The Risk Return Tradeoff in the Long-Run: 1836-2003

Christian Lundblad*

October 2006
First Draft: October 2004

Abstract

Previous studies typically find a statistically insignificant relationship between the market risk premium and its expected volatility. Further, several of these studies estimate a negative risk return tradeoff contrary to the predictions of mainstream theory. Using simulations, I demonstrate that even 100 years of data constitute a small sample that may easily lead to this finding even though the true risk return tradeoff is positive. Small sample inference is plagued by the fact that conditional volatility has almost no explanatory power for realized returns. Using the nearly two century history of U.S. equity market returns from Schwert (1990), I estimate a positive and statistically significant risk return tradeoff across every specification considered. Finally, exploratory analysis suggests a role for a time-varying risk return tradeoff linked to the changing nature of the U.S. markets and economy over the historical record.

JEL Classification: G12, C22, C15

Keywords: Asset pricing, Risk return tradeoff, GARCH

* Kenan-Flagler Business School, University of North Carolina, Chapel Hill. I have benefited from conversations with Ravi Bansal, Geert Bekaert, Tim Oversleeve, Josh Coeval, Campbell Harvey, Philippic Orion, George Taches, and discussion participants at the 2005 EFT conference in Moscow, Duke University’s financial econometrics group, the University of Michigan, Wisconsin, UC-Irvine, Emory, the Board of Governors of the Federal Reserve, and the Batten Conference at William and Mary. I also express my appreciation to the editor, G. William Schwert, and an anonymous referee for significantly improving the paper. Send correspondence to: Christian Lundblad, Department of Finance, Kenan-Flagler Business School, University of North Carolina, Campus Box # 3490, McColl Building, Chapel Hill, NC 27599. E-mail: christian_lundblad@unc.edu.
1 Introduction

The risk return tradeoff is fundamental to finance. An implication of many asset pricing models is the tradeoff between the market’s risk premium and conditional volatility. However, using the popular GARCH-in-mean framework, the empirical evidence measured over the past 50 to 75 years of the U.S. market portfolio regarding this relationship is, at best, weak.\(^1\) The evidence appears to be particularly sensitive to the volatility specification, where GARCH models that facilitate what has often been called the “leverage effect”, i.e. an asymmetric relationship between news arrival and predictable volatility, (see, for example, Nelson (1991) and Glosten, Jagannathan, and Runkle (1993)) are associated with a negative relationship between excess returns and conditional volatility, contrary to the predictions of mainstream theory (though some researchers show that the theoretical inter-temporal mean variance tradeoff is not necessarily positive – see Abel (1988) and Backus and Gregory (1993), for example)\(^2\)

In this paper, I argue that the primary challenge in estimating the mean variance relationship is not the volatility specification but one of small samples, as conditional volatility has almost no explanatory power for future realized returns. Using Monte Carlo analysis, I show that within the GARCH-in-mean framework, one requires an extremely large amount of data in order to successfully detect the risk return tradeoff regardless of the precise volatility specification. In particular, the high degree of return volatility in comparison with the extremely persistent and smooth nature of the conditional volatility drives much of this finding. The weak empirical relationship may be viewed as a statistical artifact of the small samples from which one is forced to make inference, rather than as evidence against the postulated relationship. While the documented evidence is generally based upon the last 50+ years of equity market data, Monte Carlo evidence clearly shows that data span well beyond 100 years is required. If the explanatory power of conditional volatility for returns were more pronounced, we would not

\(^1\)For example, French, Schwert, and Stambaugh (1987), Chou (1988), Campbell and Hentschel (1993), and Bansal and Lundblad (2002) find a positive relationship between the expected excess market return and conditional variance, whereas Baillie and DeGennaro (1990), Nelson (1991), and Glosten, Jagannathan, and Runkle (1993) find the opposite.

\(^2\)It should also be noted that in the exploration of the risk return tradeoff, there are several important alternatives to the GARCH-M framework (which I do not consider). Using an instrumental variables specification for conditional moments, Campbell (1987), Harvey (1991), and Whitelaw (1994) find mixed evidence on the expected return volatility tradeoff. Using non-parametric techniques, Pagan and Hong (1991) find a weak negative relationship, but Harrison and Zhang (1999) find the relationship is significantly positive at longer horizons. Ghysels, Santa-Clara, and Valkanov (2005) employ a high-frequency measure of market volatility outside of the GARCH parametric setting, and uncover a positive relationship. For a specification that facilitates regime-switching, Whitelaw (2000) documents a negative unconditional link between the mean and variance of the market portfolio. As the evidence is mixed for these alternative specifications as well, Bekaert and Wu (2000) simply pin the volatility coefficient at the value implied by the returns’ unconditional moments.
require a large data set to detect it. Once we recognize this, the focus on the historical record (what is available to us to make inference) is critical. Using the nearly two century history of the U.S. equity market experience (see Schwert (1990) and Goetzmann, Ibbotson, and Peng (2001)), I shed new light on the mean variance tradeoff. In sharp contrast to the previous literature, I find that the measured risk return relationship is both positive and significant for the available U.S. history across a variety of volatility specifications even in the presence of low explanatory $R^2$’s. Further, I show that these findings are also corroborated by U.K. historical data.

While I do explore a number of competing volatility specifications, I am agnostic as to which provides the best fit to the data. First, many specifications are non-nested, so clean tests are not readily available. Second, and more important, each specification employed here can be found throughout the risk return literature, and all are reasonable candidates that have yielded conflicting results on this issue. Instead, I argue that an alternative explanation of the inconsistency of the small-sample results across a host of different volatility specifications may be driven by an inability to detect the equilibrium relationship over the shorter period given the limited explanatory power of conditional volatility for future returns in the data. For this issue, I find that the choice of volatility specification in the GARCH-M context is second-order given these extremely low $R^2$’s. In the context of the full historical record, the risk return evidence is remarkably consistent across the various specifications employed (and alternative historical records). I obtain a positive and significant risk return tradeoff regardless of the specification. In sum, this is not a study about the volatility specification; rather, I provide collective evidence on the mean-volatility tradeoff across a number of competing specifications, all of which yield an identical story. Over the full historical record, the risk return tradeoff is remarkably consistent and always statistically significant.

The remainder of the study is organized into five sections. Section 2 provides the theoretical framework for the relationship between the market risk premium and conditional volatility. This section also includes Monte Carlo evidence regarding the small sample properties of the risk return relationship in the GARCH-M model. In section 3, the historical data and sources are described. Section 4 provides time-series evidence on the mean variance relationship, where four (possibly asymmetric) GARCH specifications are considered. In Section 5, I conduct an exploratory analysis on the possibility of a time-varying risk return tradeoff. Finally, section 6 concludes.

## 2 Market Risk and Return in Equilibrium

Merton (1973) provides the seminal work on the *dynamic* risk return tradeoff in equilibrium. He derives a linear relationship between the conditional mean of the return on the wealth portfolio,
\( E_t[r_{M,t+1}], \) in relation to its conditional variance, \( \sigma^2_{M,t}, \) and the conditional covariance with variation in the investment opportunity set, \( \sigma_{MF,t}: \)

\[
E_t[r_{M,t+1} - r_{f,t}] = \left[ \frac{-J_{WW}}{J_W} \right] \sigma^2_{M,t} + \left[ \frac{-J_{WF}}{J_W} \right] \sigma_{MF,t},
\]

where \( J(W(t), F(t), t) \) is the indirect utility function in wealth, \( W(t), \) and any variables, \( F(t), \) describing the evolution of the investment opportunity set over time; subscripts denote partial derivatives. \( \left[ \frac{-J_{WW}}{J_W} \right] \) is linked to the coefficient of relative risk aversion, denoted as \( \lambda_M, \) which is typically assumed to be positive. The second component describes the adjustment to the conditional risk premium on the wealth portfolio arising from innovations to the state variables describing variation in the investment opportunity set. As \( J_W \) is positive, the sign of the adjustment depends upon the relationship between the marginal utility of wealth and the state of the world, \( J_{WF}, \) and the conditional covariance between innovations in the wealth portfolio and the state variables. If the wealth portfolio pays off precisely in states of the world when the marginal utility of wealth is low, market participants will demand a higher risk premium to hold the wealth portfolio; the opposite can also be true if the wealth portfolio serves as a hedge against risk associated with variation in the investment opportunity set.

In order to estimate the relationship in equation (1), one needs to make assumptions about the utility function and the processes describing the evolution of the conditional variance of the wealth portfolio and the conditional covariance with innovations in the state variables. Following the majority of the articles in this literature, the specification I consider is a univariate version of equation (1):

\[
E_t[r_{M,t+1} - r_{f,t}] = \lambda_0 + \lambda_M \sigma^2_{M,t}.
\]

The general model reduces to this restricted version if one assumes that the investment opportunity set is time-invariant or if the representative market participant has log utility. While both assumptions are likely extreme (particularly when focusing on time-varying volatility), Merton (1980) argues that the general intertemporal equilibrium risk return tradeoff can still be “reasonably approximated” by equation (2), and this is certainly the specification much of this literature has employed.\(^3\) Finally, as is traditional in this literature, I add a constant, \( \lambda_0, \) which could arise in equilibrium from transaction costs or taxes.

Empirically, I explore the mean variance tradeoff in the GARCH-M context, considering a host of possible processes that have been used in the literature without taking a stand on

\(^3\)Several articles (see Scruggs (1998) and Guo and Whitelaw (2003)), for example, take issue with this claim, and argue that the resolution of the mean variance tradeoff lies in the proper accounting of hedging demands. In an earlier version of the paper, I also consider a multi-variate ICAPM formulation with bond market risks generating hedging demands similar to Scruggs and Glabadanidis (2003). For these specifications, data span is still the primary issue, and I demonstrate (results available upon request) that the positivity of the mean variance relation is still consistently uncovered only in the historical data.
which one is correct. Since the choice of volatility specification is not driven by theory and the literature has documented that different specifications imply different findings for the risk return tradeoff, I explore a battery of specifications in the broader historical context taking each as plausible.

2.1 Econometric Methodology

The economic model relating the market risk premium to conditional volatility described in the previous section naturally lends itself to the GARCH-M technology developed by Engle, Lilien, and Robins (1987) and Bollerslev, Engle, and Wooldridge (1988). To characterize the conditional volatility process, I consider several alternative specifications in the GARCH class detailed below.

Following the univariate version of the economic restriction in equation (2), the mean component in the GARCH-M framework describes the mean volatility tradeoff for the equity market return as follows:

\[
M_{t+1} - r_{f,t} = \lambda_0 + \lambda_M \sigma_{M,t}^2 + \epsilon_{t+1},
\]

where \( \epsilon_{t+1} \) is mean zero with conditional variance, \( \sigma_{t,t}^2 \).

I allow the variance component to be described by one of several different specifications for the evolution of conditional volatility: GARCH (Bollerslev (1986)), EGARCH (Nelson (1991)), TARCH (Zakoian (1994) and Glosten, Jagannathan, and Runkle (1993)), and QGARCH (Sentana (1995))\(^4\). The latter three specifications allow for asymmetry in the response of the conditional volatility to return innovations. Many studies document an asymmetric relationship between news arrival and conditional volatility, where good and bad news appear to have different implications for future volatility. Models that allow for what is sometimes called the “leverage” effect are motivated by the work of Black (1976) and Christie (1982). Hentschel (1995) describes a general family of asymmetric conditional volatility models in the GARCH class that are designed to capture this phenomenon. This feature of the data may be important; when the conditional market volatility is described by a specification that incorporates the leverage effect asymmetry, Glosten, Jagannathan, and Runkle (1993) find a weakly negative relationship between the risk premium and conditional volatility. These specifications evaluate the importance of leverage effect asymmetry and its implication for the empirical risk return tradeoff. To start, conditional volatility is assumed to evolve according to one of the following specifications:

\[
\begin{align*}
\text{GARCH}(1,1): \quad \sigma_{M,t+1}^2 &= \delta_0 + \delta_1 \epsilon_{t+1}^2 + \delta_2 \sigma_{M,t}^2, \\
\text{TARCH}(1,1): \quad \sigma_{M,t+1}^2 &= \delta_0 + \delta_1 \epsilon_{t+1}^2 + \delta_2 D_{t+1} \epsilon_{t+1}^2 + \delta_2 \sigma_{M,t}^2,
\end{align*}
\]

\(^4\)These papers discuss the conditions under which these models satisfy positivity and covariance stationarity.
QGARCH(1,1): \( \sigma_{M,t+1}^2 = \delta_0 + \delta_1 (\epsilon_{t+1} - \delta_3)^2 + \delta_2 \sigma_{M,t}^2 \),

EGARCH(1,1): \( \ln(\sigma_{M,t+1}^2) = \delta_0 + \delta_1 \left( \frac{|\epsilon_{t+1}|}{\sigma_{M,t}} \right) - \delta_3 (\frac{\epsilon_{t+1}}{\sigma_{M,t}}) + \delta_2 \ln(\sigma_{M,t}^2) \),

where \( D_t \) is an indicator that takes the value of one when \( \epsilon_t \) is negative, and zero otherwise.
The GARCH model implies a symmetric response in conditional volatility following return innovations, whereas the parameter \( \delta_3 \) in the TARCH, QGARCH, and EGARCH specifications potentially shifts or tilts the news impact curve (Engle and Ng (1993)) rightward to allow for the apparent greater effect negative return shocks have on conditional volatility over positive return shocks.

In the GARCH-M specification, the conditional variance appears in the mean equation, hence the information matrix is not block diagonal between \((\lambda_0, \lambda_M)\) on the one hand and \((\delta_0, \delta_1, \delta_2, \delta_3)\) on the other. Therefore, joint estimation of the parameters governing the mean and variance equations is required. To estimate the parameters of the four specifications under consideration, I assume that the error term is drawn from a normal density (see Campbell and Hentschel (1993) and Hentschel (1995), for example), where the conditional variance follows one of the specifications in equation (4) and the log-likelihood function, \( \mathcal{L} = \sum_{t=1}^{T} l_t(\Theta) \), is constructed by summing the (conditionally) normal density functions:

\[
l_{t+1}(\Theta) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_{M,t}^2) - \frac{1}{2}(\epsilon_{t+1}/\sigma_{M,t}^2).
\] (5)

If the errors are not normal, the estimation method can instead be viewed as a quasi-maximum likelihood procedure as in White (1982), and robust standard errors are computed according to the method of Bollerslev and Wooldridge (1992).

2.2 Monte Carlo Analysis

In this section, I demonstrate that the weak empirical mean variance tradeoff documented throughout the literature is an artifact of small samples. In the GARCH-M framework, the mean expression involves the projection of the noisy excess market return on the very persistent conditional volatility (where the volatility shock is also likely correlated with the return shock). In the exchange rate literature, Baillie and Bollerslev (2000) have demonstrated that extreme caution should be exercised when interpreting projection coefficients under these conditions. This is certain to be a challenging econometric environment. The \( R^2 (= \frac{\lambda^2 \text{Var}(\sigma_{M,t}^2)}{\text{Var}(\epsilon_{M,t})}) \) from such a regression is very low as the variance of the return is typically several orders of magnitude larger than the variance of the volatility process (this is in fact usually thought to be required of a well-behaved GARCH specification). Looking ahead, Figure 2 plots the U.S. equity return along with an estimate of the equity risk premium implied by a GARCH-M conditional volatility specification \((\lambda \sigma_t^2)\). The unconditional standard deviation of returns is 0.049, whereas the
unconditional standard deviation of the GARCH process is 0.0025. With reasonable estimates of the mean variance tradeoff, this yields an $R^2$ well below 1% (a small number even for the return predictability literature). This finding confirms that volatility has almost no explanatory power for realized returns, and that the theory fails to explain the vast majority of month-to-month return variation.

Low regression $R^2$’s do not imply that a conditional mean variance tradeoff does not exist, but they do clearly highlight why the detection of that relationship will be extremely difficult in small samples. Nearly all of the return variation is unrelated to variation in the conditional market volatility. To explore this issue further, I conduct a Monte Carlo analysis on the small sample distribution of the projection coefficient, $\lambda_M$, for several volatility specifications where I know the mean variance tradeoff is valid. As high-frequency noise masks expected return variation, this exercise highlights the extent to which detecting the relation is extremely challenging for the usual sample sizes we consider. In the simulation, expected returns are driven by conditional volatility by construction, but realized return variation is calibrated to be nearly unrelated to volatility (as we see in the data).

First, given a specification for the evolution of the conditional volatility, I simulate monthly excess returns, $\tilde{R}_t$, from the following relationship:

$$\tilde{R}_{M,t+1} = \lambda_{0,M} + 2 \cdot \sigma_{M,t}^2 + \sigma_{M,t} \varepsilon_{M,t+1},$$

(6)

where $\varepsilon_{M,t}$ is drawn from a $N(0, 1)$ distribution. The parameter $\lambda_M$ is fixed at 2 (similar to the long-run estimates presented below). The (monthly) conditional volatility, $\sigma_{M,t}^2$, in equation (6), evolves according to one of the following two calibrated GARCH models:

GARCH(1,1): $\sigma_{M,t+1}^2 = \delta_0 + (0.10) \cdot \varepsilon_{M,t+1}^2 + (0.85) \cdot \sigma_{M,t}^2$,

(7)

TARCH(1,1): $\sigma_{M,t+1}^2 = \delta_0 + (0.05) \cdot \varepsilon_{M,t+1}^2 + (0.05) \cdot D_{t+1} \varepsilon_{M,t}^2 + (0.85) \cdot \sigma_{M,t}^2$,

where $\varepsilon_{M,t+1} = \sigma_{M,t} \varepsilon_{M,t+1}$ and the intercept, $\delta_0$, is calibrated for each specification so that the unconditional variance matches the data analogue. I simulate $T$ observations on the excess returns from a GARCH-M model exactly as specified in equations (6) and (7).

Taking this simulated sample as given, a univariate GARCH-M model is estimated, as in equation (5). This procedure is repeated for a total of 5000 times to generate a distribution for $\hat{\lambda}_M$, the particular quantity of interest. The same analysis is performed for the univariate TARCH specification.\(^5\) The econometric (time-series) size for each simulation, $T$, is either 500, 2,000, or 5,000 monthly observations. Five hundred observations closely corresponds to the sample history of monthly observations available for the usual post-war sample employed in this literature, and 2,000 closely corresponds to the same for the nearly two century historical record.

\(^5\)I also perform this analysis for the two other asymmetric GARCH specifications considered. The results (available upon request) are very similar, so I exclude them in the interests of space.
The striking depiction of the small sample nature of the estimation problem is presented in Figure 3, where I plot the Monte Carlo densities of the estimated coefficient $\hat{\lambda}_M$ for the GARCH-M estimates where the data are drawn from the GARCH or TARCH processes (Panels A and B, respectively). Small sample evidence is presented for $T = 500$, $2,000$, and $5,000$ monthly observations. The estimates based upon only 500 observations are fairly uninformative about the “true” value of $\lambda_M$, in the sense that the distribution has considerable mass below zero, even when $\lambda_M = 2$ is known to be true. This is particularly pronounced when the data are drawn from the asymmetric TARCH specification, as the correlation between the volatility and return shocks increases significantly through the construction of the volatility process. Despite this, the distribution of the estimator is centered properly. Even when estimated over 2,000 observations, as exist for the longer record, a non-negligible probability mass exists below zero for the specifications considered. For instance, for $T=500$, 19% (22%) of the time the mean variance parameter, $\lambda_M$, falls below zero for the GARCH (TARCH) specification. This improves dramatically in going to 2000 observations, where this percentage falls to 3% (7%). For 5000 observations, the likelihood of a negative coefficient is less than one percent. For a sample like the post-war period, about one-fifth of the time a researcher will spuriously observe a negative mean variance coefficient. What is perhaps most relevant, under the null of a positive mean variance tradeoff, is the conditional probability of estimating a negative coefficient over a shorter period when the longer sample yields the correct positive estimate. To explore this further, for each simulation of 2000 observations, I break the draw into sub-periods exactly as described below for the observed data. The probability of obtaining a negative estimate in any one of the shorter periods conditional on obtaining a positive estimate over the full sample is 28%. That is, more than a quarter of the time, a researcher will observe a negative tradeoff in a small sample for the same draw that provides a positive estimate over the full historical record.

**Volatility Characteristics**

Given the striking Monte Carlo evidence revealing the difficulties associated with estimating the mean variance tradeoff, it is important to understand the origins of this difficulty. As suggested by the $R^2$ from the mean component of the estimation, the low explanatory power is related to the relative volatilities of the return and its conditional variance. To explore this further, I consider the degree to which these results vary by adjusting the features of the return or conditional volatility processes. Figure 4 presents evidence for three different specification adjustments to the Monte Carlo simulation: tuning return volatility (panel A), the volatility of return volatility (panel B), and the persistence of return volatility (panel C). To focus on the most challenging environment, the simulation size ($T$) is 500 monthly observations, similar to
the sample lengths we typically see in this literature. For comparison, I also present the baseline case explored above in each figure, which is identical to the solid line in Figure 3 (panel A).

In panels A and B, respectively, I show that the ability to detect the risk return relation is closely linked to the magnitude of return volatility and the volatility of return volatility. Doubling return volatility yields a considerably less informative estimate of the mean variance tradeoff, whereas halving return volatility greatly improves the precision of the estimate. In comparison to a roughly 20% frequency documented above, doubling (halving) return volatility yields a negative estimate of the relation about 35% (6%) of the time. Adjusting the volatility of return volatility also has a substantial impact on the precision of the estimates. For example, doubling the volatility of volatility yields a much tighter distribution of the mean variance relation; the econometrician will uncover a negative estimate only 5% of the time. If you quadruple the volatility of volatility, you will essentially never see a negative estimate. Despite these findings, the predictive $R^2$'s that follow from any of these adjustments remain very low (quadrupling the volatility of volatility yields $R^2$'s below 1.5%, yet deliver considerably more reliable estimates of $\lambda_M$). In all cases, nearly all of the return variation is unrelated to expected return variation. Panel C provides the small sample distribution for $\lambda_M$ across different cases where the degree of conditional volatility persistence varies. As can be seen, volatility persistence only plays a significant role in the most extreme (near unit-root) cases. Otherwise, this is not the origin of the small sample detection problem.

This auxiliary evidence highlights the challenges of this econometric environment. A regression that attempts to explain the variation of the volatile market return with the extremely quiescent conditional variance (as a GARCH specification yields) is fraught with imprecision. In the data, we can not obtain a more (less) volatile variance (return) process. Instead, as suggested by the baseline Monte Carlo evidence in Figure 3, we require more data. I explore this dimension of the GARCH-M estimator next.

**Data Span versus Frequency**

Andersen and Bollerslev (1998) show that data frequency (going to high-frequency intra-day data) dramatically improves volatility estimation. However, Merton (1980) argues that data span is required to capture expected return variation. Given the challenges documented above, I explore whether data span or frequency are required to uncover the mean variance tradeoff in the GARCH-M context; in small samples, noisy returns mask expected return variation. To explore span versus frequency, I simulate from a discretized version (at five-minute intervals) of the stochastic volatility process (8). Following Nelson (1990) and Andersen and Bollerslev (1998), I express the limiting continuous time process for a GARCH(1,1)-M specification as
follows:

\[
\frac{dP_t}{P_t} = \lambda_M \sigma^2_{M,t} dt + \sigma_{M,t} dW_{P,t},
\]

\[
d\sigma^2_{M,t} = \theta(\omega - \sigma^2_{M,t}) dt + (2\psi\theta)^{\frac{1}{2}} \sigma^2_{M,t} dW_{\sigma,t}.
\]  

(8)

\(P_t\) is the continuously determined market price level and \(\sigma^2_{M,t}\) is the stochastic instantaneous variance of the return process. \(W_{P,t}\) and \(W_{\sigma,t}\) are independent Wiener processes. Andersen and Bollerslev (1998) provide the mapping from the discrete time GARCH(1,1) model parameters \((\delta_0 = 0.0002, \delta_1 = 0.10, \text{ and } \delta_2 = 0.85)\) and the parameters \((\theta, \omega, \psi) = (0.0023, 0.0001, 0.459)\) for monthly observed data (assuming 22 trading days per month and 288 5-min observations per day). The parameter \(\lambda_M\) is fixed at 2, as for the Monte Carlo simulations presented above. To start, I simulate \(T\) monthly observations on the return from the continuous time specifications in equation (8). Taking the sample as given, a discrete time GARCH(1,1)-M model is estimated as in equation (5). This procedure is repeated 5000 times to generate a distribution for \(\hat{\lambda}_M\). By simulating from a continuous time process, I construct monthly data across different time spans (exactly as in the baseline discrete Monte Carlo), but I can also sample the data at higher (weekly and daily) frequencies. As before, the econometric (time-series) size, \(T\), is either 500, 2000, and 5000 monthly observations, exploring the span dimension. To investigate the frequency dimension, I also split the 500 monthly observations into 2000 weekly and 11000 daily observations (assuming a 22 trading day month).

Figure 5 presents the small sample distribution for \(\lambda_M\) across different data spans (Panel A) and frequencies (Panel B). Panel A yields a nearly identical (uninformative) distribution on the estimate to that shown in Figure 3 for small samples. Going to longer periods shows dramatic improvement in the ability to detect the tradeoff (again exactly as in Figure 3). This suggests data span is important, but it may simply be a matter of obtaining more data points. To rule this out, panel B shows the small sample distributions for \(\lambda_M\) for different data frequencies holding the span constant. Most important, despite having more “observations,” the small sample distributions for the risk return estimate are nearly identical (with only slight improvement with increased frequency). Further, the median (and average) values for \(\lambda_M\) are roughly 1.8 for all three measurement frequencies in Panel B, suggesting a slight downward bias, potentially contributing to negative point estimates even if the true \(\lambda_M\) is positive. Note that even over a longer data span (Panel A), the bias does not completely disappear; the median (and average) values for \(\lambda_M\) are roughly 1.85 and 1.95 for 2000 and 5000 monthly observations, respectively. Taken together, the volatility dimension (related to data frequency) is second order given the low \(R^2\); we require span to capture expected return variation in the presence of the high (low) volatility of returns (conditional variance). Even high quality intra-day trade-and-quote data would do little for the identification of the mean variance tradeoff in the GARCH-M context.
where the conditional variance process exhibits such a low level of volatility.

Essentially, a very large number (span) of time-series observations is required to estimate \( \lambda_M \) with precision in the GARCH-M context. The challenge lies in the detection of expected return variation, which is swamped by realized return noise. In fact, one can potentially conclude that the difficulty in detecting the risk return relationship in the GARCH-M context found throughout the literature may have very little to do with the particular specification under consideration. The Monte Carlo evidence suggests that the weak statistical relationship between the market risk premium and conditional volatility may be viewed as an artifact of small samples. The ambiguous time-series evidence does not necessarily cast doubt on the posited relationship; instead, we simply require more data to conduct this exercise. To that end, I turn to the historical record.

3 Data Description

To maximize the time-series content of the estimation of the risk return tradeoff, I employ the historical U.S. equity market record from 1836-2003. Over the earlier part of the sample, monthly data are collected for 1836-1925 from Schwert (1990). From 1926-2003, I use the more familiar CRSP value-weighted portfolio returns for the NYSE, AMEX, and Nasdaq markets. Combining these two sources to describe the historical U.S. equity record follows Schwert (1990). As an alternative for the earlier historical record, Goetzmann, Ibbotson, and Peng (2001) collect monthly NYSE price data beginning in the early 19th century through 1925. In my analysis, I also consider this alternative historical data to evaluate the robustness of the empirical evidence on the risk return tradeoff. Taken together, these two measures of the U.S. historical record should provide a broad indication of the asset return behavior over the earlier U.S. history. To complete the picture, I also collect total return data for the short-term bill index for the United States over the 1836-2003 period from the *Global Financial Data* provider. Short-term bill data will serve as the conditionally risk-free rate in my analysis (see Siegel (1992)). For more detail, see the data appendix.

The Monte Carlo evidence above shows that (at least for the GARCH-M technology frequently employed in this literature) the estimation of the risk return relationship is essentially a detection problem, and should prove (as it has) extremely challenging for the data span often considered. To be fair, it is clear that the U.S. equity market is a very different animal over the time period considered in terms of the trading environment and market efficiency, the link between the observed market portfolio and the theoretical wealth portfolio, and the simple raw size of the market itself. These are the drawbacks to forging ahead in this broader historical context. Nevertheless, I would argue that the exploration of these data are no less credible
than the current (voluminous) work on emerging equity markets where the data concerns are potentially even more pronounced. As a robustness check, I consider an additional historical data source for inference on the risk return tradeoff, the United Kingdom. Equity return data are also available for the United Kingdom, the predominant market in the 19th century (see data appendix).

There are also several economic reasons to consider the historical record. In addition to the extended time-series dimension, these data include several pronounced financial and macroeconomic episodes that may house important information about the equity price reaction to elevated uncertainty and associated volatility. As an additional benefit, the inclusion of these events may be helpful given the evidence on the importance of the volatility of volatility in the Monte Carlo exercise above. Schwert (1989) provides a link between documented financial crises and economic downturns and the behavior of equity prices over the longer record. This suggests that the market is linked to the broader state of the economy even for the 19th century data, and hence prices may reflect economic risks consistent with the model presented above. At the very least, it is certainly worth exploring the nature of the risk return tradeoff in this broader context given that the Monte Carlo evidence shows clearly that the relationship is hard to detect over the shorter horizon using a GARCH-M framework.

Table 1 reports summary statistics on the total returns for the U.S. equity market, $r_{M,t}$ and the short bill return, $r_{f,t}$, over several different periods for comparison. The return data for each series are also displayed in Figure 1; the break point in the figure represents the transition from the historical data to the more typical CRSP data. For the full 1836-2003 period, there are roughly 2000 monthly observations. I present summary statistics for the full 1836-2003 sample, plus breakdowns over three sub-periods: 1836-1891, 1892-1949, and 1950-2003. This allows me to explore the degree to which there are any structural shifts in the return distribution over the full sample. The latter period represents the post-war period frequently employed in this literature. Across the whole sample, the mean return on the U.S. equity market portfolio is about 79 basis points (bp) per month, ranging from 61 bp in the early period to 102 bp in the more recent. Equity return standard deviations also vary somewhat over the sample. Equity market data are more volatile during the Great Depression than during the other periods. Over this period, the mean and volatility of the risk free rate are fairly consistent, with the average risk free rate ranging between 30 and 40 basis points per month.

It is also the case that equity market return autocorrelation is somewhat more pronounced in the early period, but the difference is rather small. This could potentially be consistent with reduced market efficiency over this period, but it is much smaller than comparable coefficients observed for emerging equity markets or even portfolios of small U.S. firms in recent decades. Bill returns demonstrate a high degree of persistence across all periods considered. In sum,
the return data appear to be fairly consistent across the historical record, suggesting that the exploration of the expected return volatility tradeoff in this broader context is reasonable given the small sample challenges associated with inference in the GARCH-M context.

4 Estimating the Mean Variance Tradeoff

Using the univariate GARCH-M framework, Table 2 presents evidence on the mean variance tradeoff in equation (1) for the sample periods discussed above. In Panels A through D, I present estimates for the four different GARCH specifications detailed above, where the latter three potentially facilitate volatility asymmetry. In addition to providing estimates for the full historical record, I also compare the results across sub-periods including the more commonly used post-war data.

As a start, I reconsider the risk return evidence over the post-war data. The estimated mean variance relationship is positive (but not significant) when the latter evolves according to the symmetric GARCH specification. The point estimate, presented in panel A, for $\lambda_M$ is 3.26 (with a standard error of 2.61). This finding is in accordance with the positive relationship documented in French, Schwert, and Stambaugh (1987). However, over this period, the market portfolio exhibits asymmetry in the response of conditional volatility to positive and negative return shocks (see Nelson (1991), for example), hence Glosten, Jagannathan, and Runkle (1993) argue that the GARCH specification for the conditional volatility poorly describes U.S. equity market data. I estimate several alternative volatility specifications, the results of which are also presented in Table 2, that facilitate leverage effect asymmetry, including the TARCH (Panel B), QGARCH (Panel C), and EGARCH (Panel D) models. When the specification for conditional volatility incorporates leverage effect asymmetry, the estimated risk return relationship is negative (but, not significant) in two of the three cases. For example, the point estimate for $\lambda_M$ is -0.84 (with a standard error of 2.63) for the TARCH specification. Additionally, the parameter describing the degree of leverage effect asymmetry, $\delta_3$, is significant in all cases, suggesting that negative equity return shocks are associated with elevated market volatility. Taken at face value, this evidence is consistent with the previous research, and hence casts doubt on the positive risk return tradeoff implied by many economic models, particularly as the asymmetric specifications for conditional volatility better describe the U.S. data. However, it is worth noting that the predictive $R^2$'s in these regressions (and all throughout the article) are less than 1%. Conditional volatility has almost no explanatory power for return variation. As the Monte Carlo evidence suggests, more data than are commonly used in this exploration are required for inference regarding the economic relationship.

To highlight this issue, Table 2 also presents evidence on the risk return tradeoff for the other
samples considered, including the long historical record of U.S. equity market performance. For the period 1836-2003, the evidence on the mean variance relationship looks very different from the post-war period. First, across all four specifications, including those that allow leverage-effect asymmetry, the estimated coefficient $\lambda_M$ is positive and statically significant despite low predictive $R^2$’s. For example, when equity market volatility evolves according to either the GARCH (Panel A) or TARCH (Panel B) specification, the point estimates for $\lambda_M$ are 2.46 (with a standard error of 0.84) and 1.88 (with a standard error of 0.84), respectively. Interestingly, across all three models that facilitate volatility asymmetry, the relevant parameter, $\delta_3$, is significant, suggesting the relationship between the sign of the return innovation and conditional volatility is a feature of the full U.S. historical record, not merely the more recent CRSP data.

Over the full 1836-2003 sample, the mean excess return on the U.S. equity market is 0.0044 in monthly terms, translating to an annualized excess return of 5.3%. If you take the average risk premium implied by the GARCH-M model (the average of $\lambda_M \sigma^2_{M,t}$), you obtain a range of 0.0043 (Q-GARCH) to 0.0059 (GARCH) per month across the four volatility specifications employed. This implies a very reasonable annualized market risk premium between 5.2 and 7.1%.

Exploiting the full historical record allows for a more precise estimate of the mean variance tradeoff. Schwert (1989) highlights the links between equity market conditional volatility and documented periods of economic uncertainty and financial crises over the full historical record; hence, adding these periods may play an important role in our understanding of the economic relationships between risk and compensation for risk. Figure 6 displays the estimates of conditional market standard deviations implied by the GARCH process (the picture looks very similar for the other specifications). Elevated volatility is most pronounced during the recession of 1857, the U.S. Civil War, the recession of 1907, the Great Depression, the oil crisis in the mid-1970’s, the stock market crash of 1987, and the global financial crises of the late 1990’s. I also plot the financial crisis dates identified by Schwert (1989) (adding October 1987 and October 1998). The important (volatility) events from the earlier part of the sample suggest that the estimated relationship between the expected market return and volatility appears considerably more consistent in the broader historical context. This is in direct accordance with the Monte Carlo evidence presented in the previous section that demonstrates the difficulty in detecting the relationship in small samples.

As an interesting alternative, Campbell (1987) and Glosten, Jagannathan, and Runkle (1993) report a significant relationship between equity market variance and the level of the nominal interest rate. Further, Glosten, Jagannathan, and Runkle (1993) show that the inclusion of the risk free rate into the volatility specification reduces the magnitude of the estimated risk return tradeoff in the GARCH-M framework. To explore these features (evidence is not presented
here, but available upon request), I estimate a modified version of the univariate GARCH-M specification that allows the risk free rate to enter the volatility and/or mean specification. For the full sample, the inclusion of the risk-free security in this analysis, either through the conditional mean or volatility, has a minor effect on the estimated mean variance tradeoff. The general conclusions are completely unaffected by this alternative.

Finally, to explore the tradeoff in exclusion of the more recent period, I also estimate these four specifications across the two earlier periods. For the historical data, I find that a) volatility is very persistent; b) volatility asymmetry is an important feature of the early U.S. data; and that c) the risk return tradeoff is difficult to detect in other small samples as well. One obvious concern is that there may be a time-varying tradeoff in the U.S. data across these periods. While this may be important (and is explored further in the last section), the general volatility dynamics are quite similar across periods. Second, a joint Wald test of the null hypothesis that the estimates of $\lambda_M$ are equal to one another across periods is never rejected for any of the cases considered. Unfortunately, the Wald test is likely to have low power. If estimation of the mean variance tradeoff is challenging for the long sample, detecting within-period variation in the tradeoff coefficient is not very likely. One can not easily claim whether a failure to reject such a test is evidence of sampling error or the inability to uncover true within-sample variation. Given the magnitude of the estimation problem presented in the Monte Carlo study, the former explanation seems the most likely, but one can not rule out the latter. For that reason, I also conduct an exploratory analysis on the time-variation of the risk return tradeoff (given the potential importance of both cyclical variation or structural change). This, too, will face the same concerns that within-sample detection is a challenge. A compelling alternative is to explore the mean variance estimates obtained from other available historical records (where for example the U.K. market was primary in the 19th century).

Alternative Historical Data

To evaluate the robustness of the measurement of the risk return tradeoff in this context, I consider several alternative representations of the historical record. Table 3 presents evidence on the measurement of the risk return tradeoff for the Goetzmann, Ibbotson, and Peng (2001) capital gain series covering the historical NYSE record, appended by the CRSP value-weighted capital gain (for comparability) index of NYSE, AMEX and Nasdaq. The NYSE index shares a 0.59 correlation with the Schwert (1990) index over the pre-CRSP period (1836-1925), suggesting that while these two measures of the U.S. historical experiences are capturing similar features, the differences arising from their industrial composition may provide a useful alternative representation. First, the degree of volatility asymmetry is very similar for the NYSE
data. Most importantly, the expected return volatility tradeoff estimates, $\lambda_M$, are very similar to the estimates obtained for the Schwert (1990) index, and are highly significant across all specifications.

I also present evidence in Table 4 for the mean variance tradeoff for the U.K. across the same alternative volatility specifications explored above. The correlation between the U.S. and U.K. market indices are 0.21 and 0.51 in the historical (1836-1949) and post-war (1950-2003) periods, respectively, suggesting the U.K. data offer a useful alternative representation of the historical experience for an important market. First, the post-war U.K. data yield a positive, but insignificant, risk return tradeoff coefficient. Interestingly, this complementary evidence suggests that the negative relationship often documented in recent U.S. data should not be attributed to something special about the post-war period, despite the higher correlation between the two markets in the more recent period. Instead, the evidence lends credence to the notion that sampling error consistent with the Monte Carlo results above is driving the within-sample variation in the estimated mean variance tradeoff. In a related context, De Santis and Imrohoruglu (1997) examine a similar economic restriction for several emerging and developed economies (some with very low cross-correlations); they too find mixed evidence for shorter samples across a wide collection of countries. Taken together, it does not appear that the inconclusive evidence for the short period is unique to the U.S. experience. Second, the mean variance tradeoff is positive and significant (and right around 2, just as in the U.S. data) regardless of the volatility specification over the longer U.K. historical record. The strong tradeoff uncovered throughout the full U.S. historical record is corroborated by the U.K. experience. The relationship is not dependent upon one specific historical measurement.

5 A Time-varying Risk Return Tradeoff?

There are (at least) two concerns associated with the potentially strong assumption of a time-invariant risk return tradeoff in equation (2). First, in equilibrium, the mean variance tradeoff is described by risk aversion (ignoring intertemporal hedging demands). Risk aversion may exhibit cyclical variation through time (as implied by habit models such as Campbell and Cochrane (1999)). In a historical context, risk aversion may also vary because of structural reasons associated with the evolution of financial markets and improved risk sharing. The U.S. financial markets and economy have changed dramatically over the sample employed in this article. In the 19th century, the United States is effectively an emerging market, whereas in the 20th century the U.S. becomes the predominant world economic power. While the Monte Carlo evidence suggests any differences in the measured risk return relation could easily be attributed to sampling error, a legitimate alternative explanation is that the return distribution and the fundamental
The risk-return relationship has changed over time. For instance, Pastor and Stambaugh (2001) explore the U.S. equity premium over the historical record, and find evidence for structural breaks in the return distribution. With the development of the U.S. economy and financial markets, fundamental structural changes may have engendered a break in the mean variance relation.

Second, the evidence presented here for both the U.S. and U.K. does require an important caveat. There is also a growing body of research showing that idiosyncratic and liquidity risks play roles in equity market pricing. Campbell, Lettau, Malkiel, and Xu (2001), Malkiel and Xu (2001), and Goyal and Santa-Clara (2003) highlight the importance of market-wide measures of aggregated idiosyncratic risk. In particular, from the period for which firm-level daily equity data are available, Goyal and Santa-Clara (2003) document that idiosyncratic risk, identified as the equally weighted average of firm-level total risk (i.e., average firm level variance), is a better predictor of future market returns than overall market variance. As a result, they contend that idiosyncratic risk is priced. Also, Acharya and Pedersen (2005), Amihud (2002), Pastor and Stambaugh (2003), and Gibson and Mougeot (2004) (among many others) document that aggregate equity market liquidity risk is priced, the latter exploring the role in a GARCH-M context.

Unfortunately, for lack of firm-level or high quality trade-and-quote data, I cannot explore the relative contribution of average firm- versus market-level variance or liquidity risk in the longer historical context. Accordingly, one might view the results I document on the expected return variance tradeoff in one of these contexts. For instance, if earlier periods correspond to less-developed financial markets, it seems plausible that the index returns representing only a limited collection of stocks contain more idiosyncratic risk. The equity “market” portfolio from Schwert (1990) covering the earlier part of the sample represents a limited collection of stocks, many of which are in only a few industries (banking, insurance, and railroad from 1836-1872); while somewhat broader, the NYSE data from Goetzmann, Ibbotson, and Peng (2001) are limited to only 50-80 firms, on average. Similarly, the U.K. historical data represent only 25-75 firms in the 19th century. It may be that the power of diversification is limited during these earlier periods – that is, there may still be considerable firm or industry specific risk retained in the portfolio. The expanding breadth of the market over the sample could complicate the estimation of the risk return tradeoff if idiosyncratic risk is both important and not priced in earlier periods.

On this point, however, the theory is somewhat unclear. In a related context, Stulz (1999) argues that investors in emerging markets that are restricted from diversifying abroad due to

---

6Note, while older data are available, the U.K. portfolio becomes very thin extending back to the 18th century. The U.K. index for the 18th and early 19th centuries represents equally-weighted returns on only three shares: the Bank of England, the East India Company, and the South Sea Company. Hence, I exclude that earlier period.
legal barriers require compensation for the total risk of the domestic equity market. If the portfolio in earlier periods is more volatile (because of a less comprehensive industry representation, for example), but people cannot diversify the risk, then total volatility should be priced. In this scenario, total market volatility risk will be compensated across the entire sample, and an important source for a varying risk return relation could be variation in investor risk aversion. Looking at the estimated results across sample periods, the U.K. $\lambda_M$ estimates increase, but the comparable figures for the U.S. decline. The mixed empirical evidence over a span where the market unambiguously becomes better diversified (and more liquid) does not point towards to a clear story regarding the role for idiosyncratic risk. In fact, given the lack of firm level data and the challenging econometric environment, separating these effects (time-varying risk aversion and the magnitude and importance of idiosyncratic and liquidity risks) is nearly impossible. There is no way for me to disentangle these results for the full U.S. or U.K. historical records given the data limitations, but these are hypotheses I can not rule out.

Despite these caveats, I conduct an exploratory analysis into the potentially time-varying nature of the risk return tradeoff in this section. I directly parameterize the risk return tradeoff as a function of several observable financial and macroeconomic variables that are related to the development of the U.S. economy and financial markets over the historical record. These factors may reflect the origins of risk aversion variation from either cyclical fluctuations or long term structural developments (or may reflect general market development and the role for idiosyncratic risk).

I consider the simple GARCH(1,1) specification in equation 4, but allow the risk return tradeoff coefficient to vary with several observable financial and macroeconomic indicators of the development of the U.S. financial market and economy (presented in Table 5):\footnote{In an earlier version, I also estimate an alternative specification (similar to Gray (1996)) where states of the world are essentially defined by volatility regimes using a regime-dependent GARCH-M specification. Turner, Startz, and Nelson (1989), Whitelaw (2000) and Mayfield (2004) employ similar specifications. Across the historical sample, high volatility regimes are more likely during financial crises, and are associated with a negative risk-return tradeoff similar to previous studies. Results are available upon request.}

$$\lambda_{M,t} = \lambda_{M,0} + \lambda_{M,1}X_t. 
(9)$$

I explore the interaction along three basic dimensions: a) cyclical variation, which is captured with a binary indicator that takes the value of one when the U.S. is in an NBER documented recession and 0 otherwise; b) the development of U.S. financial markets, which is captured by the size of the U.S. equity market relative to the overall economy (MCAP/GDP), the number of listed companies, and the development of the financial sector (M3/GDP); and finally, two measures that represent the state of the overall economy and its links to the outside world, an indicator of the level of foreign trade (exports plus imports/GDP) plus an indicator of the size
of the government (GOV/GDP). Of course, the conjecture that these variables are related to the risk return relation are purely speculative, and are not motivated directly by theory. They do suggest plausible ways the tradeoff may change through time.

For the more recent equity data, Mayfield (2004) documents that the risk return tradeoff is less pronounced during documented recessions. Using alternative empirical methodologies, other papers estimate counter-cyclical risk aversion variation (higher risk premia during economic downturns); for example, see Lettau and Ludvigson (2003) for evidence based on predictability regressions. For the historical data, when the mean variance tradeoff is allowed to interact with economic recessions, the own-effect, $\lambda_{M,0}$, is positive and statistically significant. The interaction effect $\lambda_{M,1}$ is negative and significant suggesting that the risk return tradeoff is smaller during economic recessions, consistent with the evidence from Mayfield (2004). This evidence would be consistent with declining levels of risk aversion during recessions, but is not consistent with the spirit of equilibrium habit formation models. Before drawing a conclusion, there is an important concern that this specification is confounding expected returns with realized returns, particularly in the less common recession states often associated with low or negative market returns.

The development of the U.S. financial sector and the equity market in particular is likely one of the most important dimensions over which the measured relation might vary. Financial development may be associated with improved risk sharing. Also, the development of the equity market itself, through consolidation, increased transparency, regulation, liquidity, and investor participation, potentially affects the degree to which the available index represents the true aggregate wealth portfolio and the role for idiosyncratic risk. As the size and breadth (number of listed shares) of the market have changed dramatically over time, understanding the impact of these developments on the measured relationship may be important. Table 5 presents evidence where the market capitalization and number of listed companies enter the determination of the mean variance tradeoff. First, the interaction effect is significant and negative for the ratio of market capitalization to GDP (at the 10% level) and unrelated to the number of listed companies. The market capitalization, in relation to the size of the overall economy, may be linked to the relative importance or development of the equity market, and an improved equity market may indeed strengthen the link between the equity index and the aggregate wealth portfolio. At higher frequencies, market booms (declines) may also be associated with lower (higher) levels of risk aversion. Second, there does not seem to be a link with the number of companies traded. Third, a critical dimension of equity market importance is unfortunately unavailable, participation. Reliable historical data on equity market participation would greatly help to link

---

8Historical market capitalization, number of shares, trade, government expenditure and money stock data are from Rousseau and Jovanovic (2001) and Rousseau and Sylla (2001), and are kindly provided by Peter Rousseau. The number of shares requires linear interpolation to bridge the historical and CRSP periods. The remaining data are contiguous, but coverage is typically somewhat shorter.
the equity index with the aggregate wealth portfolio and may affect aggregate risk sharing. In
the absence of this, I instead use a simple time trend. While crude, over the U.S. historical
record this is likely to represent the broad trends in equity market participation over time. I
do not find a significant relationship for an interaction with time; that is, the mean variance
tradeoff does not seem to mechanically increase or decrease over the U.S. historical experience
despite the sub-sample differences presented above. This is either a simple statistical statement
about the relationship or it says that the tradeoff has not varied with participation in some
broad sense. Finally, I also consider an interaction with a measure of macroeconomic liquidity
(M3/GDP) (general financial development) which might be associated with greater risk sharing
in the domestic economy. For this variable, the effect is indeed negative, but is not statistically
significant. Collectively, the evidence qualitatively points to a lower level of risk compensation
with improved financial development, but statistical precision is low.

Finally, one could also imagine a scenario where risk aversion is affected by a greater degree
of foreign openness (international risk sharing) and a lower degree of government spending
(intervention). In Table 5, the interaction effect has the “expected” sign for each case, and is
significant for the trade dependency ratio (at the 10% level) and the size of the government (at
the 5% level). This suggests that the risk return tradeoff is larger with lower trade dependency
and more relative government spending over the historical record.

One concern is that many of these variables grow secularly through time as the U.S. economy
develops. To attempt to isolate the effect of the variation in these macro variables free of
trend, I consider two additional specifications (complete results are available upon request).
First, I run the interaction specifications including both a time trend and the macro-variable
together. Second, I statistically de-trend each macro variable, and use the de-trended series as
an alternative interaction variable. In the first specification, I find that the trend variable effect
is not statistically significant across the cases considered. In both specifications, the interaction
variables retain the same qualitative features as presented in Table 5. In fact, the negative
coefficient for the size of the market is statistically significant at the 5% level, where the trend
in long-run market growth may have masked important higher-frequency variation. That is,
the level of the market relative to the economy may play an important role in the time-varying
nature of risk compensation. Taken together, this suggests that even in the presence of trends
(or controlling for them), these variables may highlight an important role for fluctuations in risk
aversion.

In all these specifications, the first-order effect is a positive and significant mean variance
tradeoff. However, the exploratory evidence suggests that the risk return tradeoff is related to
recessions, the overall size of the equity market, external trade and government spending. The
degree to which these findings provide evidence of cyclical or structural risk aversion variation
(or the importance of idiosyncratic risk) is an interesting subject for future research.

6 Conclusion

While the risk return tradeoff is fundamental to finance, the existing empirical relationship is ambiguous at best. However, Monte Carlo evidence demonstrates that the mixed findings may be viewed as a statistical artifact of small samples, rather than as evidence against the postulated relationship itself. Conditional volatility has almost no explanatory power for realized returns. Given these limitations, a large data span is required to reliably detect the risk return tradeoff. Using information from a longer historical record of the U.S. and U.K. equity market experience, I document a significantly positive risk return relation, even when the specification for conditional volatility incorporates leverage effect asymmetry. Exploratory analysis suggests a role for a time-varying risk return tradeoff, but the overall positive mean variance tradeoff does not seem to be affected by the changing nature of the U.S. markets and economy over the historical record.
References


Data Appendix

Carefully detailed in Schwert (1990), monthly equity return data are collected for 1836-1925 from several historical sources. Equity market data before 1871 reflect the capital appreciation (without dividends) of portfolios of bank, insurance, and railroad stocks. Schwert (1990) constructs total returns by estimating the dividend yield over this period; this has little effect on the volatility and covariance patterns, but does increase the average equity return. Additionally, there are roughly 20 years over this period that suffer from a variant of the time-averaging problem addressed in Working (1960). Schwert (1990) also adjusts the price series over this period. From 1926-2003, I use the more familiar CRSP value-weighted portfolio returns for the NYSE, AMEX, and Nasdaq markets.

There are some drawbacks to the historical comparison. The earliest period in this study includes only 7 bank and 20 railroad stocks, whereas the CRSP portfolio in the latter part of the sample reflects thousands of firms operating across all aspects of the (now huge) U.S. economy. Second, Schwert’s historical market portfolio is equally weighted until 1862, value weighted from 1863 through 1885, and price weighted until 1925.

My second measure employs the market equity index from Goetzmann, Ibbotson, and Peng (2001). They collect monthly equity price data for over 600 NYSE traded firms beginning in the early 19th century through 1925, ranging from 46 firms in 1836 (the beginning of my sample) to 86 firms in 1925 – with considerable firm entry and exit in the interim. Their data has the advantage of being constructed entirely from end-of-month prices, allowing for the construction of returns free of the “Working” effect. Additionally, their data are from one unified source; however, without detailed dividend data, they can only construct the capital appreciation rather than the full total return, underestimating the risk premium on the U.S. market over this period. To update the capital gain series through 2003, I use the NYSE, AMEX and Nasdaq value-weighted capital gain series from CRSP (retx) for the 1925-2003 period. There are a number of months for which the Goetzmann, Ibbotson, and Peng (2001) constructed market index is missing data. For the missing months, I use the returns from Schwert (1990). Both data sources have missing equity data with the documented market closure during World War I. There are four months in 1914 during which the U.S. data are unavailable (data are missing for the months August to November). I set those months to an average that interpolates the pre and post index level. Since equity market data from Schwert (1990) are computed as total returns, these data will serve as the main focus of my empirical work.

As an additional robustness check, I collect U.K. equity market data over the same 1836-2003 period from the Global Financial Data provider. These data combine several historical sources on U.K. security price data, representing 25-75 firms in the early to mid-19th century, several hundred firms at the turn of the last century, to nearly 99% of the U.K. market capitalization
in more recent periods. Earlier U.K. data (extending back to the 18th and early 19th century) are available, but only represent a simple equal weighted average of three shares: the Bank of England, the East India Company, and the South Sea Company. I exclude this earlier period. Due to limited dividend data, total returns are not available for the historical data. Hence, I focus only on the capital appreciation series. For a more detailed discussion of the various historical data sources used to construct this index, see the documentation provided on the FT All-Shares Price Index (FTASD) from www.globalfindata.com. For comparison with the U.S. data, I also collect total U.K. returns for the post-war period (where reliable dividend data are available) from the same data provider. The capital appreciation and total return series are nearly perfectly correlated in the overlapping period. There are five months in 1914 during which the U.K. data are unavailable (data are missing for the months August to December). As with the U.S. data, I set those months to an average that interpolates the pre and post index level.

I also collect total return data for the short-term bill indices for the United States and the United Kingdom over the 1836-2003 period from the Global Financial Data provider. For the bill data, see www.globalfindata.com for a more detailed discussion of the data sources and construction. For the U.S. I use the commonly used short-term Treasury bill data, which are first available only in 1931 from the Federal Reserve. Between 1919 and 1931 bill return reflect the shortest-term US bond available. Short-term U.S. Treasury data are not available prior to 1919 in that there were no riskless short term securities backed by the U.S. government. From 1836 to 1918, only commercial paper data are available for the U.S., which are poor proxies for the risk-free returns over this period (see Homer and Sylla (1991), for example). Instead, Siegel (1992) constructs the riskless short rates prior to 1920 by making certain assumptions about the similarities between the U.S. and U.K. term premia over this time (see his paper for the details). I follow this construction in order to extend the US short term bill record back to the beginning of my sample. The precise construction of the U.S. risk-free rate from 1836-1919 is as follows:

\[ i_{US} = i_{UK}^A - I_{UK}^A + I_{US}^A + \beta [CP_{US} - CP_{US}^A], \]  

(10)

where \( i_{US} \) is the constructed short-term U.S. rate, \( i_{UK} \) is the short-term U.K. rate (which is available back to 1836), \( CP_{US} \) is the U.S. commercial paper rate, and \( I_{US} \) and \( I_{UK} \) are the long-term U.S. and U.K. rates, respectively. In each case, an \( A \) superscript denotes a five-year, centered moving average. Finally, let \( \beta \) be the coefficient which reduces the standard deviation of the U.S. commercial paper rate to that of the U.K. risk-free rate. To eliminate outliers in the U.S. commercial paper, I also trim the data by removing the most (1%) extreme observations. At a practical level, this does not matter as the correlation between the total and excess return on the U.S. market from 1920-2003 (when we have direct U.S. risk free data) is 0.999.
Table 1 reports summary statistics and autocorrelation coefficients for returns on the U.S. market portfolio as measured by Schwert (1990) and Goetzmann, Ibbotson, and Peng (2001). These sources are employed from 1836-1925, after which the return reflects the CRSP value-weighted index of NYSE, AMEX, and Nasdaq firms. The Schwert (1990) data are total returns, including dividends, whereas the Goetzmann, Ibbotson, and Peng (2001) are capital gains only. \( r_{ft} \) is the return on the short bill index, available from the Global Financial Data provider. More detailed information on the data sources is provided in the appendix.
Table 2
Risk-Return Tradeoff: Alternative Volatility Specifications

<table>
<thead>
<tr>
<th></th>
<th>Panel A: GARCH</th>
<th></th>
<th>Panel B: TARCH</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_0 \times 1e4 )</td>
<td>1.042</td>
<td>1.904</td>
<td>1.107</td>
<td>0.973</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>0.237</td>
<td>0.646</td>
<td>0.400</td>
<td>0.452</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>0.114</td>
<td>0.158</td>
<td>0.118</td>
<td>0.083</td>
</tr>
<tr>
<td>( \delta_3 )</td>
<td>0.016</td>
<td>0.038</td>
<td>0.025</td>
<td>0.025</td>
</tr>
<tr>
<td>( \lambda_0 \times 100 )</td>
<td>0.029</td>
<td>-0.249</td>
<td>0.184</td>
<td>0.127</td>
</tr>
<tr>
<td>( \lambda_M )</td>
<td>2.456</td>
<td>2.872</td>
<td>1.844</td>
<td>3.255</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.001</td>
<td>0.005</td>
<td>0.001</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Panel C: Q-GARCH</th>
<th></th>
<th>Panel D: E-GARCH</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_0 \times 1e4 )</td>
<td>0.969</td>
<td>1.618</td>
<td>1.086</td>
<td>2.381</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>0.241</td>
<td>0.691</td>
<td>0.407</td>
<td>3.285</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>0.110</td>
<td>0.147</td>
<td>0.111</td>
<td>0.086</td>
</tr>
<tr>
<td>( \delta_3 )</td>
<td>0.016</td>
<td>0.039</td>
<td>0.024</td>
<td>0.047</td>
</tr>
<tr>
<td>( \lambda_0 \times 100 )</td>
<td>0.074</td>
<td>-0.255</td>
<td>0.240</td>
<td>0.986</td>
</tr>
<tr>
<td>( \lambda_M )</td>
<td>1.815</td>
<td>2.366</td>
<td>1.369</td>
<td>-2.055</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.002</td>
<td>0.004</td>
<td>0.002</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 2 presents evidence on the conditional mean and volatility of the U.S. equity market portfolio implied by the GARCH-M process over three different periods. The mean equation is:

\[
\begin{align*}
\mathbb{E}(r_{M,t+1}|\mathbb{F}_t) &= \lambda_0 + \lambda_M \sigma^2_{M,t} + \epsilon_{t+1} \\
\end{align*}
\]

The conditional volatility is described by one of the following GARCH processes:

**GARCH(1,1):**

\[
\sigma^2_{M,t+1} = \delta_0 + \delta_1 \epsilon^2_{t+1} + \delta_2 \sigma^2_{M,t}
\]

**TARCH(1,1):**

\[
\sigma^2_{M,t+1} = \delta_0 + \delta_1 \epsilon^2_{t+1} + \delta_2 D_{t+1} \epsilon^2_{t+1} + \delta_3 \sigma^2_{M,t}
\]

**Q-GARCH(1,1):**

\[
\sigma^2_{M,t+1} = \delta_0 + \delta_1 (\epsilon_{t+1}/\sigma_{M,t}) + \delta_2 (\epsilon_{t+1}/\sigma_{M,t}) + \delta_3 \ln(\sigma^2_{M,t})
\]

\( D_t \) takes the value of 1 when \( \epsilon_t < 0 \) and 0 otherwise. Standard errors are provided below the parameter estimates in italics.
Table 3
Risk Return Tradeoff
Alternative U.S. Historical Data (NYSE)

<table>
<thead>
<tr>
<th></th>
<th>GARCH</th>
<th>TARCH</th>
<th>Q-GARCH</th>
<th>E-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0 \times 1e4$</td>
<td>1.279</td>
<td>1.439</td>
<td>1.220</td>
<td>-0.394</td>
</tr>
<tr>
<td></td>
<td>0.242</td>
<td>0.278</td>
<td>0.252</td>
<td>0.072</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.140</td>
<td>0.088</td>
<td>0.139</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>0.018</td>
<td>0.019</td>
<td>0.018</td>
<td>0.025</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.801</td>
<td>0.795</td>
<td>0.793</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>0.023</td>
<td>0.024</td>
<td>0.024</td>
<td>0.011</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.091</td>
<td>0.013</td>
<td>-0.258</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.028</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 \times 100$</td>
<td>0.035</td>
<td>0.065</td>
<td>0.072</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>0.157</td>
<td>0.157</td>
<td>0.157</td>
<td>0.163</td>
</tr>
<tr>
<td>$\lambda_M$</td>
<td>2.588</td>
<td>2.096</td>
<td>1.897</td>
<td>2.788</td>
</tr>
<tr>
<td></td>
<td>0.814</td>
<td>0.825</td>
<td>0.834</td>
<td>0.855</td>
</tr>
<tr>
<td><strong>R-squared</strong></td>
<td><strong>0.000</strong></td>
<td><strong>0.001</strong></td>
<td><strong>0.002</strong></td>
<td><strong>0.000</strong></td>
</tr>
</tbody>
</table>

Table 3 presents evidence on the conditional mean and volatility of the U.S. equity market portfolio implied by the GARCH-M process over an alternative index representing the U.S. historical record. The 1836-1925 data reflect the NYSE capital appreciation index from Goetzmann, Ibbotson, and Peng (2001). The comparable 1926-2003 data are taken from the value-weighted CRSP indices for the NYSE, AMEX, and Nasdaq. The mean equation is (ignoring the risk free rate as this is the capital appreciation only):

$$ r_{M,t+1} = \lambda_0 + \lambda_M \sigma^2_{M,t} + \epsilon_{t+1} $$

The conditional volatility is described by the following GARCH processes:

- **GARCH(1,1):** $\sigma^2_{M,t+1} = \delta_0 + \delta_1 \epsilon_{t+1}^2 + \delta_2 \sigma^2_{M,t}$
- **TARCH(1,1):** $\sigma^2_{M,t+1} = \delta_0 + \delta_1 \epsilon_{t+1}^2 + \delta_3 D_{t+1} \epsilon_{t+1}^2 + \delta_2 \sigma^2_{M,t}$
- **Q-GARCH(1,1):** $\sigma^2_{M,t+1} = \delta_0 + \delta_1 (\epsilon_{t+1} - \delta_3)^2 + \delta_2 \sigma^2_{M,t}$
- **EGARCH(1,1):** $\ln(\sigma^2_{M,t+1}) = \delta_0 + \delta_1 (|\epsilon_{t+1}/\sigma_{M,t}| + \delta_3 (\epsilon_{t+1}/\sigma_{M,t})) + \delta_2 \ln(\sigma^2_{M,t})$

Standard errors are provided below the parameter estimates in italics.
Table 4
Risk-Return Tradeoff: United Kingdom

<table>
<thead>
<tr>
<th></th>
<th>Panel A: GARCH</th>
<th>Panel B: TARCH</th>
<th>Panel C: Q-GARCH</th>
<th>Panel D: E-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0 \times 10^4$</td>
<td>0.085</td>
<td>1.715</td>
<td>0.076</td>
<td>5.090</td>
</tr>
<tr>
<td></td>
<td>0.028</td>
<td>1.695</td>
<td>0.028</td>
<td>4.294</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.130</td>
<td>0.139</td>
<td>0.131</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>0.019</td>
<td>0.087</td>
<td>0.020</td>
<td>0.045</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.086</td>
<td>0.806</td>
<td>0.869</td>
<td>0.650</td>
</tr>
<tr>
<td></td>
<td>0.019</td>
<td>0.130</td>
<td>0.020</td>
<td>0.221</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td></td>
<td></td>
<td>0.010</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.014</td>
<td>0.114</td>
</tr>
<tr>
<td>$\lambda_0 \times 100$</td>
<td>0.017</td>
<td>0.026</td>
<td>0.020</td>
<td>-0.456</td>
</tr>
<tr>
<td></td>
<td>0.068</td>
<td>0.580</td>
<td>0.066</td>
<td>1.367</td>
</tr>
<tr>
<td>$\lambda_M$</td>
<td>2.469</td>
<td>2.822</td>
<td>2.289</td>
<td>4.722</td>
</tr>
<tr>
<td></td>
<td>0.906</td>
<td>2.647</td>
<td>0.830</td>
<td>5.768</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.007</td>
<td>0.004</td>
<td>0.007</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_0 \times 10^4$</td>
<td>0.084</td>
<td>3.812</td>
<td>-0.090</td>
<td>-0.325</td>
</tr>
<tr>
<td></td>
<td>0.029</td>
<td>1.508</td>
<td>0.035</td>
<td>0.172</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.130</td>
<td>0.170</td>
<td>0.265</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>0.020</td>
<td>0.068</td>
<td>0.032</td>
<td>0.076</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.086</td>
<td>0.676</td>
<td>0.986</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>0.020</td>
<td>0.096</td>
<td>0.005</td>
<td>0.028</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.000</td>
<td>0.016</td>
<td>-0.030</td>
<td>-0.184</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>0.013</td>
<td>0.051</td>
<td>0.150</td>
</tr>
<tr>
<td>$\lambda_0 \times 100$</td>
<td>0.017</td>
<td>-0.193</td>
<td>-0.013</td>
<td>-0.042</td>
</tr>
<tr>
<td></td>
<td>0.066</td>
<td>1.060</td>
<td>0.078</td>
<td>0.653</td>
</tr>
<tr>
<td>$\lambda_M$</td>
<td>2.471</td>
<td>3.585</td>
<td>3.006</td>
<td>3.676</td>
</tr>
<tr>
<td></td>
<td>0.906</td>
<td>4.718</td>
<td>1.156</td>
<td>2.888</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.007</td>
<td>0.002</td>
<td>0.007</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Table 4 presents evidence on the conditional mean and volatility of the U.K. equity market portfolio implied by the GARCH-M process over two different periods: the full historical and post-war samples. The mean equation is:

$$r_{M,t+1} - r_{f,t} = \lambda_0 + \lambda_M \sigma_{M,t}^2 + \varepsilon_{t+1}$$

The post-war sample is based on U.K. total equity market returns in excess of the U.K. risk free rate. The long historical sample is based only on the capital appreciation, ignoring the risk free rate. The conditional volatility is described by one of the following GARCH processes:

**GARCH(1,1):** $$\sigma_{M,t+1}^2 = \delta_0 + \delta_1 \varepsilon_{t+1}^2 + \delta_2 \sigma_{M,t}^2$$

**TARCH(1,1):** $$\sigma_{M,t+1}^2 = \delta_0 + \delta_1 \varepsilon_{t+1}^2 + \delta_3 D_t \varepsilon_{t+1}^2 + \delta_2 \sigma_{M,t}^2$$

**Q-GARCH(1,1):** $$\sigma_{M,t+1}^2 = \delta_0 + \delta_1 (|\varepsilon_{t+1}|/\sigma_{M,t})^2 + \delta_2 \sigma_{M,t}^2$$

**EGARCH(1,1):** $$\ln(\sigma_{M,t+1}^2) = \delta_0 + \delta_1 (|\varepsilon_{t+1}|/\sigma_{M,t}) + \delta_3 (\varepsilon_{t+1}/\sigma_{M,t}) + \delta_4 \ln(\sigma_{M,t}^2)$$

$D_t$ takes the value of 1 when $\varepsilon_t < 0$ and 0 otherwise. Standard errors are provided below the parameter estimates.
Table 5

Exploratory Analysis: A Time-Varying Risk Return Tradeoff

<table>
<thead>
<tr>
<th>External Factors and the Risk Return Tradeoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recession</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>δ₀ x 1e4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>δ₁</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>δ₂</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>λ₀ x 100</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>λₕ₀</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>λₘ₁ (Interaction)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Table 5 provides estimates for the GARCH-M specification where the risk-return tradeoff coefficient, \( \lambda_{M_i} = \lambda_{M_0} + \lambda_{M_1} \times X_t \), can potentially vary through time. As interaction variables \( X_t \), I consider one of several indicators of the state of the U.S. economy or financial system: a binary indicator that takes the value of one when the U.S. is in an NBER documented recession and 0 otherwise, the size of the equity market (MCAP/GDP), the number of listed companies (divided by 100 for presentation), a simple time trend, a measure of financial development (M3/GDP), the level of trade openness (Trade/GDP), or the size of the government (Gov/GDP). Due to historical data limitations, some estimates are available for shorter time periods for which the dates are provided. Standard errors are provided below the parameter estimates.
Figure 1
Time Series Plots of Asset Return Data

Figure 1 displays total returns, including dividends, on the U.S. market portfolio as measured by Schwert (1990). This source is employed from 1836-1925, after which the return reflects the CRSP value-weighted index of NYSE, AMEX, and Nasdaq firms; the vertical line separates the two samples. The risk free return on the short bill index is also presented, available from the Global Financial Data provider. More detailed information on the data sources is provided in the appendix.
Figure 2

U.S. Equity Returns and the Risk Premium

Figure 2 displays the U.S. total return along with the equity risk premium implied by an estimated GARCH-M specification (formally presented below in Table 2). The equity risk premium is obtained as $\lambda_M \sigma^2_{M,r}$. 
Figure 3a
Monte Carlo: $\lambda_M$
Estimated: GARCH(1,1)
DGP: GARCH(1,1)
Figure 3 presents Monte Carlo densities of the estimated risk return tradeoff coefficient, $\lambda_M$. The data are drawn from either a GARCH(1,1) or TARCH (1,1) process (Panels A and B, respectively), for which the true tradeoff coefficient equals 2. Small sample evidence is presented for $T = 500, 2,000$, and $5,000$ monthly observations.
Figure 4a
Monte Carlo: $\lambda_M$
Return Volatility

- observed return volatility
- 2x return volatility
- 1/2x return volatility
Figure 4b
Monte Carlo: $\lambda_M$
Volatility of Volatility

- red: observed volatility of volatility
- blue: 2x volatility of volatility
- dashed blue: 4x volatility of volatility
Figure 4c
Monte Carlo: $\lambda_M$
Volatility Persistence

Figure 4 presents Monte Carlo densities of the estimated risk return tradeoff coefficient, $\lambda_M$. I present evidence for three different specification adjustments to the data simulation: tuning return volatility (Panels A), the volatility of return volatility (Panel B), and the persistence of return volatility (Panel C). In all cases, the true tradeoff coefficient equals 2. Small sample evidence is presented for $T = 500$ monthly observations over the various alternatives considered.
Figure 5a
Monte Carlo: $\lambda_M$
Estimated: GARCH(1,1)
DGP: Continuous Time Diffusion - Stochastic Volatility

Figure 5b

Monte Carlo: $\lambda_M$

Estimated: GARCH(1,1)

DGP: Continuous Time Diffusion - Stochastic Volatility

Figure 5 presents Monte Carlo densities of the estimated risk return tradeoff coefficient, $\lambda_M$. The data are drawn from the continuous-time process detailed in equation (8), where the true tradeoff coefficient equals 2. Small sample evidence is presented for $T = 500$, 2,000, and 5,000 discretely observed monthly observations (in Panel A), and for 500 monthly, 2,000, weekly, or 11,000 daily observations (in Panel B). The former expands the data span, but keeps observation frequency constant, whereas the latter increases the observation frequency, but holds data span constant.
Figure 6 displays the U.S. conditional market volatility implied by the estimated GARCH(1,1) specification in Table 2 over the full historical record. I also plot the financial crisis dates identified by Schwert (1989) (adding October 1987 and October 1998).