Essays on Empirical Asset Pricing

Ang Li

A thesis submitted for the degree of Doctor of Philosophy of

The Australian National University

July 2024

Declaration

I hereby confirm that this thesis represents my original work. The material has not been previously submitted to this or any other institution for a reward. To the best of my knowledge, it does not contain any material previously published or written by another author except where due reference is made.

Ang Li

July 2024

Acknowledgements

I would like to express my most heartful gratitude to the Chair of my Supervisory Committee, Prof Takeshi Yamada, for his guidance, encouragement, patience and unswerving support throughout my PhD study. Without his direction, understanding and thoughtfulness, I would not have come this far and this thesis would not have been possible. It has been the privilege and honor of my life to be his student.

My most sincere gratitude also goes to my Associate Supervisor, Dr Ding Ding. Her inspiration, encouragement, mentoring, and continuous and disinterested support helped me get through the darkest hour in my life.

I would also like to express my sincere appreciation to my Associate Supervisor, Dr Xin Liu, for her help and patient mentoring.

I am hugely appreciative of our Associate Dean (Research) of CBE, Prof Timothy Higgins, for his solicitude and support all along.

I am grateful to the RSFAS for affording me scholarships and a conducive atmosphere, which allowed me to complete my research.

I also want to thank my friends and every one of the faculty members and college staff who have supported me throughout this protracted academic journey.

Furthermore, I would also like to express my sincere gratitude to Dr Zhangjie Huang, for his careful treatment.

Lastly, I am very much indebted to my mother, for her tenacity, unconditional love, and support.

Abstract

This thesis is comprised of three Chapters. Chapter 1 investigates the pricing implication of the inequality risk in the cross-section of Chinese stock returns. We find a significant negative inequality risk premium regardless of whether tests are performed on individual stocks or portfolios. Moreover, we find evidence that the inequality risk exposure is mildly positively associated with idiosyncratic volatility and turnover rate and negatively associated with the E/P ratio and ownership concentration, supporting the conjecture that poorer investors (proxied by retail investors) tend to hold assets with higher inequality betas.

Chapter 2 investigates the pricing implications of macroeconomic factors in the cross-section of Chinese stock returns. Our results suggest that the Chinese stock market now prices macroeconomic risks similarly to more developed markets and thus can no longer be deemed a "casino." Most macroeconomic factors tested have significant loadings on the stochastic discount factor and economically significant risk premiums. Furthermore, we find that the Fama-French five factors are superfluous in the presence of the macroeconomic factors. However, the performance of the two sets of conventional macroeconomic factors tested (the PCA and Chen-Roll-Ross (1986) factors) are unstable over different test assets, thus further search for macroeconomic factors that could help better describe the cross-section of Chinese stock returns is much needed.

Chapter 3 aims to improve the explanatory power of macroeconomic factors in the cross-section of Chinese stock returns. To do so, we construct a set of machine-learning-enhanced macroeconomic factors using a novel procedure binding the theory of the ICAPM with the Elastic Net algorithm. Our results show that our machine-learning-enhanced macroeconomic factors incorporate more useful pricing information than traditional ad hoc sparse macroeconomic factors, and demonstrate superior performance to the PCA and Chen-Roll-Ross (1986) factors in most situations.

Contents

Chapter 1. Inequality and the cross-section of stock returns in China

1.1. Introduction

One day, over 2500 years ago, during a meeting with two of his students who served in the court of State Lu, Confucius made the following remark:

> I have heard that rulers of states and chiefs of families are not troubled lest their people should be few, but are troubled lest they should not keep their several places; that they are not troubled with fears of poverty, but are troubled with fears of a want of contented repose among the people in their several places. For when the people keep their several places, there will be no poverty; when harmony prevails, there will be no scarcity of people; and when there is such a contented repose, there will be no rebellious upsettings. (Confucius et al., 1971, p. 308)

Since then, the idea of Egalitarianism has become deep-rooted in the Chinese culture. Throughout Chinese history, increasing inequality has been one of the critical drivers of social instability (Yang, 2006; Knight, 2013) and dynasty changes. Many uprisings explicitly advocated for the even redistribution of lands and property. Similar phenomena in Western history have been studied by the work of Hirschman and Rothschild (1973), Turchin (2005), and Hirschman et al. (2013), who attribute dynasty changes to the "tunnel effect" – a concept related to changing tolerance for inequality. At the early stages of rapid economic development, the populace tends to be more tolerant of inequality, expecting that the wide disparities in income and wealth that accompany growth will eventually narrow. However, when these expectations are not fulfilled, this tolerance tends to diminish over time. Since the start of economic reforms in the late 1970s, China has experienced remarkable economic growth alongside a stark rise in inequality (Knight, 2014). As this trend has spanned over 40 years, the "tunnel effect" is likely becoming more prominent. Inequality is thus a critical, yet often overlooked, factor affecting economic outcomes and asset prices.

A number of studies highlight the significant impact of inequality on other macroeconomic variables, including economic growth (Shin, 2012; Rubin & Segal, 2015; Aiyar & Ebeke, 2020) and financial stability (Lysandrou, 2011). Inequality has also been linked to market risk premiums (Gollier, 2001; Markiewicz & Raciborski, 2022). However, there has been limited empirical research on the cross-sectional pricing implications of inequality. Johnson (2012) is among the few studies that directly tests whether inequality is a source of priced systematic risk in the cross-section of U.S. stock returns. Since China has a long history of Egalitarianism, it would be interesting to extend Johnson (2012) and investigate whether concerns about inequality affects the cross-section of stock returns in the world's second-largest stock market.

Johnson (2012) measures inequality using several annual proxies based on fractile ratios of taxable income proposed by Piketty and Saez (2003). The baseline measure he adopts is the ratio of income received by the top 10% to the income received by the remaining population. Since his model concerns consumption inequality, he makes various adjustments to this baseline measure (for example, incorporating transfers and computing after-tax income) to reflect consumption inequality. Compared to other alternatives, this measure can be tracked since 1913, providing a longer historical time series for robust tests.

The modern Chinese stock market has a relatively short history of over 30 years. Thus, annual inequality measures such as the Gini coefficient or the income fractile ratios as in Piketty and Saez (2003) do not yield sufficient statistical power for meaningful tests in the Chinese stock market. To address this issue, we develop three monthly indices that measure year-on-year changes in consumption and income inequality. These indices are constructed by integrating information from ten basis inequality measures, primarily based on consumption patterns. For instance, one measure is the difference between the year-on-year growth rates of luxury and economy car sales in China. We expect a positive correlation between such measures and changes in inequality. Details for the other basis inequality measures are documented in Table A.1 of Appendix A.

The three inequality indices are constructed using different dimensionality reduction methods: the dynamic factor model, the entropy weight method, and principal component analysis. Unlike Johnson (2012), we focus mainly on consumption-based measures that eliminate the need for adjustments required to model consumption inequality. The monthly frequency of our data also better aligns with the short investment horizons of Chinese retail investors, who tend to hold investments for approximately 40 days on average (Shanghai Stock Exchange, based on data between 2016 to 2019). Despite these differences, our measures share a fundamental similarity with Johnson's: both use "difference-in-difference" calculations that capture changes in income ratios between the rich and the poor.

Using the three inequality indices, we then build three corresponding inequality factors by constructing economic tracking portfolios (Lamont, 2001). We test whether inequality risk is priced in the cross-section of individual stock returns following the Fama-Macbeth (1973) procedure. We further investigate how inequality risk exposure interacts with other firm characteristics. We extend firm-level asset pricing tests to four sets of portfolios sorted by size and E/P ratio, size and idiosyncratic volatility, size and turnover, and size and ownership concentration. We further inspect the relationship between portfolio inequality exposures and portfolio returns using time-series regressions. We compare the performance of different model specifications over each set of test portfolios. We also check whether inequality is a valid candidate for ICAPM state variables following Maio and Santa-Clara (2012).

Johnson (2012) posits that in a CRRA economy, poorer investors’ greater percentage income risk necessitates hedging against rising inequality, leading to a negative inequality risk premium. The Chinese market is dominated by retail (poorer) investors (Cuthbertson et al., 1997; Docherty & Hurst, 2018; Liu et al., 2019), accounting for 86% of A-share trading volume (Securities Association of China, 2020). Our results confirm that inequality is an important systematic risk priced in the cross-section of Chinese stock returns. Consistent with Johnson (2012), we find that inequality risk bears a significant negative risk premium across individual stocks as well as portfolios, and is more pronounced than in the U.S. due to much higher proportion of retail investors in China.

Incorporating the inequality factor with Fama-French factors also reduces pricing errors and enhances model performance across different sets of test portfolios. Moreover, we find evidence that higher inequality risk exposure is mildly positively associated with idiosyncratic volatility and turnover, as well as negatively associated with the E/P ratio and ownership concentration. These results support the conjecture by Johnson (2012) that poorer investors, represented by retail investors, tend to hold assets with higher inequality risk exposures. On the other hand, while Johnson (2012) discovers a positive relationship between inequality risk exposure and market capitalization in the U.S. market, we do not observe a clear association between those two variables in the Chinese market.

Our results suggest that inequality is a viable ICAPM state variable. Specifically, the inequality indices negatively predict future market returns and positively predict future market volatilities. The results from the predictive regressions are consistent with the negative inequality risk premium found in the cross-section of stock returns. These results are generally robust across different versions of the inequality factor and remain consistent regardless of how variable are winsorized. The results also hold when excluding the sample period affected by COVID-19.

This study makes three major contributions. Firstly, we introduce novel monthly Chinese inequality indices primarily based on consumption data, distinct from traditional income-based inequality measures. Compared to traditional inequality measures such as the Gini coefficient and income fractile ratios, our indices have three advantages. First, contrary to traditional measures often built from annual survey data, our indices are constructed from monthly data, allowing for enhanced statistical power. Meyer and Sullivan (2011, 2012, 2013) and Attanasio and Pistaferri (2016) note that consumption inequality is a more appropriate gauge of inequality than income inequality, as consumption is more directly related to consumer well-being. Recent studies find evidence that income inequality, wealth inequality, and consumption inequality track each other closely (Aguiar & Bils, 2015; Abbott & Brace, 2020). Further, while households with higher income and wealth may systematically underreport their income, wealth, and expenditure (Aguiar & Bils, 2015; Abbott & Brace, 2020), our inequality indices are not based upon micro survey data, which mitigates measurement and non-response errors.

Secondly, this work adds to the few studies that empirically investigate the cross-sectional implications of inequality on asset returns. To the best of our knowledge, this study is the first to examine the influence of inequality risk on developing markets. Johnson (2012) tests for inequality risk premium in the U.S. market using a two-factor CCAPM. We extend Johnson (2012) to the largest emerging market in the world and examine inequality risk within the framework of ICAPM. We also expand the scope of focus to investigate the relationship between inequality risk exposure and several firm characteristics related to ownership structure. Our work addresses the asset pricing implications of inequality, as called for by Brunnermeier et al. (2021).

Thirdly, most studies on the cross-section of Chinese stock returns focus on fundamental factors (firm characteristics), with little attention paid to macroeconomic factors. Our work adds to the few studies that examine the cross-sectional pricing implications of macroeconomic risks in the Chinese stock market.

The rest of the chapter is organized as follows: Section 1.2 briefly reviews the theoretical background and empirical works related to our study. Section 1.3 introduces the econometric methodologies involved in factor construction and model testing. Section 1.4 provides details of the sample and variables. Section 1.5 presents the main empirical results. Section 1.6 presents results for further robustness checks. Section 1.7 concludes and summarizes the main findings of the chapter.

1.2. Literature Review

Inequality risk, which is defined by Johnson (2012) as "the risk of systematic increases in dispersion between income groups," is an essential variable that may interact with different aspects of the economy. For example, Kuznets (1955), Shin (2012), and Rubin and Segal (2015) study the relationship between inequality and economic growth. A recent study by Aiyar and Ebeke (2020) further breaks down the question and shows that inequality of opportunity moderates income inequality and economic growth. They discover a more prominent negative relationship between income inequality and future economic growth in countries with higher intergenerational mobility. Gollier (2001) and Markiewicz and Raciborski (2022) argue that inequality affects the level of market risk premium. Moreover, Lysandrou (2011) suggests that inequality is the main driving force of the financial crisis. With an even broader view, Hirschman and Rothschild (1973), Turchin (2005), and Hirschman et al. (2013) examine Western history and build models attributing dynasty changes to a shift in tolerance for inequality in society. Inequality has also attracted the attention of researchers in the field of asset pricing. This study is inspired by the works of Constantinides and Duffie (1996) and Johnson (2012).

Constantinides and Duffie (1996) introduce a theoretical framework to incorporate heterogeneity in incomplete markets. Their model proposes a pricing kernel that depends on the cross-sectional variance of log consumption growth and aggregate consumption growth in CCAPMs, in which the cross-sectional variance of log consumption growth can be regarded as a measure of inequality. Cogley (2002) generalizes the work of Constantinides and Duffie (1996) and finds only weak support for the cross-sectional dispersion of log consumption growth as an additional factor. Brav et al. (2002) find that the cross-sectional skewness, rather than the variance, of consumption growth is crucial for explaining the equity premium puzzle.

Following the intuition of Constantinides and Duffie (1996), Balduzzi and Yao (2007) develop a heterogeneous-agent model that incorporate changes in the cross-sectional variance of log consumption, rather than the cross-sectional variance of log consumption growth, as an additional factor. Johnson (2012) develops a two-agent incomplete-market model that incorporate preference over both comparative and non-comparative consumption. The model implies a negative inequality risk premium when comparative preference dominates. Utilizing several inequality proxies based on income fractile ratios proposed by Piketty and Saez (2003), Johnson (2012) tests the model over the cross-section of portfolios sorted by size and book-to-market ratio in the U.S market and finds significant negative inequality risk premium.

While inspired by Constantinides and Duffie (1996) and Johnson (2012), our study differs by treating inequality risk as a state variable in the ICAPM (Merton, 1973) rather than a CCAPM factor. This aligns our research with the strand of literature that empirically tests different specifications of the ICAPM in the cross-section of stock returns. Examples include Shanken

(1990), Brennan et al. (2004), Hahn and Lee (2006), Ang et al. (2006), Petkova (2006), Campbell and Vuolteenaho (2004), Bali (2008), Ang et al. (2009), Bali, and Engle (2010), Chen and Petkova (2012), Koijen et al. (2017), Detzel (2017), Su et al. (2023), Lin (2021) and Treepongkaruna et al. (2023).

Maio and Santa-Clara (2012) test various macroeconomic factors for consistency with the ICAPM and find that most factors are inconsistent with the ICAPM's restrictions. To address concerns about abusing the ICAPM as a "fishing license" by Fama (1991), we follow Maio and Santa-Clara (2012) and test the consistency of the inequality risk factor with the ICAPM.

Fama and French (1993, 1995, 1996) suggest that the size and value factors themselves can be regarded as state variables under the framework of the ICAPM. Based on that idea, Liew and Vassalou (2000) show that HML and SMB can predict future economic growth. Brennan et al. (2001) suggest that HML and SMB are priced because they predict changes in the investment opportunity set, as characterized by the slope and intercept of the capital market line. Vassalou (2003) tests a two-factor ICAPM and finds that the Fama-French size and value factors contain mainly news related to future GDP growth. Inspired by Vassalou (2003), Hahn and Lee (2006) discovered that the Fama-French size and value factors also capture information about term and default spread. Their discovery is supported by Petkova (2006), who finds that innovations in aggregate dividend yield, term spread, default spread, and one-month Treasury bill yield significantly correlate with the Fama-French size and value factors. Moreover, those macroeconomic factors perform even better than the Fama-French three-factor model over 25 portfolios sorted by size and book-to-market. Aretz et al. (2010) also extend the work of Vassalou (2003) and find that the Fama-French size and value factors, along with the momentum factor, capture mainly information about five macroeconomic factors and can be regarded as proxies for composite macroeconomic risk factors. Thus, a strand of literature emerges that tests variants of ICAPMs that augment certain factors of interest with the Fama-French factors. Examples include Balduzzi and Robotti (2008), Liu and Zhang (2008), Ang et al. (2006, 2009), Aretz et al. (2010), Kim et al. (2011), Chen and Petkova (2012), Bali et al. (2017), Ali et al. (2018), Leite et al. (2020) and Iania et al. (2023). Drawing upon the methodologies from those works, we also test ICAPMs augmenting the inequality risk factor with the Fama-French five factors.

1.3. Empirical methodology

1.3.1. Methods related to the construction of inequality indices

Previous studies on inequality mainly adopt measures of income inequality derived from annual data. The Gini coefficient (Benabou,1996; Cowell, 2011; Johnson, 2012; Aiyar & Ebeke, 2020) and inequality measures based on comparing income percentiles (Benabou, 1996; Piketty & Saez, 2003; Johnson, 2012; Piketty et al., 2018; Saez & Zucman, 2020) are the most common metrics. However, given the relatively short history of the modern Chinese stock market of just over 30 years, conducting statistical tests with sufficient power using annual data can be challenging. To overcome this challenge, we construct three monthly proxies that measure the year-on-year changes in Chinese consumption and income inequality. These indices are constructed by combining information from 10 basis inequality measures mainly based on consumption data. Although earlier studies have differing views on the relationships between income, wealth, and consumption inequality, more recent research (Aguiar and Bils, 2015; Abbott and Brace, 2020) finds that these inequality measures track each other closely. We employ the dynamic factor model, the entropy weight method, and principal component analysis to create our consumption and income inequality measures. More detailed descriptions of the ten basis monthly inequality measures are provided in Section 1.4.

1.3.1.1. Dynamic factor model

We construct the first version of the inequality index using the dynamic factor model (DFM). Since first introduced by Chamberlain and Rothschild (1983), the dynamic factor model has been widely adopted in Macroeconomics and Finance for dimensionality reduction due to its flexibility (see, for example, Ludvigson & Ng, 2007; Ludvigson & Ng, 2009; Neely et al., 2012; Kelly & Pruitt, 2013; Neely et al., 2014; Kelly et al., 2023).

The model assumes a set of economically related variables is driven by a vector of common latent factors, which takes the form:

$$\boldsymbol{y_t} = \boldsymbol{a} + \boldsymbol{P}\boldsymbol{f}_t + \boldsymbol{Q}\boldsymbol{x}_t + \boldsymbol{\varepsilon_t} \tag{1.1}$$

$$\boldsymbol{f_t} = \boldsymbol{c} + \Sigma_{i=1}^{p}\boldsymbol{C_i}\boldsymbol{f_{t-i}} + \boldsymbol{R}\boldsymbol{w_t} + \boldsymbol{\vartheta_t} \tag{1.2}$$

$$\boldsymbol{\varepsilon_t} = \Sigma_{j=1}^{q}\boldsymbol{D_j}\boldsymbol{\varepsilon_{t-j}} + \boldsymbol{u_t} \tag{1.3}$$

where $\boldsymbol{y_t}$ is the vector of endogenous variables comprising ten basis inequality measures. $\boldsymbol{f}_t$ is the vector of latent factors. We assume a single latent inequality factor drives the ten basis inequality measures, and thus $\boldsymbol{f}_t$ reduces to a scalar. The model assumes that the latent factor $\boldsymbol{f}_t$ follows an autoregressive process of order p, thus deemed "dynamic". The model can incorporate richer dynamic structures by allowing the residual term $\boldsymbol{\varepsilon_t}$ in (1.1) to be serially correlated. Exogenous variables $\boldsymbol{x}_t$ and $\boldsymbol{w_t}$ are also allowed to interact with the endogenous variables and latent factors. Similar to the approach of Campbell (1996), Petkova (2006), Boons (2016), and Leite et al. (2020), who model the system of state variables in asset pricing models as a VAR (1) process, we let p=1. To reduce the number of parameters, we do not add additional dynamics to the system and assume $\boldsymbol{\varepsilon_t}$ is uncorrelated with the latent factors $\boldsymbol{f}_t$ and is serially uncorrelated. We exclude exogenous variables from the system for the same purpose. The system is estimated by the Kalman Filter, and a time series of estimated values for the latent inequality factor is extracted.

1.3.1.2. Entropy weight method

We construct the second version of the inequality index using the entropy weight method (EWM), as described in Zeleny (1982), Chen and Chuang (2008), Han et al. (2015), and Li et al. (2020). The method is commonly adopted in engineering but is relatively novel in finance. In our adoption, the EWM objectively weighs each of the ten standardized basis inequality measures according to their Shannon entropy (Shannon, 1948). The final inequality index is calculated as a weighted average of basis inequality measures.

The weights for the basis inequality measures can be calculated by the following procedure:

1. Standardize each basis inequality measure.

   $$p_{kt} = \frac{x_{kt}}{\Sigma_{t=1}^{T} x_{kt}}$$

   where x_{kt} is the value of the kth basis inequality measure at period t

2. Estimate the Shannon entropy of each basis inequality measure.

   $$E_k = -\frac{\Sigma_{t=1}^{T} p_{kt*} ln p_{kt}}{\ln T}$$

3. Calculate the weight of each basis inequality measure in the inequality index. The weight is a function of Shannon entropy in Step 2. A higher entropy means the measure is less informative and is given a smaller weight.

   $$w_k = \frac{1 - E_k}{\Sigma_{k=1}^{K}(1 - E_k)}$$

Our third version of the inequality index is constructed by extracting the largest principal component of the basis inequality measures using the principal component analysis.

1.3.1.3. Economic tracking portfolio of inequality indices

In asset pricing theories, it is the news (also referred to as shocks or innovations) but not the levels of state variables that demand risk premiums. There are different ways of extracting news from levels of state variables. The two most common methods are the VAR method (Campbell, 1996; Petkova, 2006; Boons, 2016; Leite et al., 2020) and the economic tracking portfolios method (Lamont, 2001; Vassalou, 2003; Ang et al.,2006; Balduzzi & Robotti, 2008; Adrian et al., 2014; Detzel, 2017). Compared with the VAR method, the economic tracking portfolios method does not impose stringent assumptions on the exogeneity order of the state variables in the system. Thus, we extract news about future changes in inequality using the method of economic tracking portfolios. Thus, we extract news about the future changes in inequality using the method of economic tracking portfolios. For each of the three inequality indices, we run the following regression:

$$Inequality^{*}_{t+h} = \boldsymbol{b}'\boldsymbol{R}_t + \boldsymbol{c}'\boldsymbol{Z}_{t-1} + \varepsilon_{k,t+h} \tag{1.4}$$

$Inequality^{*}_{t+h}$ is the value of one of the three inequality indices h months ahead. As retail investors dominate the Chinese stock market, we focus on tracking short-horizon news about the macroeconomic variables with h=1. $\boldsymbol{R}_t$ is the vector of excess returns on the basis assets used to construct the tracking portfolios. In Vassalou (2003), $\boldsymbol{R}_t$ is comprised of the excess returns of six portfolios used to construct SMB and HML factors as in Fama and French (1993, 1995, 1996) and two fixed-income portfolios. Given this study has a relatively short time horizon, we only include the six portfolios derived from the stock market, but with a modification in the construction process by replacing B/M with E/P ratio as in Liu et al. (2019). More specifically, $\boldsymbol{R}_t$ contains returns of portfolios SV, SM, SG, BV, BM, and BG, details about which will be illustrated in Section 1.4. $\boldsymbol{Z}_{t-1}$ is the set of controlling state variables whose values are available at time t-1. In line with Lamont (2001) and Vassalou (2003), we include in $\boldsymbol{Z}_{t-1}$ the yield spread between 10-year government bonds and 1-year government bonds (Term_spread1Y), the average default spread between 3-year AA corporate bonds and 3-year government bonds (Default_spread) and 1-Month interbank repo rate (Dr1M). We also include two aggregate stock market metrics which are found to possess predictive power on future market returns in recent studies such as Gu et al. (2020) and Leippold et al. (2022): monthly market turnover (Turn_A) and monthly volatility on market returns (Svar). The estimated $\boldsymbol{b}'\boldsymbol{R}_t$ is the economic tracking portfolio of news about future changes in inequality.

1.3.2. Asset pricing tests

1.3.2.1. Fama-Macbeth (1973) procedure

We estimate the asset pricing models following the Fama-Macbeth (1973) procedure. For portfolio-level analysis, we estimate the factor risk premiums in two steps. First, we estimate the following time series regression for each test portfolio over the whole sample period:

$$R_{i,t} = c_i + \delta_i Inequality_t + \boldsymbol{\beta}_i{}'\boldsymbol{F_t} + \varepsilon_{i,t} \tag{1.5}$$

where $R_{i,t}$ is the excess return of test asset i at time t, $\boldsymbol{F_t}$ is the vector of either the market factor, the Fama-French three factors, or the Fama-French five factors. $Inequality_t$ is the economic tracking portfolio that proxies for innovations of the corresponding inequality index. Let $\widehat{\boldsymbol{x}}_{\boldsymbol{i,t}}$= $[\hat{\delta}_i \;\; \widehat{\boldsymbol{\beta}}_i{}']'$ be the vector of estimated factor exposures for asset i.

Second, we estimate the following cross-sectional regression at each point in time by GLS as suggested by Lewellen et al. (2010):

$$R_{i,t+1} = a_t + \boldsymbol{\gamma_t}'\widehat{\boldsymbol{x}}_{\boldsymbol{i,t}} + e_{i,t+1} \tag{1.6}$$

The procedure produces a time series of $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}}$.

We then estimate the average risk premium and its sampling covariance matrix by:

$$\widehat{\boldsymbol{\gamma}} = \frac{1}{T}\Sigma_{t=1}^{T}\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}} \tag{1.7}$$

$$var(\widehat{\boldsymbol{\gamma}}) = \frac{1}{T^2}\Sigma_{t=1}^{T}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}} - \widehat{\boldsymbol{\gamma}})(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}} - \widehat{\boldsymbol{\gamma}})' \tag{1.8}$$

We correct the standard errors of the estimated risk premiums for errors-in-variables bias by the Shanken (1992) correction and report Shanken's corrected t-statistics.

For firm-level analysis, we first estimate firm-level inequality betas using (1.5), with $\boldsymbol{F_t}$ being the Fama-French three factors. And estimate (1.6) with $\widehat{\boldsymbol{x}}_{\boldsymbol{i,t}}= [\hat{\delta}_i \;\; \boldsymbol{Z}_{i,t}{}']'$, where $\boldsymbol{Z_{i,t}}$ is the vector of time-varying firm characteristics.

1.3.2.2. Test of model consistency with the ICAPM

Following Maio and Santa-Clara (2012) and Cooper and Maio (2019), we test the consistency of the inequality factor against restrictions of ICAPM with the following predictive regression:

$$y_{t+h} = \mathrm{c} + \alpha Inequality_t^* + \boldsymbol{\gamma}'\boldsymbol{x_t} + \varepsilon_{t+h} \tag{1.9}$$

where y_{t+h} is either the return or volatility on the market index h months ahead. We let h=1 since myopic retail investors dominate the Chinese stock market, and the predictors experience rapid decreases in predictive powers as the forecasting horizon increases. $\boldsymbol{x_t}$ is the vector of controlling state variables, which previous studies have found to be informative about future market returns or volatilities. $Inequality_t^*$ is the value of one of three inequality indices. We are interested in the significance and sign of the coefficient α. The ICAPM implies that a necessary condition for inequality to demand a risk premium is that α is statistically significant. Moreover, if a change in inequality positively forecasts future market returns or negatively forecasts future market volatilities, "shocks/news" of change in inequality should have a positive risk premium and vice versa. We correct the standard errors of the estimated coefficients for serial correlation and heteroskedasticity with Newey and West's (1987a) standard errors.

1.4. Data and variables

The sample period of this study spans February 2005 to December 2021, constrained by the availability of the ten basis inequality measures. The pre-2000 period in the history of the Chinese stock market is often viewed as a formative stage characterized by rampant accounting frauds and market manipulations. Carpenter et al. (2021) show that only after 2004 did stock prices in the Chinese A-share market start to reflect corporate fundamentals. While excluding the 1990s reduces our sample size, it mitigates potential biases related to market underdevelopment during that period.

Our initial sample includes all stocks listed in the Chinese A-share market. The data for individual stock returns and accounting information are from CSMAR. We apply several filters to the initial sample. We exclude observations within six months of a stock's IPO. We exclude stocks in the financial industry, and we further exclude stocks with special treatments (ST, *ST, and PT). Following Liu et al. (2019), we drop the smallest 30% of stocks from the sample to alleviate the reverse merger shell-value problem associated with small-cap stocks caused by the IPO and delisting regulations for the Chinese stock market. Liu et al. (2019) find that 83% of the reverse mergers are conducted over listed "shells" among the smallest 30% of firms.

We construct the modified Fama-French size and value factors following Liu et al. (2019). Each month, we independently sort the remaining 70% of stocks by size and E/P ratio. We split the remaining collection of stocks into two size groups by the median market capitalization of these stocks, small (S) and big (B). We also split that universe into three E/P groups by the 30^{th} and 70^{th} E/P percentiles. The resulting E/P groups are denoted value (V) for the top 30% group, middle (M) for the middle 40% group, and growth (G) for the bottom 30% group. Combining the results of the two sorting procedures mentioned above, we construct six value-weighted portfolios (S/V, S/M, S/G, B/V, B/M, B/G) using the intersections of those size and value groups. The modified Fama-French size and value factors are then defined as follows:

$$SMB = \frac{1}{3}(\mathrm{S/V} + \mathrm{S/M} + \mathrm{S/G}) - \frac{1}{3}(\mathrm{B/V} + \mathrm{B/M} + \mathrm{B/G})$$
$$VMG = \frac{1}{2}(\mathrm{S/V} + \mathrm{B/V}) - \frac{1}{2}(\mathrm{S/G} + \mathrm{B/G})$$

We construct the test portfolios following a similar procedure to the construction of SMB, except that we conduct dependent sort controlling for size rather than independent sort. Each month, for test portfolios related to the four variables (E/P ratio, idiosyncratic volatility, turnover rate, and ownership concentration) that we are interested in, we first sort the remaining 70% of stocks into five size quintiles and further sort the stocks within each size quintile into five quintiles by one of the four variables mentioned above. We then form 25 value-weighted size-neutral portfolios using the intersections of those groups.

We construct the inequality index with three approaches to combine information from ten basis inequality measures. The majority of these are "difference in difference" measures that represent changes in income ratios between the rich to the poor. The first basis inequality measure equals the difference between the year-on-year change in the Retail Price Index for jewelry and the Retail Price Index for daily necessities. The second basis inequality measure equals the difference between the year-on-year change in nationwide real estate prices and rents. The third basis inequality measure equals the difference between the year-on-year change in Hong Kong luxury home prices and the Consumer Price Index. The fourth basis inequality measure equals the difference between the year-on-year change in the Urban Consumer Price Index and the Rural Consumer Price Index. The fifth basis inequality measure equals the difference between the year-on-year change in the retail sales of jewelry and the retail sales of daily necessities. The sixth basis inequality measure equals the year-on-year change in the coefficient of variation of per capita industrial profits. The seventh basis inequality measure equals the difference between the year-on-year change in the revenue for Macao's gambling industry and the retail sales of daily necessities. The eighth basis inequality measure equals the year-on-year change in the ratio of the average per capita expenditure of the top 10% of provinces to the average per capita expenditure of the rest of provinces. The ninth basis inequality measure equals the difference between the year-on-year change in the price of imported jewelry and the Retail Price Index for daily necessities. The tenth basis inequality measure equals the difference between the year-on-year growth rate of the number of luxury and economy cars sold in China. We expect such measures to correlate positively with changes in inequality.

The data for Macao's gambling industry revenue are from Macao SAR's Statistics and Census Service. The data for the change in Hong Kong luxury home prices are from the University of Hong Kong Real Estate Index Series. The data for the change in the number of luxury and economy cars sold are from the Souhu vehicle database. The rest of the data for basis inequality measures are from CSMAR and RESSET.

1.5. Empirical results

1.5.1. Descriptive statistics

This subsection presents descriptive statistics of inequality indices and factors in this study. Panel A of Table 1.1 shows the correlation matrix between the three versions of inequality indices. Inequality_fa*, Inequality_pc*, and Inequality_ent* are inequality indices constructed using the dynamic factor model, principal component analysis, and entropy weight method, respectively. Rm_f and Svar_A_f are one-month-ahead market return and volatility, respectively. The three inequality indices demonstrate high pairwise correlations. The correlation between Inequality_fa* and Inequality_pc* is 0.672, and that between Inequality_fa* and Inequality_ent* is as high as 0.563. All three indices are negatively correlated with future market return, and Inequality_fa* and Inequality_pc* are positively correlated with future market volatility. Panel B presents the augmented Dickey-Fuller (ADF) test statistics for the three inequality indices. Column (1) shows test statistics for model specifications with no drift or trend, Column (2) shows test statistics for model specifications with a deterministic time trend, and Column (3) shows test statistics for model specifications with drift but without deterministic time trend. Results show that for all three inequality measures and all model specifications, the null hypothesis of the existence of a unit root is rejected under 1% significance level. Thus, we do not need to further difference the indices for subsequent analyses.

Table 1.1. Correlations among three inequality indices and the test of stationarity

This table reports the correlations among three inequality indices and the ADF stationarity test results for the inequality indices. Inequality_fa* is the inequality index constructed using the dynamic factor model. Inequality_pc* is the inequality index constructed using the principal component analysis. Inequality_ent* is the inequality index constructed using the entropy weight method. Rm_f is the 1-month ahead return on Wind's Chinese all-share index. Svar_A_f is the 1-month ahead volatility on Wind's Chinese all-share index. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

Panel A: Correlation matrix of three inequality indices					
	Inequality_fa*	Inequality_pc*	Inequality_ent*	lnRm_f	Svar_A_f
Inequality_fa*	1.000				
Inequality_pc*	0.672	1.000			
Inequality_ent*	0.563	0.597	1.000		
Rm_f	-0.178	-0.121	-0.191	1.000	
Svar_A_f	0.211	0.163	-0.004	-0.242	1.000

Panel B: ADF tests on three inequality indices			
	No drift or trend	With trend	With drift
Inequality_fa*	-5.127^{***}	-5.141^{***}	-5.127^{***}
Inequality_pc*	-5.260^{***}	-5.251^{***}	-5.260^{***}
Inequality_ent*	-4.629^{***}	-4.617^{***}	-4.629^{***}

Figure 1.1 demonstrates how the three indices evolve through time. Consistent with the results in Panel A of Table 1.1, the three indices move closely. Inequality growth rises before the Global Financial Crisis (GFC) and drops sharply during the GFC. It rebounds to new historical high levels for two of the inequality measures during the implementation of a corresponding ¥ 4 trillion stimulus plan. However, the growth rate of inequality drops again shortly afterward, possibly due to the slowdown of the economy caused by the GFC, and started to oscillate after 2011.

Figure 1.1. The evolvement of the three inequality indices
This figure plots the development of the three inequality indices through time. Inequality_fa* is the inequality indices constructed using the dynamic factor model. Inequality_pc* is the inequality indices constructed using the principal component analysis. Inequality_ent* is the inequality indices constructed using the entropy weight method. The sample period is from 2005:M2 to 2021:M12.

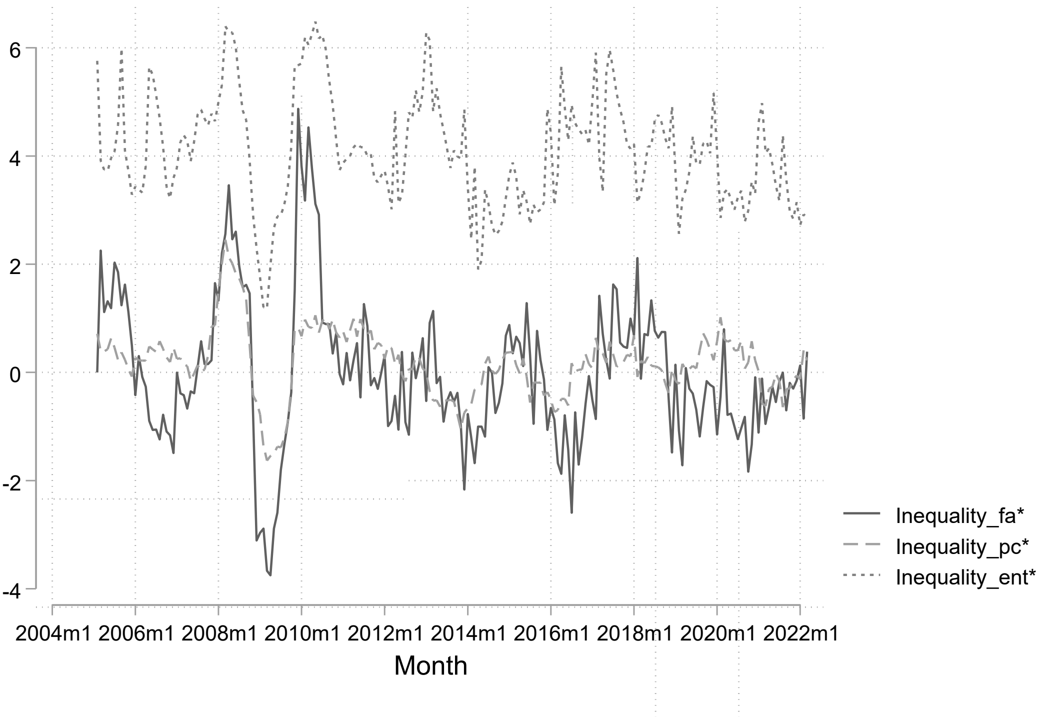


Table 1.2 presents correlations between the inequality index constructed by the dynamic factor model and other macroeconomic variables. The inequality index is most positively correlated with the National Real Estate Prosperity Index, yielding a correlation coefficient of 0.476. The result is within our expectations, as real estate constitutes the most significant proportion of Chinese household wealth. Other variables that have comparatively high correlations with the inequality index include the year-on-year change in industrial production, the year-on-year change in the Producer Price Index, and the year-on-year change in the Consumer Price Index. The results for the other two indices are qualitatively similar.

Lastly, we report the descriptive statistics and correlation matrix of factors involved in this study. Following the procedure in Section 1.3.1.3, we construct three inequality factors that proxy for innovations on three versions of inequality indices. Inequality_fa, Inequality_pc and Inequality_ent are mimicking portfolios of innovations of Inequality_fa*, Inequality_pc* and Inequality_ent* respectively. We also include the Fama-French five factors (Fama & French, 2015). Previous studies have shown that the Fama-French factors (excluding the market factor) could be regarded as innovations of certain ICAPM state variables (for example, Liew & Vassalou, 2000; Petkova, 2006; In & Kim, 2007; Hanhardt & Ansotegui, 2008). Since we aim to test the cross-sectional pricing implication of inequality under the ICAPM framework, the Fama-French factors could serve as a parsimonious set of controlling variables that proxy for a comprehensive set of latent state variables. Numerous studies also test asset pricing models that augment macroeconomic factors with the Fama-French factors. Examples include Petkova (2006), Liu and Zhang (2008), Aretz et al. (2010), Kim et al. (2011), Bali et al. (2017), Ali et al. (2018), Leite et al. (2020) and Iania et al. (2023). In Chapter 3, we will expand the universe of competing factors and investigate whether inequality retains its explanatory power.

Table 1.2. Correlations between the inequality index and other macroeconomic variables

This table reports the correlations between the inequality index constructed using the dynamic factor model and various macroeconomic variables. Ind is the year-on-year change in industrial production. PMI is the Purchasing Managers' Index. PPI_YoY is the year-on-year change in the Producer Price Index. Loan is the year-on-year change in newly created bank loans. RMB_USD is the year-on-year change in the exchange rate of RMB to USD. Realty is the National Real Estate Climate Index. Trade is the year-on-year change in total imports and exports. CPI_YoY is the year-on-year change in the Consumer Price Index. M1M2 is the year-on-year growth spread between M1 and M2. Invest is the year-on-year change in countrywide fixed asset investment. Retail_YoY is the year-on-year change in Total Retail Sales of Consumer Goods. The sample period is from 2005:M2 to 2021:M12.

	Inequality_fa*	Ind	PMI	PPI	Loan	RMB_USD	Realty	Trade	CPI_YoY	M1M2	Invest	Retail_YoY
Inequality_fa*	1.000											
Ind	0.337	1.000										
PMI	0.270	0.705	1.000									
PPI_YoY	0.313	0.294	0.242	1.000								
Loan	0.278	0.449	0.562	0.184	1.000							
RMB_USD	0.123	0.545	0.571	0.299	0.247	1.000						
Realty	0.476	0.629	0.533	0.631	0.470	0.331	1.000					
Trade	0.269	0.584	0.462	0.746	0.310	0.309	0.639	1.000				
CPI_YoY	0.284	0.331	0.132	0.408	0.024	0.054	0.435	0.331	1.000			
M1M2	0.243	0.101	0.330	0.236	0.555	0.162	0.537	0.218	0.076	1.000		
Invest	0.156	0.828	0.524	0.114	0.430	0.332	0.232	0.362	0.108	-0.055	1.000	
Retail_YoY	0.178	0.665	0.338	0.243	0.391	0.081	0.281	0.417	0.301	0.109	0.826	1.000

Panel A of Table 1.3 presents correlations between the inequality factor and Fama-French factors. The three versions of the inequality factor still preserve considerably high correlations with each other. Moreover, all three versions negatively correlate with market return and the value factor. Both Inequality_fa and Inequality_pc correlate most strongly with the value factor: the correlation between Inequality_fa and VMG is -0.243, and that between Inequality_pc and VMG is -0.257. The result is consistent with our expectations: if assets with higher inequality betas are more likely to be held by poorer (retail) investors, as postulated by Johnson (2012), we would expect these assets may be "overvalued" due to arbitrage asymmetry. We will further investigate this issue in section 1.5.4.1.

Panel B presents summary statistics for the inequality and the Fama-French five factors. All three versions of the inequality factor have negative in-sample means. Note that the means of Inequality_fa and Inequality_pc significantly differ from zero with t-statistics equal to -2.20 and -1.99, respectively. For the Fama-French factors, only the size and value factors have premiums significantly different from zero.

Table 1.3. Descriptive statistics and correlation matrix of factors

This table presents summary statistics of the factors involved in this study and the correlations among the factors. Inequality_fa, Inequality_pc and Inequality_ent are mimicking portfolios of news about Inequality_fa*, Inequality_pc* and Inequality_ent* respectively. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The sample period is from 2005:M2 to 2021:M12.

Panel A: Correlations between inequality and Fama-French factors			
	Inequality_fa	Inequality_pc	Inequality_ent
Inequality_fa	1.000		
Inequality_pc	0.483	1.000	
Inequality_ent	0.393	0.361	1.000
MKT	-0.159	-0.166	-0.207
SMB	-0.141	0.081	-0.106
VMG	-0.243	-0.257	-0.198
RMW	-0.163	-0.128	-0.262
CMA	0.125	0.058	-0.076

Table 1.3. Descriptive statistics and correlation matrix of factors *(continued)*

Panel B: Descriptive statistics of inequality and Fama-French factors					
	mean	sd	t-stat	min	max
MKT	0.693	8.324	1.183	-26.839	29.342
SMB	0.716	3.531	2.882	-13.215	17.781
VMG	0.519	3.406	2.166	-12.985	12.306
RMW	0.063	3.333	0.269	-15.273	13.570
CMA	-0.065	2.563	-0.360	-10.576	8.615
Inequality_fa	-0.628	4.057	-2.200	-17.199	15.680
Inequality_pc	-0.589	4.205	-1.991	-20.994	19.021
Inequality_ent	-0.398	3.449	-1.640	-15.848	14.813

1.5.2. Predictive regressions of the inequality index

We check whether inequality is a valid candidate for ICAPM state variables following Maio and Santa-Clara (2012). More specifically, we run 1-month ahead predictive regressions of monthly market return and volatility on inequality indices along with other control variables that are found to possess predictive power for future market return and volatility in prior studies. The set of control variables includes the price-to-book ratio (P_b), the dividend yield (Divi), the market turnover (Turn_A), the net equity expansion (Ntis), and lagged market return and volatility. The ICAPM implies that innovations of state variables that positively forecast future market returns or negatively forecast future market volatiles demand positive risk premiums and vice versa. We will compare the results in this subsection with those of the Fama-Macbeth (1973) procedures in Section 1.5.3 and assess whether inequality is a valid candidate for ICAPM state variables.

Table 1.4 reports the estimation results using the inequality index constructed using the dynamic factor model. Panel A presents results for predictive regressions of future market return. Note that the inequality index significantly negatively forecasts 1-month-ahead market return in all univariate, bivariate, and multivariate model specifications. The above result is consistent with Gomez (2016) and Toda and Walsh (2020), who find that income and wealth inequality negatively predict future market risk premiums in the U.S. market. Gomez (2016) postulates that an increase in inequality would boost the overall market risk tolerance and drive up the demand for risky assets and asset prices, resulting in a lower risk premium. Other variables that significantly forecast the market return are market turnover rate and volatility. The dividend yield is significant in the multiple regression at the 10% significance level.

Panel B presents results for regressions of future market volatility. The coefficients on the inequality index are insignificant in columns (1), (4), (5), and (6) due to the sizeable residual sum of squares. On the contrary, the inequality index significantly positively forecasts future market volatility in model specifications (2), (3), and (7) where adjusted R-squares are higher. The above results suggest that positive exposure to inequality provides a hedge against deteriorating future investment opportunities, and thus, inequality risk demands a negative risk premium.

We do not show results for regressions with longer forecasting horizons here because, contrary to previous studies on the U.S. market, such as Maio and Santa-Clara (2012), the predictors in the Chinese market show decreasing predictive power over longer forecasting horizons. This phenomenon is consistent with findings by Cuthbertson et al. (1997), Docherty and Hurst (2018), and Liu et al. (2019), who note that myopic retail investors dominate the Chinese stock market. We also repeat the analysis in Table 1.4 with Inequality_pc and Inequality_ent; the results are qualitatively similar.

Table 1.4. Predictive regressions with the inequality index

This table presents the results of 1-month ahead predictive regressions for the monthly return and volatility of the value-weighted market index. P_b is the market price-to-book ratio. Divi is the market-wide dividend yield. Turn_A is the market turnover, and Ntis is the net equity expansion. The t-statistics corrected for serial correlation and heteroskedasticity using the Newey and West (1987a) estimator are in parentheses. The sample period is from 2005:M3 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

Panel A: Predictive regressions of market return							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	F.Rm	F.Rm	F.Rm	F.Rm	F.Rm	F.Rm	F.Rm
Inequality_fa*	-0.569**	-0.554**	-0.513**	-0.596**	-0.684**	-0.646**	-0.483**
	(-2.19)	(-2.07)	(-1.98)	(-2.05)	(-2.35)	(-2.22)	(-2.31)
P_b		-0.149					0.935
		(-0.32)					(1.58)
Divi			0.782				2.523*
			(1.04)				(1.77)
Turn_A				1.381**			2.492***
				(2.07)			(3.00)
Svar_A					-0.016		-0.029**
					(-1.42)		(-2.37)
Ntis						-0.634	-0.057
						(-1.33)	(-0.83)
cons	0.306	0.605	-1.599	-1.546	1.398*	2.674	-1.055
	(0.56)	(0.49)	(-0.80)	(-1.29)	(1.77)	(1.59)	(-1.58)
adj. R^2	0.023	0.018	0.019	0.041	0.029	0.073	0.084

Table 1.4. Predictive regressions with inequality index *(continued)*

Panel B: Predictive regressions of market volatility							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	F.Svar_A	F.Svar_A	F.Svar_A	F.Svar_A	F.Svar_A	F.Svar_A	F.Svar_A
Inequality_fa*	2.714	4.087^{**}	3.357^{**}	1.248	1.539	1.279	3.714^{**}
	(1.44)	(2.30)	(1.99)	(1.37)	(1.53)	(1.31)	(2.33)
P_b		27.135^{***}					29.747^{***}
		(8.56)					(4.42)
Divi			-36.431^{***}				-0.458
			(-7.60)				(-0.09)
Turn_A				1.285^{***}			1.387^{***}
				(3.21)			(3.43)
Rm					-0.260		-1.245^{***}
					(-0.47)		(-3.02)
Ntis						-4.591	-1.469
						(-1.46)	(-0.54)
_cons	48.622^{***}	-13.218^{*}	141.605^{***}	23.558^{**}	49.305^{***}	64.033^{***}	-36.019
	(11.03)	(-1.72)	(11.49)	(2.29)	(10.84)	(6.17)	(-0.94)
adj. R^2	0.007	0.290	0.247	0.032	0.009	0.003	0.327

1.5.3. Estimation of the inequality risk premium at firm-level

In this section, we investigate whether inequality is priced in the cross-section of stock returns in China. Boons (2016) and Ang et al. (2020) argue that asset pricing tests are more efficient on individual stocks than portfolios as the efficiency gain from a larger dispersion of estimated beta outweighs the efficiency loss induced by greater measurement errors in estimated betas. Ahn et al. (2009) and Lewellen et al. (2010) also suggest that the results of portfolio-level tests are sensitive to the choice of test portfolios. Thus, in this section, we first focus on tests at the firm level. In subsequent sections, we will also present results for portfolio-level tests. We present estimates of the inequality risk premiums following the Fama-Macbeth (1973) two-step procedure illustrated in Section 1.3.2.1. We calculate the t-statistics of the estimated risk premiums using Newey and West (1987a) standard errors corrected for serial correlation and heteroskedasticity. Before conducting cross-sectional regressions, we estimate each stock's inequality beta by time series regressions. The independent variables for the time series regressions include the inequality tracking portfolio and the Fama-French three factors to partial out the size and value effect. Due to the relatively short time series, we estimate the inequality betas over the entire sample. Following the literature, we also include a set of control variables, including LnSize, LnBM, IVOL, Turn, Illiquidity, Ret_1, Ret2_12, Beta, and Leverage. LnSize is the logarithm of the firm's market capitalization at the end of last month. LnBM is the lagged logarithm of the book-to-market ratio. IVOL is the realized idiosyncratic volatility for each firm in the past month following Ang et al. (2006). A large number of studies, such as Ang et al. (2009), Chen and Petkova (2012), and Gu et al. (2018), confirm the existence of negative risk premiums associated with idiosyncratic volatilities in different markets around the globe. Turn is the average daily turnover rate in the past month. Harris and Raviv (1993), Chen et al. (2001), Hong and Stein (2003), and Barinov (2014) argue that the turnover rate is a proxy for investors' heterogeneous beliefs. On the other hand, Baker and Stein (2004), Liu et al. (2019), and Allen et al. (2024) treat the turnover rate as a proxy for investor sentiment. Another strand of literature, including Rouwenhorst (1999), Chordia et al. (2001), Eckbo and Norli (2005), and Avramov and Chordia (2006), treats turnover as a measure of liquidity risk exposure. Thus, we include turnover in our control variables and expect it to bear a negative risk premium. Since turnover conveys information other than liquidity, we also include an additional illiquidity measure following Amihud (2002) and Amihud and Noh (2021). Ret_1 equals the lagged monthly return and captures the short-term reversal effect (Jegadeesh, 1990; Lehman, 1990). At month t, Ret2_12 equals the holding period return from month t-12 to month t-2 and captures the momentum effect (Jegadeesh & Titman, 1993). Beta is the market beta estimated using daily returns during the past 250 trading days. Finally, Leverage is the lagged 1-month debt-to-equity ratio.

Table 1.5 presents the estimation results for the Fama-Macbeth regression at firm level using Inequality_fa as the proxy for inequality risk. Note that the inequality risk has negative risk premiums in Columns (1), (2), (3) and (5). In Column (5), with the complete set of control variables, the inequality risk bears a premium of -0.357% per month, which is economically significant. In Columns (2), (3), and (5), the inequality risk premiums are all statistically significant at 1% significance level. The inequality risk premium is insignificant at 5% significance level in Column (1) due to a sizable residual error. The above results coincide with those of Johnson (2012), who also documents a significant negative inequality risk premium in the U.S. market. In Column (5), we also find significant negative risk premiums for idiosyncratic volatility, log of firm size and turnover rate, and significant positive risk premiums for illiquidity. All those premiums have the right signs as we expected. The average adjusted R^2 for the full model equals 0.114, which is satisfactory at the firm level.

Table 1.5. Firm-level estimates of the inequality risk premium

This table presents the estimated risk premiums by the Fama-Macbeth (1973) procedure using Inequality_fa as a proxy for inequality risk. Inequality is the firm-level inequality risk beta estimated by augmenting the inequality index tracking portfolio with the Fama-French three factors. LnSize is the log of the firm's market capitalization at the end of last month. LnBM is the lagged 1-month log of book-to-market ratio. IVOL is idiosyncratic volatility over the past month. Turn is the average daily turnover rate during the past month. Illiquidity is Amihud's (2002) illiquidity measure. Ret_1 is a measure of the short-term reversal effect. Ret2_12 measures the momentum effect. Beta is the market beta estimated using daily returns over the past year. Leverage is the lagged 1-month debt-to-equity ratio. The Newey and West (1987a) t-statistics corrected for serial correlation and heteroskedasticity are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret_rf	(2) Ret_rf	(3) Ret_rf	(4) Ret_rf	(5) Ret_rf
Inequality	-0.254*	-0.363***	-0.312***		-0.357***
	(-1.92)	(-2.98)	(-2.91)		(-3.26)
IVOL				-0.320***	-0.238***
				(-4.30)	(-3.34)
LnSize		-0.342**	-0.339**	-0.298**	-0.251**
		(-2.37)	(-2.38)	(-2.41)	(-2.03)
LnBM		0.522***	0.215**	0.143	0.062
		(5.20)	(2.07)	(1.32)	(0.68)
Turn			-2.055***	-1.846***	-1.639***
			(-8.49)	(-6.95)	(-5.19)
Illiquidity			5.174**	3.515**	3.335***
			(2.26)	(2.30)	(2.74)
Ret_1			-0.018**	-0.008	-0.011
			(-2.34)	(-1.00)	(-1.52)
Ret2_12			0.002	0.004	0.002
			(0.77)	(1.25)	(0.83)
Beta	-0.002	-0.002	0.001	-0.001	0.002
	(-1.09)	(-0.85)	(1.03)	(-1.05)	(1.06)
Leverage			0.051	-0.123	0.080
			(0.28)	(-0.57)	(0.43)
cons	2.383**	4.270***	4.756***	4.890***	4.668***
	(2.51)	(3.21)	(3.48)	(3.55)	(3.42)
R^2	0.032	0.079	0.118	0.112	0.121
adj. R^2	0.031	0.076	0.112	0.106	0.114

Johnson (2012) postulates that poorer investors tend to hold stocks with higher inequality betas in a CRRA economy with incomplete markets. If we further assume poorer investors are typically retail investors, we expect firms with lower ownership concentration to have higher inequality betas on average. Fidrmuc et al. (2008), Byun et al. (2011), and Fan et al. (2023)

show that high ownership concentration could exacerbate information asymmetry among investors, while Ma et al. (2010) argue that high ownership concentration could enhance firm performance. Therefore, we control for ownership concentration by including Shrs, which equals the sum of the percentages held by the second largest shareholder to the tenth largest shareholder. We exclude the largest shareholder's percentage to mitigate potential confounding effects from state ownership, as many listed companies are state-owned enterprises (SOEs). A dummy variable indicating SOE status exhibits a low negative correlation (-0.106) with Shrs, alleviating concerns about government ownership bias.

Table 1.6 repeats the analysis of Table 1.5, adding Shrs as an extra control variable. The results are qualitatively similar to those in Table 1.5. We still find significant negative inequality risk premiums in Columns (2), (3) and (5). In all model specifications, Shrs is highly significant, with a positive risk premium, suggesting that investors perceive firms with higher ownership concentration as riskier. In Column (5), the estimation results for other firm characteristics are qualitatively unchanged compared with those in Table 1.5. Idiosyncratic volatility, log of firm size, turnover rate, and Amihud (2002) illiquidity still bear significant risk premiums with the right signs. In each column, the adjusted R^2 is higher than that in the corresponding columns of Table 1.5, indicating that the inclusion of Shrs enhances the explanatory power of the models. We also repeat the analysis of Tables 1.5 and 1.6 using inequality_pc and inequality_ent as proxies for inequality risk; the results are qualitatively similar, with significant negative inequality risk premiums in most of the model specifications. Those results are presented in Appendix B.

Table 1.6. Firm-level estimates of the inequality risk premium adding Shrs
This table presents the estimated risk premiums by the Fama-Macbeth (1973) procedure using Inequality_fa with Shrs as an additional control variable. Inequality is the firm-level inequality risk beta estimated by augmenting the inequality index tracking portfolio with the Fama-French three factors. Shrs is the lagged sum of the shareholding percentage of the second largest shareholder to the tenth largest shareholder. LnSize is the log of the firm's market capitalization at the end of last month. LnBM is the lagged 1-month log of book-to-market ratio. IVOL is idiosyncratic volatility over the past month. Turn is the average daily turnover rate during the past month. Illiquidity is Amihud's (2002) illiquidity measure. Ret_1 is a measure of the short-term reversal effect. Ret2_12 measures the momentum effect. Beta is the market beta estimated using daily returns over the past year. Leverage is the lagged 1-month debt-to-equity ratio. The Newey and West (1987a) t-statistics corrected for serial correlation and heteroskedasticity are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret_rf	(2) Ret_rf	(3) Ret_rf	(4) Ret_rf	(5) Ret_rf
Inequality	-0.268*	-0.375***	-0.323***		-0.363***
	(-1.88)	(-2.86)	(-2.62)		(-2.96)
Shrs	0.007**	0.010***	0.007***	0.009***	0.008***
	(2.22)	(4.25)	(3.03)	(3.59)	(3.25)
IVOL				-0.283***	-0.201***
				(-3.60)	(-3.17)
lnSize		-0.351**	-0.364**	-0.375***	-0.328**
		(-2.13)	(-2.46)	(-2.61)	(-2.26)
lnBM		0.531***	0.185*	0.063	0.142
		(5.38)	(1.87)	(0.51)	(1.45)
Turn			-1.902***	-1.696***	-1.633***
			(-8.06)	(-7.46)	(-6.34)
Illiquidity			5.069*	3.104*	2.535**
			(1.94)	(1.75)	(2.33)
Ret_1			-0.020**	-0.009	-0.013*
			(-2.44)	(-1.11)	(-1.67)
Ret2_12			0.002	0.002	0.002
			(0.63)	(0.63)	(0.67)
Beta	-0.002	-0.001	0.003	-0.001	0.004
	(-0.84)	(-0.52)	(1.13)	(-1.05)	(1.17)
Leverage			0.079	-0.028	0.109
			(0.39)	(-0.12)	(0.58)
cons	1.769*	3.955***	4.616***	4.692***	4.504***
	(1.92)	(2.97)	(3.38)	(3.43)	(3.30)
R^2	0.039	0.086	0.122	0.114	0.126
adj. R^2	0.037	0.082	0.115	0.107	0.118

Table 1.7 presents the time series averages of cross-sectional correlations between the inequality beta estimated using Inequality_fa and firm characteristics included in the Fama-Macbeth regressions. At the firm level, the inequality beta is mildly negatively correlated with the log of book-to-market ratio and ownership concentration, and positively correlated with idiosyncratic volatility and turnover rate. Johnson (2012) postulates that poorer investors tend to hold stocks with higher inequality beta in a CRRA economy with incomplete markets. Since poorer investors are often retail investors, we expect firms with higher inequality betas to have higher turnover rates, lower ownership concentration, and tend to be overpriced with lower B/M or E/P ratios and higher idiosyncratic volatilities. Table 1.7 supports our expectations. While the inequality beta is more closely correlated with the B/M ratio, ownership concentration, idiosyncratic volatility, and turnover rate than the other characteristics, these pairwise correlations are not sufficiently strong to imply robust linear relationship. We'll explore potential nonlinear relationships between the inequality beta and these four firm characteristics using portfolio analysis in the next section. Furthermore, among the firm characteristics, idiosyncratic volatility is significantly negatively correlated with the book-to-market ratio and positively correlated with the turnover rate.

Table 1.7. Correlations between the inequality beta and firm characteristics

This table reports the time series average of cross-sectional correlations among inequality beta and the firm characteristics. Inequality β is the firm-level inequality risk beta estimated by augmenting Inequality_fa with the Fama-French 3 factors. Shrs is the lagged sum of the shareholding percentage of the second largest shareholder to the tenth largest shareholder. LnSize is the log of the firm's market capitalization at the end of last month. LnBM is the lagged 1-month log of book-to-market ratio. IVOL is idiosyncratic volatility over the past month. Turn is the average daily turnover rate during the past month. Illiquidity is Amihud's (2002) illiquidity measure. Ret_1 is a measure of the short-term reversal effect. Ret2_12 measures the momentum effect. Beta is the market beta estimated using daily returns over the past year. Leverage is the lagged 1-month debt-to-equity ratio. The sample period is from 2005:M2 to 2021:M12.

	Inequality β	Shrs	IVol	LnSize	LnBM	Turn	Ret_1	Ret2_12	Illiquidity	Beta	Leverage
Inequality β	1.000										
Shrs	-0.104	1.000									
IVol	0.132	0.047	1.000								
LnSize	0.055	-0.003	-0.044	1.000							
LnBM	-0.151	0.056	-0.295	0.004	1.000						
Turn	0.126	-0.075	0.557	-0.222	-0.189	1.000					
Ret_1	-0.022	0.013	0.346	0.074	-0.121	0.154	1.000				
Ret2_12	-0.066	0.047	0.156	0.229	-0.346	0.096	-0.002	1.000			
Illiquidity	0.022	0.086	-0.060	-0.587	0.001	-0.210	-0.027	-0.177	1.000		
Beta	0.041	-0.088	0.093	-0.133	0.005	0.130	-0.009	-0.032	0.022	1.000	
Leverage	0.075	-0.036	0.025	0.029	0.069	-0.014	-0.012	-0.029	0.004	0.088	1.000

1.5.4. Inequality beta and firm characteristics

Johnson (2012) postulates that poorer investors tend to hold stocks with higher inequality beta in a CRRA economy with incomplete markets. Assuming that retail investors are often poorer investors, we expect firms with higher inequality beta to have higher turnover rates, and lower ownership concentration and tend to be overpriced with lower B/M or E/P ratios and higher idiosyncratic volatilities. Table 1.7 helps us glimpse the linear relationships between the inequality beta and those firm characteristics. In this section, we further investigate potential nonlinear relationships at the portfolio level. As Cochrane (2011) and Bali et al. (2016) point out, portfolio analysis can help uncover nonlinear relationships between variables that are difficult to detect using parametric models.

1.5.4.1. Inequality beta and the E/P ratio

In Table 1.7, the inequality beta is most negatively correlated with the log of B/M ratio with a correlation coefficient of -0.151. This subsection further probes the relationship between the inequality beta and the relative valuation of firms. Liu et al. (2019) argue that the E/P ratio is a better measure of relative valuation than the B/M ratio in the Chinese market. In Section 1.5.3, we used B/M instead of E/P to control for the value effect because logarithms cannot be negative. In this subsection, we investigate the relationship between the inequality beta and the E/P ratio. Following Chen and Petkova (2012), we first check whether inequality risk is priced over 25 size-neutral portfolios sorted by the E/P ratio using Fama-Macbeth (1973) regressions, which augment the inequality tracking portfolio with the Fama-French five factors. If the inequality beta is associated with the E/P ratio, it should help explain the return dispersion of portfolios sorted by the E/P ratio. Using time series regressions, we then inspect how inequality risk exposure interacts with the E/P ratio. Lastly, we compare various model specifications and check whether including inequality risk exposure helps reduce pricing errors on the 25 size-neutral portfolios sorted by E/P. We present results for analyses using the tracking portfolio on the inequality index estimated using the dynamic factor model.

Table 1.8 presents the summary statistics for excess returns of the 25 double-sorted portfolios. In general, the average excess returns increase monotonically with the E/P ratio within each size quintile. Moreover, the return differences between high E/P and low E/P groups are statistically significant at the 5% level across all size quintiles except for the smallest quintile.

Table 1.8. Average excess returns of the 25 Size-E/P portfolios
This table presents summary statistics for excess returns of the 25 double-sorted portfolios on firm size and the E/P ratio. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	Size				
E/P	Small	2	3	4	Big
Low	1.346	1.030	0.642	0.762	0.268
2	1.324	1.185	0.937	0.883	0.617
3	1.484	1.294	1.170	0.898	0.578
4	1.422	1.309	1.197	1.151	0.620
High	1.423	1.539	1.358	1.207	0.891
High - Low	0.077	0.509**	0.716***	0.445**	0.622**

Table 1.9 presents results for the Fama-Macbeth regressions with factor loadings estimated over the entire sample. Columns (2) and (4) present results for the Fama-French three-factor and five-factor models, respectively. Columns (3) and (5) present results for models that augment Inequality_fa with the Fama-French three and five factors, respectively. The inequality risk premiums in Columns (3) and (5) are negative and statistically significant at the 5% level. In Column (3), the estimated inequality risk premium equals -0.48% per month, close to the firm-level estimation in Table 1.6. The inequality risk premium in Column (1) is only significant at the 10% level due to high residual error. Note that the adjusted R^2 in Column (1) is much smaller than those in the other columns. Also note that Columns (3) and (5) have higher adjusted R^2s than Columns (2) and (4), respectively, indicating that inequality beta could provide incremental pricing information beyond the Fama-French five factors. We also repeat the analyses in Table 1.9 using Inequality_pc and Inequality_ent as proxies for

inequality risk. The results are presented in Appendix B. In both cases, we obtain robust negative inequality risk premiums.

Table 1.9. Estimates of the inequality risk premium on 25 Size-E/P portfolios
This table presents the estimated inequality risk premiums by augmenting the inequality tracking portfolio with the Fama-French factors using the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. Inequality is the beta of Inequality_fa. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret
Inequality	-0.577*		-0.480**		-0.405**
	(-1.87)		(-2.20)		(-2.39)
MKT	0.972	1.122	0.941	1.012*	0.719
	(1.36)	(1.62)	(1.54)	(1.73)	(1.48)
SMB		1.137***	0.686**	0.752***	0.591**
		(2.76)	(2.03)	(2.53)	(2.11)
VMG		0.804***	0.443**	0.623***	0.488**
		(2.82)	(2.15)	(2.79)	(2.16)
RMW				0.271	0.385
				(1.44)	(1.51)
CMA				0.678**	0.312*
				(2.01)	(1.67)
cons	1.992*	2.432*	1.973*	2.525**	1.277
	(1.79)	(1.76)	(1.67)	(2.23)	(0.86)
R^2	0.356	0.621	0.660	0.709	0.751
adj. R^2	0.298	0.554	0.592	0.631	0.667

Table 1.10 presents the inequality betas of the 25 Size-E/P portfolios. Following Chen and Petkova (2012), we estimate the inequality betas by augmenting the inequality risk factor with the Fama-French three factors to partial out the size and value effects. We observe an approximately monotonic relationship between inequality beta and the E/P ratio within the third, fourth, and fifth size quintiles, where inequality beta tends to decrease with the E/P ratio. However, there is no clear monotonic relationship between the inequality beta and the E/P ratio in the smaller size quintiles.

Table 1.10. Inequality betas of 25 Size-E/P portfolios

This table presents the inequality betas of 25 Size-E/P portfolios. The betas are estimated by time series regressions augmenting Inequality_fa with the modified version of Fama-French three factors as in Liu et al. (2019). The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	E/P									
Size	Low	2	3	4	High	Low	2	3	4	High
	Inequality beta (3 Factors+)					t (3 Factors+)				
Small	-0.081	-0.022	-0.019	-0.038	-0.093*	-1.48	-0.39	-0.31	-0.78	-1.88
2	0.009	-0.062	0.017	0.019	-0.122**	0.14	-1.37	0.32	0.46	-2.32
3	0.070	-0.021	-0.099*	-0.024	-0.084*	1.56	-0.56	-1.91	-0.42	-1.76
4	0.109**	0.033	-0.112**	-0.101*	-0.145***	2.19	0.44	-2.39	-1.92	-2.77
Big	0.068	0.121**	-0.075	-0.146***	-0.203***	1.55	2.27	-1.51	-2.61	-3.42

Table 1.11. Model performance comparison on 25 Size-E/P portfolios

This table compares three pricing error metrics measurements from competing asset pricing models. Each row shows the pricing error measurements of different models on 25 Size-E/P portfolios. CAPM+ stands for the model augmenting inequality risk with CAPM. 3 Factors+ stands for the model augmenting inequality risk with the Fama-French three factors. Furthermore, 5 Factors+ represents the model augmenting inequality risk with the Fama-French five factors. The sample period is from 2005:M2 to 2021:M12.

	CAPM	CAPM+	3 Factors	3 Factors+	5 Factors	5 Factors+
GRS statistics	51.607	45.804	45.790	45.788	45.784	44.744
GRS p-value	0.001	0.007	0.006	0.006	0.007	0.008
$A\|\alpha_i\|$	0.183	0.156	0.139	0.125	0.121	0.116
$A\|\alpha_i\|/ A\|\bar{r}_i\|$	0.639	0.547	0.487	0.439	0.423	0.417

It is also noteworthy that the inequality betas are mostly positive in the low E/P groups and become negative in the high E/P groups within each size quintile. This suggests that stocks with the lowest relative valuations tend to outperform the Fama-French three-factor model when inequality growth is unexpectedly high, whereas stocks with the highest relative valuations underperform. Thus, stocks with the lowest relative valuations may offer poorer investors a hedge against inequality risk, leading to lower expected returns than those predicted by the Fama-French three-factor model in a market dominated by such investors.

Following Fama and French (2015), we compare the performance of different model specifications in pricing errors over the 25 Size-E/P portfolios. The results are shown in Table 1.11. GRS statistics is the Gibbons, Ross, and Shanken (1989) test statistics for time series regressions that measures the overall magnitude of pricing errors for the model. For all model specifications, the p-values of the GRS statistics are smaller than 0.05, indicating non-negligible pricing errors. $A|\alpha_i|$ is the average absolute value of pricing errors. $A|\alpha_i|/A|\bar{r}_i|$ is the average of the absolute value of pricing errors scaled by the average of the absolute value of $\bar{r}_i$, where $\bar{r}_i$ equals the deviation of the time series average of excess returns of portfolio i from the cross-sectional mean of average excess returns. Note that models that augment inequality risk with only subsets of the five Fama-French factors generally perform better than the corresponding baseline models. For example, $A|\alpha_i|/A|\bar{r}_i|$ for the Fama-French three-factor model equals 0.487, indicating that 48.7% of the dispersion of average excess returns is left unexplained. The value drops to 0.439 for the model that augment inequality risk with the Fama-French three factors. The phenomenon indicates that inequality risk exposure may contain incremental information that could help price the 25 Size-E/P portfolios.

1.5.4.2. Inequality beta and idiosyncratic volatility

In Table 1.7, the inequality beta is most positively correlated with idiosyncratic volatility with a correlation coefficient of 0.132. Kang et al. (2014) suggest that idiosyncratic volatility is related to the behavior of retail investors. Thus, in this subsection, we further investigate a potential nonlinear relationship between the inequality beta and idiosyncratic volatility by repeating the analyses in Section 1.5.4.2 on 25 portfolios, first sorted by size and then by idiosyncratic volatility. Examining the association between the inequality beta and idiosyncratic volatility can offer deeper insights into the relationship between the inequality beta and retail investor ownership. Table 1.12 presents summary statistics for excess returns of the 25 portfolios. Generally, average excess returns decrease monotonically with idiosyncratic volatility within each size quintile. The return differences between high IVOL and low IVOL groups are all statistically significant at the 1% level within each size quintile, except for the biggest quintile. These differences (in absolute value) also decrease monotonically with size.

Table 1.12. Average excess returns of the 25 Size-IVOL portfolios
This table presents summary statistics for excess returns of the 25 double-sorted portfolios on firm size and idiosyncratic volatility. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	Size				
IVOL	Small	2	3	4	Big
Low	2.525	2.071	1.602	1.169	0.824
2	2.134	1.721	1.384	1.223	0.989
3	1.945	1.269	1.174	1.057	0.837
4	1.313	0.811	0.623	0.696	0.719
High	0.210	-0.109	-0.269	-0.031	0.313
High - Low	-2.315***	-2.180***	-1.871***	-1.200***	-0.511*

Table 1.13 presents results for the Fama-Macbeth regressions on 25 Size-IVOL portfolios. If the inequality beta is associated with idiosyncratic volatility, it should help explain the return dispersion of the portfolios sorted by IVOL. Columns (2) and (4) present results for the Fama-French three-factor and five-factor models, respectively. Columns (3) and (5) present results

for models that augment Inequality_fa with Fama-French three factors and five factors, respectively. Inequality risk premiums are significant for all model specifications incorporating the inequality risk. In Column (3), after controlling for size and value effects, the estimated inequality risk premium equals -0.609% per month, which is economically significant. Similar to the results in Table 1.9, model specifications (3) and (5) have higher adjusted R^2s than model specifications (2) and (4), respectively, indicating that the inequality beta provides incremental pricing information beyond the Fama-French factors.

Table 1.14 presents the inequality betas for 25 Size-IVOL portfolios. As in Table 1.10, we showcase the inequality betas estimated by augmenting the inequality factor with the Fama-French three factors. We observe an approximately monotonic relationship between the inequality beta and IVOL within the first, second, and fifth size quintiles. Within those quintiles, inequality beta tends to increase with idiosyncratic volatility. Stambaugh et al. (2015) argue that idiosyncratic volatility anomaly is caused by arbitrage asymmetry: stocks with higher idiosyncratic volatility often face more stringent short-sale limits and hence are overpriced. Kang et al. (2014) suggest that retail investors cause such overpricing. Their findings are consistent with the results in Table 1.14. If we presume poorer (retail) investors tend to hold stocks with higher inequality betas (Johnson, 2012), we expect stocks with higher inequality betas to have higher idiosyncratic volatility. Additionally, the inequality betas are mostly negative in the lowest idiosyncratic volatility groups and become positive in the highest idiosyncratic volatility groups within each size quintile. This suggests that stocks with the highest idiosyncratic volatilities perform better than stipulated by the Fama-French three-factor model when inequality growth is unexpectedly high, and stocks with the lowest idiosyncratic volatilities perform worse than stipulated by the Fama-French three-factor model in the same situation. Similar to the analysis for varying relative valuation of Table 1.10, stocks with the highest idiosyncratic volatilities could help poorer investors hedge against inequality risk, and in a market dominated by poorer investors, demand lower expected returns than that stipulated by the Fama-French three-factor model.

Table 1.13. Estimates of the inequality risk premium on 25 Size-IVOL portfolios
This table presents the estimated inequality risk premiums by augmenting the inequality tracking portfolio with the Fama-French factors using the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility. Inequality is the beta of Inequality_fa. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret
Inequality	-0.547***		-0.609***		-0.477***
	(-5.22)		(-7.13)		(-4.81)
MKT	0.963**	1.121***	0.885***	1.186***	0.906***
	(2.02)	(3.12)	(4.05)	(3.65)	(4.86)
SMB		1.207***	0.834***	1.113***	0.752**
		(4.54)	(3.98)	(4.38)	(2.28)
VMG		0.855***	0.980***	0.806***	1.134***
		(4.03)	(5.15)	(4.53)	(5.34)
RMW				-0.549*	-0.342*
				(-1.91)	(-1.72)
CMA				0.834**	0.759***
				(1.99)	(3.12)
cons	-4.132***	-3.244***	-3.537***	-5.252***	-2.579***
	(-4.91)	(-4.31)	(-4.01)	(-5.88)	(-4.85)
R^2	0.452	0.545	0.612	0.643	0.696
adj. R^2	0.409	0.483	0.525	0.531	0.574

Table 1.15 presents the model comparison results for 25 Size-IVOL portfolios. The p-values of the GRS statistics indicate non-negligible pricing errors for all model specifications. Similar to the results in Table 1.11, models that augment inequality risk with only the Fama-French factors perform better than the corresponding baseline models.

Table 1.14. Inequality betas of 25 Size-IVOL portfolios

This table presents the inequality betas of 25 Size-IVOL portfolios. The betas are estimated by time series regressions augmenting Inequality_fa with the modified version of Fama-French three factors as in Liu et al. (2019). The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	IVOL									
Size	Low	2	3	4	High	Low	2	3	4	High
	b (3 Factors+)					t (3 Factors+)				
Small	-0.161***	-0.075	-0.003	-0.056	0.033	-3.21	-1.58	-0.23	-0.89	0.63
2	-0.256***	-0.160***	-0.194***	-0.086*	0.057	-4.71	-2.85	-3.17	-1.84	1.03
3	-0.056	-0.085*	-0.102**	0.123**	0.255***	-0.94	-1.76	-2.21	2.41	4.14
4	0.012	-0.033	-0.130**	0.146***	0.123**	0.22	-0.56	-2.38	2.92	2.15
Big	-0.083*	-0.109**	-0.061	0.218***	0.302***	-1.87	-2.04	-0.93	3.20	4.52

Table 1.15. Model performance comparison on 25 Size-IVOL portfolios

This table compares three pricing error metrics measurements from competing asset pricing models. Each row shows the pricing error measurements of different models on 25 Size-IVOL portfolios. CAPM+ stands for the model augmenting inequality risk with CAPM. 3 Factors+ stands for the model augmenting inequality risk with the Fama-French three factors. Furthermore, 5 Factors+ represents the model augmenting inequality risk with the Fama-French five factors. The sample period is from 2005:M2 to 2021:M12.

	CAPM	CAPM+	3 Factors	3 Factors+	5 Factors	5Factors+
GRS statistics	220.124	219.806	216.275	198.215	205.079	197.831
GRS p-value	0.000	0.000	0.000	0.000	0.000	0.000
$A\|\alpha_i\|$	0.391	0.337	0.319	0.281	0.290	0.262
$A\|\alpha_i\|/ A\|\bar{r}_i\|$	0.702	0.604	0.572	0.505	0.521	0.471

1.5.4.3. Inequality beta and turnover

Table 1.7 shows that the inequality beta is also positively correlated with the turnover rate with a mild correlation coefficient of 0.126. Since a high turnover rate implies prominent retail investor behavior, it would be interesting to investigate the relationship between the inequality beta and retail investor ownership through the lens of turnover rate. In this subsection, we investigate a potential nonlinear relationship between the inequality beta and turnover by repeating the analyses in Section 1.5.4.2 on 25 portfolios, first sorted by size and then by turnover rate.

Table 1.16 presents summary statistics for excess returns of the 25 portfolios. As expected, the average excess returns generally decrease monotonically with the turnover rate within each size quintile. The return differences between high-Turn and low-Turn groups are all statistically significant at the 1% level within each size quintile, except for the biggest quintile. Similar to the case of the 25 Size-IVOL portfolios, return differences (in absolute value) between high-Turn and low-Turn groups decrease monotonically with size.

Table 1.16. Average excess returns of the 25 Size-Turn portfolios
This table presents summary statistics for excess returns of the 25 double-sorted portfolios on firm size and turnover rate. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	Size				
Turn	Small	2	3	4	Big
Low	2.492	1.974	1.482	1.228	0.700
2	2.040	1.828	1.474	1.093	1.050
3	1.875	1.190	1.065	1.149	0.934
4	1.411	0.866	0.860	0.777	0.905
High	0.214	-0.114	-0.376	-0.143	0.050
High - Low	-2.278***	-2.087***	-1.857***	-1.371***	-0.650*

Table 1.17 presents results for the Fama-Macbeth regressions on 25 Size-Turn portfolios. If the inequality beta is associated with the turnover rate, it should help price the portfolios sorted by Turn. Columns (2) and (4) present results for the Fama-French three-factor and five-factor models, respectively. Columns (3) and (5) present results for models that augment Inequality_fa with the Fama-French three factors and five factors, respectively. Similar to the results in Table 1.13, the estimated inequality risk premiums are statistically significant for all model specifications that incorporate inequality risk. Similar to the results for the 25 Size-E/P and Size-IVOL portfolios, model specifications (3) and (5) have higher adjusted R^2s than model specifications (2) and (4), respectively.

Table 1.17. Estimates of the inequality risk premium on 25 Size-Turn portfolios
This table presents the estimated inequality risk premiums by augmenting the inequality tracking portfolio with the Fama-French factors using the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and turnover rate. Inequality is the beta of Inequality_fa. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret
Inequality	-1.182**		-1.069***		-1.105***
	(-2.41)		(-3.74)		(-3.89)
MKT	0.712	1.112*	0.802	0.681	0.592
	(1.31)	(1.91)	(1.56)	(1.09)	(0.85)
SMB		0.826***	1.117***	0.684***	0.558***
		(5.65)	(6.01)	(4.56)	(5.14)
VMG		1.534***	1.201***	1.088***	0.835***
		(4.49)	(8.65)	(6.78)	(8.13)
RMW				-0.059	0.124
				(-0.56)	(0.91)
CMA				0.814***	0.443***
				(5.52)	(2.74)
cons	-4.880***	-2.678***	-3.070***	-3.393***	-2.195**
	(-3.68)	(-2.91)	(-3.59)	(-3.47)	(-2.44)
R^2	0.401	0.635	0.662	0.678	0.710
adj. R^2	0.357	0.583	0.601	0.591	0.613

Table 1.18 presents the inequality betas for the 25 Size-Turn portfolios, which are estimated by augmenting the inequality risk factor with the Fama-French three factors as in Table 1.10. We observe an approximately monotonic relationship between the inequality beta and Turn within the first, third and fifth size quintiles. In general, within those quintiles, the inequality beta tends to increase with the turnover rate. There is no clear monotonic relationship between the inequality beta and the turnover rate in the other two size quintiles. The spread in inequality betas is the largest in the smallest size quintile, where the spread in average return is also the largest, as shown in Table 1.16. The inequality betas are mostly negative in the low turnover groups and become positive in the high turnover groups within each size quintile, suggesting that stocks with the highest turnover rate outperform the Fama-French three-factor model when inequality growth is unexpectedly high, helping poorer investors hedge against inequality risk.

Table 1.19 presents model comparison results for the 25 Size-Turn portfolios. Once again, the p-values of the GRS statistics indicate significant pricing errors across all model specifications, and models that augment inequality risk with only the Fama-French factors included perform better than the corresponding baseline models.

Table 1.18. Inequality betas of 25 Size-Turn portfolios

This table presents the inequality betas of 25 Size-Turn portfolios. The betas are estimated by time series regressions augmenting Inequality_fa with the modified version of Fama-French three factors as in Liu et al. (2019). The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	Turn									
Size	Low	2	3	4	High	Low	2	3	4	High
	b (3 Factors+)					t (3 Factors+)				
Small	-0.028	-0.069*	0.072*	0.165***	0.246***	-0.71	-1.85	1.77	4.86	5.49
2	-0.169***	-0.271***	-0.037	-0.108**	0.096*	-3.24	-3.07	-0.55	-2.01	1.78
3	-0.203***	-0.187***	-0.136***	-0.156***	0.075	-4.46	-3.80	-3.12	-3.46	1.42
4	0.051	0.025	0.125**	0.091*	0.188***	0.95	0.68	2.49	1.91	3.55
Big	-0.107**	-0.013	0.189**	0.133***	0.141***	-2.32	-0.05	2.37	3.06	3.28

Table 1.19. Model performance comparison on 25 Size-Turn portfolios

This table compares three pricing error metrics measurements from competing asset pricing models. Each row shows the pricing error measurements of different models on 25 Size-Turn portfolios. CAPM+ stands for the model augmenting inequality risk with CAPM. 3 Factors+ stands for the model augmenting inequality risk with the Fama-French three factors. 5 Factors+ represents the model augmenting inequality risk with the Fama-French five factors. The sample period is from 2005:M2 to 2021:M12.

	CAPM	CAPM+	3 Factors	3 Factors+	5 Factors	5Factors+
GRS statistics	199.357	187.686	187.342	178.496	187.227	177.964
GRS p-value	0.000	0.000	0.000	0.000	0.000	0.000
$A\|\alpha_i\|$	0.317	0.286	0.278	0.272	0.276	0.264
$A\|\alpha_i\|/ A\|\bar{r}_i\|$	0.585	0.528	0.513	0.501	0.509	0.486

1.5.4.4. Inequality beta and ownership concentration

In Table 1.7, the inequality beta is also negatively correlated with ownership concentration with a mild correlation coefficient of -0.104. We expect stocks with low ownership concentration to be held mostly by retail investors thus, have higher inequality betas according to Johnson (2012). In this subsection, we further investigate the nonlinear relationship between inequality beta and ownership concentration by repeating the analyses in Section 1.5.4.2 on 25 portfolios, first sorted by size and then by Shrs. Table 1.20 presents the summary statistics for excess returns of the 25 portfolios. As expected, the average excess returns generally increase monotonically with ownership concentration within each size quintile. However, the return differences between high Shrs and low Shrs groups are only statistically significant within the third, fourth and fifth size quintiles at the 5% level.

Table 1.20. Average excess returns of the 25 Size-Shrs portfolios
This table presents summary statistics for excess returns of the 25 double-sorted portfolios on firm size and ownership concentration. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	Size				
Shrs	Small	2	3	4	Big
Low	1.660	1.092	0.619	0.587	0.526
2	1.947	1.294	0.992	1.027	0.931
3	1.752	1.345	1.084	1.310	1.034
4	1.846	1.227	1.145	1.021	0.945
High	1.841	1.383	1.236	1.003	0.938
High - Low	0.181	0.291*	0.617***	0.416**	0.411**

Table 1.21 presents results for the Fama-Macbeth regressions on 25 Size- Shrs portfolios. Columns (2) and (4) present results for the Fama-French three-factor and five-factor model, respectively. Columns (3) and (5) present results for models augmenting Inequality_fa with the Fama-French three factors and five factors, respectively. Once again, we observe consistent negative inequality risk premiums in all model specifications with inequality risk.

Table 1.22 presents the inequality betas for 25 Size-Shrs portfolios. In line with Tables 1.10, 1.14, and 1.18, we estimate the inequality betas by augmenting the inequality risk factor with the Fama-French three factors. We observe an approximately monotonic relationship between inequality beta and Shrs within the fourth and fifth size quintiles. Within those two quintiles, inequality beta tends to decrease as ownership concentration increases. The phenomenon coincides with Table 1.20, where we observe the Shrs effect to be more prominent within the larger size quintiles. The above results support the hypothesis that retail (poorer) investors tend to hold stocks with higher inequality beta.

Table 1.23 presents model comparison results for the 25 Size-Shrs portfolios. Not surprisingly, the p-values of the GRS statistics suggest significant pricing errors. Once again, for all model specifications, models that augment inequality risk with only the Fama-French factors included outperform the corresponding baseline models.

Overall, the results in Section 1.5.4 coincide with those of Johnson (2012) and confirm the existence of a negative risk premium for inequality risk not only in the cross-section of individual stock returns but also in the cross-section of managed portfolios in China. Johnson (2012) documents an inequality risk premium ranging from -8.59% to -6.34% per annum over the U.S. stock portfolios. On average, this translates to an inequality risk premium of -0.64% per month. Over the sample period (1927:M1 to 2008:M12) of Jonhson (2012), the U.S. has a market risk premium of 0.59% per month (data available from Kenneth French's Website), and the ratio of the absolute value of the U.S. inequality risk premium to the market risk premium is 1.08. In the Chinese market, within our sample period (2005:M2 to 2021:M12), the average monthly inequality risk premium across the test portfolios in Tables 1.9, 1.13, 1.17, and 1.21 equals -0.86% per month, higher than that estimated by Johnson (2012). The Chinese market risk premium equals 0.69% per month during our sample period, and the ratio of the absolute value of the Chinese inequality risk premium to the market risk premium is 1.25. Assuming the ratio of the U.S. inequality risk premium to the market risk premium does not change dramatically from 2009 onward, the above results suggest that the inequality risk premium is more prominent in China than in the U.S., which is not surprising since the Chinese investor base has a much larger proportion of retail investors than that of the U.S. Moreover, we find evidence that higher inequality risk exposure is mildly associated with higher idiosyncratic volatility and turnover rate, as well as lower E/P ratio and ownership concentration. The results are consistent with the conjecture that poorer investors tend to hold assets with higher inequality risk exposures (Johnson, 2012).

Table 1.21. Estimates of the inequality risk premium on 25 Size-Shrs portfolios
This table presents the estimated inequality risk premiums by augmenting the inequality tracking portfolio with the Fama-French factors using the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and ownership concentration. Inequality is the beta of Inequality_fa. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)	(4)	(5)
	Ret	Ret	Ret	Ret	Ret
Inequality	-1.190***		-1.314***		-1.015***
	(-3.61)		(-4.27)		(-2.98)
MKT	0.826	0.844	0.697	0.601	0.797
	(1.28)	(1.59)	(1.21)	(1.36)	(1.58)
SMB		1.033***	0.873**	1.150***	0.781**
		(2.87)	(2.32)	(3.05)	(2.37)
VMG		1.239**	0.956**	1.184**	0.867**
		(2.22)	(2.01)	(2.31)	(2.03)
RMW				0.708**	0.668**
				(2.14)	(2.02)
CMA				-0.282	0.169
				(-1.09)	(0.87)
cons	5.593***	4.586***	5.945***	2.958***	3.784***
	(4.20)	(4.06)	(5.54)	(2.76)	(3.69)
R^2	0.429	0.541	0.610	0.625	0.674
adj. R^2	0.405	0.476	0.533	0.527	0.565

Table 1.22. Inequality betas of 25 Size-Shrs portfolios

This table presents the inequality betas of 25 Size-Shrs portfolios. The betas are estimated by time series regressions augmenting Inequality_fa with the modified version of Fama-French three factors as in Liu et al. (2019). CAPM+ stands for model augmenting inequality risk with CAPM. 3 Factors+ stands for model augmenting inequality risk with the Fama-French three factors. Moreover, 5 Factors+ stands for model augmenting inequality risk with Fama-French five factors. The sample period is from 2005:M2 to 2021:M12.

	Shrs									
Size	Low	2	3	4	High	Low	2	3	4	High
	b (3 Factors+)					t (3 Factors+)				
Small	-0.028	-0.007	-0.123***	-0.089*	-0.146**	-0.52	-0.15	-2.70	-1.87	-2.28
2	-0.094*	-0.159**	-0.140***	-0.201***	-0.105**	-1.91	-2.44	-2.61	-3.68	-2.17
3	0.062	0.073*	0.102***	0.047	-0.070*	1.22	1.72	2.71	1.26	-1.76
4	-0.051	-0.048	-0.079*	-0.080*	-0.131**	-1.01	-0.88	-1.83	-1.92	-2.43
Big	0.070	0.032	-0.022	-0.028	-0.073*	1.43	0.74	-0.66	-0.75	-1.69

Table 1.23. Model performance comparison on 25 Size-Shrs portfolios

This table compares three pricing error metrics measurements from competing asset pricing models. Each row shows the pricing error measurements of models on 25 Size-Shrs portfolios. CAPM+ stands for the model augmenting inequality risk with CAPM. 3 Factors+ stands for the model augmenting inequality risk with the Fama-French three factors. 5 Factors+ represents the model augmenting inequality risk with the Fama-French five factors. The sample period is from 2005:M2 to 2021:M12.

	CAPM	CAPM+	3 Factors	3 Factors+	5 Factors	5 Factors+
GRS statistics	92.339	91.518	90.236	87.204	85.641	85.314
GRS p-value	0.000	0. 000	0.000	0.000	0.000	0.000
$A\|\alpha_i\|$	0.189	0.163	0.155	0.151	0.149	0.144
$A\|\alpha_i\|/ A\|\bar{r}_i\|$	0.632	0.544	0.519	0.506	0.498	0.483

1.6. Robustness checks

For robustness, we repeat our baseline regressions in Section 1.5.3 with winsorized inequality beta, rolling inequality beta, and over a subsample before COVID-19.

To rule out the possibility that extreme values of the estimated inequality betas drive the results in Tables 1.5 and 1.6, we repeat the analyses using winsorized inequality betas. Table 1.24 presents the results for estimated risk premiums using the Fama-Macbeth (1973) procedure, where the inequality betas are winsorized at the 2.5th and 97.5th percentiles. The results are qualitatively similar to those of the corresponding model specifications in Tables 1.5 and 1.6. We also repeat the analyses using inequality betas winsorized at the 1th and 99th percentiles and trimmed at the 1% and 2.5% tails. The results remain robust.

We repeat the analyses in Tables 1.5 and 1.6 over a sample excluding the period affected by COVID-19. The sudden outbreak of COVID-19 at the end of 2019 has exerted a profound influence on economies and societies around the globe. Studies show that COVID-19 increased equity market risk exposure (Li et al., 2022), amplified international stock market risk contagion (Liu et al., 2022), and changed central bank behavior (Cox et al., 2022). Other studies suggest that COVID-19 left a significant impact on investor psychology and behavior (Ortmann et al., 2020; Naseem et al., 2021; Bouri et al., 2021; Smales et al., 2021a; Smales et al., 2021b). To investigate whether COVID-19 has biased our results, we re-estimate the models in Tables 1.5 and 1.6 using data from February 2005 to November 2019. Once again, the results are qualitatively similar. Inequality continues to command significant negative risk premiums at the 5% level in Columns (2), (3), and (5). The risk premiums for Shrs, IVOL, LnSize, Turn, and Illiquidity also remain significant with the correct signs. The only difference is that short-term reversal Ret_1 now has a significant negative risk premium in the full model.

Table 1.24. Firm-level estimates of inequality risk premia using winsorized inequality beta

This table presents the estimated risk premiums by the Fama-Macbeth (1973) procedure using winsorized inequality beta at 2.5th and 97.5th percentile. Inequality is the firm-level inequality risk beta estimated by augmenting Inequality_fa with Fama-French 3 factors with a rolling window of 60 months. Shrs is the lagged sum of the shareholding percentage of the second largest shareholder to the tenth largest shareholder. LnSize is the log of the firm's market capitalization at the end of last month. LnBM is the lagged 1-month log of book-to-market ratio. IVOL is idiosyncratic volatility over the past month. Turn is the average daily turnover rate during the past month. Illiquidity is Amihud's (2002) illiquidity measure. Ret_1 is a measure of the short-term reversal effect. Ret2_12 measures the momentum effect. Beta is the market beta estimated using daily returns over the past year. Leverage is the lagged 1-month debt-to-equity ratio. The Newey and West (1987a) t-statistics corrected for serial correlation and heteroskedasticity are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret_rf	(2) Ret_rf	(3) Ret_rf	(4) Ret_rf	(5) Ret_rf
Inequality	-0.233*	-0.351**	-0.305***		-0.348***
	(-1.81)	(-2.48)	(-2.71)		(-2.85)
Shrs					0.008***
					(3.03)
IVOL				-0.303***	-0.200***
				(-4.12)	(-3.45)
LnSize		-0.358**	-0.344**	-0.277**	-0.331**
		(-2.48)	(-2.43)	(-2.11)	(-2.30)
LnBM		0.578***	0.230**	0.093	0.153
		(5.55)	(2.20)	(0.92)	(1.59)
Turn			-2.071***	-2.032***	-1.635***
			(-8.53)	(-7.73)	(-7.10)
Illiquidity			5.030**	4.145**	6.048**
			(2.20)	(2.48)	(2.25)
Ret_1			-0.019**	-0.009	-0.013
			(-2.24)	(-1.02)	(-1.50)
Ret2_12			0.003	0.003	0.002
			(0.81)	(1.08)	(0.48)
Beta	-0.002	-0.002	-0.001	-0.001	-0.005
	(-1.13)	(-1.07)	(-1.03)	(-1.04)	(-1.26)
Leverage			-0.110	-0.146	-0.062
			(-0.53)	(-0.82)	(-0.27)
cons	2.345**	4.224***	4.717***	4.832***	4.428***
	(2.41)	(3.17)	(3.46)	(3.36)	(3.19)
R^2	0.032	0.079	0.118	0.112	0.126
adj. R^2	0.030	0.076	0.112	0.106	0.118

Table 1.25. Firm-level estimates of inequality risk premia before COVID-19
This table presents the estimated risk premiums by the Fama-Macbeth (1973) procedure using data before COVID-19. Inequality is the firm-level inequality risk beta estimated by augmenting Inequality_fa with Fama-French 3 factors with a rolling window of 60 months. Shrs is the lagged sum of the shareholding percentage of the second largest shareholder to the tenth largest shareholder. LnSize is the log of the firm's market capitalization at the end of last month. LnBM is the lagged 1-month log of book-to-market ratio. IVOL is idiosyncratic volatility over the past month. Turn is the average daily turnover rate during the past month. Illiquidity is Amihud's (2002) illiquidity measure. Ret_1 is a measure of the short-term reversal effect. Ret2_12 measures the momentum effect. Beta is the market beta estimated using daily returns over the past year. Leverage is the lagged 1-month debt-to-equity ratio. The Newey and West (1987a) t-statistics corrected for serial correlation and heteroskedasticity are in parentheses. The sample period is from 2005:M2 to 2019:M11. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret_rf	(2) Ret_rf	(3) Ret_rf	(4) Ret_rf	(5) Ret_rf
Inequality	-0.278*	-0.306**	-0.331***		-0.314***
	(-1.72)	(-2.32)	(-2.85)		(-2.73)
Shrs					0.008***
					(2.86)
IVOL				-0.366***	-0.216***
				(-4.70)	(-3.90)
LnSize		-0.368**	-0.338**	-0.291*	-0.324**
		(-2.38)	(-2.23)	(-1.93)	(-2.07)
LnBM		0.589***	0.235**	0.075	0.155
		(5.32)	(2.11)	(0.58)	(1.46)
Turn			-2.078***	-1.822***	-1.590***
			(-7.88)	(-6.31)	(-6.21)
Illiquidity			6.237***	4.602***	3.787***
			(2.65)	(3.28)	(2.78)
Ret_1			-0.023***	-0.013*	-0.019**
			(-3.23)	(-1.81)	(-2.44)
Ret2_12			0.001	0.003	0.001
			(0.40)	(0.78)	(0.32)
Beta	0.002	0.003	0.002	0.001	0.001
	(1.00)	(1.23)	(1.09)	(1.04)	(1.03)
Leverage			-0.045	-0.121	0.031
			(-0.18)	(-0.47)	(0.11)
cons	2.733**	4.851***	4.777***	5.087***	4.879***
	(2.36)	(3.18)	(2.92)	(3.26)	(2.81)
R^2	0.030	0.077	0.118	0.110	0.124
adj. R^2	0.028	0.075	0.111	0.103	0.116

Table 1.26. Firm-level estimates of inequality risk premia using rolling inequality beta
This table presents the estimated risk premiums by the Fama-Macbeth (1973) procedure using rolling betas of Inequality_fa. Inequality is the firm-level inequality risk beta estimated by augmenting the inequality index tracking portfolio with Fama-French 3 factors with a rolling window of 60 months. Shrs is the lagged sum of the shareholding percentage of the second largest shareholder to the tenth largest shareholder. LnSize is the log of the firm's market capitalization at the end of last month. LnBM is the lagged 1-month log of book-to-market ratio. IVOL is idiosyncratic volatility over the past month. Turn is the average daily turnover rate during the past month. Illiquidity is Amihud's (2002) illiquidity measure. Ret_1 is a measure of the short-term reversal effect. Ret2_12 measures the momentum effect. Beta is the market beta estimated using daily returns over the past year. Leverage is the lagged 1-month debt-to-equity ratio. The Newey and West (1987a) t-statistics corrected for serial correlation and heteroskedasticity are in parentheses. The sample period is from 2010:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret_rf	(2) Ret_rf	(3) Ret_rf	(4) Ret_rf	(5) Ret_rf
Inequality	-0.010	-0.239	-0.258		-0.306
	(-0.04)	(-1.48)	(-1.12)		(-1.51)
Shrs					0.011***
					(3.76)
IVOL				-0.281***	-0.256***
				(-3.32)	(-2.95)
LnSize		-0.262*	-0.310**	-0.252**	-0.344**
		(-1.83)	(-2.23)	(-2.11)	(-2.20)
LnBM		0.490***	0.157	0.063	0.084
		(3.97)	(1.31)	(0.52)	(0.62)
Turn			-2.083***	-1.986***	-1.744***
			(-8.37)	(-7.62)	(-7.08)
Illiquidity			5.836**	3.621**	8.002***
			(2.24)	(2.17)	(2.94)
Ret_1			-0.019**	-0.011	-0.010
			(-2.33)	(-1.23)	(-1.11)
Ret2_12			0.003	0.003	0.001
			(1. 01)	(1.32)	(0.24)
Beta	-0.324	-0.236	-0.005	-0.011	0.929
	(-1.24)	(-0.98)	(-0.02)	(-0.03)	(1.31)
Leverage			-0.112	-0.091	-0.112
			(-0.61)	(-0.44)	(-0.51)
cons	1.939**	4.054***	4.695***	5.026***	4.693***
	(2.21)	(2.78)	(3.02)	(3.71)	(3.07)
R^2	0.023	0.070	0.112	0.101	0.114
adj. R^2	0.021	0.065	0.103	0.096	0.105

Due to concerns about statistical power, we estimate inequality betas in Tables 1.5 and 1.6 over the entire sample. Following Ferson and Harvey (1999), we also repeat the analyses using rolling inequality betas with a window of 60 months. The results are presented in Table 1.26. While the estimated inequality risk premiums are negative in Columns (1), (2), (3), and (5), they are not statistically significant at the 5% level. The R^2s also shrink compared to the corresponding model specifications in Tables 1.5 and 1.6. The phenomenon may be attributed to increased measurement errors in estimated inequality betas induced by much shorter time series.

1.7. Conclusions

This study investigates the pricing implication of inequality risk in the cross-section of Chinese stock returns. We construct three monthly proxies that measure the year-on-year changes in Chinese consumption and income inequality using three different dimensionality reduction methods: the dynamic factor model, the entropy weight method, and principal component analysis. We test whether inequality risk is priced in the cross-section of individual stock returns following Fama-Macbeth (1973). Our findings show that inequality is an important systematic risk priced in the cross-section of Chinese stock returns. Consistent with Johnson (2012), we find a significant negative inequality risk premium for both individual stocks and portfolios. Notably, our results suggest that the inequality risk premium is larger in China than in the U.S., which is unsurprising since retail investors play a more substantial role in the Chinese stock market. At the firm-level, we also find that the inequality beta is mildly negatively correlated with the log of the book-to-market ratio and ownership concentration, and positively correlated with idiosyncratic volatility and turnover rate. Aside from these characteristics, the inequality beta is not significantly correlated with other firm characteristics. These are consistent with Johnson (2012), which suggest that poorer investors tend to hold stocks with higher inequality betas in a CRRA economy with incomplete markets.

Since poorer investors are often retail investors, we expect firms with higher inequality betas to have higher turnover rates, lower ownership concentration, and tend to be overpriced with lower B/M or E/P ratios and higher idiosyncratic volatilities. To further explore these relationships, we conduct portfolio analysis to identify potential nonlinear relationships between the inequality beta and these four firm characteristics related to retail investor ownership. We extend firm-level asset pricing tests to four sets of size-neutral portfolios sorted by E/P ratio, idiosyncratic volatility, turnover rate, and ownership concentration. We investigate whether inequality beta helps explain the return dispersion of these portfolios and examine how the portfolios' inequality exposures relate to firm characteristics. We also compare the performance of different model specifications over each set of test portfolios. Moreover, we evaluate whether inequality is a valid candidate as an ICAPM state variable following Maio and Santa-Clara (2012). Our results show that inclusion of the inequality beta helps explain return dispersions of the portfolios sorted by those four firm characteristics. We find evidence that higher inequality risk beta is mildly positively associated with idiosyncratic volatility and turnover, and mildly negatively associated with the E/P ratio and ownership concentration. The results support that poorer investors tend to hold assets with higher inequality risk exposures as suggested by Johnson (2012). Our results also show that across all four sets of test portfolios, augmenting the inequality factor with the Fama-French factors helps to reduce pricing errors and improve model performance. We also find evidence that inequality may be a valid candidate as an ICAPM state variable. Our main results are robust for different versions of inequality factors and are robust to different treatments of variable winsorization, and remain robust when excluding the sample period affected by COVID-19. In Chapter 3, we will expand the universe of competing factors and investigate whether inequality remains a significant factor in pricing the cross-section of asset returns in a more extensive context.

Chapter 2. Macroeconomic risks and the cross-section of stock returns in China

2.1. Introduction

Since the CAPM's introduction over 60 years ago, academia has explored hundreds of factors to explain the cross-section of asset returns. Research such as Harvey et al. (2016) and Hou et al. (2017) highlights the vastness of this factor universe. Connor (1995) and Chan et al. (1998) divide the factors into three groups: macroeconomic factors, statistical factors, and fundamental factors.

Macroeconomic factors are derived from theoretically inspired macroeconomic shocks. Such models assume that variations in the cross-section of asset expected returns are driven by exposures to macroeconomic shocks. The Chen-Roll-Ross model (Chen et al., 1986) is a leading example. Cooper et al. (2022) extended this model to international markets, documenting the pervasive negative correlations between size and momentum factors is mainly due to their differing exposures to a global version of the Chen-Roll-Ross factors.

Statistical factors, primarily inspired by the APT, focus on the covariance structures of assets. These factors are typically constructed using maximum likelihood or principal component analysis over a panel of asset returns. Following early, unsatisfactory attempts by studies such as Roll and Ross (1980) and Connor and Korajczyk (1988) on individual assets, statistical factors gained renewed attention with recent advancements in machine learning (e.g., Kozak et al., 2018, 2020; Kelly et al., 2019; Lettau & Pelger, 2020a, b).

Fundamental factors, constructed from asset characteristics, do not require the calculation of factor exposures and have attracted significant attention in academia. The Fama-French five factors (Fama & French, 2015) are the most prominent and influential.

Connor (1995) notes that in the absence of estimation errors and data limitation, despite different constructions, the three types of factors are rotations of the same factor structure. For example, an asset’s exposure to a specific macroeconomic factor may be expressed as a linear combination of the asset's characteristics. Statistical factors may also correlate with certain macroeconomic factors (Connor & Korajczyk, 1988). Among the three types of factors, macroeconomic factors occupy a central role due to their close connection to macro-finance theories. As Cochrane (2005) puts it: “The central and unfinished task of absolute asset pricing is to understand and measure the sources of aggregate or macroeconomic risk that drive asset prices…In turn, I think that what we are learning about finance must feed back on macroeconomics” (p.xiv). Thus, this study aims to contribute to this literature by directly investigating the pricing implications of macroeconomic factors and their relationship with statistical and fundamental factors.

Our study focuses on the Chinese stock market, which has become the second-largest in the world. Despite its size, there is limited research on how macroeconomic risks affect the cross-section of Chinese stock returns. This gap may be attributed to a disconnection between the economy and the market in the early years, as noted by Girardin and Liu (2003), who likened the Chinese stock market to a casino. However, Girardin and Joyeux (2013) show that since China’s joined the WTO in 2001, the link between the market and the macroeconomy has gradually strengthened. Other studies investigating the time series relationships between the market index and macroeconomic variables also support this finding.

To address this knowledge gap, this study aims to answer the following research questions: Are macroeconomic risks priced in the Chinese stock market? Which macroeconomic factors provide incremental pricing information? Which macroeconomic factors earn significant risk premiums? Are macroeconomic factor models consistent with the ICAPM restrictions? How are macroeconomic factors, statistical factors, and fundamental factors related?

To answer these questions, we construct three sets of factors. The first set consists of statistical factors, which are derived by extracting the largest five principal components from the returns of nine main asset classes: the CSI 300 Stock Index, the CSI 500 Stock Index, the CSI 1000 Stock Index, the CSI Government Bond Index, the CSI Corporate Bond Index, the Au9999 Index traded on the Shanghai Futures Exchange, the Nanhua Industrial Commodity Index, the Nanhua Metal Commodity Index and the Nanhua Energy Commodity Index. We analyze the loading patterns on these main asst classes to understand the underlying economic meanings of the principal components. Since the economic meanings embedded in the main asset class returns are more trackable than those in individual stock returns, which are more commonly used in prior research, our principal components straddle the line between statistical and macroeconomic factors, and can be effectively regarded as a set of macroeconomic factors in a broader sense. The second set of factors consists of the Chen-Roll-Ross factors with minor modifications discussed in Section 2.3.1.4. The third set comprises the Fama-French five factors, with adjustments to the size and value factors following Liu et al. (2019) to better align with the Chinese market.

We test different specifications of factor models over three sets of test portfolios. We investigate whether each factor provides incremental pricing information by employing GMM estimation using the stochastic discount factor approach. We then evaluate whether the factors earn significant risk premiums using the Fama-Macbeth (1973) procedure. For robustness, we also estimate asset pricing models by GMM using the expected return-covariance risk premium approach. To address Fama's (1991) concern about abusing the ICAPM as a fishing license, we explicitly check the consistency of our factors with the ICAPM. We also test the relationship between the macroeconomic factors and the Fama-French five factors within the framework of GMM. Finally, we compare the performance of the three sets of factors using two different metrics.

The empirical results show that macroeconomic risks are effective for pricing stocks in the Chinese market. Most of the tested macroeconomic factors exhibit significant loadings on the stochastic discount factor. Furthermore, the majority of these factors are associated with significant risk premiums as estimated by the Fama-Macbeth procedure, though the Chen-Roll-Ross factors may not always align with the ICAPM restrictions.

The performance of the Chen-Roll-Ross model varies across test portfolios. Generally, we find significant negative risk premiums for unexpected inflation and default spread, and weaker evidence of positive risk premiums for term spread in one of the three sets of test assets. These findings are generally consistent with Cooper et al. (2022) but diverge from earlier studies on the U.S. market, such as Hahn and Lee (2006) and Aretz et al. (2010), which do not find significant default spread premiums. Additionally, we find significant positive risk premiums for expected inflation on two sets of test assets, contradicting the findings of Chan et al. (1986), Aretz et al. (2010), and Cooper et al. (2022). The industrial production factor, known to earn consistent positive risk premiums (Chan et al., 1986; Cooper et al., 2022), demonstrates inconsistent performance across different sets of test assets in our study with sign reversals on estimated risk premiums. Our results suggest that while the vanilla Chen-Roll-Ross factors contain some useful pricing information, their explanatory power for the cross-section of Chinese stock returns is relatively limited. A similar conclusion can be made on the performance of the statistical macroeconomic factors.

The test results indicate that the Fama-French five factors are redundant in the presence of the macroeconomics factors, implying that the five factors primarily capture information related to macroeconomic shocks. This finding is in line with a series of studies by Liew and

Vassalou (2000), Vassalou (2003), Hahn and Lee (2006), Hanhardt and Ansotegui Olcoz (2008), Aretz et al. (2010), Kim et al. (2011) and Leite et al. (2020), which investigate the relationship between macro and Fama-French factors in the U.S. and Eurozone. While the performances of the macro-inspired statistical factors and Chen-Roll-Ross factors are less stable across test assets compared to the Fama-French five factors, both sets of factors outperform the Fama-French three factors.

Our study contributes to the literature in the following ways. First, to the best of our knowledge, our work is the first comprehensive study of the cross-sectional implications of macroeconomic and statistical factors in the Chinese stock market. Second, our investigation sheds light on the relationship between fundamental factors and macroeconomic risk, adding to the strand of literature that directly seeks economic meanings of the Fama-French factors. Our study is also the first to compare the performance of the three types of factors within the context of the Chinese stock market.

The rest of the chapter is organized as follows: Section 2.2 briefly reviews prior studies on the three types of factors, their relationships, and the cross-section of stock returns in China. Section 2.3 outlines the econometric methodologies for constructing macroeconomic factors, estimating asset pricing models, testing model consistency with the ICAPM, and comparing model. Section 2.4 provides details of data and variables. Section 2.5 presents the empirical results of the asset pricing models and model comparisons. Section 2.6 further presents results of robustness tests using GMM with a different set of moment conditions from those in section 2.5. Finally, Section 2.7 concludes and summarizes the main findings of the chapter.

2.2. Literature Review

The main task of asset pricing is to explain the variation in the cross-section of asset returns. Early endeavors of single-factor CAPM by Sharpe (1964), Lintner (1965), and Black (1972) are unsatisfactory due to its failure to explain the cross-section of portfolios sorted by firm characteristics such as size and the book-to-market ratio. As remedies, different forms of multifactor models emerge, most of which are under the frameworks of the ICAPM (Merton, 1973), APT (Ross, 1976), CCAPM (Rubinstein, 1976; Lucas, 1978), and the closely related investment-based models (Cochrane,1991,1996).

Among the numerous efforts by academia to search for multifactor models with explanatory powers beyond that of CAPM, the main direction of efforts is to investigate the return predictabilities of firm characteristics (fundamental factors). The Fama-French three-factor model (Fama & French, 1992，1993, 1996), the Carhart four-factor model (Carhart, 1997), the Hou et al. (2015) four-factor model, and the Fama-French five-factor model (Fama & French, 2015, 2016) are crowning jewels of this strand of literature. Beyond size, value, investment, profitability, and momentum, academia also comes up with hundreds of other fundamental factors that are claimed to possess the ability to explain the cross-section of stock returns. The rapid expansion of the number of claimed fundamental factors also draws questioning and investigation by studies such as Harvey et al. (2016), in which the authors scrutinize a large number of factors, including 202 characteristics factors, under a multiple testing framework. Their findings suggest that most of the claimed factors are likely spurious. Hou et al. (2020) extend the work of Harvey et al. (2016) to 452 fundamental anomaly variables and find out that only 18% of them survive the multiple testing framework. On the contrary, a recent study by Jensen et al. (2023) reaches an opposite conclusion. They assess the replicability of 153 fundamental anomalies across 93 countries using a Bayesian model, which sacrifices much less power than the multiple testing framework of Harvey et al. (2016), and find out that most anomalies are replicable.

Another natural direction of effort is to extract factors from the covariance structure of asset returns, as Ross (1976) suggested, and explain the dispersion of expected asset returns by their exposures to those statistical factors. Roll and Ross (1980) and Chen (1983) are among the earliest of those explorations. They construct statistical factors extracted from daily returns of individual assets and find some evidence that loadings on those factors help explain the cross-section of stock returns. However, subsequent studies by Lehmann and Modest (1988) show that those statistical factors do not demonstrate superior performance over CAPM in explaining the size effect. Connor and Korajczyk (1988) construct statistical factors from individual stock returns using the method of Asymptotic Principal Components (Connor & Korajczyk, 1986). In line with Lehmann and Modest (1988), their empirical findings show that statistical factors produce significant pricing errors over portfolios formed by size.

The statistical factors have not regained much favor in academia until more recently. Kelly et al. (2019) propose an Instrumented Principal Component Analysis (IPCA) method allowing for time-varying factor loadings that depend on observable firm characteristics and find out that their statistical factors outperform the Fama-French five-factor model in terms of pricing errors both in-sample and out-of-sample. Lettau & Pelger (2020a, b) propose a Risk Premium-Principal Component Analysis (RP-PCA) method, which helps identify latent factors with small variances but helps explain the cross-section of stock returns. Their results show that the RP-PCA factors have much higher Sharpe ratios and smaller pricing errors than conventional PCA factors. Most importantly, Kozak et al. (2018, 2020) show that contrary to previous studies which build PCA factors from individual stock returns, PCA factors built from portfolios formed with firm characteristics outperform the CAPM and Fama-French six-factor (the Fama-French five factors augmented with the momentum factor) model over different sets of test portfolios out-of-sample. Thus, in this study, we construct our statistical factors from returns on several indices rather than individual stocks.

The third direction of effort is to connect cross-sectional dispersion in expected asset returns with exposures to macroeconomic risks. Chen et al. (1986) is the most influential work among those studies. They first identify five macroeconomic variables that could influence expected cash flows or discount rates, then perform Fama-Macbeth (1973) regressions with innovations of those variables on twenty portfolios sorted by size. Their results show that industrial production, default spread, and unexpected inflation are significantly priced at the 5% level, and premiums on expected inflation and term spread are significant at the 10% level. They also augment the macroeconomic factors with the market return and find that market beta does not provide incremental explanatory power. Although Shanken and Weinstein (2006) doubt the robustness of the Chen-Roll-Ross (CRR) model, Cooper et al. (2022) provide solid support to Chen et al. (1986) by extending their work to a comprehensive international dataset with a more recent sample period. They construct five global CRR factors by the GDP-weighted averages of the country-wise CRR factors (except for the global default spread factor, which is proxied by the U.S. default spread due to data limitation) and test their cross-sectional pricing implications on 48 global asset portfolios over the sample period from April 1983 to December 2018. They record significant risk premiums on industrial production, default spread, and expected inflation, although the estimated default spread premium has an opposite sign (negative) to that of Chen et al. (1986). They also conducted model comparisons and found a superior performance of the global CRR model over the global five-factor model of Fama and French (2017).

Beyond the Chen-Roll-Ross factors and consumption growth, academia has identified many other macroeconomic risks that may affect the cross-section of asset returns. A small subset of which include exchange rate risk (Jorion, 1991; Lustig & Verdelhan, 2007; Chung et al., 2015), labor income risk (Campbell, 1996; Jagannathan & Wang, 1996; Santos & Veronesi, 2006; Kim et al., 2011), macro liquidity risk (Chan et al.,1996; Jung & Kim, 2016;), monetary policy risk (Maio & Santa-Clara, 2017; Detzel, 2017; Ozdagli, & Velikov, 2020; Bu & Wu, 2021), economic policy uncertainty (Pastor & Veronesi, 2012; Brogaard & Detzel, 2015; Lee et al., 2021), commodity risks (Boons et al., 2014; Brooks et al., 2016) and inequality risk (Johnson, 2012).

Many of those factors are proposed under the theoretical framework of the ICAPM, leading to Fama's (1991) "fishing license" criticism. Maio and Santa-Clara (2012) explicitly test a variety of macroeconomic factors to verify their consistency with the ICAPM. They find that most of the factors tested are inconsistent with model restrictions imposed by the ICAPM. To address the same issue, we also draw upon the work of Maio and Santa-Clara (2012) and check the consistency of our macro factors with the ICAPM in this study.

As the number of factors accumulates, a strand of literature emerges investigating the relationships among the three types of factors. Among those studies, most of the focus is on the relationship between fundamental and macroeconomic factors. Fama and French (1993) conjecture that the size and value factors are proxies for unobservable distress risk. However, Daniel and Titman (1997) disagree and argue that no pervasive factors are embedded in the size and value factors, and it is firm characteristics but not risk exposures that drive the dispersion in the cross-section of expected stock returns. Since then, a stream of studies has emerged reinforcing Fama and French (1993).

Vassalou (2003) adopts the Lamont (2001) economic tracking portfolio approach and finds that the Fama-French size and value factors contain mainly news about future GDP growth. Following Vassalou (2003), Hahn and Lee (2006) investigate the relationship between thenFama-French and yield spread factors. They find that the Fama-French size and value factors capture information about term spread and default spread, and turn superfluous in the presence of those factors. Petkova (2006) discovers that a set of macroeconomic factors, including innovations in aggregate dividend yield, term spread, default spread, and one-month Treasury-bill yield, are significantly correlated with Fama-French size and value factors. Moreover, the macroeconomic factors perform even better than the Fama-French three-factor model over 25 portfolios sorted by size and book-to-market. Furthermore, when the macroeconomic factors are augmented with Fama-French size and value factors, the latter becomes superfluous. Aretz et al. (2010) also extend the work of Vassalou (2003) and find that the Fama-French size and value factors and the momentum factor capture mainly information about five macroeconomic factors and, thus, can be regarded as proxies for composite macroeconomic risk factors. Cooper and Maio (2019) show that risk premiums on fundamental factors are associated with their exposures to some macroeconomic factors that forecast future investment opportunities under the framework of the ICAPM. Liu and Zhang (2008) find that exposure to industrial production growth can largely explain the momentum factor. Maio and Philip (2018) find that industry momentum portfolios can be satisfactorily explained by simple two-factor models consisting of the market factors and one of the following macroeconomic factors, including industrial production, capacity utilization rate, retail sales, and a broad economic index. Amenc et al. (2019) demonstrate that common fundamental factors have significant exposures to seven macroeconomic factors, indicating that diversification strategies based on fundamental factors do not necessarily reduce portfolios' exposures to systematic risks. Cooper et al. (2022) also add to this strand of literature by providing solid evidence that the value and momentum factors are driven by their exposures to some fundamental macroeconomic factors. The pervasive phenomenon that value and momentum factors have positive risk premiums but are negatively correlated in different countries can be explained by their exposures to five global CRR macroeconomic factors. This study also adds to this strand of literature by providing supportive evidence from the Chinese stock market for the first time.

Explorations on how statistical factors relate to the other two types of factors are few. Fama (1991) criticizes that one of the most significant obstacles that hinders the development of statistical factors is the difficulty in associating statistical factors with precise economic meanings. A solid step forward is by Kelly et al. (2019). Their IPCA approach would identify the latent factors if exposures to them drive the characteristics/expected return relationship. If no latent factor is identified, the IPCA would recognize the characteristics/expected return relationship as an anomaly. Their empirical work successfully identifies five latent IPCA factors, indicating that firm characteristics are related to time-varying factor loadings on some latent factors. Connor and Korajczyk (1988) show that the largest five principal components they construct can explain a substantial portion of the variation in term spread and default spread. They deem such a relationship a necessary but not sufficient condition for the statistical factors to have economic meanings. In this study, we infuse our statistical factors with economic substance by performing PCA on main asset class returns, which ought to have a more stable relationship with macroeconomic risks than individual assets and characteristics-based portfolios.

Studies on the cross-section of stock returns in the Chinese market are overwhelmingly focused on the pricing implications of fundamental factors. One strand of literature assesses the performance of prevailing factors models such as the Fama-French three- and five-factor models on the Chinese stock market. Most studies have found the size effect to be prominent and robust. Guo et al. (2017) test the Fama-French five-factor model in the Chinese market over a sample period from July 1995 to June 2014. They find robust size, value, and profitability effects but weaker evidence of investment effect. Jiao and Lilti (2017) find that the explanatory powers of the Fama-French investment and profitability factors vary across test portfolios over a shorter sample period of July 2010 to May 2015. Huang (2019) discovers that the performance of the Fama-French five-factor model is unstable over the sample period of 1994-2016. Chen et al. (2022) compare the Fama-French three- and five-factor models over the sample period between January 2004 and December 2017 and find that the latter outperforms the former. Moreover, they suggest including a short-term-reversal factor and two ownership-based factors to improve the performance of the five-factor model in the Chinese stock market. Hu et al. (2019) show that the Fama-French value factor is insignificant in the Chinese market over the sample period of 1995-2016. Liu et al. (2019) argue that the insignificant value premium is caused by reverse merger "shell values" of the small-size firms. They address the problem by proposing the earning-to-price ratio as a better measure of relative value and dropping the smallest 30% of firms in the sample. Their new value factor, VMG, earns a significant risk premium. Another strand of literature focuses on the return predictability of other characteristics beyond the Fama-French five factors at the firm level (Wang & Di, 2007; Chen et al., 2010; Cakici et al., 2017). Nartea et al. (2013), Gu et al. (2018), and Gu et al. (2019) study the idiosyncratic volatility and discover the low-risk anomaly similar to those found in other markets. Nartea et al. (2017) document a MAX effect in the Chinese market similar to Kumar, A. (2009). An et al. (2020), Zhang et al. (2020), and Su et al. (2022) study the illiquidity premium in the Chinese market. And the list goes on.

Compared with fundamental factors, macroeconomic and statistical variables draw much less attention. Moreover, most studies investigating the impact of macroeconomic factors on the Chinese stock market aim to test time series relationships between macroeconomic variables and the market index. Few studies examine the cross-sectional pricing implications of macroeconomic factors in China. There is mixed evidence among existing studies on whether and how macroeconomic factors affect the Chinese stock market. Soenen, L., & Johnson, R. (2001) find no effect of inflation and real output on the Chinese market return. Gay Jr (2008) observes that neither the exchange rate nor the oil price affects the Chinese market return. Allen et al. (2023) find that listed stocks in the Chinese A-share market have systematically lower returns than externally listed Chinese stocks and stocks in other countries controlling for country-level macroeconomic variables. They attribute the discrepancy between Chinese stock returns and economic fundamentals to investor sentiment. Yao et al. (2013) find no effect of monetary policy on the Chinese market return, while Koivu (2012) find the opposite. Evidence shows that the bond between the Chinese stock market and macroeconomic variables was strengthened after China joined the WTO and the Global Financial Crisis (GFC). Girardin and Joyeux (2013) detect a significantly stronger relationship between Chinese market return and macroeconomic fundamentals after China joined the WTO in late 2001. Similarly, Goh et al. (2013) find that some of the Goyal and Welch (2008) style U.S macroeconomic variables can significantly predict Chinese stock return post-GFC. In contrast, those variables do not show any predictive powers before the GFC. Using the GARCH–BEKK model, Guo (2015) detects no significant relationship between real economic growth and the market return in China before the GFC. However, he records significant unidirectional causality in the mean from real economic growth to the market return during the post-GFC period. Those findings are further supported by Abbas et al. (2019), who record enhanced spillover effects from some macroeconomic variables to the Chinese market return and volatility in the post-GFC period. Thus, a comprehensive study on the cross-sectional pricing implication of macro factors is needed.

2.3. Empirical methodology

2.3.1. Methods related to the construction of macroeconomic factors

We test and compare three sets of factors using GMM and Fama-Macbeth's (1973) procedure on three sets of test assets: 25 Size-E/P portfolios, 25 Size-IVOL portfolios, and 31 industry portfolios. The three sets of factors include factors as in Chan et al. (1986) and Cooper et al. (2022) (hereafter denoted CRR factors), macroeconomic factors extracted by the principal component analysis from returns of nine main asset classes (henceforth denoted PCA factors), and the Fama-French five factors.

2.3.1.1. Principal component analysis (PCA) factors

Unlike macroeconomic factors constructed directly from macroeconomic variables and their tracking portfolios, we can also extract factors that proxy for the news of macroeconomic variables from a collection of representative main asset class returns. In line with the setup in Ross (1976), Pelger (2019), Pelger (2020), and Lopez-Lira and Roussanov (2020), suppose in the matrix form, the standardized returns of the main asset classes have the following factor structure:

$$\underset{(T\times N)}{\boldsymbol{R_{std}}} = \underset{(T\times K)}{\boldsymbol{F}}\underset{(K\times N)}{\boldsymbol{B}^T} + \underset{(T\times N)}{\boldsymbol{e}} \tag{2.1}$$

where $\boldsymbol{R_{std}}$ is a $T \times N$ matrix whose columns are time series of the standardized main asset class returns. $\boldsymbol{F}$ is a $T \times K$ matrix whose columns are the time series of K common macroeconomic factors that drive the variation in main asset class returns. $\boldsymbol{B}^T$ is the $K \times N$ factor loading matrix of main asset class returns on the K factors. Moreover, $\boldsymbol{e}$ is the $T \times N$ matrix that contains the stochastic error terms orthogonal to the common factors.

We aim to estimate $\boldsymbol{F}$ and $\boldsymbol{B}$, which can only be identified up to an invertible transformation without additional assumptions. However, if we further impose restrictions that $\frac{1}{T}\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{I}_K$ and the K factors are mutually uncorrelated, i.e., $\boldsymbol{F}^T\boldsymbol{F}$ is diagonal, then $\boldsymbol{F}$ and $\boldsymbol{B}$ can be uniquely identified. In that case, we can estimate $\boldsymbol{F}$ and $\boldsymbol{B}$ by performing a principal component analysis on the sample covariance matrix of the standardized returns. Let $\widehat{\boldsymbol{\Sigma}} \equiv \frac{1}{T}\boldsymbol{R}_{std}^{T}\boldsymbol{R}_{std}$ be the sample covariance matrix of the standardized returns. The estimator $\widehat{\boldsymbol{B}}$ for $\boldsymbol{B}$ corresponds to the eigenvectors of the largest *K* eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ multiplied by $\sqrt{T}$, and the estimator $\widehat{\boldsymbol{F}}$ for $\boldsymbol{F}$ equals $\frac{1}{T}\boldsymbol{R}_{std}\widehat{\boldsymbol{B}}$.

Some studies, such as Pelger (2019), also encounter the problem of a rank-deficient sample covariance matrix as they focus on common factors of large cross-sections of individual stock returns. However, since we are only interested in returns of a handful of main asset indices such that $T>N$, we do not have that issue in this study.

2.3.1.2. Construction of the Chen-Roll-Ross (CRR) factors

We build the Chen-Roll-Ross factors following the procedure in Chen et al. (1986) but with two modifications. First, we directly construct innovation of expected and unexpected inflation using data of analyst consensus for year-on-year change in the CPI from Wind instead of adopting the MA procedure of Fama and Gibbons (1984). Second, following Avramov and Chordia (2006), for year-on-year change in industrial production, default spread, and term spread, we extract their innovations by the method of economic tracking portfolio instead of using the first differences of the variables as innovation proxies.

2.3.1.3. Economic tracking portfolios of macroeconomic state variables

We briefly illustrate the method of economic tracking portfolios (Lamont, 2001; Vassalou, 2003; Ang et al.,2006; Balduzzi & Robotti, 2008; Adrian et al., 2014; Detzel, 2017) used to extract “shock/news” about the future values of macroeconomic variables. For each of the three CRR-related variables MP, Defy, and Term, we run the following regression:

$$y_{k,t+h} = \boldsymbol{b}_k{}'\boldsymbol{R}_t + \boldsymbol{c}_k{}'\boldsymbol{Z}_{t-1} + \varepsilon_{k,t+h} \tag{2.2}$$

$y_{k,t+h}$ is the value of one of three variables mentioned above h months ahead. As retail investors dominate the Chinese stock market, we focus on tracking short-horizon news about the variables with h=1. $\boldsymbol{R}_t$ is a vector of excess returns on a set of base assets. Following Vassalou (2003), we include in $\boldsymbol{R}_t$ the excess returns of six value-weighted portfolios that are used to construct the modified Fama-French SMB and HML factors (Fama & French, 1993, 1995, 1996) in Section 1.4 of Chapter 1, namely S/V, S/M, S/G, B/V, B/M and B/G. Note that we modify the original version of those portfolios by replacing B/M with E/P in the sorting process, as in Liu et al. (2019). $\boldsymbol{Z}_{t-1}$ is the set of control state variables with values available at time t-1. In line with Chapter 1, we include in $\boldsymbol{Z}_{t-1}$ the yield spread between the 10-year government bond and the 1-year government bond (Term_spread1Y), the average default spread between 3-year AA corporate bonds and 3-year government bond (Default_spread), 1-month interbank repo rate (Dr1M), monthly market turnover (Turn_A) and monthly volatility on market return (Svar).

The estimated $\boldsymbol{b}_k{}'\boldsymbol{R}_t$ is the economic tracking portfolio of news about future values of the macroeconomic variables. We do not apply the method to the PCA or Fama-French factors as they already proxy for news about unobserved macroeconomic forces. We also skip the procedure on CPI_YoY in the CRR model as we can construct the time series of news about expected and unexpected inflation directly from analyst consensus data.

2.3.2. Test of model consistency with the ICAPM

Following Maio and Santa-Clara (2012) and Cooper and Maio (2019), we also check the consistency of the CRR factors with implications of the ICAPM by running the following predictive regression:

$$y_{t+h} = \mathrm{c} + \boldsymbol{\beta}'\boldsymbol{x_t} + \varepsilon_{t+h} \tag{2.3}$$

where y_{t+h} is either the return or volatility on the HS300 stock index in month t+h (h=1,2,3), $\boldsymbol{x_t}$ is a vector containing the levels of raw CRR variables at month t. $\boldsymbol{\beta}$ is the vector of coefficients on the forecasting variables whose significance and signs are of interest. We correct the standard errors of the estimator of $\boldsymbol{\beta}$ for serial correlation and heteroskedasticity with Newey and West's (1987a) standard errors.

The ICAPM implies that state variables that positively forecast future market returns or negatively forecast future market volatilities should have positive risk premiums. Those that negatively forecast future market returns or positively forecast future market volatilities should have negative risk premiums because assets with positive exposure on their shocks could hedge intertemporal risk in the investment opportunity set and thus tend to be traded at higher prices. We will compare the implications of ICAPM from (2.3) with the results of GMM and Fama-Macbeth tests of asset pricing models. The test does not apply to PCA and Fama-French factors as they are regarded as news of latent state variables whose levels cannot be backed out.

2.3.3. GMM (Hansen,1982) estimation of asset pricing models

We estimate the asset pricing models using the generalized method of moments (GMM) and the Fama-Macbeth (1972) procedure. In this subsection, we illustrate the details of the GMM procedure following Greene's (2011) notation.

The GMM aims at solving the following optimization problem:

$$\mathrm{Q}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{GMM}}) \equiv min_{\{\boldsymbol{\theta}\}}\, \overline{\boldsymbol{m}}(\boldsymbol{\theta})'\boldsymbol{W}\overline{\boldsymbol{m}}(\boldsymbol{\theta}) \tag{2.4}$$

where $\boldsymbol{\theta}$ is the $K \times 1$ vector of parameters, $\mathbf{W}$ is a positive definite weighting matrix, $\overline{\boldsymbol{m}}(\boldsymbol{\theta})$ is the time series average of $L \times 1$ vector function $\boldsymbol{m_t}(\boldsymbol{\theta})$ on $\boldsymbol{\theta}$, which satisfies population moment conditions $E[\boldsymbol{m_t}(\boldsymbol{\theta_0})] = 0$ at the true parameter $\boldsymbol{\theta_0}$. In this study, $\boldsymbol{m_t}(\boldsymbol{\theta})$ corresponds to the vector of pricing errors of the asset pricing models on each test asset. Assuming that $\{\boldsymbol{m_t}(\boldsymbol{\theta})\}$ is an ergodic stationary martingale difference series and:

$$\sqrt{\boldsymbol{T}}\overline{\boldsymbol{m}}(\boldsymbol{\theta_0}) \xrightarrow{\boldsymbol{d}} N[\mathbf{0}, \boldsymbol{\Phi}]$$

we have:

$$plim(\widehat{\boldsymbol{\theta}}_{\boldsymbol{GMM}})=\boldsymbol{\theta_0}$$

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GMM}} \sim^{a} N[\boldsymbol{\theta_0}, \boldsymbol{V_{GMM}}] \tag{2.5}$$

$$\boldsymbol{V_{GMM}} = \frac{1}{T}[\overline{\boldsymbol{G}}(\boldsymbol{\theta_0})'\mathbf{W}\overline{\boldsymbol{G}}(\boldsymbol{\theta_0})]^{-\mathbf{1}}\overline{\boldsymbol{G}}(\boldsymbol{\theta_0})'\boldsymbol{W}\boldsymbol{\Phi}\mathbf{W}\overline{\boldsymbol{G}}(\boldsymbol{\theta_0})[\overline{\boldsymbol{G}}(\boldsymbol{\theta_0})'\mathbf{W}\overline{\boldsymbol{G}}(\boldsymbol{\theta_0})]^{-\mathbf{1}} \tag{2.6}$$

where $\overline{\boldsymbol{G}}(\boldsymbol{\theta_0}) = \frac{\boldsymbol{\partial}\overline{\boldsymbol{m}}(\boldsymbol{\theta_0})}{\boldsymbol{\partial}\boldsymbol{\theta}'_{\mathbf{0}}}$. If we let $\boldsymbol{W} = \boldsymbol{\Phi}^{-\mathbf{1}}$, then we have the most efficient GMM estimator with asymptotic covariance matrix:

$$\boldsymbol{V_{GMM,optimal}} = \frac{1}{T}[\overline{\boldsymbol{G}}(\boldsymbol{\theta_0})'\boldsymbol{\Phi}^{-\mathbf{1}}\overline{\boldsymbol{G}}(\boldsymbol{\theta_0})]^{-\mathbf{1}} \tag{2.7}$$

There are various estimators for the long-run covariance matrix $\boldsymbol{\Phi}$**,** which differ from their regulatory assumptions. In this study, we employ the Newey-West kernel for baseline estimation.

To test for model misspecification, Hanen (1982) proposes the J test of overidentifying restrictions if $L > K$ and $\boldsymbol{W} = \boldsymbol{\Phi}^{-1}$:

$$J(\widehat{\boldsymbol{\theta}}_{GMM,,optimal}) \equiv T \times Q(\widehat{\boldsymbol{\theta}}_{GMM,,optimal})$$

$$\sim \boldsymbol{\chi}^2(L-K)$$

However, Newey (1985) notes that the J test suffers from significant Type II errors and could easily fail to detect model misspecification. Thus, to compare factors' performance, we also employ the ΔJ test of Newey and West (1987b) and Hansen and Jagannathan (1997) distance (HJ-distance).

The ΔJ statistics takes the form:

$$\Delta\mathrm{J} \equiv T \times Q(\widehat{\boldsymbol{\theta}}_{restricted}) - T \times Q(\widehat{\boldsymbol{\theta}}_{unrestricted})$$

$$\sim \boldsymbol{\chi}^2(number\ of\ restrictions)$$

The test aims to assess the redundancy of the restrictions, and in this study, the restrictions are exclusive restrictions on the Fama-French factors. The restricted models are models with the PCA or CRR factors only, and the unrestricted models are models with the PCA or CRR factors augmented with the Fama-French factors.

We can also compare the overall pricing errors of the models using the HJ-distance. Different models have different parameter vectors θ, moment conditions, and optimal weighting matrix $\boldsymbol{W} = \boldsymbol{\Phi}^{-1}$. Thus, comparing $\mathrm{Q}(\widehat{\boldsymbol{\theta}}_{GMM,,optimal})$ across models is problematic. Hansen and Jagannathan (1997) disentangle the problem by forcing a common weighting matrix on all models. Let $\boldsymbol{R}_t^e$ be the $L \times 1$ vector of excess returns for a given set of L assets, then the common weighting matrix for all asset pricing models is $\boldsymbol{W}_{HJ} = \mathrm{E}(\boldsymbol{R}_t^e \boldsymbol{R}_t^{e\prime})^{-1}$. The resulting

$$\text{HJ-distance} = \sqrt{\overline{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}}_{GMM,HJ})' \widehat{\boldsymbol{W}}_{HJ} \overline{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}}_{GMM,HJ})}$$

are comparable across models. Note that the choice of $\boldsymbol{W}_{HJ} = \mathrm{E}(\boldsymbol{R}_t^e \boldsymbol{R}_t^{e\prime})^{-1}$ is not discretionary. Hansen and Jagannathan (1997) deduce that the HJ-distance is an estimator of the shortest distance from the model-implied stochastic discount factor (SDF) proxy to the true SDF space.

We test the asset pricing models with two sets of sample moment conditions. The first set of moment conditions (henceforth denoted moment conditions 1) follows the SDF approach (Vassalou, 2003; Cochrane, 2005; Detzel, 2017).

Moment conditions 1: $$\bar{\boldsymbol{m}}(\boldsymbol{b}) = \frac{1}{T}\Sigma_{t=1}^{T}\left[\left(1-\boldsymbol{b}'\boldsymbol{f_t^d}\right)\boldsymbol{R_t^e}\right] = 0 \tag{2.8}$$

where $\boldsymbol{R_t^e}$ $L \times 1$ is the vector of excess returns on either one of the three sets of test assets, including 25 value-weighted portfolios sorted by size and E/P, 25 value-weighted portfolios sorted by size and idiosyncratic volatility, and 31 value-weighted industry portfolios. $\boldsymbol{f_t^d}$= $[f_{1,t}^d,\ldots, f_{K,t}^d]'$ is the $K \times 1$ vector of demeaned factors in the asset pricing models, and $\boldsymbol{b}$ is the loadings of the factors on the SDF. The significance of each element of $\boldsymbol{b}$ indicates whether the corresponding factor helps price the test assets in the presence of other factors. In other words, moment conditions 1 checks whether each factor provides unique pricing information (Cochrane, 2005; Feng et al., 2020).

The second set of moment conditions (henceforth denoted moment conditions 2) follows the expected return-covariance risk premium approach (Maio & Santa-Clara, 2012; Maio & Santa-Clara, 2017; Cooper & Maio, 2019).

Moment conditions 2: $$\bar{\boldsymbol{m}}(\boldsymbol{\lambda},\boldsymbol{\mu}) = \frac{1}{T}\Sigma_{t=1}^{T}\begin{cases} R_{1,t}^e - \Sigma_{k=1}^{K}\lambda_k R_{1,t}^e\left(f_{k,t}-\mu_k\right) \\ \vdots \\ R_{L,t}^e - \Sigma_{k=1}^{K}\lambda_k R_{L,t}^e\left(f_{k,t}-\mu_k\right) \\ f_{1,t}-\mu_1 \\ \vdots \\ f_{K,t}-\mu_K \end{cases} = 0 \tag{2.9}$$

where $\boldsymbol{\lambda}$= $[\lambda_1,\ldots, \lambda_{\mathrm{K}}]'$ is the $K \times 1$ vector of factor covariance risk premiums. $\boldsymbol{\mu}$= $[\mu_1,\ldots, \mu_{\mathrm{K}}]'$

is the $K \times 1$ vector of factor means.

2.3.4. Fama-Macbeth (1973) procedure and Shanken correction

We estimate the factor risk premiums using the Fama-Macbeth (1973) two-step regression. In the first step, we estimate the following time series regression for each test portfolio over the whole sample period:

$$R_{i,t} = c_i + \boldsymbol{\beta}_i{}'\boldsymbol{f_t} + \varepsilon_{i,t} \tag{2.10}$$

where $R_{i,t}$ is the excess return of test asset i at time t, and $\boldsymbol{f_t}$ is the vector of the time t realization of either PCA, CRR, or Fama-French factors. $\boldsymbol{\beta}_i$ is the vector of factor loadings of asset i on the testing factors. We are interested in the estimated values of the loading matrix $\boldsymbol{B}$= $[\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \dots, \widehat{\boldsymbol{\beta}}_N]$ (N is the number of test assets).

In the second step, we estimate the following cross-sectional regression at each point in time by GLS, as suggested by Lewellen et al. (2010):

$$R_{i,t} = a_t + \boldsymbol{\gamma_t}'\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}} + e_{i,t} \tag{2.11}$$

At each time t, we get a point estimate of the vector of the factor risk premiums $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}}$. We then estimate the average risk premium and its sampling covariance matrix by:

$$\widehat{\boldsymbol{\gamma}} = \frac{1}{T}\Sigma_{t=1}^{T}\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}} \tag{2.12}$$

$$var(\widehat{\boldsymbol{\gamma}}) = \frac{1}{T^2}\Sigma_{t=1}^{T}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}} - \widehat{\boldsymbol{\gamma}})(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{t}} - \widehat{\boldsymbol{\gamma}})' \tag{2.13}$$

We repeat the procedure on the three sets of test portfolios.

The estimated $\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}$ in (2.11) induces an errors-in-variables problem. Shanken (1992) proposes a method to correct the measurement error bias in the standard error of the average risk premium. Following Goyal (2012), we can express the error-corrected sampling variance of the average risk premium as:

$$var_{EIV}(\widehat{\boldsymbol{\gamma}}) = (1 + c)\left[var(\widehat{\boldsymbol{\gamma}}) - \frac{1}{T}\boldsymbol{\Sigma_f}\right] + \frac{1}{T}\boldsymbol{\Sigma_f} \tag{2.14}$$

where $\boldsymbol{\Sigma_f}$ is the covariance matrix of the factors, c=$\widehat{\boldsymbol{\gamma}}'\Sigma_f^{-1}\widehat{\boldsymbol{\gamma}}$, and $var(\widehat{\boldsymbol{\gamma}})$ is the uncorrected sample covariance matrix of $\hat{\boldsymbol{\gamma}}$. We can rearrange the terms on the right-hand side of (2.14) and get:

$$var_{EIV}(\hat{\boldsymbol{\gamma}}) = var(\hat{\boldsymbol{\gamma}}) + c\left[var(\hat{\boldsymbol{\gamma}}) - \frac{1}{T}\boldsymbol{\Sigma}_{\boldsymbol{f}}\right] \tag{2.15}$$

where $c\left[var(\hat{\boldsymbol{\gamma}}) - \frac{1}{T}\boldsymbol{\Sigma}_{\boldsymbol{f}}\right]$ is the Shanken correction term. Thus, in essence, the correction measures the difference between the factors and their risk premiums. If factors are excess returns, the correction term would be negligible. We calculate and report Shanken corrected t-statistics for all Fama-Macbeth (1973) procedures in this study.

2.4. Data and variables

The sample period of this study is from December 2006 to December 2021. Due to data availability of the main asset class returns, we are restricted from performing analyses over a more extended period. However, this sample period also has its merits. First, the pre-2000 period is often depicted by market participants and scholars alike as a market establishment stage with rampant accounting frauds and market manipulations, mainly due to the inexistence of a uniform set of accounting standards and lack of market supervision. Carpenter et al. (2021) show that only after 2004 did stock prices in the Chinese A-share market start to reflect news about future profits. Girardin and Joyeux (2013) also find evidence that the tie between Chinese stock returns and macroeconomic variables gradually strengthened after China joined the WTO in late 2001. Thus, conducting tests over our sample period alleviates the influence of many confounding factors unavoidable in some earlier studies, which have to cover the period prior to 2000 due to data availability. Second, our sample covers several interesting periods under the influence of significant economic/social events, such as the 2007-2008 Global Financial Crisis and the subsequent ¥ 4 trillion stimulus plan during 2008–2009, 2015-2016 Chinese stock market turbulence, and the global COVID-19 pandemic initiated at the end of 2019.

We construct the modified Fama-French five factors and double-sorted test portfolios following the same procedure in Section 1.4 of Chapter 1. We also use 31 value-weighted industry portfolios maintained by the Shenwan Hongyuan Group Co., Ltd as test assets. Our initial sample consists of all stocks listed in the Chinese A-share market. The data for individual stock returns and accounting information are from CSMAR. We exclude observations recorded within six months since the stock's IPO. We exclude stocks in the financial industry, and we further exclude stocks with special treatments (ST, *ST, and PT). Following Liu et al. (2019), we drop the smallest 30% of stocks from the sample. Such treatment would alleviate the reverse merger shell-value problem of small stocks caused by the IPO and delisting regulations for the Chinese stock market. Liu et al. (2019) found that 83% of the reverse mergers are performed over listed "shells" in the bottom 30% size category. The data for the nine main asset class returns and consensus expected inflation are from Wind.

2.5. Empirical results

This section presents the descriptive statistics of different sets of factors and the results from the GMM and Fama-Macbeth regressions of competing models. We show results of exploratory data analysis and tests for the consistency of macroeconomic factor models with the ICAPM. We then test the significance of the macroeconomic and Fama-French factors using the GMM and Fama-Macbeth (1973) procedure. We also examine whether the Fama-French factors provide additional pricing information in the presence of macroeconomic factors using ΔJ test. Lastly, we assess the relative performance of the models by the HJ-distance and average pricing error. We will show that macroeconomic risks are priced in the Chinese stock market. The Fama-French factors lose a considerable portion of their ability to explain the cross-section of portfolio returns when macroeconomic factors are added to the models, indicating that much of the pricing information contained in the Fama-French factors is news related to the macroeconomy.

2.5.1. Exploratory data analysis and descriptive statistics

This subsection presents the results of exploratory data analysis and descriptive statistics of different sets of factors. We show factor loadings of the PCA factors on returns of nine main asset classes and probe the nexus between the PCA factors and macroeconomic variables. We also examine the correlations between the Fama-French and macroeconomic factors.

2.5.1.1. The Principal Component Analysis (PCA) factors

We build the PCA macroeconomic factors by performing principal component analysis on demeaned monthly returns of nine main asset class indices. The nine main asset class returns are chosen from three broad categories: three stock indices returns, including the monthly return of the CSI 300 Stock Index (HS300), the monthly return of the CSI 500 Stock Index (HS500), and the monthly return of the CSI 1000 Stock Index (HS1000); two bond indices returns including the monthly return of the CSI Government Bond Index (Bond_G) and the monthly return of the CSI Corporate Bond Index (Bond_C); and four commodities indices returns including the monthly return of the Au9999 Index traded at the Shanghai Futures Exchange (Gold), the monthly return of the Nanhua Industrial Commodity Index (Industry), the monthly return of the Nanhua Metal Commodity Index (Metal) and the monthly return of the Nanhua Energy Commodity Index (Energy). The indices are chosen based on their representativeness and heterogeneity in their responses to macroeconomic news. The number of assets in each category is also set to allow all categories to have a relatively balanced influence on the results of the principal component analysis.

Table 2.1 shows the correlations among returns of the nine asset indices and some macroeconomic variables. The three stock indices are highly correlated with correlation coefficients well above 0.70, although they comprise disjoint cohorts of stocks falling in different size categories. The commodities indices show a similar pattern except for Gold, which has relatively low correlations with assets within the commodities category and assets in other categories, indicating that fluctuations in gold price capture different sources of macroeconomic risks from other assets. Contrary to the case of stock and commodities, the two assets in the bond category only have a correlation of 0.278, much lower than the average level of pairwise correlations among assets within the other two categories. Also, stock indices have comparatively high positive correlations with growth variables (ID_YoY and PMI_Manu) and long-run rates. Bond indices have comparatively large negative correlations with growth and inflation variables and positive correlations with the U.S. Dollar Index. Gold is primarily related to short-term rates and the U.S. Dollar Index, indicating it is most sensitive to liquidity news. The other commodities correlate positively with growth variables and negatively with inflation, short-term rate, and the U.S. Dollar Index. Those results provide evidence that the selected assets respond to various sources of macroeconomic news differently, and it is eligible to extract those news sources using the principal component analysis.

Figure 2.1 shows the factor loadings of the largest five principal components of the nine monthly returns series. The first principal component has large positive loadings on stocks and commodities and negative loadings on bonds. Thus, we denote it as the growth factor. The component explains almost half of the total variance of the nine assets and has an explained variance ratio of 46.92%. The second principal component has large negative loadings on stocks and bonds and positive loadings on commodities except for Gold. Thus, we denote it as the stagflation factor. The component explains 22.75% of the total variance of the nine assets and has a cumulative explained variance ratio of 69.68%. The third principal component has high positive loadings on bonds and negative loadings on stocks. Thus, we denote it as the long-term rate factor. The component explains 16.27% of the total variance of the nine assets and has a cumulative explained variance ratio of 85.94%. The fourth principal component has large negative loading on Gold and small loadings on all other assets. Thus, we denote it as the liquidity factor. The component explains 10.76% of the total variance of the nine assets and has a cumulative explained variance ratio of 96.70%. The fifth principal component has a significant positive loading on Energy and negative loading on Metal. As energy commodities are more sensitive to economic growth in developed countries while metal commodities are more sensitive to economic growth in developing countries, we denote PC5 as the growth spread factor. The component explains 22.75% of the total variance of the nine assets and has a cumulative explained variance ratio of 99.49%. Our results are similar to those of Raol (2017).

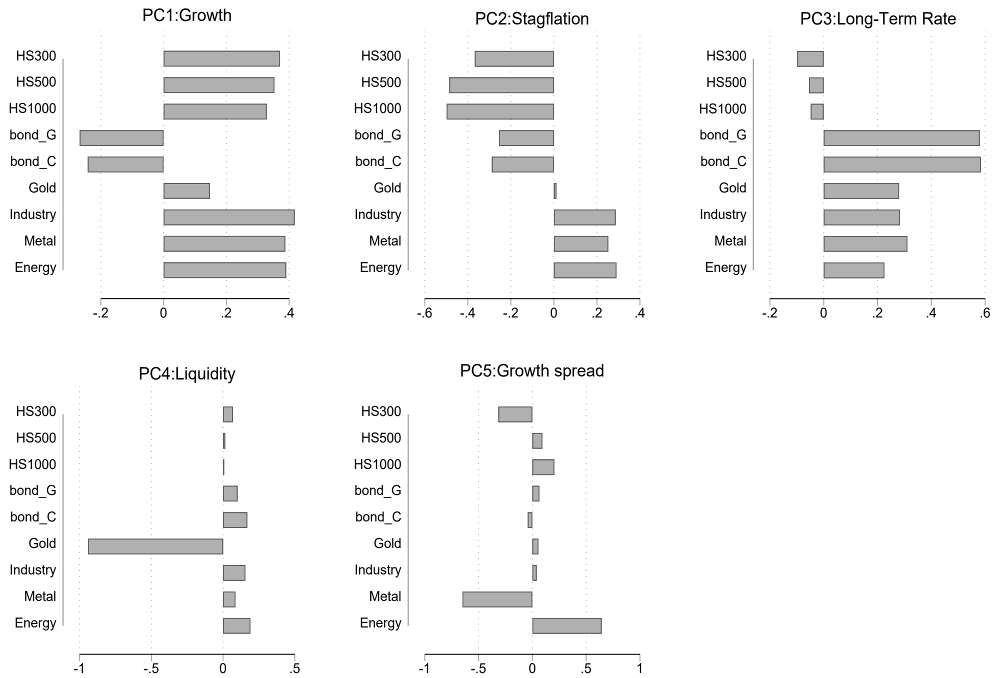
Also, note that the principal components are only identified up to a flip of the sign. In other words, suppose PCi is the i^{th} principal component of the asset returns; then $(-1) \times$ PCi accounts for the same proportion of total variance explained and thus is precisely the same principal component as PCi. We adjust the signs of the resulting principal components to make the factors more interpretable. We expect PC1, the growth factor, to have a positive risk premium in the cross-section of stock returns. In contrast, all other factors should have negative risk premiums as they each indicate a "bad state" of the economy.

Table 2.1. Correlations between returns of main asset classes and macroeconomic variables

This table reports the correlations between returns of main asset classes and macroeconomic variables. HS300 is the monthly return of the CSI 300 Stock Index. HS500 is the monthly return of the CSI 500 Stock Index. HS1000 is the monthly return of the CSI 1000 Stock Index. Bond_G is the monthly return of the CSI Government Bond Index. Bond_C is the monthly return of the CSI Corporate Bond Index. Gold is the monthly return of the Au9999 Index traded at the Shanghai Futures Exchange. Industry is the monthly return of the Nanhua Industrial Commodity Index. Metal is the monthly return of the Nanhua Metal Commodity Index. Energy is the monthly return of the Nanhua Energy Commodity Index. ID_YoY is the year-on-year change in Industrial Production. PMI_Manu is the Purchasing Managers' Index for manufacturing industries. PPI_YoY is the year-on-year change in the Retail Price Index. Yield_10Y is the yield on the 10-year government bond. Dr1M is the 1-month interbank repo rate. USDX_YoY is the year-on-year change in the U.S. Dollar Index. The sample period is from 2006:M12 to 2021:M12.

	HS300	HS500	HS1000	Bond_G	Bond_C	Gold	Industry	Metal	Energy	ID_YoY	PMI_Manu	PPI_YoY	Yield_10Y	Dr1M
HS300	1													
HS500	0.841	1												
HS1000	0.777	0.973	1											
Bond_G	-0.083	-0.073	-0.065	1										
Bond_C	-0.005	0.012	0.016	0.278	1									
Gold	0.073	0.064	0.065	0.050	0.028	1								
Industry	0.347	0.295	0.261	-0.120	-0.029	0.240	1							
Metal	0.328	0.278	0.247	-0.113	-0.036	0.230	0.901	1						
Energy	0.313	0.266	0.237	-0.111	-0.025	0.213	0.919	0.693	1					
ID_YoY	0.445	0.273	0.302	-0.342	-0.452	0.373	0.142	0.050	0.164	1				
PMI_Manu	0.470	0.415	0.399	-0.442	-0.463	0.291	0.350	0.330	0.300	0.616	1			
PPI_YoY	0.014	-0.102	-0.148	-0.474	-0.620	0.229	0.585	0.447	0.622	0.249	0.187	1		
Yield_10Y	0.275	0.258	0.258	-0.533	-0.504	-0.157	-0.136	-0.134	-0.095	0.550	0.321	0.215	1	
Dr1M	-0.021	0.003	0.005	-0.415	-0.307	-0.432	-0.197	-0.158	-0.139	-0.047	-0.082	0.077	0.627	1
USDX_YoY	-0.095	-0.073	-0.055	0.702	0.759	-0.391	-0.476	-0.491	-0.404	-0.387	-0.350	-0.507	-0.354	-0.123

Figure 2.1. Loadings of principal components on main asset class returns
This figure shows the factor loadings of the largest five principal components on nine main asset class returns. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. The sample period is from 2006:M12 to 2021:M12.



We plot the time series of the factors' monthly one-year cumulative values against corresponding macroeconomic variables to further inspect the nexus between PCA factors and macroeconomic variables.

Figure 2.2 shows the correlation between PC1 and growth variables. PMI_Manu-50 equals domestic manufacturing PMI subtracted by 50. The Keqiang Index is an indicator of aggregate economic activity advocated by Li Keqiang, former prime minister of China. The index is often regarded as a better indicator of economic growth than official GDP numbers. The Keqiang Index = 40%×year-on-year growth of industrial electricity usage + 35%×year-on-year growth of mid/long-term loans + 25%×year-on-year growth of railway freight volume. Figure 2.3 shows the correlation between PC2 and inflation variables. PPI_YoY is the year-on-year change in the Producer Price Index. PPI-CPI is the difference between the

Producer and Consumer Price Index. We gauge the general level of inflation in the economy not only by the price levels in the end markets but also in intermediary markets. The figure indicates that most of the variation in inflation comes from the PPI. Figure 2.4 shows evidence of a high correlation between PC3 and long-term interest rates. Yield_5Y is the yield to maturity of the 5-year government bond. Yield_10Y is the yield to maturity of the 10-year government bond. Figure 2.5 shows the correlation between PC4 and short-term liquidity variables. Since the original time series of the variables are of different orders of magnitude, we scale them by their historical volatilities. USDX_std is the scaled year-on-year change in the U.S. Dollar Index. DR1M_std is the scaled 1-Month Interbank Repo Rate. Figure 2.6 shows the correlation between PC5 and the relative economic performance between the U.S. and China. Similar to the case of Figure 2.5, we scale the variables to inspect the time series in the same order of magnitude. PMI_CN_std×(-1) is the negative of the Chinese Purchasing Managers' Index scaled by its historical volatility. CEAI_US_YoY is the year-on-year change in the U.S. Economic Activity Coincident Index by FRED. The results coincide with the pattern of the factor loadings of PC5: The component has large positive loadings on energy commodities and large negative loadings on metal commodities, indicating that it should have high realized values when the U.S. is expected to have relatively stronger economic performance than China and vice versa. The figures present evidence of a nexus between the PCA and macroeconomic variables and that the statistical factors are not constructed on a vague basis.

Figure 2.2. PC1 against growth variables
This figure plots the first principal component of main asset class returns against growth variables. PMI_Manu-50 equals domestic manufacturing PMI subtracted by 50. Keqiang Index = 40%×year-on-year growth of industrial electricity usage + 35%×year-on-year growth of mid/long-term loan + 25%×year-on-year growth of railway freight volume.

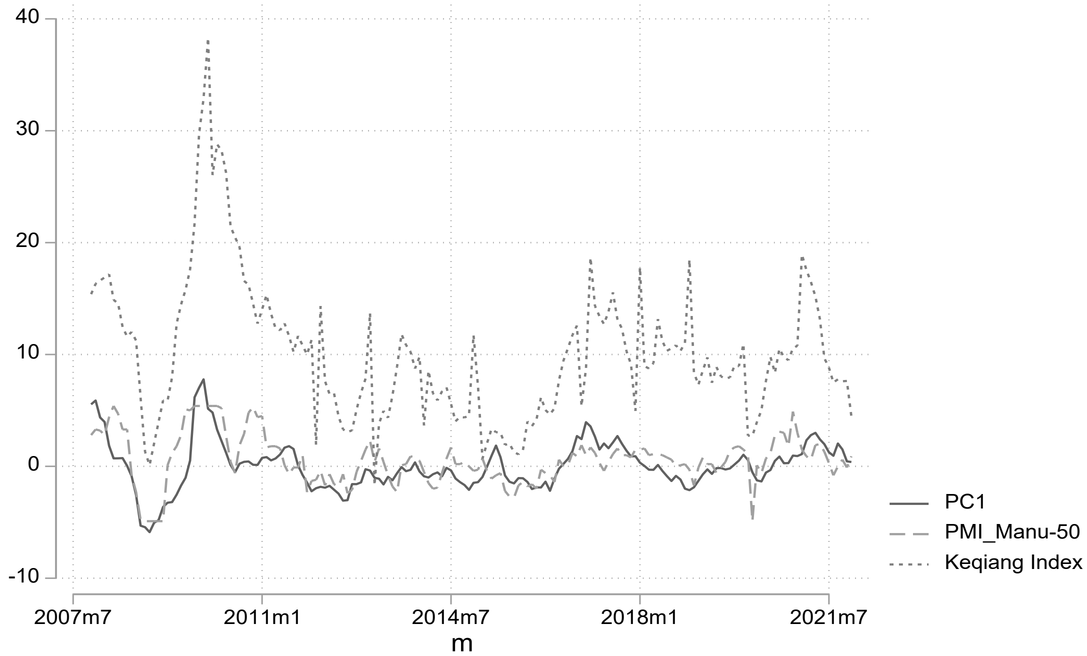


Figure 2.3. PC2 against inflation variables
This figure plots the second principal component of main asset class returns against inflation variables. PPI_YoY is the year-on-year change in the Producer Price Index. PPI-CPI is the difference between the Producer and Consumer Price Index.

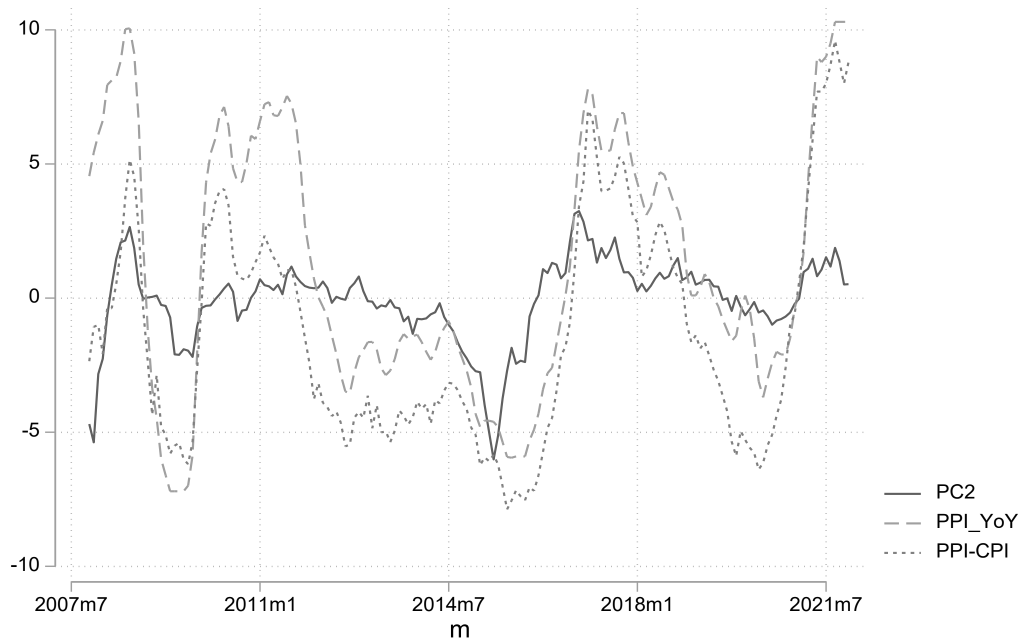


Figure 2.4. PC3 against long-term interest rates

This figure plots the negative of the third principal component of main asset class returns against long-term interest rate variables. Yield_5Y is the yield to maturity of the 5-year government bond, and Yield_10Y is the yield to maturity of the 10-year government bond.

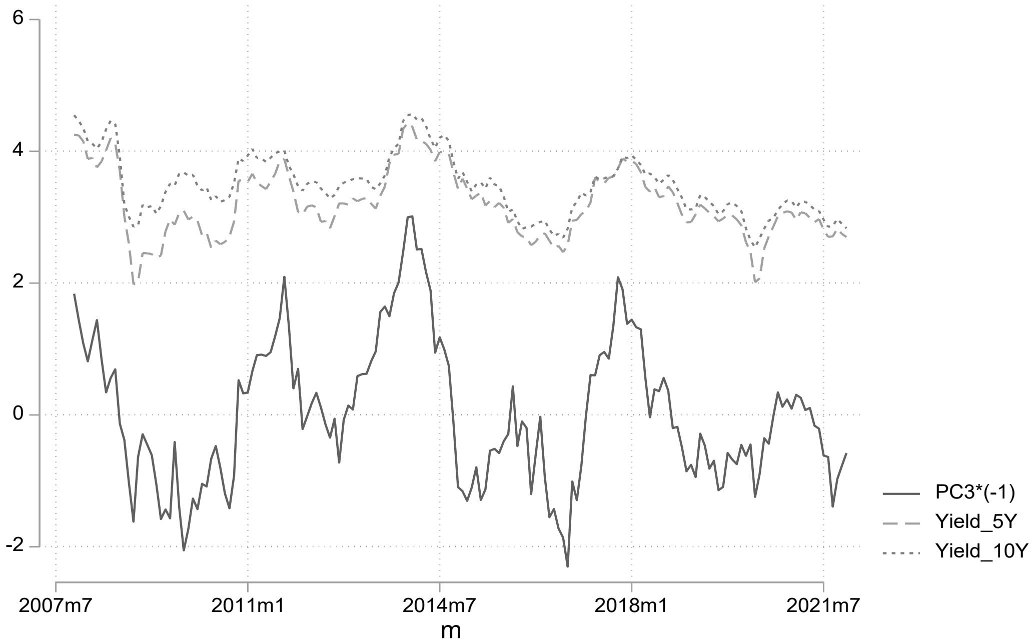


Figure 2.5. PC4 against short-term liquidity variables

This figure plots the fourth principal component of main asset class returns against short-term liquidity variables. All three variables are scaled by their historical volatilities. USDX_std is the scaled year-on-year change in the U.S. Dollar Index. DR1M_std is the scaled 1-Month Interbank Repo Rate.

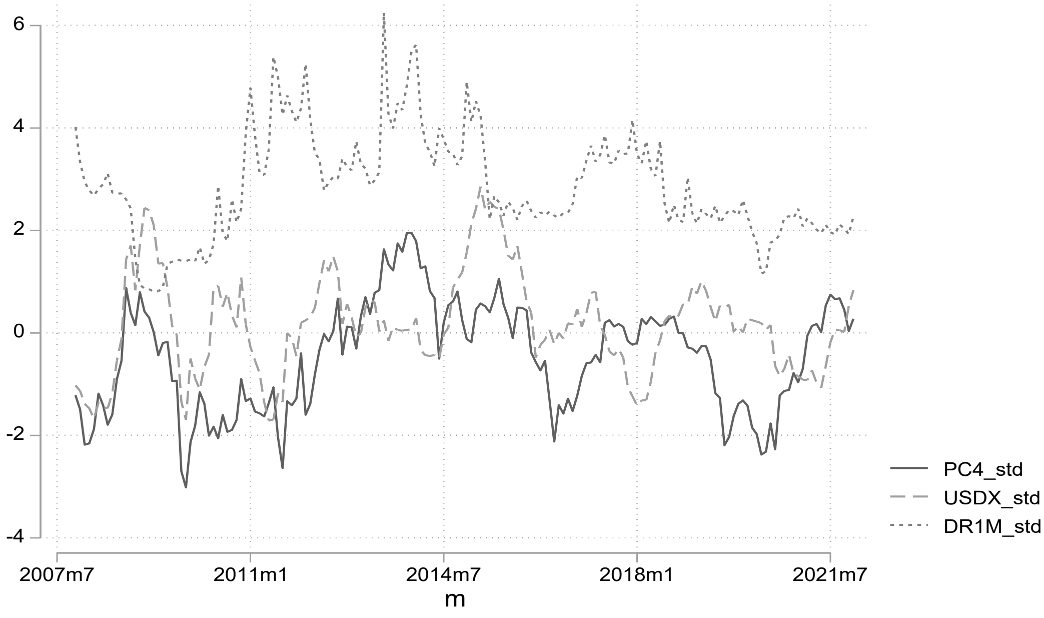


Figure 2.6. PC5 against relative economic performance between the U.S. and China
This figure plots the fifth principal component of main asset class returns scaled by its historical volatility against the relative economic performance between the U.S. and China. PMI_CN_std*(-1) is the negative of the Chinese Purchasing Managers' Index scaled by its historical volatility. CEAI_US_YoY is the year-on-year change in the U.S. Economic Activity Coincident Index by FRED.

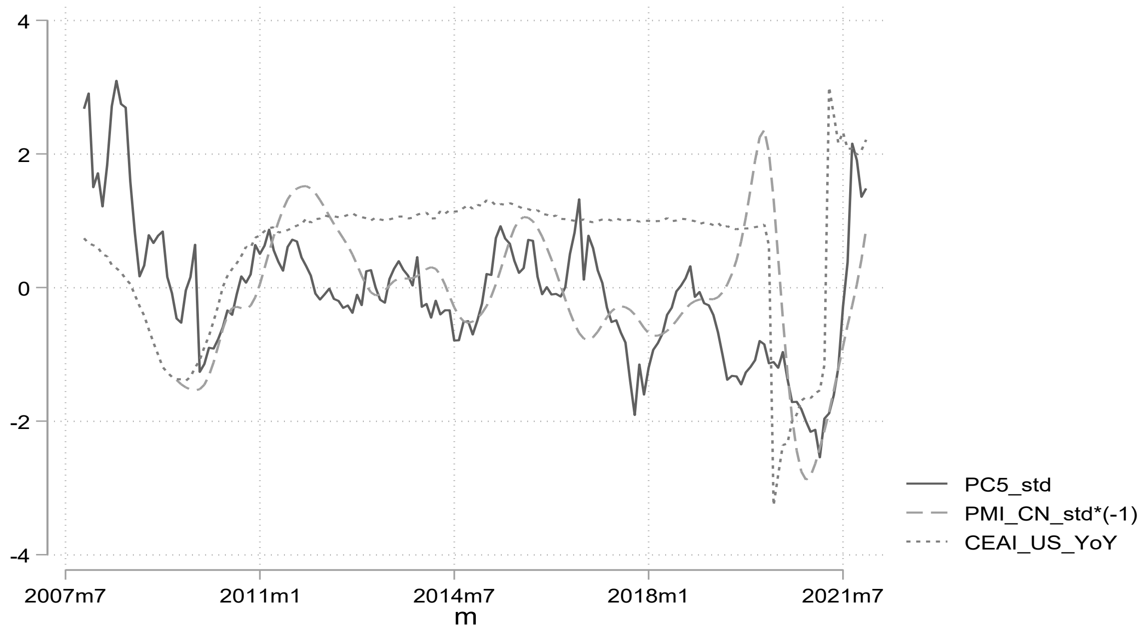


Table 2.2 presents the correlations between the PCA and Fama-French factors. MKT has a significant positive correlation with PC1 and a large negative correlation with PC2, indicating that the market factor contains mainly information related to aggregate economic growth and inflation. SMB is significantly correlated with PC2 and PC5, with correlation coefficients of -0.423 and -0.293, respectively, indicating that the size factor contains mainly information related to inflation and relative economic performance between China and developed economies. VMG and RMW have significant positive correlations with PC2 and PC5 and significant negative correlations with PC1, suggesting that value and profitability factors may contain similar information related to growth, inflation, and relative economic performance between China and developed economies. CMA is the only Fama-French factor that has a significant correlation with PC3. Contrary to our expectation, PC4 is not significantly correlated with any Fama-French factor.

The results also suggest that correlations among the Fama-French factors are caused by their correlations with macroeconomic factors. For example, VMG and RMW have comparable correlations in sign and magnitude with the five PCA factors, and the correlation coefficient between them is as high as 0.518.

Also note that there is a strong negative correlation between size (SMB) and value (VMG) factors with a correlation coefficient of -0.565, which coincides with the results of Liu et al. (2019). The negative correlation is probably because SMB and VMG have significant and comparable correlations in magnitude with PC2 and PC5, but in different directions. Overall, the results coincide with the conjecture of Connor and Korajczyk (2009) that statistical factors, macroeconomic factors, and fundamental factors are theoretically consistent and differ only by rotation. We will formally test the relationship between the PCA and Fama-French factors in Section 2.5.5 by GMM.

Table 2.2. Correlations among the PCA and Fama-French factors

This table reports the correlations between the PCA and Fama-French factors. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The sample period is from 2006:M12 to 2021:M12.

	PC1	PC2	PC3	PC4	PC5	MKT	SMB	VMG	RMW	CMA
MKT	0.764	-0.610	-0.074	-0.065	0.092	1				
SMB	0.108	-0.423	0.077	0.012	-0.293	0.159	1			
VMG	-0.150	0.394	-0.036	-0.122	0.232	-0.191	-0.565	1		
RMW	-0.239	0.375	-0.128	-0.001	0.166	-0.281	-0.390	0.518	1	
CMA	-0.025	0.029	0.324	-0.044	-0.102	-0.066	0.138	0.039	-0.540	1

2.5.1.2. The Chen-Roll-Ross (CRR) factors

Table 2.3 shows the correlations between the CRR and Fama-French factors. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios for the default and term spread respectively. Panel A of Table 2.3 shows the correlations among the CRR factors. Overall, the factors exhibit low pairwise correlations, except that growth news (MP) is significantly positively correlated with the term spread news and negatively correlated with the default spread news, which is consistent with our expectation.

Panel B shows the correlations between the CRR and Fama-French factors. MP is significantly negatively correlated with the value and profitability factors. Innovations of the expected and unexpected inflations show insignificant correlations with the Fama-French factors. The term spread news is positively correlated with the size factor and negatively correlated with the value and profitability factor. Fama and French (1992) argue that value firms tend to have more market-imposed leverage, which would benefit more from cuts in short-term rates (hence broader term spread), inducing a positive correlation between the term spread and value factor. Cornell (1999) and Campbell and Vuolteenaho (2004) reach the same conclusion from another angle. They argue that growth firms are high-duration assets like long-term bonds, and value firms have comparatively low duration and behave like short-term bonds. Thus, a widened term spread would lead to superior performance of the value stocks over growth stocks. Fama and French (1992) and Hahn and Lee (2006) provide further evidence of the positive correlation between the term spread and value factor. However, out of our expectations, our result differs from those in previous studies.

Table 2.3. Correlations among the CRR and Fama-French factors

This table presents the correlations among the CRR and Fama-French factors. Panel A shows the correlations among the CRR factors, and Panel B shows the correlations between the CRR and Fama-French factors. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread respectively. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The sample period is from 2006:M12 to 2021:M12.

Panel A: Correlations among CRR factors					
	MP	UI	DEI	Defy	Term
MP	1.000				
UI	0.002	1.000			
DEI	-0.092	0.096	1.000		
Defy	-0.450	-0.028	0.089	1.000	
Term	0.309	-0.023	-0.023	-0.058	1.000

Panel B: Correlations between CRR factors and Fama-French factors					
	MP	UI	DEI	Defy	Term
MKT	0.073	0.047	-0.024	-0.344	0.104
SMB	0.088	-0.039	0.078	-0353	0.265
VMG	-0.573	-0.009	0.042	0.250	-0.353
RMW	-0.327	-0.034	0.079	0.079	-0.270
CMA	-0.112	-0.061	0.057	0.316	-0.088

2.5.2. Predictive regressions of macroeconomic state variables

Following Maio and Santa-Clara (2012), we run multiple horizon predictive regressions of monthly market return and volatility on state variables corresponding to the CRR factors. The ICAPM postulates that shocks of state variables that positively forecast future market returns or negatively forecast future market volatiles should have a positive risk premium. And shocks of state variables that negatively forecast future market returns or positively forecast future market volatiles should have a negative risk premium. We will compare the results in this subsection with those of the asset pricing tests in the subsequent sections and assess the consistency of the CRR model with the ICAPM. The assessment does not apply to the PCA and Fama-French factors as we cannot neatly back out their underlying state variables.

Table 2.4 reports the estimation results. ID_YoY is the year-on-year change in the industrial production. CPI_YoY is the year-on-year change in the Consumer Price Index. Defy* is the default spread. Term* is the term spread. ID_YoY significantly positively forecasts market volatility at a 1-month horizon. CPI_YoY significantly negatively forecasts market return at horizons of 1,2, and 3 months ahead. Defy* significantly positively forecasts market volatility at horizons of 1,2 and 3 months ahead. Term* significantly negatively forecasts market return one month ahead and positively forecasts market volatility at horizons of 1 and 2 months ahead. Thus, according to the ICAPM, all those factors should have negative risk premiums. The result is inconsistent with our expectation as increases in ID_YoY and Term* are generally considered to signal "good states" of the economy by various studies and thus, should demand positive risk premiums. We also run univariate regressions using ID_YoY or Term* as predictors. ID_YoY does not show significant predictive power in either return or volatility univariate regressions. Term* has a positive but only marginally significant coefficient in 1-month-ahead volatility predictive regression. The results are consistent with those of the multivariate regressions.

Overall, future market volatilities are more predictable than returns. The volatility regressions generally have larger R^2s and adjusted R^2s than return regressions with the same forecasting horizons. Also, note that the predictive power decreases with the forecasting horizon. This phenomenon contrasts with previous studies on the U.S. market, such as the work of Maio and Santa-Clara (2012), who observe increased predictability as the forecasting horizon expands. The phenomenon persists probably because the Chinese stock market is dominated by retail investors who are more myopic (Cuthbertson et al., 1997; Docherty & Hurst, 2018; Liu et al., 2019).

Table 2.4. Predictive regressions of the CRR state variables

This table presents the results of multiple horizon predictive regressions for the monthly return and volatility of the value-weighted market index. The forecasting variables are current levels of state variables corresponding to the CRR factors. Columns 1, 2, and 3 show results of regressions for the market return at forecasting horizons of 1, 2, and 3 months ahead, respectively. Columns 4, 5, and 6 show results of regressions for the market volatility at forecasting horizons of 1, 2, and 3 months ahead, respectively. ID_YoY is the year-on-year change in industrial production. CPI_YoY is the year-on-year change in the Consumer Price Index. Defy* is the default spread. Term* is the term spread. The t-statistics corrected for serial correlation and heteroskedasticity using the Newey and West (1987a) estimator are in parentheses. The sample period is from 2007:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	F.Rm	F2.Rm	F3.Rm	F.Vol	F2.Vol	F3.Vol
ID_YoY	-0.266	-0.309	-0.139	3.635**	2.655	-0.746
	(-1.21)	(-1.43)	(-0.77)	(2.25)	(1.52)	(-0.45)
CPI_YoY	-0.825***	-0.613**	-0.409**	3.374	2.331	0.951
	(-3.31)	(-2.42)	(-2.11)	(1.18)	(0.71)	(0.54)
Defy*	0.502	0.502	0.644	21.985**	22.367**	23.120**
	(0.36)	(0.35)	(0.44)	(2.33)	(2.28)	(2.21)
Term*	-1.025**	-0.795	-0.269	38.452***	33.631***	4.338
	(-2.02)	(-1.46)	(-0.29)	(4.03)	(2.78)	(0.60)
_cons	3.593*	2.712	0.819	4.108	6.652	16.584
	(1.83)	(1.34)	(0.40)	(0.31)	(0.48)	(1.15)
R^2	0.051	0.039	0.025	0.118	0.090	0.071
adj. R^2	0.024	0.012	-0.003	0.095	0.067	0.047

2.5.3. Estimation of asset pricing models by GMM using moment conditions 1

This subsection presents the estimation results of different asset pricing models by GMM using moment conditions 1 in Section 2.3.3. The test assets include 25 value-weighted double-sorted portfolios on size and E/P, 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility, and 31 value-weighted industry portfolios. We estimate each model using optimal GMM with the Newy-West kernel. We also include the result of Hansen's (1982) test of overidentifying restrictions for each model specification. In contrast with tests of expected return-beta models, tests with the stochastic discount factors (SDF) aim to detect the factors' incremental pricing power. Cochrane (2005) notes that tests with moment conditions 1 are equivalent to running a multiple regression of the SDF on the factors. The significance of the estimated coefficient $\boldsymbol{b}$ in (2.8) tells us whether each factor included helps price the test assets in the presence of the other factors. The corresponding risk premiums, however, are proportional to the regression coefficients of simple regressions of the SDF on each factor. Thus, there is no simple mapping from the sign and significance of each element of $\boldsymbol{b}$ to those of the corresponding risk premium. As long as the factors are correlated, a factor may provide significant incremental pricing information while earning zero risk premium and vice versa (Feng et al., 2020). In summary, tests on moment conditions 1 check whether each factor is spurious, whereas tests on risk premiums check whether each factor is priced.

2.5.3.1. Tests on 25 Size-E/P portfolios by GMM using moment conditions 1

Our first set of test assets consists of 25 portfolios formed following a similar procedure to Fama and French (1993). A slight deviation is to replace the book-to-market ratio with the earnings-to-price ratio in the procedure. Liu et al. (2019) show that the E/P ratio is a superior measure of relative value in the Chinese stock market compared with the traditional B/M ratio. Thus, we test the competing asset pricing models on 25 Size-E/P portfolios instead of the conventional Fama-French 25 Size-B/M portfolios.

Table 2.5 presents the estimation results of the Fama-French, PCA, and CRR factors. Panel A shows the estimated coefficients of the Fama-French five factors. The results show that MKT and CMA do not provide incremental pricing information for the 25 Size- E/P portfolios when the other three factors are included. Panel B shows the estimated coefficients of the PCA factors. The results show that all principal components have significant loadings on the SDF at 1% significance level, indicating that no PCA factor is redundant when pricing the 25 Size- E/P portfolios. The results are consistent with our expectation as the PCA factors are mutually orthogonal by construction. Panel C shows the estimated coefficients of the PCA factors augmented with the market excess return. Note that the significance of the PCA factors drops as MKT is included. PC1 is only significant at the 10% level, and PC2 even becomes insignificant. The results suggest that PC2 and PC1 contain similar information to the market returns, but PC1 still provides some unique information as MKT itself is only significant at the 10% level. Panel D shows the estimated coefficients of the CRR factors. The results show that all factors provide incremental pricing information except for the innovation in expected inflation (DEI). Panel E shows estimated coefficients of the CRR factors augmented with the market excess return. Including MKT does not make a qualitative difference, with an insignificant coefficient on MKT.

Table 2.5 Tests of asset pricing models on 25 Size-E/P portfolios

This table presents the estimated coefficients of the Fama-French, PCA, and CRR factors by GMM using moment conditions 1 in Section 2.3.3. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread respectively. The estimated parameters are factor loadings on the SDF. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

Panel A: Fama-French 5 factors					
	MKT	SMB	VMG	RMW	CMA
Estimate	0.033	0.202***	0.140***	0.077**	0.015
z-stat	1.43	11.75	4.35	2.36	0.50
J-stat	7.915				
p-value	0.992				

Panel B: PCA factors						
	PC1	PC2	PC3	PC4	PC5	MKT
Estimate	0.083***	-0.142***	-0.270***	-0.225***	-0.223***	-
z-stat	3.65	-5.71	-3.55	-4.86	-2.72	-
J-stat	6.805					
p-value	0.997					

Panel C: PCA factors +MKT						
	PC1	PC2	PC3	PC4	PC5	MKT
Estimate	0.038*	0.112	-0.318**	0.107**	-0.302***	0.170*
z-stat	1.83	0.19	-2.23	1.99	-3.53	1.74
J-stat	7.044					
p-value	0.994					

Panel D: CRR factors						
	MP	UI	DEI	Defy	Term	MKT
Estimate	0.059**	0.101***	-0.063	-0.046***	-0.124***	-
z-stat	2.19	7.58	-0.82	-2.98	-2.73	-
J-stat	8.768					
p-value	0.985					

Table 2.5 Tests of asset pricing models on 25 Size-E/P portfolios *(continued)*

Panel E: CRR factors +MKT						
	MP	UI	DEI	Defy	Term	MKT
Estimate	0.046**	0.096***	0.008	-0.031***	-0.152***	0.063
z-stat	2.07	5.66	0.24	-2.86	-2.69	1.06
J-stat	8.763					
p-value	0.977					

2.5.3.2. Tests on 25 Size-IVOL portfolios by GMM using moment conditions 1

Next, we repeat the tests on 25 Size-IVOL portfolios. Lewellen et al. (2010) criticize that the Fama-French 25 Size-B/M portfolios have a strong factor structure, thus inducing a low hurdle for asset pricing tests that use them as the only set of test assets. One of their suggestions is to include portfolios sorted by other characteristics as test assets. In this case, portfolios sorted by the idiosyncratic volatility are suitable candidates. The idiosyncratic volatility puzzle is among the most prominent anomalies in the Chinese stock market (Nartea et al., 2013; Gu et al., 2018). Many studies try to disentangle the puzzle from the angle of market microstructure and investor behavior (see, for example, Brandt et al., 2010; Stambaugh et al., 2015; Hou & Loh, 2016), while Ang et al. (2009) and Chen et al. (2012) suggest that the anomaly signals the existence of omitted factors that represent neglected dimensions of systematic risk. Thus, we are interested in whether the macroeconomic factors could help price the size-neutral portfolios sorted by the idiosyncratic volatility.

Table 2.6 presents the estimation results for all three sets of factors. Panel A shows the estimated coefficients on the Fama-French five factors. Like Panel A of Table 2.5, MKT and CMA do not provide incremental pricing information for the 25 Size-IVOL portfolios when the other three factors are included. Panel B shows the estimated coefficients of the PCA factors. Contrary to Panel B of Table 2.5, only PC3 and PC5 are significant at the 5% level. Panel C shows the estimated coefficients of the PCA factors augmented with the market excess return. Including MKT does not change the coefficient estimates of the PCA factors qualitatively. MKT itself is insignificant at the 5% level. Panel D shows the estimated coefficients of the CRR factors. Out of our expectations, the results show that all factors provide incremental pricing information, even for DEI, which has an insignificant coefficient in Panel D of Table 2.5. Panel E shows the estimated coefficients of the CRR factors augmented with the market excess return. Once again, including MKT does not change the coefficient estimates on the CRR factors qualitatively.

Overall, compared with the results over the 25 Size-E/P portfolios, Table 2.6 records a reduction in the incremental pricing ability of several factors, including PC1, PC2, and PC4. The pricing ability of DEI is enhanced. It seems that except for the CRR factors, the other two sets of factors do not demonstrate better performance when pricing the 25 Size-IVOL portfolios. Similar to results over the 25 Size-E/P portfolios, Table 2.6 also documents insignificant J-statistics for all model specifications.

Table 2.6. Tests of asset pricing models on 25 Size-IVOL portfolios
This table presents the estimated coefficients of the Fama-French, PCA, and CRR factors by GMM using moment conditions 1 in Section 2.3.3. The test assets are 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread respectively. The estimated parameters are factor loadings on the SDF. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

Panel A: Fama-French 5 factors					
	MKT	SMB	VMG	RMW	CMA
Estimate	0.017	0.328***	0.505***	-0.181**	0.103
z-stat	1.38	7.48	6.12	-2.04	1.49
J-stat	11.821				
p-value	0.922				

Table 2.6. Tests of asset pricing models on 25 Size-IVOL portfolios *(continued)*

Panel B: PCA factors						
	PC1	PC2	PC3	PC4	PC5	MKT
Estimate	0.088*	-0.062*	0.262***	-0.088	-0.439***	-
z-stat	1.67	-1.75	7.17	-1.25	-3.41	-
J-stat	13.615					
p-value	0.849					

Panel C: PCA factors +MKT						
	PC1	PC2	PC3	PC4	PC5	MKT
Estimate	-0.107*	0.101	0.107**	-0.065	-0.225***	0.133*
z-stat	-1.74	1.58	2.05	-0.98	-4.55	1.78
J-stat	9.035					
p-value	0.972					

Panel D: CRR factors						
	MP	UI	DEI	Defy	Term	MKT
Estimate	-0.085**	0.101***	-0.028***	-0.023***	0.047***	-
z-stat	-2.51	5.39	-6.12	-3.14	4.12	-
J-stat	7.785					
p-value	0.993					

Panel E: CRR factors +MKT						
	MP	UI	DEI	Defy	Term	MKT
Estimate	-0.182***	0.091***	-0.023***	-0.017**	0.028***	0.082*
z-stat	-3.26	4.40	-4.79	-2.16	4.90	1.88
J-stat	7.093					
p-value	0.993					

2.5.3.3. Tests on 31 industry portfolios by GMM using moment conditions 1

Following Fama and French (1997) and Lewellen et al. (2010), we also repeat the tests on 31 industry portfolios. Compared with portfolios sorted by firm characteristics, value-weighted industry portfolios should embody a smaller factor structure and enable more powerful tests.

Table 2.7 presents the estimation results for all three sets of factors on the industry portfolios. Panel A shows the estimated coefficients of the Fama-French five factors. Contrary to the results in Tables 2.5 and 2.6, all five factors are significant at the 5% level. Panel B shows the estimated coefficients of the PCA factors. All PCA factors are significant except for PC4, which is the poxy for liquidity and exchange rate news. Panel C shows the estimated coefficients of the PCA factors augmented with the market excess return. The inclusion of MKT affects the significance of PC1 and PC2, rendering PC2 insignificant and PC1 only significant at the 10% level. The results are not surprising, since MKT has a relatively high correlation with both PC1 and PC2. Panel D shows the estimated coefficients of the CRR factors. Note that MP and DEI, which are significant in Table 2.6, are no longer significant on the industry portfolios. Panel E shows the estimated coefficients of the CRR factors augmented with the market excess return. The inclusion of MKT further weakens the informational uniqueness of Defy, which is now only marginally significant at the 10% level. Similar to the results on the 25 Size-E/P and 25 Size-IVOL portfolios, the results from Table 2.7 also show insignificant J-statistics for all model specifications.

We have also estimated conditional versions of the models involved in Tables 2.5-2.7, with scaled moments using 1-month lagged values of term spread and default spread as conditional information following Cochrane (1996). The results are qualitatively similar to those in Tables 2.5-2.7, with all model specifications having insignificant J-statistics, indicating that the models may also hold conditionally. However, this way of incorporating conditional information has received criticism for being arbitrary (Fama & French, 2020; Nagel, 2021). Thus, our focus is on the unconditional model, and the results are not presented here.

Table 2.7. Tests of asset pricing models on 31 industry portfolios

This table presents the estimated coefficients of the Fama-French, PCA, and CRR factors by GMM using moment conditions 1 in Section 2.3.3. The test assets are 31 value-weighted industry portfolios. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread respectively. The estimated parameters are factor loadings on the SDF. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

Panel A: Fama-French 5 factors					
	MKT	SMB	VMG	RMW	CMA
Estimate	0.038***	0.065***	-0.043***	0.179***	0.102***
z-stat	5.11	3.40	-2.84	7.73	5.37
J-stat	8.177				
p-value	0.999				

Panel B: PCA factors						
	PC1	PC2	PC3	PC4	PC5	MKT
Estimate	0.066***	-0.059**	0.074***	0.031	-0.218**	-
z-stat	6.20	-2.35	5.19	1.41	-2.12	-
J-stat	9.236					
p-value	0.999					

Panel C: PCA factors +MKT						
	PC1	PC2	PC3	PC4	PC5	MKT
Estimate	0.034*	0.014	-0.158***	0.028	-0.261***	0.059*
z-stat	1.82	0.33	-5.95	0.81	-3.21	1.69
J-stat	9.035					
p-value	0.972					

Panel D: CRR factors						
	MP	UI	DEI	Defy	Term	MKT
Estimate	0.011	0.096***	-0.028	-0.134**	-0.069**	-
z-stat	0.62	5.32	-0.75	-2.16	-2.09	-
J-stat	10.411					
p-value	0.997					

Table 2.7. Tests of asset pricing models on 31 industry portfolios *(continued)*

Panel E: CRR factors +MKT						
	MP	UI	DEI	Defy	Term	MKT
Estimate	0.008	0.071***	0.093*	-0.089*	-0.094**	0.005
z-stat	0.43	6.93	1.86	-1.77	-2.37	0.97
J-stat	10.082					
p-value	0.996					

2.5.4. Estimation of factor risk premiums by the Fama-Macbeth (1973) procedure

The GMM tests with moment conditions 1 in Section 2.5.3 answer whether each factor contains unique pricing information; however, whether they are priced remains to be investigated. In this subsection, we show the results of direct estimations of the factor risk premiums following the Fama-Macbeth (1973) procedure illustrated in Section 2.3.4. The tests are conducted on the same sets of test assets as in Section 2.5.3. We calculate the t-statistics of the estimated risk premiums using Shanken's (1992) standard errors corrected for errors-in-variables bias.

2.5.4.1. Risk premiums estimation on 25 Size-E/P portfolios

Table 2.8 presents the estimation results for the Fama-French five factors on 25 Size-E/P portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the Fama-French three and five factors, respectively. Most single-factor estimates are insignificant due to high standard errors caused by large residual variance. The results indicate the existence of strong size and value effects in China. The size premium is 0.904, and the value premium is 0.710 in the Fama-French five-factor model. The risk premium for the investment factor is also significant. The market risk premium, however, is statistically insignificant at the 5% level. Following a modified construction procedure for the size and value factors as in Liu et al. (2019), our findings contradict earlier studies on the Fama-French factors in the Chinese stock market by Wang and Xu (2004) and Hilliard and Zhang (2015), who record no significant value effect. However, our results are supported by several recent studies, such as Guo et al. (2017), Hu et al. (2019), and Liu et al. (2019). Also, in Table 2.5, RMW appears to provide incremental pricing information about returns of 25 Size-E/P portfolios. However, results in Column (7) indicate that RMW is not priced due to its correlation with other factors.

Table 2.9 presents the estimation results for the PCA factors on 25 Size-E/P portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the PCA factors and the PCA factors augmented with the market excess return, respectively. In the model with the PCA factors only, all factor risk premiums are significant except for that on PC5, which is only significant at the 10% level. In Column (7), the estimated risk premium on MKT is not statistically significant, and the adjusted R^2 does not benefit significantly from adding MKT. The results suggest that MKT is not priced over the 25 Size-E/P portfolios. Also, we expect PC1 to have a positive risk premium as it is positively correlated with growth variables normally considered to signal a "good state" of the economy. We expect the rest of the PCA factors to have negative risk premiums as each is positively correlated with variables that are considered to signal a "bad state" of the economy by construction. However, the signs on PC3 and PC5 are inconsistent with our expectations.

Table 2.8. Risk premiums for the Fama-French factors on 25 Size-E/P portfolios
This table presents the estimated risk premiums of the Fama-French factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and the E/P ratio. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
MKT	0.687					0.904*	0.835*
	(1.41)					(1.82)	(1.79)
SMB		0.673**				0.920***	0.904**
		(2.26)				(2.93)	(2.17)
VMG			0.305			0.725***	0.710***
			(1.31)			(2.85)	(2.61)
RMW				-0.436			0.342
				(-0.78)			(0.51)
CMA					0.945**		0.708***
					(2.22)		(2.88)
cons	0.861	0.918	1.383*	1.374*	1.355*	2.822*	4.964***
	(0.32)	(1.39)	(1.93)	(1.88)	(1.76)	(1.85)	(3.14)
R^2	0.183	0.344	0.267	0.215	0.244	0.617	0.713
adj. R^2	0.148	0.315	0.235	0.181	0.211	0.546	0.625

Table 2.9. Risk premiums for the PCA factors on 25 Size-E/P portfolios

This table presents the estimated risk premiums of the PCA factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and the E/P ratio. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
PC1	1.169**					0.783***	0.515**
	(2.24)					(2.84)	(2.31)
PC2		-0.535**				-0.647***	-0.495**
		(-2.01)				(-4.77)	(-2.23)
PC3			1.358**			0.503***	0.458**
			(2.31)			(2.61)	(2.32)
PC4				-0.149		-0.342***	-0.312***
				(-1.12)		(-2.82)	(-2.69)
PC5					0.039	0.143*	0.104
					(0.56)	(1.66)	(1.56)
MKT							0.479
							(1.60)
cons	-0.872	-1.436	-0.075	1.071	1.234	1.328	3.507***
	(-0.60)	(-1.08)	(-0.09)	(1.30)	(1.49)	(1.57)	(4.07)
R^2	0.125	0.330	0.297	0.235	0.291	0.683	0.703
adj. R^2	0.087	0.300	0.266	0.202	0.260	0.601	0.604

Table 2.10. Risk premiums for the CRR factors on 25 Size-E/P portfolios

This table presents the estimated risk premiums of the CRR factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and the E/P ratio. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread, respectively. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
MP	0.055					-0.105**	-0.221**
	(1.25)					(-1.98)	(-2.41)
UI		-0.593***				-0.234***	-0.342***
		(-3.71)				(-3.92)	(-5.87)
DEI			0.474			0.445***	0.585***
			(1.12)			(3.35)	(4.28)
Defy				-0.136		-0.531***	-0.424***
				(-1.37)		(-3.53)	(-3.20)
Term					0.011	-0.005	-0.016
					(1.33)	(-0.65)	(-0.86)
MKT							1.034
							(1.53)
cons	2.091*	0.574	0.196	0.489	0.348	4.326***	3.213**
	(1.75)	(0.83)	(0.31)	(0.62)	(0.54)	(2.87)	(2.26)
R^2	0.296	0.231	0.238	0.217	0.252	0.641	0.681
adj. R^2	0.266	0.198	0.205	0.183	0.219	0.563	0.574

Table 2.10 presents the estimation results for the CRR factors on 25 Size-E/P portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the CRR factors and the CRR factors augmented with market excess return, respectively. Once again, most single-factor estimates are insignificant due to large residual variance. Consistent with previous studies such as Chen et al. (1986), Maio and Santa-Clara (2012), Boons, M. (2016), Cooper and Maio (2019), and Cooper et al. (2022), default spread news and unexpected inflation have significant negative risk premiums as they are believed to signal "bad state" of economy in general. However, contradicting previous studies in the U.S., news about term spread earns a negative but insignificant risk premium on the Chinese 25 Size-E/P portfolios. Moreover, the risk premiums on news about industrial production growth and changes in expected inflation have signs that are opposite to our expectations. The results suggest that MP may not be a good measure of Chinese economic growth news, as it only takes into account output by firms in the industrial sector (according to statistics by the World Bank, about 24% of output in the sector is attributed to state-owned entities) that have operating revenue over 20 million RMB in the past year. In Column (7), MKT is insignificant, and including MKT does not lead to qualitatively different risk premium estimates on the CRR factors, indicating that MKT is not priced when augmented with the CRR factors.

Also, note that the ICAPM implies that MP and Term should have negative risk premiums as they positively forecast future market volatilities, as shown in Table 2.4. Thus, although the signs of the risk premiums on those two factors are opposite to those in previous studies, they are consistent with the implications of the ICAPM.

2.5.4.2. Risk premiums estimation on 25 Size-IVOL portfolios

Table 2.11 presents the estimation results for the Fama-French factors on 25 Size-IVOL portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the Fama-French three and five factors, respectively. Similar to the results on 25 size-E/P portfolios, we find significant size, value, and investment premiums. In addition, surprisingly, the market risk premium is also significant in the three- and five-factor models. Also, out of our expectation, the profitability risk premium is negative in Column (7) although only significant at the 10% level.

Table 2.12 presents the estimation results for the PCA factors on the 25 Size-IVOL portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the PCA factors and the PCA factors augmented with the market excess return, respectively. Most single-factor estimates are insignificant due to high residual variance. In multiple regressions, contrary to the results on the 25 Size-E/P portfolios, the risk premium on PC2 is insignificant with or without MKT being included. The signs on the factor risk premiums are consistent with our expectations except for PC3, which is still positive. Similar to the results in Table 2.9, MKT is not priced when augmented with the PCA factors.

Table 2.11. Risk premiums for the Fama-French factors on 25 Size-IVOL portfolios

This table presents the estimated risk premiums of the Fama-French factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret	(6) Ret	(7) Ret
MKT	0.703*					1.143**	1.290**
	(1.83)					(2.21)	(2.34)
SMB		0.786**				1.315***	1.251***
		(2.27)				(5.24)	(4.36)
VMG			0.340			0.814***	0.779***
			(0.45)			(3.91)	(3.62)
RMW				2.397***			-0.819*
				(5.08)			(-1.91)
CMA					1.843***		1.198**
					(6.61)		(2.01)
cons	2.726***	0.845	1.500**	1.428*	1.282*	-3.020***	-5.172***
	(4.48)	(1.27)	(2.34)	(1.93)	(1.75)	(-3.80)	(-4.29)
R^2	0.116	0.366	0.268	0.080	0.203	0.547	0.651
adj. R^2	0.078	0.338	0.236	0.040	0.168	0.483	0.539

Table 2.12. Risk premiums for the PCA factors on 25 Size-IVOL portfolios

This table presents the estimated risk premiums of the PCA factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret	(6) Ret	(7) Ret
PC1	-0.174					0.812**	0.436*
	(-0.42)					(2.34)	(1.71)
PC2		-0.275				-0.241	0.154
		(-1.29)				(-1.14)	(1.10)
PC3			1.150***			0.776***	0.640***
			(3.31)			(2.91)	(2.71)
PC4				-1.585***		-0.462***	-0.397**
				(-4.18)		(-3.65)	(-2.05)
PC5					-0.185	-0.078	-0.108
					(-1.60)	(-0.65)	(-0.94)
MKT							0.663
							(0.95)
cons	1.551*	0.177	2.607***	1.090	1.283*	4.371***	6.308***
	(1.81)	(0.11)	(3.78)	(1.35)	(1.79)	(4.94)	(5.30)
R^2	0.204	0.327	0.118	0.073	0.338	0.545	0.564
adj. R^2	0.170	0.290	0.079	0.033	0.308	0.426	0.438

Table 2.13. Risk premiums for the CRR factors on 25 Size-IVOL portfolios

This table presents the estimated risk premiums of the CRR factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread, respectively. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
MP	0.062					0.236***	0.210***
	(0.67)					(2.62)	(2.59)
UI		-0.673***				-0.834***	-0.891***
		(-4.34)				(-3.83)	(-5.05)
DEI			1.067**			0.312**	0.385**
			(2.31)			(2.17)	(2.34)
Defy				0.152		-0.331***	-0.262***
				(0.83)		(-3.16)	(-2.67)
Term					0.614**	0.438**	0.306**
					(2.42)	(2.44)	(2.17)
MKT							-0.791
							(-0.98)
cons	1.110*	0.930	1.197	0.351	-0.969	2.347**	-2.113**
	(1.75)	(1.35)	(1.32)	(0.37)	(-1.35)	(2.38)	(-2.24)
R^2	0.291	0.280	0.192	0.306	0.276	0.635	0.677
adj. R^2	0.263	0.248	0.157	0.276	0.244	0.559	0.569

Table 2.13 presents the estimation results for the CRR factors on 25 Size-IVOL portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the CRR factors and the CRR factors augmented with the market excess return, respectively. The risk premiums on all CRR factors are significant in multiple regressions with or without MKT. Similar to results on the 25 Size-E/P portfolios, in Column (7), the estimated risk premium on MK T is not significant, and the inclusion of MKT does not lead to qualitative different risk premium estimates on the CRR factors, indicating MKT is not priced when augmented with the CRR factors. In Column (6), MP has an estimated risk premium of 0.236, UI has an estimated risk premium of -0.834, Defy has an estimated risk premium of -0.331, and Term has an estimated risk premium of 0.438. The above results are consistent with the findings of Chen et al. (1986), Liu and Zhang (2008), Maio and Santa-Clara (2012), and Cooper et al. (2022). However, the expected inflation has an estimated risk premium of 0.312. Although the sign is consistent with the result in Table 2.10, it contradicts the findings in previous studies. Also, note that the signs on risk premiums of MP, DEI, and Term are inconsistent with the implications of the ICAPM, as shown in Table 2.4.

2.5.4.3. Risk premiums estimation on 31 industry portfolios

Table 2.14 presents the estimation results for the Fama-French factors on 31 industry portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on Fama-French three and five factors, respectively. The results indicate that the Fama-French factors perform poorly on the industry portfolios. The adjusted R^2 in Columns (6) and (7) drop considerably compared to those in Tables 2.8 and 2.11, indicating the models' deteriorated goodness of fit. Moreover, the size and value premiums have the wrong signs.

Table 2.14. Risk premiums for the Fama-French factors on 31 industry portfolios

This table presents the estimated risk premiums of the Fama-French factors by the Fama-Macbeth (1973) procedure. The test assets are 31 value-weighted industry portfolios. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
MKT	0.548					0.632	0.573
	(0.45)					(0.53)	(0.79)
SMB		-0.010				-0.638**	-1.129*
		(-0.03)				(-1.98)	(-1.82)
VMG			-0.099			-0.709**	-1.236*
			(-0.34)			(-2.08)	(-1.77)
RMW				0.235			0.146
				(1.04)			(0.14)
CMA					0.065		-0.202
					(0.20)		(-0.42)
cons	0.401	0.761	0.675	0.765	0.817*	0.439	0.161
	(0.60)	(1.62)	(1.37)	(1.56)	(1.75)	(0.64)	(0.23)
R^2	0.094	0.161	0.158	0.108	0.076	0.278	0.367
adj. R^2	0.062	0.132	0.129	0.087	0.044	0.236	0.301

Table 2.15. Risk premiums for the PCA factors on 31 industry portfolios

This table presents the estimated risk premiums of the PCA factors by the Fama-Macbeth (1973) procedure. The test assets are 31 value-weighted industry portfolios. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
PC1	0.065					0.142	0.101
	(0.09)					(1.38)	(1.24)
PC2		0.071				-0.301**	-0.173**
		(0.73)				(-2.39)	(-2.05)
PC3			-0.136			-0.289**	-0.375**
			(-1.33)			(-2.13)	(-2.38)
PC4				-0.353		0.136	0.261
				(-1.55)		(1.36)	(1.54)
PC5					0.092	0.141*	0.121*
					(1.45)	(1.73)	(1.69)
MKT							-0.393
							(-0.71)
cons	0.958	0.385	0.963	1.343**	1.037	0.794	0.917
	(1.15)	(0.42)	(1.63)	(2.01)	(1.61)	(1.04)	(1.09)
R^2	0.097	0.144	0.097	0.121	0.101	0.418	0.461
adj. R^2	0.066	0.115	0.066	0.090	0.070	0.325	0.386

Table 2.16. Risk premiums for the CRR factors on 31 industry portfolios

This table presents the estimated risk premiums of the CRR factors by the Fama-Macbeth (1973) procedure. The test assets are 31 value-weighted industry portfolios. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread, respectively. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
MP	0.038					0.104	0.073
	(1.54)					(1.57)	(1.05)
UI		-0.112				-0.047	-0.054
		(-1.41)				(-0.69)	(-0.78)
DEI			0.182			0.039	0.060
			(1.32)			(0.27)	(0.43)
Defy				-0.201		-0.485**	-0.314**
				(-1.41)		(-2.50)	(-2.13)
Term					0.049	0.053	-0.021
					(0.86)	(0.91)	(-0.68)
MKT							0.635
							(1.15)
cons	1.148	0.374	0.167	0.265	0.234	-0.343	0.067
	(1.57)	(0.61)	(0.28)	(0.34)	(0.39)	(-0.51)	(0.09)
R^2	0.109	0.081	0.058	0.136	0.036	0.233	0.285
adj. R^2	0.078	0.049	0.034	0.106	0.023	0.205	0.239

Table 2.15 presents the estimation results for the PCA factors on 31 industry portfolios. Contrary to the results in Table 2.12 and Table 2.9, the risk premiums of PC1 and PC4 are no longer significant in the last two columns. It is surprising, considering that PC1 explains almost half of the total variation in the nine main asset class returns. The risk premium of PC5 is now marginally significant at the 10% significance level but with an opposite sign to our expectation.

Table 2.16 presents the estimation results for the CRR factors on 31 industry portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the CRR factors and the CRR factors augmented with market excess return, respectively. None of the single-factor premium estimates is significant due to high residual variance. The factor risk premiums show decreased significance compared with the results on the other two sets of test assets. In columns (6) and (7), only default spread news earns a significant risk premium (-0.485 in Column (6) and -0.314 with MKT included).

Results in Tables 2.10, 2.13, and 2.16 show that the CRR factors are not always consistent with the implications of ICAPM. The inconsistency mainly comes from variables that help predict future market volatilities. The state variables corresponding to MP, Term, and Liquidity all positively forecast future market volatilities (Thus, they should earn negative risk premiums as implied by the ICAPM), whereas they are expected to earn positive risk premiums according to macroeconomic rationale (The rationale also finds some support from our empirical results).

2.5.5. Factor comparison

In this subsection, we compare the factors using three different metrics. First, we look at the results of Newey and West's (1987b) ΔJ tests for models estimated by GMM using moment conditions 1. The ΔJ statistics and the corresponding p-values for each pair of restricted and unrestricted models are presented in Table 2.17. Over each set of test assets, the restricted models are models with the PCA and CRR factors only, and the unrestricted models are models augmenting the PCA and CRR factors with the Fama-French five factors. The tests examine whether the Fama-French five factors are spurious in the presence of the macroeconomic factors. Results show that the ΔJ-statistics for each pair of the restricted and unrestricted models over each set of test assets is insignificant at the 5% level, indicating that the Fama-French five factors contain mainly information about macroeconomic news and provide little incremental power in pricing the test assets in the presence of the macro factors.

Table 2.17. Results of ΔJ tests for factor redundancy
This table presents the Newey and West (1987b) ΔJ statistics and the corresponding p-values for each pair of restricted and unrestricted models. The restricted models are models with the PCA and CRR factors only, and the unrestricted models are models augmenting the PCA and CRR factors with the Fama-French five factors. The tests examine whether the Fama-French five factors are redundant in the presence of the macroeconomic factors. The sample period is from 2006:M12 to 2021:M12.

	PCA factors	CRR factors
Size-E/P	1.941	2.043
	(0.857)	(0.843)
Size-IVOL	1.731	1.405
	(0.885)	(0.924)
Industry	4.842	0.531
	(0.435)	(0.991)

We then inspect estimates of pricing errors from the asset pricing models. Measurements of two pricing error metrics are presented in Table 2.18. Panel A shows measurements of HJ-distance calculated by 1-step GMM using moment conditions 1 with the Hansen and Jagannathan (1997) weighting matrix.

Over the 25 Size-E/P portfolios, neither of the PCA and CRR factors outperforms the Fama-French five factors, while both show superior performance to the Fama-French three factors. Surprisingly, over the 25 Size-IVOL portfolios, the CRR factors demonstrate the best performance among all sets of factors, followed by the Fama-French five factors. Similar to the results over the 25 Size-E/P portfolios, both sets of macroeconomic factors (as noted in Section 2.1, our PCA factors can be regarded as a set of macroeconomic factors in a broader sense) outperform the Fama-French three factors. The results indicate that macroeconomic factors may have captured information about some omitted systematic risks, implied by the idiosyncratic volatility anomaly (Ang et al., 2009; Chen et al., 2012), that is not contained in the size and value factors.

Like the results over the 25 Size-E/P portfolios, the Fama-French five factors regain the position of the best-performing factors over the 31 industry portfolios. The CRR factors, on the other hand, become the worst-performing factors, in stark contrast to their performance over the 25 Size-IVOL portfolios.

Panel B presents measurements of average pricing error as in Cooper et al. (2022). The metric is calculated as the square root of the mean squared cross-sectional residuals derived from the second step of the Fama-Macbeth procedure. The relative performance of the factors as indicated by their average pricing errors coincides with that indicated by HJ-distance in general, with only slight deviations. In all, results in Table 2.18 indicate that although both sets of macroeconomic factors outperform the Fama-French three factors in general, their performances over the different sets of assets are not robust.

Table 2.18. Comparison of pricing errors
This table compares measurements of two pricing error metrics from competing asset pricing models. Panel A presents results about the HJ-distance derived from a one-step GMM. Panel B presents results about the average pricing error derived from the Fama-Macbeth (1973) procedure. Each row shows the pricing error measurements of different factors on one of the three sets of test portfolios: 25 Size-E/P portfolios, 25 Size-IVOL portfolios, and 31 Industry portfolios. The sample period is from 2006:M12 to 2021:M12.

Panel A: HJ-distance				
	FF 3 factors	FF 5 factors	PCA factors	CRR factors
Size-E/P	0.425	0.389	0.395	0.413
Size-IVOL	0.781	0.622	0.704	0.297
Industry	0.595	0.562	0.586	0.601

Panel B: Average pricing error				
	FF 3 factors	FF 5 factors	PCA factors	CRR factors
Size-E/P	1.587	1.394	1.442	1.471
Size-IVOL	1.701	1.448	1.730	1.417
Industry	2.634	1.908	1.626	3.056

2.6. Robustness checks

2.6.1. Estimation of covariance risk premiums by GMM

Cochrane (2005) shows that theoretically, elements of the factor covariance risk premiums $\boldsymbol{\lambda}$ in (2.9) should be proportional to the elements of the factor loadings $\boldsymbol{b}$ in (2.8) scaled by the variance of corresponding factors. Thus, in theory $\boldsymbol{\lambda}$ and $\boldsymbol{b}$ should have the same sign and significance. In this section, we supplement the results for the estimated factor loadings in Section 2.5.3 with covariance risk premiums estimated by optimal GMM using moment conditions 2.

Table 2.19 presents the estimated factor covariance risk premiums on 25 Size-E/P portfolios. Panel A presents the results for the Fama-French factors. Consistent with the results in Panel A of Table 2.5, the coefficient on the size and value factors are still positive and significant, and those on the market return and the investment factor are still insignificant at the 5% level. However, contrary to Table 2.5, the coefficient on the profitability factor is now only significant at the 10% level. Panel B presents results for the PCA factors. Since MKT is superfluous in both the SDF test and the Fama-Macbeth procedure when augmented with the PCA factors, we exclude MKT from the final model specification. The results are generally consistent with those in Panel B of Table 2.5. All factors have significant covariance risk premiums except for PC5, which has significant factor loading on the SDF in Table 2.5. Panel C presents result for the CRR factors. Similar to the case of PCA factors, MKT is superfluous in both the SDF test and the Fama-Macbeth procedure, thus we exclude it from the final model specification. Consistent with the results in Panel D of Table 2.5, the coefficients on UI, Defy, and Term are significant and have the same signs as the corresponding factor loadings in Table 2.5. However, results for the other two factors are substantially different. The coefficient on DEI is still insignificant but with a reversed sign. Compared with the estimated SDF loading in Table 2.5, the estimated covariance risk premium on MP now turns insignificant and with a reversed sign. The J-statistics of Hansen's (1982) test of overidentifying restrictions for every model specification are insignificant at the 5% level. However, the J-statistics for the CRR factors are comparatively high, with a p-value of 0.569. In all, we do not find strong evidence of model misspecification.

Table 2.20 presents the estimated factor covariance risk premiums on 25 Size-IVOL portfolios. Panel A presents the results for the Fama-French factors. The coefficients on the size, value, and profitability factors are still significant and have the same sign as the corresponding factor loadings in Panel A of Table 2.6. However, contrary to the results in Table 2.6, the coefficient on the investment factor now becomes significant and with a reversed sign. The coefficient on the market return is still insignificant but with a reversed sign. Such results signal the possibility of model misspecification. The J-statistics is 16.343 with a p-value of 0.360. Although the statistics is not significant at the 5% level, it is relatively high compared with those of the other factors. Panel B presents the results for the PCA factors. The results are generally consistent with those in Panel B of Table 2.6, except that the coefficient on PC4 is now insignificant at the 5% level, and there is a reversed sign on the coefficient of PC5. Panel C presents the results of the CRR factors. The results are generally consistent with those in Panel D of Table 2.6, except that the coefficient on the innovation of expected inflation is no longer significant.

Table 2.19. Factor covariance risk premiums on 25 Size-E/P portfolios

This table presents the estimated covariance risk premiums of the Fama-French, PCA, and CRR factors by GMM using moment conditions 2 in Section 2.3.3. The test assets are 25 Size-E/P portfolios. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread, respectively. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

Panel A: Covariance risk premiums of the Fama-French 5 factors					
	MKT	SMB	VMG	RMW	CMA
Estimate	1.736*	1.432***	0.894***	0.623*	0.706*
z-stat	1.86	5.82	4.77	1.69	1.74
J-stat	8.100				
p-value	0.920				

Panel B: Covariance risk premiums of the PCA factors					
	PC1	PC2	PC3	PC4	PC5
Estimate	0.852***	-0.464**	-0.709***	-0.192***	0.072
z-stat	5.50	-2.37	-2.67	-3.27	1.05
J-stat	8.011				
p-value	0.923				

Panel C: Covariance risk premiums of the CRR factors					
	MP	UI	DEI	Defy	Term
Estimate	-0.113	0.584**	0.171	-1.045***	-0.546**
z-stat	-1.07	2.24	1.07	-2.84	-2.32
J-stat	13.437				
p-value	0.569				

Table 2.20. Factor covariance risk premiums on 25 Size-IVOL portfolios

This table presents the estimated covariance risk premiums of the Fama-French, PCA, and CRR factors by GMM using moment conditions 2 in Section 2.3.3. The test assets are 25 Size-IVOL portfolios. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread, respectively. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

Panel A: Covariance risk premiums of the Fama-French 5 factors					
	MKT	SMB	VMG	RMW	CMA
Estimate	-0.945	1.247***	1.092***	-0.823***	-0.477***
z-stat	-1.45	3.15	5.75	-3.49	-3.07
J-stat	16.343				
p-value	0.360				

Panel B: Covariance risk premiums of the PCA factors					
	PC1	PC2	PC3	PC4	PC5
Estimate	0.216*	-0.465	0.392***	-0.078***	0.113
z-stat	1.71	-0.33	5.70	-4.66	1.40
J-stat	15.591				
p-value	0.410				

Panel C: Covariance risk premiums of the CRR factors					
	MP	UI	DEI	Defy	Term
Estimate	-0.431**	0.876***	0.133	-0.874***	0.692**
z-stat	-2.05	2.93	1.03	-4.31	2.11
J-stat	9.585				
p-value	0.845				

Table 2.21. Factor covariance risk premiums on 31 Industry portfolios
This table presents the estimated covariance risk premiums of the Fama-French, PCA, and CRR factors by GMM using moment conditions 2 in Section 2.3.3. The test assets are 31 Industry portfolios. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread respectively. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

Panel A: Covariance risk premiums of the Fama-French 5 factors					
	MKT	SMB	VMG	RMW	CMA
Estimate	1.967***	0.515***	-0.114	1.243***	0.626***
z-stat	2.87	3.32	-1.35	4.63	3.01
J-stat	20.461				
p-value	0.492				

Panel B: Covariance risk premiums of the PCA factors					
	PC1	PC2	PC3	PC4	PC5
Estimate	-0.461*	-0.524***	1.083***	0.015	-0.061***
z-stat	-1.69	-3.86	4.93	0.47	-4.00
J-stat	7.922				
p-value	0.995				

Panel C: Covariance risk premiums of the CRR factors					
	MP	UI	DEI	Defy	Term
Estimate	-0.122	-0.437	0.261	-0.876**	-0.234
z-stat	-1.59	-1.53	1.26	-2.25	-0.96
J-stat	9.950				
p-value	0.980				

Table 2.21 presents the estimated factor covariance risk premiums on 31 industry portfolios. Panel A presents the results for the Fama-French factors. The results are generally consistent with those in Panel A of Table 2.7, except that the coefficient on VMG is now insignificant. Panel B presents the results of the PCA factors. The results are generally consistent with those in Panel B of Table 2.7. The only exception is that the coefficient on PC1 is now insignificant and with a reversed sign compared with its estimated loading on the SDF. Panel C presents the results for the CRR factors. In contrast with the results for the other two sets of factors, there is considerable inconsistency between the results in Panel C and those in Panel D of Table 2.7. The estimated coefficients on MP, UI, and DEI all have reversed signs compared with their loadings on the SDF. Moreover, the coefficients on UI and Term are no longer significant. Still, the J-statistics for every model specification above are insignificant at the 5% level, and we do not find strong evidence of model misspecification.

2.6.2. Estimation results before the introduction of the Shanghai-Hong Kong Stock Connect

Studies indicate that the tie between the Chinese stock market and macroeconomic risks has strengthened since China joined the WTO and the outbreak of the Global Financial Crisis (GFC). Girardin and Joyeux (2013) detect a significantly stronger relationship between the Chinese market return and macroeconomic fundamentals after China joined the WTO in late 2001. Similarly, Goh et al. (2013) find that some Goyal and Welch (2008) style U.S macroeconomic variables can significantly predict Chinese stock return post-GFC. Guo (2015) detects no significant relationship between real economic growth and the market return in China before the GFC but records significant unidirectional causality in the mean from real economic growth to market return during the post-GFC period. Abbas et al. (2019) also record enhanced spillover effects from some macroeconomic variables to Chinese market return and volatility during the post-GFC period. It would be interesting to check if macroeconomic risks affect the cross-section of Chinese stock returns more prominently as the Chinese stock market develops and opens up. Due to data availability, we cannot test explicitly whether macroeconomic factors have gained stronger explanatory powers in the cross-section of Chinese stock returns since China joined the WTO or the occurrence of the GFC. However, we may probe the problem indirectly using a similar event: the introduction of the Shanghai-Hong Kong Stock Connect. The Shanghai-Hong Kong Stock Connect, launched in November 2014, connects the Shanghai Stock Exchange and the Hong Kong Stock Exchange and enables investors in each market to access the other market using only local brokers and clearing houses. As an important scheme that opened the Chinese stock market, we expect it would strengthen the bond between Chinese stock returns and macroeconomic risks. Thus, we repeat the analysis in Tables 2.5, 2.9, and 2.10 over the sample period before November 2014 and check if the results are substantially different from those over the whole sample.

Table 2.22 repeats the analysis of Table 2.5 and re-estimates models with the PCA and CRR factors using GMM moment condition 1 over the sample period of December 2006 to October 2014. Panel A presents the estimated loadings of the PCA factors on the SDF. PC1, PC3, and PC4, which are significant in Table 2.5, are no longer significant at the 5% significance level. The significance level of the coefficient on PC5 also drops, although it is still significant at the 5% significance level. Compared with the result in Table 2.5, the J-statistics of the model rises from 6.805 to 10.837, although the p-value still shows no sign of model misspecification. Panel B presents the estimated loadings of the CRR factors on the SDF. The industrial production and term spread news, which are highly significant over the entire sample, are no longer significant at the 5% level. There is a reversion of sign for the loading of term spread news. The unexpected inflation and default spread are still significant at the 5% level, and the change in expected inflation remains insignificant. From Table 2.22, we observe considerable drops in the significance of the estimated factor loadings for both PCA and CRR factors.

Table 2.23 repeats the analysis of Table 2.9 and re-estimates the factor risk premiums using the Fama-Macbeth (1973) procedure over the sample period of December 2006 to October 2014. In Column (6), the adjusted R^2 for the full model drops from 0.601 over the entire sample to 0.578 over the subsample. The risk premiums for PC1, PC3, and PC4 are no longer significant at the 5% level. The risk premium for PC2 is still significant, but there is a reduction in significance compared with that over the whole sample. The risk premium for PC5 remains significant at the 10% level but with a reversed sign.

Table 2.22 Tests of asset pricing models on 25 Size-E/P portfolios before the introduction of the Shanghai-Hong Kong Stock Connect

This table presents the estimated coefficients of the PCA, and CRR factors by GMM using moment conditions 1 in Section 2.3.3. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread respectively. The estimated parameters are factor loadings on the SDF. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2014:M10. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

Panel A: PCA factors					
	PC1	PC2	PC3	PC4	PC5
Estimate	0.036*	-0.188***	0.061	-0.186*	-0.213**
z-stat	1.74	-3.25	0.79	-1.87	-2.21
J-stat	10.837				
p-value	0.950				

Panel B: CRR factors					
	MP	UI	DEI	Defy	Term
Estimate	-0.023	-0.062**	-0.012	-0.052***	0.073*
z-stat	0.91	-2.19	-0.23	-3.13	1.88
J-stat	11.621				
p-value	0.929				

Table 2.23. Risk premiums for the PCA factors on 25 Size-E/P portfolios before the introduction of the Shanghai-Hong Kong Stock Connect

This table presents the estimated risk premiums of the PCA factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. PC1 is the PCA Growth factor. PC2 is the PCA Stagflation factor.PC3 is the PCA Long-term rate factor. PC4 is the PCA Liquidity factor. PC5 is the PCA Growth spread factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2014:M10. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret	(6) Ret
PC1	0.472					0.521
	(1.44)					(1.23)
PC2		-0.623^{**}				-0.389^{**}
		(-1.99)				(-2.56)
PC3			1.054^{*}			0.412^{*}
			(1.71)			(1.83)
PC4				-0.205		-0.186
				(-0.94)		(-1.43)
PC5					-0.301^{**}	-0.172^{*}
					(-2.06)	(-1.77)
cons	-1.642^{**}	-2.363^{**}	-1.102^{*}	1.231	1.433	1.196
	(-2.06)	(-2.37)	(-1.79)	(1.58)	(1.76)	(1.36)
R^2	0.111	0.352	0.321	0.209	0.246	0.661
adj. R^2	0.072	0.323	0.291	0.175	0.213	0.578

Table 2.24. Risk premiums for the CRR factors on 25 Size-E/P portfolios before the introduction of the Shanghai-Hong Kong Stock Connect

This table presents the estimated risk premiums of the CRR factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. MP is the mimicking portfolio for industrial production growth. UI is the innovation of unexpected inflation. DEI is the innovation of expected inflation. Defy and Term are mimicking portfolios of default spread and term spread respectively. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2014:M10. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret	(6) Ret
MP	0.031					0.056
	(1.14)					(1.45)
UI		-0.516***				-0.198**
		(-3.22)				(-2.32)
DEI			0.538			0.083
			(1.44)			(0.56)
Defy				0.105		-0.326*
				(1.18)		(-1.89)
Term					0.064*	0.053
					(1.91)	(1.58)
cons	0.732	1.487*	0.671	0.625	-2.348**	1.762**
	(1.23)	(1.86)	(0.67)	(0.86)	(-2.14)	(2.38)
R^2	0.274	0.183	0.211	0.189	0.143	0.518
adj. R^2	0.242	0.147	0.177	0.153	0.105	0.391

Table 2.24 repeats the analysis of Table 2.10 and re-estimates the factor risk premiums of the CRR factors using the Fama-Macbeth (1973) procedure over the sample period of December 2006 to October 2014. Note that for all model specifications, the adjusted R^2s are smaller than those over the entire sample period. For the full model in Column (6), the risk premiums for industrial production news and change in expected inflation are no longer significant at the 5% level. The risk premium for unexpected inflation UI also suffers from a reduction in significance level. The premium for default spread news, which is significant at 1% significance level in Table 2.10, becomes significant only at the 10% level. The results in Tables 2.23 and 2.24 hint about a tighter bond between macroeconomic risks and the cross-section of Chinese stock returns as the market develops and opens up.

2.7. Conclusion

This study investigates the pricing implication of macroeconomic factors in the cross-section of Chinese stock returns. To do so, we construct two sets of macroeconomic factors: five statistical factors with macroeconomic substance extracted by PCA from nine main asset class returns and the Chen-Roll-Ross (1986) factors. We investigate whether each macroeconomic factor provides incremental pricing information by employing GMM estimation using the stochastic discount factor approach. We further test if the factors are priced over three sets of test assets by the Fama-Macbeth (1973) procedure.

The results suggest that macroeconomic risks can help price stocks in the Chinese market. Most macroeconomic factors tested have significant loadings on the stochastic discount factor. Furthermore, most are priced with statistically and economically significant risk premiums. Our results for the Chen-Roll-Ross model are generally consistent with those of Cooper et al. (2022). We find significant negative risk premiums for unexpected inflation and default spread and weaker evidence of a positive risk premium for term spread over one of the three sets of test assets. However, our results differ from some earlier studies on the U.S. market, such as Hahn and Lee (2006) and Aretz et al. (2010), in that they do not find significant default spread premiums. We also find significant positive risk premiums for expected inflation over two sets of test assets, contradicted by the findings of Chan et al. (1986), Aretz et al. (2010), and Cooper et al. (2022). Although we find evidence that macroeconomic factors are priced in the cross-section of Chinese stock returns, the performance of the PCA and Chen-Roll-Ross factors are not as robust as expected, with signs of their risk premiums flipping across different test assets. Further tests under the framework of GMM suggest that the Fama-French five factors are superfluous in the presence of both sets of macroeconomic factors, indicating that the Fama-French five factors contain mainly information about macroeconomic shocks. We also compare the pricing errors of different sets of factors using two metrics. Once again, our results show that the performances of the PCA and Chen-Roll-Ross factors are unstable over different test assets. But overall, both sets of macroeconomic factors outperform the Fama-French three factors. In all, we find evidence that the Chinese stock market now prices macroeconomic risks similarly to more developed markets and thus can no longer be deemed a "casino." However, the performance of the PCA and Chen-Roll-Ross factors are far from satisfactory in terms of robustness. Further search for macroeconomic factors that could help better describe the cross-section of Chinese stock returns is much needed.

Chapter 3. Enhancing macroeconomic factors with machine learning in the Chinese stock market

3.1. Introduction

Macroeconomic factor models use innovations of theoretically inspired macroeconomic variables as factors. Such models assume that variations in the cross-section of asset returns are induced by variations in assets' exposures on innovations of macroeconomic variables. The Chen-Roll-Ross (CRR) model, first introduced by Chen et al. (1986), is among the most prominent macroeconomic factor models. The model identifies five macroeconomic variables that could influence expected cash flows or discount rates, and the Fama-Macbeth (1973) regressions in Chen et al. (1986) show that industrial production, default spread, and unexpected inflation earn significant risk premiums. Although works such as Shanken and Weinstein (2006) doubt the validity of the Chen-Roll-Ross model, a recent study by Cooper et al. (2022) provides it with solid support.

Since the introduction of the CRR model, academia has identified many other sources of macroeconomic risks that could affect the cross-section of asset returns. A small subset includes exchange rate risk (Jorion, 1991; Lustig & Verdelhan, 2007; Chung et al., 2015), labor income risk (Campbell, 1996; Jagannathan & Wang, 1996; Santos & Veronesi, 2006; Kim et al., 2011), macro liquidity risk (Chan et al.,1996; Jung & Kim, 2016;), monetary policy risk (Maio & Santa-Clara, 2017; Detzel, 2017; Ozdagli, & Velikov, 2020; Bu & Wu, 2021), economic policy uncertainty (Pastor & Veronesi, 2012; Brogaard & Detzel, 2015; Lee et al., 2021), commodity risks (Boons et al., 2014; Brooks et al., 2016) and inequality risk (Johnson, 2012). However, none of those models has gained influence and popularity comparable to the CRR model.

There are few studies on how macroeconomic risks affect the cross-section of Chinese stock returns. Most studies examining the impact of macroeconomic variables on the Chinese stock market focus on time series relationships. Among those studies, there is mixed evidence on whether and how macroeconomic factors affect the Chinese stock market. Soenen and Johnson (2001) find no evidence that inflation and real output affect the Chinese market return. Gay Jr (2008) asserts that exchange rates and oil prices do not influence the Chinese stock market. Allen et al. (2024) find a significant discrepancy between Chinese stock returns and economic fundamentals compared with other countries. Yao et al. (2013) argue that monetary policy does not affect the Chinese market index, while Koivu (2012) finds the opposite. In the early years, the seemingly disjoint between macroeconomic fundamentals and the Chinese stock market makes Girardin and Liu (2003) call the market a “casino." However, there is evidence that the bond between the Chinese stock market and macroeconomic variables has strengthened since China joined the WTO in late 2001 and the outbreak of the Global Financial Crisis (GFC).

Girardin and Joyeux (2013) detect a significantly stronger relationship between the Chinese market index and macroeconomic fundamentals after China joined the WTO. Similarly, Goh et al. (2013) find that a set of U.S. macroeconomic variables can significantly predict Chinese stock return post-GFC. In contrast, they do not show any predictive powers before the GFC. Guo (2015) detects significant unidirectional causality in the mean from real economic growth to the Chinese market return during the post-GFC period. However, he finds no such relationship prior to the GFC. Those earlier findings are supported by Abbas et al. (2019), who recorded enhanced spillover effects from several macroeconomic variables to the Chinese market return and volatility in the post-GFC period.

Thus, a comprehensive study on the cross-sectional pricing implication of macroeconomic factors will be interesting. We cover our first trial in Chapter 2, where we test two sets of macroeconomic factors (the PCA and CRR factors) in the cross-section of the Chinese stock returns. Although we find evidence that macroeconomic risks can help price stocks in the Chinese market, the performance of those two sets of factors is not robust enough, with sign reversals on their risk premiums and loadings on the SDF across different test assets. Further results for model comparison using two different pricing error metrics also show that the relative performances of the PCA and CRR factors are unstable across different test assets. Thus, a further search for macroeconomic factors that could help better describe the cross-section of Chinese stock returns is needed. With the aid of the machine learning method, this study aims to find enhanced macroeconomic factors that could demonstrate superior performance to the two sets of macroeconomic factors in Chapter 2. We also try to answer the following research questions: Do machine-learning-enhanced macroeconomic factors provide incremental pricing information? Are they priced in the Chinese stock market? Are they consistent with the restrictions of the ICAPM? Do they perform better than the PCA and CRR factors in Chapter 2?

To do so, we first construct a set of enhanced macroeconomic factors using a novel procedure that combines the ICAPM with the Elastic Net algorithm. The final machine-learning-enhanced macro-factor model consists of five macroeconomic factors: a factor representing news about economic growth, a factor representing news about inflation, a factor representing news about macro-liquidity, a factor representing news about the default spread, and a factor representing news about change in inequality. We also include the market return as a final factor so that the model can be classified as an ICAPM.

We estimate the model by GMM following the stochastic discount factor approach and investigate whether each enhanced macroeconomic factor provides incremental pricing information. We then test whether they are priced over three sets of test assets by the Fama-

Macbeth (1973) procedure. We also probe the relationship between the machine-learning-enhanced factors and the Fama-French five factors. Finally, we compare the performance of the machine-learning-enhanced factors with those of the other two sets of macroeconomic factors constructed in Chapter 2 by two metrics.

Our results show that almost all enhanced macroeconomic factors have significant loadings on the stochastic discount factor, providing unique pricing information across all three sets of test portfolios. Furthermore, the majority of these factors earn statistically and economically significant risk premiums, as confirmed by Fama-Macbeth (1973) regressions. With the exception of liquidity, the machine-learning-enhanced macroeconomic factors are consistent with ICAPM restrictions, as their risk premiums have the expected signs based on predictive regressions of the future investment opportunity set.

Further tests under the framework of GMM suggest that the Fama-French five factors provide little incremental pricing information in the presence of the machine-learning-enhanced macroeconomics factors. It is also worth mentioning that this study expands the work of Chapter 1 and reconfirms inequality as an essential dimension of systematic risk against a large pool of competing macroeconomic variables. The inequality factor constructed in Chapter 1 using the dynamic factor model survives the screening of the Elastic net and loads significantly on the stochastic discount factor. Inequality also earns a consistent negative risk premium in the presence of other enhanced macroeconomic factors. Such results are consistent with Chapter 1 and Johnson (2012).

Our machine-learning-enhanced macroeconomic factors perform on par with the Fama-French five factors and outperform the two sets of macroeconomic factors constructed in Chapter 2 in most scenarios when evaluated by two pricing error metrics. Compared to the CRR and PCA factor models, our machine-learning-enhanced macro-factor model also demonstrate more robust performance across different sets of test portfolios in terms of the significance and consistency of signs of the estimated coefficients. Our results indicate that the machine-learning-enhanced factors incorporate more useful pricing information than traditional, ad hoc sparse macroeconomic factors such as the CRR factors.

Our study contributes to the literature in the following ways. First, to the best of our knowledge, our work is the first comprehensive study examining the effect of a large pool of potential macroeconomic state variables on the cross-section of the Chinese stock market. Second, we adopt a novel methodology that combines the theoretical framework of the ICAPM with a data-driven machine learning method to construct informative macroeconomic factors, which have improved performance over traditional, ad hoc sparse macroeconomic factor models. Our work contributes to the growing body of literature that integrates traditional asset pricing studies with machine learning tools. Third, this chapter extends the study in Chapter 1 and provides further evidence that inequality is a significant risk factor for the cross-section of stock returns. To our knowledge, this study is the first to investigate the pricing implications of inequality in the cross-section of stock returns within a broader set of macroeconomic variables (Johnson, 2012 tests a two-factor CCAPM with inequality). Our treatment of inequality as an ICAPM state variable is also a new addition to the literature.

The rest of the chapter is organized as follows: Section 3.2 outlines the econometric methodologies involved in the construction of the machine-learning-enhanced macro-factors, estimation of asset pricing models, test of model consistency with the ICAPM, and model comparisons. Section 3.3 provides details of data and variables. Section 3.4 presents the empirical results of the asset pricing models and model comparisons. Section 3.5 presents results for additional robustness tests. Finally, Section 3.6 concludes and summarizes the main findings of the chapter.

3.2. Empirical methodology

3.2.1. Methods related to the construction of machine learning (ML) factors

We test our machine learning factors using GMM and the Fama-Macbeth (1973) procedure on the same sets of test assets as in Chapter 2: 25 Size-E/P portfolios, 25 Size-IVOL portfolios, and 31 industry portfolios. We then compare the performance of the machine learning factors with those of the three sets of factors (the CRR, PCA, and Fama-French five factors) involved in Chapter 2. This section shows the methods for constructing the machine learning factors.

3.2.1.1. Hodrick–Prescott (1997) Filter

To construct the machine learning macroeconomic factors, we first need to remove the trend and irregular components of the raw macroeconomic variables. Let y_t for t=1,2,…T be the time series of a macroeconomic variable. Let us assume $y_t = \tau_t + c_t + \epsilon_t$, where τ_t is the trend component of the series, c_t is the cyclical component of the series, and ϵ_t is the irregular or "noise" component. We apply the Hodrick–Prescott filter (Hodrick & Prescott,1997) twice to y_t to extract c_t. The filter is defined as the solution to the following optimization problem:

$$\{\hat{\tau}_t\} \equiv \text{arg}min_{\{\tau_t\}}\{\Sigma_{t=1}^{T}(y_t - \tau_t)^2 + \lambda * \Sigma_{t=2}^{T-1}[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2\} \tag{3.1}$$

In essence, the objective function above breaks down the filter into two objectives. On the one hand, it aims at minimizing the discrepancy between the raw series y_t and its latent trend component τ_t. On the other, it aims to make the series of latent trend component $\{\tau_t\}$ as smooth as possible. The parameter λ determines the relative importance of the two objectives. A larger λ puts more weight on the latter objective, resulting in a smoother trend series. A smaller λ leads to a trend series closer to the original series. There is little consensus on the value of λ. For example, Morten and Harald (2001) propose that $\lambda = 14400$ for monthly time series. Ravn and Uhlig (2002) propose another rule for choosing the value of λ. The Ravn–

Uhlig rule sets λ=1600p^4, where p is the number of periods per quarter. Thus, for monthly data, λ=1600$*$ 3^4=129,600. In this study, we adopt the Ravn–Uhlig rule in determining the value of λ.

We apply the Hodrick–Prescott filter to the raw variables in two steps. In the first step, we apply the filter to y_t with λ = 129,600 and calculate the detrended series $d_t \equiv y_t - \tau_t$. We then apply the filter a second time to d_t with a small λ to remove the noise component ϵ_t from d_t. The resulting series is the cyclical component c_t of the raw macroeconomic variables. We apply the above procedure to all raw macroeconomic variables except for financial market variables. The cyclical components of the raw macroeconomic variables are then used to construct the CRR state variables and the variable pool for the Elastic Net regressions.

3.2.1.2. The Elastic Net regression and construction of the machine learning (ML) factors

Kelly and Xiu (2023) emphasize that investors' information sets are of high dimension. Thus, traditional ad hoc sparse factor models such as the Fama-French three-factor model and Chan-Roll-Ross model may not be sufficient to describe the actual data-generating process of the returns. Within the realm of the ICAPM, this translates to the existence of a large number of state variables that jointly predict the future investment opportunity set. However, macroeconomic variables are often highly collinear due to complex feedback effects, simultaneity, and serial correlations, in which case the OLS would produce highly unstable coefficient estimates. The problem is exacerbated when we only have a short time series of the variables.

Thus, we also explore constructing macroeconomic factors from a large set of overseas and domestic macroeconomic variables following a procedure that involves the Elastic Net regression, a machine learning algorithm first introduced by Zou and Hastie (2005).

In the context of this study, the method aims at solving the following optimization problem:

$$\mathrm{L}(\alpha, \lambda) \equiv min_{\{\boldsymbol{\beta}\}}\left\{\frac{1}{T}\Sigma_{t=1}^{T}(y_{t+h}-\boldsymbol{x}_{\boldsymbol{t}}'\boldsymbol{\beta})^2+\lambda*\Sigma_{k=1}^{K}[(1-\alpha)\beta_k^2+\alpha|\beta_k|]\right\} \tag{3.2}$$

where y_{t+h} is either the standardized market return or volatility at time t+h (h=1,2,3), $\boldsymbol{x}_{\boldsymbol{t}}$ is the set of K standardized candidate forecasting variables whose values are known at time t. $\boldsymbol{\beta}$ is the $K \times 1$ vector of the penalized standardized coefficients on the forecasting variables whose k^{th} element is β_k.

The penalty function $\lambda * \Sigma_{k=1}^{K}[(1-\alpha)\beta_k^2+\alpha|\beta_k|]$ in (3.2) consists of both the L1 regulation of Lasso Regression and L2 regulation of Ridge Regression. α is the hyperparameter that controls the relative importance of the two regulations. When $\alpha = 0$, the Elastic Net turns into a Ridge Regression, and all estimated coefficients are non-zero. Moreover, when $\alpha = 1$, the Elastic Net becomes a Lasso Regression. λ is the hyperparameter that controls the overall importance of the penalty function. When $\lambda = 0$, the Elastic Net reduces to an OLS Regression. Hence, the Elastic Net can strike a balance between Ridge and Lasso Regression over the task of variable selection. On the one hand, contrary to Ridge Regression, which could only produce non-zero coefficient estimates, the Elastic Net can produce sparse solutions similar to Lasso Regression. Moreover, when regressors are highly collinear, the Elastic Net produces more robust solutions than Lasso Regression due to the grouping effect. As Zou and Hastie (2005) illustrated, a regression method possesses the grouping effect if the coefficient estimates on highly correlated variables (after standardization) tend to be equal. A sufficient condition for penalized regressions' grouping effect is the penalty function's strict convexity. For Elastic Net, $\lambda * \Sigma_{k=1}^{K}[(1-\alpha)\beta_k^2+\alpha|\beta_k|]$ in (3.2) is strictly convex when $0 < \alpha < 1$. Thus, cohorts of highly correlated macroeconomic forecasting variables in (3.2) should have penalized standardized coefficients of a similar order of magnitudes and get in or out of the final set of non-zero-coefficient predictors together. The same conjecture does not apply to Lasso Regression as its L1 penalty function is not strictly convex. Even in the most extreme case with identical regressors, Lasso Regression could not guarantee identical coefficient estimates on them.

We choose the value of the hyperparameters by minimizing the Cross-Validation Function (CV) following the procedure of Hastie et al. (2015). The procedure starts with randomly partitioning the whole sample into J folds. Then, for a given set of $\{\alpha, \lambda\}$ and each fold $j \in \{1,2,\ldots J\}$, estimate (3.2) using observations not in fold j and calculate the mean squared prediction error over fold j ($MSPE_j$). For each pair of $\{\alpha, \lambda\}$, The Cross-Validation Function (CV) is the average mean squared prediction error over the J subsamples.

$$\mathrm{CV}(\alpha, \lambda) = \frac{1}{J} \Sigma_{j=1}^{J} MSPE_j \tag{3.3}$$

$$\{\alpha^*, \lambda^*\} = \ \mathrm{arg}min_{\{\alpha,\lambda\}}\, \mathrm{CV}(\alpha, \lambda) \tag{3.4}$$

We then minimize the CV function over a range of values of $\{\alpha, \lambda\}$, resulting in an optimal set of solution $\{\alpha^*, \lambda^*, \beta^*\}$. Hastie et al. (2015) illustrate that only a few values of α are worth checking, whereas scrutiny over λ is needed.

The final set of macroeconomic state variables that survive the dimensionality reduction of the Elastic Net regressions could not be used as factors directly as it is the "shock/news" of the state variables that drive the variations in asset prices. The complete procedure to construct the ML factors involves the following steps:

1. We screened out 73 variables significantly correlated with future market volatility (up to three months ahead) and 39 variables significantly correlated with future market returns (up to a horizon of 3 months ahead) from a universe of 197 domestic and overseas macroeconomic variables. We do not include aggregate stock market

variables as in Welch and Goyal (2008), Gu et al. (2020), and Leippold et al. (2022) because we wish to estimate the first-order effect of real macroeconomic activities on stock prices.

2. We conduct Elastic Net predictive regressions with future market volatilities on the 39 variables and future market returns on the 73 variables. This results in a collection of 23 macroeconomic variables with non-zero penalized standardized coefficients in either regression.

3. We further classify the resulting 23 variables into 9 categories, 5 containing more than one variable. We then use the dynamic factor model illustrated in Section 1.3.1.1 of Chapter 1 to extract the latent common factor of the variables in those five categories, resulting in a final set of 9 domestic and overseas composite state variables.

4. We finally extract the "shock/news" about the future values of the nine composite state variables as the machine learning (ML) factors utilizing the economic tracking portfolios first introduced by Lamont (2001). The coefficients of the asset pricing models are then estimated by the GMM and Fama-Macbeth (1973) procedure, as in Vassalou (2003), Ang et al. (2006), and Adrian et al. (2014).

3.2.1.3. Economic tracking portfolios of macroeconomic state variables

After we construct nine composite state variables using the dynamic factor model, we extract the "shock/news" about the future values of the state variables as factors using economic tracking portfolios, a method we have adopted in Chapters 1 and 2. For each of the nine state variables chosen by the Elastic Net regressions, we run the following regression:

$$y_{k,t+1} = \boldsymbol{b}_k{}'\boldsymbol{R}_t + \boldsymbol{c}_k{}'\boldsymbol{Z}_{t-1} + \varepsilon_{k,t+h} \tag{3.5}$$

$y_{k,t+1}$ is the value of the k[th] composite state variable one month ahead. $\boldsymbol{R}_t$ contains the excess returns of six portfolios (S/V, S/M, S/G, B/V, B/M, and B/G) used to construct the modified Fama-French size and value factors, as illustrated in Section 1.4 of Chapter 1. $\boldsymbol{Z}_{t-1}$ is the set of control state variables, which include the yield spread between the 10-year government bond and the 1-year government bond (Term_spread1Y), the average default spread between the 3-year AA corporate bonds and the 3-year government bond (Default_spread), 1-month interbank repo rate (Dr1M), monthly market turnover (Turn_A) and monthly volatility on market return (Svar). The estimated $\boldsymbol{b}_k{}'\boldsymbol{R}_t$ is the economic tracking portfolio of news about future values of the composite macroeconomic state variables.

3.2.2. Test of model consistency with the ICAPM

We test whether the ML factors are consistent with restrictions imposed by the ICAPM by repeating the analysis in Table 2.4 with the machine learning state variables. We run the following predictive regressions:

$$y_{t+h} = \mathrm{c} + \boldsymbol{\beta}'\boldsymbol{x_t} + \varepsilon_{t+h} \tag{3.6}$$

where y_{t+h} is either the return or volatility on the HS300 stock index for month t+h (h=1,2,3), $\boldsymbol{x_t}$ is the vector of the raw levels of the machine learning state variables at month t. The sign and significance of each element of $\boldsymbol{\beta}$ are of importance. The ICAPM implies that state variables that positively forecast future market returns or negatively forecast future market volatilities should have positive risk premiums and vice versa.

3.2.3. GMM (Hansen,1982) estimation of asset pricing models

In line with Chapter 2, we also estimate the asset pricing models by the generalized method of moments (GMM) using two sets of moment conditions. The first set of moment conditions (moment conditions 1) follows the SDF approach (Vassalou, 2003; Cochrane, 2005; Detzel, 2017).

Moment conditions 1: $$\overline{\boldsymbol{m}}(\boldsymbol{b}) = \frac{1}{T}\Sigma_{t=1}^{T}\left[\left(1 - \boldsymbol{b}'\boldsymbol{f_t^d}\right)\boldsymbol{R_t^e}\right] = 0 \tag{3.7}$$

where $\boldsymbol{R_t^e}$ $L \times 1$ is the vector of excess returns on either of the three sets of test assets. $\boldsymbol{f_t^d}$= $[f_{1,t}^d,\ldots, f_{K,t}^d]'$ is the $K \times 1$ vector of demeaned machine learning factors, and $\boldsymbol{b}$ contains the loadings of the factors on the SDF. Moment condition 1 checks whether each factor provides unique pricing information (Cochrane, 2005; Feng et al., 2020).

The second set of moment conditions (henceforth denoted moment conditions 2) follows the expected return-covariance risk premium approach (Maio & Santa-Clara, 2012; Maio & Santa-Clara, 2017; Cooper & Maio, 2019).

Moment conditions 2 $$\overline{\boldsymbol{m}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{T}\Sigma_{t=1}^{T}\begin{cases} R_{1,t}^e - \Sigma_{k=1}^{K}\lambda_k R_{1,t}^e\left(f_{k,t} - \mu_k\right) \\ \vdots \\ R_{L,t}^e - \Sigma_{k=1}^{K}\lambda_k R_{L,t}^e\left(f_{k,t} - \mu_k\right) \\ f_{1,t} - \mu_1 \\ \vdots \\ f_{K,t} - \mu_K \end{cases} = 0 \qquad (3.8)$$

where $\boldsymbol{\lambda}$= $[\lambda_1,\ldots, \lambda_{\mathrm{K}}]'$ is the $K \times 1$ vector of factor covariance risk premiums. $\boldsymbol{\mu}$= $[\mu_1,\ldots, \mu_{\mathrm{K}}]'$ is the $K \times 1$ vector of factor means.

3.2.4. The Fama-Macbeth (1973) procedure and Shanken correction

We further estimate the factor risk premiums by the Fama-Macbeth (1973) two-step procedure. In the first step, we estimate the following time series regression:

$$R_{i,t} = c_i + \boldsymbol{\beta}_i{}'\boldsymbol{f_t} + \varepsilon_{i,t} \qquad (3.9)$$

where $R_{i,t}$ is the excess return of test asset i at time t, $\boldsymbol{f_t}$ is the vector of the time t realization of the machine learning factors. $\boldsymbol{\beta}_i$ is the vector of factor loadings of asset i on the machine learning factors.

In the second step, we estimate the following cross-sectional regression at each point in time by GLS:

$$R_{i,t} = a_t + \boldsymbol{\gamma_t}'\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}} + e_{i,t} \tag{3.10}$$

To correct the errors-in-variables problem in (3.10), we apply the Shanken (1992) correction to the standard errors of estimated risk premiums.

3.3. Data and variables

Consistent with Chapter 2, the sample period of this study is from December 2006 to December 2021. We adopt the same set of test assets as in Chapter 2 to make meaningful model comparisons. The test assets include size-neutral portfolios sorted by the E/P ratio and idiosyncratic volatility following the same procedure in Section 1.4 of Chapter 1. In line with Chapter 2, we also conduct our analyses on 31 value-weighted industry portfolios maintained by Shenwan Hongyuan Group Co., Ltd. Our initial sample includes all stocks listed in the Chinese A-share market. The data for individual stock returns and accounting information are from CSMAR. We exclude observations recorded within six months since the stock's IPO. We exclude stocks in the financial industry, and we further exclude stocks with special treatments (ST, *ST, and PT). Following Liu et al. (2019), we drop the smallest 30% of stocks from the sample. Such treatment would alleviate the reverse merger shell-value problem of small stocks caused by the IPO and delisting regulations for the Chinese stock market. Liu et al. (2019) found that 83% of the reverse mergers are performed over listed "shells" in the bottom 30% size category.

The data for the 197 domestic and overseas macroeconomic variables that serve as raw materials for the Elastic Net regressions come from several databases, including Wind, CSMAR, CEIC, RESSET, Qianzhan, CEINET, FRED (Federal Reserve Bank of St. Louis) and World Bank Open Data. We adopt the inequality index constructed by the dynamic factor model in Chapter 1 to measure the year-on-year change in inequality. Moreover, it is worth noting that we also include two versions of the news-based Economic Policy Uncertainty (EPU) Index by Baker et al. (2016) and Davis et al. (2019), as suggested by a series of studies including Pastor and Veronesi (2012), Brogaard and Detzel (2015), and Lee et al. (2021). We further include the Geopolitical Risk (GPR) Index motivated by Balcilar et al. (2018) and Caldara and Iacoviello (2022) and oil-related variables from the U.S. Department of Energy motivated by Kang and Ratti (2013), Sim and Zhou (2015) and Narayan and Gupta (2015). However, to our disappointment, none of those variables passes our filtering procedure introduced in Section 3.3.1.3

3.4. Empirical results

3.4.1. Exploratory data analysis and descriptive statistics

This subsection presents the results of exploratory data analysis and descriptive statistics about the machine learning factors. We show results of predictive regressions of future market returns and volatilities on macroeconomic variables using the Elastic Net algorithm. We present evidence that cohorts of macroeconomic variables with non-zero penalized standardized coefficients are suitable for further processing with dimensionality reduction methods. We also glimpse the relationship between the Fama-French and machine learning macroeconomic factors by their correlations.

We build machine-learning macroeconomic factors following the procedure demonstrated in Section 3.3.1.2. The procedure benefits from the advantages of data-driven machine learning algorithms when facing a large, highly collineated regressor set while preserving the theoretical intuition of classic asset pricing models. Our practice of marrying machine learning tools and asset pricing theories is emphasized by Kelly et al. (2019), Feng et al. (2020), and Nagel (2021). As noted by Feng et al. (2020), the practice also alleviates the multiple testing concerns of Harvey et al. (2016).

Table 3.1 presents the results of the multiple horizon Elastic Net predictive regressions in step 2 of the procedure described in Section 3.3.1.2. Columns 1, 2, and 3 show results of regressions for the market return at forecasting horizons of 1, 2, and 3 months ahead, respectively. Columns 4, 5, and 6 show results of regressions for the market volatility at forecasting horizons of 1, 2, and 3 months ahead, respectively. Only regressors with non-zero penalized standardized coefficients in at least one of the specifications are included in the table. The final forecasting variable set contains 23 out of 197 domestic and overseas macroeconomic variables. We further categorize the 23 variables into nine cohorts: Domestic Growth, Domestic Inflation, Domestic Liquidity, Credibility Risk, Inequality, Exchange Rate, Overseas Growth, Overseas Inflation, and Overseas Liquidity.

The Domestic Growth category consists of six variables, including the PMI for manufacturing industries (PMI_Manu), the PMI for non-manufacturing industries (PMI_Nonmanu), the Li Keqiang Index (Keqiang Index), a monthly estimate of aggregate economic growth advocated by former premier Li Keqiang (which equals 0.40×year-on-year growth of industrial electricity usage +0.35×year-on-year growth of medium/long term loan +0.25×year-on-year growth of volume of railway freight), the month-on-month change in the OECD Chinese Business Confidence Index (BCI_CN_MoM), the month-on-month change in vehicle production (Vehicle_MoM), and year-on-year change in Total Retail Sales (Retail Sale_YoY). The Domestic Inflation category consists of three variables, including the year-on-year change in the Consumer Price Index (CPI_YoY), the year-on-year change in the Producer Price Index (PPI_YoY), and the year-on-year change in the Retail Price Index (RPI_YoY). The Domestic Liquidity category consists of three variables, including the year-on-year change in M0 (M0_YoY), the year-on-year change in M2 (M2_YoY), and the yield to maturity of the 6-month government bond (Yield_6M). The Overseas Growth category consists of four variables, including the year-on-year change in the OECD G7 countries'

Business Confidence Index (BCI_G7_YoY), the year-on-year change in the OECD U.S. Business Confidence Index (BCI_US_YoY), the monthly recession probability of the U.S. by FRED (Rec_Prob_US) and the year-on-year change in the U.S. Economic Activity Coincident Index by ECRI (ECRI_US_YoY). The Overseas Inflation category consists of 3 variables, including the year-on-year change in the U.S. Consumer Price Index (CPI_US_YoY), the year-on-year change in the U.S. Producer Price Index (PPI_US_YoY), and the U.S. inflation expectation (Inflation_Exp_US).

We find evidence of grouping effect across different cohorts. In each column, the non-zero coefficients within each category tend to have the same sign and order of magnitude. Results show that most selected variables predict future market volatilities while only a few predict future market returns. An inspection of the out-of-sample R^2s confirms the difference between return and volatility predictability. At each forecasting horizon, the return regression has much lower out-of-sample R^2 than the volatility regression. For example, at the forecasting horizon of 1 month ahead, the volatility regression has an out-of-sample R^2 as high as 0.436 while that of the return regression is only 0.134. Also, note that the out-of-sample R^2 shrinks as the forecasting horizon increases.

Table 3.1. Penalized standardized coefficients of Elastic Net regressions

This table presents the non-zero penalized standardized coefficients of multiple horizon predictive regressions for the monthly return and volatility of the value-weighted market index. Columns 1, 2, and 3 show results of regressions for the market return at forecasting horizons of 1, 2, and 3 months ahead, respectively. Columns 4, 5, and 6 show results of regressions for the market volatility at forecasting horizons of 1, 2, and 3 months, respectively. The complete set of forecasting variables includes 73 macroeconomic variables for return forecasts and 39 macroeconomic variables for volatility forecasts. The definitions of the final set of forecasting variables in the table are documented in A.3 of the Appendix. The sample period is from 2007:M2 to 2021:M12.

	F.Rm	F2.Rm	F3.Rm	F.Vol	F2.Vol	F3.Vol
Variables						
Growth						
PMI_Manu	-	-	-	-2.670	-2.320	-
PMI_Nonmanu	-	-	-	-2.437	-2.024	-
Keqiang Index	-	-	-	-1.132	-1.280	-0.201
BCI_CN_MoM	0.358	0.208	0.140	-	-	-
Vehicle_MoM	0.236	0.373	-	-	-	-
Retail Sale_YoY	-	-	-	-1.359	-1.324	-
Inflation						
CPI_YoY	-0.239	-0.106	-0.080	3.032	2.977	1.543
PPI_YoY	-0.381	-0.341	-0.228	1.544	1.223	1.005
RPI_YoY	-0.330	-0.191	-0.084	2.512	2.744	1.223
Liquidity						
M0_YoY	0.212	-	-	1.302	1.618	1.541
M2_YoY	-	-	-	1.863	1.623	1.765
Yield_6M	-	-	-	-1.148	-0.993	-
Credibility						
Defy*	-0.323	-0.167	-0.054	3.024	2.033	1.423
Inequality						
Inequality_fa*	-0.487	-0.287	-0.138	1.322	0.585	-
Exchange Rate						
USD_RMB*	-	-	-	0.352	0.327	0.129
Overseas Growth						
BCI_G7_YoY	-	-	-	-1.304	-1.209	-1.033
BCI_US_YoY	-	-	-	-1.078	-1.141	-0.532
Rec_Prob_US	-	-	-	0.869	0.641	-
ECRI_US_YoY	-	-	-	-2.014	-2.002	-
Overseas Inflation						
CPI_US_YoY	-0.544	-0.221	-0.044	-	-	-
PPI_US_YoY	-0.244	-0.056	-	-	-	-
Inflation_Exp_US	-	-	-	3.436	1.243	-
Overseas Liquidity						
Fed*	-	-	-	2.369	1.032	-
Out-of-sample R^2	0.134	0.125	0.029	0.436	0.313	0.286

Table 3.2 presents the correlations among macroeconomic variables within each category. Panel A shows the correlations among the Domestic Growth variables. PMI_Manu, PMI_Nonmanu, Keqiang Index, and Retail Sale_YoY have relatively high pairwise correlations, while BCI_CN_MoM and Vehicle_MoM show weaker correlations with the rest of the variables. The cohort is a mixture of year-on-year and month-on-month measures. The former possesses the advantage of robustness as they are less prone to the influence of seasonality and measurement errors. Whereas the latter carries more prompt information about the state of the economy. The latent factor estimated by the dynamic factor model using those measures would benefit from both advantages.

Panel B shows the correlations among the Domestic Inflation variables. All three variables, CPI_YoY, PPI_YoY, and RPI_YoY, have decent correlations with each other. Panel C shows the correlations among the Domestic Liquidity variables. The absolute values of the pairwise correlations among the three variables within this category are relatively high except for that between M0_YoY and Yield_6M. Panel D shows the correlations among the Overseas Growth variables. Note that Rec_Prob_US has comparatively low correlations with all other variables in the cohort. Panel E shows the correlations among the Overseas Inflation variables. Similar to variables within the Domestic Inflation category, all three variables in this cohort demonstrate high correlations with each other. We then use the dynamic factor model to construct latent state variables that drive the co-movements of the variables within each category mentioned above. We reverse the signs of Yield_6M and Rec_Prob_US before applying the dynamic factor model. In line with the framework of Campbell (1996), the final set of factors that enters the asset pricing model is the news of the nine composite state variables, which we extract by the method of economic tracking portfolio.

Table 3.2. Correlations among variables with non-zero Elastic Net coefficients within each category

This table reports the correlations among variables from each category. Panel A shows the correlations among the Domestic Growth variables. Panel B shows the correlations among the Domestic Inflation variables. Panel C shows the correlations among the Domestic Liquidity variables. Panel D shows the correlations among the Overseas Growth variables. Panel E shows the correlations among the Overseas Inflation variables. The definitions of the state variables shown in the table are documented in A.2 of the Appendix. The sample period is from 2006:M12 to 2021:M12.

Panel A: Correlations among the Domestic Growth variables

	PMI_Manu	PMI_Nonmanu	Keqiang Index	BCI_CN_MoM	Vehicle_MoM
PMI_Manu	1.000				
PMI_Nonmanu	0.775	1.000			
Keqiang Index	0.680	0.501	1.000		
BCI_CN_MoM	0.248	0.088	-0.082	1.000	
Vehicle_MoM	0.169	0.045	-0.090	0.199	1.000
Retail Sale_YoY	0.414	0.657	0.415	-0.130	-0.088

Panel B: Correlations among the Domestic Inflation variables

	CPI_YoY	PPI_YoY	RPI_YoY
CPI_YoY	1.000		
PPI_YoY	0.476	1.000	
RPI_YoY	0.940	0.677	1.000

Panel C: Correlations among the Domestic Liquidity variables

	M0_YoY	M2_YoY	Yield_6M
M0_YoY	1.000		
M2_YoY	0.515	1.000	
Yield_6M	-0.189	-0.473	1.000

Panel D: Correlations among the Overseas Growth variables

	BCI_G7_YoY	BCI_US_YoY	Rec_Prob_US	ECRI_US_YoY
BCI_G7_YoY	1.000			
BCI_US_YoY	0.891	1.000		
Rec_Prob_US	-0.165	-0.053	1.000	
ECRI_US_YoY	0.578	0.338	-0.183	1.000

Panel E: Correlations among the Overseas Inflation variables

	CPI_US_YoY	PPI_US_YoY	Inflation_Exp_US
CPI_US_YoY	1.000		
PPI_US_YoY	0.900	1.000	
Inflation_Exp_US	0.609	0.506	1.000

Table 3.3 presents the correlations among the nine ML factors. We also add the tracking portfolio of the yield spread between the 10-year government bonds and 1-year government bonds (Term), which is included in the CRR model in Chapter 2. Domestic growth news mostly correlates with liquidity, overseas growth, exchange rate, and inequality news. Domestic inflation news has a high correlation of 0.784 with overseas inflation news. However, it has no significant correlation with domestic liquidity news, indicating that domestic inflation is most likely imported and related to price appreciation in the international commodities market. Default spread news, often regarded as a signal of a deteriorating economic environment, tightening credit market, and increasing risk aversion (Keim & Stambaugh, 1986; Fama & French, 1988; Gertler et al., 1991; Cooper et al., 2022), is mildly negatively correlated with domestic and overseas growth and inflation news, and exchange rate news (positive value of USD_RMB means news about depreciation of RMB). Inequality news is positively correlated with domestic growth and inflation, suggesting that the risk of social injustice lurks and accumulates as the economy expands. Overall, the correlations among domestic factors are, on average, smaller than those among the Fama-French factors in absolute values, as shown in Table 2.2 of Chapter 2. The term spread news correlates most positively with domestic growth and liquidity news. The result is unsurprising as previous studies often regard a widening term spread as a "good sign" of the economy. Fama and French (1989) and Hahn and Lee (2006) note that the term spread tends to narrow down near peaks of business cycles and widen near the troughs. However, assuming a clear-cut linear relationship between term spread and economic growth is difficult. To see this, note that term spread is the differential between the long-term and short-term yields. The latter is mainly driven by monetary policy (Feroli, 2004; Hahn & Lee, 2006), while the former is mainly driven by investors' expectations of long-term economic growth and inflation according to the expectations hypothesis of term spread (See, for example, Mishkin,1988; Mishkin,1990;

Hardouvelis, 1994; Diebold et al., 2006; Ludvigson & Ng, 2009). Consider a stylized 4-phase business cycle classification; at the beginning of the recession phase, the market's adjustment to news about impending economic slowdown precedes monetary policy interference, and the long-term rate starts to drop, dragging the term spread down. When the economy continues to slow down and enters the depression phase, expansionary monetary policy is usually introduced, and the interest rates start to drop, with short-term rates dropping at a higher speed, pushing the term spread up. The economy then enters the recovery phase, during which positive news about the future economy comes in with stable monetary policy. Thus, the term spread widens up further. When the economy continues to prosper and heads towards overheating, the government tends to interfere with precautionary contractionary monetary policy, flattening the yield curve once again. Thus, the yield spread can widen during either economic expansion or contraction with monetary policy remains stable. The flight to quality effect (Longstaff, 2004) and the central bank information effect (Bu et al., 2021), which may lead to contradictory discount rate and cash flow news (Campbell & Vuolteenaho, 2004), also add more complexity to the relationship between the term spread and economic prospect. The simple model mentioned above depicts the movement of the Chinese term spread relatively well before 2012 and lost much of its explanatory power afterward due to the introduction of new monetary policy objectives such as mitigating systematic financial risks. In all, although the positive term spread news is sometimes considered an indicator of a “good state” of the economy by previous studies, its true information substance remains blurry. Also, a relatively high correlation between the domestic and foreign factors exists. We will show that the foreign factors are superfluous when augmented with the domestic factors in tests of asset pricing models. One explanation is that information about the foreign factors may have already been transmitted to the domestic factors through channels like the exchange rate. Thus, we exclude the foreign factors from subsequent studies in this section.

Table 3.3. Correlations among the machine learning macroeconomic factors

This table reports the correlations among machine learning macroeconomic factors. Growth, Inflation, and Liquidity are domestic growth, inflation, and liquidity factors constructed by the dynamic factor model and economic tracking portfolio. Growth_F and Inflation_F are overseas growth and inflation factors constructed by the dynamic factor model and economic tracking portfolio. USD_RMB, Fed, Defy, and Term are economic tracking portfolios of the change in the exchange rate of USD to RMB, the Federal Funds Rate, the default spread, and the term spread, respectively. Inequality is the economic tracking portfolios of the inequality index in Chapter 1 constructed by the dynamic factor model. The sample period is from 2006:M12 to 2021:M12.

	Growth	Inflation	Liquidity	Growth_F	Inflation_F	USD_RMB	Fed	Defy	Inequality
Growth	1.000								
Inflation	0.242	1.000							
Liquidity	0.374	0.031	1.000						
Growth_F	0.365	0.121	-0.129	1.000					
Inflation_F	0.237	0.784	-0.108	0.586	1.000				
USD_RMB	-0.430	-0.325	-0.026	-0.072	-0.266	1.000			
Fed	0.185	0.215	0.252	0.157	0.188	0.087	1.000		
Defy	-0.148	-0.111	-0.113	-0.124	-0.134	0.251	-0.238	1.000	
Inequality	0.327	0.248	0.365	0.203	0.132	-0.230	0.196	0.108	1.000
Term	0.253	-0.133	0.288	0.125	0.077	-0.039	-0.093	-0.082	-0.046

We further present the correlation between the ML factors and Fama-French factors in Table 3.4. Growth is positively correlated with the Fama-French size factor and negatively correlated with the value and profitability factor. The result is consistent with our expectation that smaller and weaker firms perform comparatively better in good times. It also echoes the studies by Petkova (2006) and Ali et al. (2018), who show that the Fama-French size and value factors contain news about future GDP growth in both the U.S. and emerging markets. Liquidity news is most positively correlated with the market return and investment factor and negatively correlated with the value factor. Similar to the result in Chapter 1, inequality news is most negatively correlated with the value factor.

The default spread news negatively correlates with the market return and size factor and positively correlates with the value and investment factors. The result coincides with several previous studies. Chan and Chen (1991) postulate that small firms tend to be marginal firms and, thus, are less likely to survive in adverse economic environments. Perez-Quiros and Timmermann (2000), Gertler and Gilchrist (1994), and Hahn and Lee (2006) argue that small firms with little collateral face more binding credit constraints during bad times and thus should have higher sensitivity to the default spread. Their findings support a negative correlation between the size and default spread factors. Chen and Zhang (1998), Hahn and Lee (2006), and Petkova (2006) also note that the value factor may be a measure of distress risk and value firms tend to have high leverage and cash flow problems, and thus the value factor may also be negatively correlated with the default spread factor. However, that intuition contradicts our result as VMG and Defy have a significant positive correlation of 0.25. The phenomenon may be attributed to the fact that small firms tend to be growth firms in the Chinese market.

Table 3.4. Correlations between the machine learning and Fama-French factors
This table presents the correlations between the ML factors and Fama-French five factors. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Defy is the mimicking portfolio of news about the default spread. Inequality is the mimicking portfolio of news about inequality growth. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor, as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The sample period is from 2006:M12 to 2021:M12.

	Growth	Inflation	Liquidity	Defy	Inequality
MKT	0.064	-0.085	0.366	-0.344	-0.134
SMB	0.341	-0.215	0.161	-0.353	-0.163
VMG	-0.372	0.243	-0.315	0.250	-0.271
RMW	-0.433	0.304	-0.199	0.079	0.156
CMA	-0.075	0.276	0.317	0.316	-0.215

3.4.2. Predictive regressions of the machine learning state variables

Following Maio and Santa-Clara (2012), we run multiple horizon predictive regressions of monthly market return and volatility on the machine learning state variables. The ICAPM postulates that shocks on state variables that positively forecast future market returns or negatively forecast future market volatiles should have a positive risk premium. And shocks on state variables that negatively forecast future market returns or positively forecast future market volatiles demand negative risk premiums.

Table 3.5 reports the estimation results using the machine learning composite state variables as forecasting variables. Growth* is the state variable representing domestic growth. Inflation* is the state variable representing domestic inflation. Liquidity* is the state variable representing domestic liquidity. Defy* is the default spread. Inequality_fa* is the inequality index constructed by the dynamic factor model in Chapter 1. Growth* significantly negatively forecasts market volatilities at horizons of 1,2, and 3 months ahead. Inflation* significantly negatively forecasts market returns at horizons of 1 and 2 months ahead and significantly positively forecasts market volatilities at horizons of 1,2 and 3 months ahead. Liquidity* significantly positively forecasts market volatilities at the horizon of 1 month. Defy* significantly positively forecasts market volatilities at horizons of 1,2 and 3 months ahead in the presence of other ML state variables. Inequality_fa* significantly negatively forecasts market returns at the horizon of 1 month ahead and significantly positively forecasts market volatilities at horizons of 1,2 and 3 months ahead. Thus, according to the ICAPM, innovations of Growth* should have a positive risk premium, and innovations of the rest of the state variables should have negative risk premiums. The result is generally consistent with our expectations. The only exception is Liquidity*: A positive shock in liquidity is usually considered a "good state" of the economy, and we expect it to have a positive risk premium. However, since it positively forecasts market volatilities, it demands a negative risk premium stipulated by the ICAPM.

Overall, similar to the results in Table 3.1, future market volatilities are more predictable than future returns. The volatility regressions generally have larger R^2s and adjusted R^2s than return regressions with the same forecasting horizon. Also, similar to Table 3.1, the predictive power decreases with the forecasting horizon. This phenomenon echoes the results in Table 2.4 of Chapter 2. However, it contrasts with the results from previous studies on the U.S. market, such as those by Maio and Santa-Clara (2012), who document increased return predictability with forecasting horizons. Once again, the phenomenon may be attributed to the fact that the Chinese stock market is dominated by retail investors who are more myopic (Cuthbertson et al., 1997; Docherty & Hurst, 2018; Liu et al., 2019). One last finding from Table 3.5 is that the ML state variables have better predictive powers than the CRR factors (the results for predictive regressions of the CRR factors are presented in Table 2.4 of Chapter 2). For regressions with the same dependent variable and forecasting horizon, the regressions on the ML state variables generally have higher R^2s and adjusted R^2s, and higher portions of statistically significant regressors. The result indicates that the dynamic factor model employed in constructing the ML state variables captures incremental pricing information from different metrics of macroeconomic risks better than traditional ad hoc macro factors.

Table 3.5. Predictive regressions of the machine learning state variables
This table presents the results of multiple horizon predictive regressions for the monthly returns and volatilities of the value-weighted market index. The predictive variables are current levels of state variables corresponding to the ML factors. Columns 1, 2, and 3 show results of regressions for the market returns at forecasting horizons of 1, 2, and 3 months ahead, respectively. Columns 4, 5, and 6 show results of regressions for the market volatilities at forecasting horizons of 1, 2, and 3 months respectively. Growth* is the state variable representing domestic growth. Inflation* is the state variable representing domestic inflation. Liquidity* is the state variable representing domestic liquidity. Defy* is the default spread. Inequality_fa* is the inequality index constructed by the dynamic factor model in Chapter 1. The t-statistics corrected for serial correlation and heteroskedasticity using the Newey and West (1987a) estimator are in parentheses. The sample period is from 2007:M2 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1) F.Rm	(2) F2.Rm	(3) F3.Rm	(4) F.Vol	(5) F2.Vol	(6) F3.Vol
Growth*	0.231	-0.143	-0.559	-4.332***	-4.239***	-1.994**
	(1.33)	(-0.12)	(-0.48)	(-2.89)	(-2.76)	(-2.44)
Inflation*	-0.658***	-0.555**	-0.227*	3.756***	3.015**	1.055
	(-2.98)	(-2.33)	(-1.82)	(2.77)	(2.02)	(1.36)
Liquidity*	0.165	0.044	0.077	3.352**	1.688*	0.107
	(1.43)	(0.30)	(0.53)	(2.08)	(1.86)	(0.15)
Defy*	-0.831	-0.423	-0.307	15.080***	15.051***	12.068***
	(-1.51)	(-0.73)	(-0.56)	(3.83)	(3.57)	(2.76)
Inequality_fa*	-0.688***	-0.439**	-0.264*	2.461***	1.925***	1.343
	(-2.87)	(-2.29)	(-1.88)	(3.47)	(2.76)	(1.54)
_cons	0.063	0.259	-0.134	3.665	5.552	2.567
	(0.04)	(0.26)	(-0.08)	(0.33)	(0.50)	(0.23)
R^2	0.099	0.067	0.039	0.231	0.194	0.159
adj. R^2	0.068	0.035	0.014	0.202	0.157	0.121

3.4.3. Estimation of factor loadings on the SDF by GMM using Moment conditions 1

This subsection presents the estimation results of the asset pricing model with machine learning factors by GMM using moment conditions 1 in Section 3.3.3. The tests are conducted over 25 value-weighted double-sorted portfolios on size and E/P, 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility, and 31 value-weighted industry portfolios. The models in the stochastic discount factors (SDF) form are estimated by optimal GMM with the Newy-West kernel. We also include the result of Hansen's (1982) test of overidentifying restrictions for each model specification. In contrast with tests of the expected return-beta models, tests with the SDF aim to detect the factors' incremental pricing power.

3.4.3.1. Tests on 25 Size-E/P portfolios by GMM using moment conditions 1

Table 3.6 presents the estimation results over the 25 Size-E/P portfolios. Panel A shows the estimated coefficients of the model with the five domestic ML factors only. All five factors are significant at 1% significance level, indicating that each contains unique pricing information for the 25 Size-E/P portfolios. Panel B shows the estimated coefficients of the model with domestic ML factors augmented with the market excess return. Note that Liquidity is no longer significant at the 5% level. The results indicate that Liquidity contains similar information as the market factor. Panel C shows the estimated coefficients of model with the domestic ML factors augmented with the market excess return and overseas ML factors. Results show that the overseas ML factors are not statistically significant in the presence of the domestic factors. However, compared to the results in Panel A, the significance of the domestic factors also declines. Growth, Inflation, and Liquidity are no longer significant at the 5% level, which indicates a multicollinearity problem. The results suggest that the pricing information in the overseas factors is incorporated into the price of the test portfolios through the domestic factors. We also tested whether the overseas factors were spurious on the other two sets of test portfolios and obtained qualitatively similar results. Therefore, we refrain from the overseas factors in the rest of the analysis. Previous studies have long considered term spread as a vital state variable, and it is incorporated into the CRR model. As such, we also test whether the term spread shock contains incremental information against the ML factors. Panel D shows estimated coefficients of the domestic ML factors augmented with the market excess return and term spread news. The coefficient on Term is not statistically significant, and the significance of Inflation and Liquidity drops compared to the result in Panel A. The results suggest that the pricing information of term spread shock is already embedded in the ML factors and market return. The results from Table 3.6 also show that the J-statistics of all model specifications are insignificant, indicating no sign of model misspecification. However, Newey (1985) notes that the tests of overidentifying restrictions are prone to Type II errors. Thus, for robustness, we will further assess the models using the HJ-distance and ΔJ test in subsequent sections.

Table 3.6. Tests with the ML factors on 25 Size-E/P portfolios

This table presents the estimated coefficients of the ML factors by GMM using moment conditions 1 in Section 3.3.3. The test assets are 25 value-weighted double-sorted portfolios on size and the E/P ratio. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Defy is the mimicking portfolio of news about the default spread. Inequality is the mimicking portfolio of news about inequality. Growth_F is the machine learning overseas growth factor. Inflation_F is the machine learning overseas inflation factor. USD_RMB is the mimicking portfolio of news about the exchange rate of USD to RMB. Fed is the mimicking portfolio of news about the Effective Federal Funds Rate. The estimated parameters are factor loadings on the SDF. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

Panel A: Domestic ML factors						
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	0.051***	-0.108***	0.044**	-0.082***	-0.054***	-
z-stat	5.90	-4.43	2.32	-3.62	-7.54	-
J-stat	8.091					
p-value	0.991					

Panel B: Domestic ML factors + MKT						
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	0.042***	-0.076**	0.024*	-0.057**	-0.041***	0.043*
z-stat	3.41	-2.47	1.79	-2.33	-4.94	1.92
J-stat	8.085					
p-value	0.986					

Table 3.6. Tests with ML factors on 25 Size-E/P portfolios *(continued)*

Panel C: Domestic ML factors + MKT+ overseas ML factors										
	Growth	Inflation	Liquidity	Defy	Inequality	Growth_F	Inflation_F	USD_RMB	Fed	MKT
Estimate	-0.036*	-0.162*	0.037*	-0.052**	-0.047***	0.019	-0.032	-0.025	-0.041	-0.053*
z-stat	-1.81	-1.75	1.82	-2.03	-5.32	0.57	-0.36	-1.18	-1.46	-1.87
J-stat	7.734									
p-value	0.934									

Panel D: Domestic ML factors + MKT+ Term							
	Growth	Inflation	Liquidity	Defy	Inequality	MKT	Term
Estimate	-0.044**	0.071*	-0.028	-0.087**	-0.063***	-0.061	0.034
z-stat	-2.03	1.81	-1.05	-2.38	-3.15	-1.16	0.98
J-stat	8.085						
p-value	0.986						

3.4.3.2. Tests on 25 Size-IVOL portfolios by GMM using moment conditions 1

Table 3.7 presents estimation results over the 25 Size-IVOL portfolios. Panel A shows the the estimated coefficients on the model with the five domestic ML factors. All factors are significant at 1% significance level except for Growth, which is only significant at the 10% level. Panel B shows estimated coefficients of the model with the domestic ML factors augmented with the market excess return. MKT is superfluous in the presence of ML factors, and the estimated coefficients for the ML factors are qualitatively unchanged compared with those in Panel A. The results record a drop in the incremental pricing ability of the Growth factor compared with its performance over the 25 Size-E/P portfolios. Similar to the results over the 25 Size-E/P portfolios, Table 3.7 also shows insignificant J-statistics for both model specifications.

Table 3.7. Tests of asset pricing models on 25 Size-IVOL portfolios

This table presents the estimated coefficients of the ML factors by GMM using moment conditions 1 in Section 2.3.3. The test assets are 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility. MKT is the market excess return. SMB is the modified Fama-French size factor. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about inequality. The estimated parameters are factor loadings on the SDF. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

Panel A: ML factors						
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	0.031*	-0.331***	-0.035***	-0.027***	-0.108***	-
z-stat	1.82	-8.03	-3.94	-2.86	-12.63	-
J-stat	7.885					
p-value	0.992					

Panel B: ML factors +MKT						
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	0.029*	-0.279***	-0.053***	-0.022***	-0.103***	0.036
z-stat	1.67	-7.66	-4.92	-2.61	-8.14	0.84
J-stat	7.856					
p-value	0.988					

3.4.3.3. Tests on 31 industry portfolios by GMM using moment conditions 1

Table 3.8 presents estimation results over the industry portfolios. Panel A shows the estimated coefficients of the model that incorporates the five domestic ML factors only. All factors are significant at 1% significance level except for Liquidity. Panel B shows the estimated coefficients of the domestic ML factors augmented with the market excess return. The coefficient estimates on the ML factors are qualitatively unchanged with the inclusion of MKT. Overall, the results show that most macroeconomic factors contain unique pricing information for the industry portfolios. Similar to results over the 25 Size-E/P and 25 Size-IVOL portfolios, Table 3.8 also shows insignificant J-statistics for both model specifications.

Table 3.8. Tests of asset pricing models on 31 industry portfolios
This table presents the estimated coefficients of the ML factors by GMM using moment conditions 1 in Section 3.3.3. The test assets are 31 value-weighted industry portfolios. MKT is the market excess return. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about inequality. The estimated parameters are factor loadings on the SDF. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

Panel A: ML factors						
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	-0.023***	-0.087***	0.017	-0.108***	-0.019***	-
z-stat	-5.06	-6.34	1.54	-7.52	-4.85	-
J-stat	8.126					
p-value	0.999					

Panel B: ML factors +MKT						
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	-0.036***	-0.053***	-0.023	-0.091***	-0.016***	0.040*
z-stat	-5.36	-4.37	-1.18	-5.29	-3.12	1.91
J-stat	8.052					
p-value	0.999					

3.4.4. Estimation of factor risk premiums by the Fama-Macbeth (1973) procedure

The GMM tests with moment conditions 1 aim to answer whether each factor is spurious; however, whether they are priced remains to be investigated. In this subsection, we show estimation results of the factor risk premiums following the Fama-Macbeth (1973) procedure over different test assets. We calculate the t-statistics of the estimated risk premiums using Shanken's (1992) standard errors corrected for the errors-in-variables bias.

3.4.4.1. Risk premiums estimation on 25 Size-E/P portfolios

Table 3.9 presents the estimation results for the ML factors on 25 Size-E/P portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the ML factors and ML factors augmented with the market excess return, respectively. All ML factor risk premiums in Columns (6) and (7) are statistically significant at the 5% level except for those on Liquidity, which are marginally significant at the 10% level. In column (7), the risk premium on MKT is marginally significant at the 10% level. However, including MKT does not lead to qualitatively different risk premium estimates for the ML factors. Growth earns an estimated premium of 0.269, Inflation earns an estimated premium of -0.202, Liquidity earns an estimated premium of 0.337, Defy earns an estimated premium of -0.363, and Inequality earns an estimated premium of -0.293. The signs on the factor risk premiums coincide with economic rationale as positive growth and liquidity news are expected to signal a “good state” of the economy ahead. In contrast, news about an increase in inflation, default spread, and inequality signals a “bad/riskier state” of the economy ahead. The signs on the risk premiums are also consistent with the implications of the ICAPM, except for that on the liquidity risk premium. In Table 3.5, Liquidity positively forecasts future market volatility and thus should earn a negative risk premium according to the ICAPM.

Table 3.9. Risk premiums for ML factors on 25 Size-E/P portfolios

This table presents the estimated risk premiums of the ML factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about inequality. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
Growth	0.171					0.303***	0.269**
	(1.34)					(2.86)	(2.10)
Inflation		-0.111***				-0.296***	-0.202***
		(-2.73)				(-3.09)	(-2.83)
Liquidity			0.325**			0.399*	0.337*
			(2.11)			(1.93)	(1.74)
Defy				-0.136		-0.414***	-0.363**
				(-1.37)		(-3.06)	(-2.12)
Inequality					-0.299**	-0.312***	-0.293***
					(-2.39)	(-4.57)	(-3.30)
MKT							0.525*
							(1.71)
cons	0.601	1.105*	0.718	0.489	-0.136	3.177***	2.897**
	(1.19)	(1.93)	(1.16)	(0.62)	(-0.21)	(2.99)	(2.13)
R^2	0.315	0.320	0.221	0.217	0.320	0.709	0.732
adj. R^2	0.285	0.291	0.187	0.183	0.291	0.658	0.680

3.4.4.2. Risk premiums estimation on 25 Size-IVOL portfolios

Table 3.10 presents the estimation results for the ML factors on 25 Size-IVOL portfolios. Columns (1)-(5) present the results of single-factor regressions, and Columns (6) and (7) present the results of multiple regressions on the ML factors and the ML factors augmented with the market excess return, respectively. All ML factor risk premiums in Columns (6) and (7) are statistically significant at the 5% level. In column (7), the risk premium on MKT is marginally significant at the 10% level. And including MKT does not lead to qualitatively different risk premium estimates for the ML factors. The signs and magnitudes of the factor risk premiums in Columns (6) and (7) are consistent with those on 25 Size-E/P portfolios and economic rationales. Similar to the results on 25 Size-E/P portfolios, the signs of the risk premiums are also consistent with the implications of ICAPM except for that on the liquidity risk premium.

Table 3.10. Risk premiums for ML factors on 25 Size-IVOL portfolios

This table presents the estimated risk premiums of the ML factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and idiosyncratic volatility. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about inequality. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
Growth	0.187					0.438***	0.306**
	(0.66)					(3.36)	(2.28)
Inflation		-0.105**				-0.233**	-0.205**
		(-2.35)				(-2.33)	(-2.14)
Liquidity			0.157**			0.207**	0.197**
			(2.12)			(2.15)	(2.05)
Defy				0.152		-0.597**	-0.565**
				(0.83)		(-2.33)	(-2.07)
Inequality					-0.117**	-0.404***	-0.326**
					(-2.41)	(-2.76)	(-2.17)
MKT							0.507*
							(1.72)
cons	1.224	1.403*	-2.549**	0.351	-0.893	-2.617***	-2.294**
	(1.57)	(1.82)	(-1.99)	(0.37)	(-0.93)	(-2.86)	(-2.47)
R^2	0.339	0.344	0.187	0.306	0.356	0.663	0.691
adj. R^2	0.298	0.315	0.151	0.276	0.329	0.574	0.588

3.4.4.3. Risk premiums estimation on 31 industry portfolios

Table 3.11 presents the estimation results for the ML factors on 31 industry portfolios. Columns (1)-(5) present results of single-factor regressions, and Columns (6) and (7) present results of multiple regressions on the ML factors and the ML factors augmented with the market excess return, respectively. The results are qualitatively similar to those in Tables 3.9 and 3.10, except that the sign of the inflation risk premium is reversed.

Results in Tables 3.9, 3.10, and 3.11 show that the performance of the ML factors is generally consistent over different test portfolios, in contrast to those of the PCA and CRR factors in Chapter 2. Besides, most of the ML factors are consistent with the implications of the ICAPM except for Liquidity. The state variable corresponding to the liquidity factor positively forecasts future market volatilities, thus, Liquidity should earn a negative risk premium as the ICAPM implies. To reconcile the problem, we could undertake explorations in future studies in two directions. Firstly, we may differentiate between "good" and "bad" market volatilities by employing realized semivariances (Barndorff-Nielsen et al., 2010; Patton & Sheppard, 2015; Bollerslev, 2020). Increases in realized volatility may not stand for deteriorated investment opportunities. Secondly, we may seek better proxies for the return on the wealth portfolio (Mayers, 1973; Roll, 1977).

Table 3.11. Risk premiums for the ML factors on 31 industry portfolios

This table presents the estimated risk premiums of the ML factors by the Fama-Macbeth (1973) procedure. The test assets are 31 value-weighted industry portfolios. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about inequality. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret	Ret	Ret	Ret	Ret	Ret	Ret
Growth	0.104					0.221**	0.217**
	(1.36)					(2.16)	(2.06)
Inflation		0.076				0.102**	0.075**
		(1.25)				(2.29)	(1.98)
Liquidity			0.127			0.084	0.096*
			(1.51)			(1.54)	(1.75)
Defy				-0.201		-0.364**	-0.291**
				(-1.41)		(-2.41)	(-2.07)
Inequality					0.067	-0.644**	-0.519**
					(0.83)	(-2.33)	(-2.20)
MKT							-0.356
							(-0.53)
cons	0.784	0.811	0.879	0.265	0.842	-0.235	0.150
	(1.63)	(1.48)	(1.61)	(0.33)	(1.23)	(-0.35)	(0.22)
R^2	0.137	0.097	0.148	0.136	0.108	0.405	0.434
adj. R^2	0.107	0.066	0.119	0.106	0.078	0.387	0.418

3.4.5. Factor comparison

In this subsection, we assess the relationship between the ML and fundamental factors and compare the performance of the ML factors with the two sets of macroeconomic factors in Chapter 2 (the PCA and CRR factors) by three different metrics.

Table 3.12 presents the Newey and West (1987b) ΔJ statistics and the corresponding p-values for the restricted and unrestricted models estimated by GMM using moment conditions 1. Over each set of test assets, the restricted model is the model with the ML factors only, and the unrestricted model is the model augmenting the ML factors with the Fama-French five factors. The test results indicate whether the Fama-French factors are spurious in the presence of the ML factors. Results show that all ΔJ-statistics over all sets of test portfolios are statistically insignificant at the 5% significance level, suggesting that the Fama-French five factors do not contain sufficient incremental pricing information over that embedded in the ML factors.

Table 3.12. Results of ΔJ tests for factor redundancy
This table presents the Newey and West (1987b) ΔJ statistics and the corresponding p-values for each pair of restricted and unrestricted models. The restricted model contains the ML factors only, and the unrestricted model augments the ML factors with the Fama-French five factors. The tests examine whether the Fama-French five factors are redundant in the presence of the ML factors. The sample period is from 2006:M12 to 2021:M12.

	Size-E/P	Size-IVOL	Industry
ΔJ statistics	0.863	0.731	0.517
p-value	(0.973)	(0.981)	(0.992)

We then compare the pricing errors of different asset pricing models. Table 3.13 presents measurements of two pricing error metrics. For convenience, we also include in Table 3.13 the results for the Fama-French, PCA, and CRR factors in Chapter 2. Panel A shows the estimated HJ-distance calculated by 1-step GMM using moment conditions 1 with the Hansen and Jagannathan (1997) weighting matrix. Over 25 Size-E/P portfolios, the ML factors outperform all other sets of factors with an HJ-distance of 0.319, followed by the Fama-

French five factors. The ML factors maintain their position as the best-performing set of factors over 31 industry portfolios, where all sets of factors show deteriorated performances compared with those over 25 Size-E/P portfolios. In contrast, the CRR factors have the worst performance among all sets of factors over the industry portfolios. The ML factors continue to show robust performance over 25 Size-IVOL portfolios with an HJ-distance of 0.485, second only to the CRR factors.

Panel B presents measurements of the average pricing error as in Cooper et al. (2022), which equals the square root of the mean squared cross-sectional residuals from the second step of the Fama-Macbeth procedure. The relative performance of the five sets of factors is similar to that measured by the HJ-distance, with only minor discrepancies. Note that the ML factors now become the best-performing set of factors over all three sets of test portfolios judged by the average pricing error. The relative performance of the PCA and CRR factors still varies across test assets. For example, the PCA factors become the worst-performing factors over the Size-IVOL portfolios, and the CRR factors come at the bottom over the industry portfolios.

The results in Table 3.13 demonstrate that the ML factors perform on par with the Fama-French five factors and have the most consistent performance pricing all three sets of assets among the three sets of macroeconomic factors. This result indicates that compared with the ad hoc sparse macroeconomic factors, such as the CRR factors, more useful pricing information has been incorporated into the ML factors.

Table 3.13. Comparison of pricing errors
This table compares measurements of two pricing error metrics from competing asset pricing models. Panel A presents results about the HJ-distance derived from a one-step GMM. Panel B presents results about the average pricing error derived from the Fama-Macbeth (1973) procedure. Each row shows the pricing error measurements of different factors on one of the three sets of test portfolios: 25 Size-E/P portfolios, 25 Size-IVOL portfolios, and 31 Industry portfolios. The sample period is from 2006:M12 to 2021:M12.

Panel A: HJ-distance					
	FF 3 factors	FF 5 factors	PCA factors	CRR factors	ML factors
Size-E/P	0.425	0.389	0.395	0.413	0.319
Size-IVOL	0.781	0.622	0.704	0.297	0.485
Industry	0.595	0.562	0.586	0.601	0.367

Panel B: Average pricing error					
	FF 3 factors	FF 5 factors	PCA factors	CRR factors	ML factors
Size-E/P	1.587	1.394	1.442	1.471	1.362
Size-IVOL	1.701	1.448	1.730	1.417	1.348
Industry	2.634	1.908	1.626	3.056	1.582

3.5. Robustness checks

3.5.1. Estimation of covariance risk premiums by GMM

In line with Chapter 2, we supplement the results for the estimated factor loadings in Section 3.4.3 with covariance risk premiums estimated by optimal GMM using moment conditions 2. According to Cochrane (2005), theoretically, elements of the factor covariance risk premiums $\boldsymbol{\lambda}$ in (3.8) should be proportional to the elements of the factor loadings $\boldsymbol{b}$ in (3.7) scaled by the variance of corresponding factors. Thus, in theory, $\boldsymbol{\lambda}$ and $\boldsymbol{b}$ should have the same sign and significance.

Table 3.14 presents the estimated factor covariance risk premiums on 25 Size-E/P portfolios. Contrary to the case of the PCA and CRR factors in Chapter 2, MKT shows marginal significance in some specifications of the SDF and Fama-Macbeth tests when augmented with the ML factors. And in Section 3.4.4, the signs on the factor risk premiums are generally consistent with the implications of the ICAPM (except for the liquidity risk premium). Thus, we retain MKT in the final model specification.

The results are generally consistent with those in Panel B of Table 3.6. The coefficients on Growth, Inflation, Defy, and Inequality are significant and have the same sign as the corresponding loadings on the SDF. The coefficients on Liquidity and MKT have reversed signs; however, consistent with the results in Table 3.6, they are not statistically significant. The J-statistics of the Hansen's (1982) test of overidentifying restrictions for the model is insignificant at the 5% significance level, and we do not find strong evidence of model misspecification.

Table 3.14. Factor covariance risk premiums on 25 Size-E/P portfolios
This table presents the estimated covariance risk premiums of the ML factors by GMM using moment conditions 2 in Section 3.3.3. The test assets are 25 Size-E/P portfolios. MKT is the market excess return. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about the change in inequality. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	Covariance risk premiums of ML factors					
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	0.613***	-0.435**	-0.367	-1.082***	-0.721***	-1.592*
z-stat	7.75	-2.38	-0.93	-3.53	-4.29	-1.78
J-stat	8.152					
p-value	0.834					

Table 3.15 presents the estimated factor covariance risk premiums on 25 Size-IVOL portfolios. The results are generally consistent with those in Panel B of Table 3.7. The signs of the estimated covariance risk premiums align with those of the corresponding factor loadings on the SDF. The only qualitative difference is that the coefficient on Liquidity is no longer significant at the 5% level, whereas that on Growth becomes significant.

Table 3.15. Factor covariance risk premiums on 25 Size-IVOL portfolios
This table presents the estimated covariance risk premiums of the ML factors by GMM using moment conditions 2 in Section 3.3.3. The test assets are 25 Size-IVOL portfolios. MKT is the market excess return. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about the change in inequality. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	Covariance risk premiums of ML factors					
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	1.027***	-0.691**	-0.386*	-0.972***	-0.451**	1.356
z-stat	7.27	-2.43**	-1.74	-5.02	-2.38	1.62
J-stat	9.232					
p-value	0.755					

Table 3.16 presents the estimated factor covariance risk premiums on 31 industry portfolios. As expected, the results are generally consistent with those in Panel B of Table 3.8. The estimated covariance risk premiums preserve the signs of the corresponding factor loadings on the SDF except for that of Growth. The J-statistics in Tables 3.14, 3.15, and 3.16 are all insignificant at the 5% significance level, and we do not find strong evidence of model misspecification. Overall, the above results show that the machine learning factors maintain consistent performance in different model specifications and over different test assets.

Table 3.16. Factor covariance risk premiums on 31 Industry portfolios
This table presents the estimated covariance risk premiums of the ML factors by GMM using moment conditions 2 in Section 3.3.3. The test assets are 31 Industry portfolios. MKT is the market excess return. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about the change in inequality. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2021:M12. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

Covariance risk premiums of ML factors						
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	0.351*	-0.412***	-0.192**	-0.607**	-0.664**	2.071**
z-stat	1.84	-3.19	-2.41	-2.18	-2.37	2.08
J-stat	8.087					
p-value	0.986					

3.5.2. Estimation results before the introduction of the Shanghai-Hong Kong Stock Connect

In line with Chapter 2, we also repeat the analysis in Tables 3.6 and 3.9 over the period before the introduction of the Shanghai-Hong Kong Stock Connect and check if the results differ substantially from those over the whole sample. Table 3.17 repeats the analysis of Panel B of Table 3.6 and re-estimates the model with the machine learning factors augmented with the market return using GMM moment condition 1 over the sample period of December 2006 to October 2014. Note that the significance of all factors drops except for Liquidity. The loading of the Growth factor drops from 0.042 to 0.031, and the estimated coefficient is no longer significant at 1% significance level. Inflation and Defy, which convey incremental pricing information over the full sample, are no longer significant at the 5% significance level. Inequality, which is highly significant at 1% significance level over the full sample, is only significant at the 10% level. The J-statistics also rises from 8.085 to 11.218, although there is still no sign of model misspecification.

Table 3.17. Tests with the ML factors on 25 Size-E/P portfolios before the introduction of the Shanghai-Hong Kong Stock Connect

This table presents the estimated coefficients of the ML factors by GMM using moment conditions 1 in Section 3.3.3. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Defy is the mimicking portfolio of news about the default spread. Inequality is the mimicking portfolio of news about inequality. The estimated parameters are factor loadings on the SDF. The z-statistics corrected for serial correlation and heteroskedasticity using the Newey West kernel are in parentheses. J-stat is Hansen's (1982) J statistics for tests of overidentifying restrictions. The sample period is from 2006:M12 to 2014:M10. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	ML factors + MKT					
	Growth	Inflation	Liquidity	Defy	Inequality	MKT
Estimate	0.031**	0.020	0.035**	0.013	-0.019*	0.027
z-stat	2.23	1.08	2.16	0.98	-1.92	1.31
J-stat	11.218					
p-value	0.916					

Table 3.18 repeats the analysis of Table 3.9 and re-estimates the factor risk premiums using the Fama-Macbeth (1973) procedure over the sample period from December 2006 to October 2014. Note that the adjusted R^2s drop in every column of Table 3.18. In column (6), the adjusted R^2 for the full model drops from 0.680 in Table 3.9 to 0.614, and the estimated risk premiums for all machine learning factors drop in magnitude compared with those over the full sample. The risk premiums for Growth and Defy, which are significant at the 5% significance level over the full sample, are no longer significant. The risk premiums for Inflation and Inequality, which are highly significant at 1% significance level over the full sample, also suffer some degree of loss in significance. However, they are still significant at the 5% level. Surprisingly, Liquidity, which has significant loadings on the SDF in Table 3.16, does not have a significant risk premium. The results in Tables 3.17 and 3.18 hint at a tighter bond between macroeconomic risks and the cross-section of Chinese stock returns.

Table 3.18. Risk premiums for the ML factors on 25 Size-E/P portfolios before the introduction of the Shanghai-Hong Kong Stock Connect

This table presents the estimated risk premiums of the ML factors by the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. Growth is the machine learning domestic growth factor. Inflation is the machine learning domestic inflation factor. Liquidity is the machine learning domestic liquidity factor. Inequality is the mimicking portfolio of news about inequality. MKT is the market excess return. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2006:M12 to 2014:M10. *, ** and *** denote significance at the 10%, 5% and 1% significance levels respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret	(6) Ret
Growth	0.081					0.096
	(1.06)					(1.53)
Inflation		-0.087**				-0.073**
		(-2.03)				(-2.26)
Liquidity			0.263*			0.235
			(1.85)			(1.43)
Defy				-0.242		0.131
				(-1.32)		(0.89)
Inequality					-0.187**	-0.235**
					(-2.08)	(-2.38)
MKT						0.336
						(1.05)
cons	0.933	2.132	1.313	-1.208	-0.818	-1.662
	(1.11)	(1.45)	(1.07)	(1.21)	(-0.96)	(-1.38)
R^2	0.300	0.315	0.193	0.187	0.288	0.682
adj. R^2	0.269	0.286	0.159	0.152	0.258	0.614

3.6. Conclusion

This study aims to improve the explanatory power of macroeconomic factors in the cross-section of Chinese stock returns. To do so, we construct a set of machine-learning-enhanced macroeconomic factors using a novel procedure binding the theory of the ICAPM with the Elastic Net regressions. The final machine learning macro-factor model consists of 5 macroeconomic factors: a factor representing news about economic growth, a factor representing news about inflation, a factor representing news about macro-liquidity, a factor representing news about the default spread, and a factor representing news about the change in inequality. Since the factors are consistent with the restrictions of the ICAPM in general, we also include the market return as a final factor.

We investigate whether each of the resulting enhanced macroeconomic factors provides incremental pricing information by employing GMM estimation following the stochastic discount factor approach. We test whether each factor is priced over three sets of test assets by the Fama-Macbeth (1973) procedure. We also probe the relationship between the machine learning factors and the Fama-French five factors and compare the pricing errors of different sets of factors using two metrics.

Our results show that almost all enhanced macroeconomic factors have significant loadings on the stochastic discount factor over different test assets, thus conveying unique pricing information. Furthermore, most are priced with statistically and economically significant risk premiums estimated by the Fama-Macbeth procedure. Except for liquidity, the machine-learning-enhanced macroeconomic factors exhibit consistency with restrictions of the ICAPM in that their risk premiums have the right signs as indicated by predictive regressions. Further tests under the framework of GMM suggest that the Fama-French five factors provide little incremental pricing information in the presence of the machine-learning-enhanced macroeconomics factors. In particular, this study expands the work of Chapter 1 and identifies inequality as a nonnegligible systematic risk from a large pool of competing macroeconomic variables. Further results show that our machine-learning-enhanced macroeconomic factors perform on par with the Fama-French five factors and better than the PCA and Chen-Roll-Ross factors constructed in Chapter 2 in most situations. Overall, evidence indicates that machine-learning-enhanced factors incorporate more useful pricing information than traditional ad hoc sparse macroeconomic factors such as the CRR factors.

Appendix A. Variable Definition

Table A.1. Definition of variables in Chapter 1

Variables	**Definition**
Macroeconomic variables	
RPI_jewelry_YoY	Year-on-year change in the jewelry subcategory of Retail Price Index
RPI_commodity_YoY	Year-on-year change in the daily necessities' subcategory of Retail Price Index
Housing_YoY	Year-on-year change in the Chinese mainland real estate prices
HK luxury _YoY	Year-on-year change in the Hong Kong luxury home prices
CPI_YoY	Year-on-year change in the Consumer Price Index
CPI_rent_YoY	Year-on-year change in the property rent subcategory of the Consumer Price Index
CPI_urban_YoY	Year-on-year change in the Urban Consumer Price Index
CPI_rural_YoY	Year-on-year change in the Rural Consumer Price Index
Retail_YoY	Year-on-year change in the Total Retail Sales of Consumer Goods
Retail_jewelry_YoY	Year-on-year change in the jewelry subcategory of Total Retail Sales of Consumer Goods
Retail_commodity_YoY	Year-on-year change in the daily necessities' subcategory of Total Retail Sales of Consumer Goods
Macao _YoY	Year-on-year change in the revenue for the Macao's gambling industry
Imported_jewelry_yoy	Year-on-year change in the price of imported jewelry
Luxury_cars_yoy	Year-on-year change in the number of luxury cars sold
Economy_cars_yoy	Year-on-year change in the number of economy cars sold
Term_spread1Y	The yield spread between the 10-year government bond and 1-year government bond
Default_spread	The average spread between the 3-year AA corporate bonds and 3-year government bond
Dr1M	1-month Interbank Repo Rate
Ind	Year-on-year change in industrial production
PMI	The Purchasing Managers' Index
PPI_YoY	Year-on-year change in the Producer Price Index
Loan	Year-on-year change in newly created bank loan
RMB_USD	Year-on-year change in the exchange rate of RMB to USD
Realty	The National Real Estate Climate Index
Trade	Year-on-year change in total import + total export
M1 _YoY	Year-on-year change in M1
M2 _YoY	Year-on-year change in M2
M1M2	Equals M1 _YoY - M2 _YoY, which measures economic prosperity. A higher value means consumers and firms are more willing to spend.
Invest	Year-on-year change in the countrywide fixed asset investment

Table A.1. Definition of variables in Chapter 1 *(continued)*

Variables	Definition
Inequality indices	
Inequality_fa*	Proxy for change in inequality constructed by dynamic factor model
Inequality_pc*	Proxy for change in inequality constructed by principal component analysis
Inequality_ent*	Proxy for change in inequality constructed by entropy weighting method
Factors	
MKT	The monthly excess return of the market portfolio
SMB	Modified Fama-French size factor as in Liu et al. (2019)
VMG	Modified Fama-French value factor as in Liu et al. (2019)
RMW	Fama-French profitability factor
CMA	Fama-French investment factor
Inequality_fa	Mimicking portfolio of news about Inequality_fa*
Inequality_pc	Mimicking portfolio of news about Inequality_pc*
Inequality_ent	Mimicking portfolio of news about Inequality_ent*
Variables in predictive regressions	
Rm	Monthly return of the market index
Svar_A	Monthly realized volatility of the market return
P_b	Market price-to-book ratio
Divi	Market dividend yield
Turn_A	Market turnover
Ntis	Net equity expansion
Firm characteristics	
Shrs	Lagged sum of shareholding percentage of the second largest shareholder to the tenth largest shareholder
IVOL	Lagged monthly idiosyncratic volatility
LnSize	Log of market capitalization of the firm at the end of last month
LnBM	Lagged 1-month log of book-to-market ratio
Turn	Average daily turnover rate during the past month
Illiquidity	Lagged monthly Amihud (2002) illiquidity
Ret_1	Lagged monthly return
Ret2_12	Holding period return from month t-12 to month t-2
Beta	Market beta estimated by time series rolling regressions using daily returns during the past 250 trading days
Leverage	Lagged 1-month debt-to-equity ratio
Basis inequality measures	
RPI_gap	RPI_jewelry_YoY - RPI_commodity_YoY
Housing_gap	Housing_YoY - CPI_rent_YoY
HK_gap	HK luxury_YoY - CPI _YoY
Urban_rural_gap	CPI_urban_YoY- CPI_rural_YoY
Retailsale_gap	Retail_jewelry_YoY - Retail_commodity_YoY
Indcv	Year-on-year change in the coefficient of variation of per capita industrial profits
Macao_gap	Macao_YoY - Retail_commodity_YoY
Expenditure_90_10	Year-on-year change in the average per capita expenditure of top 10% provinces / average per capita expenditure of bottom 90% provinces.
Imported_jewelry_gap	Imported_jewelry_yoy - Retail_commodity_YoY
Automobile_gap	Luxury_cars_YoY – Economy_cars_YoY

Table A.2. Definition of variables in Chapter 2

Variables	**Definition**
Main asset class returns	
HS300	Monthly return of the HS300 Index
HS500	Monthly return of the CSI 500 Index
HS1000	Monthly return of the CSI 1000 Index
Bond_G	Monthly return of the CSI Government Bond Index
Bond_C	Monthly return of the CSI Corporate Bond Index
Gold	Monthly return of the Au9999 Index, traded at the Shanghai Futures Exchange
Industry	Monthly return of the Nanhua Industrial Commodity Index
Metal	Monthly return of the Nanhua Metal Index
Energy	Monthly return of the Nanhua Energy & Chemical Index
Macroeconomic variables	
Domestic growth	
ID_YoY	Year-on-year change in the industrial production
Domestic inflation	
CPI_YoY	Year-on-year change in the Consumer Price Index
Stock market aggregate	
Turn_A	Monthly market turnover
Svar	Monthly realized volatility of the market return
Yield curve	
Dr1M	1-month Interbank Repo Rate
Yield_6M	Yield to maturity of 6-month government bond
Yield_5Y	Yield to maturity of 5-year government bond
Yield_10Y	Yield to maturity of 10-year government bond
Defy*	Default spread=Average yield on 3-year AA corporate bond- yield on 3-year government bond
Term*	Term spread= Yield on 10-year government bond - Yield on 1-year government bond
Factors	
MKT	Monthly excess return of the market portfolio
SMB	Modified Fama-French size factor as in Liu, Stambaugh and Yuan (2019)
VMG	Modified Fama-French value factor as in Liu, Stambaugh and Yuan (2019)
RMW	Fama-French profitability factor
CMA	Fama-French investment factor
PC1	First principal component of monthly main asset class returns, poxy for growth news
PC2	Second principal component of monthly main asset class returns, poxy for stagflation news
PC3	Third principal component of monthly main asset class returns, poxy for long-term interest rate news
PC4	Fourth principal component of monthly main asset class returns, poxy for liquidity and exchange rate news
PC5	Fifth principal component of monthly main asset class returns, poxy for relative growth news
MP	Mimicking portfolio of news about ID_YoY
DEI	First difference of analyst consensus expected value of CPI_YoY from Wind
UI	Unexpected inflation= CPI_YoY - DEI
Defy	Mimicking portfolio of news about Defy*
Term	Mimicking portfolio of news about Term*

Table A.3. Definition of variables in Chapter 3

Variables	Definition
Macroeconomic variables	
Domestic growth	
ID_YoY	Year-on-year change in industrial production
PMI_Manu	Purchasing Managers' Index for manufacturing industries
PMI_Nonmanu	Purchasing Managers' Index for non-manufacturing industries
Keqiang Index	Li Keqiang Index, an estimate of monthly economic growth, equals 40% $*$ year-on-year growth of industrial electricity usage +35% $*$ year-on-year growth of mid/long-term loan +25% $*$ year-on-year growth of volume of railway freight
BCI_CN_MoM	The month-on-month change in the OECD Chinese Business Confidence Index
Vehicle_MoM	The month-on-month change in vehicle production
Retail_Sale_YoY	Year-on-year change in total retail sales
Domestic inflation	
CPI_YoY	Year-on-year change in the Consumer Price Index
PPI_MoM	The month-on-month change in the Producer Price Index
PPI_YoY	Year-on-year change in the Producer Price Index
RPI_YoY	Year-on-year change in the Retail Price Index
Stock market aggregate	
Turn_A	Monthly market turnover
Svar	Monthly realized volatility of the market return
Domestic liquidity	
M0_YoY	Year-on-year change in M0
M2_YoY	Year-on-year change in M2
Yield curve	
Dr1M	1-month Interbank Repo Rate
Yield_6M	Yield to maturity of 6-month government bond
Yield_5Y	Yield to maturity of 5-year government bond
Yield_10Y	Yield to maturity of 10-year government bond
Defy*	Default spread=Average yield on the 3-year AA corporate bond- yield on the 3-year government bond
Term*	Term spread= Yield on the 10-year government bond - Yield on the 1-year government bond
Inequality	
Inequality_fa*	Inequality index constructed in Chapter 1 by the dynamic factor model, which measures year-on-year change in inequality in society.
Exchange rate	
USD_RMB*	Year-on-year change in the exchange rate of USD to RMB
USDX_YoY	Year-on-year change in the U.S. Dollar Index

Table A.3. Definition of variables in Chapter 3 *(continued)*

Variables	Definition
Overseas Growth	
BCI_G7_YoY	Year-on-year change in the OECD G7 countries' Business Confidence Index
BCI_US_YoY	Year-on-year change in the OECD U.S. Business Confidence Index
Rec_Prob_US	Monthly recession probability of the U.S. by FRED
ECRI_US_YoY	Year-on-year change in the U.S. Economic Activity Coincident Index by ECRI
CEAI_US_YoY	Year-on-year change in the U.S. Economic Activity Coincident Index by FRED
Overseas Inflation	
CPI_US_YoY	Year-on-year change in the U.S. Consumer Price Index
PPI_US_YoY	Year-on-year change in the U.S. Producer Price Index
Inflation_Exp_US	The U.S. inflation expectation
Overseas Liquidity	
Fed*	The Effective Federal Funds Rate
Factors	
MKT	The monthly excess return of the market portfolio
SMB	Modified Fama-French size factor as in Liu et al. (2019)
VMG	Modified Fama-French value factor as in Liu et al. (2019)
RMW	Fama-French profitability factor
CMA	Fama-French investment factor
PC1	The first principal component of monthly main asset class returns in Chapter 2, poxy for growth news
PC2	The second principal component of monthly main asset class returns in Chapter 2, poxy for stagflation news
PC3	The third principal component of monthly main asset class returns in Chapter 2, poxy for long-term interest rate news
PC4	The fourth principal component of monthly main asset class returns in Chapter 2, poxy for liquidity and exchange rate news
PC5	The fifth principal component of monthly main asset class returns in Chapter 2, poxy for relative growth news
MP	Mimicking portfolio of news about ID_YoY
DEI	The first difference of analyst consensus expected value of CPI_YoY from Wind
UI	Unexpected inflation= CPI_YoY - DEI
Defy	Mimicking portfolio of news about Defy*
Term	Mimicking portfolio of news about Term*
Growth*	Machine learning state variable representing domestic economic growth
Growth	Mimicking portfolio of news about Growth*
Inflation*	Machine learning state variable representing domestic inflation
Inflation	Mimicking portfolio of news about Inflation*
Liquidity*	Machine learning state variable representing domestic liquidity
Liquidity	Mimicking portfolio of news about Liquidity*
Inequality	Mimicking portfolio of news about Inequality_fa*
Growth_F*	Machine learning state variable representing overseas economic growth
Growth_F	Mimicking portfolio of news about Growth_F*
Inflation_F*	Machine learning state variable representing overseas inflation
Inflation_F	Mimicking portfolio of news about Inflation_F*
USD_RMB	Mimicking portfolio of news about USD_RMB*
Fed	Mimicking portfolio of news about Fed*

Appendix B. Supplementary results for Chapter 1

Table B.1. Firm-level estimates of inequality risk premiums using Inequality_pc

This table presents the estimated inequality risk premiums by the Fama-Macbeth (1973) procedure using Inequality_ pc. Inequality is the firm-level inequality risk beta estimated by augmenting the inequality index tracking portfolio with the Fama-French three factors. LnSize is the log of market capitalization of the firm at the end of last month. LnBM is the lagged 1-month log of book-to-market ratio. IVOL is idiosyncratic volatility over the past month. Turn is the average daily turnover rate during the past month. Illiquidity is Amihud's (2002) illiquidity measure. Ret_1 is a measure of the short-term reversal effect. Ret2_12 measures the momentum effect. Beta is the market beta estimated using daily returns over the past year. Leverage is the lagged 1-month debt-to-equity ratio. The Newey and West (1987a) t-statistics corrected for serial correlation and heteroskedasticity are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret_rf	(2) Ret_rf	(3) Ret_rf	(4) Ret_rf	(5) Ret_rf
Inequality	-0.513*	-0.468**	-0.457***		-0.414***
	(-1.73)	(-2.06)	(-3.51)		(-3.60)
Shrs	0.006*	0.012***	0.007***	0.009***	0.008***
	(1.88)	(4.13)	(2.81)	(3.59)	(3.15)
IVol				-0.283***	-0.247***
				(-3.60)	(-3.38)
LnSize		-0.343**	-0.395***	-0.375***	-0.359**
		(-2.25)	(-2.69)	(-2.61)	(-2.48)
LnBM		0.485***	0.131	0.063	0.084
		(3.79)	(1.07)	(0.51)	(0.71)
Turn			-1.946***	-1.696***	-1.697***
			(-8.32)	(-7.46)	(-7.32)
Illiquidity			5.654**	3.104*	6.776***
			(2.32)	(1.75)	(2.64)
Ret_1			-0.020**	-0.009	-0.014*
			(-2.46)	(-1.11)	(-1.72)
Ret2_12			0.001	0.002	0.002
			(0.49)	(0.63)	(0.54)
Beta	-0.002	-0.001	0.001	-0.001	0.002
	(-1.09)	(-0.85)	(1.03)	(-1.05)	(1.16)
Leverage			0.043	-0.028	0.078
			(0.20)	(-0.12)	(0.36)
cons	1.379**	4.662***	5.355***	4.692***	5.253***
	(2.01)	(3.02)	(3.31)	(3.43)	(3.27)
R^2	0.035	0.081	0.120	0.114	0.123
adj. R^2	0.033	0.078	0.113	0.107	0.115

Table B.2. Firm-level estimates of inequality risk premiums using Inequality_ent

This table presents the estimated inequality risk premiums by the Fama-Macbeth (1973) procedure using Inequality_ent. Inequality is the firm-level inequality risk beta estimated by augmenting the inequality index tracking portfolio with the Fama-French three factors. LnSize is the log of market capitalization of the firm at the end of last month. LnBM is the lagged 1-month log of book-to-market ratio. IVOL is idiosyncratic volatility over the past month. Turn is the average daily turnover rate during the past month. Illiquidity is Amihud's (2002) illiquidity measure. Ret_1 is a measure of the short-term reversal effect. Ret2_12 measures the momentum effect. Beta is the market beta estimated using daily returns over the past year. Leverage is the lagged 1-month debt-to-equity ratio. The Newey and West (1987a) t-statistics corrected for serial correlation and heteroskedasticity are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)	(4)	(5)
	Ret_rf	Ret_rf	Ret_rf	Ret_rf	Ret_rf
Inequality	-0.168	-0.214*	-0.226**		-0.179**
	(-1.52)	(-1.85)	(-2.36)		(-2.25)
Shrs	0.007**	0.012***	0.007***	0.009***	0.008***
	(2.23)	(4.47)	(3.28)	(3.59)	(3.38)
IVol				-0.283***	-0.242***
				(-3.60)	(-2.81)
LnSize		-0.296*	-0.360**	-0.375***	-0.327**
		(-1.75)	(-2.38)	(-2.61)	(-2.29)
LnBM		0.490***	0.137	0.063	0.094
		(3.76)	(1.09)	(0.51)	(0.78)
Turn			-1.904***	-1.696***	-1.655***
			(-8.55)	(-7.46)	(-8.03)
Illiquidity			5.119*	3.104*	6.303**
			(1.91)	(1.75)	(2.14)
Ret_1			-0.019**	-0.009	-0.012
			(-2.19)	(-1.11)	(-1.53)
Ret2_12			0.002	0.002	0.002
			(0.64)	(0.63)	(0.69)
Beta	-0.001	-0.001	-0.001	-0.001	0.002
	(-1.02)	(-1.03)	(-1.07)	(-1.08)	(1.13)
Leverage			-0.025	-0.028	0.014
			(-0.10)	(-0.12)	(0.05)
cons	1.462*	4.370**	5.147***	4.692***	5.047***
	(1.93)	(2.52)	(3.00)	(3.43)	(2.93)
R^2	0.025	0.071	0.117	0.114	0.120
adj. R^2	0.020	0.068	0.109	0.107	0.112

Table B.3. Estimates of the inequality risk premium on 25 Size-EP portfolios using Inequality_pc

This table presents the estimated inequality risk premiums by augmenting the inequality tracking portfolio with the Fama-French factors using the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. Inequality is the beta on Inequality_pc. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)	(4)	(5)
	Ret	Ret	Ret	Ret	Ret
Inequality	-0.630		-0.556**		-0.603**
	(-1.50)		(-2.05)		(-2.24)
b_MKT	1.166	0.566	0.411	0.507*	0.367
	(1.01)	(1.62)	(1.44)	(1.73)	(1.34)
b_SMB		1.137***	0.481**	0.752***	0.508**
		(2.76)	(1.99)	(2.53)	(2.03)
b_VMG		0.804***	0.388***	0.623***	0.598***
		(2.82)	(3.44)	(2.79)	(2.61)
b_RMW				0.271	0.258*
				(1.44)	(1.75)
b_CMA				0.678**	0.392
				(2.01)	(1.63)
cons	2.658*	2.432*	2.158*	2.525**	1.015
	(1.81)	(1.76)	(1.87)	(2.23)	(1.54)
R^2	0.191	0.621	0.648	0.709	0.721
adj. R^2	0.137	0.554	0.577	0.631	0.656

Table B.4 Estimates of the inequality risk premium on 25 Size-EP portfolios using Inequality_ent

This table presents the estimated inequality risk premiums by augmenting the inequality tracking portfolio with the Fama-French factors using the Fama-Macbeth (1973) procedure. The test assets are 25 value-weighted double-sorted portfolios on size and E/P ratio. Inequality is the beta on Inequality_ent. MKT is the market excess return. SMB is the modified Fama-French size factor. VMG is the modified Fama-French value factor as in Liu et al. (2019). RMW is the Fama-French profitability factor. CMA is the Fama-French investment factor. The Shanken (1992) t-statistics corrected for errors-in-variables bias are in parentheses. The sample period is from 2005:M2 to 2021:M12. *, **, and *** represent significance at the significance levels of 10%, 5%, and 1%, respectively.

	(1) Ret	(2) Ret	(3) Ret	(4) Ret	(5) Ret
Inequality	-0.296		-0.251*		-0.216**
	(-1.21)		(-1.69)		(-2.09)
b_MKT	1.311	0.566	0.454	0.507*	0.358
	(1.46)	(1.62)	(1.41)	(1.73)	(1.44)
b_SMB		1.137***	0.703**	0.752**	0.634**
		(2.76)	(2.17)	(2.43)	(2.15)
b_VMG		0.804***	0.530**	0.623***	0.471**
		(2.82)	(2.08)	(2.79)	(2.15)
b_RMW				0.271	-0.163
				(1.44)	(-1.23)
b_CMA				0.678**	0.398
				(2.01)	(1.06)
cons	2.861*	2.432*	2.373**	2.525**	1.284*
	(1.83)	(1.76)	(2.15)	(2.23)	(1.71)
R^2	0.154	0.621	0.615	0.709	0.712
adj. R^2	0.101	0.554	0.551	0.631	0.635

Bibliography

Abbas, G., Bashir, U., Wang, S., Zebende, G. F., & Ishfaq, M. (2019). The return and volatility nexus among stock market and macroeconomic fundamentals for China. *Physica A: Statistical Mechanics and its Applications*, 526, 121025.

Abbott, B., & Brace, R. (2020). Has consumption inequality mirrored wealth inequality in the survey of consumer finances? *Economics Letters*, 193, 109289.

Adrian, T., Etula, E., & Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance*, *69*(6), 2557–2596.

Aguiar, M., & Bils, M. (2015). Has Consumption Inequality Mirrored Income Inequality? *The American Economic Review*, 105(9), 2725–2756.

Ahn, D.-H., Conrad, J., & Dittmar, R. F. (2009). Basis Assets. *The Review of Financial Studies*, 22(12), 5133–5174.

Aiyar, S., & Ebeke, C. (2020). Inequality of opportunity, inequality of income and economic growth. *World Development*, 136, 105115.

Ali, F., He, R., & Jiang, Y. (2018). Size, value and business cycle variables. The three-factor model and future economic growth: Evidence from an emerging market. *Economies*, *6*(1), 14.

Allen, F., Qian, J., Shan, C., & Zhu, J. L. (2024). Dissecting the Long-Term Performance of the Chinese Stock Market. *The Journal of Finance*, 79(2), 993-1054.

Amenc, N., Esakia, M., Goltz, F., & Luyten, B. (2019). Macroeconomic risks in equity factor investing. *Journal of Portfolio Management*, *45*(6), 39–60.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.

Amihud, Y., & Noh, J. (2021). Illiquidity and stock returns II: Cross-section and time-series effects. *The Review of Financial Studies*, *34*(4), 2101–2123.

An, J., Ho, K. Y., & Zhang, Z. (2020). What drives the liquidity premium in the Chinese stock market? *The North American Journal of Economics and Finance*, 54, 101088.

Ang, A. (2014). *Asset management: A systematic approach to factor investing*. Oxford University Press.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259-299.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence. *Journal of Financial Economics*, *91*(1), 1–23.

Ang, A., Liu, J., & Schwarz, K. (2020). Using stocks or portfolios in tests of factor models. *Journal of Financial and Quantitative Analysis*, 55(3), 709-750.

Antoniou, A., Lam, H. Y., & Paudyal, K. (2007). Profitability of momentum strategies in international markets: The role of business cycle variables and behavioural biases. *Journal of Banking & Finance*, *31*(3), 955-972.

Aretz, K., Bartram, S. M., & Pope, P. F. (2010). Macroeconomic risks and characteristic-based factor models. *Journal of Banking & Finance*, 34(6), 1383–1399.

Arouri, M. E. H., Lahiani, A., & Nguyen, D. K. (2015). World gold prices and stock returns in China: Insights for hedging and diversification strategies. *Economic Modelling*, 44, 273-282.

Attanasio, O. P., & Pistaferri, L. (2016). Consumption inequality. *Journal of Economic Perspectives*, 30(2), 3-28.

Avramov, D., & Chordia, T. (2006). Asset pricing models and financial market anomalies. *The Review of Financial Studies*, *19*(3), 1001–1040.

Bai, Y., & Green, C. J. (2011). Determinants of cross-sectional stock return variations in emerging markets. *Empirical Economics*, *41*(1), 81–102.

Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets*, *7*(3), 271-299.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593-1636.

Balcilar, M., Bonato, M., Demirer, R., & Gupta, R. (2018). Geopolitical risks and stock market dynamics of the BRICS. *Economic Systems*, *42*(2), 295-306.

Balduzzi, P., & Robotti, C. (2008). Mimicking portfolios, economic risk premia, and tests of multi-beta models. *Journal of Business & Economic Statistics*, *26*(3), 354–368.

Balduzzi, P., & Yao, T. (2007). Testing heterogeneous-agent models: An alternative aggregation approach. *Journal of Monetary Economics*, 54(2), 369-412.

Bali, T. G. (2008). The intertemporal relation between expected returns and risk. *Journal of Financial Economics*, 87(1), 101-131.

Bali, T. G., & Engle, R. F. (2010). The intertemporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics*, 57(4), 377-390.

Bali, T. G., Brown, S. J., & Tang, Y. (2017). Is economic uncertainty priced in the cross-section of stock returns? *Journal of Financial Economics*, *126*(3), 471–489.

Bali, T. G., Engle, R. F., & Murray, S. (2016). *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

Barinov, A. (2014). Turnover: Liquidity or uncertainty? *Management Science*, *60*(10), 2478–2495.

Bekaert, G., & Engstrom, E. (2010). Inflation and the stock market: Understanding the "Fed Model." *Journal of Monetary Economics*, 57(3), 278–294.

Benabou, R. (1996). Inequality and growth. *NBER macroeconomics annual*, 11, 11-74.

Black, A. J., Klinkowska, O., McMillan, D. G., & McMillan, F. J. (2014). Forecasting stock returns: Do commodity prices help? *Journal of Forecasting*, *33*(8), 627–639.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business*, 45(3), 444–455.

Bollerslev, T., Li, S. Z., & Zhao, B. (2020). Good volatility, bad volatility, and the cross section of stock returns. *Journal of Financial and Quantitative Analysis*, *55*(3), 751–781.

Boons, M. (2016). State variables, macroeconomic activity, and the cross section of individual stocks. *Journal of Financial Economics*, *119*(3), 489–511.

Boons, M., de Roon, F. & Szymanowska, M. (2014). The price of commodity risk in stock and futures markets. Working paper, Nova School of Business and Economics

Bouri, E., Demirer, R., Gupta, R., & Nel, J. (2021). COVID-19 pandemic and investor herding in international stock markets. *Risks*, *9*(9), 168.

Brandt, M. W., Brav, A., Graham, J. R., & Kumar, A. (2010). The idiosyncratic volatility puzzle: Time trend or speculative episodes? *The Review of Financial Studies*, *23*(2), 863-899.

Brav, A., Constantinides, G. M., & Geczy, C. C. (2002). Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy*, 110(4), 793-824.

Brennan, M. J., Wang, A. W., & Xia, Y. (2004). Estimation and test of a simple model of intertemporal capital asset pricing. *The Journal of Finance*, 59(4), 1743-1776.

Brennan, M.J., Wang, A.W., Xia, Y., (2001). Intertemporal capital asset pricing and the Fama-French three-factor model. *Working paper*

Brogaard, J., & Detzel, A. (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science*, 61(1), 3–18.

Brooks, C., Fernandez-Perez, A., Miffre, J., & Nneji, O. (2016). Commodity risks and the cross-section of equity returns. *The British Accounting Review*, *48*(2), 134–150.

Brunnermeier, M., Farhi, E., Koijen, R. S. J., Krishnamurthy, A., Ludvigson, S. C., Lustig, H., Nagel, S., & Piazzesi, M. (2021). Review Article: Perspectives on the Future of Asset Pricing. *The Review of Financial Studies*, 34(4), 2126–2160.

Bu, C., Rogers, J., & Wu, W. (2021). A unified measure of Fed monetary policy shocks. *Journal of Monetary Economics*, 118, 331–349.

Byun, H. Y., Hwang, L. S., & Lee, W. J. (2011). How does ownership concentration exacerbate information asymmetry among equity investors? *Pacific-Basin Finance Journal*, *19*(5), 511-534.

Cakici, N., Chan, K., & Topyan, K. (2017). Cross-sectional stock return predictability in China. *The European Journal of Finance*, 23(7–9), 581–605.

Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, *112*(4), 1194-1225.

Campbell, J. Y. (1996). Understanding risk and return. *The Journal of Political Economy*, *104*(2), 298–345.

Campbell, J. Y., & Vuolteenaho, T. (2004). Bad beta, good beta. *The American Economic Review*, *94*(5), 1249–1275.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57–82.

Carpenter, J. N., Lu, F., & Whitelaw, R. F. (2021). The real value of China's stock market. *Journal of Financial Economics*, *139*(3), 679-696.

Chan, K. C., Foresi, S., & Lang, L. H. P. (1996). Does money explain asset returns? theory and empirical analysis. *The Journal of Finance*, 51(1), 345–361.

Chan, L. K. C., Karceski, J., & Lakonishok, J. (1998). The risk and return from factors. *Journal of Financial and Quantitative Analysis*, *33*(2), 159–188.

Chen, C.-C., & Chuang, M.-C. (2008). Integrating the Kano model into a robust design approach to enhance customer satisfaction with product design. *International Journal of Production Economics*, *114*(2), 667–681.

Chen, J. L., Glabadanidis, P., & Sun, M. (2022). The five-factor asset pricing model, short-term reversal, and ownership structure–the case of China. *International Review of Financial Analysis*, 82, 102147.

Chen, J., Hong, H., & Stein, J. C. (2001). Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *Journal of financial Economics*, *61*(3), 345-381.

Chen, J., Jiang, F., & Tong, G. (2017). Economic policy uncertainty in China and stock market expected returns. *Accounting & Finance*, 57(5), 1265-1286.

Chen, J., Jiang, F., Liu, Y., & Tu, J. (2017). International volatility risk and Chinese stock return predictability. *Journal of International Money and Finance*, 70, 183-203.

Chen, N., & Zhang, F. (1998). Risk and return of value stocks. *The Journal of Business*, *71*(4), 501–535.

Chen, N.-F. (1983). Some empirical tests of the theory of arbitrage pricing. *The Journal of Finance*, 38(5), 1393–1414.

Chen, N.-F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *The Journal of Business*, *59*(3), 383–403.

Chen, Q. (2015). Climate shocks, state capacity and peasant uprisings in north China during 25-1911 ce. *Economica*, *82*(326), 295–318.

Chen, S.-J., Hsieh, C.-H., & Jordan, B. D. (1997). Real estate and the arbitrage pricing theory: Macrovariables vs. derived factors. *Real Estate Economics*, 25(3), 505–523.

Chen, X., Kim, K. A., Yao, T., & Yu, T. (2010). On the predictability of Chinese stock returns. *Pacific-Basin Finance Journal*, 18(4), 403–425.

Chen, Z., & Petkova, R. (2012). Does idiosyncratic volatility proxy for risk exposure? *The Review of Financial Studies*, *25*(9), 2745–2787.

Chordia, T., & Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *The Journal of Finance*, *57*(2), 985-1019.

Chordia, T., Subrahmanyam, A., & Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, *59*(1), 3–32.

Chung, H., Majerbi, B., & Rizeanu, S. (2015). Exchange risk premia and firm characteristics. *Emerging Markets Review*, 22, 96–125.

Cochrane, J. H. (1991). Production-based asset pricing and the link between stock returns and economic fluctuations. *The Journal of Finance*, 46(1), 209–237.

Cochrane, J. H. (1996). A cross-sectional test of an investment-based asset pricing model. *The Journal of Political Economy*, *104*(3), 572–621.

Cochrane, J. H. (1999). New facts in finance. *Economic Perspectives (1977)*, *23*(3), 36–58.

Cochrane, J. H. (2005). *Asset Pricing*: *Revised edition.* Princeton University Press.

Cochrane, J. H. (2011). Presidential Address: Discount Rates. *The Journal of Finance*, *66*(4), 1047–1108.

Cochrane, J. H. (2017). Macro-Finance. *Review of Finance*, *21*(3), 945–985.

Cogley, T. (2002). Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey. *Journal of Monetary Economics*, 49(2), 309-334.

Confucius., Cao, S., & Zisi. (1971). *Confucian analects, the Great learning and the Doctrine of the mean* (J. Legge, Trans.). Dover Publications.

Connor, G. (1995). The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal*, *51*(3), 42–46.

Connor, G., & Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15(3), 373–394.

Connor, G., & Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, *21*(2), 255–289.

Connor, G., & Korajczyk, R.A. (2009). Factor models in portfolio and asset pricing theory, *handbook of portfolio construction: Contemporary Applications of Markowitz Techniques* (pp.401-418). Springer.

Constantinides, G. M., & Duffie, D. (1996). Asset pricing with heterogeneous consumers. The *Journal of Political Economy*, 104(2), 219–240.

Cooper, I., Mitrache, A., & Priestley, R. (2022). A global macroeconomic risk model for value, momentum, and other asset classes. *Journal of Financial and Quantitative Analysis*, *57*(1), 1–30.

Cooper, M., & Maio, P. (2019). Asset growth, profitability, and investment opportunities. *Management Science*, *65*(9), 3988–4010.

Cornell, B. (1999). Risk, duration, and capital budgeting: New evidence on some old questions. *The Journal of Business*, *72*(2), 183–200.

Cowell, F. A. (2011). *Measuring inequality*. Oxford University Press.

Cox, J., Greenwald, D. L., & Ludvigson, S. C. (2020). What explains the COVID-19 stock market? (No. w27784). National Bureau of Economic Research.

Cuthbertson, K., Hayes, S., & Nitzsche, D. (1997). The behaviour of UK stock prices and returns: is the market efficient? *The Economic Journal*, *107*(443), 986-1008.

Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance*, 52(1), 1–33.

Davis, S. J., Liu, D., & Sheng, X. S. (2019). Economic policy uncertainty in China since 1949: The view from mainland newspapers. Working paper.

De Haan, J., & Sturm, J. E. (2017). Finance and income inequality: A review and new evidence. *European Journal of Political Economy*, *50*, 171-195.

Detzel, A. (2017). Monetary policy surprises, investment opportunities, and asset prices. *Journal of Financial Research*, *40*(3), 315-348.

Diebold, F. X., Rudebusch, G. D., & Borag˘an Aruoba, S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1), 309–338.

Docherty, P., & Hurst, G. (2018). Investor myopia and the momentum premium across international equity markets. *Journal of Financial and Quantitative Analysis*, *53*(6), 2465–2490.

Eckbo, B. E., & Norli, Ø. (2005). Liquidity risk, leverage and long-run IPO returns. *Journal of Corporate Finance*, *11*(1), 1–35.

Fama, E. F. (1981). Stock returns, real activity, inflation, and money. *The American Economic Review*, 71(4), 545–565.

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance*, *46*(5), 1575–1617.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, *22*(1), 3–25.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25*(1), 23–49.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427-465.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, *43*(2), 153-193.

Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, *123*(3), 441–463.

Fama, E. F., & French, K. R. (2020). Comparing Cross-Section and Time-Series Factor Models. *The Review of Financial Studies*, *33*(5), 1891–1926.

Fama, E. F., & Gibbons, M. R. (1984). A comparison of inflation forecasts. *Journal of Monetary Economics*, *13*(3), 327–348.

Fama, E.F., & French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.

Fama, E.F., & French, K.R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* 116, 1–22.

Fama, E.F., & French, K.R. (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies* 29, 69–103.

Fama, E.F., & Macbeth, J.D. (1973). Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 71, 607–636.

Fama, E.F., French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.

Fama, E.F., French, K.R. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, 50, 131-155.

Fama, E.F., French, K.R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51, 55-184.

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, *75*(3), 1327–1370.

Feroli, M. (2004). Monetary policy and the information content of the yield spread. In *Policy File*. The Federal Reserve Board.

Ferson, W. E., & Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4), 1325-1360.

Fidrmuc, J., Goergen, M., & Renneboog, L. (2008). Insider trading, news releases, and ownership concentration. In *Insider Trading* (pp. 309-370). CRC Press.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, *16*(2), 127-.

Gay Jr, R. D. (2008). Effect of macroeconomic variables on stock market returns for four emerging economies: Brazil, Russia, India, and China. *International Business & Economics Research Journal*, 7(3).

Gertler, M., & Gilchrist, S. (1994). Monetary policy, business cycles, and the behavior of small manufacturing firms. *The Quarterly Journal of Economics*, *109*(2), 309–340.

Gertler, M., Hubbard, R. G., & Kashyap, A. (1991). Interest rate spreads, credit constraints, and investment fluctuations: An empirical investigation. In *Financial markets and financial crises* (pp. 11-32). University of Chicago Press.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. Econometrica: *Journal of the Econometric Society*, 1121-1152.

Girardin, E., & Joyeux, R. (2013). Macro fundamentals as a source of stock market volatility in China: A GARCH-MIDAS approach. *Economic Modelling*, *34*, 59–68.

Girardin, E., & Liu, Z. (2003). The Chinese stock market: a casino with 'buffer zones'? *Journal of Chinese Economic and Business Studies*, *1*(1), 57-70.

Glaeser, E., Scheinkman, J., & Shleifer, A. (2003). The injustice of inequality. *Journal of Monetary Economics*, *50*(1), 199–222.

Goh, J. C., Jiang, F., Tu, J., & Wang, Y. (2013). Can US economic variables predict the Chinese stock market? *Pacific-Basin Finance Journal*, 22, 69-87.

Gollier, C. (2001). Wealth inequality and asset pricing. *The Review of Economic Studies*, *68*(1), 181–203.

Gomez, M. (2016). Asset prices and wealth inequality. *Working paper*.

Goyal, A. (2012). Empirical cross-sectional asset pricing: A survey. *Financial Markets and Portfolio Management*, *26*(1), 3–38.

Greene, W. (2012) *Econometric Analysis*. *7th Edition*, Prentice Hall, Upper Saddle River.

Gu, M., Jiang, G. J., & Xu, B. (2019). The role of analysts: an examination of the idiosyncratic volatility anomaly in the Chinese stock market. *Journal of Empirical Finance*, 52, 237-254.

Gu, M., Kang, W., & Xu, B. (2018). Limits of arbitrage and idiosyncratic volatility: Evidence from China stock market. *Journal of Banking & Finance*, *86*, 240-258.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, *222*(1), 429–450.

Guo, B., Zhang, W., Zhang, Y., & Zhang, H. (2017). The five-factor asset pricing model tests for the Chinese stock market. *Pacific-Basin Finance Journal*, *43*, 84–106.

Guo, J. (2015). Causal relationship between stock returns and real economic growth in the pre- and post-crisis period: Evidence from China. *Applied Economics*, 47(1), 12–31.

Hahn, J., & Lee, H. (2006). Yield spreads as alternative risk factors for size and book-to-market. *Journal of Financial and Quantitative Analysis*, *41*(2), 245–269.

Han, B., Liu, H., & Wang, R. (2015). Urban ecological security assessment for cities in the Beijing–Tianjin–Hebei metropolitan region based on fuzzy and entropy methods. *Ecological Modelling*, *318*, 217–225.

Hanhardt, A., & Ansotegui Olcoz, C. (2008). Do the Fama and French Factors proxy for state variables that predict macroeconomic growth in the Eurozone? Available at SSRN 1098225.

Hansen, L.P. (1982). Large sample properties of generalized methods of moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L.P., Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance* 52, 557–590.

Hardouvelis, G. A. (1994). The term structure spread and future changes in long and short rates in the G7 countries: Is there a puzzle? *Journal of Monetary Economics*, *33*(2), 255–283.

Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. *The Review of Financial Studies*, *6*(3), 473-506.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ...and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Hastie, T., Tibshirani, R., & Wainwright, M. (Martin J.). (2015). *Statistical learning with sparsity: the lasso and generalizations* (1st.). CRC Press.

Hilliard, J., & Zhang, H. (2015). Size and price-to-book effects: Evidence from the Chinese stock markets. *Pacific-Basin Finance Journal*, 32, 40-55.

Hirschman, A. O., & Rothschild, M. (1973). The changing tolerance for income inequality in the course of economic development: With a mathematical appendix. *The Quarterly Journal of Economics*, *87*(4), 544-566.

Hirschman, A. O., Adelman, Jeremy., Rothschild, Emma., & Sen, A. (2013). *The essential Hirschman*. Princeton University Press.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, 1-16.

Hong, H., & Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies*, *16*(2), 487-525.

Hou, K., & Loh, R. K. (2016). Have we solved the idiosyncratic volatility puzzle? *Journal of Financial Economics*, *121*(1), 167–194.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3), 650–705.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating Anomalies. *The Review of Financial Studies*, *33*(5), 2019–2133.

Hu, G. X., Chen, C., Shao, Y., & Wang, J. (2019). Fama–French in China: Size and value factors in Chinese stock returns. *International Review of Finance*, *19*(1), 3–44.

Hu, L., Han, J., & Zhang, Q. (2018). The impact of monetary and fiscal policy shocks on stock markets: Evidence from China. *Emerging Markets Finance & Trade*, *54*(8), 1856–1871.

Huang, T. L. (2019). Is the Fama and French five-factor model robust in the Chinese stock market? *Asia Pacific Management Review*, 24(3), 278-289.

Iania, L., Collage, R., & Vereycken, M. (2023). The impact of uncertainty in macroeconomic variables on stock returns in the USA. *Journal of Risk and Financial Management*, *16*(3), 189.

In, F., & Kim, S. (2007). A note on the relationship between Fama–French risk factors and innovations of ICAPM state variables. *Finance Research Letters*, *4*(3), 165–171.

Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *The Journal of Finance*, 51(1), 3–53.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65-91.

Jensen, G. R., Mercer, J. M., & Johnson, R. R. (1996). Business conditions, monetary policy, and expected security returns. *Journal of Financial Economics*, *40*(2), 213-237.

Jensen, T. I., Kelly, B., & Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, 78(5), 2465-2518.

Jiang, Y., Fu, T., Long, H., Zaremba, A., & Zhou, W. (2022). Real estate climate index and aggregate stock returns: Evidence from China. *Pacific-Basin Finance Journal*, *75*, 101841.

Jiao, W., & Lilti, J. J. (2017). Whether profitability and investment factors have additional explanatory power comparing with Fama-French Three-Factor Model: empirical evidence on Chinese A-share stock market. *China Finance and Economic Review*, 5, 1-19.

Johnson, T. C. (2012). Inequality risk premia. *Journal of Monetary Economics*, *59*(6), 565–580.

Jorion, P. (1991). The pricing of exchange rate risk in the stock market. *Journal of Financial and Quantitative Analysis*, 26(3), 363–376.

Jung, H., & Kim, D. (2016). Macro liquidity risk, money growth, and the cross-section of stock returns: The case of Korea. *Emerging Markets Finance and Trade*, 52(6), 1438-1454.

Kang, W., & Ratti, R. A. (2013). Oil shocks, policy uncertainty and stock market return. *Journal of International Financial Markets, Institutions and Money*, *26*, 305-318.

Kang, J., Lee, E., & Sim, M. (2014). Retail Investors and the Idiosyncratic Volatility Puzzle: Evidence from the Korean Stock Market. *Asia-Pacific Journal of Financial Studies, 43*(2), 183-222.

Keim, D. B., & Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics*, *17*(2), 357–390.

Kelly, B., & Xiu, D. (2023). Financial machine learning. *Foundations and Trends in Finance*, *13*(3-4), 205-363.

Kelly, B., and S. Pruitt. (2013). Market expectations in the cross-section of present values. *Journal of Finance* 68:1721–56.

Kelly, B., Palhares, D., & Pruitt, S. (2023). Modeling corporate bond returns. *The Journal of Finance*, 78(4), 1967-2008.

Kelly, B., S. Pruitt, & Y. Su. (2019). Characteristics *are* covariances: A unified model of risk and return. *Journal of Financial Economics* 134(3), 501–524.

Kim, D., Kim, T. S., & Min, B. K. (2011). Future labor income growth and the cross-section of equity returns. *Journal of Banking & Finance*, *35*(1), 67-81.

Knight, J. (2013). The economic causes and consequences of social instability in China. *China Economic Review*, *25*, 17-26.

Knight, J. (2014). Inequality in China: an overview. *The World Bank Research Observer*, 29(1), 1-19.

Koijen, R. S., Lustig, H., & Van Nieuwerburgh, S. (2017). The cross-section and time series of stock and bond returns. *Journal of Monetary Economics*, 88, 50-69.

Koijen, R., & S. V. Nieuwerburgh. (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics* 3:467–91.

Koivu, T. (2012). Monetary policy, asset prices and consumption in China. *Economic Systems*, *36*(2), 307–325.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *Journal of Finance* 73(3), 1183–1223.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics* 135(2), 271–292.

Kumar, A. (2009). Who gambles in the stock market? *The Journal of Finance*, 64(4), 1889-1933.

Kuznets, S. (1955). Economic Growth and Income Inequality. *The American Economic Review*, 1-28.

Labys, W. C., Achouch, A., & Terraza, M. (1999). Metal prices and the business cycle. *Resources Policy*, *25*(4), 229-238.

Lamont, O. A. (2001). Economic tracking portfolios. *Journal of Econometrics*, 105(1), 161–184.

Lee, K., Jeon, Y., & Nam, E.-Y. (2021). Chinese economic policy uncertainty and the cross-section of U.S. asset returns. *International Review of Economics & Finance*, *76*, 1063–1077.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1), 1-28.

Lehmann, B. N., & Modest, D. M. (1988). The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics*, 21(2), 213–254.

Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, *145*(2), 64–82.

Leite, A. L., Klotzle, M. C., Pinto, A. C. F., & da Silveira Barbedo, C. H. (2020). The Fama-French's five-factor model relation with interest rates and macro variables. *The North American Journal of Economics and Finance*, *53*, 101197.

Lettau, M. and M. Pelger (2020a). Estimating latent asset-pricing factors. *Journal of Econometrics* 218, 1–31.

Lettau, M. and M. Pelger (2020b). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies* 33(5), 2274–2325.

Lettau, M., & Pelger, M. (2020). Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies*, 33(5), 2274–2325.

Lewellen, J. (2015). The cross-section of expected stock returns. *Critical Finance Review* 4:1–44.

Lewellen, J., Nagel, S., & Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, *96*(2), 175–194.

Li, J., Zhang, W., & Kong, E. (2018). Factor models for asset returns based on transformed factors. *Journal of Econometrics*, 207(2), 432–448.

Li, M. C., Lai, C. C., & Xiao, L. (2022). Did COVID-19 increase equity market risk exposure? Evidence from China, the UK, and the US. *Applied Economics Letters*, *29*(6), 567–571.

Li, Y., Zhang, Q., Wang, L., & Liang, L. (2020). Regional environmental efficiency in China: An empirical analysis based on entropy weight method and non-parametric models. *Journal of Cleaner Production*, 276, 124147.

Liew, J., & Vassalou, M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics*, *57*(2), 221-245.

Lin, Q. (2021). The q5 model and its consistency with the intertemporal CAPM. *Journal of Banking & Finance*, 127, 106096.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1), 13–37.

Liu, J., Stambaugh, R. F., & Yuan, Y. (2019). Size and value in China. *Journal of Financial Economics*, *134*(1), 48–69.

Liu, L. X., & Zhang, L. (2008). Momentum profits, factor pricing, and macroeconomic risk. *The Review of Financial Studies*, *21*(6), 2417–2448.

Liu, Y., Wei, Y., Wang, Q., & Liu, Y. (2022). International stock market risk contagion during the COVID-19 pandemic. *Finance Research Letters*, *45*, 102145–102145.

Longstaff, F. A. (2004). The flight-to-liquidity premium in U.S. treasury bond prices. *The Journal of Business*, *77*(3), 511–526.

Lopez-Lira, A., & Roussanov, N. L. (2020). Do common factors really explain the cross-section of stock returns? *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.*

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica*, 46(6), 1429–1445.

Ludvigson, S. C., & Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, *83*(1), 171–222.

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, *22*(12), 5027–5067.

Lustig, H., & Verdelhan, A. (2007). The cross section of foreign currency risk premia and consumption growth risk. *American Economic Review, 97*(1), 89-117.

Lysandrou, P. (2011). Global inequality as one of the root causes of the financial crisis: A suggested explanation. *Economy and Society*, *40*(3), 323–344.

Maio, P., & Philip, D. (2018). Economic activity and momentum profits: Further evidence. *Journal of Banking & Finance*, *88*, 466–482.

Maio, P., & Santa-Clara, P. (2012). Multifactor models and their consistency with the ICAPM. *Journal of Financial Economics*, *106*(3), 586-613.

Maio, P., & Santa-Clara, P. (2017). Short-term interest rates and stock market anomalies. *Journal of Financial and Quantitative Analysis*, *52*(3), 927–961.

Markiewicz, A., & Raciborski, R. (2022). Income inequality and stock market returns. *Review of Economic Dynamics*, *43*, 286–307.

Mayers, D. (1973). Nonmarketable assets and the determination of capital asset prices in the absence of a riskless asset. *The Journal of Business*, *46*(2), 258–267.

Mei, J., Scheinkman, J. A., & Xiong, W. (2009). Speculative Trading and Stock Prices: Evidence from Chinese A-B Share Premia. *Annals of Economics and Finance*, *10*(2), 225–255.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867-887.

Meyer, B. D., & Sullivan, J. X. (2011). Further Results on Measuring the Well-Being of the Poor Using Income and Consumption. *The Canadian Journal of Economics*, 44(1), 52–87.

Meyer, B. D., & Sullivan, J. X. (2012). Identifying the Disadvantaged: Official Poverty, Consumption Poverty, and the New Supplemental Poverty Measure. *The Journal of Economic Perspectives*, 26(3), 111–135.

Meyer, B. D., & Sullivan, J. X. (2013). Consumption and income inequality and the great recession. *American Economic Review*, 103(3), 178-183.

Mishkin, F. S. (1988). The information in the term structure: Some further results. *Journal of Applied Econometrics*, *3*(4), 307–

Mishkin, F. S. (1990). What does the term structure tell us about future inflation? *Journal of Monetary Economics*, *25*(1), 77–95.

Nagel, S. (2021). *Machine Learning in Asset Pricing*. Princeton University Press.

Narayan, P. K., & Gupta, R. (2015). Has oil price predicted stock returns for over a century? *Energy economics*, *48*, 18-23.

Nartea, G. V., Kong, D., & Wu, J. (2017). Do extreme returns matter in emerging markets? Evidence from the Chinese stock market. *Journal of Banking & Finance*, *76*, 189–197.

Nartea, G. V., Wu, J., & Liu, Z. (2013). Does idiosyncratic volatility matter in emerging markets? Evidence from China. *Journal of International Financial Markets, Institutions and Money*, *27*, 137-160.

Naseem, S., Mohsin, M., Hui, W., Liyan, G., & Penglai, K. (2021). The investor psychology and stock market behavior during the initial era of COVID-19: A study of China, Japan, and the United States. *Frontiers in Psychology*, *12*, 626934.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management science*, 60(7), 1772-1791.

Newey, W. (1985). Generalized method of moments specification testing. *Journal of Econometrics* 29,229–256.

Newey, W. K., & West, K. D. (1987a). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Newey, W. K., & West, K. D. (1987b). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*(3), 777–787.

Ortmann, R., Pelster, M., & Wengerek, S. T. (2020). COVID-19 and investor behavior. *Finance research letters*, *37*, 101717.

Ozdagli, A., & Velikov, M. (2020). Show me the money: The monetary policy risk premium. *Journal of Financial Economics*, 135(2), 320-339.

Pástor, Ľ., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *The Journal of Political Economy*, *111*(3), 642–685.

Pastor, L., & Veronesi, P. (2012). Uncertainty about government policy and stock prices. *The Journal of Finance*, *67*(4), 1219-1264.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics*, *97*(3), 683–697.

Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, 208(1), 23–42.

Pelger, M. (2020). Understanding systematic risk: A high-frequency approach. *The Journal of Finance*, 75(4), 2179–2220.

Peng, J., Cui, J., Qin, F., & Groenewold, N. (2009). Stock prices and the macro economy in China. *Economics Discussion Paper no*, *9*.

Perez-Quiros, G., & Timmermann, A. (2000). Firm size and cyclical variations in stock returns. *The Journal of Finance*, *55*(3), 1229–1262.

Petkova, R. (2006). Do the Fama-French factors proxy for innovations in predictive variables? *The Journal of Finance*, *61*(2), 581–612.

Piketty, T. (2014). *Capital in the twenty-first century*. Harvard University Press.

Piketty, T., & Saez, E. (2003). Income inequality in the United States, 1913–1998. *The Quarterly Journal of Economics*, 118(1), 1–41.

Piketty, T., Saez, E., & Zucman, G. (2018). Distributional national accounts: methods and estimates for the United States. *The Quarterly Journal of Economics*, 133(2), 553-609.

Raol, J. (2017). How macro factors can aid asset allocation. *Invesco Fixed Income White Paper Series*.

Rapach, D., & G. Zhou. (2013). Forecasting stock returns, *Handbook of economic forecasting*, eds. G. Elliott and A. Timmermann, 328–83. Amsterdam, the Netherlands: Elsevier.

Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics*, *4*(2), 129–176.

Roll, R. and S. A. Ross (1980). An empirical investigation of the arbitrage pricing theory. *Journal of Finance* 35(5), 1073–1103.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341–360.

Rouwenhorst, K. G. (1999). Local return factors and turnover in emerging stock markets. *The Journal of Finance*, *54*(4), 1439–1464.

Rubin, A., & Segal, D. (2015). The effects of economic growth on income inequality in the US. *Journal of Macroeconomics*, 45, 258-273.

Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. *The Bell Journal of Economics*, 7(2), 407–425.

Saez, E., & Zucman, G. (2020). The rise of income and wealth inequality in America: Evidence from distributional macroeconomic accounts. *Journal of Economic Perspectives*, 34(4), 3-26.

Santos, T., & Veronesi, P. (2006). Labor income and predictable stock returns. *The Review of Financial Studies*, 19(1), 1-44.

Shang, H., Yuan, P., & Huang, L. (2016). Macroeconomic factors and the cross-section of commodity futures returns. *International Review of Economics & Finance*, *45*, 316–332.

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, 45(1-2), 99-120.

Shanken, J., & Weinstein, M. I. (2006). Economic forces and the stock market revisited. *Journal of Empirical Finance*, 13(2), 129–144.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423.

Sharpe, W. F. (1964). Capital Asset Prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425.

Shephard, N., Kinnebrock, S., & Barndorff-Nielsen, O. E. (2010). Measuring downside risk – Realized Semivariance. In *Volatility and Time Series Econometrics*. Oxford University Press.

Shin, I. (2012). Income inequality and economic growth. *Economic Modelling*, 29(5), 2049–2057.

Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance*, *52*(2), 737-783.

Sim, N., & Zhou, H. (2015). Oil prices, US stock return, and the dependence between their quantiles. *Journal of Banking & Finance*, *55*, 1-8.

Smales, L. A. (2021a). Investor attention and the response of US stock market sectors to the COVID-19 crisis. *Review of Behavioral Finance*, *13*(1), 20–39.

Smales, L. A. (2021b). Investor attention and global market returns during the COVID-19 crisis. *International Review of Financial Analysis*, *73*, 101616.

Soenen, L., & Johnson, R. (2001). The interrelationship cbetween macroeconomic variables and stock prices-The case of China. *Journal of Asia-Pacific Business*, *3*(2), 67–81.

Stambaugh, R. F., J. Yu, and Y. Yuan (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, Vol. 70(5), 1903 – 1948.

Su, C. W., Qin, M., Tao, R., Shao, X. F., Albu, L. L., & Umar, M. (2020). Can Bitcoin hedge the risks of geopolitical events? *Technological Forecasting and Social Change*, 159, 120182.

Su, Z., Lyu, T., & Yin, L. (2022). China's illiquidity premium: Due to risk-taking or mispricing? *Pacific-Basin Finance Journal*, 76, 101861.

Toda, A. A., & Walsh, K. J. (2020). The equity premium and the one percent. *The Review of Financial Studies*, 33(8), 3583-3623.

Treepongkaruna, S., Chan, K. F., & Malik, I. (2023). Climate policy uncertainty and the cross-section of stock returns. *Finance Research Letters*, 55, 103837.

Turchin, P. (2005). *War and peace and war: The life cycles of imperial nations*. New York: Pi.

Vassalou, M. (2003). News related to future GDP growth as a risk factor in equity returns. *Journal of financial economics*, *68*(1), 47-73.

Wang, F., & Xu, Y. (2004). What determines Chinese stock returns? *Financial Analysts Journal*, 60, 65-77.

Wang, Y., & Di Iorio, A. (2007). The cross section of expected stock returns in the Chinese A-share market. *Global Finance Journal*, 17(3), 335–349.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

White, H., (1980). A heteroskedasticity-consistent covariance matrix estimator and direct test for heteroskedasticity. *Econometrica* 48, 817–838.

Wong, K. A., Tan, R. S. K., & Liu, W. (2006). The cross-section of stock returns on the Shanghai Stock Exchange. *Review of Quantitative Finance and Accounting*, 26, 23-39.

Yang, D. L. (2006). Economic transformation and its political discontents in China: Authoritarianism, unequal growth, and the dilemmas of political development. *Annual Review of Political Science*, *9*(1), 143–164.

Yao, S., Luo, D., & Loh, L. (2013). On China's monetary policy and asset prices. *Applied Financial Economics*, *23*(5), 377–392.

Zeleny, M. (1982) Multiple Criteria Decision Making. McGraw Hill, New York.

Zhang, Y., & Wang, G. (2020). Compensation for illiquidity in China: Evidence from an alternative measure. *The North American Journal of Economics and Finance*, *53*, 101187.

Zhao, X. Q. (1999). Stock prices, inflation and output: Evidence from China. *Applied Economics Letters*, *6*(8), 509–511.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *67*(2), 301–32