Data-generating process uncertainty: What difference does it make in portfolio decisions? [☆]

Jun Tu[a], Guofu Zhou[a,b,]*

[a] Olin School of Business, Washington University, St. Louis, MO 63130, USA
[b] China Center for Financial Research, Tsinghua University, Beijing, China

Abstract

As the usual normality assumption is firmly rejected by the data, investors encounter a data-generating process (DGP) uncertainty in making investment decisions. In this paper, we propose a novel way to incorporate uncertainty about the DGP into portfolio analysis. We find that accounting for fat tails leads to nontrivial changes in both parameter estimates and optimal portfolio weights, but the certainty–equivalent losses associated with ignoring fat tails are small. This suggests that the normality assumption works well in evaluating portfolio performance for a mean-variance investor.
© 2003 Elsevier B.V. All rights reserved.

1. Introduction

Recently a great number of Bayesian studies have emerged on a range of finance problems, mainly because the conceptual set-up of Bayesian analysis is well suited for analyzing financial decisions, and the associated integral valuations have become numerically tractable with high speed computers. For example, Pástor and Stambaugh (1999, 2000) investigate how the mispricing uncertainty associated with an asset pricing model can fundamentally change the way we estimate the cost of capital and the way we make portfolio decisions. Wang (2001) provides a Bayesian shrinkage solution to an asset allocation problem, while Barberis (2000) extends the static model of Kandel and Stambaugh (1996) and shows how evidence of predictability may improve allocation decisions. Brennan and Xia (2001), Avramov (2002), and Cremers (2002) each offers interesting Bayesian assessments on predictability in light of the model uncertainty, and both Cremers (2001) and Avramov and Chao (2002) derive insightful Bayesian tests of the capital asset pricing model.

However, none of the recent papers has examined the data-generating process (DGP) uncertainty that occurs when an investor does not know whether the data is generated from a population model with a normal distribution. In other words, existing Bayesian studies generally assume a normal distribution for the DGP, and an investor's belief in other likely DGPs is not incorporated into the analysis.[1] Does an informed investor have such a belief? The answer seems to be in the affirmative because the departure of monthly security returns from normality is well known, as documented by Fama (1965), Affleck-Graves and McDonald (1989), and Richardson and Smith (1993), among others. More precisely, consider the assets used by Pástor and Stambaugh (2000). As shown in Section 2, the multivariate kurtosis of the data is as large as 220.756, implying that the *p*-value is less than 0.00001 for the validity of the normality hypothesis! On the other hand, a multivariate *t* distribution with eight degrees of freedom has a *p*-value of 71.06%. Statistically, a multivariate *t* distribution can be viewed as a mixture of an infinite number of multivariate normal distributions; different mixtures may be described by different degrees of freedom. In particular, the multivariate normal distribution is a trivial mix that has an infinite degrees of freedom. From the completely inadequate normality assumption to that of a reasonable multivariate *t*, there must lie numerous other mixtures of the multivariate normal. Hence, an investor who is unsure about which mixture represents the true DGP clearly faces the problem of DGP uncertainty.

The objective of this paper is to draw the attention of empirical finance research to the importance of DGP uncertainty and to show how this uncertainty can be incorporated into financial decision-making. In particular, we focus on the portfolio

---

[1] The classical approaches, like Hansen's (1982) GMM which was well applied by MacKinlay and Richardson (1991), allow much more general distributions than the normal. But the focus of these approaches is on hypothesis testing rather than decision-making. In addition, the analysis usually relies on inefficient parameter estimates, as explained later.

choice problem of an expected utility maximizing investor, the same problem analyzed by Pástor and Stambaugh (2000). Although it might be interesting to re-examine many other influential studies that rely on the unequivocally rejected normality assumption as well, we choose the Pástor and Stambaugh framework for three reasons. First, the portfolio choice problem is one of the most commonly asked questions in investment practice. Second, it is this problem for which asset pricing theory has the most assertions and suggestions. Third, the Pástor and Stambaugh set-up is well established, incorporating not only the uncertainty in parameter estimates, but also the uncertainty in model mispricing. Based on this framework, we add DGP uncertainty so that an investor can make a more realistic decision in light of all three uncertainties: parameter, asset pricing model, and DGP. Hence, in a certain sense, this paper complements the earlier work of Pástor and Stambaugh (2000).

To model the DGP uncertainty, we assume that the investor has knowledge of a set of probability distributions that are possible candidates for the true data-generating process. The set for our applications consists of 31 multivariate t distributions with the degrees of freedom taking the values from $2.1, 3, 4, 5, \ldots, 31$, to 32. The smaller the degrees of freedom, the greater the deviation from the normal distribution. Theoretically, the normal distribution is the limiting case of the t distribution as the degrees of freedom approach infinity and is virtually the same as a t with 32 degrees of freedom. We can thus use the latter in place of the normal to simplify the notation. To reflect the possibility that the true DGP could be far different from the normal, we allow the degrees of freedom to be as small as 2.1, but exclude 2 because it implies an infinite variance. A diffuse or uninformative prior on the true distribution will assign an equal probability mass to the 31 values of v, the unknown degrees of freedom, whereas an informative prior can assign different probabilities across v. Given a prior belief, we let the data tell us which one of the distributions is the best approximation of the true distribution through Bayesian updating. If the posterior mass of v concentrates on the smaller values, then the data tells us that the normality assumption is a poor approximation. Following Pástor and Stambaugh (2000), we also consider priors on the mispricing uncertainty of three asset pricing models: the CAPM, the Fama and French (1993) three-factor model, and the Daniel and Titman (1997) characteristic-based model. Interestingly, regardless of the level of mispricing uncertainty, the Bayesian updating always transforms the diffuse prior into a posterior belief with almost all of the probability mass centered around $v = 8$. This is true not only for the entire data set, from July 1963 through December 1997 (the same data as in Pástor and Stambaugh, 2000), but also for subperiod data. Hence, the data seems informative to indicate strong evidence against using the usual normal DGP, favoring instead the use of a t distribution with about eight degrees of freedom. Indeed, the same conclusion holds even if one has an informative prior for which the normality assumption is true with 50%, 75%, or 99% probability, respectively.

Once the DGP uncertainty is introduced, three other major findings follow. First, the estimates of expected asset returns and standard deviations can be quite different from those obtained under the normality assumption. Generally speaking, the more

skewed or leptokurtic the asset returns, the greater the difference. For example, the estimated expected excess market return across various prior beliefs is about 6.36% per year under normality, but is substantially lower at a value of around 5.28% per year under DGP uncertainty. This difference might well be important in the computation of the cost of capital. The intuition for this result is that normality weights all tail observations equally in the estimation of the mean, while the *t* distribution gives the tails fewer weights than the center as outliers are more likely to occur under the *t*. Historically, the market is described by a distribution with thick tails that are fatter on the positive side, and thus the *t* distribution must shift the estimated population mean leftward to better fit the data. Interestingly, the standard deviation of the market excess return is also lowered, from about 15.00% per year under normality to around 14.58% per year under DGP uncertainty. The intuition here is that a normal distribution usually requires a larger variance than a *t* distribution to fit the data, due to the greater influence the normal attributes to outliers.

Our second finding is that the optimal portfolio weights under DGP uncertainty can be substantially different from those obtained under the normality assumption. For example, for an investor who believes that the Fama-French model is a good approximation of the underlying asset pricing model, some of her non-benchmark portfolio weights decrease or increase by more than 50% if she has a relative risk coefficient of 2.83 and faces no margin requirements. More generally, across all the models and priors, investors are much more conservative about taking market risk under DGP uncertainty than under the normal assumption.

Due to correlations among the payoffs of risky positions, the performance of two portfolio selections can be similar even though they are quite different in position-by-position allocations. One way to gauge the differences among alternative portfolios is through the certainty–equivalent loss suggested by Pástor and Stambaugh (2000). The third finding of our paper is that, in the presence of DGP uncertainty and in terms of the certainty–equivalent return, the loss for an investor who is forced to hold the portfolio that is optimal under the misspecified normality assumption is actually small. For instance, under a 50% margin requirement, the maximum loss across an array of model mispricing priors of the three asset pricing models is only 0.54% per year, implying that Pástor and Stambaugh's (2000) Bayesian certainty–equivalent analysis is surprisingly robust to the underlying distributional assumptions. That is, despite the fact that the normality assumption is unequivocally rejected by the data, the expected utility achieved based on a much more reasonable DGP is not significantly different from that obtained under normality. To assess the sensitivity of this result to the use of test assets, we also conduct the same analysis for Fama and French's (1993) 25 assets and 20 industry portfolios, and reach the same conclusion. However, this does not suggest that the DGP uncertainty will not affect other aspects of the model. For example, as pointed out earlier, the impacts on risk and return estimations as well as on optimal portfolio weights can be substantial. In addition, whether or not stock return predictability is robust to the normality assumption is an open question and a topic for future research.

The remainder of the paper is organized as follows. Section 2 provides the empirical evidence which suggests the necessity of modelling DGP uncertainty, and then discusses a framework for making Bayesian portfolio decisions under such uncertainty. Section 3 applies the proposed approach to the data and reports the empirical results. Section 4 concludes.

2. Investing under DGP uncertainty

In this section, we provide first some empirical understanding of the data by showing that the usual normality assumption is strongly rejected while a suitable multivariate t is not. Next, we consider investment decisions under a given data-generating process (DGP), and then we model the DGP uncertainty by using a class of multivariate t distributions with varying degrees of freedom and study the associated investment decisions. Finally, we outline an economic measure for evaluating the performance of investment decisions.

2.1. Data properties

Following Pástor and Stambaugh (2000), we consider an investment universe that contains $n = 12$ risky positions. Three of the risky positions are the Fama and French (1993) benchmark positions, SMB, HML, and MKT. The other nine non-benchmark positions are formed as spreads between portfolios selected from a larger universe of 27 equity portfolios created by a three-way sorting on size, book-to-market, and HML beta. More precisely, at the end of June of year t, all NYSE, AMEX, and NASDAQ stocks listed on both the CRSP and Compustat files are sorted based on market capitalization ('size') and assigned to one of three categories, where each category has an equal number of stocks. Independently, the same stocks are then sorted based on the ratio of book value of equity to market capitalization ('book-to-market') and, again, assigned to one of three categories, where each category has the same number of stocks. The intersection of these two sets of categories produces nine groups of stocks. The stocks within each group are then sorted by HML beta and assigned to one of three subgroups containing an equal number of stocks. This three-way grouping creates 27 value-weighted portfolios that are identified by a combination of three letters which designate increasing values of size (S, M, B), book-to-market (L, M, H), and HML beta (l, m, h). The portfolios on the long side and the short side of a spread, one of the nine non-benchmark positions, are different in HML (high versus low) beta but are matched in size and book-to-market. Pástor and Stambaugh (2000) describe further details on the construction of the data.[2] The data are monthly returns from July 1963 through December 1997.

Means, standard deviations, and autocorrelations of the data are presented in Table 1. The means range from −0.42% per month in position ML(l-h) to 0.53% per

---

[2] We are grateful to Ľuboš Pástor for providing us the data as used in Pástor and Stambaugh (2000).

Table 1
Summary statistics

The table reports the sample means (in percentage points), standard deviations (in percentage points), and autocorrelations of the nine spread positions created by Pástor and Stambaugh (2000) and the three Fama-French benchmark portfolios based on monthly returns from July 1963 through December 1997. The nine positions are spreads between value-weighted portfolios constructed by a three-way sorting and are identified by a combination of three letters which designate increasing values of size (S,M,B), book-to-market (L,M,H), and HML beta (l, m, h).

	Mean	Std	Autocorrelation						
			ρ_1	ρ_2	ρ_3	ρ_4	ρ_{12}	ρ_{24}	ρ_{36}
SL(l-h)	−0.00	4.11	−0.14	0.07	−0.11	0.15	−0.06	−0.05	−0.00
SM(l-h)	−0.05	2.45	−0.06	0.00	−0.03	0.03	−0.00	0.01	−0.04
SH(l-h)	−0.02	2.21	0.10	0.08	−0.03	−0.03	0.04	−0.02	0.05
ML(l-h)	−0.42	3.00	−0.00	−0.01	0.01	0.03	−0.02	−0.02	0.03
MM(l-h)	−0.05	2.51	0.08	0.01	−0.03	−0.04	0.06	−0.03	0.07
MH(l-h)	−0.05	2.90	0.08	0.05	−0.04	0.07	−0.03	−0.02	−0.04
BL(l-h)	−0.14	3.03	0.12	0.01	0.03	−0.04	0.02	−0.01	0.02
BM(l-h)	−0.19	3.39	0.18	−0.11	−0.00	−0.02	0.00	0.02	0.11
BH(l-h)	−0.12	3.28	0.08	−0.02	−0.05	0.03	−0.05	−0.06	−0.04
SMB	0.22	2.88	0.17	0.03	−0.04	0.02	0.20	0.05	0.07
HML	0.43	2.53	0.17	0.03	−0.03	−0.02	0.08	0.06	0.03
MKT	0.53	4.31	0.03	−0.02	−0.01	−0.02	0.02	−0.01	−0.02

month in the market portfolio. The lowest standard deviation is found in SH(l-h) and the highest in the market portfolio. Although the first-order autocorrelations are not zero, they are small, with a maximum value of only 0.18. There does not exist any evidence for seasonality as most of the high-order autocorrelations are very close to zero. Most studies, such as Fama and French (1993) and many related papers, assume that the asset returns are jointly normal and independently and identically distributed (i.i.d.) over time. Although Fama and French (1993) use also the bootstrap method, their parameter estimates are obtained under normality. Consistent with these studies, Pástor and Stambaugh (2000) assumes i.i.d. normal. In this paper, we allow non-normal i.i.d. distributions. While the i.i.d. assumption rules out predictability and time-varying parameters, we maintain it for three reasons. First, though there are numerous studies claiming some degrees of predictability, Lamoureux and Zhou (1996) and Bossaerts and Hillion (1999) find that the evidence for predictability is quite weak, and Han (2002) questions its economical significance. On the other hand, even if one models time-varying expected returns, Ghysels (1998) shows that model fitting becomes much worse than a constant-parameter model. Second, imposing a dynamic structure on the DGP as well as assuming a t distribution are likely to make the problem too complex to solve, and this line of research goes beyond the scope of the current paper. Third, the impact of DGP uncertainty might be better isolated and understood when all aspects of the models are cast in the earlier framework, with the one exception of replacing the commonly used normal distribution by a set of t distributions.

Now we turn our attention to the skewness and kurtosis of the data to understand why we must replace the normal by a t distribution. Following Mardia (1970) and many multivariate statistics books, we define multivariate skewness and kurtosis by

$$D_1 = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} r_{ts}^3 \quad \text{and} \quad D_2 = \frac{1}{T} \sum_{t=1}^{T} r_{tt}^2, \tag{1}$$

respectively, where $r_{ts} = (X_t - \bar{X})' S^{-1}(X_s - \bar{X})$, X_t is an n-vector of the returns in period t, $\bar{X}$ is the sample mean, and S is the sample covariance matrix. There are two important properties of both D_1 and D_2. First, as the sample size increases without bound, they converge to their population counterparts

$$\delta_1 = E([(X - \theta)' \Sigma^{-1}(Y - \theta)]^3) \quad \text{and} \quad \delta_2 = E([(X - \theta)' \Sigma^{-1}(X - \theta)]^2), \tag{2}$$

respectively, where Y is independent of X but has the same distribution. Under the normality assumption, δ_1 is equal to zero and δ_2 is equal to $n(n + 2)$. The second property of D_1 and D_2 is that they are invariant to any linear transformations of the data. In other words, any nonsingular repackaging of the assets will be unable to alter the skewness and kurtosis. Due to this invariance property, one can assume without any loss of generality that the true distribution has zero mean and unit covariance matrix for the purpose of computing the exact distributions of D_1 and D_2. As demonstrated by Zhou (1993), the exact distribution can be computed up to any desired level of accuracy by simulating samples from the standardized hypothetical true distribution of the data without specifying any unknown parameters.

Table 2 provides both the univariate and multivariate sample skewness and kurtosis of the data. The univariate skewness of the assets varies from a low of only 0.001 (of the sixth MH(l-h) asset) to a high of 0.258 (of the eighth BM(l-h) asset). Under the null hypothesis that the assets individually follow a normal distribution, the associated p-values based on 100,000 draws of Monte Carlo integration are provided in the third column of the table. We observe that univariate normality is rejected for five of the assets at the usual 5% significant level. However, it is multivariate normality that is assumed by most existing studies. The third column of the lower panel of Table 2 provides a p-value of 0.00 (less than 0.00001) for the multivariate skewness test, implying an overwhelming rejection of the multivariate normality hypothesis. In comparison with the univariate rejection, the multivariate one is much stronger which may not be surprising as multivariate normality implies univariate normality but the converse does not hold. Results based on the kurtosis test, which are reported in the ninth column, tell the same story. Except for three assets, univariate normality is rejected by the univariate kurtosis test at the usual 5% level. Closer examination shows that the p-values of the univariate kurtosis test are substantially smaller than those of the univariate skewness test, resulting in four more rejections of the normality hypothesis. This indicates that the violation of the normality assumption seems to come more often from the kurtosis analysis. Similar to the results based on skewness tests, the multivariate kurtosis test strongly rejects the joint normality of the asset returns.

Now consider the hypothesis that the data follows a t distribution with degrees of freedom $v = 4, 8, 12$, and 16, respectively. When $v = 8$, the p-values of the skewness

Table 2

Normality test of the Pástor and Stambaugh (2000) portfolios

The table reports the sample univariate and multivariate skewness and kurtosis measures of the nine spread positions created by Pástor and Stambaugh (2000) and the three Fama-French benchmark portfolios based on monthly returns from July 1963 through December 1997. In addition, it also reports the *p*-values (in percentage points) of the skewness and kurtosis tests if the data is assumed to be drawn from a univariate or multivariate normal distribution, or a univariate or multivariate *t* distribution with degrees of freedom equal to 16, 12, 8, and 4, respectively.

	Skewness						Kurtosis					
	Statistic	Normal	Student-*t* with df				Statistic	Normal	Student-*t* with df			
			16	12	8	4			16	12	8	4
Univariate												
SL(l-h)	0.024	19.06	33.63	38.94	50.81	77.38	8.353	0.00	0.03	0.14	1.59	33.52
SM(l-h)	0.036	10.90	24.23	29.71	42.33	72.69	3.280	9.83	58.96	74.58	92.93	99.98
SH(l-h)	0.036	11.19	24.61	30.08	42.69	72.89	4.172	0.04	6.29	14.61	41.48	96.56
ML(l-h)	0.072	2.53	10.54	15.02	26.99	62.90	3.675	0.73	23.16	39.38	71.29	99.51
MM(l-h)	0.003	65.36	73.64	76.53	81.72	91.98	3.367	5.76	49.28	66.46	89.28	99.95
MH(l-h)	0.001	82.19	86.65	88.04	90.73	95.94	3.323	7.61	54.14	70.68	91.31	99.96
BL(l-h)	0.095	1.05	6.63	10.38	21.01	58.17	4.377	0.01	3.85	9.80	32.45	94.19
BM(l-h)	0.258	0.00	0.56	1.54	6.22	39.41	6.072	0.00	0.25	0.89	5.87	62.02
BH(l-h)	0.010	39.62	53.21	57.62	66.60	85.16	3.453	3.36	40.66	58.56	84.97	99.90
SMB	0.057	4.56	14.58	19.66	32.15	66.50	4.283	0.02	4.76	11.73	36.29	95.33
HML	0.003	63.34	72.15	75.17	80.63	91.52	3.968	0.11	10.68	21.90	52.77	98.27
MKT	0.154	0.13	2.31	4.54	12.41	49.31	5.581	0.00	0.45	1.51	8.92	71.41
Multivariate												
	13.974	0.00	0.33	4.37	56.03	100.00	220.756	0.00	0.06	2.75	71.06	100.00

test are all greater than 5% for each of the twelve assets, and the multivariate test obtains a high *p*-value of 56%. This is an astonishing finding, as the *t* distribution is theoretically symmetric and has zero population skewness. The reason that we cannot reject the *t* is because the finite sample variation of the skewness of a *t* distribution is so large that it implies large probabilities for observing the large sample skewness. However, if the skewness is too large to be explained by the standard *t* distribution, a skewed *t* can be developed (see, e.g., Branco and Dey (2001)). When $v = 12$ or 16, there are two univariate rejections from the skewness test. Moreover, the multivariate skewness test rejects both $v = 12$ and $v = 16$, though the *p*-value when $v = 12$ is 4.37%, just slightly below the 5% level. On the other hand, the *p*-values are much larger for $v = 4$. Similar results hold for the kurtosis test as well. Theoretically, the kurtosis of a *t* distribution can go to infinity if one lowers the degrees of freedom. As a result, it is not surprising that a suitable *t* distribution should pass the kurtosis test, and therefore, if the data truly follows a *t* distribution, the degrees of freedom should be equal to eight or less. However, it should be noted that although $v = 4$ better explains the large sample skewness and kurtosis, it may not be a better model than $v = 8$ because its tails are extremely fat and hence may not explain the average data well. This is confirmed later by our Bayesian posterior

analysis. In short, we find that the multivariate normality assumption is strongly rejected by the data while a suitable t is well supported.[3]

The above results show that $v = 8$ is sufficient to explain the sample skewness and kurtosis of the data. But to what extent does this result depend on the choice of investment universe? To address this question, we apply the same tests to 20 industry portfolios constructed by sorting their two-digit SIC codes, following Moskowitz and Grinblatt (1999). This set of industry portfolios plus the earlier three factors form an interesting alternative investment universe because, as shown by King (1966), industry groupings well proxy the investment opportunity set: industry groupings maximize intragroup and minimize intergroup correlations. The left panel of Table 3 provides test results on the industry portfolios. Although there are many univariate rejections for $v = 12$, both multivariate skewness and kurtosis tests suggest that this value of v appears sufficient for the industry returns. For another example, not reported here, we apply the same tests to French's (1993) 25 assets, and find results similar to those of the industry portfolios. Thus, based on our multivariate tests, we see that the degrees of freedom varies as the investment universe changes.

A related question is whether or not $v = 8$ is sensitive to trading frequency. One can compound the above monthly data into quarterly returns. A re-run of the tests on the quarterly data shows that the normality assumption is still strongly rejected with multivariate p-values less than 0.01%. In addition, while $v = 20$ is sufficient for both the industry portfolios and the Fama and French assets, $v = 16$ is necessary for the Pástor and Stambaugh assets. This indicates that normality works better for the lower frequency data. The remaining question is how the earlier monthly results vary with a higher frequency. While daily series are not available for either the Fama and French assets or Pástor and Stambaugh assets, daily industry returns are available, as are the daily returns on the one-month Treasury bill and the factors.[4] The right panel of Table 3 provides test results. For higher frequency data, the normality assumption is rejected much more strongly, and v has to go down substantially to fit the data. When $v = 5$, the p-value of the multivariate skewness test is only 3.6%, but that of the kurtosis test is 11.4%. This suggests that $v = 5$ is close to fitting the skewness and kurtosis of the daily returns. Notice also that, for the industry returns, the p-values of the skewness test can be smaller than those of the kurtosis test. In short, compared with the monthly returns, deviations from normality increase substantially for the daily returns.

Finally, we construct weekly industry returns as the holding period returns from the end of Wednesday of one week through the end of the subsequent Wednesday by compounding daily returns in this holding period, and we then examine normality by using both the skewness and kurtosis tests. While the daily returns require that v be

---

[3] Zellner (1976) seems to be the first to conduct a Bayesian t analysis. However, there exist few studies in finance which use the multivariate t distribution (see, e.g., MacKinlay and Richardson (1991); Zhou (1993); Geczy (2001)), and these are all based on non-maximum likelihood methods. Statistically, these methods are very inefficient, as shown by Kan and Zhou (2003).

[4] The latter are downloaded from French's homepage. We are grateful to Ken French for making this data available at www.mba.tuck.dartmouth.edu/pages/faculty/ken.french.

Table 3

Normality test of 20 industry portfolios

The table reports the sample univariate and multivariate skewness and kurtosis measures of industry portfolios with monthly returns from July 1963 to December 1997 and daily returns from July 1, 1963 to December 31, 1997. In addition, it also reports the p-values (in percentage points) of the skewness and kurtosis tests if the data is assumed to be drawn from a univariate or multivariate normal distribution, or a univariate or multivariate t distribution with degrees of freedom equal to 16, 12, 8, or 5, 4, 3, respectively.

	Monthly										Daily									
	Skewness					Kurtosis					Skewness					Kurtosis				
	Stat.	Nor.	t with df			Stat.	Nor.	t with df			Stat.	Nor.	t with df			Stat.	Nor.	t with df		
			16	12	8			16	12	8			5	4	3			5	4	3
Univariate																				
Misc.	0.189	0.0	1.4	3.0	9.6	5.022	0.0	1.1	3.3	15.5	2.706	0.0	0.5	3.3	24.3	43.537	0.0	0.6	4.1	35.6
Mining	0.112	0.6	4.8	8.0	17.9	5.121	0.0	0.9	2.8	14.0	0.765	0.0	1.4	8.7	42.8	20.441	0.0	1.8	13.4	78.0
Food	0.004	59.9	69.5	72.7	78.8	5.648	0.0	0.4	1.4	8.4	1.418	0.0	0.9	5.2	32.6	34.820	0.0	0.8	5.5	45.3
Apparel	0.064	3.5	12.5	17.2	29.5	5.871	0.0	0.3	1.1	6.8	2.914	0.0	0.4	3.1	23.4	58.496	0.0	0.4	2.8	26.0
Paper	0.005	56.1	66.4	69.9	76.5	5.211	0.0	0.8	2.5	12.8	3.364	0.0	0.4	2.9	21.9	51.352	0.0	0.4	3.3	29.6
Chemical	0.005	56.5	66.8	70.3	76.8	5.831	0.0	0.3	1.1	7.1	1.804	0.0	0.8	4.5	29.3	39.011	0.0	0.7	4.8	40.2
Petroleum	0.011	38.0	51.8	56.3	65.5	4.782	0.0	1.6	4.8	20.1	0.715	0.0	1.5	9.2	43.7	27.861	0.0	1.1	7.9	57.8
Construction	0.069	2.9	11.3	15.9	28.0	6.174	0.0	0.2	0.8	5.4	3.612	0.0	0.4	2.7	21.1	58.989	0.0	0.3	2.8	25.7
Prim. Metals	0.020	23.3	38.0	43.3	54.5	5.058	0.0	1.0	3.1	14.9	4.304	0.0	0.3	2.4	19.3	64.738	0.0	0.3	2.6	23.6
Fab. Metals	0.220	0.0	0.9	2.3	7.8	5.670	0.0	0.4	1.4	8.2	2.094	0.0	0.7	4.1	27.5	35.155	0.0	0.8	5.5	44.7
Machinery	0.001	82.6	87.0	88.3	90.9	4.157	0.0	6.5	15.0	42.2	1.227	0.0	1.0	5.8	34.8	28.790	0.0	1.1	7.4	55.6
Electrical Eq.	0.017	27.5	42.0	47.3	57.9	4.383	0.0	3.8	9.7	32.2	0.560	0.0	1.9	10.7	47.8	19.137	0.0	2.1	15.1	82.3
Transport Eq.	0.065	3.3	12.2	16.9	29.2	5.836	0.0	0.3	1.1	7.0	0.567	0.0	1.9	10.6	47.6	20.468	0.0	1.8	13.4	77.9
Manufacturing	0.054	5.2	15.8	20.9	33.5	4.221	0.0	5.6	13.2	39.1	0.954	0.0	1.2	7.3	39.1	24.750	0.0	1.3	9.6	65.1
Railroads	0.000	91.0	93.2	94.0	95.4	4.270	0.0	4.9	12.0	36.9	0.970	0.0	1.1	7.1	38.8	29.349	0.0	1.0	7.2	54.5
Other Trans.	0.036	11.0	24.4	29.9	42.5	3.949	0.1	11.2	22.8	53.8	0.094	0.0	11.6	36.1	73.9	11.887	0.0	6.7	43.0	99.5
Utilities	0.093	1.1	7.0	10.7	21.6	4.601	0.0	2.4	6.5	24.9	1.407	0.0	0.9	5.3	32.7	44.753	0.0	0.5	4.0	34.6
Dept. Stores	0.011	36.9	50.9	55.4	64.8	4.579	0.0	2.5	6.8	25.6	0.367	0.0	3.0	14.8	54.8	20.392	0.0	1.8	13.5	78.2
Retail	0.170	0.1	1.8	3.7	11.0	5.470	0.0	0.5	1.7	9.9	1.205	0.0	1.0	5.9	35.0	24.653	0.0	1.3	9.7	65.3
Financial	0.009	41.5	54.7	59.1	67.8	4.222	0.0	5.6	13.2	39.0	1.378	0.0	0.9	5.3	33.0	28.241	0.0	1.1	7.7	56.8
SMB	0.057	4.6	14.6	19.7	32.1	4.283	0.0	4.8	11.7	36.3	4.204	0.0	0.3	2.5	19.4	57.723	0.0	0.4	2.9	26.4
HML	0.003	63.3	72.2	75.2	80.6	3.968	0.1	10.7	21.9	52.8	0.007	0.2	61.3	78.8	93.0	7.487	0.0	37.0	95.8	100.0
MKT	0.154	0.1	2.3	4.5	12.4	5.581	0.0	0.4	1.5	8.9	1.569	0.0	0.9	5.0	31.2	33.169	0.0	0.9	5.9	47.8
Multivariate																				
	56.830	0.0	4.8	52.8	99.9	671.962	0.0	3.6	64.3	100.0	202.795	0.0	3.6	34.0	99.3	1690.901	0.0	11.4	97.6	100.0

equal to 5 or smaller, the weekly returns are sufficiently modelled for $v = 8$. However, compared with $v = 12$ for the monthly industry returns, this is still substantially smaller, which confirms the conventional wisdom that weekly returns are closer to the normality than daily returns, but still farther from it than monthly returns.

2.2. Investing under a given DGP

Following the well-established framework of Pástor and Stambaugh (2000), consider an investment universe which contains cash plus n spread positions. The Pástor and Stambaugh set-up defines spread position i, constructed at the end of period $t - 1$, as a purchase of one asset coupled with an offsetting short sale of another. The two assets are denoted as L_i and S_i, and their rates of return in period t are denoted as $R_{L_i,t}$ and $R_{S_i,t}$. Thus, a spread position of size X_i has a dollar payoff $X_i(R_{L_i,t} - R_{S_i,t})$. Since regulation T requires the use of margins for risky investments, a constant $c > 0$ is used to characterize the degree of margin requirements. The spread position involves at least one risky asset which, without loss of generality, is designated as asset L_i. If the other asset of position i, S_i of size X_i, is risky as well, then $(2/c)|X_i|$ dollars of capital are required. Otherwise, $(1/c)|X_i|$ dollars of capital are required. For example, if $c = 2$, the set-up implies a 50% margin.

The total capital required to establish the spread positions must be less than or equal to the investor's wealth, W_{t-1}. That is

$$\sum_{i \in \Lambda} (2/c)|X_i| + \sum_{i \notin \Lambda} (1/c)|X_i| \leqslant W_{t-1}, \tag{3}$$

where Λ denotes the set of positions in which S_i is risky, or alternatively,

$$\sum_{i \in \Lambda} 2|w_i| + \sum_{i \notin \Lambda} |w_i| \leqslant c, \tag{4}$$

where $w_i = X_i/W_{t-1}$. Let r_t denote an n-vector such that its ith element $r_{i,t}$ ($= R_{L_i,t} - R_{S_i,t}$) represents the return of the ith risky position at time t. If there is a riskless asset with a rate of return $R_{f,t}$, then the excess return of this portfolio is

$$R_{p,t} - R_{f,t} = \sum_{i=1}^{n} w_i r_{i,t}. \tag{5}$$

The investor is assumed to choose w so as to maximize the mean-variance objective function

$$U = \mathrm{E}(R_{p,t}) - \tfrac{1}{2} A \, Var(R_{p,t}), \tag{6}$$

subject to the wealth constraint in Eq. (4), where A is interpreted as the coefficient of relative risk aversion. If we denote the mean vector and covariance matrix of r_t as E and V, respectively, then the investor's optimal portfolio choice problem can be

rewritten as the solution to

$$\max_w \left(w'E - \frac{1}{2} A w' V w \right) \quad \text{s.t.} \quad \sum_{i \in A} 2|w_i| + \sum_{i \notin A} |w_i| \leqslant c. \tag{7}$$

Assume that r_t follows a well-defined probability distribution,

$$r_t \sim P(E, \ V), \tag{8}$$

where P is the DGP of the data. In the classical framework, the investor solves problem (7) with E and V replaced by their estimators in light of the assumed distribution P. The density function of P is usually available, and E and V are then estimated using the maximum likelihood method. However, the estimation errors are often ignored. In the Bayesian framework, on the other hand, both parameter E and V are viewed as random variables, and the parameter uncertainty is captured by their posterior distributions in light of the data. Incorporating the parameter uncertainty amounts to integrating over the posterior distributions. As detailed in the next subsection, this is mathematically equivalent to replacing E and V in problem (7) by the predictive mean vector and covariance matrix of r_t conditional on available data.

2.3. Incorporating DGP uncertainty

As demonstrated earlier, the t distribution with $v \approx 8$ is maintained by the (monthly) Pástor and Stambaugh (2000) data, while the multivariate normality assumption is strongly rejected. The question here is that if investors do not have such knowledge about the DGP, how do we model their uncertainty with regard to which distributional assumption is suitable? To address this question, we assume that the investors know a set of probability distributions that are likely candidates for the true DGP. In our applications, the set consists of 31 multivariate t distributions with degrees of freedom varying from $2.1, 3, 4, 5, \ldots, 31$, to 32. The smaller the degrees of freedom, the greater the deviation from the normal distribution. The DGP uncertainty is captured by v, the degrees of freedom of the t distribution. Theoretically, a normal distribution has an infinite v, but is well modelled by a t with $v = 32$. If investors are unsure about the degree of nonnormality in the data, they may place an equal probability on all possible values of v.

Formally, we specify a set of 31 multivariate t distributions, with degrees of freedom varying from $2.1, 3, 4, 5, \ldots, 31$, to 32, as the likely candidates for the DGP. The returns thus follow

$$r_t \sim DGP(E, V, v), \tag{9}$$

where the DGP is a t distribution with an unknown parameter v. One may argue that this t assumption is unnecessary because, with suitable priors, the predictive density of the returns under the normality assumption is also t. However, the degrees of freedom of the predictive density is too high to be any different from normal. As a

result, to model the fat tails of the data, we have to assume a t distribution for the DGP, as we did in (9).

In the Bayesian framework, v is viewed as a random variable which takes values from the set $S_v = \{2.1, 3, 4, 5, \ldots, 31, 32\}$. The diffuse prior on v can be written as

$$p_0(v) = \frac{1}{|S_v|}, \tag{10}$$

where $|S_v|$ is the number of elements in the set S_v, which is 31 in this case. Mathematically, the DGP can be any distribution, and v can then simply be an index parameter. However, as we restrict the candidate distributions to the t class, we can interpret the parameter v as an indicator of data's kurtosis. Because the kurtosis is a decreasing function of v, the smaller its value, the more leptokurtic the data.

It is of interest to examine some alternative and perhaps informative priors other than the diffuse prior. In particular, as the multivariate normality is used so extensively in finance, we also consider a prior of $p_0(v = 32) = 50\%$, 75%, or 99%, implying that one has the prior belief that a normal distribution has, respectively, a 50%, 75%, or 99% probability of being the true DGP.

To conduct the necessary Bayesian analysis of the posterior distribution of v, we need to impose priors on other parameters of the model as well. For a more intuitive understanding of the priors, it is useful to cast the DGP into a regression setting. Let $r_t = (y_t, x_t)$, where y_t contains the excess returns of m non-benchmark positions and x_t contains the excess returns of k $(= n - m)$ benchmark positions. Consider the familiar multivariate regression

$$y_t = \alpha + B x_t + u_t, \tag{11}$$

where u_t is an $m \times 1$ vector with zero means and a non-singular covariance matrix. To relate α and B to the earlier parameters E and V, consider the corresponding partition

$$E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix}, \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}. \tag{12}$$

Under the usual multivariate normal distribution, it is clear that the distribution of y_t conditional on x_t is also normal, and the conditional mean is a linear function of x_t. Hence,

$$\mathrm{E}(y_t|x_t) = E_1 + V_{12} V_{22}^{-1}(x_t - E_2), \tag{13}$$

$$Var(y_t|x_t) = V_{11} - V_{12} V_{22}^{-1} V_{21}. \tag{14}$$

Therefore, the parameters α and B obey the following relationship with the earlier parameters E and V:

$$\alpha = E_1 - B E_2, \quad B = V_{12} V_{22}^{-1}, \tag{15}$$

and

$$\Sigma = V_{11} - B V_{22} B'. \tag{16}$$

The Σ matrix is the familiar notation for the covariance matrix of u_t under the usual normality assumption. However, here the DGP is in the class of the t distributions. While the mean of y_t conditional on x_t is still a linear function of x_t as above, its conditional covariance matrix is no longer independent of x_t

$$Var(y_t|x_t) = k(x_t, \ v)(V_{11} - V_{12}V_{22}^{-1}V_{21}), \tag{17}$$

where $k(x_t, \ v)$ is a function of both the quadratic form $(x_t - E_2)'V_{22}^{-1}(x_t - E_2)$ and the degrees of freedom v. Correspondingly, the covariance matrix of the residual term u_t is $k(x_t, \ v)\Sigma$. Hence, while we continue to use Σ as in the earlier definition that $\Sigma = V_{11} - BV_{22}B'$, this term is no longer equal to the covariance matrix of u_t (unless $v = \infty$). As a result, we must take special care in drawing the parameters from their posterior distribution even though the mapping of the parameters can be done in exactly the same way as in the normal case.

We turn our attention to priors on α, B, and Σ. By allowing for uncertainty about mispricing in asset pricing models, Pástor and Stambaugh (2000) and Pástor (2000) develop ways to form useful priors. We follow this development here completely. Different pricing models impose different restrictions on α. For examples, a factor-based pricing model such as the CAPM or the Fama and French (1993) restricts α to zero. To allow for mispricing uncertainty, however, the prior distribution of α is specified as a normal distribution conditional on Σ,

$$\alpha|\Sigma \sim \mathrm{N}\left(0, \sigma_\alpha^2\left(\frac{1}{s^2}\Sigma\right)\right), \tag{18}$$

where s^2 is a suitable prior estimate for the average of the diagonal elements of Σ. This alpha–sigma link is also explored by MacKinlay and Pástor (2000) in the frequentist set-up. The numerical value of σ_α represents an investor's prior degree of uncertainty with regards to a given model's pricing ability. When $\sigma_\alpha = 0$, the investor believes dogmatically in the model, that is, that there is no mispricing uncertainty. On the other hand, when $\sigma_\alpha = \infty$, the investor believes that the pricing model is entirely useless.

However, for the Daniel and Titman's (1997) characteristic-based model, the prior on α requires a different treatment. This is because the pricing model now imposes the restriction that $E_1 = 0$ given that the characteristics of the long and short positions of the nine non-benchmark risky spreads are matched (approximately). The implication is that $\alpha = -BE_2$ in Eq. (11). To allow for mispricing uncertainty, we use Pástor and Stambaugh's prior on α which follows the normal distribution:

$$\alpha|\Sigma \sim N(-BE_2, \sigma_\alpha^2 I_m). \tag{19}$$

The remaining priors are fairly standard and are detailed in the appendix. Thus, the complete prior on all the parameters can be written as

$$p_0(\theta) = p_0(\alpha|\Sigma)p_0(\Sigma)p_0(B)p_0(E_2)p_0(V_{22})p_0(v). \tag{20}$$

An important noteworthy fact about the priors is that we assume independence between the DGP and the mispricing uncertainties. This may not be true in general, since a given DGP may not allow the validity of an asset pricing model and hence

can disallow a dogmatic belief about the model. However, Chamberlain (1983) and Owen and Rabinovitch (1983) show that the elliptical class, of which t is a special case, is the largest class of distributions that validates the mean-variance framework, and hence the independence assumption is innocuous for both the CAPM and the Fama and French (1993) models. For the Daniel and Titman (1997) model, the t assumption on returns does not seem to have any impact on the characteristics relation, and hence the independence assumption between the DGP and the mispricing uncertainties also appears reasonable.

With the priors given above, the investor forms her posterior belief $p(\theta|R)$ in light of the data $\{R: r_t, \ t = 1, \dots, T\}$. The predictive distribution of r_{T+1} can then be computed as

$$p(r_{T+1}|R) = \int_\theta p(r_{T+1}|R, \theta) p(\theta|R) \, \mathrm{d}\theta. \tag{21}$$

When the investor solves (7) by integrating over the parameter uncertainty, it is easy to see that this is equivalent to replacing E and V with E^* and V^*, the mean and covariance matrix of the predictive distribution. By the law of iterated expectations, it is clear that the predictive mean obeys the relation

$$E^* = \mathrm{E}(r_{T+1}|R) = \mathrm{E}(E(r_{T+1}|\theta, R)|R) = \mathrm{E}(E|R). \tag{22}$$

This says that the predictive mean is identical to the posterior mean. Hence, to the extent that the difference between the classical estimated mean and the posterior mean is small, the classical approach of plugging in the estimates for the mean should have little effect on the valuation of the quadratic utility function. However, the predictive covariance matrix obeys the relation

$$\begin{aligned} V^* &= Var(r_{T+1}|R) \\ &= \mathrm{E}(Var(r_{T+1}|\theta, R)|R) + Var(\mathrm{E}(r_{T+1}|\theta, R)|R). \end{aligned} \tag{23}$$

This says that the predictive variance is a sum of two components, the posterior mean of the variance of a given model (which is the posterior mean of V under normality and a diffuse prior), and the variance of E due to its uncertainty. Since the classical estimation usually provides only an estimate for the first component, simply plugging-in this estimate is likely to over-value the quadratic utility function. Finally, as the DGP is of the class of t distributions, the implementation of both the posterior and the predictive evaluations will be more complex than in the case of normality. The details are provided in the Appendix.

2.4. Performance measures

Given a combination of a prior on the DGP and a prior on the mispricing of an asset pricing model, there will be a unique predictive distribution that the investor uses to make her investment decisions. For example, with a given prior on the DGP, an investor who centers her prior belief on the Fama and French (1993) model would use the predictive distribution from this model to form an optimal portfolio that maximizes her expected utility. Clearly, different predictive distributions may imply

different moments of the returns, which in turn results in different portfolio selections. Of course, once different optimal portfolios are derived from alternative pricing models, it is of interest to judge the economic importance of the differences. However, the differences cannot be simply evaluated based on the portfolio weights alone. Because of the correlations among the payoffs of risky positions, the performance of two portfolio choices can be similar even when they are quite different in position-by-position allocations.

Pástor and Stambaugh (2000) introduce a useful measure to evaluate the economic value of portfolio differences. Consider, for example, the problem of measuring the economic importance of two competing asset pricing models, the Fama-French model and the CAPM, for a given prior on the DGP. A portfolio allocation, w_O, which is optimal under the predictive distribution from the Fama-French model with a certain mispricing prior and a certain prior on the DGP, is easily computed, and implies an expected utility of

$$\mathrm{E}U_O = w_O' E^* - \frac{1}{2} A w_O' V^* w_O, \tag{24}$$

where E^* and V^* are the predictive mean and covariance matrix of the Fama-French model. Another allocation, w_S, which is optimal under the predictive distribution from the CAPM with the same priors, gives rise to an expected utility of

$$\mathrm{E}U_S = w_S' E^* - \frac{1}{2} A w_S' V^* w_S. \tag{25}$$

Notice that this expected utility is evaluated based on the same E^* and V^* as that of the Fama-French model. Because $\mathrm{E}U_O$ and $\mathrm{E}U_S$ are the certainty–equivalent excess returns of the two allocations, the difference $CE = \mathrm{E}U_O - \mathrm{E}U_S$ is the "perceived" loss in terms of the certainty–equivalent return to an investor who is forced to accept the optimal portfolio selection based on the CAPM. Since w_O is optimal in the Fama-French model, CE is positive or zero by construction. In the same spirit, Fleming, Kirby and Ostdiek (2001) provide a similar but different measure in the classical framework.

The focus of this paper is on the impact of making a false normality assumption versus incorporating the DGP uncertainty into consideration. Hence, our use of the above certainty–equivalent loss will center on the case in which the two expected utilities (certainty–equivalent excess returns) are based on two optimal allocations in the same asset pricing model such that one is obtained under normality and the other under DGP uncertainty. More precisely, let w_t be the optimal allocation in an asset pricing model under suitable priors and under DGP uncertainty, and let w_n be the optimal allocation in the same asset pricing model under normality. Then the difference of the certainty–equivalent excess returns,

$$\mathrm{E}U^t - \mathrm{E}U^n = [w_t' E^* - \tfrac{1}{2} A w_t' V^* w_t] - [w_n' E^* - \tfrac{1}{2} A w_n' V^* w_n], \tag{26}$$

is the perceived loss due to the false normality belief, where E^* and V^* are the predictive mean and covariance matrix under DGP uncertainty. Intuitively, if an investor believes in normality, she chooses her optimal portfolio according to the model under normality. However, the true state of the world is that of DGP

uncertainty, and thus her true utility should be evaluated under DGP uncertainty. As a result, the investor's utility under normality should be no greater than that under DGP uncertainty, i.e., EU^t. The issue that remains is how large this value can be. Generally speaking, if the difference is over a couple of percentage points per year, we can interpret it as a signal that a false belief in the normality assumption results in an economically significant loss in terms of the certainty–equivalent return. For easy reference, we refer to $EU^t - EU^n$ as the certainty–equivalent gain below.

3. Empirical results

In this section, we report the empirical results of applying the proposed methodology in Section 2 to the same data used by Pástor and Stambaugh (2000). First, we examine how the data sheds light on which one of the 31 DGPs is the most likely candidate to be the true process. Second, we provide estimates of the means and variances under DGP uncertainty, showing in particular that these estimates can be drastically different from those obtained under the normality assumption and that they can thus be important for computing the cost of capital. Third, we report the optimal portfolio weights under DGP uncertainty and compare them with those obtained under normality. Finally, we investigate portfolio performance in terms of the certainty–equivalent gains. To examine the sensitivity of this performance measure to the use of alternative assets, we also report the certainty–equivalent gains when we replace the investment universe first by the 20 industry portfolios, and then by Fama and French's (1993) three factors.

3.1. Which data-generating process?

As discussed in Section 2.1, the classical tests suggest a t distribution with approximately eight degrees of freedom for returns on Pástor and Stambaugh's (2000) spread positions and Fama and French's (1993) three factors. In our Bayesian analysis, we allow a wide range of degrees of knowledge on the true DGP by using t distributions with degrees of freedom varying from $2.1, 3, 4, 5, \ldots, 31$ to 32. We thereby allow distributions to range from those with too much excess kurtosis to a distribution virtually identical to the normal. Viewing the true degrees of freedom, v, as a random parameter drawn from the 31 possible states, the Bayesian analysis translates any prior belief on v into a posterior belief.

Consider four priors on v. The first one is a diffuse prior that assigns an equal probability of $\frac{1}{31}$ to v for each of the 31 states. The other three priors are all informative and assign a probability of $\frac{1}{2}$, $\frac{3}{4}$, and $\frac{99}{100}$, respectively, to a normal distribution (as represented here by the last state, $v = 32$), and an equal probability to the other states. In addition, a diffuse prior on the mispricing of asset pricing models is used to isolate the evidence on DGP uncertainty from that on model mispricing. However, even if informative priors on mispricing were used, the results would change little. Fig. 1 plots both the prior and posterior distributions of v. With the diffuse prior, the posterior has a huge concentration centered around $v = 8$, some
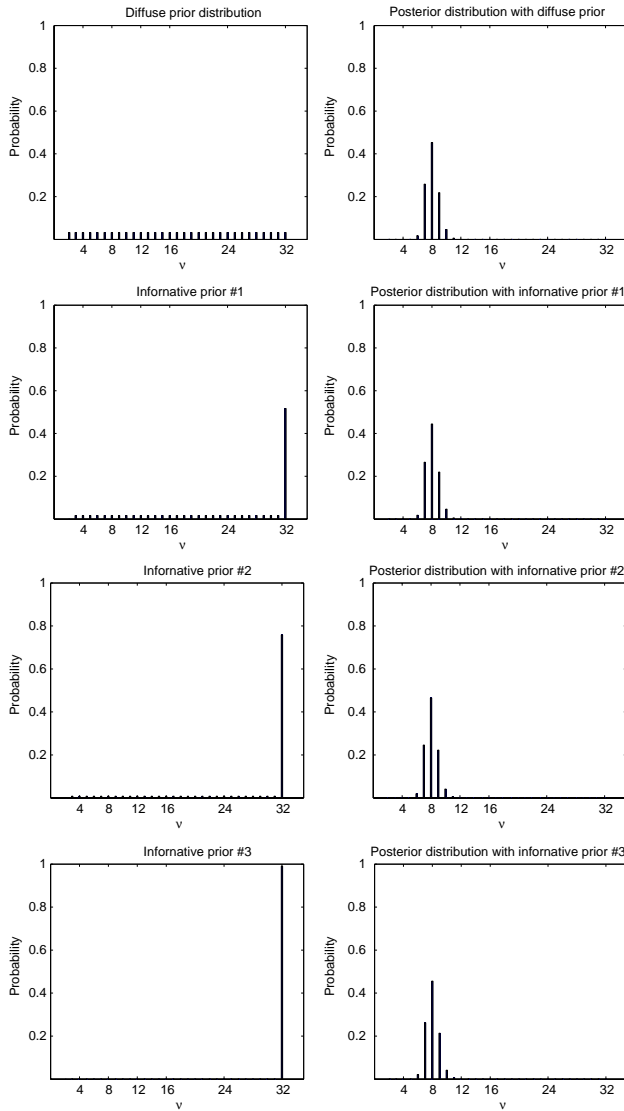
Fig. 1. Priors and posteriors of degrees of freedom. The figure displays both the prior and posterior densities with respect to the 12 assets of Pástor and Stambaugh (2000) for the degrees of freedom, v, which takes values from $S_v = \{2.1, 3, 4, 5, \ldots, 31, 32\}$. Of the four priors on v, the first is a diffuse prior that assigns a probability of $\frac{1}{31}$ to v for each of the 31 states, and the other three priors are informative, assigning a probability of $\frac{1}{2}$, $\frac{3}{4}$, and $\frac{99}{100}$, respectively, to a normal distribution (as represented here by the last state, $v = 32$), and an equal probability to the other states.

mass (about 4%) on $v = 6$ and $v = 10$, and virtually null on the other degrees of freedom. It is striking that the shape of the posterior density is almost unchanged even if one has a prior belief in a normal distribution of 99%! Intuitively, as the

earlier kurtosis test implies a *p*-value of less than 0.00001, the normal must fit the data too poorly to be changed by a 99% prior. Statistically, the difference in the log-likelihoods between the *t* and the normal is over 65, so a 99% prior will be unable to alter the posterior concentration. On the other hand, the posterior standard deviation of *v* is about 0.90 across the priors, implying that a strong prior on $v = 9$ or $v = 7$ should have a much greater impact on the posterior distribution of *v* than a strong prior on $v = 32$. In short, the data speaks loudly against a normal DGP and favors the DGP of a *t* distribution with approximately eight degrees of freedom, consistent with the earlier results based on the classical analysis. This is also consistent with Kan and Zhou's (2003) asymptotic analysis. The econometric and statistical literature lacks empirical studies on the *t* model when *v* is unknown and x_t is random. To obtain additional insight, we simulate 10,000 data sets from a *t* with the Bayesian posterior mean under the diffuse prior as the true parameters and $v = 8$, and we compute the exact maximum likelihood estimates. We find that the mean and standard deviation are $\bar{v} = 8.23$ and 0.91, respectively, re-affirming the informativeness of the data on *v* from the classical perspective.

One may ask why a value of $v = 4$ is not chosen by the Bayesian analysis since it obtains much greater *p*-values to easily pass both the multivariate skewness and kurtosis tests. The reason is that the skewness and kurtosis tests examine the tail distribution of the data. While $v = 4$ is a much better choice to describe the tails, it is not a good choice to model the middle distribution of the data. This can be verified in the classical framework by showing that the maximum likelihood value associated with $v = 4$ is much smaller than that associated with $v = 8$. In the Bayesian framework, this is done by examining the posterior distribution of *v*, as shown above.

3.2. Means and variances of the assets

Before analyzing the portfolio decisions, it is of interest to examine the fundamental parameters of the model. If there are no substantial differences in the parameter estimates, it would be unlikely to find any differences in portfolio choices; on the other hand, any differences found in parameter estimates are useful for understanding differences in portfolio selections.

The fundamental parameters of the model are the means and covariances of the assets. In a Bayesian decision context, as demonstrated by Zellner and Chetty (1965), the predictive densities are of primary interest. By Eq. (22), the predictive mean is the same as the posterior mean which is, under normality and a diffuse prior, simply the sample mean. However, by (23), the predictive variance has an additional component that incorporates measurement error into consideration. The empirical estimates are reported in Tables 4 and 5. Panel A of Table 4 provides results of the means under the hypothesis that the data is generated from a normal distribution–the common DGP assumption. It is seen that by varying both model mispricing errors and asset pricing models, the predictive market mean excess return is virtually unchanged at 0.53% per month, the same as the sample mean, as reported in Table 1, except for a couple values with the Daniel and Titman (1997) model.

Table 4
Predictive means

The first panel of the table reports the predictive means (in percentage points) of the twelve Pástor and Stambaugh (2000) assets when the data-generating process is assumed to be multivariate normally distributed, while the second panel provides the differences between these results and the predictive means when the data-generating process is assumed to be uncertain (and falling into a class of t distributions, of which one is essentially the normal distribution).

	$\sigma_\alpha = 0$			$\sigma_\alpha = 1\%$			$\sigma_\alpha = 2\%$			$\sigma_\alpha = \infty$		
	DT	FF	CM	DT	FF	CM	DT	FF	CM	DT	FF	CM
Normality												
SL(l-h)	−0.00	−0.18	0.12	0.03	−0.13	0.09	0.05	−0.08	0.05	0.00	−0.00	−0.00
SM(l-h)	0.00	−0.13	0.05	0.02	−0.11	0.03	0.01	−0.08	−0.01	−0.05	−0.05	−0.05
SH(l-h)	0.00	−0.23	−0.04	0.01	−0.17	−0.03	0.02	−0.11	−0.03	−0.02	−0.02	−0.02
ML(l-h)	−0.00	−0.23	0.10	−0.11	−0.28	−0.03	−0.23	−0.34	−0.20	−0.42	−0.42	−0.42
MM(l-h)	−0.00	−0.11	0.07	0.02	−0.09	0.04	0.02	−0.07	0.00	−0.05	−0.05	−0.05
MH(l-h)	0.00	−0.18	−0.00	0.00	−0.15	−0.01	0.00	−0.10	−0.03	−0.05	−0.05	−0.05
BL(l-h)	0.00	−0.35	0.02	−0.01	−0.29	−0.02	−0.04	−0.23	−0.07	−0.14	−0.14	−0.14
BM(l-h)	−0.00	−0.23	0.02	−0.03	−0.22	−0.04	−0.07	−0.21	−0.10	−0.20	−0.19	−0.19
BH(l-h)	−0.00	−0.13	0.07	−0.01	−0.13	0.02	−0.03	−0.13	−0.04	−0.12	−0.12	−0.12
SMB	0.21	0.22	0.10	0.26	0.22	0.13	0.24	0.22	0.17	0.22	0.22	0.22
HML	0.42	0.43	−0.11	0.30	0.43	0.02	0.34	0.43	0.20	0.43	0.43	0.43
MKT	0.53	0.53	0.53	0.63	0.53	0.53	0.59	0.53	0.53	0.53	0.53	0.53
DGP uncertainty minus normality												
SL(l-h)	0.00	0.05	−0.03	−0.01	0.02	−0.04	−0.04	−0.03	−0.06	−0.09	−0.09	−0.09
SM(l-h)	0.00	0.02	−0.00	−0.01	0.01	−0.00	−0.01	−0.00	−0.01	−0.02	−0.02	−0.02
SH(l-h)	−0.00	0.03	0.01	0.03	0.03	0.02	0.05	0.04	0.04	0.06	0.06	0.07
ML(l-h)	−0.00	0.02	−0.01	−0.02	−0.01	−0.01	−0.05	−0.03	−0.03	−0.09	−0.08	−0.09
MM(l-h)	0.00	0.02	−0.01	0.02	0.03	0.01	0.03	0.03	0.02	0.03	0.04	0.04
MH(l-h)	−0.00	0.02	0.00	0.01	0.02	0.01	0.02	0.02	0.02	0.03	0.03	0.04
BL(l-h)	−0.00	0.02	0.01	−0.01	0.00	−0.00	−0.03	−0.02	−0.02	−0.05	−0.05	−0.05
BM(l-h)	0.00	0.02	0.02	0.00	0.02	0.01	0.01	0.02	0.02	0.02	0.02	0.02
BH(l-h)	0.00	0.01	−0.01	−0.02	−0.00	−0.01	−0.03	−0.02	−0.02	−0.05	−0.05	−0.04
SMB	0.03	−0.06	−0.03	−0.05	−0.06	−0.04	−0.05	−0.06	−0.05	−0.06	−0.06	−0.06
HML	0.02	−0.02	0.01	−0.05	−0.02	−0.01	−0.04	−0.02	−0.02	−0.02	−0.02	−0.02
MKT	0.02	−0.09	−0.09	−0.03	−0.09	−0.09	−0.06	−0.09	−0.09	−0.09	−0.09	−0.09

However, for both the SMB and HML factors, this is not the case. Under the informative priors centered around the CAPM, the mean returns of the SMB and HML factors deviate from their sample means substantially. Similarly, with the informative priors centered around any of the three models, the mean excess returns of the non-factor assets also deviate from their sample means substantially.

For easier assessment of the changes caused by DGP uncertainty, Panel B of Table 4 reports, under the subtitle "DGP uncertainty minus normality", the differences between the predictive means under DGP uncertainty and those under normality. It is interesting to note that among the three factors, the greatest difference about 1.1% per year occurs for the market excess return. This difference is large enough that it may indeed be economically important in the estimation of the

Table 5
Predictive standard deviations

The first panel of the table reports the predictive standard deviations (in percentage points) of the twelve Pástor and Stambaugh (2000) assets when the data-generating process is assumed to be multivariate normally distributed, while the second panel provides the differences between these results and those obtained when the data-generating process is assumed to be uncertain (and falling into a class of t distributions, of which one is essentially the normal distribution).

	$\sigma_\alpha = 0$			$\sigma_\alpha = 1\%$			$\sigma_\alpha = 2\%$			$\sigma_\alpha = \infty$		
	DT	FF	CM	DT	FF	CM	DT	FF	CM	DT	FF	CM
Normality												
SL(l-h)	4.13	4.12	4.11	4.13	4.13	4.11	4.13	4.13	4.11	4.14	4.13	4.12
SM(l-h)	2.47	2.47	2.46	2.47	2.47	2.46	2.47	2.47	2.46	2.47	2.47	2.46
SH(l-h)	2.23	2.22	2.21	2.23	2.22	2.21	2.23	2.22	2.22	2.23	2.23	2.22
ML(l-h)	3.05	3.04	3.03	3.04	3.04	3.03	3.03	3.03	3.02	3.03	3.03	3.01
MM(l-h)	2.54	2.53	2.52	2.54	2.53	2.52	2.54	2.54	2.52	2.54	2.54	2.52
MH(l-h)	2.92	2.91	2.91	2.92	2.92	2.91	2.92	2.92	2.91	2.93	2.92	2.91
BL(l-h)	3.05	3.03	3.04	3.05	3.04	3.04	3.05	3.04	3.04	3.06	3.05	3.04
BM(l-h)	3.41	3.41	3.40	3.41	3.41	3.40	3.41	3.41	3.40	3.41	3.41	3.40
BH(l-h)	3.30	3.30	3.29	3.30	3.30	3.29	3.30	3.30	3.29	3.31	3.30	3.29
SMB	2.89	2.89	2.89	2.89	2.89	2.89	2.89	2.89	2.89	2.89	2.89	2.88
HML	2.54	2.55	2.58	2.55	2.55	2.57	2.55	2.55	2.56	2.55	2.55	2.55
MKT	4.32	4.33	4.32	4.33	4.33	4.32	4.33	4.33	4.32	4.33	4.33	4.32
DGP uncertainty minus normality												
SL(l-h)	−0.40	−0.39	−0.42	−0.40	−0.39	−0.42	−0.39	−0.40	−0.41	−0.39	−0.40	−0.42
SM(l-h)	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.03	0.02	0.01
SH(l-h)	0.01	0.01	−0.00	0.01	0.01	−0.00	0.01	0.01	−0.00	0.01	0.01	−0.01
ML(l-h)	0.13	0.14	0.12	0.13	0.13	0.12	0.12	0.13	0.12	0.12	0.12	0.10
MM(l-h)	0.03	0.03	0.02	0.03	0.03	0.02	0.03	0.03	0.02	0.03	0.03	0.02
MH(l-h)	−0.00	−0.00	−0.02	−0.00	−0.00	−0.01	0.00	0.00	−0.01	−0.00	−0.00	−0.02
BL(l-h)	0.03	0.03	0.02	0.03	0.03	0.02	0.03	0.03	0.02	0.03	0.03	0.01
BM(l-h)	0.02	0.02	0.00	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.00
BH(l-h)	0.08	0.08	0.06	0.08	0.08	0.06	0.08	0.08	0.06	0.08	0.08	0.06
SMB	−0.09	−0.09	−0.10	−0.09	−0.09	−0.10	−0.09	−0.09	−0.09	−0.09	−0.09	−0.10
HML	0.01	0.00	−0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	−0.00
MKT	−0.12	−0.12	−0.13	−0.12	−0.12	−0.13	−0.12	−0.12	−0.13	−0.12	−0.12	−0.14

cost of capital; and may be explained by the market's high skewness and kurtosis, as reported in Table 2. Both the SMB and HML factors also exhibit similar patterns. Under DGP uncertainty, not only is the expected return on the market lower, but the expected returns on the SMB and HML factors are lower also. The impact on the HML factor is very minimal, however, apparently because it deviates little from the normal. Theoretically, with the diffuse prior, the estimate under normality is essentially the same as the sample mean, and the estimate under DGP uncertainty is the maximum likelihood estimator based on the t density functions. Although both converge asymptotically to the true parameters (viewed from the classical statistical perspective), the latter is more accurate given it has smaller variances (see, e.g., Kan and Zhou (2003)). There are two determinants for the estimation accuracy. The

fatter the tails, the more accurate the latter because the t distribution weights outliers less heavily than the normal does. Another determinant is dimensionality. Under normality, the efficient expected return estimation depends only on an asset's own history, but under the t, it also depends on its covariations with other assets. The higher the dimensionality, the more accurate the t method. These two determinants help us understand why the factor mean estimates are different under normality than under DGP uncertainty. For the non-factor assets, the pattern of expected return estimates is not so closely related to the skewness and kurtosis measures. This is because non-factor assets' tail properties are different, and the tail properties are what determine how the expected return estimates may shift away from those of the normal (see Kan and Zhou (2003) for an analytical proof of this point). In addition, the estimated predictive means are affected not only by their own deviations from the normal, but also by both the factors and the given asset regression model.

Table 5 provides a similar analysis for the predictive standard deviations whose patterns are qualitatively similar to those found on the predictive means. Like the predictive means, the predictive standard deviations of the factors under the normality assumption are higher than those under the DGP uncertainty. The intuition for this result is that, due to outliers, a normal distribution usually requires a larger variance to fit the data than a t distribution does. However, the percentage differences are small, especially for the non-factor assets.

3.3. Portfolio choice

In this subsection, we investigate how incorporation of the DGP affects portfolio choice decisions. Consider first the case in which the pricing model is that of Fama and French (FF) and there are no pricing errors. It is well known that an exact-factor pricing model like the FF tells investors that expected excess returns are linear combinations of risk exposures (betas) to K sources of risk (factors). Any mean-variance efficient portfolio is a combination of K benchmark positions that mimic those factors. Therefore, a mean-variance investor with a dogmatic belief in the FF should hold only a combination of the three benchmark positions, and zero amounts in the nine non-benchmark positions. Following Pástor and Stambaugh (2000), Table 6 reports the optimal weights per \$100 in the benchmark positions under the normal DGP with varying mispricing priors for a mean-variance optimizing investor with relative risk aversion equal to 2.83. Our results match those of Pástor and Stambaugh with minimal differences caused perhaps by random errors in simulations and minor data-updating. As Pástor and Stambaugh (2000) have already offered excellent explanations for the results under the normal DGP, we will focus on the results under DGP uncertainty. When there is a 50% margin requirement ($c = 2$), the optimal allocations on the market under DGP uncertainty fall by roughly 10% to 12%. The percentage fall reduces when $c = 5$ or 10 (not reported), but is still substantial even for $c = \infty$ with $\sigma_\alpha \leqslant 2\%$. Similarly, the optimal allocations on the HML factor change substantially. However, the optimal allocations on the SMB factor change little when $c = 2$, though they do change more for $c = \infty$.

Table 6

Portfolio choice under normality and under DGP uncertainty

The first panel of the table reports the optimal portfolio weights per \$100 of a mean-variance optimizing investor with relative risk aversion equal to 2.83. The maximum value of risky positions that can be established per dollar of wealth is denoted by c. The panels under normality assumption provide the results when the data-generating process is assumed to be multivariate normally distributed, while the other panels provide the differences between these results and those obtained when the data-generating process is assumed to be uncertain (and falling into a class of t distributions, of which one is essentially the normal distribution).

	$\sigma_\alpha = 0$			$\sigma_\alpha = 1\%$			$\sigma_\alpha = 2\%$			$\sigma_\alpha = \infty$		
	DT	FF	CM	DT	FF	CM	DT	FF	CM	DT	FF	CM
$c = 2$												
Normal												
SL(l-h)	0.0	0	0	−0.0	−0.0	0.0	0.0	0.0	0.0	0.0	0.0	−0.0
SM(l-h)	−0.0	0	0	−0.0	0.0	0.0	−0.0	−0.0	−0.0	0.0	−0.0	−0.0
SH(l-h)	−0.0	0	0	−0.0	−0.0	0.0	0.0	−0.0	−0.0	−0.0	0.0	0.0
ML(l-h)	0.0	0	0	−0.0	−0.0	−12.6	−0.0	−0.0	−12.4	−11.5	−11.6	−11.9
MM(l-h)	0.0	0	0	−0.0	0.0	0.0	0.0	−0.0	−0.0	−0.0	−0.0	0.0
MH(l-h)	−0.0	0	0	0.0	−0.0	0.0	0.0	0.0	0.0	0.0	−0.0	0.0
BL(l-h)	0.0	0	0	0.0	−0.0	0.0	−0.0	−0.0	0.0	−0.0	−0.0	−0.0
BM(l-h)	0.0	0	0	0.0	−0.0	−0.0	−0.0	−0.0	−0.0	−0.0	−0.0	−0.0
BH(l-h)	−0.0	0	0	−0.0	0.0	−0.0	−0.0	−0.0	0.0	0.0	−0.0	−0.0
SMB	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HML	61.9	62.1	0	50.2	62.2	35.3	54.2	62.2	41.7	50.8	50.8	50.5
MKT	76.3	75.8	100.0	99.5	75.6	104.3	91.6	75.6	91.8	75.2	75.1	75.2
Cash	23.7	24.2	0.0	0.5	24.4	−4.3	8.4	24.4	8.2	24.8	24.9	24.8
DGP uncertainty minus normality												
SL(l-h)	−0.0	0	0	0.0	−0.0	0.0	−0.0	−0.0	−0.0	−0.0	0.0	−0.0
SM(l-h)	0.0	0	0	0.0	−0.0	−0.0	−0.0	0.0	0.0	−0.0	−0.0	−0.0
SH(l-h)	0.0	0	0	0.0	0.0	0.1	0.0	0.0	−0.0	0.0	−0.0	0.0
ML(l-h)	0.0	0	0	−0.0	−0.0	−10.9	−16.8	−3.8	−28.9	−53.5	−53.0	−55.2
MM(l-h)	−0.0	0	0	0.0	−0.0	0.1	0.0	0.0	−0.0	0.0	0.0	0.0
MH(l-h)	0.0	0	0	0.0	0.0	0.0	−0.0	−0.0	−0.0	−0.0	−0.0	−0.0
BL(l-h)	0.0	0	0	−0.0	−0.0	0.0	−0.0	−0.0	0.0	0.0	−0.0	−0.0
BM(l-h)	−0.0	0	0	−0.0	−0.1	−0.0	−0.0	−0.0	0.0	0.0	−0.0	0.0
BH(l-h)	0.0	0	0	0.0	−0.0	−0.0	−0.0	−0.0	0.0	−0.0	0.0	−0.0
SMB	0.0	0.0	0	−0.0	−0.0	0.1	0.0	0.0	0.0	−0.0	−0.0	0.0
HML	−2.7	4.1	0	−1.9	4.3	−6.3	−15.8	0.5	−24.0	−47.9	−47.4	−49.9
MKT	5.4	−8.2	−11.5	3.8	−9.0	−9.5	−2.0	−8.5	−9.9	−11.1	−11.3	−10.7
Cash	−5.4	8.2	11.5	−3.8	9.0	9.5	2.0	8.5	9.9	11.1	11.3	10.7
$c = \infty$												
Normal												
SL(l-h)	16.1	0	0	27.1	16.5	15.3	49.7	37.2	35.8	65.2	64.9	64.9
SM(l-h)	59.7	0	0	57.9	20.0	18.4	75.2	44.7	43.0	78.1	78.5	78.4
SH(l-h)	176.6	0	0	168.5	46.4	42.4	177.7	104.7	100.8	182.1	182.5	183.1
ML(l-h)	62.5	0	0	−48.1	−48.0	−45.7	−117.7	−108.3	−106.4	−188.3	−188.7	−192.5
MM(l-h)	−29.4	0	0	−9.2	4.1	3.9	10.7	8.7	8.6	17.0	16.0	15.9
MH(l-h)	36.0	0	0	29.1	8.1	7.1	31.4	17.9	17.3	31.4	31.5	31.3
BL(l-h)	176.9	0	0	156.2	37.7	34.3	160.0	84.8	81.3	148.2	148.4	147.3

Table 6. (*Continued*)

	$\sigma_\alpha = 0$			$\sigma_\alpha = 1\%$			$\sigma_\alpha = 2\%$			$\sigma_\alpha = \infty$		
	DT	FF	CM	DT	FF	CM	DT	FF	CM	DT	FF	CM
BM(l-h)	2.5	0	0	−2.8	−8.9	−8.3	−14.8	−19.6	−19.2	−34.8	−35.0	−35.4
BH(l-h)	−1.0	0	0	−10.5	−5.3	−5.0	−11.6	−11.7	−11.7	−21.0	−20.8	−21.5
SMB	79.4	63.2	0	95.4	71.1	22.7	96.1	81.3	53.1	96.1	96.1	96.2
HML	608.3	352.5	0	462.8	378.4	107.7	464.7	412.5	253.5	465.4	463.9	460.6
MKT	209.3	168.2	100.0	212.4	178.2	124.8	211.5	192.0	160.1	212.5	212.4	213.1
Cash	−109.3	−68.2	0.0	−112.4	−78.2	−24.8	−111.5	−92.0	−60.1	−112.5	−112.4	−113.1
DGP uncertainty minus normality												
SL(l-h)	−9.1	0	0	−15.8	−9.1	−9.0	−21.6	−20.1	−20.2	−35.0	−34.3	−35.0
SM(l-h)	−14.5	0	0	−26.0	−10.3	−9.6	−32.6	−20.4	−20.0	−35.6	−36.9	−37.5
SH(l-h)	4.5	0	0	5.0	2.8	3.9	17.6	11.1	12.0	31.6	31.1	34.6
ML(l-h)	−10.9	0	0	−15.5	−0.1	1.1	−14.9	−4.0	−2.4	−17.9	−17.5	−21.3
MM(l-h)	−10.4	0	0	12.5	8.0	5.9	24.5	17.6	16.2	29.6	32.1	36.4
MH(l-h)	1.0	0	0	−6.5	−0.7	−0.7	−5.4	−0.5	−0.8	−1.6	0.8	0.1
BL(l-h)	1.4	0	0	−35.5	−14.7	−13.4	−43.7	−31.1	−30.1	−50.6	−50.1	−50.6
BM(l-h)	16.6	0	0	10.9	5.0	4.8	12.8	10.3	9.8	19.4	18.3	17.9
BH(l-h)	−8.0	0	0	−13.8	−5.0	−4.6	−19.0	−12.5	−12.5	−23.8	−24.7	−24.9
SMB	64.6	22.0	0	15.6	19.4	0.8	14.4	14.8	4.1	13.4	13.6	14.2
HML	23.7	−9.3	0	−67.4	−25.1	−24.5	−70.9	−45.7	−49.6	−70.1	−68.7	−68.1
MKT	15.8	−10.5	−11.5	0.5	−10.3	−12.8	−1.5	−7.7	−9.9	−1.2	−1.7	0.3
Cash	−15.8	10.5	11.5	−0.5	10.3	12.8	1.5	7.7	9.9	1.2	1.7	−0.3

Consider now the case in which the pricing model is still the Fama and French three-factor model, but there exits mispricing uncertainty of 1%, i.e., $\sigma_\alpha = 1\%$. The panel in Table 6 under the subtitle "DGP Uncertainty Minus Normality" reports the differences between the optimal allocations under DGP uncertainty and those under normality. When there are no margin requirements, we see that among investors' non-benchmark long positions, there are huge decreases in three positions, SL(l-h) (55.2%), SM(l-h) (51.5%), and BL(l-h) (39.0%). There is also a huge increase in MM(l-h) (195.1%), followed by a significant decrease in BM(l-h) (56.2%) and a significant increase in BH(l-h) (94.3%) among the non-benchmark short positions when we switch from the normal DGP to DGP uncertainty.

Besides these significant changes in the weights of the non-benchmark positions, the weight of the benchmark position SMB also increases significantly (27.3%). The intuition behind this is as follows: Investors choose substantial nonzero allocations in SMB since it offers diversification, though it has a significantly lower mean than the other two benchmark positions. Moreover, the variances of the optimal portfolios are reduced at the costs of their expected returns under DGP uncertainty. In other words, investors invest less aggressively, favoring lower variance portfolios in spite of lower expected returns. Therefore, SMB becomes more attractive than under the normality assumption since it offers a greater diversification benefit. As a result, investors are willing to hold more of SMB than the normal case while reducing the weights of other positions to balance the budget. The changes of the

other two benchmark positions are also large, though not as much so as for the SMB.

As reported in Table 6, the results are more or less similar for other pricing models with varying degrees of mispricing uncertainty and varying margin requirements. Overall, there are some substantial differences in the optimal allocations under the normal assumption versus DGP uncertainty, but these differences do not occur simultaneously for all of the assets, but rather occur only for a few of them (with small differences present in others). However, portfolio performance cannot be evaluated simply based on the portfolio weights alone. Because of the correlations among the payoffs of risky positions, the performance of two portfolios can be similar even though the portfolios are quite different in position-by-position allocations. Therefore, we turn our attention now to the performance measure discussed earlier in Section 2.4.

3.4. Performance measure

We examine now the certainty–equivalent loss for the optimal portfolio choice under the normality assumption as compared with the optimal portfolio choice under DGP uncertainty. With varying pricing models, varying degrees of pricing errors, and varying margin requirements, we assume, as before, that the mean-variance optimizing investor has a relative risk aversion equal to 2.83. Table 7 reports the annualized percentage points of $EU^t - EU^n$, referred to as certainty–equivalent gains in the table. Again, certainty–equivalent gains are the "perceived" gains of utilizing DGP uncertainty, or the amount of loss to the investor when she is forced to hold the portfolio that is optimal under the normality assumption.

It is seen that with a 50% margin requirement ($c = 2$), the maximum loss is only 0.54% across all the models and mispricing errors. This is clearly a small amount per year. When the margin requirement is relaxed, however, in general the loss becomes greater. Still, under the dogmatic belief (the belief of zero pricing errors), the largest loss is only 0.63% which occurs for Daniel and Titman's (1997) characteristic-based model. With a reasonable 3% model mispricing error, the maximum loss increases to 0.75%, again a small number. At a 10% model mispricing error, the maximum loss is still below 1% for all margin levels. Indeed, the maximum loss increases only to 1.06% (barely greater than 1%) when investors believe that the models are completely useless. Moreover, once a margin of 50% is imposed, the maximum loss is only 0.54%. Hence, the utility loss is economically small and negligible when the investment universe consists of Pástor and Stambaugh's (2000) nine spread positions and Fama and French's (1993) three factors, for the monthly data from July 1963 through December 1997. These results show that Pástor and Stambaugh's (2000) Bayesian certainty–equivalent analysis is surprisingly robust to normality. That is, although the normality assumption is unequivocally rejected by the data, the expected utility achieved, based on a much more reasonable DGP, is not that much different from that obtained under the normality assumption.

As the performance measure is an important summary, it is of interest to examine its sensitivity to the use of alternative test assets. For this purpose, we apply the

Table 7
Certainty-equivalent gains with the Pástor and Stambaugh's (2000) portfolios

The table reports the annualized certainty–equivalent gains (in percentage points) for a mean-variance optimizing investor with relative risk aversion equal to 2.83 under the DGP uncertainty versus under normality when the investment universe consists of Pástor and Stambaugh's (2000) nine spread positions and Fama and French's (1993) three factors. That is, the table provides $EU^t - EU^n$ for varying pricing models, degrees of mispricing priors and margin requirements, where EU^t is the maximized certainty–equivalent return under DGP and EU^n is the maximized certainty–equivalent return under normality.

	$\sigma_\alpha = 0$	$\sigma_\alpha = 1\%$	$\sigma_\alpha = 2\%$	$\sigma_\alpha = 3\%$	$\sigma_\alpha = 6\%$	$\sigma_\alpha = 10\%$	$\sigma_\alpha = \infty$
$c = 2$							
DT	0.01	0.01	0.01	0.15	0.38	0.38	0.40
FF	0.03	0.03	0.03	0.05	0.32	0.38	0.40
CM	0.04	0.05	0.13	0.22	0.47	0.54	0.41
$c = 5$							
DT	0.02	0.07	0.27	0.31	0.34	0.35	0.37
FF	0.01	0.02	0.13	0.26	0.30	0.33	0.29
CM	0.04	0.07	0.16	0.21	0.05	0.38	0.37
$c = 10$							
DT	0.22	0.16	0.21	0.52	0.56	0.56	0.51
FF	0.08	0.10	0.21	0.37	0.49	0.50	0.58
CM	0.04	0.10	0.23	0.39	0.49	0.64	0.60
$c = \infty$							
DT	0.63	0.46	0.65	0.75	0.89	0.98	1.00
FF	0.08	0.14	0.36	0.58	0.83	0.90	1.00
CM	0.04	0.10	0.33	0.57	0.85	0.97	1.06

above analysis to the 20 industry portfolios. Panel A of Table 8 provides the results. In general, the losses for the industry portfolios are larger. With a 50% margin requirement, the maximum loss is 0.72% across models and mispricing errors. Notice that the maximum loss is still 0.72% if the model mispricing error does not exceed 2%, regardless of margin requirements. Now, if we impose a fairly relaxed margin of $c = 10$ and allow the mispricing error to be as large as 3%, the maximum loss shrinks to 0.56%. Hence, results for the industry returns still support the robustness of the certainty–equivalent analysis to the normality assumption. This is also the case with Fama and French's (1993) 25 assets (results omitted here).

Similar to the normality tests, one can ask how the robustness result may change with trading frequency. Apart from the monthly frequency, the weekly frequency is of the greatest interest for examination. Panel B of Table 8 reports the certainty–equivalent gains for weekly industry returns. The losses are now clearly much greater than those obtained from the monthly frequency. This might be expected because the weekly data deviates much more from normality than the monthly series. However, after imposing a 50% margin requirement, the losses are capped at 1.52%. The maximum loss across all the scenarios can be as high as 4.56%, but such a large value

Table 8

Certainty-equivalent gains with industry returns

The table reports the annualized certainty–equivalent gains (in percentage points) for a mean-variance optimizing investor with relative risk aversion equal to 2.83 under the DGP uncertainty versus under normality when the investment universe consists of 20 industry portfolios and Fama and French's (1993) three factors. That is, the table provides $EU^t - EU^n$ for varying pricing models, degrees of mispricing priors and margin requirements, where EU^t is the maximized certainty–equivalent return under DGP and EU^n is the maximized certainty–equivalent return under normality.

	$\sigma_\alpha = 0$	$\sigma_\alpha = 1\%$	$\sigma_\alpha = 2\%$	$\sigma_\alpha = 3\%$	$\sigma_\alpha = 6\%$	$\sigma_\alpha = 10\%$	$\sigma_\alpha = \infty$
Monthly							
$c = 2$							
FF	0.42	0.08	0.07	0.16	0.51	0.60	0.59
CM	0.04	0.07	0.24	0.35	0.53	0.68	0.72
$c = 5$							
FF	0.11	0.08	0.14	0.32	0.54	0.66	0.70
CM	0.04	0.09	0.27	0.46	0.65	0.70	0.81
$c = 10$							
FF	0.20	0.25	0.36	0.47	0.73	0.82	0.92
CM	0.02	0.14	0.34	0.56	0.77	0.96	1.03
$c = \infty$							
FF	0.21	0.30	0.72	1.22	1.91	2.17	2.28
CM	0.04	0.15	0.71	1.29	2.15	2.45	2.64
Weekly							
$c = 2$							
FF	0.13	1.12	1.45	1.40	1.48	1.45	1.52
CM	0.94	1.45	1.50	1.42	1.43	1.52	1.52
$c = 5$							
FF	0.94	1.78	1.89	1.83	1.91	1.83	1.77
CM	0.94	1.70	2.05	1.96	1.89	2.06	1.92
$c = 10$							
FF	2.11	1.95	2.57	2.38	2.23	2.43	2.46
CM	0.94	1.75	2.53	2.27	2.64	2.34	2.35
$c = \infty$							
FF	2.28	3.80	4.17	4.22	4.28	4.27	4.27
CM	0.93	3.67	4.16	4.35	4.44	4.56	4.30

is unrealistic as it occurs when there are no margin requirements. In addition, the issue of transaction costs is likely to become more important for rebalancing portfolios weekly. Given these considerations, the significance of the large losses should be greatly discounted for cases in which the margin is less than 50%. Indeed, based on the results associated with a reasonable 50% margin, we do not see strong evidence against the robustness result, even at the weekly frequency.

Finally, we comment on the statistical reason behind the results as well as on future research. In contrast with the huge *p*-value gains going from the normal to *t* distribution, the gains in terms of portfolio performance are small, as shown above. This is because the normality tests examine the tail properties of the distribution through skewness and kurtosis, whereas the performance measure focuses mostly on the average behavior, that is, the global properties of the data. As a result, an enormously strong rejection of the normality assumption for the DGP is transformed into no rejection of the normal model in terms of the performance measure. This intuition also explains why asset pricing tests can make a huge difference if a *t* distributional assumption is used instead of the normal. Indeed, Zhou (1993) shows that while the Gibbons et al. (1989) test rejects the CAPM strongly with the normality assumption, it can no longer reject the model when a suitable *t* distributional assumption is made. This is because, while the test depends on the aggregate behavior of the data, it is the tail property of the test that determines the *p*-value. Clearly, this suggests that any financial applications involving the tail properties, such as risk management, should be much more sensitive to the underlying assumption on the DGP than the portfolio decision should. In particular, future research may apply the tools here for value-at-risk evaluation in the mean-variance framework (see Jorion, 2001). In addition, it is noteworthy that the method proposed in this paper is particularly useful for analyzing large-dimensional problems. In contrast, the Bayesian non-normality approach recently proposed by Eraker, Johannes, and Polson (2003) is limited to only a few variables though it is a nice framework for examining returns with possible jumps. In fact, wherever multivariate normality is assumed, our approach can be easily applied by replacing the normal DGP with distributions of the *t* class to possibly better fit the data and to assess the robustness of the model to the normality assumption.

4. Conclusion

Many important empirical questions in finance are studied under a multivariate normal DGP, though the normality assumption is firmly rejected by statistical tests. Although the generalized method of moments (GMM) of Hansen (1982), one of the most widely used classical methods, allows for a much more general distributional assumption than the normal, its estimates of important parameters such as the expected asset returns, alphas, and betas in the standard linear factor regression model are the same as if the data were assumed normal, except that it enlarges the standard errors to account for non-normality. The modeling of multivariate non-normality is a nontrivial task, as evidenced by the fact that the well-known GARCH type framework is difficult to implement in practice for high-dimensional problems. Although the alternative Bayesian paradigm, as applied by Pástor and Stambaugh (1999, 2000), Barberis (2000), Wang (2001), Lewellen and Shanken (2002), Cremers (2002), Avramov (2002), Avramov and Chao (2002) and many others, provides

profound insights into a variety of financial issues, none of these Bayesian studies allows for non-normality of the data.

This paper makes a first Bayesian attempt to model multivariate non-normality of the data in a high-dimensional portfolio allocation problem. Replacing the normal data-generating process by a set of 31 plausible t distributions which range from those close to normal to those far away from normal, we let the data tell which distribution is most likely the true DGP. Based on Pástor and Stambaugh (2000), we provide a framework to account for parameter uncertainty, mispricing uncertainty, and DGP uncertainty in investment decisions. Once DGP uncertainty is introduced, we find that both the parameter estimates and the optimal portfolio weights can be substantially different from those obtained under normality. However, in terms of Pástor and Stambaugh's (2000) certainty–equivalent (CE) measure, the loss to an investor who is forced to hold the portfolio that is optimal under normality is small, suggesting that the CE measure is robust to the normality assumption. From a wider perspective, our proposed framework applies not only to portfolio choice problems, but also to many existing studies, to examine the robustness of their conclusions to the normality assumption. In addition, this framework is particularly useful for analyzing problems in which the tail distributions are important, such as risk management and option pricing, where the usual normality assumption can be replaced by t distributions to better model the tail distribution of the data.

Appendix A. posterior evaluation

A.1. *Predictive distribution in risk-based models*

Define $Y = (y_1, \dots, y_T)'$, a $T \times m$ matrix, $X = (x_1, \dots, x_T)'$, a $T \times k$ matrix, and $Z = (\iota_T \ X)$, a $T \times (k+1)$ matrix, where ι_T denotes a T-vector of ones. Also define $A = (\alpha, B)'$, a $(k+1) \times m$ matrix and $a = vec(A)$. Then the regression model (11) can be written as

$$Y = ZA + U, \tag{A.1}$$

where $U = (u_1, \dots, u_T)'$. Unlike the case of the normal distribution, the likelihood function of the t distribution does not combine with the prior to yield analytical posterior distributions in terms of a few summary statistics of the data. Fortunately, with the Bayesian technique of data augmentation, this difficulty can be resolved. By augmenting the data space from (r_t) to (r_t, w_t), where w_t follows a gamma distribution

$$w_t \propto G\left(\frac{v}{2}, \frac{v}{2}\right), \tag{A.2}$$

we can decompose a t distribution into a product of a normal and a gamma (see, e.g., Fang et al. (1990) for a proof), with $\Omega = (v-2)V/v$,

$$p(y_t, \ x_t \mid E, \ \Omega, \ v) = \int p(y_t, \ x_t \mid E, \ \Omega, \ v, \ w_t) p(w_t|v) \, dw_t. \tag{A.3}$$

Thus, conditioning on w_t, r_t follows a multivariate normal distribution and the regression model is essentially a weighted version of the standard one analyzed by Harvey and Zhou (1990). Therefore, a multivariate t distribution can be viewed as a mixture of an infinite number of multivariate normal distributions such that the density of w_t determines the weights. In the classical framework, the EM algorithm is a similar data augmentation approach, for which Kan and Zhou (2003) provide the associated asymptotic theory on parameter estimation.

Define

$$w = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_T \end{pmatrix}. \tag{A.4}$$

Conditioning on the augmented data, the likelihood function can be factored as

$$p(Y, \ X \,|\, E, \ \Omega, \ v, \ w) = p(Y \,|\, A, \ \Sigma, \ w, \ X) p(X \,|\, E_2, \ \Omega_{22}, \ w), \tag{A.5}$$

where, letting $\Sigma = \Omega_{11} - \Omega_{12} \, \Omega_{22}^{-1} \, \Omega_{21}$,

$$\begin{aligned} p(Y \,|\, A, \ \Sigma, \ w, \ X) &\propto |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr[(Y_w - Z_w A)'(Y_w - Z_w A)\Sigma^{-1}] \right\} \\ &\propto |\Sigma|^{-T/2} \exp\left\{ -\frac{T}{2} tr\hat{\Sigma}_w \Sigma^{-1} - \frac{1}{2} tr((A - \hat{A}_w)' Z_w' Z_w (A - \hat{A}_w)\Sigma^{-1}) \right\} \\ &\propto |\Sigma|^{-T/2} \exp\left\{ -\frac{T}{2} tr\hat{\Sigma}_w \Sigma^{-1} - \frac{1}{2} tr[(a - \hat{a}_w)'(\Sigma^{-1} \otimes Z_w' Z_w)(a - \hat{a}_w)] \right\} \end{aligned} \tag{A.6}$$

and

$$\begin{aligned} p(X \,|\, E_2, \ \Omega_{22}, \ w) &\propto |\Omega_{22}|^{-T/2} \exp\left\{ -\frac{1}{2} tr((X - \iota_T E_2')' w(X - \iota_T E_2'))\Omega_{22}^{-1} \right\} \\ &\propto |\Omega_{22}|^{-T/2} \exp\left\{ -\frac{T}{2} tr\widehat{\Omega_{22w}}\Omega_{22}^{-1} - \frac{\sum_{t=1}^{T} w_t}{2} tr((E_2 - \widehat{E_{2w}})(E_2 - \widehat{E_{2w}})'\Omega_{22}^{-1}) \right\}, \end{aligned} \tag{A.7}$$

with $Y_w = w^{1/2} Y$, $Z_w = w^{1/2} Z$, $\hat{a}_w = vec(\hat{A}_w)$, and

$$\hat{A}_w = (Z_w' Z_w)^{-1} Z_w' Y_w = (Z' w Z)^{-1} Z' w Y,$$

$$\hat{\Sigma}_w = (Y_w - Z_w \hat{A}_w)'(Y_w - Z_w \hat{A}_w)/T = (Y - Z\hat{A}_w)' w(Y - Z\hat{A}_w)/T,$$

$$\hat{E}_{2w} = (X' w \iota_T)/\iota_T' w \iota_T$$

and

$$\widehat{\Omega_{22w}} = (X - \iota_T \widehat{E_{2w}})' w(X - \iota_T \widehat{E_{2w}})/T.$$

The joint prior distribution of all parameters is

$$p_0(\theta) = p_0(\alpha|\Sigma)p_0(\Sigma)p_0(B)p_0(E_2)p_0(\Omega_{22})p_0(v), \tag{A.8}$$

where

$$p_0(\alpha|\Sigma) \propto |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}\alpha'\left(\frac{\sigma_\alpha^2}{s^2}\Sigma\right)^{-1}\alpha\right\}, \tag{A.9}$$

$$p_0(\Sigma) \propto |\Sigma|^{-(v_\Sigma+m+1)/2} \exp\left\{-\frac{1}{2}trH\Sigma^{-1}\right\}, \tag{A.10}$$

$$p_0(B) \propto 1, \tag{A.11}$$

$$p_0(E_2) \propto 1, \tag{A.12}$$

$$p_0(\Omega_{22}) \propto |\Omega_{22}|^{-(k+1)/2}, \tag{A.13}$$

and

$$p_0(v) = \frac{1}{|S_v|}, \tag{A.14}$$

$|S_v|$ is defined as in Section 2.3, $H = s^2(v_\Sigma - m - 1)I_m$, $v_\Sigma = 15$, $s^2 = tr((Y - Z\hat{A})'$ $(Y - Z\hat{A})/T)/m$, and $\hat{A} = (Z'Z)Z'Y$. In addition, consider the transformation

$$\alpha'\left(\frac{\sigma_\alpha^2}{s^2}\Sigma\right)^{-1}\alpha = a'(\Sigma^{-1} \otimes D)a, \tag{A.15}$$

where $a = \text{vec}(A)$ and D is a $(k+1) \times (k+1)$ matrix whose $(1,1)$ element is s^2/σ_α^2 and whose other elements are all zero. Then, it follows that the likelihood in (A.5)–(A.7) can be combined with the prior in (A.9)–(A.14) to obtain the posterior distribution

$$p(E, \Omega, v|R) \propto p(R|E, \Omega, v)p_0(E, \Omega, v).$$

Hence, the joint posterior of the regression parameters is

$$p(a, \Sigma \,|\, w, R) \propto |\Sigma|^{-(k+1)/2} \exp\left\{-\frac{1}{2}a'(\Sigma^{-1} \otimes D)a - \frac{1}{2}tr((a - \hat{a}_w)'(\Sigma^{-1} \otimes Z'_w Z_w)(a - \hat{a}_w))\right\}$$
$$\times |\Sigma|^{-(T+v_\Sigma+m-k+1)/2} \exp\left\{-\frac{1}{2}tr(H + T\hat{\Sigma}_w)\Sigma^{-1}\right\}. \tag{A.16}$$

Let

$$F_w = D + Z'_w Z_w = D + Z'wZ,$$

and

$$Q_w = Z'wZ - Z'wZF_w^{-1}Z'wZ.$$

By completing the square on a, we have

$$p(a, \Sigma \,|\, w, R) \propto |\Sigma|^{-(k+1)/2} \exp\left\{-\frac{1}{2}[(a - \tilde{a}_w)'(\Sigma^{-1} \otimes F_w)(a - \tilde{a}_w)]\right\}$$
$$\times |\Sigma|^{-(T+v_\Sigma+m-k+1)/2} \exp\left\{-\frac{1}{2}tr(H + T\hat{\Sigma}_w + \hat{A}'_w Q_w \hat{A}_w)\Sigma^{-1}\right\}, \tag{A.17}$$

where $\tilde{a}_w = (I_m \otimes F_w^{-1} Z' w Z)\hat{a}_w$. Hence,

$$\Sigma^{-1} \mid R, \ w \sim W(T + v_\Sigma - k, \ (H + T\hat{\Sigma}_w + \hat{A}'_w Q_w \hat{A}_w)^{-1}), \tag{A.18}$$

and

$$a \mid \Sigma^{-1}, \ R, \ w \sim N(\tilde{a}_w, \ \Sigma \otimes F_w^{-1}). \tag{A.19}$$

The joint posterior distribution of E_2 and Ω_{22} is

$$p(E_2, \ \Omega_{22}|w, \ R) \propto |\Omega_{22}|^{-(T+k+1)/2} \exp\left\{ -\frac{T}{2} tr \ \widehat{\Omega_{22w}} \Omega_{22}^{-1} - \frac{\sum_{t=1}^T w_t}{2} tr((E_2 - \widehat{E_{2w}}) \right.$$

$$\left. \times (E_2 - \widehat{E_{2w}})' \Omega_{22}^{-1}) \right\}. \tag{A.20}$$

As a result, we have

$$\Omega_{22}^{-1} \mid R, \ w \sim W(T - 1, \ (T\widehat{\Omega_{22w}})^{-1}) \tag{A.21}$$

and

$$E_2 \mid \Omega_{22}^{-1}, \ R, \ w \sim N\left(\widehat{E_{2w}}, \ \left(\sum_{t=1}^T w_t \right)^{-1} \Omega_{22} \right). \tag{A.22}$$

Both the likelihood function conditioning on the augmented data (R, w) and the prior can be factored into two independent parts on (a, Σ) and (E_2, Ω_{22}), respectively. However, after w is marginalized out, the posteriors on (a, Σ) and (E_2, Ω_{22}) are no longer independent, as under the common multivariate normal distribution.

To carry out the posterior evaluation, we draw a sample from the joint posterior distribution of w and θ according to their joint posterior distribution. Ignoring the w part of this sample, the θ part of the sample is also a sample drawn from its marginal posterior distribution. The following steps show how to implement this idea:

(1) draw v from the prior $p_0(v)$,
(2) draw iid series w_t from gamma $(v/2, v/2)$,
(3) $\Sigma^{-1} \mid R, \ w \propto W(T + v_\Sigma - k, (H + T\hat{\Sigma}_w + \hat{A}'_w Q_w \hat{A}_w)^{-1})$,
(4) $a \mid \Sigma^{-1}, \ R, \ w \propto N(\tilde{a}_w, \ \Sigma \otimes F_w^{-1})$,
(5) $\Omega_{22}^{-1} \mid R, \ w \propto W(T - 1, (T\widehat{\Omega_{22w}})^{-1})$,
(6) $E_2 \mid \Omega_{22}^{-1}, \ R, \ w \propto N(\widehat{E_{2w}}, (\sum_{t=1}^T w_t)^{-1}\Omega_{22})$,
(7) draw v from the posterior: $p(v| \ w) \propto p_0(v)\Pi_{t=1}^T \frac{(\frac{v}{2})^{v/2}}{\Gamma(\frac{v}{2})} w_t^{(v/2)-1} \exp\left(-\frac{v}{2}w_t\right)$,
(8) draw iid series w_t from gamma

$$\left(\frac{v + m + k}{2}, \frac{v + (y_t - z_t A)' \Sigma^{-1}(y_t - z_t A) + (x_t - E_2)' \Omega_{22}^{-1}(x_t - E_2)}{2} \right),$$

(9) Repeat steps (3)–(8).

We can, following Geweke and Zhou (1996), start the above Gibbs sampling procedure from any arbitrary initial value in the support of the posterior density.

Based on Appendix A.2, it is easy to see that each of the 31 values in S_v has a nonzero posterior probability (though some of the posterior probabilities can be very small). Therefore, we can start the Gibbs sampling procedure by drawing v from the prior $p_0(v)$ first. Let $g = M + Q$ denote the total number of iterations of the above loop. Disregarding the first M draws of the burning period, we obtain the mean vector and covariance matrix of the predictive distribution using the other Q draws as follows. From (13), we first obtain the mean vector

$$E^* = \mathrm{E}(\alpha + BE_2, E_2)'$$ (A.23)

and then compute the numerical approximations as:

$$\mathrm{E}(\alpha + BE_2) = \frac{1}{Q} \Sigma_{i=1}^{Q}(\alpha^i + B^i E_2^i)$$ (A.24)

and

$$\mathrm{E}(E_2) = \frac{1}{Q} \Sigma_{i=1}^{Q} E_2^i.$$ (A.25)

Secondly, we obtain the covariance matrix,

$$V^* = Var(r_{T+1}|R) = \mathrm{E}(Var(r_{T+1}|\theta, R)|R) + Var(\mathrm{E}(r_{T+1}|\theta, R)|R)$$
$$= \mathrm{E}\left(\frac{v}{v-2}\Omega|R\right) + Var(E|R),$$ (A.26)

and compute the numerical approximations as

$$V^* = \frac{1}{Q} \Sigma_{i=1}^{Q}\left(\frac{v^i}{v^i - 2}\Omega^i\right) + \frac{1}{Q} \Sigma_{i=1}^{Q}(E^i - \bar{E})^2,$$ (A.27)

where

$$\Omega^i = \begin{pmatrix} \Omega_{11}^i & \Omega_{12}^i \\ \Omega_{21}^i & \Omega_{22}^i \end{pmatrix}$$ (A.28)

with $\Omega_{11}^i = \Sigma^i + B^i \Omega_{22}^i (B^i)'$, $\Omega_{12}^i = B^i \Omega_{22}^i$, and $\bar{E} = 1/Q \Sigma_{i=1}^{Q} E^i$.

A.2. Predictive distribution in the characteristic-based model

The likelihood function and the prior on $(B, \Sigma, E_2, \Omega_{22}, v)$ are the same as in the factor-based models presented in Section A.1. The only difference here is the prior for α,

$$p(\alpha|B, E_2) \propto \exp\left\{ -\frac{1}{2\sigma_\alpha^2}(\alpha + BE_2)'(\alpha + BE_2) \right\}.$$

Hence, the conditional prior on α is normal and centered at the pricing restriction. To transform this prior on (α, B) to a prior on $A = (\alpha, B)$ or

$a = vec(A)$, note that

$$\frac{1}{\sigma_\alpha^2}(\alpha + BE_2)'(\alpha + BE_2) = \frac{1}{\sigma_\alpha^2}(1\ E_2')\begin{pmatrix} \alpha' \\ B' \end{pmatrix}(\alpha, B)\begin{pmatrix} 1 \\ E_2 \end{pmatrix}$$

$$= \frac{1}{\sigma_\alpha^2}(1\ E_2')AA'\begin{pmatrix} 1 \\ E_2 \end{pmatrix}$$

$$= a'\left(\frac{1}{\sigma_\alpha^2}I_m \otimes \Psi\right)a, \tag{A.29}$$

where

$$\Psi = \begin{pmatrix} 1 & E_2' \\ E_2 & E_2 E_2' \end{pmatrix}.$$

The full conditional posterior distribution of a is

$$p(a|\Sigma,\ E_2,\ R, w) \propto \exp\left\{-\frac{1}{2}\left(a'\left(\frac{1}{\sigma_\alpha^2}I_m \otimes \Psi\right)a + ((a - \hat{a}_w)'(\Sigma^{-1} \otimes Z_w'Z_w)(a - \hat{a}_w))\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(a - \bar{a}_w)G_w(a - \bar{a}_w)'\right\}, \tag{A.30}$$

where

$$G_w = \left(\frac{1}{\sigma_\alpha^2}I_m \otimes \Psi\right) + (\Sigma^{-1} \otimes Z_w'Z_w),$$

and

$$\bar{a}_w = G_w^{-1}(\Sigma^{-1} \otimes Z_w'Z_w)\hat{a}_w.$$

Hence, the full conditional posterior distribution of a is a normal distribution,

$$a\,|\,\Sigma,\ E_2,\ R, w \sim N(\bar{a}_w,\ G_w^{-1}).$$

The full conditional posterior distribution of Σ is

$$p(\Sigma|a, R, w) \propto |\Sigma|^{-(T+v_\Sigma+m+1)/2} \exp\left\{-\frac{1}{2}tr[(Y_w - Z_wA)'(Y_w - Z_wA) + H]\Sigma^{-1}\right\},$$

which follows a usual inverted Wishart distribution,

$$\Sigma\,|\,a, R, w \sim W(T + v_\Sigma,\ [(Y_w - Z_wA)'(Y_w - Z_wA) + H]^{-1}).$$

The full conditional posterior distribution of E_2 is

$$p(E_2|B, \Omega_{22}, R, w) \propto \exp\left\{-\frac{1}{2}[E_2'B'BE_2 + 2E_2'B'\alpha + tr\,E_2\iota_T'w\iota_T E_2'\Omega_{22}^{-1}\right.$$

$$\left. - 2E_2'\Omega_{22}^{-1}X'w\iota_T]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(E_2 - \bar{E}_{2w})'P_w(E_2 - \bar{E}_{2w})\right\}, \tag{A.31}$$

where

$$P_w = \frac{B'B}{\sigma_\alpha^2} + \left(\sum_{t=1}^{T} w_t \right) \Omega_{22}^{-1},$$

$$\bar{E}_{2w} = P_w^{-1} \left(\left(\sum_{t=1}^{T} w_t \right) \Omega_{22}^{-1} \hat{E}_{2w}^{-1} - \frac{B'\alpha}{\sigma_\alpha^2} \right).$$

Then, the full conditional posterior distribution of E_2 is a normal distribution,

$$E_2 | \alpha, B, \Omega_{22}, R, w \sim N(\bar{E}_{2w}, \ P_w^{-1}).$$

Finally, the full conditional posterior distribution of Ω_{22} is

$$p(\Omega_{22} | E_2, R, w) | \propto |\Omega_{22}|^{-(T+k+1)/2} \exp \left\{ -\frac{1}{2} tr((X - \iota_T E_2')' w (X - \iota_T E_2')) \Omega_{22}^{-1} \right\},$$

which implies an inverted Wishart distribution,

$$\Omega_{22}^{-1} | E_2, R, w \sim W(T, [(X - \iota_T E_2')' w (X - \iota_T E_2')]^{-1}).$$

Based on the above, E^* and V^* are obtained by implementing the sampling procedure in Section A.1 by replacing the full conditional posteriors with their counterparts in this section.

A.3. The posterior density of v

The posterior density of the degrees of freedom, $\pi(v|R)$, is the marginal density of the joint posterior density $\pi(E, \Omega, v|R)$,

$$\pi(v|R) = \int_\Omega \int_E \pi(E, \Omega, v|R) \, \mathrm{d}E \, \mathrm{d}\Omega. \tag{A.32}$$

Then the joint posterior density is

$$\pi(E, \Omega, v|R) \propto \Pi_{t=1}^{T} p(y_t, \ x_t | E, \Omega, v) p_0(E, \Omega, v)$$
$$= p_0(E, \ \Omega, \ v) \int_{w_T} \cdots \int_{w_1} (\Pi_{t=1}^{T} p(y_t, \ x_t | E, \ \Omega, \ v, \ w_t) p(w_t|v))$$
$$\times \mathrm{d}w_1 \, \mathrm{d}w_2 \cdots \mathrm{d}w_T. \tag{A.33}$$

Therefore,

$$\pi(v|R) \propto p_0(v) \int_{w_T} \cdots \int_{w_1} \left(\int_\Omega \int_E (\Pi_{t=1}^{T} p(y_t, \ x_t | E, \ \Omega, \ v, \ w_t)) p_0(E, \ \Omega) \, \mathrm{d}E \, \mathrm{d}\Omega \right)$$
$$\times p(w_t|v) \, \mathrm{d}w_1 \, \mathrm{d}w_2 \cdots \mathrm{d}w_T. \tag{A.34}$$

Let

$$Q([w_t]_{t=1}^{T}) = \int_\Omega \int_E (\Pi_{t=1}^{T} p(y_t, \ x_t | E, \ \Omega, \ v, \ w_t)) p_0(E, \ \Omega) \, \mathrm{d}E \, \mathrm{d}\Omega, \tag{A.35}$$

and

$$J(v) = \int_{w_T} \cdots \int_{w_1} Q([w_t]_{t=1}^T) p(w_t|v) \, \mathrm{d}w_1 \, \mathrm{d}w_2 \cdots \mathrm{d}w_T. \qquad (A.36)$$

Then

$$\pi(S_v(i)|R) = \frac{p_0(S_v(i))J(S_v(i))}{\sum_{i=1}^{|S_v|} (p_0(S_v(i))J(S_v(i)))}, \quad i = 1, 2, \ldots, |S_v|. \qquad (A.37)$$

References

Affleck-Graves, J., McDonald, B., 1989. Nonnormalities and tests of asset pricing theories. Journal of Finance 44, 889–908.

Avramov, D., 2002. Stock return predictability and model uncertainty. Journal of Financial Economics 64, 423–458.

Avramov, D., Chao, J., 2002. An exact Bayes test of asset pricing with application to international markets. University of Maryland, Working paper.

Barberis, N., 2000. Investing for the long run when returns are predictable. Journal of Finance 55, 225–264.

Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: what do we learn? Review of Financial Studies 12, 405–428.

Branco, M.D., Dey, D.K., 2001. A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis 79, 99–113.

Brennan, M., Xia, Y., 2001. Assessing asset pricing anomalies. Review of Financial Studies 14, 905–942.

Chamberlain, G., 1983. A characterization of the distributions that imply mean-variance utility functions. Journal of Economic Theory 29, 185–201.

Cremers, M., 2001. Multifactor efficiency and Bayesian inference. New York University, Working paper.

Cremers, M., 2002. Stock return predictability: a Bayesian model selection perspective. Review of Financial Studies 15, 1223–1249.

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

Eraker, B., Johannes, M.J., Polson, N.G., 2003. The impact of jumps in returns and volatility. Journal of Finance 58, 1269–1300.

Fama, E.F., 1965. The behavior of stock market prices. Journal of Business 38, 34–105.

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

Fang, K., Kotz, S., Ng, K., 1990. Symmetric multivariate and related distributions. Chapman and Hall, New York.

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56, 329–352.

Geczy, C., 2001. Some generalized tests of mean-variance efficiency and multifactor model performance. University of Pennsylvania, Working paper.

Geweke, J., Zhou, G., 1996. Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies 9, 557–587.

Ghysels, E., 1998. On stable factor structures in the pricing of risk: do time-varying betas help or hurt? Journal of Finance 53, 549–573.

Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Han, Y., 2002. Can an investor benefit from return predictability? Washington University, St. Louis, Working paper.

Hansen, L.P., 1982. Large sample properties of the generalized method of moments estimators. Econometrica 50, 1029–1054.

Harvey, C.R., Zhou, G., 1990. Bayesian inference in asset pricing tests. Journal of Financial Economics 26, 221–254.

Jorion, P., 2001. Value at risk: the new benchmark for managing financial risk. McGraw-Hill, New York.

Kan, R., Zhou, G., 2003. Modeling non-normality using multivariate *t*: implications for asset pricing. Washington University, St. Louis, Working paper.

Kandel, S., Stambaugh, R.F., 1996. On the predictability of stock returns: an asset-allocation perspective. Journal of Finance 51, 385–424.

King, B., 1966. Market and industry factors in stock price behavior. Journal of Business 39, 139–190.

Lamoureux, C., Zhou, G., 1996. Temporary components of stock returns: what do the data tell us? Review of Financial Studies 9, 1033–1059.

Lewellen, J., Shanken, G., 2002. Learning, asset pricing tests, and market efficiency. Journal of Finance 57, 1113–1145.

MacKinlay, A.C., Pástor, Ľ., 2000. Asset pricing models: implications for expected returns and portfolio selection. Review of Financial Studies 13, 883–916.

MacKinlay, A.C., Richardson, M., 1991. Using generalized method of moments to test mean-variance efficiency. Journal of Finance 46, 511–527.

Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530.

Moskowitz, T., Grinblatt, M., 1999. Do industries explain momentum? Journal of Finance 54, 1249–1290.

Owen, J., Rabinovitch, R., 1983. On the class of elliptical distributions and their applications to the theory of portfolio choice. Journal of Finance 38, 745–752.

Pástor, Ľ., 2000. Portfolio selection and asset pricing models. Journal of Finance 55, 179–223.

Pástor, Ľ., Stambaugh, R.F., 1999. Costs of equity capital and model mispricing. Journal of Finance 54, 67–121.

Pástor, Ľ., Stambaugh, R.F., 2000. Comparing asset pricing models: an investment perspective. Journal of Financial Economics 56, 335–381.

Richardson, M., Smith, T., 1993. A test for multivariate normality in stock returns. Journal of Business 66, 295–321.

Wang, Z., 2001. A shrinkage theory of asset allocation using asset pricing models. Columbia University, Working paper.

Zellner, A., 1976. Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. Journal of the American Statistical Association 71, 400–405.

Zellner, A., Chetty, V.K., 1965. Prediction and decision problems in regression models from the Bayesian point of view. Journal of the American Statistical Association 60, 608–616.

Zhou, G., 1993. Asset pricing tests under alternative distributions. Journal of Finance 48, 1927–1942.