Summer 2020

Fundamental Anomalies

Guoliang Ma

Fundamental Anomalies *

Guoliang Ma
Advisor: Cindy Yu

Abstract

This paper quantifies to what extent stock market anomalies are driven by firm fundamentals. We estimate the parameters of a 2-capital q-model (Congalves, Xue and Zhang, 2019) by matching the entire time series of stock returns at firm level using Markov Chain Monte Carlo (MCMC), instead of matching the average anomaly returns as prior studies (Liu, Whited and Zhang, 2009; Congalves, Xue and Zhang, 2019) do. Our paper resolves the critique of the prior studies that the parameter values of the model are chosen to fit a specific set of anomalies and different values are required for different anomalies. Because anomaly premiums are not the moment conditions of our estimation, our methodology provides a true test on the capability of the q-theory in explaining anomalies. We show that the model is able to generate sizable premiums for Momentum, ROE, and Asset growth.

JEL Classification:
Keywords: q-theory, Investment, BMCMC estimation

*All errors are our own.

1 Introduction

Production-based asset pricing literature is pioneered by Cochrane (1991) and further developed by Zhang (2005), Liu, Whited and Zhang (2009), and Hou, Xue and Zhang (2015) among others. This line of research studies returns from the supply side of the economy and formulate returns based on firm fundamentals, under the assumptions that firm operates to optimize its market equity. Hayashi (1982) showed that with homogeneous and degree-one production technology, a firm's return on investment, which is a function of its fundamentals, equals to its weighted average of cost of equity and cost of debt, i.e., the WACC. The empirical success of this identity in explaining the cross section of stock returns are two folds. On one hand, Liu, Whited and Zhang (2009) showed that a simple q-model fits Value, SUE and Investment anomalies well, when the parameters of the model are estimated based on this identity using the average returns of these anomaly decile portfolios as the target moments. On the other hand, Hou, Xue and Zhang (2015) developed a reduced-form q-factor model based on the insight that investment return is a function of investment-to-asset ratio and profitability. They showed that q-factor model can explain a large set of anomalies.

Our paper aims at resolving a critical weakness in the approach of Liu, Whited and Zhang (2009). In their study, and in a subsequent study by Congalves, Xue and Zhang (2019), the identity between investment return and the WACC is estimated at portfolio level. For example, targeting the average returns of 10 Book-to-Market equity portfolios, one can estimate a set of parameters and compute model-implied portfolio returns. However, when using a different set of target moments (for example, average returns of the 10 Momentum portfolios), the estimation would lead to a very different set of parameters. As Campbell (2017) (page 213) puts it: "This problem, that different parameters are needed to fit each anomaly, is a pervasive one in the q-theoretic asset pricing literature."

In this paper, we instead consider the Markov Chain Monte Carlo (MCMC) method that fully exploits the conditional information in the time series of firm-level stock returns. Our method offers two advantages in comparison to prior studies. First, our estimates of parameters are independent of any specific testing portfolios. This feature is critical for resolving the aforementioned critique by Campbell (2017). Our parameter estimates are chosen to fit the time series of firm-level stock returns and do not depend on any specific portfolios. Potentially, we can test the capability of the q-theory to explain any stock return anomaly.

The second advantage of our methodology is that given the vast amount of information exploited by the MCMC estimation, we are able to identify much more detailed information on model parameters than prior studies with only a few moment conditions. Because technology varies drastically across industries and changes over time, it is natural to allow model parameters to vary cross industry and over time.

Our estimation generates the estimates of technology parameters γ_{jt} and physical investment adjustment costs a_{jt} in the production function for each Fama-French 10-industry and for each year between 1967 to 2016. Based on the parameter estimation, we compute firm-level fundamental[1] stock returns per annum. In total, we generate 136,598 firm-year fundamental return observations. Not surprisingly, realized returns have a much wider distribution than fundamental returns at both left and right tails. The standard deviation of realized returns is 72.6% per annum, while only 51.6% for estimated fundamental returns. However, the means of realized and fundamental returns are closely matched, 15.6% vs. 15.9%.

At the market level, the value-weighted realized and fundamental market returns have a highly significant correlation of 77.1%. Moreover, the fundamental market returns successfully capture every significant ups and downs of the realized index during the sample period, such as the Internet Bubble around 2000 and the financial crisis around 2008.

[1]To clarify the terminologies from economic modeling perspective and from the perspective of statistical inference, by "fundamental", we mean model-implied when introducing models and the estimation is all about fundamental returns.

To examine the capability of the q-theory in explaining stock market anomalies, we construct 4 anomalies, each with 10 decile portfolios and one hedging portfolio (the difference between the two extreme decile portfolios, see Section 6). These 4 anomalies cover 4 major categories of anomalies: Value, Momentum, Physical investment, captured by asset growth (I/K), and Profitability, captured by ROE. We gauge the model's fit of anomalies in three ways. First, we follow Congalves, Xue and Zhang (2019) and compute the difference between average realized and fundamental portfolio returns, defined in Section 6 as α. The better the model fit, the smaller the average absolute α across decile portfolios ($\overline{|\alpha|}$) and across hedging portfolios ($\overline{|\alpha_{H-L}|}$) are. The average absolute value α's of the 40 decile portfolios is 1.38% (per annum). Second, the fundamental vs. realized portfolio returns are plotted to study the overall pattern of decile portfolios. All decile portfolios exhibit clear trend as expected except for Value portfolio. Third, the premia time series are plotted to reflect detailed information about realized premia and fundamental premia. All four premia series follow the realized premia closely.

Our paper is most closely related to Liu, Whited and Zhang (2009), Liu and Zhang (2014), and Congalves, Xue and Zhang (2019). These papers conducted GMM estimation of the q-Theory models with average anomaly portfolio returns as target moments. Liu, Whited and Zhang (2009) showed that a one-capital q-model can match the average returns of portfolios sorting on earnings surprises, book-to-market equity, and capital investment, when estimated separately. However, the parameters are significantly different when targeting portfolios are sorted on different variables. Liu and Zhang (2014) used the same model and estimation strategy to explain momentum premium. Congalves, Xue and Zhang (2019) estimated a 2-capital q-model to match the average returns of 40 decile portfolios sorted on book-to-market equity, investment-to-assets, return-on-equity, and stock returns from past 12 to 2 months. Different from Liu and Zhang (2014) and Congalves, Xue and Zhang (2019), in our paper, fundamental returns are computed at firm level rather than at portfolio level. Different from all these papers, our estimation does not depend on any specific portfolios.

Our paper is also related to Belo, Xue and Zhang (2013) and Belo et al. (2019), both of which conducted GMM estimation of a q-theory model but instead of explaining stock returns, they tried to understand market equity value. In a broader sense, our paper belongs to the large literature that try to explain the cross section of stock returns, theoretically and empirically, using production-based asset pricing models, such as Cochrane (1996), Zhang (2005), and Hou, Xue and Zhang (2015) among others.

Finally, our paper is related to the large macroeconomic literature, e.g., Smets and Wouters (2007) among others, that used MCMC method to estimate structural models and forecast macroeconomic variables based on the estimation. Li et al. (2019) extracted latent macroeconomic shocks using MCMC and test whether these macro risks can explain the cross section of stock returns.

The rest of the paper is organized as follows. Section 2 outlines the model. Section 3 discusses our estimation strategy. Section 4 explains the data used in estimation and the testing portfolios. Section 5 demonstrates the power of MCMC by analyzing simulation results. Section 6 presents the empirical estimation and analysis. Section 7 concludes.

2 The Model

We adopt the 2-capital model in Congalves, Xue and Zhang (2019), in which firms use three inputs in production: labor (L), long-term physical capital (K), and short-term working capital (W). Operating profits of firm i in a specific industry at time t is $\Pi(K_{it}) = \Pi(K_{it}, W_{it}, L_{it})$, which exhibits constant-return-to-scale. Under the assumption of a perfect competitive and frictionless labor market, labor is chosen to maximize contemporaneous operating profits. With Cobb-Douglass production technology, marginal products of physical and working capital are given by $\partial\Pi_{it}/\partial K_{i,t} = \gamma_{K,jt}\Pi_{it}/K_{it}$ and $\partial\Pi_{it}/\partial W_{i,t} = \gamma_{W,jt}Y_{it}/W_{it}$, respectively, in which $\gamma_{K,jt}$, $\gamma_{W,jt} > 0$ are the corresponding shares of capital (technology parameter components) in sales Y_{it} with technology parameter $\gamma_{jt} \equiv \gamma_{K,jt} + \gamma_{W,jt} < 1$. We allow $\gamma_{K,jt}$ and $\gamma_{W,jt}$ to be industry- and time-specific.

Firms choose investments in physical and working capital to maximize the market equity. Physical capital evolves as $\dot{K}_{it+1} = I_{it} - \delta_{it}K_{it}$. In discrete form it is $K_{it+1} = (1-\delta_{it})K_{it} + I_{it}$ in which I_{it} is the investment in physical capital, and δ_{it} is the depreciation rate. Investment in physical capital incur quadratic adjustment costs:

$$\Phi_{it} \equiv \Phi(I_{it}, K_{it}) = \frac{a_{jt}}{2}\left(\frac{I_{it}}{K_{it}}\right)^2 K_{it}, \tag{1}$$

where physical adjustment costs parameter a_{jt} is industry-specific and time varying. Working capital evolves as $W_{it+1} = W_{it} + \Delta W_{it}$, in which ΔW_{it} is investment in working capital. Following Congalves, Xue and Zhang (2019), we assume that working capital does not depreciate and is not accompanied with adjustment costs[2].

In addition to equity financing, firm i in industry j issues debt B_{it} with interest rate r_{it}^B at beginning of time t, which is repaid at the beginning of $t+1$. At tax rate τ_t, firm i's net payout is given by $D_{it} \equiv \Phi(I_{it}, K_{it}) = (1-\tau_{it})(\Pi_{it} - \Phi_{it}) - I_{it} - \Delta W_{it} + B_{it+1} - r_{it}^{Ba}B_{it} + \tau_t\delta_{it}K_{it}$, in which $r_{it}^{Ba} \equiv r_{it}^B - \tau_t(r_{it}^B - 1)$ is the after-tax interest rate. Taking the stochastic pricing kernel, M_{t+1} as given, firm i chooses I_{it}, K_{it+1}, ΔW_{it}, W_{it+1}, and B_{it+1} to maximize its cum-dividend market equity, $V_{it} = \sum_{s=0}^{\infty} M_{t+s}D_{t+s}$. The first-order condition for physical investment implies that $\mathbb{E}_t[M_{t+1}r_{it+1}^{KF}] = 1$, in which r_{it+1}^{KF} is the fundamental asset return:

$$r_{it+1}^{KF} = \frac{(1-\tau_{it+1})\left[\gamma_{jt+1}\left(\frac{Y_{it+1}}{K_{it+1}}\right) + \frac{a_{jt+1}}{2}\left(\frac{I_{it+1}}{K_{it+1}}\right)^2\right] + \tau_{t+1}\delta_{it+1} + (1-\delta_{it+1})\left[1 + (1-\tau_{t+1})a_{jt+1}\left(\frac{I_{it+1}}{K_{it+1}}\right)\right]}{1 + (1-\tau_{it})a_{jt}\left(\frac{I_{it}}{K_{it}}\right)}. \tag{2}$$

Similarly, the first-order condition for working capital investment implies that $\mathbb{E}_t[M_{t+1}r_{it+1}^{WF}] = 1$,

[2] Note there is a slight abuse of notations, we use subscript j to index industries and i to index firms. But we keep track of these subscripts separately so they are not mixed. For example, when there is need to compare industries, j_1, j_2 should be used instead of saying industry i and industry j.

in which r^{WF}_{it+1} is the fundamental return on working capital investment:

$$r^{WF}_{it+1} = 1 + (1 - \tau_{t+1})\gamma^{W}_{jt+1}\frac{W_{it+1}}{K_{it+1}}\,. \tag{3}$$

As shown in Congalves, Xue and Zhang (2019), the weighted average of the two fundamental investment returns equals the weighted average of cost of equity and after-tax cost of debt:

$$w^K_{it} r^{KF}_{it+1} + (1 - w^K_{it}) r^{WF}_{it+1} = w^B_{it} r^{Ba}_{it+1} + (1 - w^B_{it}) r^{SF}_{it+1}\,, \tag{4}$$

in which $w^B_{it} \equiv B_{it+1}/(V_{it} - D_{it} + B_{it+1})$ is the firm's market leverage, $r^{SF}_{it+1} \equiv V_{it+1}/(V_{it} - D_{it})$ is the fundamental stock return, $w^K_{it} \equiv q_{it}K_{it+1}/(q_{it}K_{it+1} + W_{it+1})$ is the weight of firm's market value attributed to physical capital and $q_{it} \equiv 1 + a_{jt}(1 - \tau_t)I_{it}/K_{it}$ is the shadow price of capital, i.e., the marginal q. The marginal q of working capital is one.

Solving for the stock market from equation (4) leads to:

$$r^{SF}_{it+1} = \frac{w^K_{it} r^{KF}_{it+1} + (1 - w^K_{it}) r^{WF}_{it+1} - w^B_{it} r^{Ba}_{it+1}}{1 - w^B_{it}}\,. \tag{5}$$

Define fundamental return of firm i from t to $t+1$ in functional form as

$$r^{SF}_{it+1} \equiv f\left(\boldsymbol{X}_{it}, \boldsymbol{X}_{it+1} | \boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}\right) = \frac{w^K_{it} r^{KF}_{it+1} + (1 - w^K_{it}) r^{WF}_{it+1} - w^B_{it} r^{Ba}_{it+1}}{1 - w^B_{it}} \tag{6}$$

where $\boldsymbol{X}_{it}$ is the set of accounting variables used in equation (2) that measure firm i's economic nature, and $\boldsymbol{\theta}_t \equiv \{(\gamma_{jt}, a_{jt}), j = 1, \cdots, 10\}$ is the set of latent variables at time t for Fama-French 10 industries. Equation (5) implies that the equality between realized stock return and fundamental return holds for any firm i and for any period from t to $t+1$. Note that this equality also holds for any linear combination of returns, e.g., the equality holds at the portfolio level.

We define r^S_{pt+1} as realized stock return at portfolio level, and r^F_{pt+1} for fundamental stock return at portfolio level. The estimated version of r^F_{pt+1} is denoted by $\hat{r}^F_{pt+1}$.

3 Estimation methodology

Prior studies (Liu, Whited and Zhang, 2009; Congalves, Xue and Zhang, 2019) matched the unconditional moment conditions derived from equation (5): $E_T[r^S_{pt+1} - \hat{r}^F_{pt+1}]$ of testing portfolio p using the General Method of Moments (GMM), where $E_T[\cdot]$ refers to the operation of taking time series average. We consider MCMC method that fully exploits the conditional information in the time series of firm-level stock returns. Next we explain our methodology in details.

In our estimation, we assume that technology parameters in production function, γ_{jt}, and physical adjustment costs, a_{jt}, are industry specific and time varying. These parameters are referred to "latent variables" in MCMC estimation and are assumed to evolve independently as random walk processes:[3]

$$\begin{bmatrix} \gamma_{jt+1} \\ a_{jt+1} \end{bmatrix} = \begin{bmatrix} \gamma_{jt} \\ a_{jt} \end{bmatrix} + \begin{bmatrix} \sigma_\gamma & 0 \\ 0 & \sigma_a \end{bmatrix} \begin{bmatrix} e^\gamma_{jt+1} \\ e^a_{jt+1} \end{bmatrix}, \tag{7}$$

where σ_γ and σ_a are the standard deviations of random noises, e^γ and e^a, which follow standard normal distribution. Imposing a random walk process on the deep parameters not only encourages persistence, but also enables us to borrow information across time in estimation, leading to more efficient estimates.

Realized stock return of firm i (in industry j) at time $t+1$ is modeled as:

$$r^S_{it+1} = r^{SF}_{i,t+1} + \sigma^r_{it} e^r_{it+1}, \tag{8}$$

where the estimation error e^r_{it+1} follows i.i.d. standard normal distribution, and the standard

[3] We also estimate an autoregressive process with order one. The estimated persistence is very close to one for both parameters. Thus, we use random walk process in our baseline model for simplicity.

deviation, σ_{it}^{r}, of estimation error is specified as:

$$\sigma_{it}^{r} = \varpi_{it}^{-1/2}\sigma_r \,, \qquad \varpi_{it} \equiv \frac{V_{it}}{\sum_{i=1}^{N_t} V_{it}} \,, \tag{9}$$

in which N_t is the number of firms at time t in our sample and σ_r is a parameter that needs to be estimated. With this specification, we introduce heteroscedasticity into the observational errors of realized stock returns by allowing the variance to be a decreasing function of firms' market equity. This reflects the fact that stock returns of larger firms are less noisy and better unveil these firms' fundamentals than the returns of small firms.[4] More importantly, specified this way, the estimated model are more economically relevant in the sense that it captures the regularity of the majority of the economy. The same rationale motives the use of NYSE breakpoints in constructing portfolios and regressions with weighted least squares in asset pricing studies (e.g., Hou, Xue and Zhang, 2015). Specifically, we set the variance function $(\sigma_{it+1}^{r})^2$ to be proportional to the inverse of the firm's relative value of market equity. Effort (not reported here) has been made to investigate other kinds of functional forms relating the variability in the observational error of a firm's stock return to its market equity; the relationship specified in the above equation (9) fits the data the best in terms of mean absolute error (mae) of firm-level stock returns.

We use inverse gamma distributions for the priors of standard deviations: $\sigma_{\gamma}^2 \sim IG(\kappa_1^{\gamma}, \kappa_2^{\gamma})$, $\sigma_a^2 \sim IG(\kappa_1^{a}, \kappa_2^{a})$, and $\sigma_r^2 \sim IG(\kappa_1^{r}, \kappa_2^{r})$, where κ_1 and κ_2 are hyper-parameters of the inverse gamma distribution (shape-scale parameterizations). The values of κ_1^{r}, κ_1^{γ}, and κ_1^{a} are specified to be 0.01, 1, and 1, respectively; the values of κ_2^{r}, κ_2^{γ}, and κ_2^{a} are chosen to be 0.02, 5, and 5, respectively. The values of κ_1's are chosen relatively small so that the data information is more likely to dominate (see Appendix A). The values of κ_2's are set relatively large so that the variances of the priors are large and thus less informative. Although these values are chosen seemingly arbitrarily, as

[4]Larger firms have more analysts following than smaller firms, thus their values is under much closer scrutiny. Moreover, stocks of larger firms are generally more liquid and their market values are less likely to be manipulated or affected by a small group of investors.

MCMC runs (so information from the data gets entered into the posterior draws), the choices of the hyper-parameters weigh less and less. Most importantly, the information from data dominates the posterior draws when MCMC converges.

Finally, the time series of latent variables $\boldsymbol{\theta} \equiv \{\theta_t, t = 1, \cdots, T\}$ and distribution parameters $\boldsymbol{\sigma} \equiv \{\sigma_\gamma, \sigma_a, \sigma_r\}$ are drawn from the joint posterior distribution, in an iterative manner from each complete conditional posterior distribution. The joint posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ is:

$$\begin{aligned} \mathcal{P}(\boldsymbol{\theta}, \boldsymbol{\sigma} | \boldsymbol{X}, \boldsymbol{r^S}, \boldsymbol{r^{Ba}}) &= \prod_{t=1}^{T} \prod_{i=1}^{N_t} \mathcal{P}^{SN} \left(e_{it}^r | X_{it}, X_{it+1}, \theta_t, \theta_{t+1}, \sigma_r^2 \right) \\ &\quad \times \prod_{t=1}^{T} \prod_{j=1}^{10} \mathcal{P}^{SN} \left(e_{jt}^\gamma | \gamma_{jt}, \gamma_{jt+1}, \sigma_\gamma^2 \right) \mathcal{P}^{SN} \left(e_{jt}^a | a_{jt}, a_{jt+1}, \sigma_a^2 \right) \\ &\quad \times \mathcal{P}^{IG} \left(\sigma_\gamma^2 | \kappa_1^\gamma, \kappa_2^\gamma \right) \mathcal{P}^{IG} \left(\sigma_a^2 | \kappa_1^a, \kappa_2^a \right) \mathcal{P}^{IG} \left(\sigma_r^2 | \kappa_1^r, \kappa_2^r \right) \end{aligned} \tag{10}$$

where $\boldsymbol{X}$ is the panel of fundamental observables, $\boldsymbol{r}^S$ and $\boldsymbol{r}^{Ba}$ are the panels of realized stock and bond returns, and $\mathcal{P}^{SN}(\cdot)$ and $\mathcal{P}^{IG}(\cdot)$ refer to the probability density functions of standard normal and inverse gamma distributions, respectively. We run 20,000 MCMC iterations, burning in the first 10,000 and confirm the convergence of the posterior draws. Appendix A details the sampling algorithm and posterior derivations.

Note that our MCMC estimation approach is fundamentally different from the estimation method in Congalves, Xue and Zhang (2019) and it offers several advantages. First, our estimates of parameter values are independent of any specific testing portfolios. We utilize the entire *distribution* of stock returns to estimate time-varying and industry specific model parameters, while Congalves, Xue and Zhang (2019) use the *moments* (aggregate weighted averages) of certain specific portfolios only to estimate constant model parameters. This feature is critical for resolving the critique of Campbell (2017) that the parameter values of the model are chosen to fit a specific set of anomalies and different values are required for different anomalies. Our parameter estimates are chosen to fit the time series of firm-level stock returns and do not depend on any specific portfolios. Potentially, we can test the capability of the q-theory to explain any stock return anomaly.

Second, our MCMC algorithm generates random draws of model parameters from their joint *posterior distribution* given the observations on firms' stock and bond returns and fundamentals, while the GMM method in Congalves, Xue and Zhang (2019) output point estimates of model parameters, which are *deterministic* given the same observations. One advantage of our Bayesian approach is that uncertainties of the estimated parameters as well as the implied stock returns can be easily quantified based on posterior draws after the burn-in period of MCMC iterations, thus inferences for estimated parameters and predicted stock returns can be readily made. We conduct simulation studies, in which we know the true parameter values, to evaluate our method and found that our MCMC approach can accurately estimate those dynamic deep parameters with small relative mean absolute errors of 3.64% and 2.68%, for γ and a respectively, and predicts annualized stock returns with mean estimation error of 0.0049.

Third, given the vast amount of information exploited by our Bayesian MCMC estimation, we are able to identify much more detailed information on model parameters than prior studies with only a few moment conditions. Because technology varies drastically across industries and changes over time, it is natural to allow model parameters to vary cross industry and over time. With the Fama-French 10-industry classification, we estimate 2×530 latent variables using the sample between June 1967 to December 2016. Although the number of parameters and latent variables are small compared to the number of observations used in our estimation $(136, 598)$, it is infeasible to estimate them via either GMM or a maximum likelihood based approach.

Lastly, although model parameters are industry and time specific, the posterior of any specific industry-time parameter utilizes the information of the entire data sample because of the random walk process imposed on parameters in equation (7). Moreover, adjustment costs at adjacent periods, a_{jt+1} and a_{jt}, enter into the probability distribution of estimation error e^r_{it+1} via equations

(2) and (10), which connects information of returns across time. In addition, firms enter and exit specific industries over time. Thus, the physical adjustment cost parameters a_{j_2t+1} and a_{j_1t} enter into the probability distribution of estimation error e^r_{it+1} if firm i switches from industry j_1 to j_2 at time $t+1$, which connects information of returns across industries. Consequently, the identification of any specific latent variable γ_{jt} and a_{jt} utilize the information of the entire data sample.[5]

4 Data

Our full sample includes all publicly traded firms in Compustat. We obtain monthly and daily stock return data from the Center for Research in Security Prices (CRSP). We include only common stocks traded on NYSE, Amex, and NASDAQ with available accounting and return data. Firms with primary standard industrial classifications between 6000 and 6999 (financial firms), firms with negative book equity, and firms with nonpositive total assets, net property, plant, and equipment, or sales at the portfolio formation are excluded. These data items are needed to calculate firm-level fundamental returns.

4.1 Measures and timing alignment

Model-implied fundamental returns are constructed in annual frequency because the needed fundamental variables such as investments are only available at annual frequency for the long sample starting from 1967. In our model, time-t stock variables are at the beginning of year t, and time-t flow variables are over the course of year t. Thus, time-t stock variables are obtained from the balance sheet of fiscal year $t-1$ and flow variables from the balance sheet of fiscal year t.

We adopt the same measures used by Congalves, Xue and Zhang (2019) for the variables needed to construct the fundamental returns. Specifically, output, Y_{it}, is measured as sales (Compustat

[5]Even though γ's at different time periods and different industries do not enter into the posterior distribution of the same estimation error, information on γ is linked with that of a, through which they are interlinked across industry and over time.

annual item SALE). Physical capital, K_{it}, is gross property, plant, and equipment (item PPEGT). Short-term working capital, W_{it}, is current assets (item ACT). Total debt, B_{it+1}, is long-term debt (item DLTT, zero if missing) plus short-term debt (item DLC, zero if missing) from time t balance sheet. Tax rate τ is the statutory corporate income tax rate from the commerce clearing house's annual publications. The depreciate rate of physical capital, δ_{it}, is the amount of depreciation and amortization (item DP) minus the amortization of intangibles (item AM, zero if missing) divided by physical capital (item PPENT). Physical investment, I_{it}, is measured as $K_{it+1} - (1 - \delta_{it})K_{it}$. The market leverage, w^B_{it}, is the ratio of total debt to the sum of total debt and market equity. The pre-tax cost of debt, r^B_{it}, is the ratio of total interest and related expenses (item XINT) scaled by total debt, B_{it+1}. The descriptive properties and the correlation matrix of these aforementioned accounting variables are provided in Table 1. All the statistics closely match those reported in Congalves, Xue and Zhang (2019).

In the model, the fundamental stock return of firm i from year t to $t+1$, r^{SF}_{it+1}, is constructed using both stock and flow variables. Therefore, we align r^{SF}_{it+1} with the observed annual return of firm i from the middle of fiscal year t to the middle of fiscal year $t+1$, following Liu, Whited and Zhang (2009) and Congalves, Xue and Zhang (2019). Specifically, if the fiscal end of year t is month τ, the observed counterpart of r^{SF}_{it+1}, r^{S}_{it+1}, is the 12-month accumulated return between month $\tau - 5$ and $\tau + 6$. The fundamental returns for these 12 months are identical and are annualized returns albeit in monthly frequency. Note that the estimated model parameters are industry and time specific. One subtle timing mismatch arises when firms in the same industry have different fiscal ends. In that case, the parameters of year t, $\boldsymbol{\theta}_t$, could apply to slightly different periods in calendar time for firms with different fiscal ends.

The posterior means of parameter draws, generated from equation (10), are used to construct firm-level fundamental returns, based on which fundamental portfolio returns are constructed.[6]

[6]In theory, we could generate the posterior distribution of the fundamental stock return for firm i at time $t+1$, r^{SF}_{it+1}, for every firm-year observation. In turn, we can also generate the posterior distribution of time-t returns on any portfolio. However, doing so requires months of computational time even on our server with multiple cores.

Even though firm-level fundamental returns change annually (in fiscal year), fundamental portfolio returns change monthly because of the variation in fiscal year ends across firms. Fundamental portfolio returns are still annualized returns.

While realized portfolio stock returns are for a given month, the matching fundamental returns are based on annual accounting variables both prior to and after the month. To better align the timing and make fair comparison, we follow Congalves, Xue and Zhang (2019) and compound the realized portfolio stock returns within a 12-month rolling window with the month in question in the middle of the window. Specifically, we multiply gross returns from month $t-5$ to month $t+6$ to match the fundamental returns constructed in month t. Applying this rolling procedure to the monthly portfolio returns (January 1967-June 2017) yields the monthly observations of annualized portfolio returns from June 1967 to December 2016.

4.2 Anomalies

We ask whether model-implied fundamental stock returns exhibit the same anomalous patterns observed in the data. We explore four anomalies in three categories: one value-versus-growth anomalies sorted on book-to-market equity ratio (Bm), one momentum anomaly sorted on the prior 11-month returns skipping the most recent month (R^{11}), one investment anomaly sorted on asset growth (dA/A) and one profitability anomaly sorted on return-on-equity (ROE). The detailed definition of each variable and the construction of testing portfolios are described in Appendix C.

Table 2 presents the monthly average excess returns of the 10 decile portfolios sorted on each of the 4 anomaly variables. The t-statistics adjusted for heteroscedasticity and autocorrelations are reported in parentheses. "L" denotes the lowest decile, "H" the highest decile, and "H-L" the high-minus-low decile. As in Hou, Xue and Zhang (2019), decile portfolios are formed with NYSE breakpoints and value-weighted returns to control for microcaps. The sample period is from January 1967 to June 2016 for all anomaly variables. It is clear that all 4 anomalies are all statistically and economically significant in our sample period.

5 Simulation Study

5.1 Simulation Setup

We combine information from the observed explanatory variables and the model (2) to generate the simulation data. A subset (instead of customized set) of the full data set is chosen to approximate the real-world behavior of the fundamental variables. We restrict the subset to a balanced panel so that the random walk sequences of the latent variables, as implied by the model structure, are preserved. We use simulated technology parameters and adjustment costs to generate the stock returns. Among the 15,780 observations in the subset, 2,750 contain missing values[7] so we exclude them from the simulation set. At last, our simulation data panel contains 15 years of 13,030 observations, spanning from 1991 to 2005. The simulation data set consists of 1,052 firms from 7 industries. Since the fundamental variables were winsorized once for the full data, we didn't further winsorize for the simulation subset.

We simulate random walk processes for technology parameters and adjustment costs separately for each of the 7 industries. Technology parameter is a non-negative value between 0 and 1 by its economic nature. The first-year technology parameter γ_{j0} for industry j is simulated as a logistic transformation of a standard normal random variable to ensure the range constraint. Also note that since first-year technology parameters are not estimated, the generating mechanism has no real effect on the whole simulation setup. We then simulate 7 random walk sequences of length 15 from the first-year technology parameters based on equation (7). A small variance, $\sigma_\gamma^2 = 0.01$, is used so that no technology parameters exceed the $(0, 1)$ boundary. The whole sequences are

[7] The missing rate is about 17.4%, which is much lower than the actual missing rate of our full data compared to its "balanced" panel, 84.2%.

generated as a whole; if one number gets outside the boundaries, the whole sequence is discarded and regenerated. Thus the random walk structure is preserved.

Unlike technology parameters, adjustment costs have fewer restrictions. According to the literature (e.g., Congalves, Xue and Zhang (2019)), adjustment costs should be positive without a certain upper bound. To reflect the fact that adjustment costs are larger than technology parameters, we generate the first-year adjustment costs, a_{j0} from a iid $\mathcal{N}(5, 0.09)$ distribution. The 7 adjustment cost sequences then evolve according to equation (7) with variance $\sigma_a^2 = 0.09$. The whole sequences are generated in the same way as technology parameters, i.e. retaining only sequences that are entirely in the restricted range.

Table 3 presents the time-series mean of the simulated technology parameters and adjustment costs, along with the estimated values for each industry. We also report the variability, measured by standard deviation within each sequence. We see both larger magnitude in level and in variability of adjustment costs than of technology parameters, as expected from the way they are generated. The technology parameters vary from 0.33 to 0.65, and the adjustment costs vary from 3.38 to 6.22. The within-sequence variability for technology parameters $\gamma_{j\cdot}$ is about 0.1 and the within-sequence variability of $a_{j\cdot}$ is about 0.5, larger than the expected 0.3. Because the sequence length is rather short (15 years), we acknowledge this larger variability, and keep the generated adjustment cost sequences.

Expected asset returns are then calculated from equation (2), with the real-world fundamental variables in the simulation subset and the simulated latent variables. Heteroscedasticity is also considered to reflect that larger firms has stabler stock returns. Equation (8) shows that

$$r_{it+1}^{S}/\sigma_{it}^{r} = r_{it+1}^{SF}/\sigma_{it}^{r} + e_{it+1}^{r}.$$

Thus we first transform the simulated asset return to a *weighted scaled asset return*, ret_{it+1}, which is the sum of the weighted scaled *stock* return and a constant, as

$$ret_{it+1}^F = \sqrt{\varpi_{it}} \times \left(r_{i,t+1}^{SF} + \frac{w_{it}^B r_{i,t+1}^{Ba}}{1 - w_{it}^B} \right) = \frac{\sqrt{\varpi_{it}}}{1 - w_{it}^B} r_{i,t+1}^{KF}.$$

Noting that ret_{it+1} is a function of the fundamental variables and latent variables, for ease of later discussion, we denote the quantity ret_{it+1} by

$$ret_{it+1} \coloneqq \Lambda(\gamma_{it+1}, a_{it+1}, a_{it}, \varpi_{it}) + e_{it+1}^r. \tag{11}$$

We then add normal white noise with standard deviation of 5% to the calculated ret_{it+1}^F (i.e. $e_{it+1}^r \sim \mathcal{N}(0, 0.05^2)$)

$$ret_{it+1} = ret_{it+1}^F + e_{it+1}^r.$$

The number 5% is chosen is chosen to mimic the reality that most firms earn 5%-10% annual returns. The estimation is then carried out based on ret_{it+1} instead of asset return itself for statistical convenience.

We compare the simulated version of weighted scaled asset return, ret_{it+1} to the observed counterpart ret_{it+1}^{real} and found very similar behavior. Specifically for skewness, which is of more interest due to the behavior of the stock returns, the realized skewness of ret_{it+1}^{real} is 3.0459 and that of the simulated version is 3.3303. It is also worth noting that in reality, all returns are non-negative, but in simulation, 671 of the observations are less than 0, mounting to 4.25% of the simulation data set. We keep all these observations because they also account for the variability of the parameters.

5.2 Choices of Initial Values

We use a "rolling" least squares to calculate the initial values for MCMC algorithm. The least squares consists two steps. In the first step, a_{j0} is assumed to be the same as a_{j1}, $\forall j \in \{1, \ldots, 7\}$.

For specific choices of γ and $a_{j0} = a_{j1} \equiv a$, the residual sum of squares (RSS) for industry j is calculated as

$$\sum_{i \in E_{1j}} (ret_{it+1} - \Lambda(\gamma, a, a, \varpi_{it}))^2$$

where E_{tj} is a set of firms which belong to industry j at time $t-1$ so E_{1j} consists firms belonging to industry j at time 0, (year 1990). We minimize the RSS over (γ, a) to get the initial values $\gamma_{j1}^{(0)}$ and $a_{j0}^{(0)}$. In the next step, a rolling window is applied to calculate the initial values for other latent variables in their random walk sequences. From equation (2), once a_{jt} is given, the function is linear in γ_{jt+1} and a_{jt+1} so in this step ordinary least squares is adopted. In particular, to get the initial values for γ_{j1} and a_{j1}, we minimize

$$\sum_{i \in D_{1j}} \left(ret_{it+1} - \Lambda(\gamma_{j1}, a_{j1}, a_{j0}^{(0)}, \varpi_{it}) \right)^2$$

over γ_{1j} and a_{1j}, where D_{1j} is the set of firms that belong to industry j at time 1. Continuing with the rolling scheme, we get all the initial values $\gamma_{jt+1}^{(0)}$, $t = 1, 2, \ldots, T$ and $a_{jt+1}^{(0)}$, $t = 0, 1, \ldots, T$ to initiate MCMC.

We set the initial values for σ_γ^2 and σ_a^2 to 0.01 and 0.50 respectively so that they are away from the true values. The initial value for σ_ε^2 is set to be 0.25[8]. For econometric robustness, we also use another set of initial values. In that settings, we set γ_{jt} and a_{jt} to be time-invariant, industry-invariant constants. We ran MCMC algorithms with different (though arbitrary) chosen "constant" initials and the algorithm still converges to the true value, showing our method is robust to the choices of initials.

[8] It turns out that the initial value for σ_ε^2 has little effect on estimation results as MCMC runs

5.3 Simulation Results

We ran MCMC algorithm for 20,000 iterations and discard the first 10,000 iterations as burn-in. The results reported in later discussion are based on the posterior distribution formed from the 10,001 to 20,000 iterations. Columns $\hat{\gamma}$ and $\hat{a}$ in Table 3 present the estimated values of latent variables aggregated at industry level (the reported values are time-series averages of posterior means). The largest deviation of technology parameters is 0.07, with relative deviation 9.23%. The average deviation across industries is 0.03. For adjustment costs, the largest deviation from true value is 0.32, with corresponding relative deviation 5.19%; the average deviation across industries is 0.10, suggesting the precision of latent variables estimated by MCMC algorithm.

We plot the true latent variables described in Section 5.1, along with the posterior means, posterior 2.5% and 97.5% quantiles of posterior draws in Figure 1. The blue curves show the true value. The thick black curves presents the posterior means and the dashed curves delineate the upper and lower 2.5% quantiles. We also present the initial values as indicated by "GMM"[9]. The posterior means captures the trend for all industries for both technology parameters and adjustment costs. Most of the posterior 95% bands also contain the true values of latent variables. Comparing to the initial values, we also see great improvements as few points deviates farther from the true values.

We can also see some latent variables estimates by MCMC deviate more than the others. This is caused by outliers from the simulated weighted scaled asset returns. The denominator of equation (2) shows that when the value of $(1-\tau_{it})a_{jt}\frac{I_{it}}{K_{it}}$ is close to -1 the asset return will be substantially inflated[10]. When this is the case, the mean of the posterior distribution will be dragged away from its "normal" value, as can be seen from Appendix A. As the MCMC algorithm proceeds, smaller deviations are magnified by the influenced mean, resulting in less precise estimates compared to places without such outliers. Although there are some true latent variables falling outside the

[9] As the name suggests, we use quotation marks because the rolling nature distinguishes our method from the standard GMM methods. The terms "initial" and "GMM" are used interchangeably in later sections.

[10] This could happen when firms have negative investments.

credential bands, the majority of true values are bounded by the 95% MCMC boundaries. This shows that our estimation is precise and overall significant.

Next we study the estimated asset and stock returns. Since our estimation is carried out in the form of weighted scaled asset return, we first transform back to asset/stock returns. With the posterior means $\hat{\gamma}_{jt}$ and $\hat{a}_{jt}$, we calculate the estimated weighted scaled asset return as $\widehat{ret}_{it+1} = \Lambda(\hat{\gamma}_{jt+1}, \hat{a}_{jt+1}, \hat{a}_{jt}, \varpi_{it})$. The fundamental (estimated) asset return is then calculated as

$$\hat{r}^K_{it+1} = \widehat{ret}_{it+1} \times \frac{1 - w^B_{it}}{\varpi_{it}}. \tag{12}$$

The fundamental (estimated) stock return is

$$\hat{r}^S_{it+1} = \frac{\hat{r}^K_{it+1} - w^B_{it+1} r^{Ba}_{it+1}}{1 - w^B_{it+1}}. \tag{13}$$

Panel A, Table 4 presents the mean absolute error (column "MAE", calculated as $\frac{1}{T}\frac{1}{N}\sum_{t=0}^{T-1}\sum_{i=1}^{N} |r^h_{it+1} - \hat{r}^h_{it+1}|$ where $h \in \{K, S\}$ for asset and stock returns respectively) and residuals (column "residual", calculated as $\frac{1}{T}\frac{1}{N}\sum_{t=0}^{T-1}\sum_{i=1}^{N}(r^h_{it+1} - \hat{r}^h_{it+1})$ where $h \in \{K, S\}$ for asset and stock returns respectively) of 3 estimation methods. The row "MCMC" indicates our MCMC estimation. "Initial" presents the estimated returns calculated from the rolling "GMM" methods, which is also the initial value of our MCMC algorithm. Congalves, Xue and Zhang (2019) adopted GMM estimates, assuming that the technology parameters and adjustment costs are both time- and industry-invariant. Row "flat" in Table 4 adopts their assumption and estimate the time- and industry-invariant latent variables by minimizing RSS. The results show our MCMC estimates outperform both rolling least squares and flat GMM.

In terms of MAE, MCMC wins by a narrow margin. Nonetheless, in terms of residuals, on average, the other two methods provide residuals one order of magnitude of MCMC. Figure 2 illustrates this pattern. Although residuals from MCMC and residuals from "GMM" look very alike, MCMC residuals are a little more concentrated at 0, resulting in the numeric differences. In other words, the MCMC residuals are distributed more symmetrically around 0, than the "GMM" residuals are. In anyways, since our main goal is to find the weighted returns of portfolios, residuals provide us with a better metric for precision. This simulation study validates that our MCMC method is able to capture the dynamics of time varying and industry varying technology and adjustment cost parameters reasonably well, and can estimate asset and stock returns with precision, provided that the model is correctly specified.

6 Estimation Results

6.1 Parameter estimates

MCMC algorithm outputs the posterior distributions of technology parameters in the production function γ_{jt} and physical investment adjustment costs a_{jt} for each Fama-French 10-industry and for each year between 1967 to 2016. For illustration purposes, we plot the time series of the posterior means and 95% credential bounds for γ_{+t} and a_{+t} in Figure 3, where the symbol $+$ means averaged across industries. Credential bounds/intervals distinguish themselves from confidence intervals in that prediction intervals are *constant* while confidence intervals are *random* (or realizations of random intervals). We assume parameters to be themselves *random variables*, thus can easily calculate the probability that they fall in certain ranges. Notationally, let p be the probability of interest, and I_P denote the symmetric posterior prediction interval for θ, then it satisfies $\mathbb{P}(\theta \in I_P|\boldsymbol{X}) = p$. The narrow prediction bounds around the posterior means in Figure 3 indicate that the model latent variables can be accurately identified by the data under the assumed model structure. In general, there are no noticeable trends in both latent variables, suggesting that the contribution of physical capital to output relative to labor has not changed substantially over the past 50 years. It is likely that as capital becomes more advanced, required human capital becomes more critical to production as well. As such, the contribution of capital and labor to output stays relatively stable. At the same time, we do see fluctuations in the estimated latent variables.

Table 5 shows the time series average of the posterior means and 95% prediction intervals of γ_{j+} and a_{j+} for each industry, where + means average by time. In the table we explicitly distinguish the industries by their economic meaning instead of indicator variables to better explain our findings. The intervals are calculated by averaging γ_{j+}'s within each MCMC iteration, and then calculating the posterior quantiles from each γ_{j+} posterior distribution. The notation $[\cdot]_{TS}$ is used to denote time-series standard deviation, which measures the variability across time. It is the time-series standard deviation of the posterior means for each industry. In general, parameter values can differ greatly across industries, echoing our intuition that different industries adopt very different production technologies. The magnitude of these estimates are largely consistent with our economic intuitions. For example, γ_{++} is estimated to be 0.08 on average, compared to 0.28 for telecom sector, consistent with the fact that capital is less important for wholesales and retail sector than for telecom sector. The other noteworthy point is that within each industry, the technology parameters and adjustment costs vary a lot, as is shown by the aggregated plot in Figure 3.

6.2 Overall fit of the estimation

Similar transformations as shown in equations (12) and (13) (from $\widehat{ret}_{it+1}$ to asset returns $\hat{r}^K_{it+1}$ and $\hat{r}^S_{it+1}$) are carried out for empirical analysis. Panel B, Table 4 provides the estimated MAE and residuals for empirical analysis. Again we see narrow margins between MCMC estimates and "GMM"/flat estimates in MAE. A notable difference from simulation study is the differences between residuals, where we see only slight improvements of MCMC. Nonetheless, we still see the residual drops between 40% and 67% comparing with "GMM" and flat estimation ,though the improvements are not as stunning as in simulation study.

Based on the posterior means of the model parameters, we compute firm-level fundamental stock returns per annum. In total, we have 136,598 firm-year observations. Figure 4 plots the histograms of realized (in orange) and fundamental (in blue) firm-level returns. Not surprisingly, realized returns have a much wider distribution than fundamental returns at both left and right tails. The standard deviation of realized returns is 71.5% per annum, while only 51.4% for fundamental returns. However, the means of realized and fundamental returns are closely matched, 15.5% vs. 15.9%.

Figure 5 plots the time series of value-weighted realized (in orange) and fundamental (in blue) market returns. We adopt common practice by weighting firm level returns with their relative capitalization, measured by their size/market equity relative to the market total. Size/market equity is calculated as the total shares outstanding multiplied by the share prices. These two time series have a highly significant correlation of 77.1%. Noticeable, the fundamental market returns successfully capture every significant ups and downs of the realized index during the sample period, such as the Internet Bubble around 2000 and the financial crisis around 2008.

6.3 Anomalies

Another goal of ours is to identify anomalies documented by literature. Starting from Fama and French (1992), mimicking portfolios are used to identify factor anomalies. The identification relies on forming portfolios by sorting firms. For instance, the value investment suggests investing on value firms, which are captured by high book-to-market equity. To identify the value anomaly, we sort firms by their book-to-market equity each year, and then form 10 portfolios for each decile partitioned by NYSE breakpoints. These 10 portfolios are called decile portfolios. The other portfolio which longs the H(ighest) decile while shorts the L(owest) decile simultaneously is called hedging portfolio. The weighted returns of the hedging portfolio are calculated as the Value anomaly. The hedging portfolio returns are also referred to as Value premium. For Value decile portfolios, literature has also identified a growing trend in weighted returns against the order of the 10 decile portfolios, as shown in the first row of Table 2. Based on our estimation, we can also calculate the estimated fundamental Value anomalies, as opposed to the realized Value premium.

If the patterns of the fundamental anomalies line up with the realized Value anomaly, it would indicate that our methods help capture the value premium, contributing to the literature that anomalies exist fundamentally.

We gauge the model's fitting of anomalies in three ways. First, we follow Congalves, Xue and Zhang (2019) and compute the difference between average realized and fundamental portfolio returns, defined as $\alpha \equiv E_T[r^S_{pt+1} - \hat{r}^F_{pt+1}]$. The better the model fits the data, the smaller the average absolute α across decile portfolios, $\overline{|\alpha|}$, and across hedging portfolios, $\overline{|\alpha_{H-L}|}$, are. Table 6 exhibits the α's of the 40 decile portfolios and the 4 hedging portfolios. All returns are in percent per annum. 4 out of 10 α's are significantly different from 0 for Value anomaly. 2, 1, and 3 out of 10 α's are different from 0 for momentum, asset growth and profitability. Overall, our estimation matches the realized trend, except that the Value pattern is not matched as well as others.

The second way is to plot the realized vs. fundamental portfolio returns. As shown in Figure 6. The horizontal axes show the rank of decile portfolios sorted by different variables (book-to-market equity for Value, Asset Growth, Momentum, and roe for Profitability). The vertical axes show the realized/fundamental returns of the portfolios. Lower ranking portfolios are plotted on the left while higher ranking portfolios are plotted on the right. The orange curves show the realized returns for decile portfolios and the blue curves show the fundamental returns. The vertical distances between the right most point and left most point on same curves are the factor premia. The most ideal estimation would yield two closely matched curves. As shown in Figure 6, the trends and premia are matched well except for book-to-market equity.

The third way is to plot the premia by factors over time, as is shown in Figure 7. Figure 6 shows us the aggregated returns of decile portfolios averaged over time. We can also take a closer look at the premia over time because our portfolios are formed annually. Plotting each of the 10 decile portfolio returns over time would be daunting and not at all more informative. Thus we choose to plot the factor premia time series instead. As shown in Figure 7, except for some specific points, the estimated fundamental premia follow the realized premia closely, with more moderate volatility.

7 Conclusion

In this paper, we propose a full-information Bayesian MCMC estimation of a simple 2-capital q-model. The estimation is able to generate estimates of model deep parameters that are industry-specific and time varying. Firm-level fundamental returns imputed based on the parameter estimates largely match the realized returns. More importantly, the model is able to generate sizable premiums for Momentum, ROE, Asset growth.

In future work, we expect three potential extensions. In trying to solve the book-to-market mismatching, we'll extend the basic model to incorporate working capital directly. To investigate the sources of anomalies more precisely, time-variation and industry-variation effects will be studied individually. Last but not least, one-step portfolio returns prediction is also one topic that will be studied.

Acknowledgement

Dr. Erica Li from CKGSB has provided invaluable inputs and guidance from the onset of this project. We would like to thank Dr. Shujing Wang from Tongji University for her contributions in data collection, pre-processing and calculation of portfolio returns.

References

Belo, Frederico, Chen Xue and Lu Zhang. 2013. "A supply approach to valuation." Review of Financial Studies 26:3029–3067.

Belo, Frederico, Vito Gala, Juliana Salomao and Maria Ana Vitorino. 2019. "Decomposing Firm Value." Working paper, University of Minnesota.

Campbell, John Y. 2017. Financial Decisions and Markets: A Course in Asset Pricing. Princeton University Press.

Cochrane, John H. 1991. "Production-based asset pricing and the link between stock returns and economic fluctuations." Journal of Finance 46:209–237.

Cochrane, John H. 1996. "A Cross-Sectional Test of an Investment-Based Asset Pricing Model." Journal of Political Economy 104(3):572–621.

Congalves, Andrei, Chen Xue and Lu Zhang. 2019. "Aggregation, Capital Heterogeneity, and Investment CAPM." The Review of Financial Studies p. forthcoming.

Fama, Eugene F. and Kenneth R. French. 1992. "The cross-section of expected stock returns." Journal of Finance 47:427–465.

Hayashi, Fumio. 1982. "Tobin's marginal q and average q: A neoclassical interpretation." Econometrica 8:213–224.

Hou, Kewei, Chen Xue and Lu Zhang. 2015. "Digesting Anomalies: An Investment Approach." Review of Financial Studies 28(3):650–705.

Hou, Kewei, Chen Xue and Lu Zhang. 2019. "Replicating anomalies." forthcoming.

Li, Erica X.N., Haitao Li, Shujing Wang and Cindy Yu. 2019. "Macroeconomic risks and asset pricing: evidence from a dynamic stochastic general equilibrium model." Management Science 65(8):3449–3947.

Liu, Laura X. L. and Lu Zhang. 2014. "A neoclassical interpretation of momentum." Journal of Monetary Economics 67:109–128.

Liu, Laura X.L., Toni W. Whited and Lu Zhang. 2009. "Investment-Based Expected Stock Returns." Journal of Political Economy 117(6):1105–1139.

Smets, Frank and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." American Economic Review 97(3):586–606.

Zhang, Lu. 2005. "The Value Premium." Journal of Finance 60(1):67–103.

Figure 1: **Histogram of Residuals from MCMC and "GMM" algorithms**

The blue curve shows the true latent variables. Thick Black curves indicate posterior means, while dashed black curves show upper and lower 2.5% quantiles of the posterior distribution. Green curves are the initial values.

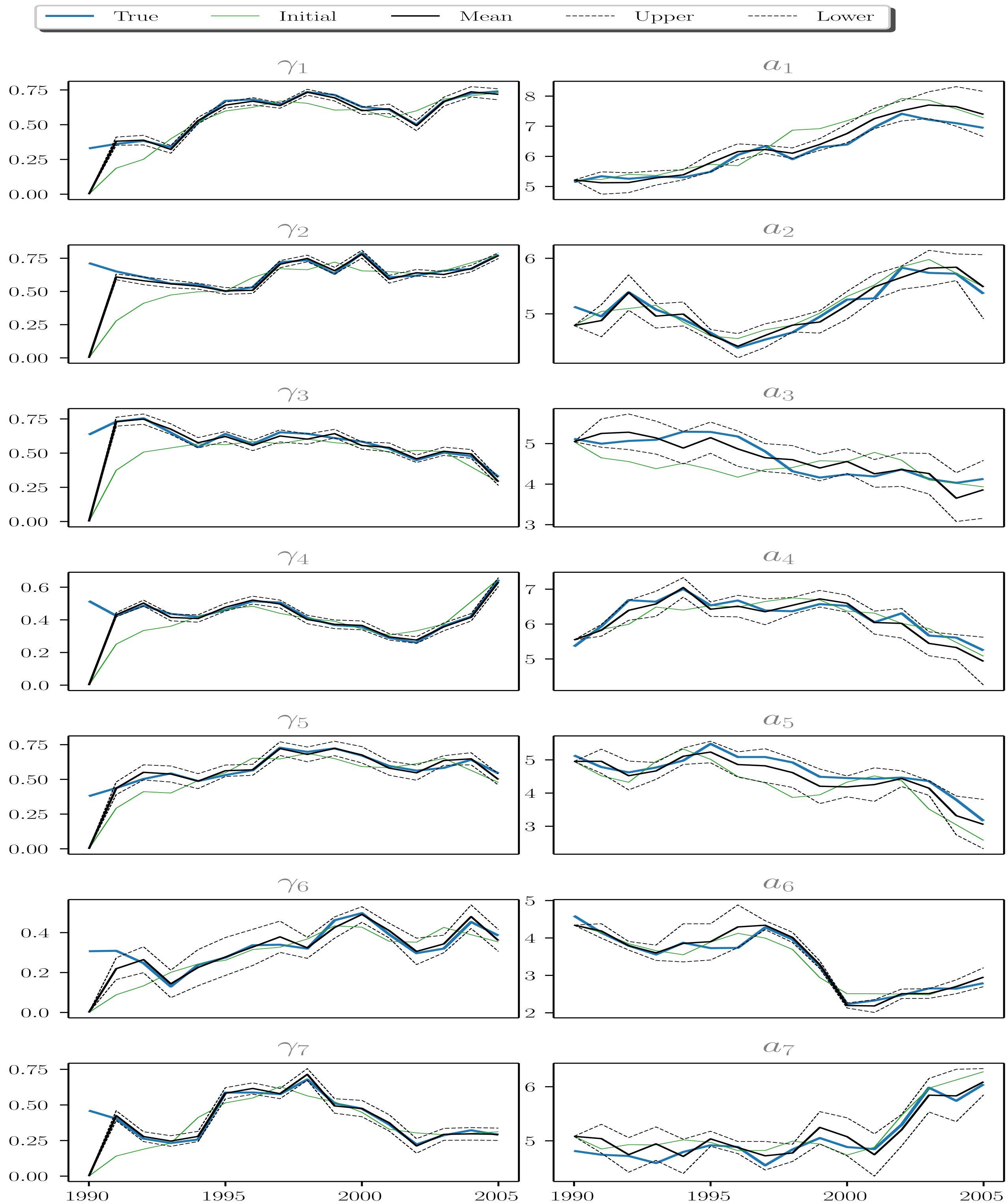


Figure 2: **Histogram of Residuals from MCMC and "GMM" algorithms**

This figure shows overlapped histograms for both residuals obtained from MCMC estimates and "GMM" estimates. The residuals are of firm-level stock returns. The blue histogram is for MCMC, which is more concentrated and symmetric around 0 than the orange histogram for "GMM".

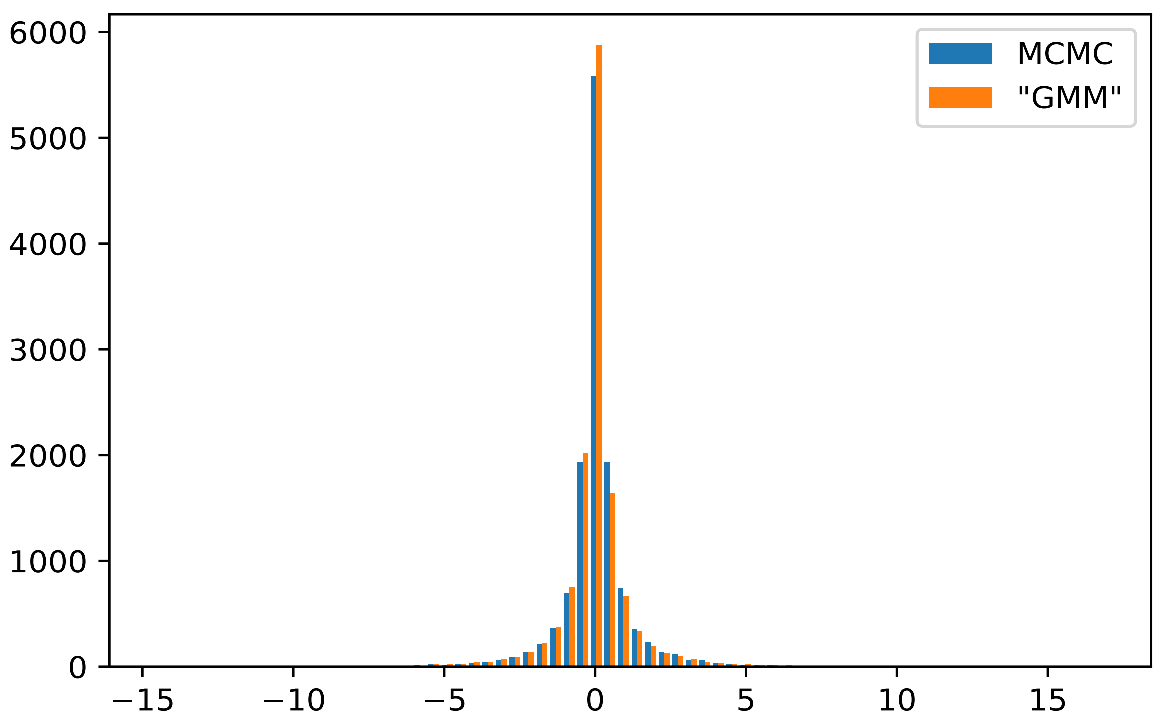
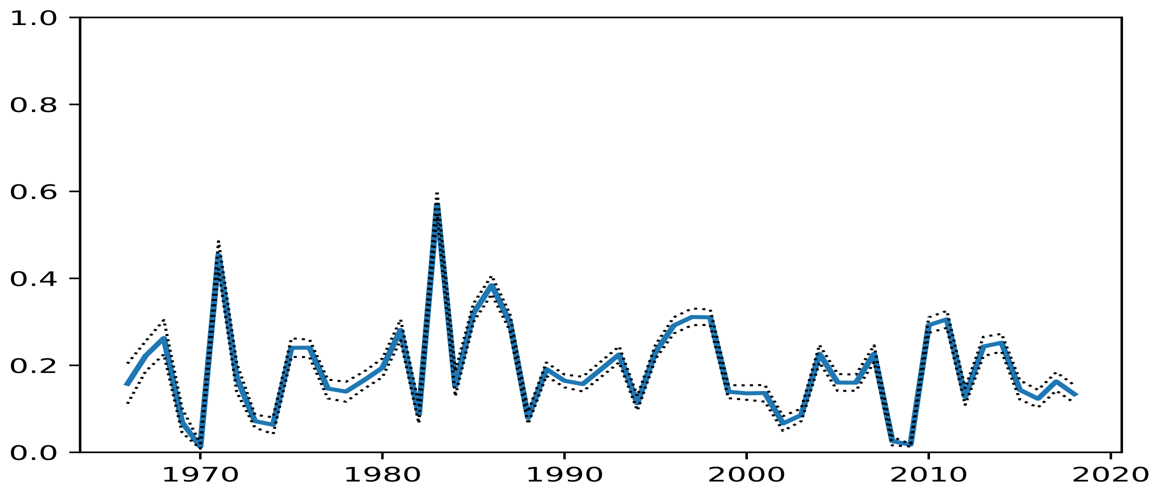
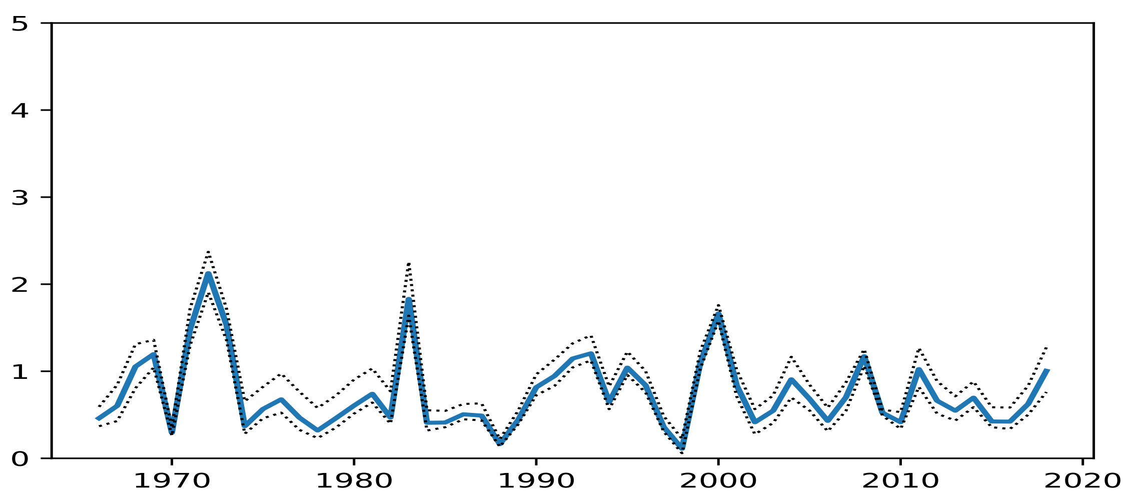


Figure 3: **Time series of estimated parameters**

This figure shows the time series (1967-2016) of the posterior means of the technology parameters in the production function, γ_{+t}, and physical investment adjustment costs a_{+t}, averaged across the Fama-French 10 industries. Parameter values are plotted in solid lines and the 95% prediction intervals are in dotted lines.



(a) Capital-to-ouput ratio: γ



(b) Investment adjustment cost parameter: a

Figure 4: **Distribution of firm-level returns: realized vs. fundamental**

This figure shows the histograms of realized (in orange) and fundamental (in blue) firm-level returns from June 1967 to December 2016. Returns are in annual frequency and in percentage. The number of observations is 136,598. Observations are trimmed at 99 percentile for illustration purposes.

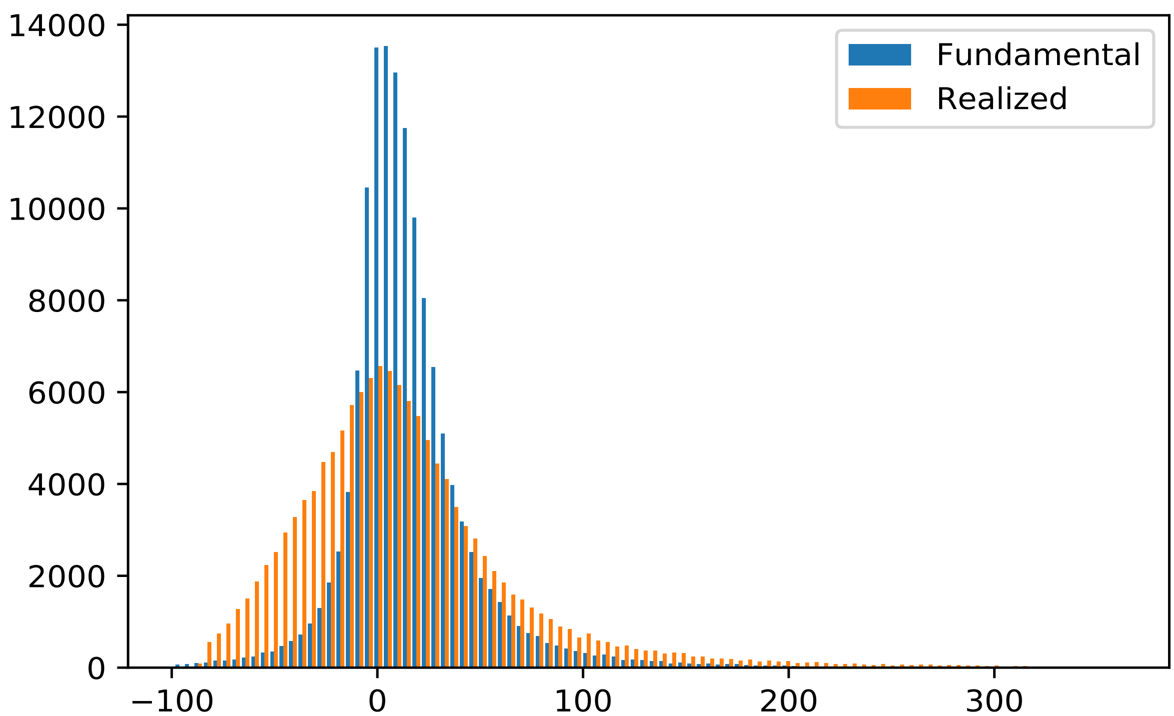


Figure 5: **Market return: realized vs. fundamental**

This figure shows the time series of market returns computed from realized (in orange) and fundamental (in blue) firm-level returns, respectively, from June 1967 to December 2016. Returns are annualized and in percent. For each time point, the plotted values are market capitalization-weighted returns of firms from the whole market.

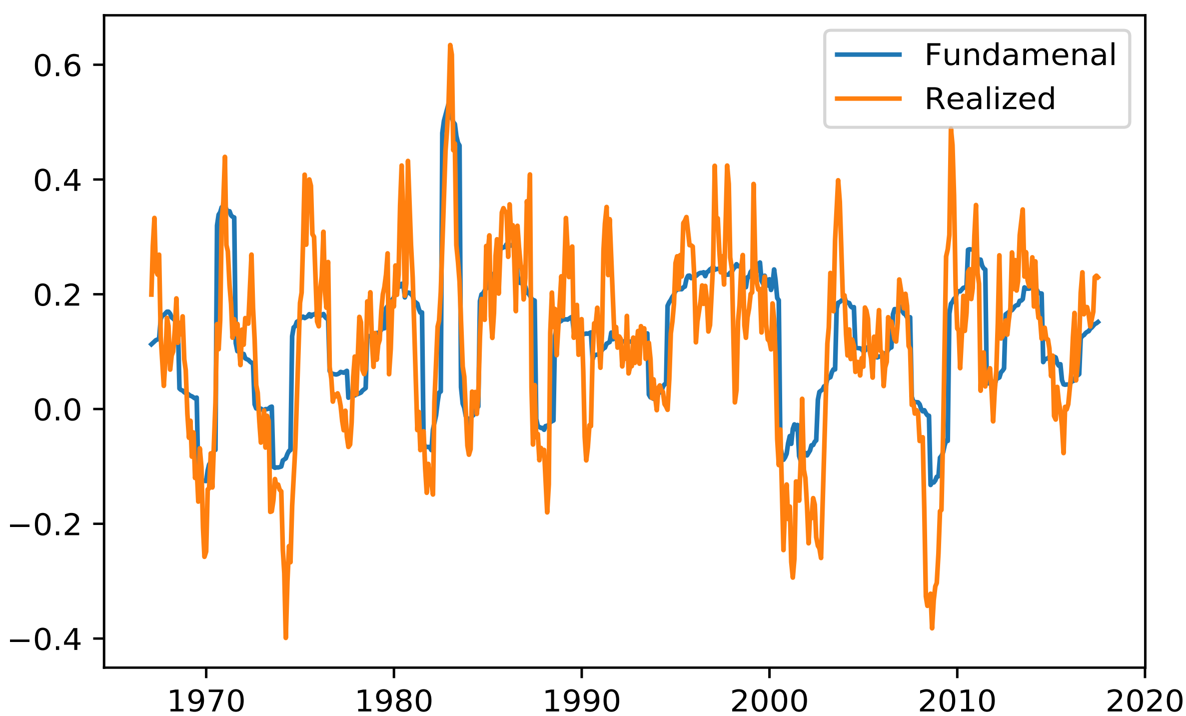
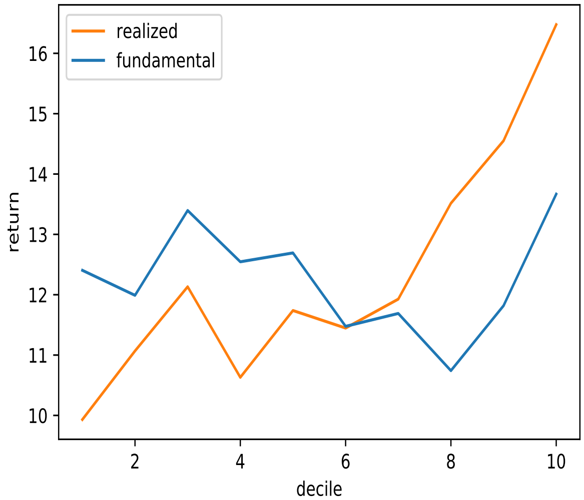
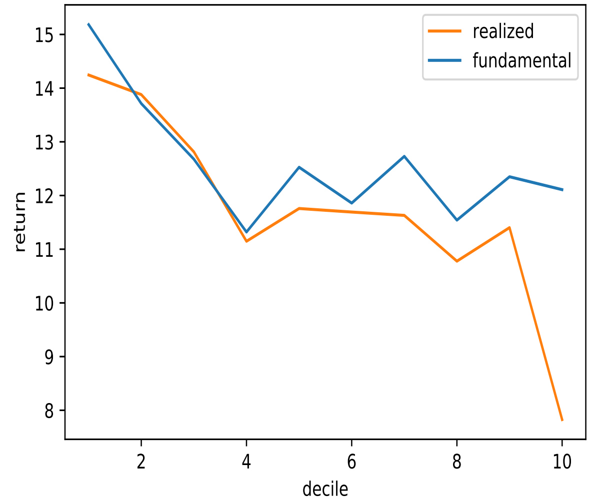


Figure 6: **Anomalies: realized vs. fundamental**

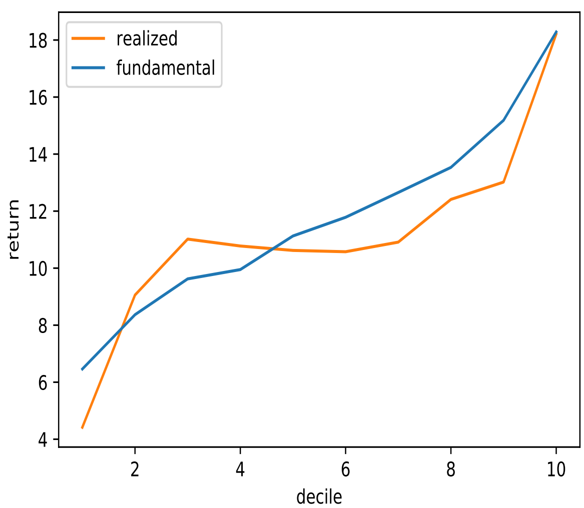
This figure plots the 40 decile portfolio returns for Value, Asset growth, Momentum and ROE sorted stocks. Realized decile portfolios are in orange while the fundamental are in blue. A perfect match would be spotted if the two lines match up. The vertical distance between the right most point and the left most point on one curve gives the corresponding factor premium.



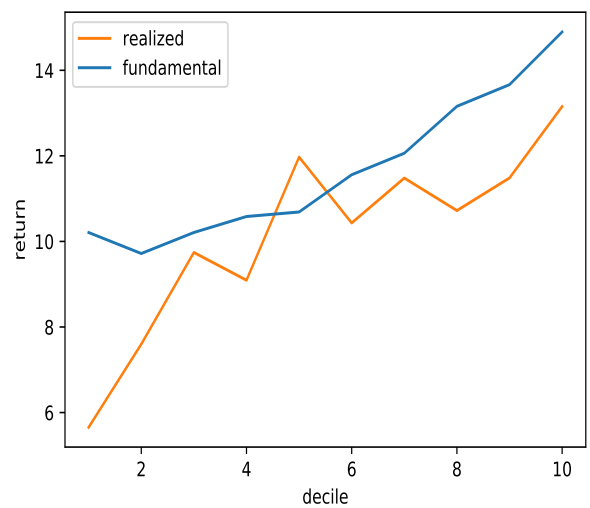
(a) Value



(b) Asset growth



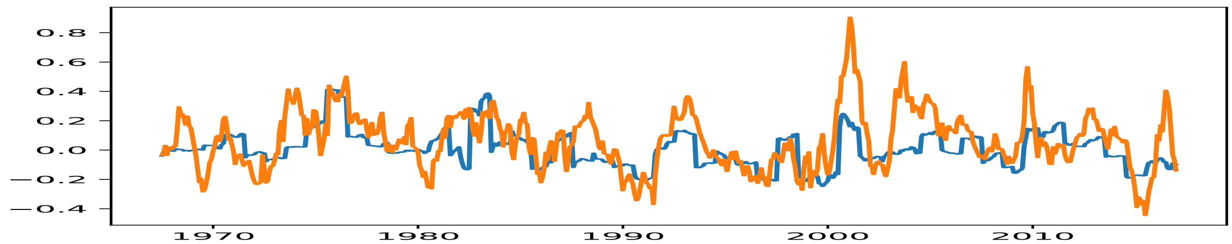
(c) Momentum



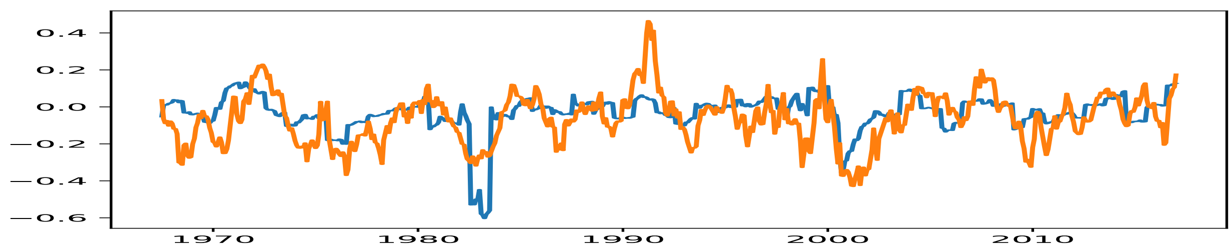
(d) Profitability (ROE)

Figure 7: **Time series of factor premia: realized vs. fundamental**

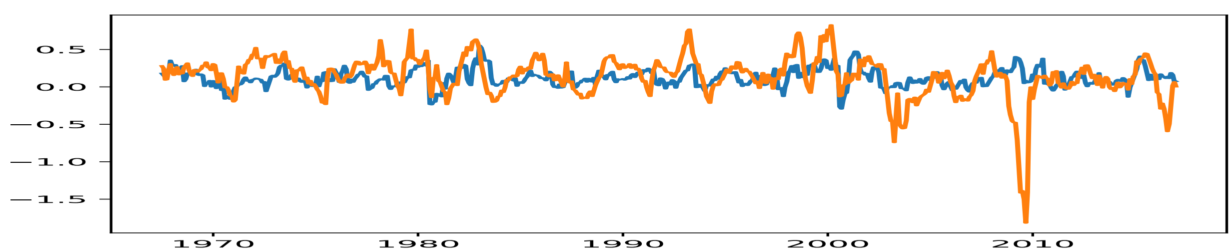
This figure plots the time series of realized (in orange) and fundamental (in blue) factor premiums. Returns are in percent per annum and in monthly frequency.



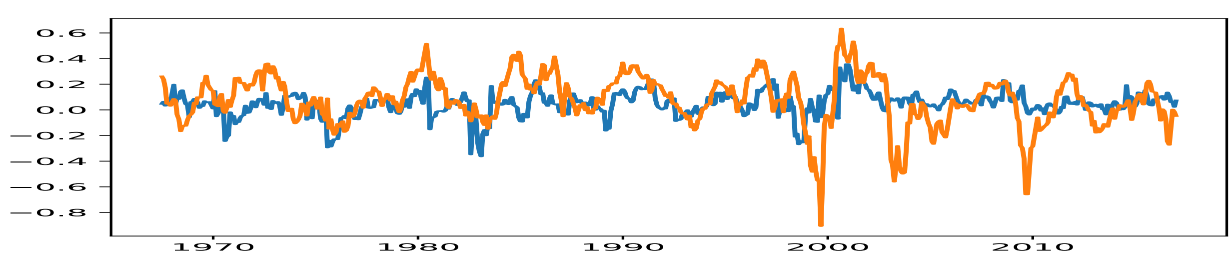
(a) Value



(b) Asset growth



(c) Momentum



(d) Profitability (ROE)

Table 1: **Descriptive statistics of firm-level accounting variables in the fundamental returns**

This table reports the time series averages of cross-sectional summary statistics (Panel A), including mean, standard deviation, percentiles (5th, 25th, 50th, 75th, and 95th), and pairwise correlations (Panel B) for firm-level annual accounting variables. The sample for the fundamental return is from June 1967 to December 2016.

Panel A: Summary statistics

	Mean	StdDev	p5	p25	p50	p75	p95
I_{it}/K_{it}	0.368	0.481	-0.041	0.108	0.220	0.432	1.418
$\Delta W_{it}/W_{it}$	0.135	0.361	-0.335	-0.055	0.075	0.229	0.878
Y_{it}/K_{it}	7.832	8.730	0.433	2.230	4.923	9.436	32.068
Y_{it}/W_{it}	3.095	1.793	0.698	1.863	2.722	3.953	7.474
$Y_{it}/(K_{it}+W_{it})$	1.640	0.937	0.296	0.972	1.533	2.146	3.780
$K_{it}/(K_{it}+W_{it})$	0.394	0.254	0.062	0.193	0.337	0.568	0.882
w_{it}^B	0.289	0.233	0.006	0.088	0.240	0.449	0.734
δ_{it+1}	0.191	0.133	0.047	0.104	0.150	0.236	0.523
r_{it+1}^B	0.099	0.064	0.017	0.061	0.086	0.118	0.284

Panel B: Pairwise correlations

	$\frac{I_{it+1}}{K_{it+1}}$	$\frac{\Delta W_{it}}{W_{it}}$	$\frac{\Delta W_{it+1}}{W_{it+1}}$	$\frac{Y_{it+1}}{K_{it+1}}$	$\frac{Y_{it+1}}{W_{it+1}}$	$\frac{Y_{it+1}}{(K_{it+1}+W_{it+1})}$	$\frac{K_{it+1}}{(K_{it+1}+W_{it+1})}$	w_{it}^B	δ_{it+1}	r_{it+1}^B
I_{it}/K_{it}	0.322	0.322	0.094	0.164	-0.073	0.050	-0.175	-0.182	0.296	0.096
I_{it+1}/K_{it+1}		0.240	0.322	0.366	0.020	0.212	-0.272	-0.272	0.536	0.257
$\Delta W_{it}/W_{it}$			0.027	0.069	-0.049	0.004	-0.072	-0.087	0.061	0.038
$\Delta W_{it+1}/W_{it+1}$				0.094	0.267	0.199	0.076	-0.130	0.079	0.166
Y_{it+1}/K_{it+1}					0.072	0.599	-0.645	-0.164	0.540	0.104
Y_{it+1}/W_{it+1}						0.571	0.466	0.182	-0.192	0.046
$Y_{it+1}/(K_{it+1}+W_{it+1})$							-0.349	-0.068	0.211	0.129
$K_{it+1}/(K_{it+1}+W_{it+1})$								0.319	-0.561	-0.079
w_{it}^B									-0.320	-0.106
δ_{it+1}										0.138

Table 2: **Descriptive properties of decile portfolios**

This table reports the monthly average excess returns of decile portfolios for 12 anomaly variables, including book-to-market equity ratio (Bm), momentum (R^{11}), asset growth (I/A), and return-on-equity(ROE). The t-statistics adjusted for heteroscedasticity and autocorrelations are reported in parenthesis. Decile portfolios are formed with NYSE breakpoints and value-weighted returns. L denotes the low decile, H the high decile, and H-L the high-minus-low decile. The sample period is from January 1967 to June 2017 for all anomaly variables.

	L	2	3	4	5	6	7	8	9	H	H-L
Bm	0.403	0.533	0.603	0.506	0.495	0.539	0.651	0.625	0.716	0.900	0.497
	(1.83)	(2.67)	(3.14)	(2.48)	(2.64)	(2.92)	(3.41)	(3.33)	(3.58)	(3.77)	(2.39)
R^{11}	-0.013	0.344	0.518	0.481	0.476	0.459	0.485	0.604	0.642	1.038	1.052
	(-0.04)	(1.25)	(2.40)	(2.39)	(2.54)	(2.38)	(2.84)	(3.30)	(3.11)	(3.91)	(3.51)
I/A	0.713	0.700	0.628	0.500	0.548	0.541	0.560	0.500	0.535	0.262	-0.451
	(2.97)	(3.64)	(3.74)	(2.92)	(3.19)	(2.96)	(3.06)	(2.52)	(2.27)	(1.02)	(-3.06)
ROE	0.037	0.266	0.425	0.418	0.553	0.480	0.529	0.493	0.548	0.676	0.638
	(0.12)	(1.08)	(1.95)	(2.30)	(3.04)	(2.47)	(2.76)	(2.55)	(2.88)	(3.28)	(3.10)

Table 3: **Description of Simulation Setup**

This table summarizes the simulation setup and the estimation results obtained from MCMC. Each row presents the statistics of one of 7 industries used in simulation, averaged over time. The columns γ and a are the true values of latent variables, averaged over time so we can better see how they vary across industries. The columns $std(\gamma)$ and $std(a)$ provide their time-series variability within each industry. The other columns show the estimated average of the estimates within each industry.

industry	γ	$std(\gamma)$	$\hat{\gamma}$	a	$std(a)$	$\hat{a}$
1	0.58	0.15	0.55	6.16	0.79	6.32
2	0.65	0.09	0.59	5.12	0.43	5.11
3	0.58	0.11	0.54	4.65	0.49	4.63
4	0.43	0.09	0.40	6.22	0.52	6.14
5	0.57	0.10	0.55	4.63	0.56	4.46
6	0.33	0.09	0.31	3.38	0.76	3.42
7	0.41	0.15	0.39	5.04	0.47	5.12

Table 4: **Aggregated estimation results from Simulation**

This table summarizes the estimation results in terms of asset returns and stock returns. Panel A is for simulation studies and Panel B is for empirical results. Three methods are compared. Columns "MAE" and "residual" provides two precision metrics, mean absolute error and residual. Row "MCMC" indicates estimation based on MCMC algorithm. Row "Initial" presents the estimation based on "GMM" method, detailed in Section 5. Row "flat" is obtained from GMM assuming time-invariant industry-invariant parameters γ and a. MCMC gives the smallest errors among the three, measured in either MAE or residuals.

Panel A: Summary statistics for simulation study

	source	MAE	residual
asset	MCMC	0.45	0.001
	Initial	0.46	0.03
	flat	0.49	0.02
stock	MCMC	0.64	0.005
	Initial	0.66	0.06
	flat	0.7	0.02

Panel B: Summary statistics for empirical analysis

	source	MAE	residual
asset	MCMC	0.29	0.01
	Initial	0.31	0.02
	flat	0.31	0.03
stock	MCMC	0.42	-0.003
	Initial	0.46	0.006
	flat	0.44	0.005

Table 5: **Parameter estimates**

This table reports the summary statistics of the posterior distributions of the technology parameter in the production function γ and investment adjustment cost parameter a. Columns under γ (a) and $[2.5\%, 97.5\%]_{\gamma(a)}$ report the time series averages of the posterior means of the estimated technology parameter in the production function (investment adjustment cost parameter) and its 95% prediction interval for the Fama-French 10 industries. Column under $[\gamma]_{TS}$ ($[a]_{TS}$) reports the time-series standard deviation of the posterior means for each industry.

Industry	γ	$[2.5\%, 97.5\%]_\gamma$	$[\gamma]_{TS}$	a	$[2.5\%, 97.5\%]_a$	$[a]_{TS}$
Consumer Nondurables	0.13	[0.12, 0.14]	0.09	0.43	[0.29, 0.63]	0.45
Consumer Durables	0.16	[0.14, 0.19]	0.18	1.14	[0.83, 1.50]	1.12
Manufacturing	0.16	[0.15, 0.18]	0.11	0.58	[0.51, 0.68]	0.99
Energy	0.20	[0.19, 0.22]	0.14	0.46	[0.41, 0.54]	0.53
Business Equipment	0.23	[0.21, 0.25]	0.19	1.83	[1.71, 2.14]	2.03
Telecom	0.28	[0.25, 0.32]	0.21	0.76	[0.66, 1.07]	0.69
Wholesale & Retail	0.08	[0.08, 0.09]	0.06	0.89	[0.78, 1.06]	0.99
Healthcare	0.19	[0.17, 0.21]	0.16	0.60	[0.46, 0.79]	0.68
Utilities	0.29	[0.26, 0.33]	0.17	0.27	[0.20, 0.39]	0.33
Others	0.17	[0.15, 0.19]	0.13	0.50	[0.47, 0.56]	0.52

Table 6: **Realized stock returns, fundamental returns, and alphas for portfolio deciles**

Panel A reports the annualized alpha of decile portfolios and the average absolute alpha over decile portfolios for each of the 4 anomaly variables. The alpha is the average difference between portfolio stock returns and constructed fundamental returns. Panel B reports the ratio of fundamental return to realized return for decile portfolios and high-minus-low decile portfolios. The sample period is from June 1967 to December 2016.

	L	2	3	4	5	6	7	8	9	H	H-L
Bm	-2.60***	-0.84	-1.49	-1.93	-1.48	0.22	1.04	2.40**	2.38**	3.67***	6.27***
	(-2.63)	(-0.96)	(-1.71)	(-1.83)	(-1.62)	(0.23)	(0.96)	(2.42)	(2.20)	(2.63)	(4.23)
R11	-2.09	0.67	1.36	0.80	-0.53	-1.25	-1.80**	-1.19	-2.26**	-0.22	1.87
	(-0.87)	(0.36)	(1.12)	(0.74)	(-0.51)	(-1.26)	(-2.11)	(-1.33)	(-1.99)	(-0.15)	(0.76)
I/A	-1.22	0.10	0.17	-0.23	-0.82	-0.35	-1.09	-0.76	-1.08	-4.37***	-3.16***
	(-0.94)	(0.11)	(0.20)	(-0.23)	(-0.99)	(-0.39)	(-1.33)	(-0.79)	(-0.97)	(-3.27)	(-2.90)
ROE	-4.77***	-1.73	-0.40	-0.97	0.74	-1.02	-0.97	-2.45***	-2.17**	-1.72	3.05
	(-2.63)	(-1.20)	(-0.34)	(-1.02)	(0.70)	(-0.89)	(-1.02)	(-2.67)	(-2.49)	(-1.59)	(1.80)

Appendix

A Bayesian MCMC

Markov Chain Monte Carlo (MCMC) is a generic method for approximating arbitrary distributions (in our case, the 1063-dimensional joint distribution). MCMC estimation is done by drawing sequentially by *samplers*, which draw from posterior distributions. Under certain circumstances, the sequence drawn by samplers "mixes" over the sampling space in a proper manner, such that the joint distribution can be approximated. There are two advantages of using MCMC algorithm: (1) MCMC sampler does not require a closed form of the joint distribution and (2) MCMC sampler needs only the posterior up to a constant proportion. In essence, even without a closed form of the posterior distribution, MCMC can still provide valid estimates.

In our model, there are in total 1063 parameters and latent variables to be estimated over the 53 years observation period across 10 industries. For each industry, the latent variables are highly dependent on its previous values. Besides, as firms switching between industries, different latent variables may depend on varying sets of observable fundamental variables. Due to the complicated nature and the high dimensionality of parameter space, moment based methods and likelihood maximization fail in practicability. Even if local maxima are achievable, there is no guarantee that such maxima are global. To address these issues, we take a Bayesian approach by MCMC.

In implementing MCMC, we derive posterior distributions for both parameters and latent variables. Metropolis-Hastings embedded Gibbs sampler is used for estimation. Whenever the closed form is not directly attainable, we use Metropolis-Hastings algorithm. Metropolis-Hastings algorithm was originally due to Metropolis (1953) and then generalized by Hastings (1970). It is a propose-reject method which first propose a *candidate* draw and then decide whether a *jump* is made from the current state to the proposed. Depending on the difference of proposal distributions, there are many variations under this generic heading. In our paper, the candidate is chosen in a manner that exploits as much as possible information from the posterior distributions. For a thorough discussion of Gibbs sampling and Metropolis-Hastings, see Robert and Casella (2004).

Using subscript i for firms; t for time points and k for industries, we derive the posterior distributions of latent variables $\beta_{\gamma tk}$ and β_{atk} and parameters σ_ε^2, σ_γ^2 and σ_a^2. For the time point $t = s-1$, $s \in [1, T]$ and the industry $k = j$, $j \in [1, D]$ let D_{sj} be the set of firms that belong to industry j at time s. Denote by E_{sj} the set of firms that belong to industry j at time $s-1$ and exist at time s.

In the following, we use w_{is}^m to indicate the relative market equity of firm i at time s. In order to address the heteroscedasticity problem, we first transform the return by scaling it with a constant quantity. When we are done with estimation, we convert it back to stock returns.

Since all weights are known variables, we can define a *weighted scaled asset return*

$$ret_{i,t+1} = \sqrt{w_{i,t+1}^m} \times \left(r_{i,t+1}^s + \frac{w_{it} r_{i,t+1}^B}{1 - w_{it}} \right) = \frac{\sqrt{w_{i,t+1}^m}}{1 - w_{it}} r_{i,t+1}^a.$$

Also define $\varepsilon_{i,t+1} = \sqrt{w^m_{i,t+1}}\eta_{i,t+1}$ so that $\varepsilon_{i,t+1} \overset{iid}{\sim} \mathcal{N}\Big(\cdot; 0, \sigma^2_\varepsilon\Big)$. The newly defined ret can be seen as a function of latent variables. Define

$$ret_{i,t+1} \equiv \Lambda_{i,t+1}\Big(\beta_{\gamma,t+1,k}, \beta_{a,t+1,k}, a_{it}, w^m_{i,t+1}\Big) + \varepsilon_{i,t+1}. \tag{14}$$

The joint distribution of parameters and latent variables can be written as

$$\begin{aligned}
&\prod_{t=0}^{T-1}\prod_{i=1}^{N}\mathcal{N}\Big(r_{i,t+1}; \Lambda_{i,t+1}(\beta_{\gamma,t+1,k}, \beta_{a,t+1,k}, a_{it}, w^m_{i,t+1}), \sigma^2_\varepsilon\Big)\\
&\cdot\prod_{t=0}^{T-1}\prod_{k=1}^{D}\mathcal{N}\Big(\beta_{\gamma,t+1,k}; \beta_{\gamma tk}, \sigma^2_\gamma\Big)\\
&\cdot\prod_{t=0}^{T-1}\prod_{k=1}^{D}\mathcal{N}\Big(\beta_{a,t+1,k}; \beta_{atk}, \sigma^2_a\Big)\\
&\cdot\mathcal{IG}\Big(\sigma^2_\varepsilon; \gamma_\varepsilon, \beta_\varepsilon\Big)\cdot\mathcal{IG}\Big(\sigma^2_\gamma; \gamma_\gamma, \beta_\gamma\Big)\cdot\mathcal{IG}\Big(\sigma^2_a; \gamma_a, \beta_a\Big).
\end{aligned} \tag{15}$$

Note that in the joint distribution, no prior distributions for latent variables are assigned because we treat the initial latent variables $\beta_{\gamma 0k}$ and β_{a0k} as unknown constants. The driver for latent variables' evolvement is fully explained by the variances of $\nu_{\gamma,t+1,k}$ and $\nu_{a,t+1,k}$ so we do not assign prior to the other latent variables.

We assign inverse gamma distributions as priors for the parameters σ^2_ε, σ^2_γ and σ^2_a because they are conjugate.

For each latent variable and each parameter, we derive its posterior distribution as follows:

For the latent variables $\beta_{\gamma sj}$, the posterior is normal:

$$\begin{aligned}
&p\left(\beta_{\gamma sj}\middle|\ \{\beta_{\gamma tk}\}, \{\beta_{\gamma sk}\}, \{\beta_{\gamma sk}\}, \{\beta_{\gamma tk}\}, \{\beta_{a0k}\}, \{\beta_{atk}\}\,\sigma^2_\varepsilon, \sigma^2_\gamma, \sigma^2_a\right)\\
&= \exp\left\{-\frac{1}{2}\underbrace{\left(\frac{1}{\sigma^2_\varepsilon}\sum_{i\in D_{sj}} A^2_{is} + \frac{1+\mathbb{1}_{s\notin\{1,T\}}}{\sigma^2_\gamma}\right)}_{u_1}\beta^2_{\gamma sj} + \underbrace{\left(\frac{1}{\sigma^2_\varepsilon}\sum_{i\in D_{sj}}\varphi_{is}A_{is} + \frac{1}{\sigma^2_\gamma}(\beta_{\gamma,s-1,j}\mathbb{1}_{s\neq 1} + \beta_{\gamma,s+1,j}\mathbb{1}_{s\neq T})\right)}_{v_1}\beta_{\gamma sj}\right\}\\
&\sim\mathcal{N}\left(\beta_{\gamma sj}; \frac{v_1}{u_1}, \frac{1}{u_1}\right).
\end{aligned}$$

where $\varphi_{is} := ret_{is} - \sqrt{w^m_{is}} \times \dfrac{\tau_{is}\delta_{is} + \frac{W_{is}}{K_{is}} + (1-\delta_{is})}{(1-w_{i,s-1})\left[1 + (1-\tau_{i,s-1})a_{i,s-1}\frac{I_{i,s-1}}{K_{i,s-1}} + \frac{W_{is}}{K_{is}}\right]}$

$$-\sqrt{w_{is}^m} \times \frac{\frac{1}{2}(1-\tau_{is})\left(\frac{I_{is}}{K_{is}}\right)^2 + (1-\delta_{is})(1-\tau_{is})\frac{I_{is}}{K_{is}}}{(1-w_{i,s-1})\left[1+(1-\tau_{i,s-1})a_{i,s-1}\frac{I_{i,s-1}}{K_{i,s-1}} + \frac{W_{is}}{K_{is}}\right]}\beta_{asj}$$

and $A_{is} := \sqrt{w_{is}^m} \times \dfrac{(1-\tau_{is})\frac{Y_{is}}{K_{is}}}{(1-w_{i,s-1})\left[1+(1-\tau_{i,s-1})a_{i,s-1}\frac{I_{i,s-1}}{K_{i,s-1}} + \frac{W_{is}}{K_{is}}\right]}$.

For β_{asj}, there is no clear closed form of the posterior distribution so we implement Metropolis-Hastings. We first consider the posterior of β_{asj} although it is not clear what distribution it follows:

$$\begin{aligned} &p\left(\beta_{asj} \middle| \left\{\beta_{\gamma tk}\right\}, \left\{\beta_{\gamma Tk}\right\}, \left\{\beta_{atk}\right\}, \left\{\beta_{ask}\right\}, \left\{\beta_{ask}\right\}, \left\{\beta_{atk}\right\}, \sigma_\varepsilon^2, \sigma_\gamma^2, \sigma_a^2\right) \\ &= \prod_{t=0}^{T-1}\prod_{i=1}^{N} \mathcal{N}\left(ret_{i,t+1}; \Lambda_{i,t+1}, \sigma_\varepsilon^2\right) \prod_{t=0}^{T-1} \mathcal{N}\left(\beta_{a,t+1,j}; \beta_{atj}, \sigma_a^2\right). \end{aligned} \tag{17}$$

We propose from

$$\begin{aligned} &\prod_{i\in D_{sj}} \mathcal{N}\left(ret_{is}; \Lambda_{is}, \sigma_\varepsilon^2\right)\mathcal{N}\left(\beta_{asj}; \beta_{a,s-1,j}, \sigma_a^2\right)\mathcal{N}\left(\beta_{asj}; \beta_{a,s+1,j}, \sigma_a^2\right) \\ &= \exp\left\{-\frac{1}{2}\underbrace{\left(\frac{1}{\sigma_\varepsilon^2}\sum_{i\in D_{sj}} B_{is}^2 + \frac{1+\mathbb{1}_{s\notin\{1,T\}}}{\sigma_a^2}\right)}_{u_2}\beta_{asj}^2 + \frac{2}{2}\underbrace{\left(\frac{1}{\sigma_\varepsilon^2}\sum_{i\in D_{sj}}\psi_{is}B_{is} + \frac{1}{\sigma_a^2}\left(\beta_{a,s-1,j}\mathbb{1}_{s\neq 1} + \beta_{a,s+1,j}\mathbb{1}_{s\neq T}\right)\right)\beta_{asj}}_{v_2}\right\} \\ &\sim \mathcal{N}\left(\beta_{asj}; \frac{v_2}{u_2}, \frac{1}{u_2}\right) \end{aligned}$$

where $\psi_{is} := ret_{is} - \sqrt{w_{is}^m} \times \dfrac{\tau_{is}\delta_{is} + \frac{W_{is}}{K_{is}} + (1-\delta_{is})}{(1-w_{i,s-1})\left[1+(1-\tau_{i,s-1})a_{i,s-1}\frac{I_{i,s-1}}{K_{i,s-1}} + \frac{W_{is}}{K_{is}}\right]}$

$$-\sqrt{w_{is}^m} \times \frac{(1-\tau_{is})\frac{Y_{is}}{K_{is}}}{(1-w_{i,s-1})\left[1+(1-\tau_{i,s-1})a_{i,s-1}\frac{I_{i,s-1}}{K_{i,s-1}} + \frac{W_{is}}{K_{is}}\right]}\beta_{\gamma sj}$$

and $B_{is} := \sqrt{w_{is}^m} \times \dfrac{\frac{1}{2}(1-\tau_{is})\left(\frac{I_{is}}{K_{is}}\right)^2 + (1-\delta_{is})(1-\tau_{is})\frac{I_{is}}{K_{is}}}{(1-w_{i,s-1})\left[1+(1-\tau_{i,s-1})a_{i,s-1}\frac{I_{i,s-1}}{K_{i,s-1}} + \frac{W_{is}}{K_{is}}\right]}$.

To decide whether to accept the candidate, let

$$\pi(x) = \prod_{i\in D_{sj}} \mathcal{N}\Big(ret_{is}; \Lambda_{is}\big(\beta_{\gamma sj}, x, a_{i,s-1}, w^m_{is}\big), \sigma^2_\varepsilon\Big)$$
$$\cdot \prod_{i\in E_{s+1,j}} \mathcal{N}\Big(r_{i,s+1}; \Lambda_{i,s+1}\big(\gamma_{i,s+1}, a_{i,s+1}, x, w^m_{i,s+1}\big), \sigma^2_\varepsilon\Big)$$
$$\cdot\mathcal{N}\Big(x; \beta_{a,s+1,j}, \sigma^2_a\Big) \cdot \mathcal{N}\Big(x; \beta_{a,s-1,j}, \sigma^2_a\Big).$$

The acceptance rate α is then

$$\alpha = \frac{\pi(\beta^{prop}_{asj})}{\pi(\beta_{asj})} \cdot \frac{\mathcal{N}\big(\beta_{asj}; \frac{v_2}{u_2}, \frac{1}{u_2}\big)}{\mathcal{N}\big(\beta^{prop}_{asj}; \frac{v_2}{u_2}, \frac{1}{u_2}\big)} = \frac{\prod_{i\in E_{s+1,j}} \mathcal{N}\Big(r_{i,s+1}; \Lambda_{i,s+1}\big(\gamma_{i,s+1}, a_{i,s+1}, \beta^{prop}_{asj}, w^m_{i,s+1}\big), \sigma^2_\varepsilon\Big)}{\prod_{i\in E_{s+1,j}} \mathcal{N}\Big(r_{i,s+1}; \Lambda_{i,s+1}\big(\gamma_{i,s+1}, a_{i,s+1}, \beta^{(g-1)}_{asj}, w^m_{i,s+1}\big), \sigma^2_\varepsilon\Big)}.$$

The posterior distributions for parameters are:

$$p\left(\sigma^2_\varepsilon \middle| \left\{\beta_{\gamma tk}\right\}, \left\{\beta_{atk}\right\}, \sigma^2_\gamma, \sigma^2_a\right) \sim \mathcal{IG}\left(k_\varepsilon + \frac{NT}{2}, \theta_\varepsilon + \frac{1}{2}\sum_{t=0}^{T-1}\sum_{i=1}^{N}\left(ret_{i,t+1} - \Lambda_{i,t+1}\right)^2\right). \tag{19}$$

$$p\left(\sigma^2_\gamma \middle| \left\{\beta_{\gamma tk}\right\}, \left\{\beta_{atk}\right\}, \sigma^2_\varepsilon, \sigma^2_a\right) \sim \mathcal{IG}\left(k_\gamma + \frac{DT}{2}, \theta_\gamma + \frac{1}{2}\sum_{t=0}^{T-1}\sum_{k=1}^{D}\left(\beta_{\gamma,t+1,k} - \beta_{\gamma tk}\right)^2\right). \tag{20}$$

$$p\left(\sigma^2_a \middle| \left\{\beta_{\gamma tk}\right\}, \left\{\beta_{atk}\right\}, \sigma^2_\varepsilon, \sigma^2_\gamma\right) \sim \mathcal{IG}\left(k_a + \frac{DT}{2}, \theta_a + \frac{1}{2}\sum_{t=0}^{T-1}\sum_{k=1}^{D}\left(\beta_{a,t+1,k} - \beta_{atk}\right)^2\right). \tag{21}$$

where k_ε, θ_ε, k_γ, θ_γ, k_a and θ_a are prior parameters, which are defined in detail in appendix. where k_γ, θ_γ, k_a and θ_a are prior parameters, which are defined in detail in appendix.

In each MCMC iteration, systematic scan is used. In each iteration, we sample by a pre-specified order the parameters/latent variables from the above posterior distribution conditional on the most updated information. After all the parameters and latent variables are updated, a new iteration is started. We run 20000 iterations in total and find the estimate and upper/lower 2.5% quantiles based on the empirical distribution of the 20000 iterations.

B Algorithms

Algorithm 1: MH embedded Gibbs Sampling

Procedure
  MH embedded Gibbs Sampling
  Initial values: Latent variables and parameters: $\underline{\boldsymbol{\beta}}^{(0)}$, $\sigma_{\varepsilon}^{2(0)}$, $\sigma_{\gamma}^{2(0)}$ and $\sigma_{a}^{2(0)}$
  for $\underline{g = 1 \text{ to } G}$ **do**
    for $\underline{j = 1 \text{ to } D}$ **do**
      Sample $\beta_{a0j}^{(g)}$ with MH algorithm and update
    end
    for $\underline{s = 1 \text{ to } T-1}$ **do**
      for $\underline{j = 1 \text{ to } D}$ **do**
        Sample $\beta_{\gamma sj}^{(g)}$ with Gibbs sampling and update
      end
    end
    for $\underline{s = 1 \text{ to } T-1}$ **do**
      for $\underline{j = 1 \text{ to } D}$ **do**
        Sample $\beta_{asj}^{(g)}$ with MH algorithm and update
      end
    end
    for $\underline{j = 1 \text{ to } D}$ **do**
      Sample $\beta_{\gamma Tj}^{(g)}$ with Gibbs sampling and update
    end
    for $\underline{j = 1 \text{ to } D}$ **do**
      Sample $\beta_{aTj}^{(g)}$ with MH algorithm and update
    end
    Sample $\sigma_{\varepsilon}^{2(g)}$ with Gibbs sampling and update
    Sample $\sigma_{\alpha}^{2(g)}$ with Gibbs sampling and update
    Sample $\sigma_{a}^{2(g)}$ with Gibbs sampling and update
  end
end

Algorithm 2: Metropolis-Hastings algorithm (β_{asj} in our model) as an example

```
Procedure
    Metropolis-Hastings algorithm
    for g = 1 to G do
        Other parts of MH embedded Gibbs sampling ···
        for s = 1 to T − 1 do
            for j = 1 to D do
                Choose proposal:
                    N(·; v_2/u_2, 1/u_2)
                Calculate acceptance rate:
                    α = min{ π_2s(β_asj^prop)/π_2s(β_asj) · N(β_asj; v_2/u_2, 1/u_2) / N(β_asj^prop; v_2/u_2, 1/u_2) }
                if min{1, α} > u ~ U(0,1) then
                    cand = β_asj^(g−1)
                end
                β_asj^(g) = cand
                Update β_a^(g−1)'s
            end
        end
        Other parts of MH embedded Gibbs sampling ···
    end
end
```

Where the proposal and acceptance rate are:

$$\mathcal{N}\Big(\cdot; \frac{v_2}{u_2}, \frac{1}{u_2}\Big)$$

$$\alpha = \min\left\{ \frac{\pi_{2s}(\beta_{asj}^{prop})}{\pi_{2s}(\beta_{asj})} \cdot \frac{\mathcal{N}\big(\beta_{asj}; \frac{v_2}{u_2}, \frac{1}{u_2}\big)}{\mathcal{N}\big(\beta_{asj}^{prop}; \frac{v_2}{u_2}, \frac{1}{u_2}\big)} \right\}$$

Algorithm 3: Choices of Initial Values

```
Procedure
    Choices of Initial Values
    Assume β_a0k = β_a1k for k = 1, 2, ..., D
    for j = 1 to D do
        (β_γ1k^(0), β_a1k^(0)) = arg min(Σ_{s=min{t}} Λ_i,s(β_γ1k, β_a1k, β_a1k, w_i,s^m))^2
    end
    for s = 2 to T do
        (β_γsk^(0), β_ask^(0)) = arg min(Σ_s Λ_i,s(β_γ1k, β_a1k, β_a,s−1,k, w_i,s^m))^2
    end
end
```

$$(\beta^{(0)}_{\gamma 1k}, \beta^{(0)}_{a1k}) = \arg\min\big(\textstyle\sum_{s=\min\{t\}} \Lambda_{i,s}(\beta_{\gamma 1k}, \beta_{a1k}, \beta_{a1k}, w^m_{i,s})\big)^2$$

$$(\beta^{(0)}_{\gamma sk}, \beta^{(0)}_{ask}) = \arg\min\big(\textstyle\sum_{s} \Lambda_{i,s}(\beta_{\gamma 1k}, \beta_{a1k}, \beta_{a,s-1,k}, w^m_{i,s})\big)^2$$

C Definition of Sorting Variables

Bm Book-to-market equity ratio, defined as the book value of equity for fiscal year end in the previous calender year $t-1$ divided by the market value of equity at the end of December of previous calender year $t-1$. Following Davis, Fama, and French (2000), we measure book equity as stockholders' book equity, plus balance sheet deferred taxes and investment tax credit (item TXDITC or the sum of item TXDB and item ITCB) if available, minus the book value of preferred stock. Stockholders' equity is the value reported by Compustat (item SEQ) if it is available. If not, we measure stockholdersÂ¡Â¯ equity as the book value of common equity (item CEQ) plus the par value of preferred stock (item PSTK), or the book value of assets (item AT) minus total liabilities (item LT). Depending on availability, we use redemption (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for the book value of preferred stock.

$\boldsymbol{R^{11}}$ Prior 11-month returns from month t-12 to t-2.

I/A Following Cooper, Gulen, and Schill (2008), we measure I/A as change in total assets (Compustat annual item AT) scaled by lagged total assets. At the end of June of each year t, we use NYSE breakpoints to split stocks into deciles based on I/A for the fiscal year ending in calendar year t-1 and calculate monthly value-weighted decile returns from July of year t to June of t+1.

ROE *ROE* is income before extraordinary items (Compustat quarterly item IBQ) divided by 1-quarter-lagged book equity. From 1972 onward, quarterly book equity is shareholders' equity, plus balance sheet deferred taxes and investment tax credit (item TXDITCQ) if available, minus the book value of preferred stock (item PSTKQ). Depending on availability, we use stockholders' equity (item SEQQ), or common equity (item CEQQ) plus the book value of preferred stock (item PSTKQ), or total assets (item ATQ) minus total liabilities (item LTQ) in that order as shareholders' equity. Prior to 1972, we expand the sample coverage by using book equity from Compustat annual files and imputing quarterly book equity with clean surplus accounting (Hou, Xue, and Zhang 2019).

At the beginning of each month t, we sort stocks into deciles on their most recent past *ROE*. Before 1972, we use the most recent *ROE* computed with quarterly earnings from the fiscal quarter ending at least four months ago. From 1972 onward, we use Roe computed with quarterly earnings from the most recent quarterly earnings announcement date (item RDQ). For a firm to enter the portfolio formation, we require the end of the fiscal quarter corresponding to its most recent Roe to be within six months prior to the portfolio formation and its earnings announcement date to be after the corresponding fiscal quarter end. Monthly decile returns are calculated for the current month t, and the deciles are rebalanced at the beginning of month t + 1.