Valuation ratios and shape predictability in the distribution of stock returns *

Paolo Giordani † Michael Halling‡

ABSTRACT

While a large literature on return predictability has shown a link between valuation levels and expected rates of aggregate returns in-sample, we document a link between valuation levels and the shape of the distribution of cumulative (for example, over 12 and 24 months) total returns. Return distributions become more asymmetric and negatively skewed when valuation levels are high. In contrast, they are roughly symmetric when valuation levels are low. These results turn out to be very robust to alternative (a) measures of valuation levels, (b) model specifications and (c) equity markets (international and industry-level). Importantly, these findings shed light on how equity prices regress back to their means conditional on valuation levels, have important practical implications for risk measurement and asset management, and refine the well-known finding of negative skewness in aggregate returns. The model with conditional skewness also outperforms benchmark models assuming a symmetric or constant-skewness distribution in an out-of-sample setup. Our empirical results support theoretical asset pricing models that have asymmetric responses to shocks, such as stochastic bubbles, liquidity spirals or models with time-varying risk aversion.

JEL Classifications: G12, G17, C22.

Keywords: return predictability, valuation ratios, skewness.

*Acknowledgments: We are grateful to Jules van Binsbergen, John Cochrane, Magnus Dahlquist, Raymond Kan (discussant), Jan-Peter Krahnen, Sydney Ludvigson, Lasse Pedersen, Otto Randl, Christian Schlag, Paul Söderlind, Pietro Veronesi, Grigori Vilkov, Jessica Wachter, Amir Yaron, conference participants at the NFA (2019), and seminar participants at Aarhus University, BI Oslo, Cass Business School, Copenhagen Business School, Durham University, Frankfurt School of Finance and Management, Goethe University Frankfurt, Örebro, Stockholm School of Economics, the Swedish Central Bank, the Trinity Webinar, the University of St. Gallen (econometrics as well as finance) and the Vienna University of Business and Economics. All errors are our responsibility.

†Email: paolo.giordani@bi.no. BI Norwegian Business School.

‡Corresponding author. Email: michael.halling@hhs.se. Stockholm School of Economics and Swedish House of Finance, Drottninggatan 98, 111 60 Stockholm.

1 Introduction

A large literature has looked at the time-variation in expected rates of returns and the link between prices, dividends and discount rates (see, for example, Cochrane (2011) and Fama (2013) for recent discussions of that literature; Golez and Koudijs (2016) evaluate this link using four centuries of data). The goal of this paper is to provide a fresh view on this important question in empirical asset pricing. Specifically, we look beyond the mean of the return distribution and focus on the shape of the predictive return distribution. The key innovation of the paper is that we model the shape to depend on a valuation ratio such as the cyclically-adjusted price-earnings ratio or the book-to-market ratio.

Our research idea is illustrated by a quick look into the data. Figure 2 reports histograms of observed cumulative 12-month log-returns conditional on valuation ratios being HIGH (top quartile) and LOW (bottom quartile) and highlights a pronounced shift in the shape of the distribution: while it looks symmetric in the case of low valuation ratios, it becomes negatively skewed for higher valuation ratios. This pattern suggests time-variation in market skewness conditional on valuation levels. It extends the currently prevailing view in the literature that aggregate returns exhibit, in general, negative skewness (see Campbell and Hentschel (1992) and Duffee (1995) for early evidence and Oh and Wachter (2019) for a recent paper building on this view). We are not aware of any academic study that has documented and carefully modeled this pattern. This represents one of the contributions of our paper.

Such an asymmetry has important economic implications. While the existing literature on return predictability helps us understand the dynamics of time-varying expected rates of return, it does not explain how the reversion to that mean will actually occur. For example, when valuations are high (low), how will prices adjust reflecting the expected low (high) returns? Is this adjustment more likely to happen smoothly or rather abruptly? Another central question in this literature is about the interpretation of long-lasting deviations of market values from fundamentals: do these patterns reflect (rational) bubbles? Put differently, why is the timing of market reversals so difficult even at extremely high valuations? These are precisely the questions we will address in this paper.

Specifically, we propose an econometric framework that is simple but flexible enough to model the asymmetry of the return distribution as a function of a valuation ratio and, at the same time, nests the standard, linear predictive regressions as a special case. In more detail, we compare the model with conditional skewness to two benchmark models; one that implies a symmetric distribution, and one that implies a distribution with constant skewness. To reflect the well-known fact that equity returns have fat tails and to avoid any confounding effects between excess kurtosis and skewness of the estimated return distributions, we use a (skew) T-distribution to model returns instead of a (skew) normal distribution. For some comparisons, however, we also refer back to the standard Gaussian model.

Relative to the standard, linear predictive regression, our main empirical model with conditional skewness has similar implications for mean prediction. However, the model is powerful enough to help us understand how regression to the mean works. Using this framework and the standard US data, we find strong statistical evidence that the shape of the return distribution varies conditional on the valuation ratio and that the distribution becomes more negatively skewed when valuation ratios are high. Put differently, our empirical evidence documents that if valuations are high, regression to the mean is more likely to happen with strongly negative returns; in contrast, if valuations are low it is more likely to happen smoothly.

The model with conditional skewness is well-supported by the data. Its log likelihood exceeds those of the competing benchmark models; and the parameter governing the link between valuation levels and the shape of the return distribution is statistically significant. These results are very robust across different subsamples (pre-1945 and post-1945 samples), returns horizons (12-month and 24-month returns), proxies for valuation ratios (the cyclically-adjusted price-earnings CAPE ratio, the margin-adjusted CAPE ratio, the book-to-market ratio and past 5-year returns), model specifications (also allowing conditional dispersion to depend on the valuation level), international equity markets (the UK and a global portfolio of international equity indices) and industry portfolios.

Interestingly, when valuation ratios are very high the most likely value of the future return (i.e., the mode of the predictive distribution) still remains positive (in fact roughly unchanged) in our empirical analysis showing that timing the peak of a bull market is made inherently more difficult as a consequence of time-varying skewness. Conversely, since at low valuations our predictive distributions become approximately symmetric, very low valuations have higher power to forecast market direction.

The model with conditional skewness also shows promising out-of-sample performance. The model significantly outperforms the competing benchmark models with a symmetric or a constant-skewness distribution, both using standard statistical as well as utility-based metrics. With respect to the no-predictability benchmark, we find no evidence of predictability, as one would expect given the simplicity of the model. However, looking at prediction errors as well as utility gains over time shows that the model with conditional skewness experiences long periods during which it consistently outperforms the no-predictability benchmark. However, during pronounced periods of high valuations (for example, the 90ties or the most recent 10 years) it is punished for not being able to forecast the market downturns.

Our empirical results have important implications for investors, asset managers and risk managers. Obviously, ignoring that the return distribution becomes very negatively skewed when valuation levels are high leads to severe underestimation of risk measures such as volatility, value-at-risk and expected-tail-loss. These issues of underestimation of risk hold for the standard Gaussian model as well as the two benchmark models that we evaluate empirically, a model assuming a symmetric T-distribution and a model assuming a T-distribution with constant skewness. For example, while a symmetric T-distribution, with parameters estimated from the full sample, implies a 1% value-at-risk of -45% (-36% simple returns) for 12-month cumulative total log returns when valuation levels are high, our model with conditional skewness implies a 1% value-at-risk of -71% (-51% simple returns) in this case.

Interestingly, we observe the mirror-image of this pattern, albeit to a less extreme extent, when valuation levels are low. In this case, the Gaussian model and our benchmark models overestimate risk for any risk measure that we look at. For example, while a symmetric T-distribution estimates the 1% value-at-risk to be -28% for 12-month cumulative total returns when valuation levels are low, our model with conditional skewness implies a 1% value-at-risk of -23% in this case. Thus, from an investor's point of view ignoring the conditional skewness is a lose-lose situation. For example, a mean-variance investor using a Gaussian model would invest too aggressively in the market when valuation ratios are already high but too conservatively when valuation ratios are low.[1]

The remainder of the paper is organized as follows. In Section 2 we summarize the related literature focusing on theoretical models consistent with our empirical results. In Section 3, we describe the empirical model and the predictive framework. Section 4 describes the data used in our analysis and provides some descriptive statistics. Section 5 summarizes the empirical results including robustness tests. Section 6 concludes.

2 Related Literature

Several theories have been proposed to rationalize negative skewness in asset returns. Among these the "leverage effect" (a drop in market valuations increases leverage ratios and, as a consequence, increases volatility of subsequent returns) and the "volatility feedback effect" (bad news lowers future expected cash-flows and increases the risk premium; good news, in contrast, increases future expected cash-flows but, again increases the risk premium resulting in a dampened overall

[1] Recently asset management strategies based on volatility, such as risk parity and volatility targeting, have become increasingly popular (see, for example, Moreira and Muir (2017)). Obviously, such strategies are very sensitive to accurate estimates of future volatilities.

effect) have been found to lack the quantitative importance to explain the data (see, for example, Bekaert and Wu (2000) and Poterba and Summers (1986)). Chen, Hong, and Stein (2001) propose and evaluate an alternative explanation based on heterogeneous investors, differences in opinions and short-sale constraints for some investors. Hueng and McDonald (2005), however, find no support for this explanation in the case of aggregate stock market returns.

Importantly, however, the theories discussed in the previous paragraph fail to rationalize that the shape of the return distribution varies with valuation ratios. A theoretical motivation that overcomes this shortcoming is linked to stochastic rational bubbles, as first developed by Blanchard and Watson (1982). In these models, the stock price is the sum of a fundamental price and a bubble component. The bubble is stochastic, as it continues with a given probability p and bursts with probability $(1\text{-}p)$. Importantly, the model explicitly links the shape of the predictive distribution to the valuation ratio. If one, for example, assumes that the fundamental price follows a symmetric distribution then the Blanchard-Watson model implies a symmetric predictive distribution at low valuations (in this case, the bubble component is zero); at high valuations, however, the predictive distribution becomes increasingly left skewed as a mixture of two distributions. This would be consistent with our empirical results.[2]

Another theoretical framework that fits our empirical results is the one on funding liquidity and liquidity spirals proposed in Brunnermeier and Pedersen (2009) and evaluated for carry trades in Brunnermeier, Nagel, and Pedersen (2008). In that framework, assets that speculators invest in feature negative skewness arising from an asymmetric response to fundamental shocks: losses of speculators are amplified when they hit funding constraints (e.g., margin calls); as a consequence they unwind their positions and further depress prices only deepening their funding constraints and leading the asset market into a liquidity spiral; positive shocks to the positions of speculators, in contrast, are not amplified. Importantly, Figure 7 in Brunnermeier and Pedersen (2009) shows that skewness becomes more negative with initial funding levels. This model prediction fits our empirical patterns well, as valuation levels of asset markets should be related to funding levels.[3]

A third stream in the theoretical asset pricing literature that features asymmet-

---

[2]Note, however, that we do not view our empirical results as evidence in support of the existence of price bubbles. Instead, we view them as being merely consistent with some, but not all, of these models' predictions. If **rational** bubbles existed, prices would change while expected returns would not (see, for example, Cochrane (2011)). In our case, however, we find both, time-varying asymmetry, consistent with the Blanchard-Watson model, and predictability in mean returns, consistent with Campbell and Shiller (1988) and Cochrane (2008).

[3]Jorda, Schularick, and Taylor (2016) make a related argument showing that the business cycles in countries with high leverage (proxied by private credit to GDP) tend to be more negatively skewed featuring more pronounced crashes.

ric responses of returns to fundamental shocks is built around time-varying risk aversion (e.g., habit-based models). A recent example of that literature that is in line with our empirical results is Greenwald, Lettau, and Ludvigson (2016). They propose a model to explain stock price fluctuations in which investors are close to risk-neutral most of the time but subject to rare spikes in their risk aversion that generate a "flight-to-safety"and, as a consequence, a rapid drop in the price of the risky asset. An important advantage of that framework is that in addition to the shape predictability it also matches our results for expected rates of returns, which are essentially zero when valuation levels are high and positive when valuations are low. In contrast, stochastic bubble models — having a constant expected rate of return — and the liquidity spiral framework — featuring an always positive expected rate of return to compensate for liquidity risk — seem to be at odds with our empirical estimates of expected rates of returns.

Finally, David and Veronesi (2014) develop a dynamic equilibrium model of learning that also provides a rationalization for the link between valuation ratios and the shape of the return distribution. In their model investors learn about different regimes in the fundamental value. During a boom period, positive news about fundamentals has little impact on investors' beliefs; negative news, however, may lead to a large downward revision in beliefs; thus, in that situation investors perceive greater downside risk than in bad times. As a consequence, stock returns will be negatively skewed in good times.

Interestingly, David and Veronesi (2014) also provide evidence from option markets that is consistent with our empirical results. They find that the ratio of the implied volatilities of out-of-the-money puts over out-of-the-money calls, an indicator of the market's assessment of downside risk versus upside risk, raises during expansions and drops during recessions (i.e., it is pro-cyclical). Their analyses focuses on three-months options and the sample period of 1988 to 2011. Our analysis, instead, documents consistent patterns for long-term returns up to two years and is based on a much longer sample period.

Similarly, Veldkamp (2005) develops a model of endogenous information flow to study slow booms and sudden crashes in lending markets in emerging markets. In the model, agents undertake more economic activity in good times than in bad time. Thus, economic activity generates public information about the state of the economy. If the economic state changes when times are good and information is abundant, asset prices adjust quickly and a sudden crash occurs. When times are bad, scarce information and high uncertainty slow agents' reactions as the economy improves; a gradual boom ensues.

In terms of empirical literature, the following two recent papers are closely related to our work. Greenwood, Shleifer, and You (2019) study industry-returns and use an ad-hoc definition of bubbles based on past returns. Looking exclusively at those bubble periods, they document results that are consistent with the ones we report; such as, for example, that a sharp price increase predicts a substantially higher probability of a crash. They, however, do not study the predictive relation between valuation levels and the shape of the return distribution in a comprehensive econometric framework that also allows for predictability in other characteristics of the return distribution.

Gormsen and Jensen (2017) study higher-order moments of monthly and quarterly returns using estimates extracted from option markets. The main advantage of those estimates is that they are forward looking. Relying on option markets, however, also comes at a cost, such as, for example, the lack of options with long-horizon maturities and an overall relatively short sample size. While their empirical setup is quite different from ours, some of their results are qualitatively consistent with our analysis. For example, they also show that higher-order risks are time-varying and tend to increase during good times.

Our paper also relates, more broadly speaking, to the literature on the non-normality of asset returns. Looking at daily or even higher-frequency returns, this literature finds excess kurtosis and negative skewness. It usually models conditional skewness as following an autoregressive process and models it jointly with conditional volatility (see, for example, Harvey and Siddique (1999) and Jondeau and Rockinger (2003)). Our approach is very different as we look at longer horizon returns (in particular, 12-month and 24-month cumulative returns) because we share the view of Fama and French (2017) and Bessembinder (2017) that the literature has so far been focusing on the distributional characteristics of short-horizon returns (daily and monthly) rather than on the characteristics of long-horizon returns. Studying longer horizons, of course, comes with its own econometric challenges, such as fewer observations. We focus on 12-month and 24-month log returns because the asymmetry in the predictive distribution typically peaks somewhere in this interval. It is also important to emphasize that skewness in daily returns may have little or no connection to the skewness of cumulative returns (see Appendix A for a detailed discussion of this issue).

Finally, our analysis is related to the extensive literature on the predictability of expected returns; in particular, to studies applying regime-switching models. Henkel, Martin, and Nardari (2011), for example, estimate a Markov switching multivariate model for returns and four predictive variables. For the simplified case of one predictive variable, namely a valuation ratio, it is possible to show that their empirical model together with the result that predictability exists mostly during recessions implies that the future return distribution is a mixture of two distributions, one with fixed mean (no predictability regime) and one with a mean decreasing in the valuation ratio. Under reasonable parameterizations, the future return distribution then exhibits time-varying skewness consistent with our evidence: skewness is negative and large in absolute terms when valuation levels are high, and small and potentially even positive when valuation levels are low.

3 Model Specification

A skew-T distribution with deterministically varying parameters. The standard predictive regression is a linear projection of cumulative log returns on a valuation (ratio), also in logs, so the implicit model is

$$y_{t,t+h} = \beta_0 + \beta_1 x_t + \varepsilon_t,$$

where $y_{t,t+h} = \log((P_{t+h} + D_{t+1:t+h})/P_t)$ are cumulative total log returns over h periods, x_t is a log valuation ratio, and OLS estimation is optimal under the assumption that ε_t is Gaussian.

The most parsimonious and interpretable way to extend this model to capture the idea that valuation ratios may also affect the shape of the distribution is to move from a symmetric to an asymmetric distribution, where the asymmetry is a function of valuation levels. A skew-normal distribution would be the most immediate extension of the regression model, but we prefer to be slightly more general and opt for a skew-T distribution. Allowing for fat tails is always good practice, particularly with financial data, and in our case it is particularly important to mitigate the risk of interpreting one or a few outliers as asymmetry or time-varying asymmetry. Forcing a Gaussian distribution on fat-tailed data results in extremely noisy estimates of skewness in repeated samples, particularly if skewness is measured as the centered third moment. In our sample the key results are little changed (t-statistics are even higher) if we force a high value for the degrees of freedom. However, since this restriction is strongly rejected by the data, we show results for the more general and robust model, which is

$$y_{t,t+h} \sim skewt(m_t, \sigma, v, \gamma_t)$$

where *skewt* is the skew-T distribution of Fernandez and Steel (1998). Here m_t is the mode (location parameters), σ is the dispersion parameter, v are the degrees of freedom, $0 < \gamma_t < \infty$ is the asymmetry (shape) parameter, and the model parameters are deterministic functions of a constant and x_t as follows:

$$\begin{aligned} m_t &= \beta_{0,m} + \beta_{1,m} x_t \\ \log \sigma &= \beta_{0,\sigma} \\ \log v &= \beta_{0,v} \\ \log \gamma_t &= \beta_{0,\gamma} + \beta_{1,\gamma} x_t. \end{aligned}$$

Notice that we work with logs of the dispersion and degrees-of-freedom parameters, σ and v, and also model $\log \gamma_t$ rather than γ_t as a linear function of log-valuation x_t. This makes the distribution $p(y_{t,t+h}|x_t)$ well-defined for any value of $\beta_{1,m}$ and $\beta_{1,\gamma}$.

This model nests the standard predictive regression as a special case with $\beta_{1,\gamma} = \beta_{0,\gamma} = 0$ and v fixed at a large number. If v is freely estimated, we have a regression with T rather than Gaussian errors. We will refer to this model as the Symmetric-T Model. An interesting comparison is with a model where the skew is fixed, so $\beta_{0,\gamma}$ is freely estimated but $\beta_{1,\gamma} = 0$. We will refer to this model as the Constant-Skew-T model. The main model of interest, however, is one in which we also estimate $\beta_{1,\gamma}$ to see whether valuation ratios affect the shape of the return distribution. We will refer to this model as the Conditional-Skew-T model. In this paper we present results for a simplified Conditional-Skew-T model by imposing $\beta_{1,m} = 0$. In our sample this restriction is never rejected using any standard selection criteria like BIC or AIC, and when $\beta_{1,m}$ and $\beta_{1,\gamma}$ are estimated jointly $\beta_{1,m}$ is always small with t-statistics much lower than one. What this implies is that the mode of the distribution is fixed, and as the distribution becomes more left (right) skewed, its mean is lower (higher). Of course this does not have to be the case for other assets or samples, where m_t could either move left or, as in the Blanchard-Watson rational bubble model (Blanchard and Watson (1982)), shift right at higher valuations.

We also consider an even more general version of the model by estimating $\beta_{1,\sigma}$ — also allowing the variance to be a function of valuations. Details on this model implementation can be found in the robustness section. This improves the fit to the data but does not have important implications for the analysis of the shape of the return distribution which represents the focus of this paper. Thus, we decided to focus on the simpler model throughout the paper. If our goal was to maximize the fit to the data we would indeed need to model the dispersion as time-varying, and include more variables; see Li and Villani (2010) for an example of such a model fitted to daily stock return data (without including any measure of valuation).

Skewness, asymmetry, and some features of the skew-T distribution. There are several skew-T distributions available in the literature. Jones (2014), with a univariate emphasis, and Lee and McLachlan (2013), with a multivariate emphasis, provide excellent reviews. Most proposals are fairly recent and there is still very little applied work to guide a choice (Jones (2014)). We have opted for the version of Fernandez and Steel (1998) because, in their model, the role of each parameter is easy to interpret; in particular, our main hypothesis — that asymmetry varies with valuations — is captured by just one parameter. It also nests the standard regression equation and its likelihood is available in closed form, which aids in the estimation.

The idea of Fernandez and Steel (1998) is to introduce an inverse scale factor in the positive and negative orthants, so that if the distribution $f(\varepsilon_t)$ is unimodal and symmetric around zero, then we can create a skewed distribution p indexed by γ_t

$$p(\varepsilon_t|\gamma_t) = \frac{2}{\gamma_t + \frac{1}{\gamma_t}} \left\{ f(\frac{\varepsilon_t}{\gamma_t} I_{[0,\infty)}(\varepsilon_t) + f(\gamma_t \varepsilon_t) I_{(-\infty,0)}(\varepsilon_t) \right\}.$$

In our case $\varepsilon_t = f(y_{t,t+h} - m_t)$ and $f(\varepsilon_t)$ is a student T distribution with dispersion σ and degrees of freedom v. In Fernandez and Steel (1998) γ is fixed, but the extension is fairly straightforward. In our experience, this two-piece transformation fits moderate skewness well and is very convenient and robust in estimation, but may not be the best choice for severe skewness.

Each parameter has a fairly straightforward interpretation: m_t is the mode, σ is the dispersion, v controls the fatness of the tails, and γ_t determines the amount of asymmetry. However, each statistical moment is in general a function of all four parameters (see Fernandez and Steel (1998) for closed-form expressions). In particular, m_t is the mode, which differs from the mean unless $\gamma_t = 1$, the variance is a function of σ, v, and γ_t, and the most common measure of skewness as the centered third moment divided by the cubed standard deviation is also a function of v, and γ_t.

For unimodal distributions, Arnold and Groeneveld (2010) propose a measure of skewness defined as one minus twice the probability mass left of the mode, which in our case is

$$\frac{\gamma_t^2 - 1}{\gamma_t^2 + 1},$$

since in the skew-T distribution of Fernandez and Steel (1998), γ alone controls the allocation of mass to each side of the mode as

$$\frac{P(y_t \geq m_t|\gamma_t)}{P(y_t < m_t|\gamma_t)} = \gamma_t^2.$$

Given our use of the Fernandez and Steel skew-T distribution, ”asymmetry” in this paper is a one-to-one function of the amount of probability mass on each side of the mode. This definition is of course not free from shortcomings, but it is intuitive and far more stable than the centered third moment, particularly for fat-tailed distributions.

Estimation. Since the likelihood and all derivatives are available in closed form, estimation by maximum likelihood is convenient and works well for the small models considered in this paper. When using overlapping data,[4] the assumption of conditionally independent observations is incorrect and results can be interpreted as quasi-ML. A correction for autocorrelation should then be made to compute standard errors and t-statistics.

A very effective Markov Chain Monte Carlo algorithm (Gamerman (1997)) exists for generalized linear models, of which ours is a special case. Our version is taken from Li and Villani (2010). The problem is broken into sequential steps of estimating the coefficients associated with each parameter in separate blocks, with tailored proposal distributions obtained by maximizing the conditional likelihood at each step. The computational cost is compensated by increased reliability: in more complex problems and/or in less informative data, there can be multiple modes that the MCMC is able to explore in our experience. The general version of the model in which explanatory variables can affect both the mode and the asymmetry is particularly prone to multimodality, requiring either MCMC or great care in optimization.

For all results presented in this paper, the posterior means from MCMC (which we report and which are obtained with very disperse priors) for the key parameters of interest, namely $\beta_{0,\gamma}$ and $\beta_{1,\gamma}$, are nearly identical to ML estimates. Maximum likelihood estimation gives consistently lower estimates of the degrees-of-freedom parameter than MCMC. This is not surprising: the data contain some very large outliers, and ML can only accommodate them with a fairly low v. In contrast, MCMC results in a posterior distribution for v. This distribution has a higher mean and mode than the ML estimate, but also a tail of very low draws of v which induce very fat tails in the distribution of returns. We view this as a highly desirable feature of fully Bayesian MCMC inference, since it allows for very fat tails without forcing the spikes in the center of the distribution that are associated with low degrees-of-freedom in a student-t distribution.

We are not aware of any fully Bayesian approach to inference with overlapping observations: the likelihood is technically misspecified. Common practice is to either work with non-overlapping observations, which throws away a lot of useful data, or work as if observations were independent, which gives incorrect posteriors and over-confident results. We have employed an ad-hoc fix inspired by autocorrelation-consistent standard errors computed in a frequentist approach: the log-likelihood within each MCMC step is divided by $1+0.5(h-1)$ in an attempt to account for overlapping observations. For the results reported in this paper, this produces standard deviations extremely close to autocorrelation consistent standard

[4]Main results reported in the paper are based on overlapping data. In the robustness section, we also present results from non-overlapping data that are qualitatively as well as quantitatively very similar to our main results.

errors when the posterior mean is close to the posterior mode (the ML estimate), as in the case for the key coefficients of interest $\beta_{0,\gamma}$ and $\beta_{1,\gamma}$. We report results from these MCMC draws, but emphasize that the key findings are equally strong from ML estimates. For the full sample, t statistics for the main parameter of interest $\beta_{1,\gamma}$ are close to three even if non-overlapping observations are used.

4 Data and Descriptive Statistics

The key variables of interest are the cumulative, overlapping (i.e., all possible 12-month and 24-month periods are considered) 12-month and 24-month total log returns. The only predictive variable is the cyclically adjusted price-to-earnings ratio (CAPE); as a robustness test, we replicate our main results using the market-to-book (MB) ratio, the margin-adjusted CAPE (Hussman (2017)), or the 5-year return (e.g., Asness, Moskowitz, and Pedersen (2013), Greenwood, Shleifer, and You (2019)) as the only predictive variable. The sample period is from January 1881 (January 1921) to December 2014 in the case of CAPE (MB). All variables are in logs. We also standardize the valuation ratios in the model such that the mean is zero and the standard deviation is one, which makes parameters easier to interpret. Data is at the monthly frequency and is taken from Amit Goyal's webpage.

Figure 1 shows the time-series graph of market-to-book and CAPE. As one would expect, the two series are very closely related — noticeable differences can be observed up to the 30ties and during the 60ties and 70ties. Negative values correspond to periods of time during which book values exceed market prices while large positive values correspond to market booms. One clearly observes the stock market downturn before the great depression and the run-up and subsequent correction associated with the boom in technology stocks at the end of the last century.

Table 1 presents the summary statistics of 12-month (Panel A) and 24-month (Panel B) total returns including means, standard deviations and skewness. We report these statistics for the full sample period, and the pre-1945 and post-1945 subperiods separately. Furthermore, we report them separately for the first, the pooled second and third, and the fourth valuation quartiles. There are several measures of skewness available in the literature, each attempting to quantify the asymmetry in a distribution. The most common measure is the centered and standardized third moment. This statistic is known to suffer from large sampling errors in the case of distribution with fat tails and is therefore very susceptible to outliers.

The expected 12-month return is 16.2% in the case of low (lowest quartile of CAPE) and 3.9% in the case of high (highest quartile of CAPE) valuation ratios (the unconditional mean is 8.7% with a standard deviation of 18.8%). The standard deviation of 12-month returns is 16.9% for the case of lowest valuation ratios and only slightly higher at 19.4% for highest valuation ratios. In the case of 24-month returns the full sample average is 17.3% with a standard deviation of 26.3%; expected 24-month returns amount to 31.7% for periods with lowest and 7.3% for periods with highest valuation ratios; the corresponding standard deviations are 18.7% and 32.9%, respectively. Note the mild increase in standard deviation (between 12-month and 24-month returns) in the case of lowest valuation ratios and the comparatively very stark increase in the case of highest valuation ratios. The patterns in average realized returns observed across valuation quartiles are consistent with the standard Campbell-Shiller argument that low (high) valuation ratios predict high (low) expected rates of returns.

In terms of skewness, Table 1 shows that, as expected and consistent with existing literature, cumulative 12-month and 24-month returns are, in general, negatively skewed. Most importantly, however, we find that they are more negatively skewed when valuation ratios are high (top quartile) than when valuation ratios are low (bottom quartile). In the case of bottom-quartile valuation ratios, we frequently find even positive or close-to-zero skewness. Thus, these simple descriptive statistics already suggest a link between valuation ratios and the shape of the return distribution.

In some cases, however, skewness, as measured by the standardized third moment, does not monotonically decrease when valuation levels increase; i.e., in some cases we find even lower skewness when valuation ratios are in the middle quartiles. While this seems to be at odds with our story, it is most likely related to the previously discussed shortcomings of the standard skewness measure that we report in Table 1. To get a better idea of the shapes of the empirical return distributions, Figure 2 (12-month returns) and Figure 3 (24-month returns) show histograms of realized returns for the full sample and conditional on valuation ratios at the beginning of the return observation period. In both cases, we clearly see that the shape of the distribution of observed returns changes substantially conditional on the valuation ratio. While it looks slightly positively skewed in the case of the lowest valuation quartile, it becomes increasingly asymmetric and negatively skewed for higher valuation quartiles. These patterns also appear to be somewhat more pronounced for realized 24-month than 12-month returns.

Finally, we explore the term-structure of skewness over various horizons. As discussed before (and in detail in Appendix A), the skewness in daily returns might have little to do with the skewness in cumulative returns. Figure 4 shows estimates of the asymmetry in cumulative returns for horizons of one month up to 35 months. To address the issue of noisiness in standard skewness measures, we plot the Arnold and Groeneveld measure of skewness (see Arnold and Groeneveld (2010)) for skew-t distributions estimated on the data. Again, we do this separately for the entire sample as well as for observations in the lowest and highest valuation quartiles.

Focusing on the entire sample, we see that skewness becomes more negative when the horizon increases and then reverts back to zero, but the convergence is slow. A similar pattern is observed when valuation levels are high but the initial increase in negative skewness up to horizons of 20 months or so is even more pronounced. The case of observations when valuation levels are low looks somewhat different because skewness increases (i.e., becomes less negative or even positive) up to horizons of 6 months, then drops (i.e., becomes increasingly more negative) up to horizons of 19 months, and then reverts back to zero or even positive values at a faster speed than in the case of all observations or high-valuation observations. Importantly, these results show that the basic relation between valuation levels and skewness of cumulative returns is not horizon-dependent. They also illustrate that skewness does not converge to zero for those horizons that we consider in the paper (see, for example, Neuberger (2012), Neuberger and Payne (2018) and Johansson (2019)).

5 Empirical Results

In this section, we summarize our empirical results focusing on 12-month and 24-month cumulative total log returns and the cyclically-adjusted-price-earnings ratio (CAPE) as proxy for the valuation ratio.

5.1 Model Parameters

Table 2 summarizes parameter estimates for the three models of interest — the Symmetric-T model, the Constant-Skew-T model and the Conditional-Skew-T model — when 12-month returns are modeled. Panel A reports results based on the full sample of data, Panel B focuses on the pre-1945 and Panel C on the post-1945 sub-period.

The Symmetric-T model represents the standard, simple linear regression model with the only difference that we assume a T-distribution instead of a normal distribution for the residuals. Consistent with the literature we find that valuation ratios predict expected returns: a one-standard deviation increase in $log(CAPE)$ results in a, statistically and economically significant, drop in expected 12-month returns of 4.226%. The corresponding full sample OLS estimates assuming a normal distribution for the errors are 8.667 for $\beta_{0,m}$ and -4.817 for $\beta_{1,m}$. Thus, the impact of the valuation ratio is slightly smaller once residuals are modeled to follow a fat-tailed distribution. Note also that parameters $\beta_{0,m}$ and $\beta_{1,m}$ predict the mode of the predictive distribution rather than the mean in our framework. However, as long as the predictive distribution is symmetric the mode is, obviously, equal to the mean.

The Constant-Skew-T model extends the basic model by allowing the predictive distribution to be skewed. This results in a substantial increase in fit as measured by the log-likelihood. The constant asymmetry parameter of -0.316 implies a negatively skewed predictive distribution, as one would expect given available empirical evidence. The coefficient of the valuation ratio in predicting the mode of the distribution stays essentially unchanged. Interestingly, however, the skewness parameter becomes insignificant and is cut in half, to -0.160, once we focus on the pre-1945 sample period implying that returns were less negatively skewed early-on according to that model specification.

Finally, Table 2 summarizes the parameter estimates for the Conditional-Skew-T model which models the predictive distribution's asymmetry as a function of the valuation ratio. We find that conditioning on the valuation ratio in the shape equation improves the model fit (i.e., the log likelihood increases). We find a value of -0.175 for $\beta_{1,\gamma}$ indicating that the distribution becomes more negatively skewed when valuation ratios increase. It also shows that the total estimate of the shape parameter, including the constant term, becomes essentially zero, implying a symmetric distribution, for valuation ratios that are close to two standard deviation below zero (i.e., low valuation ratios).

As discussed before, we observe substantial variation in $\beta_{0,\gamma}$ across sub-periods. Interestingly, however, estimates of $\beta_{1,\gamma}$ do not share this behavior. Thus, while the overall asymmetry in the return distribution has changed somewhat over time, the link between valuation levels and skewness appears to have been stable and statistically significant. Changes in the unconditional asymmetry (captured by $\beta_{0,\gamma}$) across sub-periods do not necessarily imply a break in the relation we are interested in, since $\beta_{1,\gamma}$ is stable and valuation proxies also have different sample means across sub-periods. The model then implies that periods of higher (lower) average valuations should have more (less) pronounced average skewness. A formal test for the null hypothesis that $\beta_{0,\gamma}$ is constant versus the alternative that it has changed post-1945 has a borderline t-statistic of 1.9, with the AIC criterion picking the extension and the more stringent BIC criterion choosing constant parameters.

Note that in the Conditional-Skew-T model we do not include the valuation ratio in the mode equation (i.e., we set $\beta_{1,m} = 0$). The main motivation to do so is for simplicity. Allowing CAPE to appear in both the mode and the asymmetry equation results in a lower t-statistic for our key parameter $\beta_{1,\gamma}$, even when its point estimate is little affected or even increases in absolute terms. The reason for this effect is that the correlation between $\beta_{1,\gamma}$ and $\beta_{1,m}$, whether measured from MCMC draws or from the asymptotic covariance matrix of the maximum likelihood estimator, is around 0.9. As a consequence, the t-statistics are much smaller when including CAPE in both equations, while any test rejects setting both parameters to zero strongly.[5]

Nevertheless, these results do have some further noteworthy implications: while the return distribution becomes much more negatively skewed when valuations are high, the most likely return of the distribution (i.e, the mode) is essentially unaffected by valuation levels. Given that $\beta_{0,m} > 0$, this also means that the most likely return is positive even at very high valuation ratios reflecting the difficulty of timing market reversals.

Table 3 contains parameter estimates of the three models when 24-month returns are used as dependent variable. Results look qualitatively very similar in this case. It is noteworthy to point out that the estimates of unconditional skewness in the Constant-Skew-T model, $\beta_{0,\gamma}$, are in all sample periods statistically insignificant and smaller than in the case of 12-month returns. In contrast, however, the coefficients capturing the conditional impact of valuation ratios on the shapes of the predictive distributions, $\beta_{1,\gamma}$, are always statistically significant and increase relative to Table 2.

Bottom line, we find — across all return definitions and sample periods considered — that valuation ratios have a statistically significant impact on the shape of the return distribution. Specifically, distributions become more negatively skewed when valuation ratios increase. In the following sections, we analyze the resulting shapes of the return distributions in more detail.

5.2 Predictive Distributions

The parameter estimates discussed in the previous section already document that valuation ratios help predict the shape of the distributions of 12-month and 24-month returns. Judging, however, how large this impact is in terms of the resulting asymmetry of the distributions directly from the parameter estimates is difficult. Thus, we take a detailed view at the predictive distributions implied by the various models in this section.

Figures 5 and 6 represent the main results of the paper. They show the conditional predictive distributions for 12-month (Figure 5) and 24-month (Figure 6)

[5] To further understand this behavior, it is useful to notice that $\beta_{1,m} = 0$ affects only the conditional mean of the predictive distribution, while $\beta_{1,\gamma}$ affects all moments, though mainly mean and skewness. The impact of $\beta_{1,\gamma}$ on the mean is almost perfectly linear; that is, as far as point predictions are concerned, the two parameters are nearly unidentified. When two parameters are nearly unidentified, their t-statistics approach zero even if they have very strong effects and very high t-statistics when included individually. Since $\beta_{1,\gamma}$ also has a strong impact on the asymmetry of the distribution, both parameters end up being identified.

returns implied by the Conditional-Skew-T model[6] using the full sample parameter estimates. The top graph in each figure represents the case of high and the bottom graph the case of low valuation ratios. While the modes of the two distributions are identical by design, the model implies very different shapes of the distribution depending on the level of the valuation ratios: while predictive distributions look pretty much symmetric for low valuation ratios, they become asymmetric and negatively skewed in the case of high valuation ratios. As discussed before, results are slightly stronger for 24-month returns than for 12-month returns.

Table 4 provides some further information on the conditional distributions implied by our models, namely the mean, the standard deviation, the normalized third moment (skewness), the probability mass left to the mode (asymmetry), the 1% Value-at-Risk and the 1% Expected Tail Loss. Most importantly, we are interested in skewness and asymmetry. By construction, skewness is zero and asymmetry is equal to 50% in the case of the Symmetric-T Model. When we allow the distribution to be skewed in the Constant-Skew-T Model, we find that the implied distributions become asymmetric and that probability mass shifts to the left of the mode; in the case of 12-month (24-month) returns 65% (61%) of the probability mass end up being below the mode.

Finally in the Conditional-Skew-T Model, we observe that valuation levels have a strong impact on the shape of the distributions. In the case of 12-month returns, we find that, for low valuation levels, the distribution is nearly symmetric with a skewness of zero and 47% of the probability mass being to the left of the mode. In stark contrast, for high valuation levels, we find that nearly 80% of the probability mass is below the mode and skewness is equal to -0.97. A similarly pronounced pattern prevails in the case of 24-month returns with the only difference that, in the case of low valuation levels, the implied distribution seems to be positively skewed with only 39% of the probability mass to the left of the mode.

These results illustrate, yet again, that the shape of the return distribution depends strongly on valuation levels. An equally important question, however, is whether this shape dependence also has implications for other characteristics of the return distribution such as means or standard deviations. In general, across all models, we find that expected returns are considerably lower when valuation levels are high, as one would expect. Estimated conditional expected returns (as a function of valuation levels) are very similar across models.

We also observe an interesting pattern for model-implied standard deviations. Both, the Symmetric-T and the Constant-Skew-T Model, show a tendency to overestimate volatility when valuation levels are low and, at the same time, underesti-

[6] Note that the two benchmark models, by design, do not model conditional skewness and, thus, it makes no sense to draw these graphs for the two benchmark models.

mate volatility when valuation levels are high, relative to the standard deviations implied by the Conditional-Skew-T Model. That means that the two benchmark models are on the wrong side in both cases: when valuations are very low, a mean-variance investor using those standard deviation estimates would invest too cautiously while the same investor would invest too aggressively when valuation levels are very high. Note that in our econometric setup these volatility patterns only arise as a consequence of the change in the asymmetry of the distribution, as dispersion and degrees of freedom are modeled identically across all three models.

Similar patterns are observed when we move to risk measures beyond volatility, such as value-at-risk and expected tail loss. In both cases, we find that the two benchmark models overestimate risk in the case of low valuation levels but underestimate risk in the case of high valuation levels. For example, in the case of 12-month returns the Constant-Skew-T Model implies a 1% value-at-risk of -37% (-53%) when valuation levels are low (high) while the Conditional-Skew-T Model implies values of -23% (-71%). While it is difficult to judge economic importance of these differences without having a specific application or portfolio in mind, they certainly look sizable and noteworthy to us.[7]

So far, we have focused on full sample evidence in the discussion of the predictive distributions. Figures 7 and 8 illustrate the model-implied distributions separately for the pre-1945 and post-1945 sample periods while Tables 5 and 6 provide the corresponding characteristics. Most importantly, the patterns that we discussed above based on the full sample also hold for each sub-sample separately. Thus, our results are robust across different sample periods and do not seem to be driven by individual years or particular events.

5.3 Out-of-sample Predictability

The results so far have focused on in-sample analyses and have shown that the Conditional-Skew-T Model is supported in the data and outperforms the Symmetric-T and the Constant-Skew Model. In this section, we evaluate how well the Conditional-Skew-T Model performs in an out-of-sample framework. Monthly predictions of 12-month ahead expected returns are generated using an expanding window using the period up to January 1945 as a burn-in phase.

Figure 9, in a first step, shows the maximum-likelihood point estimates of the key parameters of the Conditional-Skew-T model, $\beta_{0,\gamma}$ (labeled b4(1) in the figure) and $\beta_{1,\gamma}$ (labeled b4(2) in the figure), together with +2/-2 standard deviation bands. Given that we also demeaned the valuation ratio in the model, coefficient

[7] In the case of the Conditional-Skew-T Model and 24-month returns, we find value-at-risk and expected-tail-loss estimates of -103% and -126%, respectively. Note that throughout the paper we use log-returns. Thus, these estimates correspond to -64% and -72% in terms of simple returns.

$\beta_{0,\gamma}$ can be directly interpreted as the unconditional skewness. As expected, the unconditional skewness of the market is consistently negative throughout the period. Interestingly, however, we find that there is a decreasing trend in unconditional skewness over time.

In case of parameter $\beta_{1,\gamma}$ that captures the predictive relation between valuation ratios and the shape of the return distribution, we find that the estimate is negative throughout the entire period and that the two standard deviation confidence interval never includes the zero line. Most importantly, the figure illustrates that the parameter estimate varies very little over time. Only during the 90ties leading into the burst of the "tech-bubble"we observe a brief period with more pronounced changes in the parameter.

Figure 10 shows the corresponding monthly predictions of 12-month returns using the Conditional-Skew-T Model over time together with CAPE.[8] It shows, as one would expect given our earlier discussions, that predicted returns vary considerably over time and move against market valuations. For example, while valuations increased consistently between the early 80ties and the early 2000s, predicted expected returns consistently dropped reaching zero or even negative value when valuation levels were at the peak.

5.3.1 Statistical Evaluation of Out-of-sample Predictability

To evaluate the out-of-sample predictive performance of the Conditional-Skew-T model, we follow the predictability literature and perform standard statistical tests in a first step. In the next subsection, we will then complement these results with an economic analysis. We start with a statistical evaluation based on mean squared prediction errors (MSPEs). The no-predictability benchmark is an expanding window estimate of the average 12-month return. It is well-known that the Symmetric-Model does not yield statistically significant out-of-sample predictability on this data — nevertheless we still include it in our analysis to facilitate the comparison with the literature.[9]

In terms of MSPEs, the Conditional-Skew-T model performs surprisingly well. It statistically outperforms the Symmetric-G model (with a p-value of 1%) and

[8] In this section, we focus on simple returns in the analysis while the model is estimated for log-returns. Qualitatively results are very similar when doing the assessment of out-of-sample predictability in log-returns. Given that the idea here is to evaluate whether one could trade using the model, discrete returns appear to be the more natural choice.

[9] To further simplify the comparison with the literature in this section, we replace the Symmetric-T benchmark model with a Symmetric-G(aussian) model. The reason why we do not also use that Symmetric-G model in the earlier in-sample analysis is that a model with normally distributed returns would not stand a fair chance in a likelihood comparison given that the distribution of log returns is known to exhibit large excess kurtosis.

the Constant-Skew-T model (with a p-value of 2%) during the out-of-sample period.[10] Interestingly, relative to these competing models the out-of-sample predictability comes mostly from expansionary observations while during recessions the Symmetric-G and the Constant-Skew-T model seem to predict returns somewhat more accurately than the Conditional-Skew-T model.

With respect to the no-predictability benchmark, we find overall a positive (i.e., the Conditional-Skew-T model predicts more accurately on average) but statistically not significant effect. In this case, a very pronounced result is that the Conditional-Skew-T model predicts much more accurately than the no-predictability benchmark during recessions. However, during expansions this pattern is reversed.

The above discussion of the predictive performance is based on point estimates of average prediction quality, which can be driven by individual outliers. Thus, to get a better idea of the extent of predictability and its consistency over time, we plot the cumulative difference in squared prediction errors between models. Panel A of 11 shows the corresponding results when we compare the Conditional-Skew-T model to the Symmetric-G model (solid line) or to the Constant-Skew-T model (dashed line). In both cases, we observe that the line is increasing consistently across time indicating that the strong performance of the Conditional-Skew-T model relative to these benchmark models is not driven by a few, individual observations.

Panel B repeats this analysis looking at the difference in cumulative squared errors between the Conditional-Skew-T and the no-predictability model. The graph actually shows that the Conditional-Skew-T model outperforms the no-predictability benchmark during extended periods of time but is punished very substantially during periods of high valuations. This is just another representation of the earlier-discussed challenge of timing stock market downturns.

5.3.2 Economic Evaluation of Out-of-sample Predictability

To corroborate the statistical evidence on out-of-sample predictability, we also perform an economic evaluation assuming an investor with power utility and risk aversion coefficient equal to 3. Each month the investor calculates the optimal weight of investing in the market and the risk-free asset by maximizing expected utility. The investor rebalances monthly even though the predictions used in the optimization are for the 12-month horizon. We calculate ex-post realized utility based on monthly returns as well as the monthly volatility of the resulting strategy. We focus on power utility in the economic evaluation as an investor using power utility is automatically averse to skewness because the disutility of a negative return –R is

[10] Reported p-values have to be taken with a grain of salt as they are not adjusted for the overlapping nature of the returns.

(much) larger than the utility of a similar-sized positive return +R for moderately large return realizations.[11]

In terms of realized utility gains, the Conditional-Skew-T model again clearly outperforms the two considered benchmark models. For example, relative to the Constant-Skew-T model it yields an average, annualized utility gain of 0.25% which is small in economic terms but still statistically significantly (p-value of 2%) different from zero. Relative to the symmetric model, the average utility gain is 0.43%. Relative to the no-predictability benchmark, the Conditional-Skew-T model underperforms but the average utility loss is not statistically significant. Interestingly, the Conditional-Skew-T model yields a very substantial utility gain of 4.3% relative to the no- predictability benchmark during recessions.

Figure 12 plots cumulative utility gains over time to assess the consistency of the average utility differences just discussed. Panel A, again, compares the Conditional-Skew-T to the two competing benchmark models and shows that the Conditional-Skew-T model outperforms the benchmark models consistently during most time periods. In the most recent history, say the last twenty years, the Constant-Skew-T model shows strong performance resulting in a noticeable decrease in cumulative utility gains relative to the Conditional-Skew-T model.

Panel B of Figure 12 shows the cumulative utility gains of the Conditional-Skew-T relative to the no-predictability benchmark. The graph shows extended periods, for example from the early 60ties to the early 90ties, during which the Conditional-Skew-T model yields consistent utility gains. On the other side, there are several, pronounced periods, especially the run-up in valuations during the 90ties but also the most recent 10 years, during which the utility gains collapse, as an investor unaware of (or not responding to) high valuations would obviously outperform a valuation-sensitive investor during prolonged periods of high and rising valuations.

5.4 Robustness

To make sure our results about the predictability of the shape of the return distribution are not driven by discretionary choices we made along the way, we perform an extensive set of robustness tests. First, we rerun our main analysis using non-overlapping returns. Second, we repeat the empirical analysis for three alternative proxies for valuation levels and also study the impact of margin debt on the shape of the return distribution. Third, we extend the model to allow for a link between the valuation level and return dispersion. Fourth, we provide international and,

[11] Our results are robust to different choices of utility functions and look qualitatively as well as quantitatively similar if we consider, for example, an investor with quadratic utility. Our results are also robust to different choices of reasonable values of risk aversion.

fifth, industry-level evidence. In all cases, the empirical analysis shows that our results are very robust to these changes.

5.4.1 Non-overlapping Returns

The results discussed in the paper are based on overlapping returns. Using overlapping returns yields a gain in efficiency, as all available information is exploited in the estimation, but comes at the cost of making the assumption of conditionally independent observations incorrect. In this robustness section, we re-estimate our models using non-overlapping data (specifically, returns from January to January) to ensure that our results are not driven by the choice of using overlapping data.

Table 7 shows the corresponding parameter estimates for 12-month (Panel A) and 24-month (Panel B) returns. Several interesting observations can be made. First, point estimates of individual coefficients across models and horizons are only marginally affected by using non-overlapping instead of overlapping returns. Second, t-statistics drop, as one would have expected but remain above standard thresholds for main coefficients of interest. Third, all results discussed before also hold up when using non-overlapping data. Most importantly, we also find a statistically significant and negative relation between valuation levels and the shape of the distribution of future returns. Thus, our results are robust to the use of non-overlapping data.

5.4.2 Alternative Valuation Ratios and Margin Debt

As the first set of robustness tests, we replicate the main steps of our empirical analysis using the market-to-book (MB) ratio, the margin-adjusted CAPE (Hussman (2017)), and the 5-year return (e.g., Asness, Moskowitz, and Pedersen (2013), Greenwood, Shleifer, and You (2019)) instead of the CAPE ratio as our predictive variable. While the market-to-book ratio and the 5-year return are straightforward to construct (when using the market-to-book ratio the sample only starts in January 1921), the margin-adjusted CAPE needs some more explanation.

The simple but intuitive idea of the margin-adjusted CAPE is that margins are embedded into every earnings-based valuation ratio, including CAPE. As margins vary themselves over time and over the business cycle, it might be useful and important to explicitly account for them. Hussman (2017) argues that adjusting CAPE for that embedded margin significantly improves the relationship between CAPE and subsequent market returns. To construct margin-adjusted CAPE one would ideally like to have information on aggregate sales of the firms in the S&P 500 to relate them to S&P 500 earnings but, unfortunately, that data does not seem to be available sufficiently long back in time. Thus, we use data on corporate profits for the entire economy divided by GDP using Federal Reserve Economic Data as a proxy for S&P500 earnings divided by sales.

Specifically, we calculate margin-adjusted CAPE in the following way. We first collect annual data on corporate profits after tax and on GDP (available from 1929) and then switch to quarterly data for these two data series in 1947. We then compute profits-to-GDP for each quarter (or year), and set all months in that quarter (year) equal to that value. Then, we use a 10-year sliding window to compute a smoothed value of profits/GDP. To get the margin-adjusted CAPE, the standard CAPE at each time t is multiplied by the ratio of this 10-year smoothed value of profits-to-GDP to its full sample mean.

Table 8 shows the corresponding parameter estimates if we use 12-month (Panel A) and 24-month (Panel B) returns as dependent variables. Note that for simplicity and readability we focus on the Conditional-Skew-T model in the table. This choice does not represent a limitation, as we also find for these alternative valuation proxies, similar to the main results, that the the Conditional-Skew-T model gets most support in the data compared to the Symmetric-T and Constant-Skew-T model (as reflected by maximum log-likelihoods).

Most importantly, Table 8 shows that the coefficients of the alternative valuation proxies in the shape equation are all negative and statistically significant. Thus, as valuation levels — proxied by any of these three alternative proxies — increase, the shapes of the return distributions, for both 12-month and 24-month returns, become more asymmetric and, in particular, more negatively skewed.

Table 9 reports detailed characteristics of the model-implied distributions: the patterns described in the main result section also prevail when we use these alternative proxies for valuation levels. Most importantly, we observe a stark change in the shape of the distributions conditional on valuation levels. For example, in the case of 24-month returns and the market-to-book ratio the probability mass to the left of the mode is 49% (83%) when valuation levels are low (high). Differences between the implied distributions when valuation levels are low and high are somewhat less pronounced when we use the 5-year return as a proxy for valuation levels. In this case, we also find that our empirical models fit the data considerably worse in terms of likelihood. This might not be too surprising given that the 5-year return is a very different and most-likely more noisy proxy for valuation levels compared to the market-to-book ratio or CAPE-based measures. Overall, however, our main results about the shape predictability of the return distribution are very robust to different proxies for valuation levels.

In a recent paper, Asness, Frazzini, Gormsen, and Pedersen (2018) argue that margin debt plays an important role in empirical asset pricing. Specifically, they argue that high margin debt means low financial constraints and low margin debt means tight financial constraints (i.e., the interpretation of margin debt is supply driven). The intuition is that at the end of an expansion margin debt is high and financial constraints are low (non-binding). However, when negative economic news or other signals of a turnaround arise, the supply of funding for intermediaries dries up (e.g., margin requirements are adjusted) and financial constraints tighten up. As Brunnermeier and Pedersen (2009) show such dynamics can result in liquidity spirals where funding liquidity as well as asset liquidity jointly deteriorate dramatically and quickly due to amplification effects.

To empirically evaluate this specific mechanism, we construct a time-series of monthly margin debt, normalized by GDP, for NYSE-designated clearing firms from Eikon. Figure 13 shows standardized CAPE and margin debt (MD) since January 1960. It highlights that there is a strong relation between the two time-series (the correlation is 0.72) but that there are also periods when they actually differ quite substantially. If we use margin debt to predict the shape of the future 12-month return distribution we find qualitatively similar results to the ones discussed above for CAPE and other valuation measures. MD negatively predicts the asymmetry in the return distribution; i.e., periods of high MD predict future returns with negative skewness. The coefficient, however, is smaller in magnitude than the one we find for CAPE; it also has a smaller t-statistic and the fit of the model overall becomes worse.

Thus, the empirical evidence suggests that margin debt, as a proxy for financial constraints of financial intermediaries, is associated with the time-variation in the asymmetry of market return distributions. Given, however, that the empirical support is weaker than for CAPE and other valuation ratios, it does not seem to be the only channel.

5.4.3 Model with Conditional Dispersion

The third robustness test considers an extension of our econometric specification that allows the return dispersion to depend on the valuation level. Specifically, we will estimate the following set of equations (consistent with the Conditional-Skew-T model we will also set the link between the mode of the return distribution and the valuation level to zero, as this link does not receive support in the data):

$$\begin{aligned} m_t &= \beta_{0,m} \\ \log \sigma &= \beta_{0,\sigma} + \beta_{1,\sigma} x_t \\ \log \nu &= \beta_{0,\nu} \\ \log \gamma_t &= \beta_{0,\gamma} + \beta_{1,\gamma} x_t . \end{aligned}$$

Table 10 summarizes the parameter estimates of this model for 12-month returns (Panel A) and 24-month returns (Panel B). Similar to the main results, we distinguish three samples — the full sample, the pre-1945 sample and the post-1945 sample. Across all specifications, we find that our main result still holds; i.e., $\beta_{1,\gamma}$ is negative and statistically significant. Comparing the point estimates to those reported in Table 2 and 3 shows very minor changes; basically the estimates of $\beta_{1,\gamma}$ are unaffected by allowing $\beta_{1,\sigma}$ to be different from zero.

In contrast, estimates of $\beta_{1,\sigma}$ are not significantly different from zero across all specifications. Thus, there does not seem to be a strong association between current valuation levels and the return dispersion of future returns. Ignoring the lack of significance for a second, it is interesting to point out that point estimates are consistently negative for 12-month returns while being consistently positive for 24-month returns.

Table 11, finally, characterizes model-implied distributions for low and high valuations. Not surprisingly, we do not observe any significant changes with respect to our main empirical specification (refer to Table 4 for the details). Again, we find that the implied distributions become much more negatively skewed when valuation levels are high. In the case of 24-month returns, we do observe that the positive but insignificant estimates of $\beta_{1,\sigma}$ have some noticeable negative impact on the expected rate of return (decreases) and the standard deviation (increases) when valuation levels are high. Bottom line, however, is that our main results are unaffected by whether return dispersion is allowed to depend on valuation levels or not.

5.4.4 International Evidence

As the fourth robustness test for our empirical results, we repeat our analysis on a sample of international equity markets. We obtain data on total returns, dividend yields, consumer price indices, price-earnings ratios and short-term interest rates from Global Financial Data and construct Shiller's CAPE for each market.[12] We use data at the monthly frequency but set monthly observations equal to the last available PE-ratio when price-earnings ratios are only available at the annual frequency (i.e., we do not perform any interpolation).

Table 12 lists the individual countries included in the international sample together with the dates when the data starts for each country. We end up with an unbalanced panel of 29 countries.[13] We leave out the US from this analysis to avoid any confounding effects. While we model returns in the case of US, we

[12] Shiller's CAPE is real price divided by the ten year moving average of real earnings. To construct it, we build returns from total returns by subtracting the dividend yield from the total returns; then we construct the real equity index and back-out real earnings using the price-earnings ratio.

[13] We started with a sample of 41 countries but had to drop countries which did not have sufficient data to construct CAPE.

focus on excess returns in the case of the international sample because inflation and interest levels vary considerably across countries in our sample (some of the countries included are emerging market countries).[14]

Using the international data, we run several robustness tests. First, we replicate our main results for the UK, which is the only country in the GFD data, for which we are able to construct a data history that is comparable in length to the one we used for the US. Second, we pool all countries and estimate a common model (i.e., common model parameters) across countries. Third, we allow for country-specific fixed effects. Fourth, we use trailing 5-year cumulative returns as proxies for valuation levels. Fifth, we add US CAPE to the country-specific CAPE in the model.

Figure 14 compares valuation levels for the US and the UK (top graph) and the US and the equal-weighted global portfolio (bottom graph). In both cases, one observes periods during which valuation levels seem to be closely related as well as periods during which they evolve rather independently from each other. Correlations are 0.64 (0.58) for the US and the UK (the US and the global portfolio) series. Another interesting pattern is that neither valuation levels in the UK nor in the global portfolio show positive spikes comparable to the ones we have seen in the US.

Our main results all hold up in these robustness tests. First, the Conditional-Skew-T model is most supported by the data (i.e., shows largest log-likelihoods) outperforming the Symmetric-T and Constant-Skew-T model in terms of fitting the data, both for 12-month and 24-month future returns.

Second, estimates of $\beta_{1,\gamma}$ are highly significant and negative confirming that valuation levels and skewness are robustly negatively related. In the case of the UK and the 12-month horizon, the coefficient estimate is -0.27 — compared to an estimate of -0.18 for the US (see Table 2) — with a t-statistic of -4.8 implying a substantially more pronounced effect of valuation levels on the shape of the return distribution. For the pooled international data, the coefficient estimate is very similar to the one we found for the US and amounts, for example, to -0.17 with a t-statistic of -8.6 in the case of 12-month returns[15].

When we allow for country-specific fixed effects, the same coefficient drops even further to -0.19 (with a t-statistic of -10.9). Panel A of Table 13, as an example, shows the detailed parameter estimates in this particular case. Adding then

[14]Note that results for the US are basically unchanged when modeling excess returns. In the case of the international sample, results are also the same qualitatively but the fit of the models deteriorates noticeable when we model returns instead of excess returns.

[15]The t-statistics that we report for the estimates of the pooled international sample have to be interpreted with a grain of salt, as they assume that countries are independent from each other, which is obviously not the case.

also US Cape leads to an increase of the coefficient associated with domestic CAPE to -0.16 with a t-statistic of -7.9. The coefficient on US CAPE itself is small and insignificant suggesting that US valuation levels do not add information beyond domestic CAPE.[16]

5.4.5 Industry-Level Evidence

In a similar spirit to the robustness tests using international data, we also repeat our main analysis using the 30 US industry portfolios from Ken French's data library as the final robustness test. This industry-level data is available since July 1926. When we estimate our models using pooled data as well as pooled data with industry fixed effects, we confirm our earlier results at the market-level and find results that suggest an even more pronounced negative association between valuation levels and the skewness of future returns.

Panel B of Table 13 summarizes the results when controlling for industry fixed effects. It shows that the Conditional-Skew-T model outperforms the Symmetric-T and the Constant-Skew-T models in terms of likelihood. The coefficient of interest, $\beta_{1,\gamma}$, is estimated to equal -0.117, with a t-statistic of -5.5., suggesting a pronounced and statistically significant negative association between valuation levels and skewness even in the case of industry-based portfolios.

6 Conclusion

In this paper, we document a robust link between valuation levels and the shape of the distribution of cumulative (up to 24 months) total log returns in the SP500 and in international equity indices. Our key result is that return distributions become considerably more asymmetric and negatively skewed when valuation levels are high; in contrast they tend to be symmetric, sometimes even slightly positively skewed, when valuation levels are low. These patterns are very robust across return horizons, proxies for valuation levels and sample periods.

While the emphasis of the literature is usually on predicting expected rates of return (point prediction), we focus on the novel and important question of how asset prices actually revert back to these time-varying means. Our empirical results indicate that this reversion is rather smooth and gradual when valuation levels are low and potentially abrupt when valuation levels are high.

The dependence of the shape of the return distribution on valuation levels has several further interesting practical implications. Most importantly, it implies that

[16]We abstain from reporting all those results to avoid overwhelming the reader with similar, and thus somewhat repetitive, results. Detailed results are available from the authors upon request.

measures of risk (e.g., standard deviation, value-at-risk, expected-tail-loss), derived from symmetric distributions or distributions with constant skewness, are underestimated when valuation levels are high and overestimated when valuation levels are low relative to a model with conditional skewness. This indicates a lose-lose situation for risk managers and asset managers relying on these risk measures. Importantly, magnitudes of these deviations are sizable.

Another noteworthy result of our empirical analysis is that we find the mode of the return distribution to be consistently positive and essentially unaffected by valuation levels. This implies that even when valuation levels are extremely high, the most likely return over the next 12 to 24 months remains positive reflecting the well-known difficulty of predicting turning points and market downturns. Overall, our empirical evidence on how valuations affect the asymmetry of the predictive distribution of returns is qualitatively consistent with stochastic rational bubbles in the spirit of Blanchard and Watson (1982). However, expected returns are constant in models of rational bubbles, whereas the introduction of time varying skewness does nothing to change the relation between valuations and expected returns that has been the focus of so much attention in the literature. The finding that the mode of the predictive distribution (the most likely outcome) does not change and remains positive even at extremely high valuations may however provide useful insights into why such valuations can be reached at all.

References

Arnold, B. C., Groeneveld, R. A., 2010. Measuring skewness with respect to the mode. The American Statistician 49 (1), 34–38.

Asness, C., Frazzini, A., Gormsen, N. J., Pedersen, L. H., 2018. Betting against correlation: Testing theories of the low-risk effect. Journal of Financial Economics p. forthcoming.

Asness, C. S., Moskowitz, T. J., Pedersen, L. H., 2013. Value and momentum everywhere. The Journal of Finance 68, 929–985.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. The Review of Financial Studies 13, 1–42.

Bessembinder, H., 2017. Do stocks outperform treasury bills. The Journal of Financial Economics Forthcoming.

Blanchard, O. J., Watson, M. W., 1982. Bubbles, rational expectations, and financial markets. In: Crises in Economic and Financial Structure. Lexington Books 12, 295–315.

Brunnermeier, M. K., Nagel, S., Pedersen, L. H., 2008. Carry trades and currency crashes. NBER Macroeconomics Annual 23, 313–347.

Brunnermeier, M. K., Pedersen, L. H., 2009. Market liquidity and funding liquidity. The Review of Financial Studies 22, 2201–2238.

Campbell, J. Y., Hentschel, L., 1992. No news is good news: An asymmetric modelof changing volatility in stock returns. Journal of Financial Economics 31(3), 281–318.

Campbell, J. Y., Shiller, R. J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies 1, 195–228.

Chen, J., Hong, H., Stein, J. C., 2001. Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. The Journal of Financial Economics 61, 345–381.

Cochrane, J. H., 2008. The dog that did not bark: A defense of return predictability. The Review of Financial Studies 21 (4), 1533–1575.

Cochrane, J. H., 2011. Presidential address: Discount rates. The Journal of Finance LXVI (4), 1047–1108.

David, A., Veronesi, P., 2014. Investor's and central bank's uncertainty embedded in index options. The Review of Financial Studies 27 (6), 1661–1716.

Duffee, G. R., 1995. Stock returns and volatility a firm-level analysis. Journal of Financial Economics 37(3), 399–420.

Fama, E. F., 2013. Two pillars of asset pricing, nobel Prize Lecture.

Fama, E. F., French, K. R., 2017. Long-horizon returns, working paper.

Fernandez, C., Steel, M. F. J., 1998. On bayesian modeling of fat tails and skewness. The Journal of the American Statistical Association 93 (441), 359–371.

Gamerman, D., 1997. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing 7, 57–68.

Golez, B., Koudijs, P., 2016. Four centuries of return predictability, working Paper.

Gormsen, N. J., Jensen, C. S., 2017. Higher-moment risk, working paper.

Greenwald, D. L., Lettau, M., Ludvigson, S. C., 2016. Origins of stock market fluctuations, working paper.

Greenwood, R., Shleifer, A., You, Y., 2019. Bubbles for fama.

Harvey, C. R., Siddique, A., 1999. Autoregressive conditional skewness. The Journal of Financial and Quantitative Analysis 43 (4), 465–487.

Henkel, S. J., Martin, J. S., Nardari, F., 2011. Time-varying short-horizon predictability. Journal of Financial Economics 99, 560 – 580.

Hueng, J. C., McDonald, J. B., 2005. Forecasting asymmetries in aggregate stock market returns: Evidence from condtional skewness. The Journal of Empirical Finance 12, 666–685.

Hussman, J. P., 2017. Valuations, sufficient statistics, and breathtaking risks, market comment at www.hussmanfunds.com.

Johansson, A. S., 2019. Characterizing the tail-risk of factor mimicking portfolios, working paper.

Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. The Journal of Economic Dynamics & Control 27, 1699–1737.

Jones, M., 2014. On families of distributions with shape parameters. International Statistical Review 83 (2), 175–192.

Jorda, O., Schularick, M., Taylor, A. M., 2016. Macrofinancial history and the new business cycle facts, working paper.

Lee, S. X., McLachlan, G. J., 2013. On mixtures of skew normal and skew t-distributions. Advances in Data Analysis and Classification 7 (3), 241–266.

Li, F., Villani, M., 2010. Flexible modelling of conditional distributions using smooth mixtures of asymmetric student t densities. The Journal of Statistical Planning and Inference 140, 3638–3654.

Moreira, A., Muir, T., 2017. Volatility managed portfolios. The Journal of Finance LXXII, 1611–1644.

Neuberger, A., 2012. Realized skewness. The Review of Financial Studies 25 (11), 3424–3455.

Neuberger, A., Payne, R., 2018. The skewness of the stock market at long horizons, working paper.

Oh, S., Wachter, J., 2019. Cross-sectional skewness, working paper.

Poterba, J. M., Summers, L. H., 1986. The persistence of volatility and stock market fluctuations. The American Economic Review 76, 1142–1151.

Veldkamp, L. L., 2005. Slow boom, sudden crash. Journal of Economic Theory 124, 230 – 257.

Appendix A Skewness in Cumulative Returns versus Skewness in Daily Returns

The approach taken in our paper is to model the asymmetry (or skewness) in the distribution of cumulative log returns directly. A more common approach is to build rolling measures of skewness using one or two quarters of daily data. We now show that the connection between the two measures of skewness may be weak or even non-existent, and argue that the better approach is to measure skewness in cumulative returns at the horizon of interest.

Chen, Hong, and Stein (2001) build in-sample measures of skewness in a cross section of daily returns (six-months periods), and then regress these on trading volume, recent returns, and valuation ratios. The skew at time t is estimated from the last six months of daily returns as

$$SKEW_t = \frac{\sum r_{t-i}^3}{(\sum r_{t-i}^2)^{3/2}}.$$

Brunnermeier, Nagel and Pedersen (2009) compute a similar measure for currencies, and regress it on interest rate differentials. Both papers find some relation between this measure of skewness and explanatory variables. However, we see two shortcomings of this approach. The first problem is that this estimate of the the Pearson's moment skewness has (at best) a very high variance for financial market returns, which are strongly leptokurtic. The second problem is that the distribution of cumulative returns one month or one year ahead is empirically far more asymmetric than the distribution of daily returns, at least for an equity index like the SP500. In fact, the vast majority of GARCH and stochastic volatility models with leverage (stronger effect on future variance of negative shocks compared to positive shocks) produce symmetric daily log returns but asymmetric cumulative returns at longer horizons, so that a measure of skew built from daily return has an expected value of zero.

To illustrate, let's consider a plain-vanilla stochastic volatility model with leverage,

$$\begin{aligned} r_t &= \beta e^{0.5x_t}\varepsilon_t, \\ x_{t+1} &= \phi x_t + \rho\sigma\varepsilon_t + \sigma\sqrt{1-\rho^2}u_t \\ \varepsilon_t &\sim N(0,1),\ u_t \sim N(0,1), \end{aligned}$$

where $\rho < 0$ implies a leverage effect. By setting ρ appropriately it is possible to obtain strongly skewed distribution of cumulative returns (see Figure A1 for an example with an horizon of $h = 60$), yet the distribution of daily returns is leptokurtik but always symmetric, and a measure of skew built as $\sum_{i=1}^{h} r_{t-h}^3 / (\sum_{h=1}^{h} r_{t-h}^2)^{3/2}$ has an expected value of zero (with high variance).

Figure A.1: **Skewness of daily returns versus 60-day cumulative returns**
The figure shows the distributions of daily log returns and 60-day cumulative returns generated from a simple stochastic volatility model.

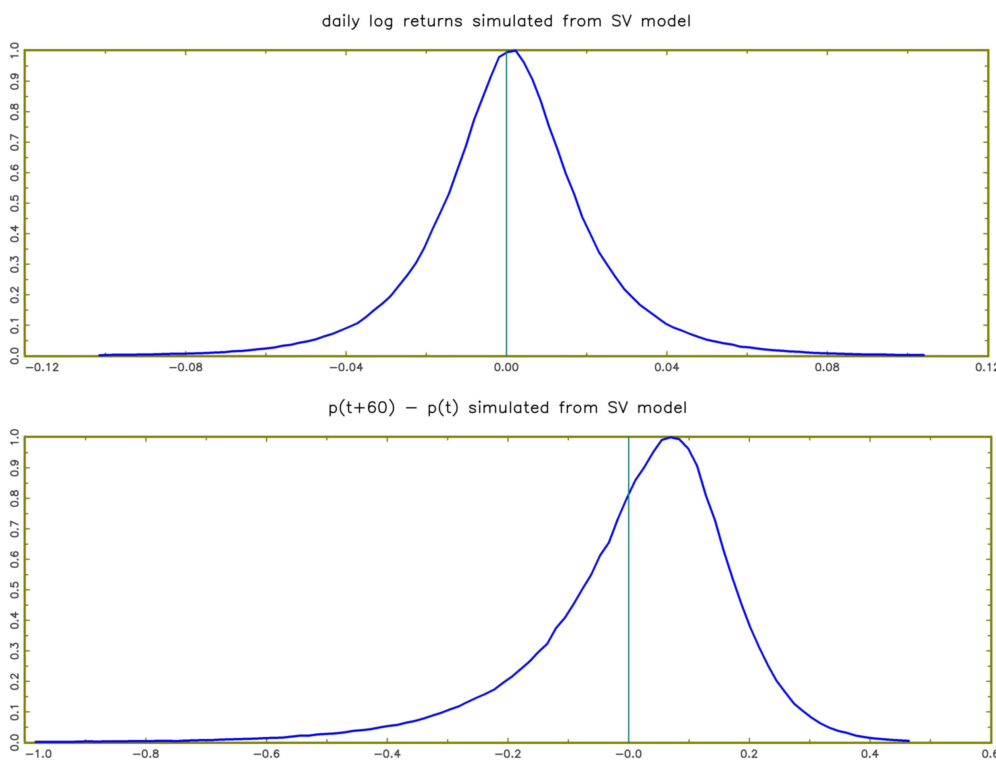


Figure 1: **Market-to-book and Cyclically-Adjusted-Price-Earnings (CAPE) Ratios**

The figure shows the standardized — mean equal to zero, standard deviation equal to 1 — log market-to-book (dashed line) and log cyclically-adjusted-price-earnings ratio (solid line).

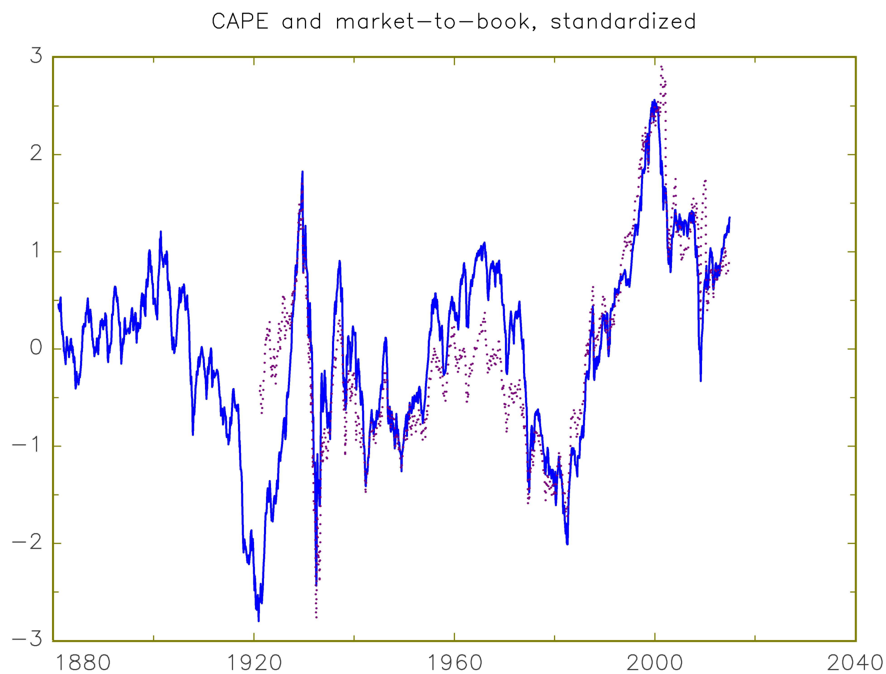


Table 1: **Summary Statistics**

This table provides summary statistics — means, standard deviations and skewness (the normalized third moment) — of 12-month and 24-month total equity returns. All variables are in logs. In addition to unconditional estimates, we also report those statistics conditional on valuation quartiles. We further report those statistics separately for the pre-1945 and post-1945 periods. Means and standard deviations are reported in percentage terms.

Panel A: 12-month returns

	Full Sample			**Pre-1945**			**Post-1945**		
	Mean (%)	Std. Dev. (%)	Skew.	Mean (%)	Std. Dev. (%)	Skew.	Mean (%)	Std. Dev. (%)	Skew.
All Returns	8.67	18.75	-0.83	6.34	21.65	-0.66	10.40	15.65	-0.82
1st Valuation Quartile	16.21	16.85	0.30	16.80	20.65	0.35	15.47	13.03	-0.15
2nd & 3rd Valuation Quartile	7.30	18.21	-1.25	5.05	18.86	-1.56	10.46	14.59	-0.56
4th Valuation Quartile	3.88	19.43	-0.95	-1.47	23.85	-0.50	5.24	18.29	-1.02

Panel B: 24-month returns

	Full Sample			**Pre-1945**			**Post-1945**		
	Mean (%)	Std. Dev. (%)	Skew.	Mean (%)	Std. Dev. (%)	Skew.	Mean (%)	Std. Dev. (%)	Skew.
All Returns	17.33	26.28	-1.13	12.31	29.95	-1.09	20.92	21.89	-0.74
1st Valuation Quartile	31.70	18.60	-0.13	30.93	21.42	-0.03	30.86	17.29	0.13
2nd & 3rd Valuation Quartile	15.17	22.58	-0.37	12.27	22.52	0.23	21.61	17.76	-0.69
4th Valuation Quartile	7.30	32.90	-1.41	-6.15	37.88	-1.26	9.66	27.64	-0.38

Table 2: **Model Parameters when Predicting 12-month Returns**
This table provides parameter estimates of three different models — the Symmetric-T Model, the Constant-Skew-T Model and the Conditional-Skew-T Model — when 12-month returns are used as dependent variable. All variables are in logs. The table reports mean parameter estimates and corresponding t-values.

Panel A: Full Sample

	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	9.686	8.707	16.825	7.122	16.233	6.480
$\beta_{1,m}$	-4.226	-3.721	-4.185	-4.007		
$\beta_{0,\sigma}$	2.724	38.296	2.668	34.686	2.680	35.343
$\beta_{0,\nu}$	2.025	4.666	2.163	4.384	2.407	4.058
$\beta_{0,\gamma}$			-0.316	-3.195	-0.297	-2.900
$\beta_{1,\gamma}$					-0.175	-4.254
Log-Likeli.	-6790.7		-6756.7		-6752.6	

Panel B: Pre-1945

	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	7.081	3.795	11.356	2.708	10.256	2.180
$\beta_{1,m}$	-6.029	-3.117	-6.081	-3.159		
$\beta_{0,\sigma}$	2.849	25.523	2.846	26.119	2.853	27.924
$\beta_{0,\nu}$	2.062	3.059	2.229	2.981	2.476	3.122
$\beta_{0,\gamma}$			-0.160	-1.124	-0.124	-0.790
$\beta_{1,\gamma}$					-0.197	-3.345
Log-Likeli.	-3230.6		-3227.5		-3225.8	

Panel C: Post-1945

	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	11.081	8.078	19.543	6.452	18.420	6.344
$\beta_{1,m}$	-4.083	-3.007	-3.903	-3.343		
$\beta_{0,\sigma}$	2.617	27.417	2.506	23.661	2.516	25.639
$\beta_{0,\nu}$	2.649	3.216	2.984	3.493	3.048	3.576
$\beta_{0,\gamma}$			-0.433	-2.833	-0.378	-2.632
$\beta_{1,\gamma}$					-0.190	-3.368
Log-Likeli.	-3357.4		-3328.3		-3326.8	

Table 3: **Model Parameters when Predicting 24-month Returns**
This table provides parameter estimates of three different models — the Symmetric-T Model, the Constant-Skew-T Model and the Conditional-Skew-T Model — when 24-month returns are used as dependent variable. All variables are in logs. The table reports mean parameter estimates and corresponding t-values.

Panel A: Full sample						
	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	18.395	8.874	25.350	5.328	27.230	5.756
$\beta_{1,m}$	-8.623	-4.030	-8.032	-3.765		
$\beta_{0,\sigma}$	3.020	29.884	3.006	29.644	2.994	31.445
$\beta_{0,v}$	2.059	3.307	2.281	3.154	2.658	3.276
$\beta_{0,\gamma}$			-0.219	-1.583	-0.263	-1.895
$\beta_{1,\gamma}$					-0.240	-4.191
Log-Likeli.	-7201.8		-7188.8		-7174.5	

Panel B: Pre-1945						
	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	13.137	3.796	16.208	1.920	18.675	2.072
$\beta_{1,m}$	-11.276	-3.142	-11.080	-3.004		
$\beta_{0,\sigma}$	3.116	20.491	3.098	20.313	3.094	22.220
$\beta_{0,v}$	2.081	2.502	2.121	2.373	2.602	2.751
$\beta_{0,\gamma}$			-0.079	-0.359	-0.133	-0.565
$\beta_{1,\gamma}$					-0.299	-3.314
Log-Likeli.	-3366.8		-3366.5		-3362.2	

Panel C: Post-1945						
	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	21.364	8.603	24.906	4.482	28.090	4.707
$\beta_{1,m}$	-8.369	-3.210	-7.927	-2.932		
$\beta_{0,\sigma}$	2.907	21.366	2.892	22.477	2.851	22.012
$\beta_{0,v}$	2.694	2.662	2.728	2.767	2.966	2.994
$\beta_{0,\gamma}$			-0.129	-0.719	-0.209	-1.025
$\beta_{1,\gamma}$					-0.255	-3.247
Log-Likeli.	-3540.5		-3536.9		-3524.3	

Table 4: **Predictive Distributions (Full Sample)**

This table provides means, standard deviations (SD), the normalized third moment (skewness, SKEW), the probability mass below the mode (asymmetry, ASY), the 1% Value-at-Risk and the 1% Expected Tail Loss of predictive distributions of 12-month and 24-month total equity returns implied by our models. Parameter estimates are based on the full sample. All variables are in logs. All values reported are in percentage terms except for skewness. Specifically, we report those characteristics separately for high (+2 standard deviations) and low (-2 standard deviations) valuation levels. We further report those statistics separately for the Symmetric-T Model (Panel A), the Constant-Skew-T Model (Panel B) and the Conditional-Skew-T Model (Panel C).

Panel A: Symmetric-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	18.13	18.31	0.00	50.00	-28.26	-40.76	35.46	25.01	0.00	50.00	-28.30	-46.68
High valuation	1.28	18.31	0.00	50.00	-45.09	-57.22	1.23	25.01	0.00	50.00	-62.53	-80.02

Panel B: Constant-Skew-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	17.06	18.43	-0.75	65.29	-36.83	-51.83	33.57	25.32	-0.56	60.78	-37.59	-58.27
High valuation	0.29	18.43	-0.75	65.29	-53.35	-67.74	1.40	25.32	-0.56	60.78	-69.21	-88.69

Panel C: Conditional-Skew-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	17.47	16.60	0.00	47.35	-23.04	-32.74	34.14	22.74	0.16	39.32	-17.47	-28.99
High valuation	-1.13	21.68	-0.97	78.48	-70.76	-86.44	-0.64	32.09	-1.05	81.55	-103.37	-125.70

Table 5: **Predictive Distributions (Pre-1945)**

This table provides means, standard deviations (SD), the normalized third moment (skewness, SKEW), the probability mass below the mode (asymmetry, ASY), the 1% Value-at-Risk and the 1% Expected Tail Loss of predictive distributions of 12-month and 24-month total equity returns implied by our models. Parameter estimates are based on the pre-1945 sample. All variables are in logs. All values reported are in percentage terms except for skewness. Specifically, we report those characteristics separately for high (+2 standard deviations) and low (-2 standard deviations) valuation levels. We further report those statistics separately for the Symmetric-T Model (Panel A), the Constant-Skew-T Model (Panel B), and the Conditional-Skew-T Model (Panel C).

Panel A: Symmetric-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	19.12	21.29	0.00	50.00	-35.08	-50.32	35.54	28.80	0.00	50.00	-38.96	-62.25
High valuation	-4.98	21.29	0.00	50.00	-59.10	-73.12	-9.63	28.80	0.00	50.00	-83.97	-104.37

Panel B: Constant-Skew-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	18.46	21.48	-0.37	57.93	-40.66	-57.22	34.58	29.40	-0.31	53.94	-45.32	-69.26
High valuation	-5.51	21.48	-0.37	57.93	-64.24	-79.22	-9.20	29.40	-0.31	53.94	-88.53	-109.58

Panel C: Conditional-Skew-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	18.34	21.33	0.46	36.82	-27.65	-38.50	36.76	30.65	0.66	28.29	-22.38	-35.88
High valuation	-5.93	24.18	-0.83	73.81	-81.73	-97.47	-11.82	36.89	-1.06	81.18	-130.87	-150.72

Table 6: **Predictive Distributions (Post-1945)**

This table provides means, standard deviations (SD), the normalized third moment (skewness, SKEW), the probability mass below the mode (asymmetry, ASY), the 1% Value-at-Risk and the 1% Expected Tail Loss of predictive distributions of 12-month and 24-month total equity returns implied by our models. Parameter estimates are based on the post-1945 sample. All variables are in logs. All values reported are in percentage terms except for skewness. Specifically, we report those characteristics separately for high (+2 standard deviations) and low (-2 standard deviations) valuation levels. We further report those statistics separately for the Symmetric-T Model (Panel A), the Constant-Skew-T Model (Panel B) and the Conditional-Skew-T Model (Panel C).

Panel A: Symmetric-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	19.27	15.52	0.00	50.00	-19.05	-27.62	38.13	21.39	0.00	50.00	-15.11	-27.54
High valuation	2.90	15.52	0.00	50.00	-35.45	-43.23	4.54	21.39	0.00	50.00	-48.35	-57.92

Panel B: Constant-Skew-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	18.22	15.62	-0.71	70.39	-26.74	-36.66	36.61	21.76	-0.28	56.41	-22.05	-35.87
High valuation	2.58	15.62	-0.71	70.39	-41.85	-50.40	5.23	21.76	-0.28	56.41	-51.43	-61.31

Panel C: Conditional-Skew-T Model

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	18.22	13.44	-0.19	49.90	-15.69	-23.19	37.08	21.41	0.44	35.39	-8.14	-18.11
High valuation	1.93	18.11	-0.72	81.99	-55.14	-62.05	5.91	25.86	-0.75	80.81	-68.11	-74.42

Table 7: **Robustness Test: Non-overlapping Data**

This table provides parameter estimates of three different models — the Symmetric-T Model, the Constant-Skew-T Model and the Conditional-Skew-T Model — when 12-month (Panel A) or 24-month (Panel B) non-overlapping returns are used as dependent variables. The results are based on returns from January to January. All variables are in logs. The table reports mean parameter estimates and corresponding t-values.

Panel A: 12-Month Returns

	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	9.332	6.416	17.697	5.143	18.172	5.399
$\beta_{1,m}$	-3.994	-2.820	-3.761	-2.778		
$\beta_{0,\sigma}$	2.727	30.045	2.659	28.523	2.639	27.470
$\beta_{0,v}$	2.638	3.250	3.096	3.658	3.132	3.758
$\beta_{0,\gamma}$			-0.374	-2.527	-0.400	-2.702
$\beta_{1,\gamma}$					-0.163	-2.823
Log-Likeli.	-575.2		-571.5		-571.4	

Panel B: 24-Month Returns

	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	17.932	5.838	26.055	4.418	27.795	4.467
$\beta_{1,m}$	-8.906	-2.846	-7.196	-2.520		
$\beta_{0,\sigma}$	3.119	26.152	3.053	23.033	3.024	23.195
$\beta_{0,v}$	2.743	3.085	2.694	2.964	2.812	3.123
$\beta_{0,\gamma}$			-0.254	-1.512	-0.287	-1.616
$\beta_{1,\gamma}$					-0.231	-2.751
Log-Likeli.	-308.5		-307.147		-306.9	

Table 8: **Robustness Test: Model Parameters when using Alternative Valuation Ratios**

This table provides parameter estimates of the Conditional-Skew-T Model when 12-month returns and 24-month returns are used as dependent variable. We consider three alternative valuation ratios: (i) the market-to-book ratio, (ii) the margin-adjusted CAPE, and (iii) the past 5-years of returns. All variables are in logs. All results summarized in the table are based on the full sample of data. The table reports mean parameter estimates and corresponding t-values.

Panel A: 12-Month Returns

	Market-to-Book		**Margin-adjusted CAPE**		**Past 5-Year Return**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	19.602	7.177	18.282	7.292	17.753	6.779
$\beta_{0,\sigma}$	2.639	26.396	2.643	27.202	2.686	36.323
$\beta_{0,v}$	1.924	3.732	2.533	3.346	2.595	4.143
$\beta_{0,\gamma}$	-0.376	-3.134	-0.391	-3.491	-0.351	-3.223
$\beta_{1,\gamma}$	-0.166	-3.111	-0.267	-4.826	-0.142	-3.697

Panel B: 24-Month Returns

	Market-to-Book		**Margin-adjusted CAPE**		**Past 5-Year Return**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	34.757	5.880	32.064	5.879	27.826	5.485
$\beta_{0,\sigma}$	2.999	22.768	2.933	23.015	3.022	31.221
$\beta_{0,v}$	2.362	2.691	2.883	2.897	2.602	3.190
$\beta_{0,\gamma}$	-0.393	-2.207	-0.398	-2.251	-0.274	-1.885
$\beta_{1,\gamma}$	-0.208	-2.857	-0.313	-4.362	-0.170	-2.965

Table 9: **Robustness Test: Predictive Distributions using Alternative Valuation Ratios as Predictive Variable (Full Sample)**

This table provides means, standard deviations (SD), the normalized third moment (skewness, SKEW), the probability mass below the mode (asymmetry, ASY), the 1% Value-at-Risk and the 1% Expected Tail Loss of predictive distributions of 12-month and 24-month total equity returns implied by the Conditional-Skew-T Model using three alternative valuation ratios (the market-to-book ratio, the margin-adjusted CAPE, and the past 5-year return) as predictive variable. Parameter estimates are based on the full sample. All variables are in logs. All values reported are in percentage terms except for skewness. Specifically, we report those characteristics separately for high (+2 standard deviations) and low (-2 standard deviations) valuation levels.

Panel A: Market-to-Book Ratio

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	18.51	17.77	-0.17	52.20	-28.75	-43.18	34.32	23.13	-0.48	48.85	-27.95	-46.29
High valuation	0.12	24.07	-1.25	80.47	-79.47	-96.50	3.04	35.61	-1.11	83.45	-110.67	-127.72

Panel B: Margin-adjusted CAPE

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	21.75	16.36	0.19	42.90	-16.03	-25.16	37.45	20.07	-0.16	38.79	-10.68	-22.76
High valuation	-10.35	28.30	-1.14	86.41	-100.37	-114.41	-11.74	41.69	-1.14	88.57	-144.35	-163.36

Panel C: Past 5-Year Return

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	3.21	20.34	-0.99	53.34	-58.16	-73.04	8.80	28.40	-1.01	45.36	-77.42	-99.70
High valuation	-0.69	22.43	-1.08	78.07	-70.17	-86.12	2.29	32.01	-1.13	76.39	-98.58	-122.24

Table 10: **Robustness Test: Model with Predictability in Dispersion**
This table provides parameter estimates of a model that allows for a link between return dispersion and valuation levels. We call that model the Conditional-Skew-T-Vola Model. All variables are in logs. Details on the estimation of these parameters are available from the authors upon request. The table reports mean parameter estimates and corresponding t-values.

Panel A: 12-Month Returns

	Full Sample		**Pre-1945**		**Post-1945**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	16.983	6.537	10.472	2.029	18.501	6.034
$\beta_{0,\sigma}$	2.669	34.104	2.854	28.374	2.514	24.159
$\beta_{1,\sigma}$	-0.053	-0.965	-0.013	-0.175	-0.016	-0.221
$\beta_{0,v}$	2.384	3.988	2.416	3.153	3.004	3.456
$\beta_{0,\gamma}$	-0.332	-3.051	-0.135	-0.780	-0.383	-2.474
$\beta_{1,\gamma}$	-0.188	-4.244	-0.203	-3.329	-0.197	-3.405
Log-Likeli.	-6749.7		-3225.8		-3326.8	

Panel B: 24-Month Returns

	Full Sample		**Pre-1945**		**Post-1945**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	25.729	5.550	14.529	1.451	25.052	4.129
$\beta_{0,\sigma}$	3.005	31.000	3.076	20.140	2.856	23.043
$\beta_{1,\sigma}$	0.080	1.064	0.108	0.844	0.152	1.420
$\beta_{0,v}$	2.618	3.171	2.372	2.574	2.893	3.054
$\beta_{0,\gamma}$	-0.211	-1.568	-0.007	-0.024	-0.087	-0.418
$\beta_{1,\gamma}$	-0.233	-4.057	-0.302	-2.890	-0.261	-3.055
Log-Likeli.	-7166.2		-3359.3		-3509.5	

Table 11: **Robustness Test: Predictive Distributions implied by the Conditional-Skew-T-Vola Model**

This table provides means, standard deviations (SD), the normalized third moment (skewness, SKEW), the probability mass below the mode (asymmetry, ASY), the 1% Value-at-Risk and the 1% Expected Tail Loss of predictive distributions of 12-month and 24-month total equity returns implied by our models. Parameter estimates are based on the full sample. All variables are in logs. All values reported are in percentage terms except for skewness. Specifically, we report those characteristics separately for high (+2 standard deviations) and low (-2 standard deviations) valuation levels.

	12-Month Returns						**24-Month Returns**					
	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL	Mean	SD	SKEW	ASY	1%-VaR	1%-ETL
Low valuation	18.23	18.65	0.03	47.80	-27.59	-39.05	32.48	19.82	0.15	37.52	-12.40	-23.20
High valuation	-1.80	21.82	-1.27	80.47	-71.84	-89.79	-8.62	40.38	-1.14	79.48	-138.06	-162.45

Table 12: **Robustness Test: Sample of international countries**
This table lists all countries included in our international sample including the start date of the observations, the end date and the number of months using in the estimation.

Country	Start Date	End Date	Months
aus	197906	201103	382
aut	199109	201103	235
bel	197906	201103	382
bra	199801	200401	69
can	196512	201103	544
che	197906	201103	382
dnk	198001	201103	375
esp	198911	201103	257
gbr	193711	201103	881
ger	197906	201103	382
grc	198701	201103	291
hkg	198212	201103	340
ind	199801	201103	159
isr	200905	201103	23
jap	196512	201103	544
kor	198402	201103	326
mys	198212	201103	340
nld	197907	201103	381
nor	198001	200009	249
nzl	199712	201103	160
pak	199801	200709	117
phl	199201	201103	231
rsa	197002	201103	494
sgp	198212	201103	340
swe	197906	201104	383
tai	199801	201103	159
tha	199712	201103	160
tur	199602	201103	182
ven	199801	200403	75

Table 13: **Robustness Test: Model Parameters for an International Sample and Industry-Level Portfolios in the US**

This table provides parameter estimates of three different models — the Symmetric-T Model, the Constant-Skew-T Model and the Conditional-Skew-T Model — for a sample of international stock market indices (Panel A) and a sample of industry-level portfolios in the US (Panel B). Overlapping 12-month returns are used as dependent variables (excess returns in the case of the international data). All variables are in logs. The table reports mean parameter estimates and corresponding t-values.

Panel A: International Sample

	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	0.742	1.097	6.751	4.494	6.897	5.119
$\beta_{1,m}$	-6.751	-10.986	-6.606	-9.764		
$\beta_{0,\sigma}$	3.086	107.694	3.058	90.631	3.031	95.162
$\beta_{0,v}$	2.067	11.613	2.054	11.394	2.034	10.869
$\beta_{0,\gamma}$			-0.180	-4.302	-0.187	-4.880
$\beta_{1,\gamma}$					-0.194	-10.869
Log-Likeli.	-40982		-40914		-40853	

Panel B: Industry-Level Analysis

	Symmetric-T		**Constant-Skew-T**		**Conditional-Skew-T**	
	Mean	t-stat	Mean	t-stat	Mean	t-stat
$\beta_{0,m}$	11.262	14.343	17.108	11.132	16.561	10.878
$\beta_{1,m}$	-3.358	-4.502	-3.665	-5.117		
$\beta_{0,\sigma}$	2.906	62.453	2.890	61.090	2.882	61.291
$\beta_{0,v}$	1.332	8.770	1.404	8.479	1.396	8.537
$\beta_{0,\gamma}$			-0.218	-4.298	-0.206	-4.032
$\beta_{1,\gamma}$					-0.117	-5.487
Log-Likeli.	-23515		-23453		-23444	

Figure 2: **Histograms of Realized 12-month Returns**
The figure shows four histograms of realized 12-month returns. All returns are in logs. The left graph in the top row shows the full sample unconditional distribution of realized 12-month returns. The remaining graphs show, in clock-wise direction, the full sample distributions conditional on being in the (i) lowest valuation quartile, (ii) the top valuation quartile, and (iii) the two middle valuation quartiles.

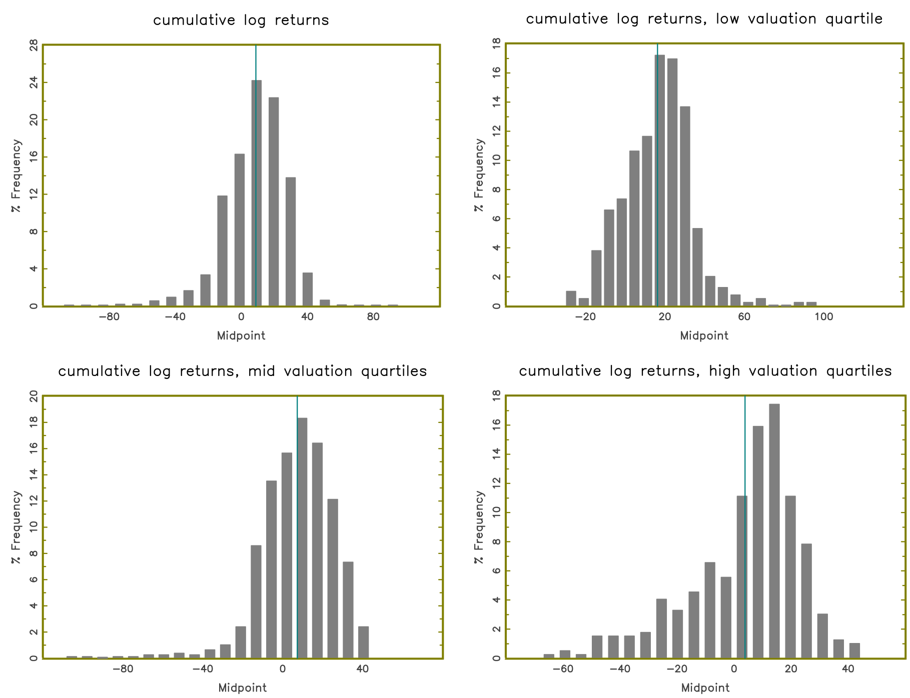


Figure 3: **Histograms of Realized 24-month Returns**
The figure shows four histograms of realized 24-month returns. All returns are in logs. The left graph in the top row shows the full sample unconditional distribution of realized 24-month returns. The remaining graphs show, in clock-wise direction, the full sample distributions conditional on being in the (i) lowest valuation quartile, (ii) the top valuation quartile, and (iii) the two middle valuation quartiles.

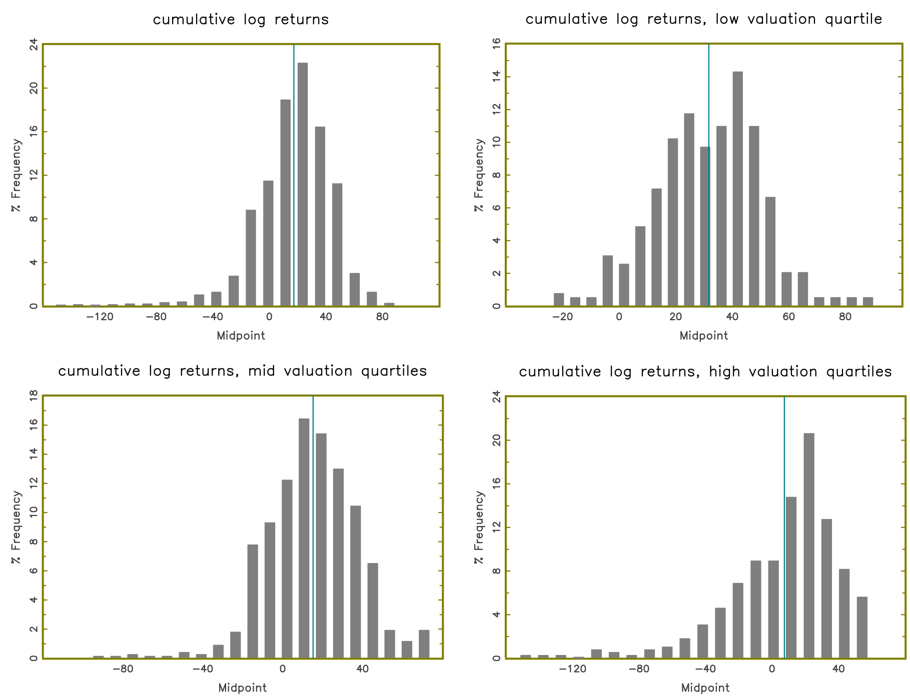


Figure 4: **Term-Structure of Skewness**

The figure shows the term-structure of skewness for cumulative returns up to 35 months. The skewness measure is the Arnold and Groeneveld measure of skewness (see Arnold and Groeneveld (2010)) from skew-t distributions calibrated to the empirical data. The graph shows results for the full unconditional sample as well as for samples conditional on being in the bottom or the top valuation quartile.

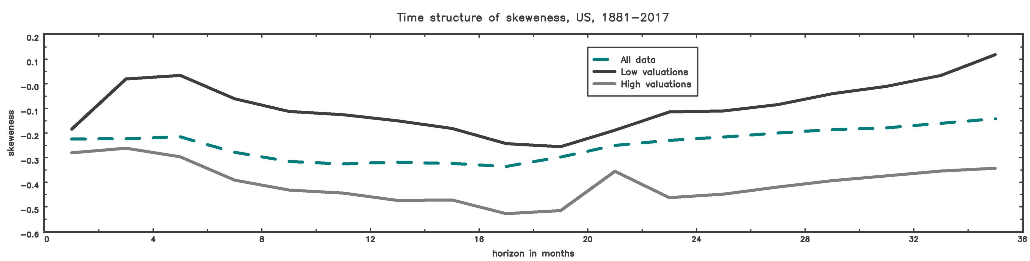


Figure 5: **Model-implied Conditional 12-month Return Distribution**
The figure shows the model-implied return distributions of 12-month returns for low (two standard deviations below the mean) and high (two standard deviations above the mean) valuation levels. All returns are in logs. The parameters governing the distributions are summarized in Table 2.

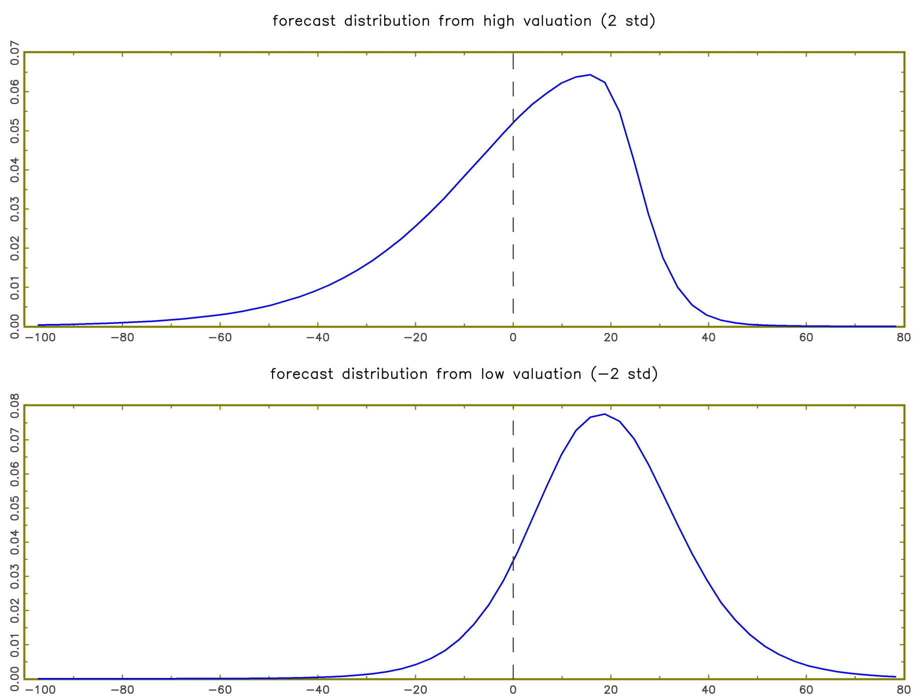


Figure 6: **Model-implied Conditional 24-month Return Distribution**
The figure shows the model-implied return distributions of 24-month returns for low (two standard deviations below the mean) and high (two standard deviations above the mean) valuation levels. All returns are in logs. The parameters governing the distributions are summarized in Table 3.

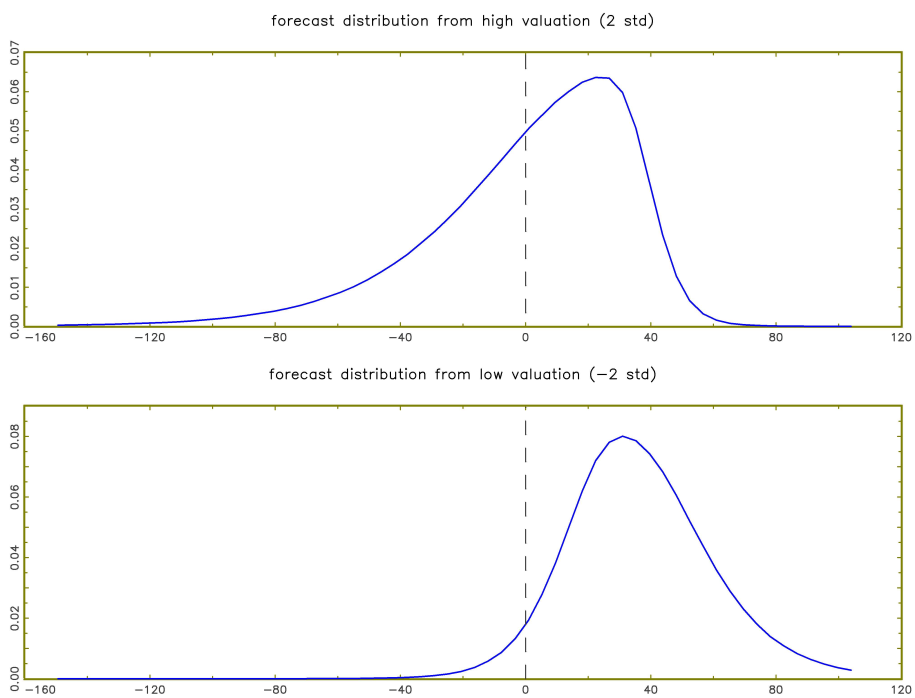
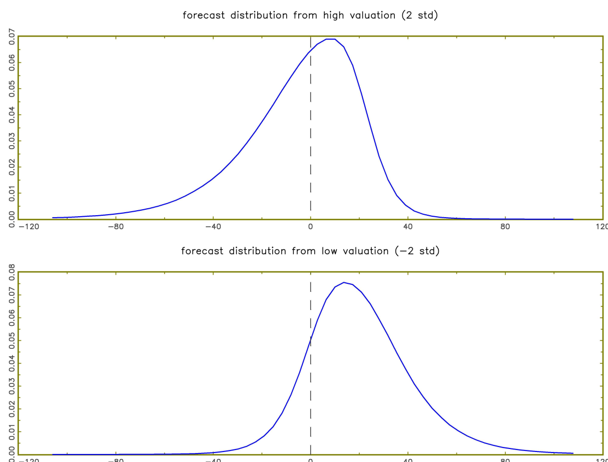


Figure 7: **Sub-sample Results: Model-implied Conditional 12-month Return Distribution**
The figure shows the model-implied return distributions of 12-month returns for low (two standard deviations below the mean) and high (two standard deviations above the mean) valuation levels separately for the pre-1945 and the post-1945 sample periods. All returns are in logs. The parameters governing the distributions are summarized in Table 2.

Panel A. Pre-1945



Panel B. Post-1945

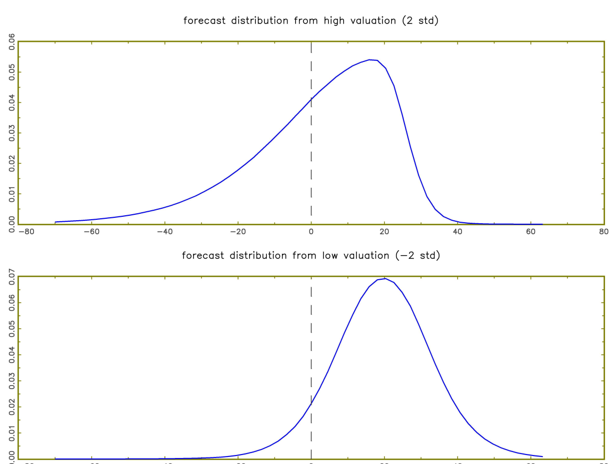
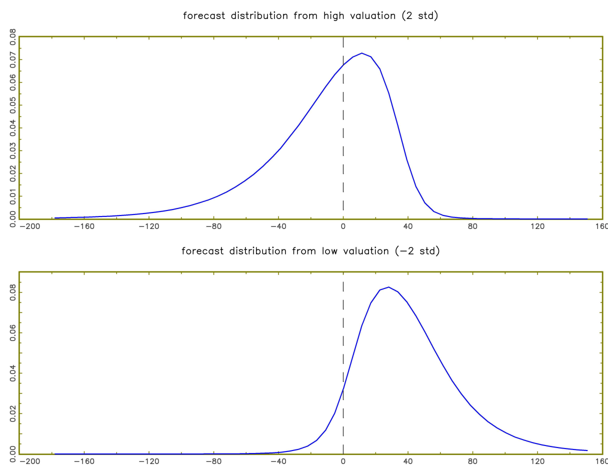


Figure 8: **Sub-sample results: model-implied, conditional 24-month return distribution**

The figure shows the model-implied return distributions of 24-month returns for low (two standard deviations below the mean) and high (two standard deviations above the mean) valuation levels separately for the pre-1945 and the post-1945 sample periods. All returns are in logs. The parameters governing the distributions are summarized in Table 3.

Panel A. Pre-1945



Panel B. Post-1945

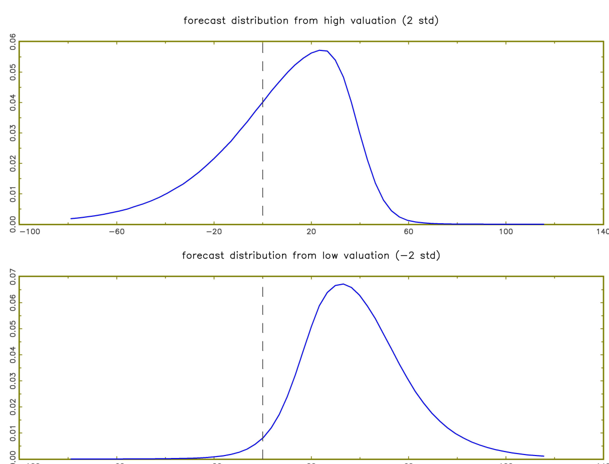


Figure 9: **Recursive Estimates of Core Model Parameters**
The figure shows recursive estimates of $\beta_{0,\gamma}$ (labeled b4(1) in the figure) and $\beta_{1,\gamma}$ (labelled b4(2) in the figure) together with +2 standard deviation and -2 standard deviation bands for the Conditional-Skew-T model. The model is reestimated monthly using an expanding window. The period 1881 to 1945m1 is used as burn-in phase.

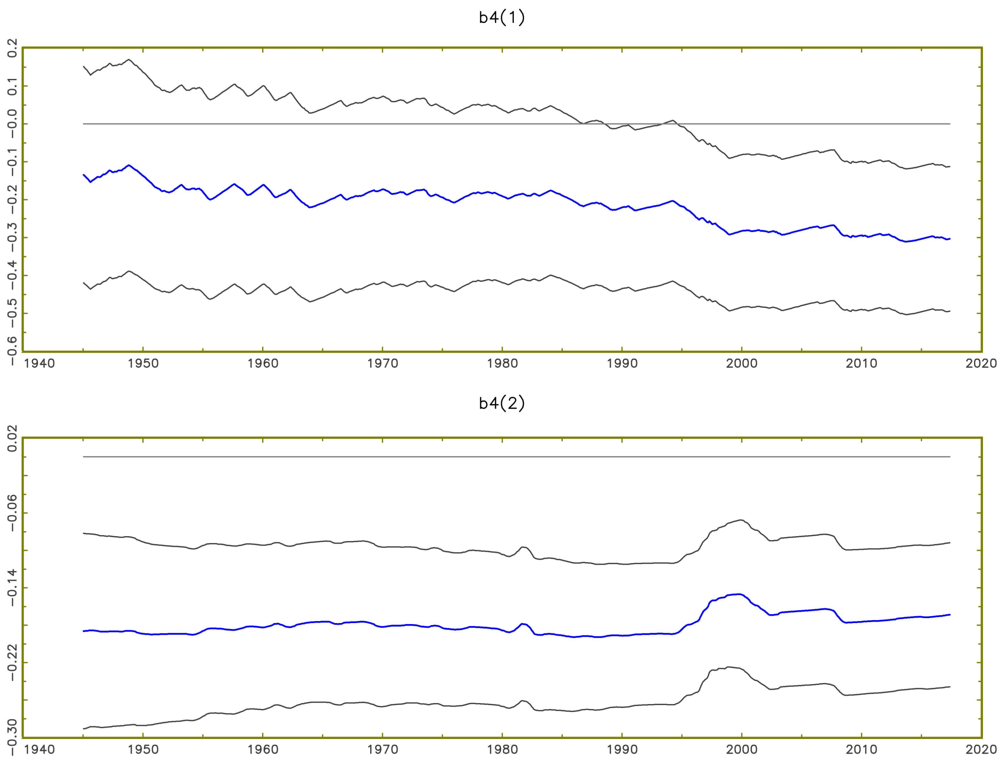


Figure 10: **Monthly Predictions of 12-Month Expected Returns**
The figure shows monthly predictions of 12-month returns using the Conditional-Skew-T Model over time. While the model is estimated using log-returns, the graph shows predicted simple returns.

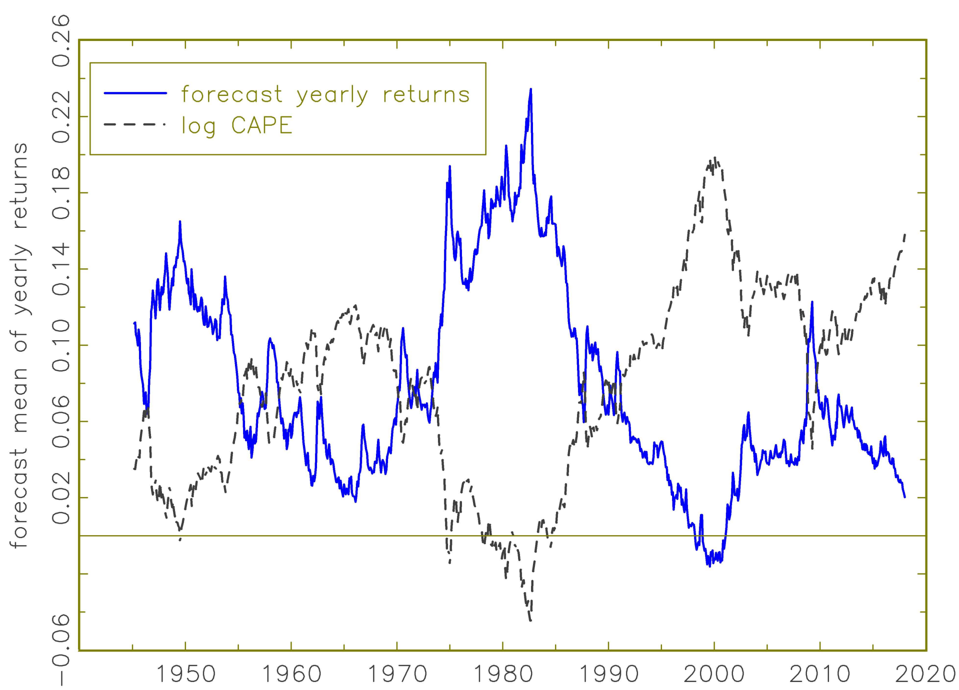


Figure 11: **Cumulative Differences in Squared Prediction Errors**
The figure shows cumulative differences in squared prediction errors. Panel A compares predictions from the Conditional-Skew-T model to predictions from the Symmetric-G model (dashed line) or from the Constant-Skew-T model (solid line). Panel B shows cumulative squared errors between the Conditional-Skew-T and the no-predictability model.

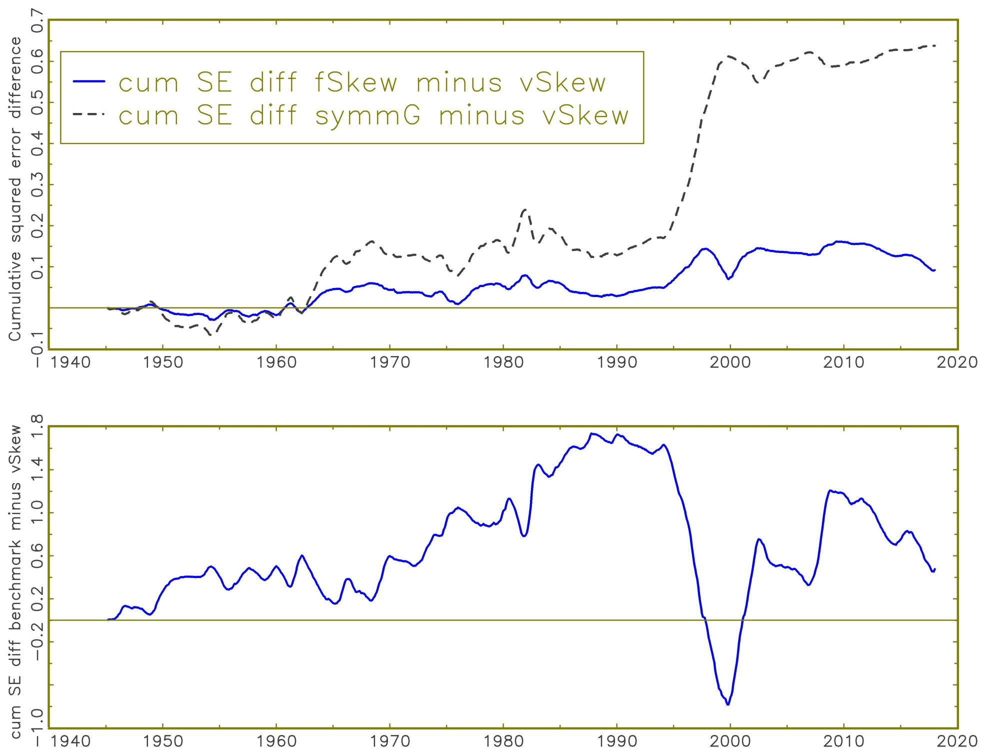


Figure 12: **Cumulative Utility Gains**

The figure shows cumulative utility gains. Panel A compares utility gains from the Conditional-Skew-T model to utility gains from the Symmetric-G model (dashed line) or from the Constant-Skew-T model (solid line). Panel B shows cumulative utility gains between the Conditional-Skew-T and the no-predictability model.

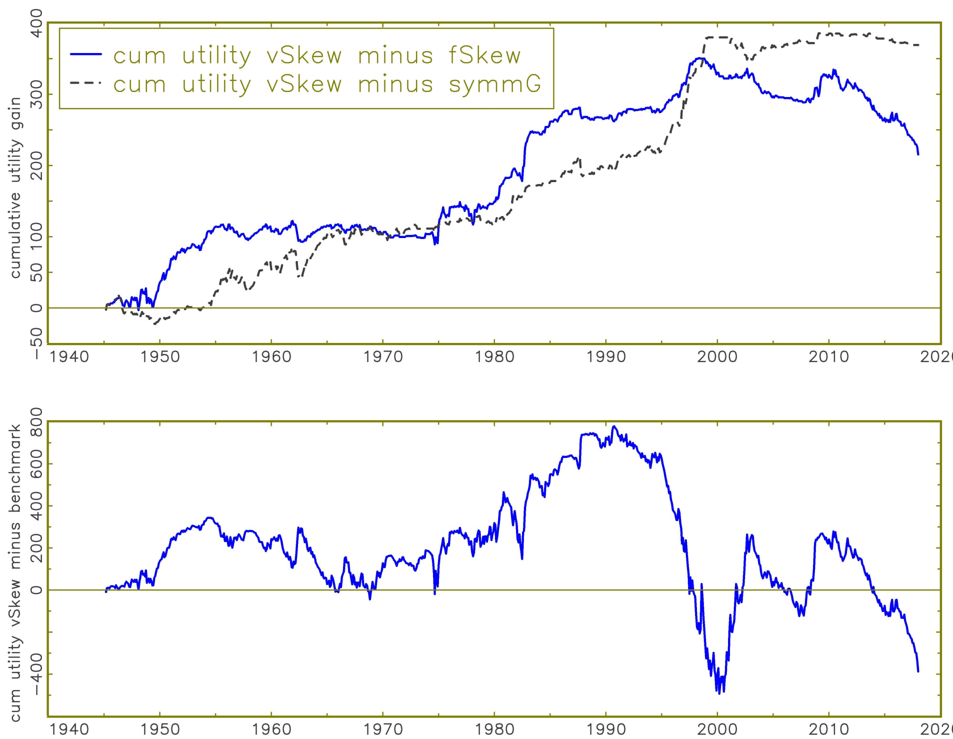


Figure 13: **Margin Debt and Cyclically-Adjusted-Price-Earnings (CAPE) Ratios**

The figure shows the standardized — mean equal to zero, standard deviation equal to 1 — margin debt and log cyclically-adjusted-price-earnings ratio.

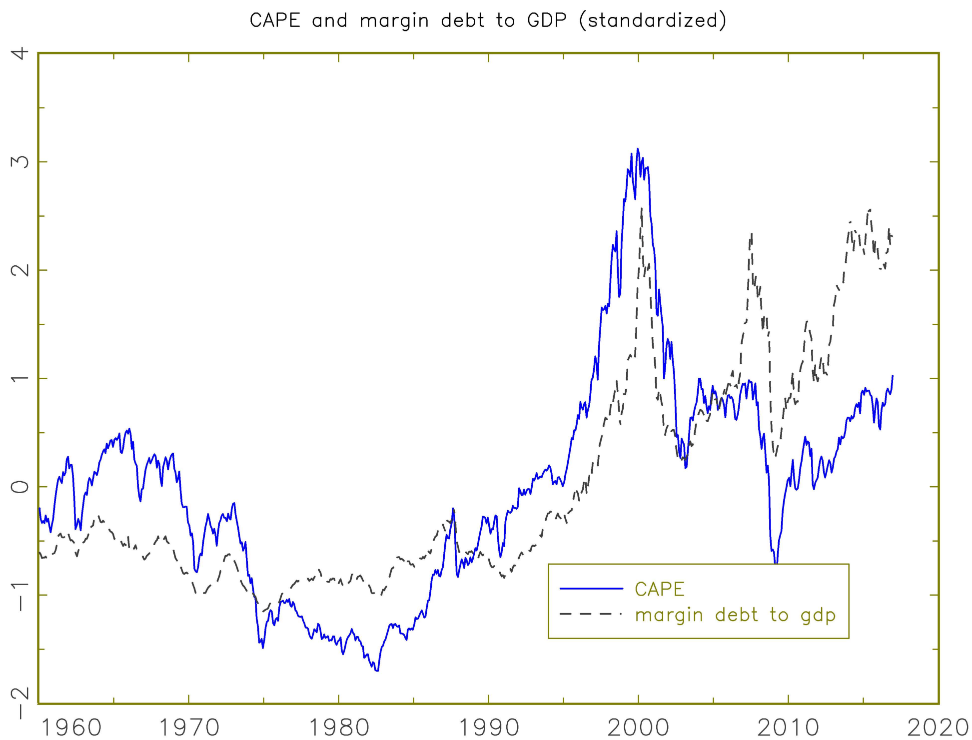
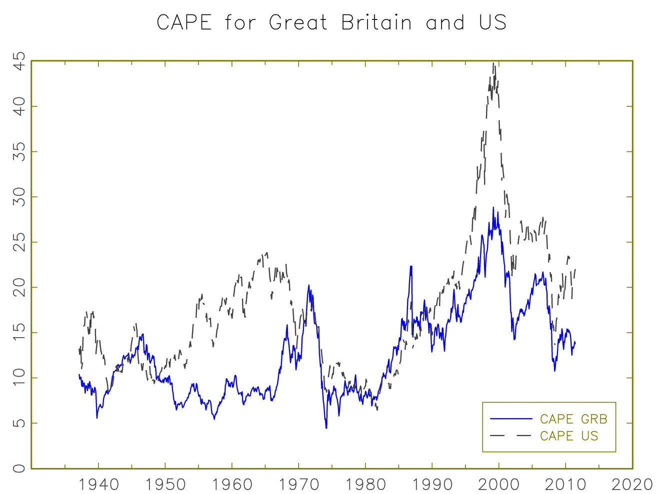


Figure 14: **Cyclically-Adjusted-Price-Earnings (CAPE) Ratios**
The figure shows the CAPE for the US and the UK (top panel) and for the US and the equal-weighted global portfolio (bottom panel).

Panel A. US versus UK CAPE



Panel B. US versus Global Portfolio CAPE