testing asset pricing models include the maximum likelihood method and the generalized method of moments. Such methods, however, do not seem suitable for dealing with large cross sections of individual assets over short horizons.

[2] The long list of related papers includes, among others, Gibbons (1982), Shanken (1985), Connor and Korajczyk (1988), Lehmann and Modest (1988), Gibbons et al. (1989), Harvey (1989), Lo and MacKinlay (1990), Zhou (1991), Shanken (1992), Connor and Korajczyk (1993), Zhou (1993), Zhou (1994), Berk (1995), Kim (1995), Hansen and Jagannathan (1997), Ghysels (1998), Jagannathan and Wang (1998), Kan and Zhou (1999), Jagannathan and Wang (2002), Chen and Kan (2004), Lewellen and Nagel (2006), Shanken and Zhou (2007), Kan and Robotti (2009), Hou and Kimmel (2010), Lewellen et al. (2010), Nagel and Singleton (2011), Ang and Kristensen (2012), Kan et al. (2013a) and Kan et al. (2013b).

[3] Litzenberger and Ramaswamy (1979) develop such an estimator under the assumptions that the disturbance variance–covariance matrix is diagonal and known. Theorem 5 in Shanken (1992) relaxes these assumptions and provides an N-consistent estimator of the ex-post risk premia under sufficiently weak cross-sectional dependence between the disturbances. Jagannathan et al. (2010), in subsection 3.7, provide detailed assumptions under which such an N-consistent ex-post risk premia estimator is obtained (see their Theorem 7).