Essays in Asset Pricing and the Econometrics of Risk

by

Bryan T. Kelly

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
Department of Finance
Stern School of Business
New York University

April 2010

Robert F. Engle
## Contents

1. Risk Premia and the Conditional Tails of Stock Returns ................................................................. 1
   1.1 Introduction .................................................................................................................................. 2
   1.2 Asset Pricing Theory and Conditional Return Tails ..................................................................... 7
      1.2.1 Economy I: Long Run Risks and Tail Risk in Cash Flows .................................................... 8
      1.2.2 Economy II: Time-Varying Rare Disasters ........................................................................... 12
      1.2.3 Testable Implications ............................................................................................................. 15
   1.3 Empirical Methodology .................................................................................................................. 16
      1.3.1 The Dynamic Power Law Model ............................................................................................ 16
      1.3.2 Estimating the Dynamic Power Law Model ........................................................................... 20
      1.3.3 Volatility and Heterogeneous Exceedence Probabilities ..................................................... 24
      1.3.4 Monte Carlo Evidence ........................................................................................................... 25
   1.4 Empirical Results ............................................................................................................................ 26
      1.4.1 Tail Risk Estimates .................................................................................................................. 26
      1.4.2 Predicting Stock Market Returns ............................................................................................ 29
      1.4.3 Tail Risk and the Cross Section of Expected Returns ........................................................... 33
   1.5 Conclusion ..................................................................................................................................... 35

2. Dynamic Equicorrelation (with Robert Engle) .................................................................................. 57
   2.1 Introduction .................................................................................................................................... 58
2.2 The Dynamic Equicorrelation Model

2.2.1 Estimation

2.2.2 Differences Between DECO and DCC

2.2.3 DECO as a Feasible DCC Estimator

2.2.4 Estimation Structure Versus Fit Structure

2.2.5 The Block Dynamic Equicorrelation Model

2.3 Correlation Monte Carlos

2.3.1 Equicorrelated Processes

2.3.2 Non-Equicorrelated Processes

2.4 Empirical Analysis

2.4.1 Data

2.4.2 Dynamic Equicorrelation in the S&P 500, 1995-2008

2.4.3 Factor ARCH DECO

2.4.4 Comparing DECO and DCC Correlations

2.4.5 Block DECO

2.4.6 Equicorrelation and Implied Correlations, Dow Jones Index

2.4.7 Out-of-Sample Hedging Performance

2.5 Conclusion

3 Testing Asymmetric-Information Asset Pricing Models (with Alexander Ljungqvist)

3.1 Introduction

3.2 The Model

3.2.1 Equilibrium Effects

3.2.2 Discussion

3.2.3 Robustness
4 A Practical Guide to Volatility Forecasting (with Christian Brownlees and Robert Engle)
Chapter 1

Risk Premia and the Conditional Tails of Stock Returns

Abstract: Theory suggests that the risk of infrequent yet extreme events has a large impact on asset prices. Testing models of this hypothesis remains a challenge due to the difficulty of measuring tail risk fluctuations over time. I propose a new measure of time-varying tail risk that is motivated by asset pricing theory and is directly estimable from the cross section of returns. My procedure applies Hill’s (1975) tail risk estimator to the cross section of extreme events each day. It then optimally averages recent cross-sectional Hill estimates to provide conditional tail risk forecasts. Empirically, my measure has strong predictive power for aggregate market returns, outperforming all commonly studied predictor variables. I find that a one standard deviation increase in tail risk forecasts an increase in excess market returns of 4.4% over the following year. Cross-sectionally, stocks that highly positively covary with my tail risk measure earn average annual returns 6.0% lower than stocks with low tail risk covariation. I show that these results are consistent with predictions from two structural models: i) a long run risks economy with heavy-tailed consumption and dividend growth shocks, and ii) a time-varying rare disaster framework.

\[ \text{1} \]

\[ \text{1} \]

I thank my thesis committee, Robert Engle (chair), Xavier Gabaix, Alexander Ljungqvist and Stijn Van Nieuwerburgh for many valuable discussions. I also thank Mikhail Chernov, Itamar Drechsler, Marcin Kacperczyk, Anthony Lynch and Seth Pruitt for insightful comments.
1.1 Introduction

The mere potential for infrequent events of extreme magnitude can have important effects on asset prices. Tail risk, by nature, is an elusive quantity, which presents economists with the daunting task of explaining market behavior with rarely observed phenomena. This crux has led to notions such as peso problems (Krasker 1980) and the rare disaster hypothesis (Rietz 1988; Barro 2006), as well as skepticism about these theories due to the difficulty in testing them.

The goal of this paper is to investigate the effects of time-varying extreme event risk in asset markets. The chief obstacle to this investigation is a viable measure of tail risk over time. To overcome this, I devise a panel approach to estimating economy-wide conditional tail risk. Working from standard asset pricing models, I show that tail risks of all firms are driven by a common underlying process. Because individual returns contain information about the likelihood of market-wide extremes, the cross section of firms can be used to accurately measure prevailing tail risk in the economy. I elicit a conditional tail estimate by turning to the cross section of extreme events at each point in time, rather than waiting to accumulate a sufficient number of extreme observations in univariate time series. This bypasses data limitations faced by alternative estimators, for example those relying on options prices or intra-daily data.

My framework, which fuses asset pricing theory with extreme value econometrics, distills to a central postulate for the tail distribution of returns. Define the tail as the set of return events exceeding some high threshold $u$. I assume that the tail of asset return $i$ behaves according to

$$P(R_{i,t+1} > r \mid R_{i,t+1} > u \text{ and } \mathcal{F}_t) = \left(\frac{r}{u}\right)^{-a_i \zeta_t}. \quad (1.1)$$

Equation (1.1) states that extreme return events obey a power law. Since at least Mandelbrot (1963) and Fama (1963), economists have argued that unconditional tail distributions of
financial returns are aptly described by a power law. The key parameter of the model, $-a_i \zeta_t$, determines the shape of the tail and is referred to as the tail exponent. High values of $-a_i \zeta_t$ correspond to “fat” tails and high probabilities of extreme returns.

In contrast to past power law research, Equation [1.1] is a statement about the conditional return tail. The exponent varies over time because $\zeta_t$ is a function of the conditioning information set $\mathcal{F}_t$. While different assets have different levels of tail risk (determined by the constant $a_i$), dynamics are the same for all assets because they are driven by a common conditional process. Thus, $\zeta_t$ may be thought of as economy-wide extreme event risk in returns. I refer to the tail structure in [1.1] as the dynamic power law model.

The tail generating process in [1.1] arises naturally from at least two structural models: i) a long run risks economy (Bansal and Yaron 2004) modified to include heavy-tailed shocks, and ii) a time-varying rare disaster framework. In my long run risks modification, non-Gaussian tails in consumption and dividend growth are governed by a new tail risk state variable, $\Lambda_t$. I show that expected excess returns depend linearly on the tail risk process. I then prove that the tail distribution of returns behaves according to the dynamic power law model. In particular, tail exponent fluctuations are the same for all assets and driven by $\Lambda_t$.

An attractive feature of the dynamic power law structure is that it emerges in varied theoretical settings rather than being tied to a single modeling paradigm. I demonstrate that a second structural model with unpredictable consumption growth and time-varying rare disasters, in the spirit of Gabaix (2009b) and Wachter (2009), delivers the dynamic power law structure for the lower tail of returns.

These structural models tightly link the time-varying tail exponent to expected excess returns on risky assets since both are driven by the tail risk process $\Lambda_t$. This generates two key testable implications. First, tail risk should positively forecast excess aggregate market returns. Aggregate dividends have substantial exposure to consumption tail risk. Thus, a positive tail risk shock increases the return required by investors to hold the market. Because
the tail process is persistent, its shocks have a long-lived effect. Return forecastability arises because future expected excess returns remain high until the expectations effect of a tail risk shock dies out.

The second testable prediction applies to the cross section of expected returns. High tail risk is associated with bad states of the world and high marginal utility. This implies that the price of tail risk is negative, hence assets with high betas on the tail risk process will have lower expected returns than assets with low tail risk betas. Intuitively, an asset whose return covaries highly with tail risk has a tendency to payoff in adverse states, serving as an effective tail risk hedge. As a result, it commands a high price and earns relatively low average returns.

I build an econometric estimator for the dynamic power law structure suggested by these economic frameworks. The intuition from structural models is that tail risks of individual assets are closely related to aggregate tail risk. In a sufficiently large cross section, enough stocks will experience tail events each period to provide accurate information about the prevailing level of tail risk. I use this cross-sectional extreme return information to estimate economy-wide tail risk at each point in time. This avoids having to accumulate years of tail observations from the aggregate market time series, and therefore avoids using stale observations that carry little information about current tail risk.

My procedure applies Hill’s (1975) tail risk estimator to the cross section of extreme events each day. The model then optimally averages recent cross-sectional Hill estimates to provide conditional tail risk forecasts. A major obstacle in estimation is the model’s potentially enormous number of nuisance parameters. I overcome this with a strategy based on quasi-maximum likelihood theory. The idea is to find a simpler version of the infeasible model that has the same maximum likelihood first order conditions. Estimation may then be based on this mis-specified, yet feasible, model. I reduce the complexity of the problem to three parameters by treating observations as though they are i.i.d. I then prove that maximizing
the resulting quasi-likelihood provides consistent and asymptotically normal estimates of the data’s true dynamics, and show how to calculate standard errors for inference.

I implement the dynamic power law estimator using daily returns from the cross section of CRSP stocks. I find that the cross-sectional average tail exponent is highly persistent and fluctuates between -4 and -1.5. This range is consistent with a survey by Gabaix (2009a), who finds that estimates of unconditional tail exponents consistently hover around -3 based on data for a variety of domestic and international equities. My estimates of lower and upper tail risk are positively correlated (56% monthly). There is evidence of cyclicality in lower (upper) tail risk as it shares a monthly correlation of 53% (39%) with unemployment, 15% (14%) with the aggregate log dividend-price ratio, and -10% (-7%) with the Chicago Fed National Activity Index.

Using the fitted tail series, I first test the model prediction that tail risk should forecast aggregate stock market returns. Predictive regressions show that a one standard deviation increase in the risk of negative tail events forecasts an increase in annualized excess market returns of 6.7%, 4.4%, 4.5% and 5.0% at the one month, one year, three year and five year horizons, respectively. These are all statistically significant with t-statistics of 2.9, 2.1, 2.2 and 2.3, based on Hodrick’s (1992) standard error correction. At the monthly frequency, tail risk achieves an $R^2$ of 1.6% in-sample and 1.3% out-of-sample, outperforming the price-dividend ratio and other common predictors. Cochrane (1999) and Campbell and Thompson (2008) argue that a monthly $R^2$ of this magnitude has large economic significance. A heuristic calculation suggests that a 1.3% $R^2$ can generate a 50% improvement in Sharpe ratio over a buy-and-hold investor. My tail risk measure outperforms all commonly considered forecasting variables in terms of predictive power, including the log dividend-price.

\[ s^* = \frac{\sqrt{\hat{R}^2 + R^2}}{\sqrt{1 - R^2}} \]

Cochrane shows that the Sharpe ratio ($s^*$) earned by an active investor exploiting predictive information (summarized by a regression $\hat{R}^2$) and the Sharpe ratio ($s_0$) earned by a buy-and-hold investor are related by $s^* = \frac{\sqrt{\hat{R}^2 + R^2}}{\sqrt{1 - R^2}}$. Campbell and Thompson estimate a monthly equity buy-and-hold Sharpe ratio of 0.108 using data back to 1871. Therefore, a predictive $R^2$ of 1.3% implies $s^* = 0.162$, an improvement of 50% over the buy-and-hold value.
ratio and fourteen other variables surveyed by Goyal and Welch (2008). Estimated tail risk coefficients and their statistical significance are robust to controlling for these alternative predictors. They are also robust to using an alternative tail series based on factor model residuals rather than raw returns.

The tail exponent has large, significant explanatory power for the cross section of returns in the direction predicted by theory. In my first test, I sort stocks into quintiles based on their estimated beta on the tail risk process. I find that stocks with the highest tail risk betas earn average annual returns 6.0% lower than the lowest tail risk beta stocks ($t$-statistic=2.6). This negative tail risk premium is robust to double sorts based on tail risk beta and i) market beta, ii) size or iii) book-to-market ratio. To simultaneously control for multiple alternative factors, I next test the tail risk premium with Fama and MacBeth (1973) regressions. Based on NYSE, AMEX and NASDAQ stocks, I find that a stock whose tail beta is one standard deviation above the cross-sectional mean has an annual expected return 5.6% lower than the average stock ($t$-statistic=2.9) after controlling for market volatility and Fama and French (1993) factors, consistent with theory and with results based on portfolio sorts. These findings are also robust to estimating the tail exponent with factor model residuals, and to using alternative sets of test assets.

My research draws on several literatures. Theoretically, I build on the long run risks literature of Bansal and Yaron (2004). My framework is most closely related to recent long run risks extensions that accommodate more sophisticated descriptions of consumption growth, particularly Eraker and Shaliastovich (2008), Bansal and Shaliastovich (2008, 2009), and Drechsler and Yaron (2009). The Rietz-Barro hypothesis and its extensions to dynamic settings by Gabaix (2009b) and Wachter (2009) are also important predecessors of the ideas developed here.

A large literature has modeled extreme returns using jump processes with time-varying intensities that depend on observable state variables, including the widely used affine class
of Duffie, Pan and Singleton (2000). My approach, which models conditional tails in discrete
time and uses observable parameter updates based on the history of extreme returns, is new.
Furthermore, the notion of extracting information about common, time-varying tails from
the cross section of returns is novel, though similar in spirit to the identification strategy in

Recently, economists have extracted tail risk estimates from options prices or intra-daily
returns to assess the relation between rare events and equity prices. Bollerslev, Tauchen and
Zhou (2009) examine how the variance risk premium implicit in index option prices relates
to the equity premium. Backus, Chernov and Martin (2009) use equity index options to
extract higher order return cumulants and draw inferences about disaster risk premia over
calculate realized and risk neutral tail risk measures to explain equity and variance risk
premia. In these cases, researchers are bound by data limitations that restrict the sample
horizon to at most twenty years. In contrast, my tail risk series is estimated using data
from 1962 to 2008, and in general may be used whenever a sufficiently large cross section of
returns is available.

Lastly, I contribute to a literature that attempts to jointly explain behavior of returns
in the time series and cross section, including (among others) Ferson and Harvey (1991),
Lettau and Ludvigson (2001a,b), Lustig and Van Nieuwerburgh (2005) and Koijen, Lustig
and Van Nieuwerburgh (2009).

1.2 Asset Pricing Theory and Conditional Return Tails

In this section I develop two consumption-based asset pricing models. Both generate a
dynamic power law structure in the tail distribution of returns and produce clear testable
implications for the link between tail risk and risk premia. The first is a modification of the
long run risks economy of Bansal and Yaron (2004). In addition to the standard aspects of the Bansal-Yaron formulation, I allow for non-Gaussian shocks to both consumption growth and idiosyncratic dividend growth. My specification of tail risk leads to tractable expressions both for prices and for the tail distribution of returns.

The second example economy is a version of the time-varying rare disaster model of Gabaix (2009b) and Wachter (2009). This example is included to demonstrate the flexibility of the dynamic power law structure for encompassing a broad set of distinct economic models of returns.

1.2.1 Economy I: Long Run Risks and Tail Risk in Cash Flows

Investor preferences over consumption are recursive (Epstein and Zin 1989). These are summarized by the economy-wide intertemporal marginal rate of substitution, which is also the stochastic discount factor that prices assets in the economy. Written in its log form, this is

\[ m_{t+1} = \theta \ln \beta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1) r_{c,t+1}, \]

where \( \theta = \frac{1-\gamma}{1-\psi} \), \( \gamma \) is the risk aversion coefficient, \( \psi \) is the intertemporal elasticity of substitution (IES), \( \Delta c_{t+1} \) is log consumption growth, and \( r_{c,t+1} \) is the log return on an asset paying aggregate consumption each period. Throughout the paper I assume \( \gamma > 1 \) and \( \psi > 1 \), which implies \( \theta < 0 \). These parameter restrictions ensure that risk aversion is greater than the reciprocal of IES, therefore agents have a preference for early uncertainty resolution.
Dynamics of the real economy are

\[
\Delta c_{t+1} = \mu + x_t + \sigma_c \sigma_t z_{c,t+1} + \sqrt{\Lambda_t} W_{c,t+1}
\]

\[
x_{t+1} = \rho x_t + \sigma x_t z_{x,t+1}
\]

\[
\sigma^2_{t+1} = \tilde{\sigma}^2 (1 - \rho_s) + \rho_s \sigma^2_t + \sigma_s z_{\sigma,t+1}
\]

\[
\Lambda_{t+1} = \tilde{\Lambda} (1 - \rho_{\Lambda}) + \rho_{\Lambda} \Lambda_t + \sigma_{\Lambda} z_{\Lambda,t+1}
\]

\[
\Delta d_{i,t+1} = \mu_i + \phi_i \Delta c_{t+1} + \sigma_i \sigma_t z_{i,t+1} + q_i \sqrt{\Lambda_t} W_{i,t+1}.
\]  

(1.2)

Included in this specification are standard elements of a long run risks model: log consumption growth (\(\Delta c_{t+1}\)), its persistent conditional mean (\(x_t\)) and volatility (\(\sigma_t\)), and dividend growth for asset \(i\) (\(\Delta d_{i,t+1}\)). The \(z\) shocks are standard normal and independent. In addition to their Gaussian shocks, consumption and dividend growth depend on non-Gaussian shocks, \(W\). These are independent unit Laplace variables with mean zero. Their density results from splicing the densities of independent positive and negative unit exponentials together at zero,

\[f_W(w) = \frac{1}{2} \exp(-|w|), \ w \in \mathbb{R}.
\]

The Laplace shocks dominate the tails of cash flow growth. To see this, consider the tail distribution of \(Z = X + Y\), where \(X \sim N(0, \tau)\) and \(Y\) is an independent Laplace variable with scale parameter \(\alpha\). It may be shown that the tail behavior of \(Z\) is determined solely by the Laplace summand, \(P(Z > u + \eta | Z > u) \sim \exp(-\alpha \eta)\). The relation \(f(u) \sim g(u)\), read “\(f\) is asymptotically equivalent (or tail equivalent) to \(g\)” denotes \(\lim_{u \to \infty} f(u)/g(u) = 1\).

Heavy-tailed consumption growth and dividend growth shocks \(W_{c,t+1}\) and \(W_{i,t+1}\) are scaled by \(\sqrt{\Lambda_t}\) and \(q_i \sqrt{\Lambda_t}\), respectively. As the Gaussian stochastic process \(\Lambda_t\) evolves, the risk of extreme cash flow events fluctuates. High values of \(\Lambda_t\) fatten the tails of cash flow shocks while low values shrink cash flow tails; consequently, I refer to \(\Lambda_t\) as the tail risk.
I solve the model with procedures commonly employed in consumption-based affine pricing models, following Bansal and Yaron (2004), Eraker and Shaliastovich (2008), and Bollerslev, Tauchen and Zhou (2009), among others. The first result proves that log valuation ratios in the economy are linear in the tail risk process.

Proposition 1. The log wealth-consumption ratio and log price-dividend ratio for asset $i$ are linear in state variables,

$$\begin{align*}
wc_{t+1} &= A_0 + A_x x_{t+1} + A_\sigma \sigma_{t+1}^2 + A_\Lambda \Lambda_{t+1} \\
\end{align*}$$

(1.3)

$$\begin{align*}
pd_{i,t+1} &= A_{i,0} + A_{i,x} x_{t+1} + A_{i,\sigma} \sigma_{t+1}^2 + A_{i,\Lambda} \Lambda_{t+1}.
\end{align*}$$

Proofs are relegated to Appendix A.1 (including expressions for the $A$ constants).

Risk premia for an asset are determined by covariation between $m_{t+1}$ and the asset’s log return. Hence, the coefficients on shocks to the stochastic discount factor take on the interpretation of risk prices. Discount factor shocks are

$$m_{t+1} - E_t [m_{t+1}] = -\lambda_c (\sigma_c \sigma_{t+1} + \sqrt{\Lambda_t W_{c,t+1}}) - \lambda_x \sigma_x \sigma_{t+1} - \lambda_\sigma \sigma_{t+1} - \lambda_\Lambda \Lambda_{t+1}.$$  

The term $\lambda_\Lambda$ captures the price of tail risk (risk price expressions are also shown in Appendix A.1). The tail risk price is negative since an increase in uncertainty decreases agents’ utility. Assets that covary positively with tail risk behave as hedges since they tend to pay off when marginal utility is high. Since tail risk hedges are valuable to investors, they command higher prices and lower expected returns, ceteris paribus. The next result shows that an asset’s risk premium is a linear function of variance and tail risk.

---

3Analytical results are stated subject to linear approximations such as the log return identity of Campbell and Shiller (1988), as used in the aforementioned articles.
Proposition 2. The expected return on asset \( i \) in excess of the risk free rate is

\[
E_t[r_{i,t+1} - r_{f,t}] = \beta_{i,c} \lambda_c (\sigma_c^2 \sigma_i^2 + 2 \Lambda_t) + \beta_{i,x} \lambda_x \sigma_x^2 \sigma_i^2 + \beta_{i,\sigma} \lambda_\sigma \sigma_\sigma^2 + \beta_{i,\Lambda} \lambda_\Lambda \sigma_\Lambda^2 - \frac{1}{2} Var(r_{i,t+1}). \tag{1.4}
\]

Proposition 2 describes both the time series and cross-sectional relation between tail risk and expected returns. First, Equation 1.4 is a predictive regression that implies returns are forecastable by variance and tail risk in equilibrium. This is a natural consequence of predictable changes in compensation that investors require in order to bear these risks, and is not an arbitrage opportunity or violation of efficient markets. Let subscript \( m \) denote the asset that pays aggregate dividends (i.e., the market portfolio). In the modified long run risks model, the predictive regression coefficient on tail risk is \( 2 \beta_{m,c} \lambda_c \). The constant \( \beta_{m,c} \) is positive as long as the exposure of aggregate dividend growth to consumption growth is greater than the reciprocal of IES (\( \phi_m > 1/\psi \)). This is typically assumed in calibrations of long run risks models, where aggregate dividends are treated as levered claims on consumption (\( \phi_m > 1 \), as in Bansal and Yaron 2004). The price of transitory consumption risk \( \lambda_c \) is also positive, which delivers the intuitive implication that the predictive coefficient is positive. Higher tail risk increases the return investors require to hold the market portfolio going forward.

The term \( \beta_{i,\Lambda} \lambda_\Lambda \sigma_\Lambda^2 \) governs cross-sectional differences in expected return due to differential exposure to the tail risk process. Because tail risk carries a negative price of risk, Equation 1.4 generates the prediction that stocks with high tail risk betas (\( \beta_{i,\Lambda} \)) earn a negative risk premium over low tail risk beta stocks. High tail risk beta stocks perform well when tail risk is high. Because tail risk is utility decreasing, these stocks serve as effective hedges. They therefore command a high price and earn low expected returns.

---

4 The predictive regression will also be affected by the Jensen term \( \frac{1}{2} Var(r_{i,t+1}) \) since it is a function of \( \Lambda_t \). As long as the Jensen effect is small relative to the risk premium effect (\( 2 \beta_{m,c} \lambda_c \)), the positive sign of the predictive coefficient will still hold.
Using the price expressions in Proposition 1, I show that the tail distribution of arithmetic returns for each stock satisfies the dynamic power law structure in Equation 1.1.

**Proposition 3.** The lower and upper tail distributions of arithmetic returns are asymptotically equivalent to a power law,

\[
P_t(R_{i,t+1} < r \mid R_{i,t+1} < u) \sim \left( \frac{r}{u} \right)^{a_i \zeta_t},
\]

\[
P_t(R_{i,t+1} > r \mid R_{i,t+1} > u) \sim \left( \frac{r}{u} \right)^{-a_i \zeta_t},
\]

where \(a_i = \max(\phi_i, q_i)^{-1}\) and \(\zeta_t = 1/\sqrt{\Lambda_t}\).

Here the relation \((\sim)\) describes tail equivalence at the lower and upper support boundaries of \(R_{i,t+1}\) (i.e., \(\lim_{u \downarrow 0} f(u)/g(u) = 1\) for the lower tail or \(\lim_{u \to \infty} f(u)/g(u) = 1\) for the upper tail). The value \(r\) is assumed to vary in fixed proportion with \(u\) (i.e., \(r = u\eta, \eta > 0\)). As \(a_i \zeta_t\) decreases, both the upper and lower tails become fatter. This proposition demonstrates the link between tail risk in cash flows and returns. The process governing consumption and dividend growth tail risk, \(\Lambda_t\), drives time variation in the parameter governing the tail distribution of returns, \(\zeta_t\). The key insight of this result is that extreme event risk in the real economy may be estimated from the tail distribution of returns. It also means that the model implications stated above may be tested based on parameter estimates for the conditional tail distribution of returns.

### 1.2.2 Economy II: Time-Varying Rare Disasters

In this subsection I develop an economy with variable rare disasters in consumption growth and idiosyncratic dividend growth. I am brief in my exposition of the rare disaster model as

\footnote{Note that with a trivial reformulation, the lower tail of the distribution can be written identically to the upper tail distribution. This is done by reversing the sign of the lower tail of log returns before exponentiating, which is clear from the proof in Appendix A.1}
much of the intuition carries over from the first model economy.

Investor preferences again take the Epstein-Zin form. In contrast to the long run risks model, consumption growth is unpredictable in a rare disaster economy. In most periods, consumption growth is Gaussian. Upon the rare occurrence of a disaster, consumption growth experiences a heavy-tailed negative shock. The severity of disasters varies through time, so that consumption growth takes the form

\[ \Delta c_{t+1} = \mu + \sigma_c \sigma_t z_{c,t+1} - \iota_{c,t+1} \Lambda_t V_{c,t+1}. \]

The first shock \( z_{c,t+1} \) is standard normal. In non-disaster times, variations in the consumption growth distribution arise only from heteroskedasticity (\( \sigma_t^2 \), which is unchanged from system 1.2). The second shock, \( V_{c,t+1} \), is the disaster shock. It is drawn from the relatively heavy-tailed unit exponential distribution. \( V_{c,t+1} \) is first multiplied by a Bernoulli(\( \delta \)) random variable, \( \iota_{c,t+1} \), that determines whether or not a disaster occurs at \( t + 1 \). The occurrence of a disaster therefore follows the distribution

\[ \iota_{c,t+1} = \begin{cases} 1 & \text{with probability } \delta \\ 0 & \text{with probability } 1 - \delta. \end{cases} \]

Second, \( V_{c,t+1} \) is multiplied by \( \Lambda_t \), which determines the severity of a disaster. In this context, I refer to \( \Lambda_t \) as the disaster risk process. It is stochastic and evolves as in system 1.2.

Dividend growth of stock \( i \) is given by

\[ \Delta d_{i,t+1} = \mu_i + \phi_i \Delta c_{t+1} + \sigma_i \sigma_t z_{i,t+1} - q_{i,t+1} \Lambda_t V_{i,t+1}. \]

As in the time-varying rare disaster model of Gabaix (2009b), individual stock dividends

---

6It is straightforward to allow for a time-varying disaster probability \( \delta_t \).
have exposure to aggregate consumption, and hence aggregate disaster risk. I also allow
for the possibility of idiosyncratic dividend disasters. These are associated with severe neg-
ative shocks to firms’ idiosyncratic payoffs that are independent of the broader economy.
Rare idiosyncratic disasters occur through the shock \( q_i \iota_{i,t+1} \Lambda_t V_{i,t+1} \), where \( \iota_{i,t+1} \) is an i.i.d. \( \text{Bernoulli}(\delta) \) variable and \( q_i \) determines the magnitude of stock \( i \)’s idiosyncratic disasters relative to consumption disasters.

The following result demonstrates that log valuation ratios are linear in disaster risk.

**Proposition 4.** The log wealth-consumption ratio and log price-dividend ratio of asset \( i \) are linear in state variables,

\[
wc_{t+1} = A_0 + A_\sigma \sigma^2_{t+1} + A_\Lambda \Lambda_{t+1}
\]

\[
pd_{i,t+1} = A_{i,0} + A_{i,\sigma} \sigma^2_{t+1} + A_{i,\Lambda} \Lambda_{t+1}.
\]

Expressions for the \( A \) coefficients are found in Appendix A.1. As in the long run risks model, \( (1 - \theta)_{\kappa_1} A_\Lambda \) may be interpreted as the price of disaster risk. The proof of Proposition 4 shows that \( A_\Lambda < 0 \), which implies that disaster risk has a negative price. The next result shows that the risk premium for each asset is linear in disaster risk.

**Proposition 5.** The expected return on asset \( i \) in excess of the risk free rate is

\[
E_t[r_{i,t+1} - r_{f,t}] = r_{i,0} + b_{i,\sigma} \sigma^2_t + b_{i,\Lambda} \Lambda_t.
\]

Lastly, I prove that the lower tail structure of returns generated by the time-varying rare disaster model is consistent with the dynamic power law model of Equation 1.1.

**Proposition 6.** The lower tail distribution of arithmetic returns is asymptotically equivalent to a power law\(^7\)

\[
P_t(R_{i,t+1} < r \mid R_{i,t+1} < u) \sim \left( \frac{r}{u} \right)^{a_i \zeta_t}
\]

\(^7\)As in Proposition 3, the value \( r \) is assumed to vary in fixed proportion with \( u \) \( (r = u\eta, \eta > 0) \).
where \( a_i = \max(\phi_i, q_i)^{-1} \) and \( \zeta_t = 1/\Lambda_t \).

Because the disaster shock is strictly negative, only the lower tail of returns exhibits power law behavior. The upper tail is lognormal since large upside moves result only from Gaussian shocks. The model can be easily reformulated to accommodate time-varying rare booms alongside rare disasters. The qualitative effects of tail risk in this case are similar to those from the disaster and long run risks models I’ve presented.

### 1.2.3 Testable Implications

To summarize, structural economic models predict a close link between the risk of extreme events in the real economy, \( \Lambda_t \), and risk premia across assets and over time. Direct estimation of conditional tail risk from consumption and dividend data is essentially infeasible due to their infrequent observation and poor measurement. The two structural models presented here highlight the path to an alternative estimation strategy since the power law tail exponent of stock returns, \( -\zeta_t \), is also driven by \( \Lambda_t \). Because returns are frequently and precisely observed, estimates of their tail distribution identify the \( \Lambda_t \) process. Most importantly, the tight structure that these models place on the return tail distribution implies that the cross section can be exploited to extract conditional tail risk estimates at high frequencies.

The model-implied pricing effects of tail risk can be tested with estimates of the time-varying component in return power law exponents. The exponent series \( -\zeta_t \) is an increasing function of \( \Lambda_t \), so that when cash flow tail risk rises, return tails become fatter.\(^8\) A preliminary assumption of the economic model is that economy-wide tail risk \( \Lambda_t \) varies persistently through time, which implies that \( -\zeta_t \) should as well. This assumption is testable.

---

\(^8\)Different models imply different relations between the tail exponent, \( -\zeta_t \), and tail risk in fundamentals, \( \Lambda_t \). In the long run risks model \( \Lambda_t = 1/\zeta_t^2 \), while in the rare disaster model \( \Lambda_t = 1/\zeta_t \). Specifications of the \( \Lambda_t \) process are flexible and can generate a wide range of functional forms linking \( \Lambda_t \) and \( -\zeta \). Under any specification, however, it is the case that \( -\zeta_t \) is increasing in \( \Lambda_t \). Rather than rely on a model-specific functional link between these two, I test model implications using the estimated tail exponent \( -\zeta_t \) without further transformation. My empirical results are robust to transformations based on the economic models here.
Testable Implication 1. The dynamic power law exponent $-\zeta_t$ is time-varying and persistent.

The next implication applies to the equity premium time series. Equations 1.4 and 1.6 imply that aggregate tail risk forecasts excess returns on the market portfolio. Because aggregate dividends have substantial exposure to consumption tail risk, a positive tail risk shock increases the return required by investors to hold the market. Since the tail process is persistent, future expected excess returns remain high until the expectations effect of a tail risk shock dies out.

Testable Implication 2. The tail risk series $-\zeta_t$ positively forecasts excess market returns.

Next, because tail risk detracts from utility, it carries a negative price of risk. This generates the cross-sectional prediction that stocks with positive tail risk betas earn a negative risk premium. This should also be true of the tail risk proxy $-\zeta_t$.

Testable Implication 3. Stock with high betas on the tail risk process $-\zeta_t$ earn a negative risk premium in relation to those with low tail risk betas.

1.3 Empirical Methodology

1.3.1 The Dynamic Power Law Model

In this section I propose a procedure for estimating the dynamic power law model. My approach exploits the comparatively rich information about tail risk in the cross section of returns, as opposed to relying, for example, on short samples of high frequency univariate data or options prices.

Estimating fully-specified versions of the tail models from Section 2 is extremely difficult, and essentially infeasible without multi-step estimation. It requires specifying a dependence
structure among return tails and estimating stock-specific $a_i$ parameters. Incorporating both considerations adds an enormous number of parameters: Estimating the $a_i$ constants adds $n$ parameters while imposing dependence structures like those implied by the models in Section 2 adds another $nK$ parameters, where $K$ is the number of factors in a given model. Furthermore, these parameters are nuisances since the goal is to measure the common element of tail risk, not univariate distributions. The stochastic nature of the tail risk process further complicates estimation. Contemporaneous shocks to the tail exponent can be thought of as the extreme value equivalent of stochastic volatility. It rules out simple likelihood methods, instead requiring computation-intensive procedures like simulation-based estimation.

I propose several simplifications of the dynamic power law model that isolate the common component of tail risk with a tractable, accurate procedure. My simplification requires estimating only three parameters instead of several thousand, and reduces estimation time to under one minute despite working with a daily cross section of several thousand stocks. To begin, I more fully specify the statistical model, including an evolution equation for the tail exponent (which was left unspecified in Equation 1.1).

Assumption 1 (Dynamic Power Law Model). Let $R_t = (R_{1,t}, \ldots, R_{n,t})'$ denote the cross section of returns in period $t$. Let $K_t$ denote the number of $R_t$ elements exceeding threshold $u$ in period $t$. The tail of individual returns for stock $i$ ($i = 1, \ldots, n$), conditional upon

---

9In Section 2, returns obey a tail factor structure due to the factor structure in heavy-tailed cash flow shocks. The common heavy-tailed shock enters via each firm’s exposure to consumption growth while the heavy-tailed idiosyncrasy comes from firm-specific dividend shocks. In that case, $K = 1$.

10$R$ denotes arithmetic return, which directly maps the tail distribution here with theoretical results in Section 2. This is without loss of generality as the model is equally applicable to log returns. In estimation, I work with daily returns. Because of the small scale of daily returns, the approximation $\ln(1 + x) \approx x$ is accurate to a high order and the distinction between arithmetic and log returns is negligible.

11I assume for notational simplicity that these are the first $K_t$ elements of $R_t$. This is immaterial since, in the treatment here, elements of $R_t$ are exchangeable.
exceeding \( u \) and given information \( \mathcal{F}_t \), obeys the probability distribution\(^{12}\):

\[
F_{u,i,t}(r) = P(R_{i,t+1} > r \mid R_{i,t} > u, \mathcal{F}_t) = \left( \frac{r}{u} \right)^{-a_i \zeta_{t+1}}
\]

with corresponding density

\[
f_{u,i,t}(r) = \frac{a_i \zeta_{t+1}}{u} \left( \frac{r}{u} \right)^{-(1+a_i \zeta_{t+1})}.
\]

The common element of exponent processes, \( \zeta_{t+1} \), evolves according to\(^{13}\):

\[
\frac{1}{\zeta_{t+1}} = \pi_0 + \pi_1 \frac{1}{\zeta_t^{\text{upd}}} + \pi_2 \frac{1}{\zeta_t}
\]

and the observable update of \( \zeta_{t+1} \) is

\[
\frac{1}{\zeta_t^{\text{upd}}} = \frac{1}{K_t} \sum_{k=1}^{K_t} \ln \frac{R_{k,t}}{u}.
\]

The evolution of \( \zeta_t \) is designed to capture autoregressive time series behavior with a parsimonious parameterization. The update term \( 1/\zeta_t^{\text{upd}} \) is a summary statistic calculated from the cross section of tail observations on date \( t \), which I discuss in more detail below. Recursively substituting for \( \zeta_t \) shows that

\[
\frac{1}{\zeta_{t+1}} = \frac{\pi_0}{1 - \pi_2} + \pi_1 \sum_{j=0}^{\infty} \pi_2^j \frac{1}{\zeta_{t-j}^{\text{upd}}},
\]

Thus, \( 1/\zeta_{t+1} \) is simply an exponentially-weighted moving average of daily updates based on observed extreme returns. When the \( \pi \) coefficients are estimated with maximum likelihood,
this moving average is an optimal forecast of future tail risk.

The role of the update is to summarize information about prevailing tail risk from recent extreme return observations. This conditioning information enters the evolution of $1/\zeta_{t+1}$ via the summary statistic to refresh the conditional tail measure. I calculate a summary of tail risk from the cross section each period using Hill’s (1975) estimator. The Hill estimator is a maximum likelihood estimator of the cross-sectional tail distribution. It takes the form

$$\frac{1}{\zeta^\text{Hill}_t} = \frac{1}{K_t} \sum_{k=1}^{K_t} \ln \frac{R_{k,t}}{u}. \tag{1.7}$$

To see why this makes sense as an update, note that when $u$-exceedances (i.e., $R_{i,t}/u$) obey a power law with exponent $-a_i\zeta_t$, the log exceedence is exponentially distributed with scale parameter $a_i\zeta_t$. By the properties of an exponential random variable, $E_{t-1}[\ln(R_{i,t}/u)] = 1/(a_i\zeta_t)$. As a consequence, the expected value of update $1/\zeta^\text{upd}_t$ is the cross-sectional harmonic average tail exponent,

$$E_{t-1}\left[\frac{1}{K_t} \sum_{k=1}^{K_t} \ln \frac{R_{k,t}}{u}\right] = \frac{1}{\bar{a} \zeta_t}, \quad \text{where} \quad \frac{1}{\bar{a}} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i}. \tag{1.8}$$

The left hand side is an average over the entire cross section due to the fact that the identities of the $K_{t+1}$ exceedences are unknown at time $t$.\footnote{While the identities of the exceedences are unknown, the number of exceedences is known since the tail is defined by a fixed fraction of the cross section size.} This important property will be used to establish consistency and asymptotic normality of the dynamic power law estimation procedure that follows.

Before proceeding to the estimation approach, note that Equation (1.7) is a stochastic process because $\zeta^\text{upd}_{t+1}$ is a function of time $t$ returns. However, $\zeta_{t+1}$ is deterministic conditional upon time-$t$ information. This is different than the specification in the structural models of Section 2, which imply that $\zeta_{t+1}$ is subject to a $t+1$ shock. I argue that this discrepancy
is largely innocuous. In the limit of small time intervals, tail risk processes in the structural models and the exponent process in the econometric model can be specified to line up exactly. A conditionally deterministic tail exponent process, then, can be thought of as a discrete time approximation to a continuous time stochastic process. The advantage of the approximation is that straightforward likelihood maximization procedures can be used for estimation. This property is the tail analogue to the relation between GARCH models (in which volatility is conditionally deterministic) and stochastic volatility models. Nelson (1990) shows that a discrete GARCH(1,1) return process converges to a stochastic volatility process as the time interval shrinks to zero. An important result of Drost and Werker (1996) proves that estimates of a GARCH model at any discrete frequency completely characterize the parameters of its continuous time stochastic volatility equivalent. The same notion lies behind treating the process in (1.7) as a discrete time approximation to a continuous time stochastic process for the tail exponent.\footnote{I conduct a Monte Carlo experiment to examine the tail analogue of the Drost and Werker result. I find that the dynamic power law estimator based on the model in Assumption \ref{assumption:dynamic_power_law} continues to provide accurate estimates of the tail exponent process when the exponent follows a Gaussian autoregression. I discuss this more at the end of the section, and provide a detailed description of the experiment and its results in Appendix \ref{appendix:monte_carlo}.}

1.3.2 Estimating the Dynamic Power Law Model

My estimation strategy uses a quasi-likelihood technique, and is an example of a widely used econometric method with early examples dating at least back to Neyman and Scott (1948), Berk (1966), and the in-depth development of White (1982). The general idea is to use a partial or even mis-specified likelihood to consistently estimate an otherwise intractable model. The proofs that I present can also be thought of as a special case of Hansen’s (1982) GMM theory. To avoid the nuisance parameter problem, I treat assets as though they are independent with identical tail distributions each period. The independence assumption avoids the need to estimate factor loadings for each stock, and the identical assumption avoids
having to estimate each $a_i$ coefficient. These simplifications, however, alter the likelihood from the true likelihood associated with Assumption 1 to a “quasi”-likelihood, written below. I show that maximizing the quasi-likelihood produces consistent and asymptotically normal estimates for the parameters that govern tail dynamics, $\pi_1$ and $\pi_2$. Ultimately, the estimated $\zeta_t$ series is shown to be the fitted cross-sectional harmonic average tail exponent. Since the average exponent series differs from $\zeta_t$ only by the multiplicative factor $\bar{a}$, the two are perfectly correlated.

Before stating the main proposition I discuss two important objects, the log quasi-likelihood and the “score” function (the derivative of the log quasi-likelihood with respect to model parameters). I refer to the tail model in Assumption 1 as the “true” model. Suppose, counterfactually, that all returns in the cross section share the same exponent, which is equal to the cross-sectional harmonic average exponent. Then the tail distribution of all assets becomes

$$\tilde{F}_{u,i,t}(R_{t+1}) = \left(\frac{R_{t+1}}{u}\right)^{-\bar{a}\zeta_t+1}$$

with corresponding density

$$\tilde{f}_{u,i,t}(R_{t+1}) = \frac{\bar{a}\zeta_t+1}{u} \left(\frac{R_{t+1}}{u}\right)^{-1-\bar{a}\zeta_t+1}.$$  

Tildes signify that this distribution is different than the true marginal, $F_{u,i,t}$. Under cross-sectional independence, the corresponding (scaled) log quasi-likelihood is

$$\mathcal{L}(\{R_t\}_{t=1}^T; \pi) = \frac{1}{T} \sum_{t=0}^{T-1} \ln \tilde{f}(R_{t+1}; \pi, F_t) = \frac{1}{T} \sum_{t=0}^{T-1} \sum_{k=1}^{K_{i+1}} \left( \ln \frac{\bar{a}\zeta_t+1}{u} - (1 + \bar{a}\zeta_t+1) \ln \frac{R_{k,t+1}}{u} \right), \quad (1.9)$$

where $u$-exceedences are included in the likelihood and non-exceedences are discarded. Define
the gradient of \( \ln f_t(R_{t+1}; \pi) \) with respect to \( \pi \) (the time-\( t \) element of the score function) as

\[
s_t(R_{t+1}; \pi) \equiv \nabla_\pi \ln f_t(R_{t+1}; \pi) = \frac{\partial \ln f_t(R_{t+1}; \pi)}{\partial \xi_{t+1}} \nabla_\pi \xi_{t+1} = \left( \frac{K_{t+1}}{\xi_{t+1}} - \bar{a} \sum_{k=1}^{K_{t+1}} \ln \frac{R_{k,t+1}}{u} \right) \nabla_\pi \xi_{t+1}. \tag{1.10}
\]

With these expressions in place, I present my central econometric result.

**Proposition 7.** Let the true data generating process of \( \{R_t\}_{t=1}^T \) satisfy the dynamic power law model of Assumption 1 with parameter vector \( \pi^* \). Define the quasi-likelihood estimator \( \hat{\pi}_{QL} \) as

\[
\hat{\pi}_{QL} = \arg \max_{\pi \in \Pi} \mathcal{L}(\{R_t\}_{t=1}^T; \pi).
\]

If the following conditions are satisfied

i. \( \pi^* \) is interior to the parameter space \( \Pi \) over which maximization occurs;

ii. for \( \pi \neq \pi^* \), \( E[s_t(R_{t+1}; \pi)] \neq 0 \), and

iii. \( E[\sup_{\pi \in \Pi} \|s_t(R_{t+1}; \pi)\|] < \infty \),

Then \( \hat{\pi}_{QL} \xrightarrow{p} \pi^* \).

Furthermore, if

iv. \( E[\sup_{\pi \in \Pi} \|\nabla_\pi s_t(R_{t+1}; \pi)\|] < \infty \),

v. \( \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} s_t(R_{t+1}; \pi^*) \xrightarrow{d} N(0, G) \) and

vi. \( E[\nabla_\pi s_t(R_{t+1}; \pi^*)] \) is full column rank,

Then \( \sqrt{T}(\hat{\pi}_{QL} - \pi^*) \xrightarrow{d} N(0, \Psi) \), where \( \Psi = S^{-1}GS^{-1} \), \( S = E[\nabla_\pi s_t(R_{t+1}; \pi^*)] \), and \( G = E[s_t(R_{t+1}; \pi^*)s_t(R_{t+1}; \pi^*)'] \).

**Proof.** The proof follows Newey and McFadden (1994). Before proceeding, I establish a key lemma upon which the remainder of the proposition relies. It shows that \( s_t(R_{t+1}; \pi) \) (which
is based on the mis-specified model \( \tilde{F}_{u,t} \) has expectation equal to zero given that the true data generating process satisfies Assumption [1]

**Lemma 1.** Under Assumption [1] \( E[s_t(R_{t+1}; \pi)] = 0 \).

By the law of iterated expectations,

\[
E[s_t(R_{t+1}; \pi)] = E[ E_{t-1}[s_t(R_{t+1}; \pi)] ]
\]

\[
= E \left[ E_t \left( \frac{K_{t+1}}{\zeta_{t+1}} - \bar{a} \sum_{k=1}^{K_{t+1}} \ln \frac{R_{k,t+1}}{u} \right) \nabla_{\pi} \zeta_{t+1} \right]
\]

\[
= E \left[ \left( \frac{K_{t+1}}{\zeta_{t+1}} - \frac{K_{t+1}}{\zeta_{t+1}} \right) \nabla_{\pi} \zeta_{t+1} \right]
\]

\[
= 0.
\]

The second equality follows from expression [10] and the t-measurability of \( \zeta_{t+1} \). The third equality follows from Equation [8] proving the lemma.

Observe that the first order condition for maximization of [9] is \( \frac{1}{T} \sum_{t=0}^{T-1} s_t(R_{t+1}; \pi) = 0 \). That is, maximization of the log quasi-likelihood produces a valid moment condition upon which estimation may be based. With this insight in hand, the approach of Newey and McFadden may be employed to establish asymptotic properties of the dynamic power law quasi-likelihood estimator. By condition (i) and the fact that the true generating process satisfies Assumption [1], Lemma [1] shows that the moment condition arising from maximization of \( L(\{R_t\}_{t=1}^T; \pi) \) is satisfied. Adding condition (ii), \( \pi^* \) is uniquely identified. By the dominance condition (iii), the uniform law of large numbers may be invoked to establish convergence in probability.

To establish convergence to normality, I use a mean value expansion of the moment condition sample analogue around \( \bar{\pi} \) (a value between \( \hat{\pi} \) and \( \pi^* \)), which gives

\[
\frac{1}{\sqrt{T}}(\hat{\pi} - \pi) = - \left[ \frac{1}{T} \sum_t \nabla_{\pi} s_t(R_{t+1}; \bar{\pi}) \right]^{-1} \frac{1}{\sqrt{T}} s_t(R_{t+1}; \pi^*). 
\]
This expansion is performed noting that $\pi^*$ is interior to the parameter space and, by the functional form of the quasi-likelihood, $s_t(R_{t+1}; \pi)$ is continuously differentiable over $\Pi$. Because $\bar{\pi}$ is between $\hat{\pi}$ and $\pi^*$, $\bar{\pi}$ is also consistent for $\pi^*$ by the convergence in probability result just shown. Using this fact together with condition (iv) delivers $\frac{1}{T} \sum_t \nabla_{\pi} s_t(R_{t+1}; \pi) \xrightarrow{p} S$. By condition (v), $\frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} s_t(R_{t+1}; \pi^*) \xrightarrow{d} N(0, G)$. This, together with Slutzky’s theorem and assumption (vi), proves the result. 

\[ \square \]

1.3.3 Volatility and Heterogeneous Exceedence Probabilities

Implicit in the formulation above is that each element of the vector $R_t$ has an equal probability of exceeding threshold $u$. However, heterogeneity in individual stock volatilities affects the likelihood that a particular stock will experience an exceedence. Let $X$ be a power law variable such that $P(X > u) = bu^{-\zeta}$. The $u$-exceedence distribution of $X$ is $P(X > x|X > u) = (\frac{x}{u})^{-\zeta}$. Now consider a volatility rescaled version of this variable, $Y = \sigma X$. The exceedence probability of $Y$ equals $b (\frac{u}{\sigma})^{-\zeta}$, different than that of $X$. When $\sigma > 1$, $Y$ has a higher exceedence probability than $X$. However, the shape of $Y$’s $u$-exceedence distribution is identical to that of $X$.

A reformulation of the estimator to allow for heterogeneous volatilities is easily established. Let each stock have unique $u$-exceedence probability $p_i$, and consider the effect of this heterogeneity on the expectation of the tail exponent update. In this case, the expectation is no longer the harmonic average tail exponent, but is instead the exceedence probability-weighted average exponent,

$$ E_t \left[ \frac{1}{K_{t+1}} \sum_{k=1}^{K_{t+1}} \ln \frac{R_{k,t+1}}{u} \right] = \frac{1}{\zeta_{t+1}} \sum_{i=1}^{n} \omega_i a_i, $$

where $\omega_i = p_i / \sum_j p_j$. The entire estimation approach and consistency argument outlined above proceeds identically after establishing this point. The ultimate result is that the fitted
ζ_t series is no longer an estimate of the equal-weighted average exponent, but takes on a volatility-weighted character due to the effect that volatility has on the probability of tail occurrences.

Another potential concern is contamination of tail estimates due to time-variation in volatility. I address this by allowing the threshold u to vary over time. My procedure selects u as a fixed q\% quantile,

$$
\hat{u}_t(q) = \inf \left\{ R_{(i),t} \in R_t : \frac{q}{100} \leq \frac{(i)}{n} \right\}
$$

where (i) denotes the i^{th} order statistic of (n x 1) vector $R_t$. In this case, u expands and contracts with volatility so that a fixed fraction of the most extreme observations are used for estimation each period, nullifying the effect of volatility dynamics on tail estimates. My estimates are based on $q = 5$ (or 95 for the upper tail).

### 1.3.4 Monte Carlo Evidence

Appendix A.2 describes a series of Monte Carlo experiments designed to assess finite sample properties of the dynamic power law estimator. Table 11 shows results confirming that the asymptotic properties derived above serve as accurate approximations in finite samples. They also demonstrate the estimator’s robustness to dependence among tail observations and volatility heterogeneity across stocks, both of which are suggested by the structural models. Table 12 explores the estimator’s performance when the true tail exponent is conditionally

---

Threshold choice can have important effects on results. An inappropriately low threshold will contaminate tail exponent estimates by using data from the center of the distribution, whose behavior can vary markedly from tail data. A very high threshold can result in noisy estimates resulting from too few data points. While sophisticated methods for threshold selection have been developed (Dupuis 1999; Matthys and Beirlant 2000; among others), these often require estimation of additional parameters. In light of this, Gabaix et al. (2006) advocate a simple rule fixing the $u$-exceedence probability at 5% for unconditional power law estimation. I follow these authors by applying a similar simple rule in the dynamic setting. Unreported simulations suggest that $q = 1$ to 5 (or 95 to 99 for the upper tail) is an effective quantile choice in my dynamic setting.
stochastic. Even though the estimator presented here relies on a conditionally deterministic exponent process, its estimates achieve over 80% correlation with the true tail series on average.

1.4 Empirical Results

1.4.1 Tail Risk Estimates

Estimates for the dynamic power law model use daily CRSP data from August 1962 to December 2008 for NYSE/AMEX/NASDAQ stocks with share codes 10 and 11. Accuracy of extreme value estimators typically requires very large data sets because only a small fraction of data is informative about the tail distribution. Since the dynamic power law estimator relies on the cross section of returns, I require a large panel of stocks in order to gather sufficient information about the tail at each point in time. Figure 1 plots the number of stocks in CRSP each month. The sample begins with just under 500 stocks in 1926, and has fewer than 1,000 stocks for the next 25 years. In July 1962, the sample size roughly doubles to almost 2,000 stocks with the addition of AMEX. In December 1972, NASDAQ stocks enter the sample and the stock count leaps above 5,000, fluctuating around this size through 2008.\(^{17}\) The dramatic cross-sectional expansion of CRSP beginning in August 1962 leads to my focus on the 1963 to 2008 sample.

Other data used in my analysis are daily Fama-French return factors, monthly risk free rates and size/value-sorted portfolio returns from Ken French’s Data Library\(^{18}\) market return predictor variables from Ivo Welch’s website\(^{19}\) variance risk premium estimates from Hao Zhou’s website\(^{20}\) and macroeconomic data from the Federal Reserve.

\(^{17}\) The dynamic power law estimator in Section 1.3 accommodates changes in size of the cross section over time, highlighting another attractive feature of the estimator.

\(^{18}\) URL: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

\(^{19}\) URL: http://welch.econ.brown.edu/.

\(^{20}\) URL: http://sites.google.com/site/haozhouspersonalhomepage/.
I focus my empirical analysis on the tails of raw returns. In each test I consider tail risk estimated using data from the lower tail only, from the upper tail only, and from combining lower and upper tail data. I refer to the latter case as “both” tails, which in the tests that follow should not be understood as simultaneously including separate estimates of the lower and upper tail in regressions.

For robustness, I also explore how results change when residuals from the Fama-French three-factor model are used to estimate tail risk. Factor model residuals offer a means of mitigating the effects of dependence on the estimator’s efficiency. Threshold \( u_t \) is chosen to be the 5% cross-sectional quantile each period. Standard errors are estimated based on the form of the asymptotic covariance matrix \( \Psi \) derived in Proposition 7. In particular, \( S \) is the log quasi-likelihood Hessian evaluated at \( \hat{\pi} \) parameters and \( G \) is the sample average log quasi-likelihood gradient outer product evaluated at \( \hat{\pi} \). Further details on standard error estimation via likelihood Hessians and gradient outer products may be found in Hayashi (2000).

Table 1 reports estimates for the dynamic power law evolution (1.7). The lower tail exponent of raw returns \( \zeta_t \) varies around a mean of 2.20, with \( \hat{\pi}_1 = 0.072 \) and \( \hat{\pi}_2 = 0.923 \). The \( p \)-values for all statistics reported in Table 2 are below 0.001, rejecting the null hypothesis of constant tail risk and supporting Testable Implication 1. The fact that \( \hat{\pi}_1 + \hat{\pi}_2 > 0.99 \) implies that the tail exponent is highly persistent. The upper tail is slightly fatter (consistent with the results of Jansen and de Vries 1991) and less persistent, though with more time series variability (given the higher value of \( \hat{\pi}_1 \)). When stock returns are converted to factor model residuals, estimation results are qualitatively unchanged.

I plot the fitted tail series \((-\zeta_t)\) based on raw returns in Figure 2. The estimates from

\[\text{\textsuperscript{21}}\text{As my asymptotic theory results and Monte Carlo evidence show, abstracting from dependence does not affect the estimator’s consistency. It may, however, affect the variance of estimates. The asymptotic covariance derived in Proposition \( \text{\textsuperscript{7}} \) accounts for this decreased efficiency, hence test statistics maintain appropriate size.}\]

27
both tails together are shown in Figure 2a, and separate estimates for the lower and upper tails are shown in Figure 2b. These are plotted alongside the log price-dividend ratio for the aggregate market. The estimated tail risk series appears moderately countercyclical, sharing a monthly correlation of -14%, -15% and -14% with the log price-dividend ratio based on estimates from both tails, the lower tail and the upper tail, respectively. The beginning of the sample sees high tail risk, immediately following a drop in the value-weighted index of 28% in the first half of 1962 (the first major market decline during the post-war period). Tail risk declines steadily until December of 1968, when it reaches its lowest levels in the sample. This corresponds to a late 1960’s bull market peak, the level of which is not reached again until the mid-1970’s. Tail risk rises throughout the 1970’s, accelerating its ascent during the oil crisis. It remains around its mean value of -2 to -2.5 for most of the remaining sample. Tail risk recedes in the four bull market years leading up to the 1987 crash, rising quickly in the following months. During the technology boom tail risk retreats sharply but briefly, spiking to its highest post-2000 level amid the early 2003 market trough. At this time the value-weighted index was down 49% from its 2000 high and NASDAQ was 78% of its peak. Finally, during the late 2008 financial crisis, tail risk sees a modest climb after falling in the first half of 2008. Figure 3 plots the tail series estimated from Fama-French three-factor model residuals. The broad dynamic patterns are the same as those in Figure 2.

Figure 4 shows the threshold series \( u_t \) for raw returns alongside monthly realized volatility of the CRSP value-weighted index. The threshold based on both tails has a 60% correlation with volatility. As the figure and correlation show, the threshold appears to successfully absorb volatility changes.

In Table 2, I report monthly correlations between tail risk estimates and macroeconomic variables. Lower (upper) tail estimates have correlation with unemployment of 53% (39%) and -10% (-7%) with the Chicago Fed National Activity Index (CFNAI), once again suggesting some cyclicality in tail risk.
As a brief exploration into the determinants of tail risk, I estimate a regression of the tail process on its own lag and lags of a collection of macroeconomic variables, including realized equity volatility, unemployment, inflation, growth in industrial production, CFNAI and the aggregate stock market return. I show the estimated impacts of these variables on future tail risk in Table 3. Coefficients have been scaled to be interpreted as the response of tail risk ($-\zeta_t$, in number of standard deviations) to a one standard deviation increase in the dependent variable. The most important variable for determining tail risk is the past market return. A return one standard deviation above its mean predicts that the lower tail becomes thinner by 0.175 standard deviations (Newey-West $t=7.1$), while a high past return increases the weight of the upper tail by 0.227 standard deviations ($t=4.6$). Other potentially important determinants of tail risk are past unemployment and equity volatility.

1.4.2 Predicting Stock Market Returns

Testable Implication 2 predicts that the estimated tail risk process ($-\zeta_t$) forecasts risk premia over short and long horizons. I investigate this hypothesis with a series of predictive regressions.

The dependent variable is the excess return on the CRSP value-weighted index at frequencies of one month, one year, three years and five years. The regressor of interest is the tail exponent. Since my estimator is available daily, forecasts are based on $-\zeta_t$ on the last day prior to the forecast horizon.

To avoid look ahead bias, I do not use the tail series estimated in Section 1.4.1 for forecasting since its construction relies on parameters estimated from the full sample. Instead, I use an exponentially-weighted moving average of the period-by-period update $\zeta_{\text{upd}}$. The weighting parameter is fixed ex ante at 0.94, the value used by RiskMetrics for its “exponential smoother”. As pointed out in Section 3, the dynamic power law estimator is simply an optimal exponentially-weighted moving average of updates. Thus, the RiskMetrics exponential smoother applied to the daily cross-sectional Hill estimate is a special case of my model. The RiskMetrics smoothing parameter is analogous to $\pi_2 = 0.94$ in the tail exponent evolution 1.7. For comparison, my maximum likelihood estimates of $\pi_2$ are 0.923, 0.683 and 0.798 for the lower tail, upper tail and both tails, respectively. My predictive results are quite insensitive to the choice of the weighting parameter. Using values ranging from 0.70 to 0.98, including my estimated values, produces only small differences in results, in many cases improving upon the results reported here.
For comparison, I consider two additional predictor variables motivated by my theory model. (Then, in robustness checks, I consider a much larger set of alternative predictors.) The first is monthly realized volatility of the aggregate stock market. Along with the tail risk process, my modified long run risks model suggests that conditional volatility drives expected excess returns. Second, I consider the log dividend-price ratio. The Campbell-Shiller identity implies that variables driving risk premia necessarily enter into the log dividend-price ratio, hence it too should forecast future excess returns as long as one of the state variables does.  

All regressions are conducted at the monthly frequency, meaning that observations are overlapping for the one, three and five year analyses. Richardson and Smith (1991), Hodrick (1992) and Boudoukh and Richardson (1993) (among others) have noted the inferential problems concomitant with overlapping horizon predictive regressions. Overlapping return observations induce a moving average structure in prediction errors, distorting the size of tests based on OLS, and even Newey-West, standard errors. Ang and Bekaert (2007) demonstrate in a Monte Carlo study that the standard error correction of Hodrick (1992) provides the most conservative test statistics relative to other commonly employed standard error estimators, and maintains appropriate test size over horizons as long as five years. I therefore calculate all statistics with Hodrick standard errors.

Predictive regressions based on tail estimates from raw returns are reported in Table 4. To illustrate economic magnitudes, coefficients are scaled to be interpreted as the effect of a

23While true in theory, it may be the case in practice that some state variables have forecasting power while valuation ratios do not. A failure to find return predictability by valuation ratios despite predictability by state variables may arise from poor measurement of aggregate fundamentals such as dividends or other payouts, while purely price-based measures of state variables like tail risk may be more accurate.

24A second statistical consideration is high persistence in the daily tail exponent series. While the daily series appears nearly integrated, regressions are run at the monthly frequency, and ζ-based forecasts only use its value on the last day of each month. Persistence in the month-end series is substantially weaker, with an AR(1) coefficient of 0.890. While this value is mild compared to a monthly AR(1) coefficient of 0.996 for the log dividend-price ratio, it nonetheless calls for cautious inference. I find that Stambaugh bias has little effect on my results. The lower tail risk regression using a one month horizon is a representative case. Stambaugh bias accounts for no more than 3% of the predictive coefficient magnitude in that regression, thus the estimate remains statistically significant at the 1% level.
one standard deviation increase in the regressor on future annualized returns. The lower tail risk series has large, significant forecasting power over all horizons. A one standard deviation increase in lower tail risk predicts an increase in future excess returns of 6.7%, 4.4%, 4.5% and 5.0% per annum, based on data for one month, one year, three year and five year horizons, respectively. The corresponding \( t \)-statistics are 2.9, 2.1, 2.2 and 2.3. The effect of lower tail risk retains the same economic and statistical magnitudes when included alongside realized volatility and the log dividend-price ratio. Similarly, a one standard deviation increase in risk measured from both tails forecasts an increase in excess market returns of between 3.7% and 5.6%. Upper tail risk carries the same sign premium as lower tail risk, which is consistent with the long run risks model of Section 2. The forecasting power of the upper tail is weaker than the lower tail, though its predictive power is impressive at long horizons.

I also evaluate the out-of-sample predictive ability of tail risk. Using data only through month \( t \) (beginning with \( t = 60 \) to provide a sufficiently large initial estimation period), I run univariate predictive regressions of market returns on lower tail risk. This coefficient is used to forecast the \( t + 1 \) return. Because the coefficient is based only on data through \( t \), this procedure mimics the information set an investor would work with in real-time. I plot the resulting sequence of estimated predictive coefficients in Figure 5. Estimates are remarkably stable and are significant at the 95% level in most months, despite having few observations in the early part of the analysis. Moreover, the out-of-sample \( R^2 \) is 1.30%, only slightly lower than the in-sample value.

In Table 5, I explore robustness of these results to estimating tail risk from Fama-French three-factor model residuals rather than raw returns. This mitigates dependence among tail observations, which can potentially improve the finite sample properties of the estimator. The analysis proceeds as in Table 4, and conclusions are unchanged. Lower tail risk demonstrates the strongest ability to forecast market returns, and upper tail risk predicts returns in the same direction as lower tail risk. Residual tails provide even more
explanatory power than raw returns over one, three and five year horizons, and upper tail risk becomes a statistically successful predictor at the one month horizon.

In a second robustness check, I compare the forecasting power of my tail risk measure to a large set of alternative forecasting variables that have been shown to successfully predict returns. Goyal and Welch (2008) conduct a comprehensive analysis of forecasting performance using an extensive set of predictors from the literature. Table 6 reports results of univariate forecasting regressions using fifteen Goyal and Welch variables that are available at the monthly frequency over the 1963-2007 sample. No variable forecasts returns as strongly or consistently over all horizons as the lower tail risk process. Inflation strongly predicts one month returns, but its effect dies out at longer horizons. After tail risk, the most successful long horizon univariate predictors are the cross-sectional premium (Polk, Thompson and Vuolteenaho 2006, data available through 2002), the term spread, and the long term yield (Campbell 1987; Fama and French 1989). I also run bivariate regressions using lower tail risk alongside each Goyal and Welch variable from Table 6. Table 7 presents these results. The predictive ability of tail risk is broadly unaffected by including alternative regressors. In addition to the Goyal and Welch predictors, I compare the performance of my tail risk measure to the variance risk premium (VRP). Bollerslev, Tauchen and Zhou (2009) demonstrate the outstanding predictive power of VRP, which is the difference between risk neutral variance extracted from options data and realized variance calculated from ultra-high frequency data (see the last row of Table 6 for univariate VRP regressions). These authors, as well as Drechsler and Yaron (2009), demonstrate how VRP is fundamentally linked to

25In one of the fifteen cases, coefficient estimates suggest a potential multicollinearity problem: This is when tail risk and the cross-sectional premium are included together. The monthly correlation between these two variables is 64%. Multicollinearity presents an inferential problem for individual regressors, but not for the total explanatory power of the model. The monthly predictive $R^2$ for the two variables together is 7.0%, which is an extremely high degree of predictability (see Campbell and Thompson 2008). At the five year horizon, the cross section premium effect dies out, but the tail risk effect remains large and significant (scaled coefficient of 4.2% with $t$-statistic 2.0).
tail risk in the real economy and thus naturally related to my measure. The shortcoming of VRP, however, is that it is only available starting in 1990 due to intra-daily data limitations. I compare the predictive power of my tail risk estimate and VRP in bivariate predictive regressions over the 1990-2008 subsample. Results are reported in the last row of Table 7. The monthly $R^2$ increases from 3.5% and 2.1% when VRP and tail risk, respectively, are included alone, to 5.2% when they are included together. At the annual horizon, the univariate VRP and tail risk $R^2$ increases from 3.8% and 18.5%, respectively, to 22.8% when they are included together. The coefficient on $-\zeta_t$ is slightly larger than its estimate from the full sample, and is stable over all horizons. These results suggest that my tail risk measure and VRP are complementary descriptors of tail risk. In summary, I conclude that increases in my tail risk measure significantly predict increases in excess returns on the market over short and long horizons, consistent with Testable Implication 2.

1.4.3 Tail Risk and the Cross Section of Expected Returns

In Section 2, I show that the heavy-tailed long run risks model and time-varying rare disaster model predict a negative price of tail risk (Testable Implication 3). As a first test of this hypothesis, I examine average returns of portfolios formed on the basis of stocks’ exposure to my estimated tail risk process. I proceed as follows. In each month $t + 1$, I estimate tail risk betas by regressing monthly excess returns over the 60 months ending at $t$ on contemporaneous innovations in the month-end tail risk ($-\zeta_t$, estimated from both tails). I then sort stocks into tail risk beta quintile portfolios, including all NYSE/AMEX/NASDAQ stocks with share codes 10 and 11 and at least 36 months of non-missing returns out of the previous 60 months. Table 8, Panel A reports time series average post-formation portfolio returns for tail risk beta-sorted portfolios. Stocks that covary least with tail risk (quintile 1) earn average annualized returns of 6.40%, while stocks with the highest tail risk betas (quintile 5) earn only 0.36% per annum. The difference in average annualized returns between
quintile 5 and quintile 1 is -6.03% with $t$-statistic 2.6, supporting the hypothesis of a negative tail risk premium.

To check whether this result is driven by a particular subset of stocks, I construct double-sorted portfolios. In Panel B of Table 8, I first sort stocks into quintiles by market beta (estimated in the same rolling manner as tail risk beta). Then, within market beta quintiles, I sort again by tail risk beta. The negative tail risk premium appears in each quintile, with a difference in average returns across high and low tail risk beta quintiles ranging from -3.06% to -4.43% per annum. Similarly, I find a negative tail risk premium across size quintiles (Panel C) and book-to-market quintiles (Panel D), with the exception of the very smallest stocks and very highest book-to-market stocks.

To test the negative tail risk price hypothesis while controlling for multiple alternative explanatory characteristics, I run a series of Fama-MacBeth regressions. Table 9 shows results from second stage regressions using NYSE/AMEX/NASDAQ stocks as test assets. I scale coefficients so that they may be interpreted as the change in annual expected excess returns for a one (cross-sectional) standard deviation change in risk factor exposure. When regressions include only tail risk, the risk price estimate is -6.2% ($t$=2.9) based on the lower tail and -5.3% ($t$=2.5) using both tails. After controlling for realized equity market volatility (as prescribed by the long run risks model) and Fama-French factors, lower tail and both tail risk price estimates are -5.6% and -4.4%, respectively ($t$=2.9 and 2.4).

Repeating the Fama-MacBeth analysis with tail risk estimated from Fama-French factor model residuals (Panel B of Table 9) leads to the same conclusions. When all four additional explanatory variables are included, lower tail and both tail risk prices are estimated at -3.9% and -3.0% ($t$=2.2 and 1.9). Finally, I repeat the analysis using three alternative sets of test assets. First, to determine if results are being driven by smaller NASDAQ stocks, I consider NYSE stocks separately. Results are presented in Table 10. I find similar tail risk price magnitudes as in the full CRSP cross section (-5.1% and -4.5% for the lower tail and both
tails) with stronger significance ($t=3.5$ and $3.3$). Table 10 also includes test results using 100 size and value-sorted portfolios as test assets (from Ken French’s website) and 25 market beta and tail risk beta-sorted portfolios. In both cases, the estimated price of lower tail risk is significantly negative. More generally, the results of this section support the prediction of Testable Implication 3 that the price of tail risk is negative, hence high tail risk betas stocks earn comparatively low average returns.

1.5 Conclusion

A measure of extreme event risk is essential to understanding the behavior of asset prices. If this risk changes through time, extreme value techniques based on aggregate data will be incapable of providing conditional tail measures. I present a new dynamic tail risk model that overcomes this difficulty. The model uses the cross section of individual stock returns to inform estimates of conditional tail risk at each point in time. Tests show that extremal risk is significantly time varying and persistent. This conclusion holds for raw returns as well as factor model residuals.

Evidence suggests that tail risk has large predictive power for excess aggregate stock market returns over horizons of one month to five years, outperforming all alternative predictors commonly consider in the literature. A one standard deviation increase in tail risk forecasts a 4.4% increase in excess market returns over the following year. Furthermore, individual stock tails have large, significant explanatory power for the cross section of average stock returns. Stocks that covary highly with my estimated tail risk series earn average annual returns 6.0% lower than stocks with low tail risk covariation.

These results can be understood from the perspective of (at least) two structural models: i) a long run risks model modified to include time-varying heavy-tailed shocks and ii) a time-varying rare disaster model. I show that in both of these models the price of tail risk is
negative. Because tail risk is detrimental to investors’ utility, assets that pay off in high tail risk states are valuable hedges and thus earn low average returns, consistent with empirical results in the cross section. In the time series, high tail risk increases the return required by investors to hold the market portfolio, which results in return forecastability, also consistent with my empirical findings.
Bibliography


Table 1. Dynamic Power Law Estimates for the NYSE/AMEX/NASDAQ Panel.
The table reports estimates for the dynamic power law model using the panel of NYSE/AMEX/NASDAQ stocks from August 1962 to December 2008. Estimation follows the quasi-maximum likelihood procedure described in Section 3. The model is estimated separately using data from both tails together, the lower tail alone and the upper tail alone. Results in Panel A are based on raw returns and results in Panel B are based on residuals from the Fama-French three-factor model. Standard errors of coefficients are reported in parentheses, and are calculated from estimates of the quasi-maximum likelihood asymptotic covariance structure derived in Section 3. The parameter $\zeta$ is the implied unconditional mean tail exponent based on estimates of $\pi_0$, $\pi_1$ and $\pi_2$ (the reported standard error estimate for $\zeta$ has been appropriately transformed based on standard errors of $\pi_0$, $\pi_1$ and $\pi_2$).

<table>
<thead>
<tr>
<th></th>
<th>Both Tails</th>
<th>Lower Tail</th>
<th>Upper Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Raw Returns</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\zeta$</td>
<td>2.110</td>
<td>2.201</td>
<td>1.872</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.044)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>$\pi_1$</td>
<td>0.188</td>
<td>0.072</td>
<td>0.239</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.010)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>0.798</td>
<td>0.923</td>
<td>0.683</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.011)</td>
<td>(0.092)</td>
</tr>
<tr>
<td><strong>Panel B: Factor Model Residuals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\zeta$</td>
<td>2.090</td>
<td>2.145</td>
<td>2.055</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.052)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>$\pi_1$</td>
<td>0.211</td>
<td>0.118</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.007)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>0.780</td>
<td>0.879</td>
<td>0.801</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.007)</td>
<td>(0.017)</td>
</tr>
</tbody>
</table>

Table 2. Tail Risk Correlation with Macroeconomic Variables.
The table reports correlations between tail risk estimates and macroeconomic variables. Tail risk ($-\zeta$) is estimated for the upper and lower tail separately and for both tails together by the dynamic power law estimator on raw return data for NYSE/AMEX/NASDAQ stocks. Macroeconomic variables included are the log dividend-price ratio, unemployment rate, inflation rate, growth rate of industrial production, Chicago Fed National Activity Index and the variance risk premium (VIX^2 minus realized S&P 500 variance). Since tail risk is measured daily, correlations are calculated based on month-end values. The sample horizon is 1963 to 2008 (1990-2008 for the variance risk premium).

<table>
<thead>
<tr>
<th></th>
<th>Lower $-\zeta$</th>
<th>Upper $-\zeta$</th>
<th>Both $-\zeta$</th>
<th>ln D/P</th>
<th>Unemp.</th>
<th>Infl.</th>
<th>Growth</th>
<th>CFNAI</th>
<th>VRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower $-\zeta$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper $-\zeta$</td>
<td>0.56</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both $-\zeta$</td>
<td>0.91</td>
<td>0.64</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln D/P</td>
<td>0.15</td>
<td>0.14</td>
<td>0.14</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemp.</td>
<td>0.53</td>
<td>0.39</td>
<td>0.47</td>
<td>0.57</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.05</td>
<td>0.02</td>
<td>-0.07</td>
<td>0.36</td>
<td>0.08</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind. Prod. Gr.</td>
<td>-0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.01</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFNAI</td>
<td>-0.10</td>
<td>-0.07</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.14</td>
<td>0.03</td>
<td>0.46</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>VRP</td>
<td>0.03</td>
<td>0.29</td>
<td>0.14</td>
<td>-0.11</td>
<td>-0.10</td>
<td>0.15</td>
<td>0.06</td>
<td>0.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3. Macroeconomic Determinants of Tail Risk.
The table reports results from predictive regressions for lower and upper tail risk ($-\zeta$) estimated from raw returns on the NYSE/AMEX/NASDAQ cross section. Regressions are run using monthly data. In addition to the lagged dependent variable, regressors include lagged realized equity market volatility (calculated as the realized daily volatility for the CRSP value-weighted index each month), unemployment rate, inflation rate, growth rate of industrial production, Chicago Fed National Activity Index and return on the aggregate market. Test statistics are reported below coefficients in italics and use Newey-West standard errors with twelve lags. The sample horizon is 1963 to 2008.

<table>
<thead>
<tr>
<th></th>
<th>$-\zeta_{t}$ (lower)</th>
<th></th>
<th>$-\zeta_{t}$ (upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. Vol$_{t-1}$</td>
<td>0.027 0.107</td>
<td></td>
<td>0.098 0.009</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td></td>
<td>3.3 0.4</td>
</tr>
<tr>
<td>Unemploy$_{t-1}$</td>
<td>0.041 0.055</td>
<td></td>
<td>0.053 0.083</td>
</tr>
<tr>
<td></td>
<td>3.2 3.4</td>
<td></td>
<td>2.6 3.3</td>
</tr>
<tr>
<td>Inflation$_{t-1}$</td>
<td>0.002 0.016</td>
<td></td>
<td>0.010 0.022</td>
</tr>
<tr>
<td></td>
<td>0.2 1.5</td>
<td></td>
<td>0.5 1.1</td>
</tr>
<tr>
<td>Ind. Pro. Gr.$_{t-1}$</td>
<td>-0.002 -0.017</td>
<td></td>
<td>-0.027 -0.020</td>
</tr>
<tr>
<td></td>
<td>0.1 0.9</td>
<td></td>
<td>0.7 0.7</td>
</tr>
<tr>
<td>CFNAI$_{t-1}$</td>
<td>0.008 -0.017</td>
<td></td>
<td>0.046 0.000</td>
</tr>
<tr>
<td></td>
<td>0.4 1.2</td>
<td></td>
<td>1.0 0.0</td>
</tr>
<tr>
<td>Market Ret$_{t-1}$</td>
<td>-0.175 -0.188</td>
<td></td>
<td>0.227 0.204</td>
</tr>
<tr>
<td></td>
<td>7.1 8.0</td>
<td></td>
<td>4.6 4.6</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.239 0.004 -0.316 0.003 0.003 0.006 0.003</td>
<td>-0.312 0.005 -0.479 0.006 0.005 0.014 0.006</td>
<td>2.4 0.1 2.9 0.1 0.1 0.3 0.2</td>
</tr>
<tr>
<td></td>
<td>2.8 0.2 3.0 0.1 0.1 0.2 0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.842 0.808 0.802 0.797 0.797 0.804 0.829</td>
<td>0.684 0.607 0.619 0.607 0.607 0.628 0.648</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Predicting Excess Aggregate Stock Market Returns (Raw Return Tails).

The table reports predictive regressions of CRSP value-weighted market index excess returns over one month, one year, three year and five year horizons. The first three predictors are the dynamic power law model tail risk process (\(-z_t\)) estimated using both tails, the lower tail and the upper tail of raw returns for NYSE/AMEX/NASDAQ stocks (using RiskMetrics moving average weighting parameters). Since the tail process is estimated daily, forecasts use the value of \(-z_t\) on the last day before the forecast period. Other regressors are the log dividend-price ratio and monthly realized volatility of the CRSP value-weighted market index. Test statistics are calculated using Hodrick (1992) standard errors with lag length equal to the number of months in each horizon. The sample horizon is 1963 to 2008.

<table>
<thead>
<tr>
<th></th>
<th>One month horizon</th>
<th>One year horizon</th>
<th>Three year horizon</th>
<th>Five year horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-z_t) (both)</td>
<td>5.63</td>
<td>4.92</td>
<td>3.72</td>
<td>3.67</td>
</tr>
<tr>
<td>(-z_t) (lower)</td>
<td>6.69</td>
<td>6.14</td>
<td>4.44</td>
<td>4.24</td>
</tr>
<tr>
<td>(-z_t) (upper)</td>
<td>3.11</td>
<td>2.94</td>
<td>3.56</td>
<td>3.15</td>
</tr>
<tr>
<td>ln D/P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs.</td>
<td>565</td>
<td>565</td>
<td>532</td>
<td>532</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.011</td>
<td>0.016</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Three year horizon</th>
<th>Five year horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-z_t) (both)</td>
<td>3.96</td>
<td>4.76</td>
</tr>
<tr>
<td>(-z_t) (lower)</td>
<td>4.52</td>
<td>5.02</td>
</tr>
<tr>
<td>(-z_t) (upper)</td>
<td>3.87</td>
<td>4.73</td>
</tr>
<tr>
<td>ln D/P</td>
<td>2.60</td>
<td>2.31</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.36</td>
<td>6.50</td>
</tr>
<tr>
<td>Obs.</td>
<td>565</td>
<td>532</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.011</td>
<td>0.148</td>
</tr>
</tbody>
</table>
Table 5. Predicting Excess Aggregate Stock Market Returns (Factor Model Residual Tails).
The table reports predictive regressions of CRSP value-weighted market index excess returns over one month, one year, three year and five year horizons. The first three predictors are the dynamic power law model tail risk process ($-\zeta_t$) estimated using both tails, the lower tail and the upper tail of Fama-French three-factor model residuals for NYSE/AMEX/NASDAQ stocks (using RiskMetrics moving average weighting parameters). Since the tail process is estimated daily, forecasts use the value of $-\zeta_t$ on the last day before the forecast period. Other regressors are the log dividend-price ratio and monthly realized volatility of the CRSP value-weighted market index. Test statistics are calculated using Hodrick (1992) standard errors with lag length equal to the number of months in each horizon. The sample horizon is 1963 to 2008.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>One month horizon</th>
<th>One year horizon</th>
<th>Three year horizon</th>
<th>Five year horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-\zeta_t$ (both)</td>
<td>4.23 1.9</td>
<td>4.25 1.7</td>
<td>4.84 2.1</td>
<td>5.45 2.4</td>
</tr>
<tr>
<td>$-\zeta_t$ (lower)</td>
<td>4.29 2.0</td>
<td>4.61 2.0</td>
<td>5.27 2.0</td>
<td>5.65 2.0</td>
</tr>
<tr>
<td>$-\zeta_t$ (upper)</td>
<td>6.74 2.8</td>
<td>6.46 2.7</td>
<td>2.0 2.0</td>
<td>4.04 2.2</td>
</tr>
<tr>
<td>ln D/P</td>
<td>2.60 1.0</td>
<td>2.03 0.8</td>
<td>2.2 1.2</td>
<td>2.3 1.2</td>
</tr>
<tr>
<td>R. Volatility</td>
<td>-3.05 1.2</td>
<td>-3.09 1.2</td>
<td>-3.43 1.3</td>
<td>1.06 0.8</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.56 4.56</td>
<td>4.33 4.33</td>
<td>5.32 5.32</td>
<td>6.30 6.30</td>
</tr>
<tr>
<td>Obs.</td>
<td>565 565</td>
<td>565 565</td>
<td>565 565</td>
<td>565 565</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.006 0.006</td>
<td>0.016 0.012</td>
<td>0.013 0.021</td>
<td>0.016 0.019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Three year horizon</th>
<th>Five year horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-\zeta_t$ (both)</td>
<td>4.67 2.1</td>
<td>5.19 2.4</td>
</tr>
<tr>
<td>$-\zeta_t$ (lower)</td>
<td>5.19 2.2</td>
<td>5.49 2.4</td>
</tr>
<tr>
<td>$-\zeta_t$ (upper)</td>
<td>3.82 1.8</td>
<td>4.47 2.4</td>
</tr>
<tr>
<td>ln D/P</td>
<td>1.93 1.69</td>
<td>1.5 1.6</td>
</tr>
<tr>
<td>R. Volatility</td>
<td>-0.20 -0.05</td>
<td>0.51 0.63</td>
</tr>
<tr>
<td>Intercept</td>
<td>6.27 6.18</td>
<td>7.32 7.27</td>
</tr>
<tr>
<td>Obs.</td>
<td>532 532</td>
<td>508 508</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.155 0.181</td>
<td>0.374 0.412</td>
</tr>
</tbody>
</table>

The table reports predictive regressions of CRSP value-weighted market index excess returns over one month, one year, three year and five year horizons. The first row repeats forecasting results for the lower tail risk series from Table 4. Results in all other rows but the last are from univariate regressions for the 1963-2007 sample using predictors considered in Goyal and Welch (2007) (data from Ivo Welch’s website). The last row reports results using the variance risk premium (Bollerslev, Tauchen and Zhou 2009, data from Hao Zhou’s website). In this case, regressions use data from 1990-2008 due to the unavailability of the variance risk premium prior to 1990. Test statistics are calculated using Hodrick (1992) standard errors with lag length equal to the number of months in each horizon.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>One month horizon</th>
<th>One year horizon</th>
<th>Three year horizon</th>
<th>Five year horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>t-stat</td>
<td>R²</td>
<td>Coeff.</td>
</tr>
<tr>
<td>Tail Risk (-(\zeta) lower)</td>
<td>6.70</td>
<td>2.9</td>
<td>0.016</td>
<td>4.44</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>0.81</td>
<td>0.3</td>
<td>0.000</td>
<td>1.76</td>
</tr>
<tr>
<td>Cross section premium</td>
<td>5.53</td>
<td>2.4</td>
<td>0.010</td>
<td>-4.19</td>
</tr>
<tr>
<td>Default return spread</td>
<td>1.73</td>
<td>0.7</td>
<td>0.001</td>
<td>-0.08</td>
</tr>
<tr>
<td>Default yield spread</td>
<td>4.65</td>
<td>2.0</td>
<td>0.008</td>
<td>2.26</td>
</tr>
<tr>
<td>Dividend return spread</td>
<td>-0.27</td>
<td>0.1</td>
<td>0.000</td>
<td>0.88</td>
</tr>
<tr>
<td>Dividend yield spread</td>
<td>2.55</td>
<td>1.0</td>
<td>0.002</td>
<td>3.19</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>2.65</td>
<td>1.1</td>
<td>0.003</td>
<td>3.18</td>
</tr>
<tr>
<td>Earnings price ratio</td>
<td>2.80</td>
<td>1.1</td>
<td>0.003</td>
<td>2.86</td>
</tr>
<tr>
<td>Inflation</td>
<td>-6.72</td>
<td>2.6</td>
<td>0.016</td>
<td>-1.99</td>
</tr>
<tr>
<td>Long term return</td>
<td>5.35</td>
<td>2.3</td>
<td>0.010</td>
<td>2.21</td>
</tr>
<tr>
<td>Long term yield</td>
<td>-0.70</td>
<td>0.3</td>
<td>0.000</td>
<td>1.50</td>
</tr>
<tr>
<td>Net equity expansion</td>
<td>-4.34</td>
<td>2.1</td>
<td>0.007</td>
<td>-2.11</td>
</tr>
<tr>
<td>Stock volatility</td>
<td>0.15</td>
<td>0.1</td>
<td>0.000</td>
<td>0.44</td>
</tr>
<tr>
<td>Term Spread</td>
<td>4.49</td>
<td>2.0</td>
<td>0.007</td>
<td>3.95</td>
</tr>
<tr>
<td>Treasury bill rate</td>
<td>-3.08</td>
<td>1.3</td>
<td>0.003</td>
<td>-0.81</td>
</tr>
<tr>
<td>Variance risk premium</td>
<td>9.73</td>
<td>3.3</td>
<td>0.035</td>
<td>4.34</td>
</tr>
</tbody>
</table>

The table repeats the analysis of Table 6, but instead uses bivariate regressions that include each alternative predictor alongside the lower tail risk process estimated with the dynamic power law model (using raw returns of NYSE/AMEX/NASDAQ stocks). Test statistics are calculated using Hodrick (1992) standard errors with lag length equal to the number of months in each horizon. For each horizon, the first two columns are the coefficient estimate and t-statistic for the alternative predictor, while the third and fourth columns are the coefficient and t-statistic for the tail risk process.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>One month horizon</th>
<th>One year horizon</th>
<th>Three year horizon</th>
<th>Five year horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.  t-stat</td>
<td>Tail Coeff.  t-stat</td>
<td>Coeff.  t-stat</td>
<td>Tail Coeff.  t-stat</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>0.73  0.3</td>
<td>6.47  2.8</td>
<td>0.015</td>
<td>1.70  0.7</td>
</tr>
<tr>
<td>Cross section premium</td>
<td>16.9  4.7</td>
<td>16.7  4.6</td>
<td>0.070</td>
<td>-2.17  0.9</td>
</tr>
<tr>
<td>Default return spread</td>
<td>1.45  0.6</td>
<td>6.42  2.8</td>
<td>0.016</td>
<td>-0.28  0.5</td>
</tr>
<tr>
<td>Default yield spread</td>
<td>3.36  1.5</td>
<td>5.72  2.5</td>
<td>0.019</td>
<td>1.33  0.5</td>
</tr>
<tr>
<td>Dividend payout ratio</td>
<td>0.49  0.2</td>
<td>6.54  2.8</td>
<td>0.015</td>
<td>1.42  0.7</td>
</tr>
<tr>
<td>Dividend price ratio</td>
<td>1.80  0.7</td>
<td>6.27  2.7</td>
<td>0.016</td>
<td>2.70  1.1</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>1.78  0.7</td>
<td>6.23  2.7</td>
<td>0.016</td>
<td>2.61  1.0</td>
</tr>
<tr>
<td>Earnings price ratio</td>
<td>1.66  0.6</td>
<td>6.18  2.6</td>
<td>0.016</td>
<td>2.11  0.8</td>
</tr>
<tr>
<td>Inflation</td>
<td>-6.31  2.5</td>
<td>6.05  2.6</td>
<td>0.030</td>
<td>-1.70  1.0</td>
</tr>
<tr>
<td>Long term return</td>
<td>4.56  2.0</td>
<td>5.86  2.5</td>
<td>0.023</td>
<td>1.64  2.5</td>
</tr>
<tr>
<td>Long term yield</td>
<td>-3.45  1.3</td>
<td>7.71  3.2</td>
<td>0.019</td>
<td>-0.10  0.0</td>
</tr>
<tr>
<td>Net equity expansion</td>
<td>-2.34  1.1</td>
<td>5.66  2.3</td>
<td>0.017</td>
<td>-0.63  0.3</td>
</tr>
<tr>
<td>Stock volatility</td>
<td>1.02  0.4</td>
<td>6.62  2.9</td>
<td>0.016</td>
<td>1.04  0.6</td>
</tr>
<tr>
<td>Term Spread</td>
<td>2.46  1.0</td>
<td>5.59  2.3</td>
<td>0.017</td>
<td>2.70  1.3</td>
</tr>
<tr>
<td>Treasury bill rate</td>
<td>-3.92  1.7</td>
<td>6.96  3.0</td>
<td>0.021</td>
<td>-1.37  0.5</td>
</tr>
<tr>
<td>Variance risk premium</td>
<td>9.19  2.9</td>
<td>6.73  2.0</td>
<td>0.052</td>
<td>4.63  3.3</td>
</tr>
</tbody>
</table>

48
Table 8. Returns on Tail Risk Beta-Sorted Portfolios.
The table reports equal-weighted monthly returns (in annualized percentages) for portfolios of NYSE/AMEX/NASDAQ stocks from 1963-2008. At the beginning of month \(t+1\), I form portfolios based on tail risk beta, market beta, size, and book-to-market ratio, estimated using data from the 60 months beginning at date \(t-59\) and ending at date \(t\). In Panel A, stocks are sorted into quintile portfolios by their beta on the tail risk measure, estimated by regressing monthly portfolio returns on monthly innovations from an AR(1) model for the tail risk series \(\zeta\) (which is itself estimated in a preliminary step with the dynamic power law model using both tails). In Panels B, C and D, stocks are first sorted into quintiles based on market beta \(\beta_{MKT}\), market equity and book-to-market, respectively. Then, within each quintile, stocks are further sorted into tail risk beta quintiles. Also, within each quintile (or, in the case of Panel A, among all stocks), I calculate the difference in average returns between high tail risk beta stocks and low tail risk beta stocks, as well as the associated \(t\)-statistic. For a stock to be included in a portfolio at \(t+1\), I require that it has at least 36 months of non-missing return data out of the previous 60 months.

<table>
<thead>
<tr>
<th>Tail Risk Beta</th>
<th>Low</th>
<th></th>
<th>High</th>
<th></th>
<th>Diff.</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>(5-1)</td>
</tr>
</tbody>
</table>

**Panel A: Tail Risk Beta Only**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>6.40</td>
<td>7.13</td>
<td>6.23</td>
<td>4.44</td>
<td>0.36</td>
<td>-6.03</td>
</tr>
</tbody>
</table>

**Panel B: Market Beta / Tail Risk Beta**

<table>
<thead>
<tr>
<th></th>
<th>Low (\beta_{MKT})</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.71</td>
<td>7.41</td>
<td>7.11</td>
<td>6.40</td>
<td>3.65</td>
<td>-3.06</td>
</tr>
<tr>
<td>2</td>
<td>5.91</td>
<td>6.19</td>
<td>5.97</td>
<td>4.50</td>
<td>2.27</td>
<td>-3.64</td>
</tr>
<tr>
<td>3</td>
<td>4.32</td>
<td>5.12</td>
<td>4.34</td>
<td>3.25</td>
<td>0.44</td>
<td>-3.88</td>
</tr>
<tr>
<td>4</td>
<td>2.54</td>
<td>3.36</td>
<td>2.53</td>
<td>1.06</td>
<td>-1.01</td>
<td>-3.55</td>
</tr>
<tr>
<td>High (\beta_{MKT})</td>
<td>-0.02</td>
<td>1.78</td>
<td>-0.35</td>
<td>-1.57</td>
<td>-4.45</td>
<td>-4.43</td>
</tr>
</tbody>
</table>

**Panel C: Market Equity / Tail Risk Beta**

<table>
<thead>
<tr>
<th></th>
<th>Small</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.16</td>
<td>9.27</td>
<td>10.15</td>
<td>10.67</td>
<td>13.98</td>
<td>3.82</td>
</tr>
<tr>
<td>2</td>
<td>2.76</td>
<td>4.48</td>
<td>3.46</td>
<td>0.51</td>
<td>-4.33</td>
<td>-7.10</td>
</tr>
<tr>
<td>3</td>
<td>4.81</td>
<td>5.81</td>
<td>4.99</td>
<td>0.86</td>
<td>-5.32</td>
<td>-10.13</td>
</tr>
<tr>
<td>4</td>
<td>6.71</td>
<td>7.72</td>
<td>6.45</td>
<td>4.79</td>
<td>-2.18</td>
<td>-8.89</td>
</tr>
<tr>
<td>Big</td>
<td>5</td>
<td>6.76</td>
<td>6.82</td>
<td>6.80</td>
<td>5.68</td>
<td>1.43</td>
</tr>
</tbody>
</table>

**Panel D: Book-to-Market / Tail Risk Beta**

<table>
<thead>
<tr>
<th></th>
<th>Growth</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.75</td>
<td>6.16</td>
<td>5.23</td>
<td>3.05</td>
<td>-1.88</td>
<td>-7.63</td>
</tr>
<tr>
<td>2</td>
<td>7.50</td>
<td>7.34</td>
<td>7.07</td>
<td>5.23</td>
<td>1.94</td>
<td>-5.56</td>
</tr>
<tr>
<td>3</td>
<td>9.14</td>
<td>8.78</td>
<td>8.11</td>
<td>7.22</td>
<td>4.71</td>
<td>-4.43</td>
</tr>
<tr>
<td>4</td>
<td>10.35</td>
<td>9.93</td>
<td>9.00</td>
<td>8.98</td>
<td>7.11</td>
<td>-3.24</td>
</tr>
<tr>
<td>Value</td>
<td>5</td>
<td>11.09</td>
<td>10.66</td>
<td>11.64</td>
<td>12.01</td>
<td>14.06</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9. Cross-Sectional Return Regressions.
The table reports slope and intercept estimates from second-stage regressions of NYSE/AMEX/NASDAQ stock returns on factor exposures estimated in first-stage regressions. The first three factors are innovations from AR(1) models of tail risk series estimated using the dynamic power law model for both tails, the lower tail and the upper tail of raw returns (Panel A) and of Fama-French three-factor model residuals (Panel B). Other factors included are AR(1) innovations to the monthly realized volatility of the CRSP value-weighted market portfolio, the excess market return and Fama and French's (1993) SMB and HML returns. Coefficients are standardized to represent the change in average annualized percentage excess returns resulting from a one standard deviation increase in the regressor. The t-statistic is reported below each coefficient in italics. The sample horizon is 1963-2008.

**Panel A: Raw Return Tails**

<table>
<thead>
<tr>
<th>ζ (both)</th>
<th>ζ (lower)</th>
<th>ζ (upper)</th>
<th>R. Volatility</th>
<th>R_mkt</th>
<th>SMB</th>
<th>HML</th>
<th>Intercept</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.303</td>
<td>-6.197</td>
<td>-0.112</td>
<td>6.001</td>
<td>2.420</td>
<td>-3.787</td>
<td>-0.940</td>
<td>0.030</td>
<td>0.024</td>
</tr>
<tr>
<td>-5.582</td>
<td>-5.782</td>
<td>-0.005</td>
<td>6.309</td>
<td>2.268</td>
<td>-3.596</td>
<td>-0.900</td>
<td>0.029</td>
<td>0.024</td>
</tr>
<tr>
<td>-4.399</td>
<td>-5.586</td>
<td>-1.617</td>
<td>6.863</td>
<td>0.645</td>
<td>-3.583</td>
<td>0.616</td>
<td>0.020</td>
<td>0.025</td>
</tr>
<tr>
<td>2.5</td>
<td>2.3</td>
<td>2.7</td>
<td>5.209</td>
<td>4.082</td>
<td>2.2</td>
<td>2.2</td>
<td>4.802</td>
<td>0.406</td>
</tr>
<tr>
<td>2.4</td>
<td>2.4</td>
<td>2.9</td>
<td>4.663</td>
<td>4.023</td>
<td>2.2</td>
<td>2.2</td>
<td>4.356</td>
<td>0.392</td>
</tr>
<tr>
<td>2.9</td>
<td>2.7</td>
<td>0.1</td>
<td>4.978</td>
<td>4.561</td>
<td>2.3</td>
<td>2.3</td>
<td>5.035</td>
<td>4.295</td>
</tr>
</tbody>
</table>

**Panel B: Factor Model Residual Tails**

<table>
<thead>
<tr>
<th>ζ (both)</th>
<th>ζ (lower)</th>
<th>ζ (upper)</th>
<th>R. Volatility</th>
<th>R_mkt</th>
<th>SMB</th>
<th>HML</th>
<th>Intercept</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.688</td>
<td>-3.199</td>
<td>-5.339</td>
<td>6.379</td>
<td>0.406</td>
<td>-3.596</td>
<td>-0.923</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>-3.847</td>
<td>-2.605</td>
<td>-4.479</td>
<td>6.801</td>
<td>-1.171</td>
<td>-3.506</td>
<td>0.6</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>-3.016</td>
<td>-3.881</td>
<td>-2.577</td>
<td>7.196</td>
<td>0.616</td>
<td>-3.591</td>
<td>0.4</td>
<td>0.024</td>
<td>0.025</td>
</tr>
<tr>
<td>2.5</td>
<td>2.0</td>
<td>1.9</td>
<td>4.830</td>
<td>4.042</td>
<td>2.0</td>
<td>2.0</td>
<td>4.392</td>
<td>0.406</td>
</tr>
<tr>
<td>2.0</td>
<td>1.6</td>
<td>2.2</td>
<td>4.435</td>
<td>3.928</td>
<td>1.9</td>
<td>1.9</td>
<td>4.295</td>
<td>0.398</td>
</tr>
<tr>
<td>1.9</td>
<td>2.2</td>
<td>2.3</td>
<td>5.305</td>
<td>4.295</td>
<td>2.1</td>
<td>2.1</td>
<td>4.897</td>
<td>0.409</td>
</tr>
</tbody>
</table>

The table reports slope and intercept estimates from second-stage regressions of NYSE/AMEX/NASDAQ stock returns on factor exposures estimated in first-stage regressions. The first three factors are innovations from AR(1) models of tail risk series estimated using the dynamic power law model for both tails, the lower tail and the upper tail of raw returns (Panel A) and of Fama-French three-factor model residuals (Panel B). Other factors included are AR(1) innovations to the monthly realized volatility of the CRSP value-weighted market portfolio, the excess market return and Fama and French's (1993) SMB and HML returns. Coefficients are standardized to represent the change in average annualized percentage excess returns resulting from a one standard deviation increase in the regressor. The t-statistic is reported below each coefficient in italics. The sample horizon is 1963-2008.
The table reports slope and intercept estimates from second-stage regressions of portfolio returns on factor exposures estimated in first-stage regressions. The first three factors are innovations from AR(1) models of tail risk series estimated using the dynamic power law model for both tails, the lower tail and the upper tail of raw returns. Other factors included are AR(1) innovations to the monthly realized volatility of the CRSP value-weighted market portfolio, the excess market return and Fama and French's (1993) SMB and HML returns. I consider three alternative sets of test assets: 1) NYSE-listed stocks, 2) 100 size and value-sorted portfolios (from Ken French’s Data Library) and 3) 25 portfolios sorted on market beta and tail risk beta. Coefficients are standardized to represent the change in average annualized percentage excess returns resulting from a one standard deviation increase in the regressor. The t-statistic is reported below each coefficient in italics. The sample horizon is 1963-2008.

<table>
<thead>
<tr>
<th></th>
<th>NYSE Stocks</th>
<th>Size / Value-Sorted Portfolios</th>
<th>Market Beta / Tail Beta-Sorted Portfolios</th>
</tr>
</thead>
<tbody>
<tr>
<td>-(\zeta) (both)</td>
<td>-4.367</td>
<td>-1.200</td>
<td>-3.176</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>1.6</td>
<td>3.0</td>
</tr>
<tr>
<td>-(\zeta) (lower)</td>
<td>-5.273</td>
<td>-1.802</td>
<td>-4.238</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>3.3</td>
<td>3.7</td>
</tr>
<tr>
<td>-(\zeta) (upper)</td>
<td>-0.856</td>
<td>-0.360</td>
<td>-0.441</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>1.4</td>
<td>1.1</td>
</tr>
<tr>
<td>R. Vol.</td>
<td>5.785</td>
<td>-0.453</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>5.913</td>
<td>-0.701</td>
<td>-0.100</td>
</tr>
<tr>
<td></td>
<td>6.348</td>
<td>0.724</td>
<td>4.266</td>
</tr>
<tr>
<td></td>
<td>4.201</td>
<td>0.321</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>3.738</td>
<td>0.372</td>
<td>0.252</td>
</tr>
<tr>
<td></td>
<td>4.137</td>
<td>0.322</td>
<td>0.233</td>
</tr>
<tr>
<td>(R_{mkt})</td>
<td>-1.458</td>
<td>-1.307</td>
<td>-2.185</td>
</tr>
<tr>
<td></td>
<td>-1.596</td>
<td>-1.237</td>
<td>-2.114</td>
</tr>
<tr>
<td></td>
<td>-1.193</td>
<td>-1.232</td>
<td>-2.218</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>4.2</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>4.0</td>
<td>3.2</td>
</tr>
<tr>
<td>SMB</td>
<td>-3.338</td>
<td>0.716</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>-3.039</td>
<td>0.665</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>-3.118</td>
<td>0.692</td>
<td>-0.030</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>0.9</td>
<td>0.0</td>
</tr>
<tr>
<td>HML</td>
<td>-1.936</td>
<td>2.365</td>
<td>2.223</td>
</tr>
<tr>
<td></td>
<td>-1.982</td>
<td>2.321</td>
<td>2.011</td>
</tr>
<tr>
<td></td>
<td>-1.772</td>
<td>2.353</td>
<td>1.967</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td>3.7</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>1.8</td>
<td>3.6</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>3.7</td>
<td>3.2</td>
</tr>
<tr>
<td>Intercept</td>
<td>7.196</td>
<td>13.127</td>
<td>20.545</td>
</tr>
<tr>
<td></td>
<td>7.473</td>
<td>16.977</td>
<td>19.674</td>
</tr>
<tr>
<td></td>
<td>7.400</td>
<td>12.943</td>
<td>26.604</td>
</tr>
<tr>
<td></td>
<td>8.378</td>
<td>23.886</td>
<td>14.583</td>
</tr>
<tr>
<td></td>
<td>8.386</td>
<td>23.021</td>
<td>14.829</td>
</tr>
<tr>
<td></td>
<td>7.598</td>
<td>23.085</td>
<td>14.918</td>
</tr>
<tr>
<td></td>
<td>7.6</td>
<td>4.3</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>7.6</td>
<td>4.7</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>4.0</td>
<td>5.3</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.060</td>
<td>0.200</td>
<td>0.523</td>
</tr>
<tr>
<td></td>
<td>0.058</td>
<td>0.199</td>
<td>0.522</td>
</tr>
<tr>
<td></td>
<td>0.059</td>
<td>0.180</td>
<td>0.527</td>
</tr>
<tr>
<td></td>
<td>0.137</td>
<td>0.367</td>
<td>0.702</td>
</tr>
<tr>
<td></td>
<td>0.137</td>
<td>0.369</td>
<td>0.699</td>
</tr>
<tr>
<td></td>
<td>0.134</td>
<td>0.365</td>
<td>0.708</td>
</tr>
</tbody>
</table>

51
Table 11. Dynamic Power Law Monte Carlo Results.
The table reports simulation results based on the Monte Carlo experiment described in Appendix B. In all cases, data is generated by the process $R_{i,t} = b_i R_{m,t} + e_{i,t}$, where $R_{m,t}$ and $e_{i,t}$, $i=1,...,n$ are independent Student $t$ variates with $a_i\zeta_t$ degrees of freedom. I consider four cases: i) Independent and identically distributed observations: $b_i=0$ and $a_i=1$ for all $i$, ii) dependent and identically distributed observations: $b_i \sim N(1,.5)$ and $a_i=1$ for all $i$, iii) independent and heterogeneously distributed observations: $b_i=0$ and $a_i \sim N(1,.2)$ for all $i$, and iv) dependent and heterogeneously distributed observations: $b_i \sim N(1,.5)$ and $a_i \sim N(1,.2)$ for all $i$. The cross section size is $n=1,000$ or $2,500$ and the time series length is $T=1,000$ or $5,000$. Parameters used to generate the data are shown in the “True Value” row. I report the mean, median and standard deviation of parameter estimates across simulations, as well as the mean asymptotic standard error estimate. In the last row of each set of results, I report the mean absolute error and the correlation between the fitted and true $\zeta$ series. The column heading d.o.f. denotes estimates of the intercept parameter (transformed to be interpreted as the time series mean of $\zeta$). Results are based on 1,000 replications.

<table>
<thead>
<tr>
<th></th>
<th>Independent, Identical</th>
<th>Dependent, Identical</th>
<th>Independent, Heterogeneous</th>
<th>Dependent, Heterogeneous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\pi_1$ $\pi_2$ d.o.f.</td>
<td>$\pi_1$ $\pi_2$ d.o.f.</td>
<td>$\pi_1$ $\pi_2$ d.o.f.</td>
<td>$\pi_1$ $\pi_2$ d.o.f.</td>
</tr>
<tr>
<td>True Value</td>
<td>0.050 0.930 3.000</td>
<td>0.050 0.930 3.000</td>
<td>0.050 0.930 3.000</td>
<td>0.050 0.930 3.000</td>
</tr>
<tr>
<td>$n=1,000$ Mean</td>
<td>0.042 0.914 2.910</td>
<td>0.042 0.907 3.248</td>
<td>0.050 0.922 2.358</td>
<td>0.049 0.920 2.640</td>
</tr>
<tr>
<td></td>
<td>0.041 0.925 2.909</td>
<td>0.041 0.919 3.255</td>
<td>0.050 0.925 2.343</td>
<td>0.048 0.926 2.637</td>
</tr>
<tr>
<td>Median</td>
<td>0.015 0.040 0.069</td>
<td>0.015 0.049 0.085</td>
<td>0.016 0.029 0.172</td>
<td>0.014 0.029 0.091</td>
</tr>
<tr>
<td>Mean ASE</td>
<td>0.014 0.049 0.067</td>
<td>0.015 0.074 0.100</td>
<td>0.014 0.029 0.157</td>
<td>0.015 0.031 0.115</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.053 0.980</td>
<td>0.086 0.980</td>
<td>0.166 0.969</td>
<td>0.104 0.981</td>
</tr>
<tr>
<td>MAE, Corr.</td>
<td>0.044 0.920 2.700</td>
<td>0.043 0.912 3.019</td>
<td>0.057 0.923 2.257</td>
<td>0.052 0.925 2.460</td>
</tr>
<tr>
<td></td>
<td>0.043 0.926 2.699</td>
<td>0.042 0.923 3.023</td>
<td>0.057 0.924 2.153</td>
<td>0.051 0.927 2.401</td>
</tr>
<tr>
<td></td>
<td>0.014 0.032 0.048</td>
<td>0.015 0.037 0.051</td>
<td>0.014 0.023 0.125</td>
<td>0.014 0.025 0.144</td>
</tr>
<tr>
<td></td>
<td>0.015 0.041 0.057</td>
<td>0.014 0.046 0.078</td>
<td>0.016 0.030 0.230</td>
<td>0.015 0.028 0.178</td>
</tr>
<tr>
<td></td>
<td>0.085 0.973</td>
<td>0.033 0.977</td>
<td>0.202 0.960</td>
<td>0.153 0.955</td>
</tr>
<tr>
<td>$n=2,500$ Mean</td>
<td>0.041 0.927 2.900</td>
<td>0.040 0.927 3.255</td>
<td>0.049 0.929 2.297</td>
<td>0.049 0.928 2.605</td>
</tr>
<tr>
<td></td>
<td>0.042 0.928 2.901</td>
<td>0.040 0.928 3.255</td>
<td>0.049 0.930 2.299</td>
<td>0.049 0.929 2.609</td>
</tr>
<tr>
<td></td>
<td>0.006 0.013 0.031</td>
<td>0.006 0.014 0.035</td>
<td>0.006 0.010 0.036</td>
<td>0.006 0.010 0.041</td>
</tr>
<tr>
<td></td>
<td>0.006 0.013 0.029</td>
<td>0.007 0.013 0.037</td>
<td>0.006 0.010 0.053</td>
<td>0.006 0.010 0.065</td>
</tr>
<tr>
<td></td>
<td>0.043 0.996</td>
<td>0.076 0.996</td>
<td>0.170 0.994</td>
<td>0.105 0.997</td>
</tr>
<tr>
<td>$n=1,000$ Mean</td>
<td>0.044 0.927 2.684</td>
<td>0.043 0.927 3.021</td>
<td>0.054 0.933 2.232</td>
<td>0.053 0.927 2.352</td>
</tr>
<tr>
<td></td>
<td>0.044 0.929 2.685</td>
<td>0.043 0.927 3.016</td>
<td>0.054 0.931 2.083</td>
<td>0.052 0.928 2.345</td>
</tr>
<tr>
<td></td>
<td>0.006 0.012 0.019</td>
<td>0.006 0.012 0.024</td>
<td>0.006 0.008 0.186</td>
<td>0.006 0.009 0.033</td>
</tr>
<tr>
<td></td>
<td>0.007 0.013 0.021</td>
<td>0.006 0.012 0.024</td>
<td>0.005 0.013 0.267</td>
<td>0.005 0.008 0.078</td>
</tr>
<tr>
<td></td>
<td>0.085 0.995</td>
<td>0.020 0.997</td>
<td>0.203 0.980</td>
<td>0.158 0.990</td>
</tr>
</tbody>
</table>

$T=1,000$  

$T=5,000$
Table 12. Stochastic Tail Exponent Monte Carlo Results.
The table reports simulation results based on the Monte Carlo experiment described in Appendix B. In all cases, data is generated as a vector of \( n \) i.i.d. Student \( t \) variates with \( \zeta \) degrees of freedom over \( T \) periods, where \( \zeta_{t+1} = \zeta(1-\rho) + \rho \zeta_t + \sigma \eta_{t+1}, \eta_t \) is standard normal, \( n=1,000, T=1,000, \) and \( \rho=0.999. \) The standard deviation of the tail risk process is \( \sigma=0.005 \) or \( 0.010 \) (Panels A and B, respectively). I report summary statistics for the true and fitted tail processes, as well as their mean absolute error and correlation, averaged over all simulations. I also report summary statistics of parameter estimates and the mean asymptotic standard error estimate. The column heading \( d.o.f. \) denotes estimates of the intercept parameter (transformed to be interpreted as the time series mean of \( \zeta_t \)). Results are based on 1,000 replications.

### Panel A: \( \sigma=0.005 \)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>True ( \zeta_t )</td>
<td>3.194</td>
<td>0.517</td>
<td>4.688</td>
<td>2.346</td>
</tr>
<tr>
<td>Fitted ( \zeta_t )</td>
<td>3.045</td>
<td>0.311</td>
<td>3.805</td>
<td>2.428</td>
</tr>
<tr>
<td>True/Fitted MAE</td>
<td>0.264</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True/Fitted Correlation</td>
<td>0.818</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>( \pi_1 )</th>
<th>( \pi_2 )</th>
<th>( d.o.f. )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.036</td>
<td>0.954</td>
<td>2.968</td>
</tr>
<tr>
<td>Median</td>
<td>0.036</td>
<td>0.957</td>
<td>2.950</td>
</tr>
<tr>
<td>Mean ASE</td>
<td>0.010</td>
<td>0.015</td>
<td>0.239</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.010</td>
<td>0.020</td>
<td>0.209</td>
</tr>
</tbody>
</table>

### Panel B: \( \sigma=0.010 \)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>True ( \zeta_t )</td>
<td>3.464</td>
<td>1.179</td>
<td>7.390</td>
<td>1.909</td>
</tr>
<tr>
<td>Fitted ( \zeta_t )</td>
<td>3.103</td>
<td>0.634</td>
<td>4.752</td>
<td>1.972</td>
</tr>
<tr>
<td>True/Fitted MAE</td>
<td>0.525</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True/Fitted Correlation</td>
<td>0.875</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>( \pi_1 )</th>
<th>( \pi_2 )</th>
<th>( d.o.f. )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.070</td>
<td>0.923</td>
<td>2.955</td>
</tr>
<tr>
<td>Median</td>
<td>0.070</td>
<td>0.924</td>
<td>2.948</td>
</tr>
<tr>
<td>Mean ASE</td>
<td>0.013</td>
<td>0.015</td>
<td>0.340</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.015</td>
<td>0.017</td>
<td>0.303</td>
</tr>
</tbody>
</table>
Figure 1. CRSP Cross Section Size, 1926-2008.
This figure plots the number of NYSE/AMEX/NASDAQ stocks in the CRSP database each month from 1926 to 2008. The count jumps in July 1962 with the addition of AMEX and in December 1972 with the addition of NASDAQ.

Figure 2. Tail Exponent Estimates (Raw Returns).
Plotted is the fitted tail risk time series ($-\zeta_t$) on the last day of each month. Estimates are found using the dynamic power law model and raw returns for the cross section of NYSE/AMEX/NASDAQ stocks from 1963 to 2008. The tail series is estimated using both tails together (Figure 2a) and the lower and upper separately (Figure 2b). Tail exponent values are shown on the right vertical axis. Shown in the grey shaded region is the aggregate log price-dividend ratio, whose scale is shown on the left vertical axis.

Figure 2a. Exponent Estimated from Both Tails.
Figure 2. Continued.

Figure 2b. Exponent Estimated from Lower and Upper Tails Separately.

Figure 3. Daily Tail Exponent Estimates (Factor Model Residuals).
Plotted is the fitted tail risk time series ($-\zeta_t$) on the last day of each month. Estimates are found using the dynamic power law model and Fama-French three-factor model residuals for the cross section of NYSE/AMEX/NASDAQ stocks from 1963 to 2008. The tail series is estimated using both tails. Tail exponent values are shown on the right vertical axis. Shown in the grey shaded region is the aggregate log price-dividend ratio, whose scale is shown on the left vertical axis.
Figure 4. Tail Threshold and Aggregate Market Volatility (Raw Returns).
Plotted is the tail threshold series on the last day of each month. Estimates are found using the dynamic power law model and raw returns for the cross section of NYSE/AMEX/NASDAQ stocks from 1963 to 2008. The tail series is estimated using both tails together. Also shown is the monthly realized volatility of the CRSP value-weighted index.

Figure 5. Out-of-Sample Aggregate Stock Market Predictive Regressions.
In each month \( t \) (beginning at \( t=60 \) to allow for a sufficiently large initial estimation period), I estimate rolling univariate forecasting regressions of monthly CRSP value-weighted index returns on tail risk \( \zeta_t \), estimated from the lower tail of raw returns. Estimates only use data through date \( t \); these are then used to forecast returns at \( t+1 \). The thick black line shows the sequence of estimated coefficients, and dotted lines represent 95% confidence intervals. The out-of-sample \( R^2 \) from this procedure is 1.30%.
Chapter 2

Dynamic Equicorrelation (with Robert Engle)

Abstract: A new covariance matrix estimator is proposed under the assumption that at every time period all pairwise correlations are equal. This assumption, which is pragmatically applied in various areas of finance, makes it possible to estimate arbitrarily large covariance matrices with ease. The model, called DECO, involves first adjusting for individual volatilities and then estimating correlations. A quasi-maximum likelihood result shows that DECO provides consistent parameter estimates even when the equicorrelation assumption is violated. We demonstrate how to generalize DECO to block equicorrelation structures. DECO estimates for US stock return data show that (block) equicorrelated models can provide a better fit of the data than DCC. Using out-of-sample forecasts, DECO and Block DECO are shown to improve portfolio selection compared to an unrestricted dynamic correlation structure.
2.1 Introduction

Since the first volatility models were formulated in the early eighties there have been efforts to estimate multivariate models. The specification of these models developed over the past 25 years with a range of papers surveyed by Bollerslev, Engle and Nelson (1994) and more recently by Bauwens et al. (2006) and Silvennoinen and Terasvirta (2008). A general conclusion from this analysis is that it is difficult to estimate multivariate GARCH models with more than a half dozen return series because the specifications are so complicated.

Recently, Engle (2002) proposed Dynamic Conditional Correlation (DCC), greatly simplifying multivariate specifications. DCC is designed for high-dimensional systems but has only been successfully applied to up to 100 assets by Engle and Sheppard (2005). Even though there are few parameters to estimate, the maximum likelihood estimator must invert a $100 \times 100$ matrix thousands of times and this is time consuming. In addition, with 100 assets there are 4,950 correlation time series; this output is large and difficult to store or plot. As the size of the system grows, estimation becomes increasingly cumbersome. For cross sections of hundreds or thousands of stocks, as are common in asset pricing applications, estimation can break down completely.

Approaches exist to address the high dimension problem, though each has limitations. One type of approach is to impose structure on the system, such as simplifying the model to take advantage of a factor structure. Univariate GARCH dynamics in factors can generate time-varying correlations while keeping the residual covariance matrix constant through time. This idea motivated the Factor ARCH models of Engle, Ng and Rothschild (1990, 1992). Engle’s (2008) Factor Double ARCH adds volatility dynamics to residuals, which allows the residual covariance to move through time while holding residual correlations constant. The benefit of these models is their feasibility for large numbers of variates: If there are $n$ dependent variables and $k$ factors, estimation requires only $n + k$ GARCH models.
Furthermore, it uses a full likelihood and will be efficient under appropriate conditions. One drawback is that it is not always clear what the factors are, or factor data may not be available. Another is that correlation dynamics can exist in residuals even after controlling for the factors, as in the case of US equity returns (Engle 2008; Engle and Rangel 2008). Addressing either of these problems leads back to the unrestricted DCC specification, and thus to the dimensionality dilemma.

A second solution uses the method of composite likelihood. This method was recently proposed by Engle, Shephard and Sheppard (2008) to estimate unrestricted DCC for vast cross sections. Composite likelihood overcomes the dimension limitation by breaking a large system into many smaller sub-systems in a way that generalizes the “MacGyver” method of Engle (2008). This approach possesses great flexibility by avoiding structural assumptions, but will generally be inefficient due to its reliance on a partial likelihood.

The contrast of Factor ARCH and composite likelihood highlights a fundamental trade-off in large system conditional covariance modeling. Imposing structure on the covariance can make estimation feasible and, if correctly specified, efficient; but it sacrifices generality and can suffer from break-downs due to misspecification. On the other hand, less structured models like composite likelihood break the curse of dimensionality while maintaining a general specification. However, its cost is a loss of efficiency from using a partial likelihood.

We propose a new way to address this trade-off that can selectively combine simplifying structural assumptions and composite likelihood versatility. We consider a system in which all pairs of returns have the same correlation on a given day but this correlation varies over time. The model, called Dynamic Equicorrelation (DECO), eliminates the computational and presentational difficulties of high dimension systems. Because equicorrelated matrices have simple analytic inverses and determinants, likelihood calculation is dramatically simplified and optimization becomes feasible for vast numbers of assets.

Intuitively, DECO will work well when there is a dominating component of pairwise
correlations inducing all pairwise correlations to move together. As discussed below, the alarming noisiness in pairwise correlation estimates often renders them unreliable for financial applications. Since DECO effectively averages correlations, noise can be greatly reduced and the dominant element of their time variation more easily discerned.

In a one-factor world, the relation between the return on an asset and the market return is

\[ r_j = \beta_j r_m + e_j, \quad \sigma_j^2 = \beta_j^2 \sigma_m^2 + v_j. \]

If the cross sectional dispersion of \( \beta_j \) is small and idiosyncrasies have similar variance each period, then the system is well-described by Dynamic Equicorrelation. As we demonstrate in empirical results, such a model provides a reasonable description of correlations in a large cross section of US stocks.

A natural application for this one-factor structure lies in the market for credit derivatives such as collateralized debt obligations, or CDO’s. A key feature of the risk in loan portfolios is the degree of correlation between default probabilities. A simple industry valuation model allows this correlation to be one number if firms are in the same industry and a different and smaller number if they are in different industries. Hence, within each industry an equicorrelation assumption is being made.

More broadly, to price CDO’s, an assumption is often made that these are large homogeneous portfolios (LHP’s) of corporate debt. As a consequence, each asset will have the same variance, the same covariance with the market factor and the same idiosyncratic variance. Thus, in an LHP, the \( j \) subscripts disappear. The correlation between any pair of assets then becomes

\[ \rho = \frac{\beta_j^2 \sigma_m^2}{\beta_j^2 \sigma_m^2 + v}. \]

In fact, the LHP assumption implies equicorrelation.

DECO’s structure can be substantially weakened by using block equicorrelated matri-
ces, while maintaining the simplicity and robustness of the basic DECO formulation. A block model may capture, for instance, industry correlation structures. All stocks within an industry share the same correlation while correlations between industries take another value. In the two-block setting, analytic inverses and determinants are still available and fairly simple, thus optimization for the two-Block DECO model is as easy as the one block case. Adding more blocks quickly confounds DECO’s analytic simplicity, however. To overcome this, we show that the two-block structure can be combined with the method of composite likelihood to estimate Block DECO with an arbitrary number of blocks. In this sense, DECO incorporates the estimation advantages of a structural approach as well as the flexibility of composite likelihood. Since the subsets of assets used in Block DECO are pairs of blocks rather than pairs of assets, a larger portion of the likelihood (and therefore more information) is used for optimization. As a result, the estimator can be more efficient than unrestricted composite likelihood DCC.

Another way to enrich dependence beyond equicorrelation is to combine DECO with Factor (Double) ARCH. To understand how this may work, consider a model in which one factor is observable and the dispersion of loadings on this factor is high. Further, suppose each asset loads roughly the same on a second, latent, factor. The first factor contributes to diversity among pairwise correlations, and is clearly not driven by noise. Thus, DECO will be a poor candidate for describing raw returns. However, residuals from a regression of returns on only the first factor will be well described by DECO. One way to model this data set is to use Factor Double ARCH with DECO residuals. Such a model is estimated by first estimating GARCH regression models for each stock, then estimating DECO on the standardized residuals.

What will occur if DECO is applied to variables that are not equicorrelated? If (block) equicorrelation is violated, DECO can still provide consistent estimates. In particular, we prove quasi-maximum likelihood results showing that if DCC is a consistent estimator, then
DECO and Block DECO will be consistent also. However, like composite likelihood, it will be inefficient. The relative efficiency of DECO and composite likelihood DCC, in this case, will depend on the severity of efficiency loss due to misspecification (DECO) versus the loss due to partial likelihood (composite likelihood DCC).

There is a substantial history of the use of equicorrelation in economics. In early studies of asset allocation, Elton and Gruber (1973) found that assuming all pairs of assets had the same correlation reduced estimation noise and provided superior portfolio allocations over a wide range of alternative assumptions. Ledoit and Wolf (2004) use Bayesian methods for shrinking the sample correlation matrix to an equicorrelated target and show that this helps select portfolios with low volatility compared to those based on the sample correlation. Further prescriptions for avoiding the notorious noisiness of unrestricted sample correlations abound in the literature (Ledoit and Wolf 2003, 2004; Michaud 1989; Jagannathan and Ma 2003; and Jobson and Korkie 1980, among others). Ledoit and Wolf’s Bayesian shrinkage and Elton and Gruber’s parameter averaging are different approaches to noise reduction in unconditional correlation estimation. While the Bayesian method has not yet been employed for conditional variances, DECO makes it possible to incorporate Elton and Gruber’s noise reduction technique into a dynamic setting. By averaging pairwise correlations, (Block) DECO smooths correlation estimates within groups. As long as this reduces estimation noise more than it compromises the true correlation structure, smoothing can be beneficial. Our empirical results suggest that the benefits of smoothing indeed extend to the conditional case. Across a range of first-stage factor models, (Block) DECO selects out-of-sample portfolios that have significantly lower volatilities than those chosen by unrestricted DCC.

The equicorrelation assumption also surfaces in derivatives trading. For instance, a popular position is to buy an option on a basket of assets and then sell options on each of the components, sometimes called a dispersion trade. By delta hedging each option, the value of this position can be seen to depend solely on the correlations. Let the basket have
weights given by the vector $w$, and let the implied covariance matrix of components of the basket be given by the matrix $S$. Then the variance of the basket can be expressed as

$$\sigma^2 = w' Sw.$$ 

In general we only know about the variances of implied distributions, not the covariances. Hence it is common to assume that all correlations are equal, giving

$$\sigma^2 = \sum_{j=1}^{n} w_j^2 s_j^2 + \rho \sum_{i \neq j} w_i w_j s_i s_j$$

which can be solved for the implied correlation

$$\rho = \frac{\sigma^2 - \sum_{j=1}^{n} w_j^2 s_j^2}{\sum_{i \neq j} w_i w_j s_i s_j}.$$ 

As a consequence, the value of this position depends upon the evolution of the implied correlation. When each of the variances is a variance swap made up of a portfolio of options, the full position is called a correlation swap. As the implied correlation rises, the value of the basket variance swap rises relative to the component variance swaps.

The next section develops the DECO model and theoretically demonstrates its robustness to violations of equicorrelation by appealing to quasi-maximum likelihood theory. In particular, when the true model is DCC, DECO provides a feasible estimation method when the dimension of the system may be otherwise too large for DCC to handle. We then explore loosening the equicorrelation structure to allow block equicorrelation and discuss how it may be estimated with composite likelihood.

While DECO is closely related to DCC, the two models are non-nested: DECO is not simply a restricted version of DCC, but a competing model. Indeed, DECO possesses some subtle, though important, features lacking in DCC. A key example is that DECO
correlations between any pair of assets \( i \) and \( j \) depend on the return histories of all pairs. For the analogous DCC specification (ie., using the same number of parameters), the \( i,j \) correlation depends on the histories of \( i \) and \( j \) alone. In this sense, DECO parsimoniously draws on a broader information set when formulating the correlation process of each pair. To the extent that true correlations are affected by realizations of all assets, the failure of DCC to capture the information pooling aspect of DECO can disadvantage it as a descriptor of the data generating process.

Section 2.3 presents Monte Carlo experiments that assess the model’s performance under equicorrelated and non-equicorrelated generating processes. Simulations demonstrate the optimality of the DECO estimator when data is truly equicorrelated. DCC performs poorly when estimating DECO processes due to differences between the two models’ likelihoods, discussed in more detail in Section 2.2.

When the data obeys DCC, thus violating equicorrelation, DECO appears to remain consistent, though is substantially less efficient than DCC, as QML theory suggests. We see that when the time series length is short relative to the size of the cross section, DECO has a tendency to produce a downward bias in its two key parameters. As the ratio \( T/N \) grows, however, this bias becomes imperceptible.

In Section 2.4, we apply DECO, Block DECO and DCC models to US stock return data. We find that DCC correlations between pairs of stocks have a large degree of comovement, suggesting that DECO may be beneficial in describing the system’s correlation. Indeed we find that basic DECO, and DECO with 10 industry blocks, provide a better fit of the data than DCC. Next, we extract residuals using the CAPM and Fama-French three-factor model. We find that pairwise correlations in factor model residuals behave much more independently, without any clear comovement patterns outside of certain episodes such as the technology bubble of the late nineties and the credit crisis in the second half of 2008. As a result, DECO does a poorer job than DCC of fitting the data after accounting for factors.
Finally, we analyze the ability of (Block) DECO to construct optimal out-of-sample hedge portfolios. As we will show, (Block) DECO and DCC share the same basic covariance evolution, and thus the same evolution parameters. In light of this, we ask: What is the best estimator for forecasting the correlation evolution in terms of its ability to form minimum variance (MV) portfolios. Then, we discuss how the block structure of fitted correlation matrices can be altered after estimation. For instance, DECO can be used to estimate model parameters and generate an unrestricted fitted correlation process with each pair taking unique values. Next, the fitted correlations can be used to construct any other block form by simply averaging correlations within blocks. This leads to another question. After estimating any given DCC or DECO model, is there a best ex post block structure to use for portfolio formation?

We find that (Block) DECO is the model that most often delivers MV portfolios with the lowest sample variance. This is true whether the first-stage structure for returns has no factors, is the one-factor CAPM, or is the Fama-French three-factor model.

When we vary the ex post block structure using each model’s fitted parameters, we see that results depend importantly on the first-stage factor structure. When no factor structure is imposed, there is a clear benefit to ex post correlation averaging within blocks. This is true whether Block DECO or DCC is the correlation model. However, when factors are accounted for, the benefit to averaging fitted residual correlations within blocks is no longer clear. In summary, for US equity returns, there appears to be a benefit to using (Block) DECO for both estimation and ex post portfolio formation in terms of in-sample and out-of-sample performance.
2.2 The Dynamic Equicorrelation Model

We begin by defining an equicorrelation matrix and present a result for its invertibility and positive definiteness that will be useful throughout the paper.

**Definition 2.2.1.** A matrix $R_t$ is an equicorrelation matrix of an $n \times 1$ vector of random variables if it is positive definite and takes the form

$$R_t = (1 - \rho_t)I_n + \rho_t J_n \quad (2.1)$$

where $\rho_t$ is the equicorrelation, $I_n$ denotes the $n$-dimensional identity matrix and $J_n$ is the $n \times n$ matrix of ones.

**Lemma 2.** The inverse and determinant of the equicorrelation matrix, $R_t$, are given by

$$R_t^{-1} = \frac{1}{1 - \rho_t} I_n + \frac{-\rho_t}{(1 - \rho_t)(1 + [n - 1]\rho_t)} J_n \quad (2.2)$$

and

$$\det(R_t) = (1 - \rho_t)^{n-1}(1 + [n - 1]\rho_t). \quad (2.3)$$

Further, $R_t^{-1}$ exists if and only if $\rho_t \neq 1$ and $\rho_t \neq -\frac{1}{n-1}$, and $R_t$ is positive definite if and only if $\rho_t \in (-\frac{1}{n-1}, 1)$.

**Proof:** For Equations 2.2 and 2.3 see Graybill (1983), Theorems 8.3.4 and 8.4.4. Existence of an inverse relies on non-zero denominators in Equation 2.2. The positive definiteness condition derives from its equivalence with all eigenvalues being positive; this equivalence can be seen in Equation 2.3 which is the product of the eigenvalues of $R_t$. *Q.E.D.*

**Definition 2.2.2.** A time series of $n \times 1$ vectors $\{\tilde{r}_t\}$ obeys a Dynamic Equicorrelation (DECO) model if $\text{Var}_{t-1}(\tilde{r}_t) = D_t R_t D_t$, where $R_t$ is given by Equation 2.1 for all $t$ and $D_t$
is the diagonal matrix of conditional standard deviations of \( \tilde{r}_t \). The dynamic equicorrelation is \( \rho_t \).

### 2.2.1 Estimation

Like many covariance models, a two-stage quasi-maximum likelihood (QML) estimator of DECO will be consistent and asymptotically normal under broad conditions including many forms of model misspecification. We provide asymptotic results here for a general framework that includes several standard multivariate GARCH models as special cases, including DECO and the original DCC model. The development is a slightly modified reproduction of the two-step QML asymptotics of White (1994). After presenting the general result, we elaborate on practical estimation of DECO using Gaussian returns and GARCH covariance evolution.

First, define the (scaled) log quasi-likelihood of the model as
\[
L \left( \{ \tilde{r}_t \}, \theta, \phi \right) = \frac{1}{T} \sum_{t=1}^{T} \log f_{1,t}(\tilde{r}_t, \theta, \phi),
\]
which is parameterized by vector \( \gamma = (\theta, \phi) \in \Gamma = \Theta \times \Phi \). The two-step estimation problem may be written as
\[
\max_{\theta \in \Theta} L_1(\{ \tilde{r}_t \}, \theta) = \frac{1}{T} \sum_{t=1}^{T} \log f_{1,t}(\tilde{r}_t, \theta) \tag{2.4}
\]
\[
\max_{\phi \in \Phi} L_2(\{ \tilde{r}_t \}, \hat{\theta}, \phi) = \frac{1}{T} \sum_{t=1}^{T} \log f_{2,t}(\tilde{r}_t, \hat{\theta}, \phi) \tag{2.5}
\]
where \( \hat{\theta} \) is the solution to (2.4), and \( \hat{\phi} \) is the second stage maximizer solving (2.5) given \( \hat{\theta} \). The full two-stage QML estimator for this problem is \( \hat{\gamma} = (\hat{\theta}, \hat{\phi}) \). Under the technical assumptions listed in Appendix B.1, White (1994) proves the following result for consistency and asymptotic normality of \( \hat{\gamma} \).

**Proposition 2.2.1.** **Under Assumptions [3] through [8],**
\[
\sqrt{T}(\hat{\gamma} - \gamma^*) \overset{A}{\sim} N(0, A^*^{-1}B^*A^*-1),
\]
where

\[
A^* = \begin{pmatrix}
\nabla_{\theta} E[L_1(\tilde{r}_t, \theta^*)] & 0 \\
\nabla_{\theta} E[L_2(\tilde{r}_t, \theta^*, \phi^*)] & \nabla_{\phi} E[L_2(\tilde{r}_t, \theta^*, \phi^*)]
\end{pmatrix}
\]

and

\[
B^* = \text{Var}(T^{-1/2} \sum_t (s^*_{1,t}, s^*_{2,t}))
\]

where \(s^*_{1,t} = \nabla_{\theta} L_1(\tilde{r}_t, \theta^*)\) and \(s^*_{2,t} = \nabla_{\phi} L_2(\tilde{r}_t, \theta^*, \phi^*)\).

DECO is adopted for individual applications by specifying a conditional volatility model (i.e., defining the process for \(D_t\)) and a \(\rho_t\) process. We assume that each conditional volatility follows a GARCH model. We work with volatility-standardized returns, denoted by omitting the tilde, \(r_t = D_t^{-1} \tilde{r}_t\), so that \(\text{Var}(r_t) = R_t\).

The basic \(\rho_t\) specification we consider derives from the DCC model of Engle (2002) and its cDCC modification proposed by Aielli (2006). The correlation matrix of standardized returns, \(R_t^\text{DCC}\), is given by

\[
Q_t = \tilde{Q}(1 - \alpha - \beta) + \alpha \tilde{Q}_{t-1}^{1/2} r_{t-1} r_{t-1}' \tilde{Q}_{t-1}^{1/2} + \beta Q_{t-1}.
\]

\[
R_t^\text{DCC} = \tilde{Q}_t^{-1/2} Q_t \tilde{Q}_t^{-1/2}
\]

where \(\tilde{Q}_t\) replaces the off-diagonal elements of \(Q_t\) with zeros but retains its main diagonal.

DECO sets \(\rho_t\) equal to the average pairwise DCC correlation.

\[
R_t^{\text{DECO}} = (1 - \rho_t) I_n + \rho_t J_{n \times n}
\]

\[
\rho_t = \frac{1}{n(n-1)} \left( \ell' R_t^{\text{DCC}} - n \right) = \frac{2}{n(n-1)} \sum_{i > j} \frac{q_{i,j,t}}{\sqrt{q_{i,i,t} q_{j,j,t}}}
\]

where \(q_{i,j,t}\) is the \(i,j\)th element of \(Q_t\). The following assumption and lemma ensure that DECO possesses certain properties important for dynamic correlation models.

Assumption[section]
Assumption 2. The matrix $\bar{Q}$ is positive definite, $\alpha + \beta < 1$, $\alpha > 0$ and $\beta > 0$.

Lemma 3. Under Assumption 2, the correlation matrices generated by every realization of a DECO process according to Equations 2.6 through 2.9 are positive definite and the process is mean reverting.

Proof: Following from Lemma 2, it is sufficient for positive definiteness (and hence invertibility) to show that $\rho_t \in (\frac{-1}{n-1}, 1) \forall t$. To this end, note that $Q_t$ is a weighted average of positive definite matrices and is therefore positive definite. This and (2.7) imply $R_t^{DCC}$ is also positive definite with ones along the main diagonal. It follows that $\rho_t > \frac{-1}{n-1}$. Since $\rho_t$ is an average of terms bounded above by one, we obtain the upper bound, $\rho_t < 1$.

As discussed by Aielli (2006), $Q_t$ will be mean reverting under the condition $\alpha + \beta < 1$, which implies that $\rho_t$ and the DECO correlation matrix, which are well-behaved functions of $Q_t$, will also be mean reverting. Q.E.D.

The result states that, for any correlation matrix, the transformation to equicorrelation shown in Equations 2.8 and 2.9 results in a positive definite matrix. The bounds $(\frac{-1}{n-1}, 1)$ for $\rho_t$ are not assumptions of the model, but are guaranteed by this transformation as long as $R_t^{DCC}$ is positive definite. As will be seen in our later discussion of Block DECO, bounds on permissible correlation values become looser as the number of blocks increases.

We estimate DECO with Gaussian quasi-maximum likelihood, which embeds it in the framework of Proposition 2.2.1. Conditional on past realizations, the return distribution is $\tilde{r}_{t|t-1} \sim N(0, H_t)$, $H_t = D_t R_t D_t$. To establish notation, we use superscripted densities $f_t^{DECO}$ and $f_t^{DCC}$ to indicate the Gaussian density of $\tilde{r}_{t|t-1}$ assuming the covariance specifically obeys DECO or DCC, respectively. Similarly, superscripted $L$’s represent the log likelihood of the corresponding model. Omission of superscripts will be used to discuss densities and log likelihoods without specific assumptions on the dynamics or structure of covariance matrices.

69
The multivariate Gaussian log likelihood function $L$ can be decomposed (suppressing constants) as

\[
L = \frac{1}{T} \sum_{t} \log(f_t)
\]

\[
= -\frac{1}{T} \sum_{t} (\log |H_t| + \tilde{r}_t' H_t^{-1} \tilde{r}_t)
\]

\[
= -\frac{1}{T} \sum_{t} (\log |D_t|^2 + \tilde{r}_t' D_t^{-2} \tilde{r}_t - r_t' r_t) - \frac{1}{T} \sum_{t} (\log |R_t| + r_t' R_t^{-1} r_t)
\]

Let $\theta \in \Theta$ and $\phi \in \Phi$ denote the vector of univariate volatility parameters and the vector of correlation parameters, respectively. The above equation says that the log likelihood can be separated additively into two terms. The first term, which we call $L_{Vol}(\theta)$, depends on the parameters of the univariate GARCH processes which affect only the $D_t$ matrices and are independent of $R_t$ and $\phi$. The second term, which we call $L_{Corr}(\phi, \theta)$, depends on both the univariate GARCH parameters (embedded in the $r_t$ terms) as well as the correlation parameters. Engle (2002) and Engle and Sheppard (2005) note that correlation models of this form satisfy the assumptions of Proposition 2.2.1. In particular, $L_{Vol}(\theta)$ corresponds to $L_1$ in the proposition and $\hat{\theta}$ is the vector of first-stage volatility parameter estimates. $L_{Corr}(\hat{\theta}, \phi)$ corresponds to the second-step likelihood, $L_2$, so $\hat{\phi}$ is the maximizer of $L_{Corr}(\hat{\theta}, \phi)$.

It is easily verified that the Gaussian model will automatically satisfy several of the stated assumptions necessary for consistency and asymptotic normality as long as the covariance matrix follows standard, well-behaved multivariate GARCH evolutions. In particular, it is important that the model ensures stationarity and the covariance matrix remains positive definite in all periods. This will be the case when Assumption 2 holds and the univariate GARCH models are stationary.

The vector $\hat{\gamma}^{DECO}$ is the two-stage Gaussian estimator when the second stage likelihood obeys DECO. It assumes returns are Gaussian and the correlation process obeys Equations
Corollary 2.2.1. If $\hat{\gamma}^{DECO}$ satisfies the identification condition in Assumption 7 with corresponding unique maximizer $\gamma^*$, then $\hat{\gamma}^{DECO}$ is consistent and asymptotically normal for $\gamma^*$.

The analogous corollary for $\hat{\gamma}^{DCC}$, the two-stage Gaussian DCC estimator, will be useful in developing our key proposition in the following subsection.

Corollary 2.2.2. If $\hat{\gamma}^{DCC}$ satisfies the identification condition in Assumption 7 with corresponding unique maximizer $\gamma^*$, then $\hat{\gamma}^{DCC}$ is consistent and asymptotically normal for $\gamma^*$.

Due to the Gaussianity of the models, the remaining assumptions of Proposition 2.2.1 are also satisfied, and the above results follow.

To appreciate the payoff from making the equicorrelation assumption, consider the second step likelihood under DECO.

$$L_{\text{Cor},t}^{DECO}(\hat{\theta}, \phi) = -\frac{1}{T} \sum_t \left( \log |R_t^{DECO}| + \hat{\gamma}_t D_t^{\hat{\theta}} \right)$$

$$= -\frac{1}{T} \sum_t \left[ \log \left( [1 - \rho_t] \sum_{i} \hat{\gamma}_{i,t}^2 \right) - \frac{\rho_t}{1 + (n-1)\rho_t} \left( \sum_{i} \hat{\gamma}_{i,t} \right)^2 \right]$$

where $\hat{\gamma}_t$ are returns standardized for first stage volatility estimates, $\hat{\gamma}_t = D_t(\hat{\theta})^{-1} \tilde{\gamma}_t$, and $\rho_t$ obeys Equation 2.9. In DCC, the conditional correlation matrices must be recorded and inverted for all $t$ and their determinants calculated; further, these $T$ inversions and determinant calculations are repeated for each of the many iterations required in a numeric optimization program. This is costly for small cross sections and potentially infeasible for very large ones. In contrast, DECO reduces computation to $n$-dimensional vector outer products with no
matrix inversions or determinants required, rendering the likelihood optimization problem manageable even for vast-dimensional systems.

A remaining question is how to estimate the potentially large number of parameters in $\tilde{Q}$. A common simplification in the DCC literature is “correlation targeting,” meaning that $\tilde{Q}$ is taken to be the sample correlation matrix estimated in a preliminary step rather than jointly with $\alpha$ and $\beta$ in the second stage. The motivation for this approach is that an unrestricted $\tilde{Q}$ contains $n(n - 1)/2$ parameters, another curse of dimensionality that may preclude ML estimation. Correlation targeting can be used in the DECO specification; however, alternative forms that limit the parameterization of $\tilde{Q}$, such as the the restriction that all off-diagonal elements are equal, can greatly reduce the total parametrization and be easily estimated in the second stage.

### 2.2.2 Differences Between DECO and DCC

The transformation from DCC correlations to DECO in Equation $2.9$ introduces subtle differences between the DECO and DCC likelihoods. The two models, in fact, are non-nested; they share the same number of parameters and there is no parameter restriction that makes the models identical. The correlation matrix for DCC is non-equicorrelated in all realizations, while, by definition, DECO is always exactly equicorrelated in all samples.

Both models build off of the $Q$ process in Equation $2.6$. On a given day, $Q$ is updated as a function of the lagged return vector and the lagged $Q$ matrix. From here, $Q$ is transformed to $R^{DCC} = \tilde{Q}_t^{-\frac{1}{2}}Q_t\tilde{Q}_t^{-\frac{1}{2}}$. $R^{DCC}$ is the correlation matrix that enters the DCC likelihood. Note that the $i,j$ element of this matrix is $q_{i,j,t}/\sqrt{q_{i,i,t}q_{j,j,t}}$. Clearly, the information about pair $i,j$’s correlation at time $t$ depends on the history of assets $i$ and $j$ alone. On the other
hand, the correlation between $i$ and $j$ under DECO is

$$
\rho_t = \frac{2}{n(n-1)} \sum_{i>j} \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}}
$$

which depends on the history of all pairs. The failure of DCC to capture this information pooling aspect of DECO correlations hinders the ability of the DCC likelihood to provide a good description of the data generating process, resulting in poor estimation performance, as will be seen in the Monte Carlo results of Section 2.3.

There is also a key difference that arises from the need to estimate DCC with a partial, composite likelihood. When DECO is the true model, DCC estimation is akin to estimating the correlation of a single pair, sampled $n(n-1)/2$ times. The difference between each pair is measurement error. DECO, by averaging pairwise correlation at each step, attenuates this measurement error. It also uses the full cross sectional likelihood rather than stringing bivariate likelihoods together like composite likelihood. The outcome is a more accurate estimate of correlations for all pairs.

### 2.2.3 DECO as a Feasible DCC Estimator

Often the equicorrelation assumption fails so that there is cross-sectional variation in pairwise correlations, as in DCC. In this case the DECO model remains a powerful tool. The following result shows that as long as DCC is a Gaussian QML estimator, DECO will be also.

**Proposition 2.2.2.** If $\hat{\gamma}_{DCC}$ satisfies Corollary 2.2.2 so that it is a consistent and asymptotically normal QML estimator of $\gamma^*$, then $\hat{\gamma}_{DECO}$ is also consistent and asymptotically normal.

**Proof:** By assumption, the correlation process of the second stage DCC model is correct when $\phi = \phi^*$. We proceed by demonstrating that this implies the DECO score has expected value zero only when $\phi = \phi^*$, thus DECO is identified and satisfies Corollary 2.2.1.
The expectation of the DECO score is

\[
E\left[ \frac{\partial \log f^{DECO}_{2,t}(\tilde{r}, \theta^*, \phi)}{\partial \theta_k} \right] = E\left[ E_{t-1}\left[ \frac{\partial \log f^{DECO}_{2,t}(\tilde{r}, \theta^*, \phi)}{\partial \rho_t} \right] \frac{\partial \rho_t}{\partial \phi_k} \right].
\]  

(2.10)

This equality follows from \( \partial \rho_t/\partial \phi_k \) being \((t - 1)\)-measurable. Based on the second stage likelihood in Equation 2.10, the derivative inside the conditional expectation can be written

\[
\frac{\partial \log f^{DECO}_{2,t}(\tilde{r}, \theta^*, \phi)}{\partial \rho_t} = (1 - \rho_t)^{-2}(1 + [n - 1] \rho_t)^{-2}\left[ (n - 1)(1 - \rho_t)^2(1 + [n - 1] \rho_t) - (n - 1)(1 - \rho_t)(1 + [n - 1] \rho_t)^2 + (1 + [n - 1] \rho_t)^2 \sum_i r_{i,t}^2 - (1 + [n - 1] \rho_t)^2(\sum_i r_{i,t}^2) \right].
\]

This expression uses \( r_t \) rather than \( \hat{r}_t \) because the function is evaluated at \( \theta^* \) rather than \( \hat{\theta} \). When \( \phi = \phi^* \), that is, when DCC is the true correlation process, \( \sum_i r_{i,t}^2 \) and \( (\sum_i r_{i,t})^2 \) have \((t - 1)\)-conditional expectations of \( n \) and \( \sum_{i,j} \rho_{i,j,t} = n(n - 1) \rho_t + n \), respectively. In this case only, \( E_{t-1}[\partial \log f^{DECO}_{2,t}(\tilde{r}, \hat{\theta}, \phi)/\partial \rho_t] \) reduces to zero, and as a result (2.10) is also zero. 

Q.E.D.

2.2.4 Estimation Structure Versus Fit Structure

How useful is Proposition 2.2.2 in practice? Suppose, for instance, the system is so large that DCC estimation based on full maximum likelihood is infeasible. The result says that one can consistently estimate DCC parameters using DECO despite its misspecification. The estimated parameters can then be plugged into Equation 2.6 to reconstruct the unrestricted DCC fitted process. In short, DECO, like composite likelihood, provides feasible estimates for a DCC model that may be otherwise computationally infeasible.

The flexibility of DECO goes beyond its ability to fit unrestricted DCC processes. The
logic of Proposition 2.2.2 ensures that DECO can consistently estimate block equicorrelation processes as well. To do this, $\alpha$ and $\beta$ are estimated with DECO, then data is run through the evolution equation in (2.6), plugging in $\hat{\alpha}^{DECO}$ and $\hat{\beta}^{DECO}$. The resulting DCC correlation series, based on DECO estimates, can be used to construct any fitted block correlation structure after the fact by averaging pairwise DCC correlations within blocks (see schematic in Figure 1).

Throughout the remaining sections we will refer to “estimation structures” and “fit structures,” and it is important to draw the distinction between them. The estimation structure is the structure that the correlation matrix takes within the likelihood. When DECO is used, the estimation structure is a single block. The fit structure, on the other hand, refers to the structure of the final, fitted correlation matrices. It is achieved by averaging DCC correlations within blocks after estimation. The resulting block structure can be different than the estimation structure and have one block, many blocks, or be unrestricted (as in DCC).

In the next section, we present an alternative estimation approach called Block DECO. Block DECO directly models the block correlation structure ex ante and makes use of it within the estimation procedure. In this case, the estimation structure will be allowed to have multiple blocks. As with DECO, ex post block averaging can be used to generate a different desired correlation fit structure. With Block DECO as the estimator, fitted correlations can have the same, more, or fewer blocks than the estimation structure.

Using DECO with ex post averaging to achieve block correlations is, from an implementation standpoint, simpler than using full-fledged Block DECO estimation. As will be shown, Block DECO estimation involves composite likelihood and thus is operationally more complex. Ex post averaging achieves the same outcome of dynamic block correlations with the simplicity of DECO’s Gaussian QML estimation. The advantage of more complicated Block DECO estimation is that it can potentially be more efficient. We turn to that model
2.2.5 The Block Dynamic Equicorrelation Model

While DECO will be consistent even when equicorrelation is violated, it is possible that a loosening of the structure to block equicorrelation can improve maximum likelihood estimates. In this vein, we extend DECO to take the block structure into account ex ante and thus incorporate it into the estimation procedure.

As an example of Block DECO’s usefulness, consider modeling correlation of stock returns with particular interest in intra- and inter-industry correlation dynamics. This may be done by imposing equicorrelation within and between industries. Each industry has a single dynamic equicorrelation parameter and each industry pair has a dynamic cross-equicorrelation parameter. With block equicorrelation, richer cross-sectional variation is accommodated while still greatly reducing the effective dimensionality of the correlation matrix.

This section presents the class of block Dynamic Equicorrelation models and examines their properties.

Definition 2.2.3. $R_t$ is a $K$-block equicorrelation matrix if it is positive definite and takes the form

$$R_t = \begin{pmatrix} (1 - \rho_{1,1,t})I_{n_1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & 0 & (1 - \rho_{K,K,t})I_{n_K} \end{pmatrix} + \begin{pmatrix} \rho_{1,1,t}J_{n_1} & \rho_{1,2,t}J_{n_1 \times n_2} & \cdots \\ \rho_{2,1,t}J_{n_2 \times n_1} & \ddots & \vdots \\ \vdots & \ddots & \rho_{K,K,t}J_{n_K} \end{pmatrix}$$

(2.11)

where $\rho_{l,m,t} = \rho_{m,l,t}$ $\forall l, m$.

Block DECO specifies that, conditional on the past, each variable is Gaussian with mean zero, variance one, and correlations taking the structure in Equation (2.11). The return vector
$r_t$ is partitioned into $K$ sub-vectors; each sub-vector $r_t$ contains $n_t$ returns. The Block DECO correlation matrix, $R_{t}^{BD}$, allows distinct processes for each of the $K$ diagonal blocks and $K(K - 1)/2$ unique off-diagonal blocks. Blocks on the main diagonal have equicorrelations following

$$\rho_{l,l,t} = \frac{1}{n_t(n_t - 1)} \sum_{i \in l, j \in l, i \neq j} \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}}, \tag{2.12}$$

and blocks off the main diagonal follow

$$\rho_{l,m,t} = \frac{1}{n_t n_m} \sum_{i \in l, j \in m} \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}}. \tag{2.13}$$

The $i,j$th element of the matrix in (2.6) is $q_{i,j,t}$, so Block DECO correlations are calculated as the average DCC correlation within each block.

The following results show the consistency and asymptotic normality of Block DECO. In analogy to DECO, $\hat{\gamma}_{BD}$ is the two-stage Gaussian Block DECO estimator assuming returns are Gaussian and the correlation process obeys Equations 2.12 and 2.13.

**Corollary 2.2.3.** If $\hat{\gamma}_{BD}$ satisfies the identification condition in Assumption 7 with corresponding unique maximizer $\gamma^*$, then $\hat{\gamma}_{BD}$ is consistent and asymptotically normal for $\gamma^*$.

Further, like DECO, Block DECO is a QML estimator of DCC models.

**Proposition 2.2.3.** If $\hat{\gamma}_{DCC}$ satisfies Corollary 2.2.2 so that it is a consistent and asymptotically normal QML estimator of $\gamma^*$, then $\hat{\gamma}_{BD}$ is also consistent and asymptotically normal, though with different asymptotic covariance matrix.

The proof follows the same argument as the proof of Proposition 2.2.2.

Block DECO balances the flexibility of unrestricted correlations with the structural simplicity of DECO. However, when the number of blocks is greater than two, the analytic forms for the inverse and determinant of the Block DECO matrix begin to lose their tractability. A
special case that remains simple regardless of the number of blocks occurs when each of the blocks on the main diagonal are equicorrelated, but all off-diagonal block equicorrelations are forced to zero. Each diagonal block constitutes a small DECO submodel, and therefore its inverse and determinant are known. The full inverse matrix is the block diagonal matrix of inverses for the DECO sub-models, and its determinant is the product of the sub-model determinants.

Conveniently, the composite likelihood method can be used to estimate Block DECO in more general cases. The composite likelihood is constructed by treating each pair of blocks as a sub-model, then calculating the quasi-likelihoods of each sub-model, and finally summing quasi-likelihoods over all block pairs. As discussed in Engle, Shephard and Sheppard (2008), each pair provides a valid, though only partially informative, quasi-likelihood. A model for any number of blocks requires only the analytic inverse and determinant for a two-block equicorrelation matrix when using the method of composite likelihood. The following lemma establishes the analytic tractability provided by two-block equicorrelation. We suppress $t$ subscripts as all terms are contemporaneous.

Lemma 4. If $R$ is a two-block equicorrelation matrix, that is, if

$$R = \begin{pmatrix} (1 - \rho_{1,1})I_{n_1} & 0 \\ 0 & (1 - \rho_{2,2})I_{n_2} \end{pmatrix} + \begin{pmatrix} \rho_{1,1,1}J_{n_1 \times n_1} & \rho_{1,2,1}J_{n_1 \times n_2} \\ \rho_{2,1,1}J_{n_2 \times n_1} & \rho_{2,2,2}J_{n_2 \times n_2} \end{pmatrix}$$

then,

i. the inverse is given by

$$R^{-1} = \begin{pmatrix} b_1I_{n_1} & 0 \\ 0 & b_2I_{n_2} \end{pmatrix} + \begin{pmatrix} c_1J_{n_1 \times n_1} & c_3J_{n_1 \times n_2} \\ c_3J_{n_2 \times n_1} & c_2J_{n_2 \times n_2} \end{pmatrix}$$
where
\[ b_i = \frac{1}{1 - \rho_{i,i}}, \quad i = 1, 2 \]

and
\[ c_1 = \frac{\rho_{1,1}(\rho_{2,2}(n_2 - 1) + 1) - \rho_{1,2}^2 n_2}{(\rho_{1,1} - 1)([\rho_{1,1}(n_1 - 1) + 1][\rho_{2,2}(n_2 - 1) + 1] - n_1 n_2 \rho_{1,2}^2)} \]
\[ c_2 = \frac{\rho_{2,2}(\rho_{1,1}(n_1 - 1) + 1) - \rho_{1,2}^2 n_1}{(\rho_{2,2} - 1)([\rho_{1,1}(n_1 - 1) + 1][\rho_{2,2}(n_2 - 1) + 1] - n_1 n_2 \rho_{1,2}^2)} \]
\[ c_3 = \frac{\rho_{1,2}}{n_1 n_2 \rho_{1,2}^2 - (\rho_{1,1}(n_1 - 1) + 1)(\rho_{2,2}(n_2 - 1) + 1)} \]

ii. the determinant is given by
\[ \det(R) = (1 - \rho_{1,1})^{n_1 - 1}(1 - \rho_{2,2})^{n_2 - 1}[(1 + [n_1 - 1] \rho_{1,1})(1 + [n_2 - 1] \rho_{2,2}) - \rho_{1,2}^2 n_1 n_2] \]

iii. \( R \) is positive definite if and only if
\[ \rho_i \in \left( \frac{-1}{n_i - 1}, 1 \right), \quad i = 1, 2 \]

and
\[ \rho_{1,2} \in \left( -\sqrt{\frac{(\rho_{1,1}(n_1 - 1) + 1)(\rho_{2,2}(n_2 - 1) + 1)}{n_1 n_2}}, \sqrt{\frac{(\rho_{1,1}(n_1 - 1) + 1)(\rho_{2,2}(n_2 - 1) + 1)}{n_1 n_2}} \right). \]

Proof: Part (i) can be shown by multiplying \( R \) and \( R^{-1} \). The condition \( RR^{-1} = I \) produces a system of five equations in five unknowns, and the result follows as the solution to this system. Part (ii) follows from Graybill (1983), Theorem 8.2.1. Part (iii) is equivalent to the statement that all eigenvalues of \( R \) are positive, which is equivalent to \( R \) being positive.
definite. Q.E.D.

With this result in hand, the likelihood function of a two-Block DECO model can, as in the simple equicorrelation case, be written to avoid costly inverse and determinant calculations.

\[
L = -\frac{1}{2} \sum_{t} \left( \log |R_t| + r_t' R_t^{-1} r_t \right)
\]

\[
= -\frac{1}{2} \sum_{t} \left[ \log \left( (1 - \rho_{1,1,t})^{n_1-1}(1 - \rho_{2,2,t})^{n_2-1} \times [(1 + [n_1 - 1]\rho_{1,1,t})(1 + [n_2 - 1]\rho_{2,2,t}) - \rho_{1,2,t}^2 n_1 n_2] \right) + r_t' \left( \begin{bmatrix} b_1 I_{n_1} & 0 \\ 0 & b_2 I_{n_2} \end{bmatrix} + \begin{bmatrix} c_1 J_{n_1 \times n_1} & c_3 J_{n_1 \times n_2} \\ c_3 J_{n_2 \times n_1} & c_2 J_{n_2 \times n_2} \end{bmatrix} \right) r_t \right] \]

\[
= \log \left( (1 - \rho_{1,1,t})^{n_1-1}(1 - \rho_{2,2,t})^{n_2-1} \times [(1 + [n_1 - 1]\rho_{1,1,t})(1 + [n_2 - 1]\rho_{2,2,t}) - \rho_{1,2,t}^2 n_1 n_2] \right) + \left( b_1 \sum_{i}^{n_1} r_{i,1}^2 + b_2 \sum_{i}^{n_2} r_{i,2}^2 + c_1, t (\sum_{i}^{n_1} r_{i,1})^2 + 2c_3, t (\sum_{i}^{n_1} r_{i,1})(\sum_{i}^{n_2} r_{i,2}) + c_2, t (\sum_{i}^{n_2} r_{i,2})^2 \right)
\]

In the multi-block case, the above two-block log likelihood is calculated for each pair of blocks, and then these submodel likelihoods are summed, forming the objective function to be maximized.

### 2.3 Correlation Monte Carlos

#### 2.3.1 Equicorrelated Processes

This section presents results from a series of Monte Carlo experiments that allow us to evaluate the performance of the DECO framework when the true data generating process is
known. We begin by exploring the model’s estimation ability when DECO is the generating process. Asset return data for 10, 30 or 100 assets are simulated over 1,000 or 5,000 periods according to Equations 2.6-2.9. We also consider a range of values for $\alpha$ and $\beta$. For each simulated data set, we estimate DECO and composite likelihood DCC. Here and throughout, we use a subset of $n$ randomly chosen pairs of assets to form the composite likelihood in order to speed up computation. In unreported results, we run a subset of our simulations estimating composite likelihood with all $n(n - 1)/2$ pairs, and results were virtually indistinguishable. Engle, Shephard and Sheppard (2008) find that the loss from using a subset of $n$ pairs is negligible.

Simulations are repeated 2,500 times and summary statistics for the maximum likelihood parameter estimates are calculated. Table 1 reports the mean, median and standard deviation of $\alpha$ and $\beta$ estimates, their average QML asymptotic standard errors (calculated using the “sandwich” covariance estimator of Bollerslev and Wooldridge 1992), and the root mean squared error (RMSE) for the true versus fitted average pairwise correlation process. Both models use correlation targeting, thus the intercept matrix is the same for both models and not reported.

The results show that across parameter values, cross section sizes, and sample lengths, DECO outperforms unrestricted DCC in terms of both accuracy and efficiency. Depending on simulation parameters, DECO is between two to ten times more accurate than DCC at matching the simulated average correlation path, as measured by RMSE. In small samples, DCC can fare particularly poorly. For example, when $T = 1000$, $n = 10$ and the $\beta = 0.97$, DCC’s mean $\beta$ estimate is 0.55, versus 0.93 for DECO. Also, DCC’s QML standard errors can grossly underestimate the true variability of its estimates for all sample sizes. The simulations also show that when the data generating process is equicorrelation, increasing the number of assets in the cross section improves estimates.

What explains the poor performance of DCC? Some characteristics of the DECO like-
lihood are lacking in DCC, as discussed in Section 2.2. DCC updates pairwise correlations using pairwise data histories, rather than using the data history of all series as in DECO. Also, the partial information nature of composite likelihood DCC makes it further difficult for DCC to estimate the parameters of a DECO process.

### 2.3.2 Non-Equicorrelated Processes

Proposition 2.2.2 highlights DECO’s ability to consistently estimate DCC parameters despite violation of equicorrelation. To demonstrate the performance of DECO in this light, we simulate series using DCC as the data generating process (Equations 2.6 and 2.7). Thus, while equicorrelation is violated, the average pairwise correlation behaves according to DECO and the assumptions of Proposition 2.2.1 are satisfied. In the correlation evolution we use an intercept matrix that is non-equicorrelated; the standard deviation of off-diagonal elements is 0.33, demonstrating that the differences in pairwise correlations for the simulated cross sections are substantial.

Again, we generate return data for 10, 30 or 100 assets over 1,000 or 5,000 periods using a range of values for \(\alpha\) and \(\beta\). Next, we estimate both DECO and composite likelihood DCC. Table 2 reports summary statistics.

DECO exhibits a downward bias in its \(\beta\) estimates that is exacerbated at low values of \(T/N\). For large \(T/N\), the \(\beta\) bias nearly disappears. Composite likelihood performs comparatively well, though the difference in accuracy versus DECO is almost indistinguishable when \(T\) is 5,000. The superior performance of DCC is perhaps most clearly seen in its excellent precision. In all cases, the variability of DCC estimates are a fraction of DECO’s.

It appears that DECO’s performance under misspecification (Table 2) is overall better than DCC’s performance under misspecification (Table 1). Small samples generally result in a downward bias in DECO estimates, but these estimates always manage to stay within one standard error of the estimates achieved by the correctly specified model. This is in
contrast to the severe downward biases displayed by DCC in Table 1. Similarly, DECO’s QML standard errors, while understated by an order of magnitude of roughly two, are only mildly biased compared to the performance of DCC’s standard errors in Table 1.

2.4 Empirical Analysis

2.4.1 Data

Since DECO is motivated primarily as a means of estimating dynamic covariances for large systems, our sample includes constituents of the S&P 500 Index. A stock is included if it was traded over the full horizon 1995-2008 and was a member of the index at some point during that time. This amounts to 466 stocks. Data on returns and SIC codes (which will be used for block assignments) come from the CRSP daily file. In our Factor ARCH regressions and Block DECO estimation we use Fama-French three-factor return data and industry assignments (based on SICs) from Ken French’s website. Precise definitions of portfolios can be found there.

We also compare average correlations for (Block) DECO and DCC to option implied correlations. For this analysis we use a 36 stock subset of the S&P sample that were continuously traded over 1995-2008 and were members of the Dow Jones Industrials at some point in that period. We also use daily option implied volatilities on these constituents and the index from October 1997 through September 2008 from the standardized options file of OptionMetrics.

Before proceeding to the results, we include a brief aside regarding estimation that will be important for the information criterion comparisons we make throughout. All second-stage correlation models that we estimate have the same number of parameters: an \( \alpha \) estimate, a \( \beta \) estimate, and \( n(n-1)/2 \) unique elements of the intercept matrix. Each factor structure, however, has a different number of parameters. Residual GARCH models contain
a total of $5n$ parameters. In addition, the loadings in a $K$-factor model (including a constant as one of the $K$ factors) contribute an additional $nK$ parameters. Also, the likelihoods from different composite likelihood methods are not directly comparable because they use sub-models of differing dimensions. Therefore, we use composite likelihood fitted parameters to evaluate the full joint Gaussian likelihood after the fact, which is directly comparable to the DECO likelihood.

### 2.4.2 Dynamic Equicorrelation in the S&P 500, 1995-2008

Our appraisal of DECO has thus far relied on simulated data, now we assess DECO estimates for the S&P 500 sample. As discussed in the section on model estimation, we use a consistent two-step procedure to estimate correlations. In the first stage we regress individual stock returns on a constant and specify residuals to be asymmetric GARCH(1,1) processes with Student-$t$ innovations (Glosten, Jagannathan and Runkle 1993). GARCH regressions are estimated stock-by-stock via maximum likelihood, and then volatility-standardized residuals are given as inputs to the second-stage DECO model. Here and throughout, second-stage models are estimated using correlation targeting for the intercept matrix $\bar{Q}$. The first column of Panel A in Table 3a shows estimates for the basic DECO specification, their standard errors, and the Akaike information criterion (AIC) for the full two-stage log likelihood.

We find $\hat{\alpha} = .021$ and $\hat{\beta} = .979$, thus the DECO parameters are in the range of typical estimates from GARCH models. To three decimals places, $\hat{\alpha}$ and $\hat{\beta}$ sum to one, indicating that the equicorrelation is nearly integrated. Figure 2a plots the fitted S&P DECO series against the price level of the S&P 500 Index. The clearest feature of the plot is the tendency for the average correlation to rise when the market is decreasing and fall when the market is increasing. This inverse relationship between market value and correlations has been documented previously in the literature. Longin and Solnik (1995, 2001) find that correlations between country level indices are higher during bear markets and in volatile
periods. Ang and Chen (2002) find the same result for correlations between portfolios of US stocks and the aggregate market. Our results show that, over the past 15 years, correlations reached their highest level during the global crisis in the last four months of 2008, when the average correlation between S&P 500 stocks reached nearly 60%.

2.4.3 Factor ARCH DECO

As discussed in the introduction, DECO may be used to model residuals from a factor model of returns. The log likelihood of a factor model decomposes additively since $\log f_{r,t}(r_t) = \log f_{r,t}(r_t|Factors_t) + \log f_{Factors,t}(Factors_t)$. An additive log likelihood can be maximized by maximizing each element of the sum separately, thus the volatility and correlations of factors can be estimated separately from the volatility and correlations of residuals and estimates will be consistent.

Our next empirical result demonstrates the usefulness of DECO in capturing lingering dynamics among correlations of factor model residuals. We consider two factor structures for returns: the Sharpe-Lintner CAPM and the Fama-French (1993) three-factor model. In both cases, the first-stage models are regressions of individual stock returns on a set of factors where, as before, all factors and idiosyncrasies are modeled as asymmetric GARCH(1,1). The second-stage is estimated with the basic DECO specification. Estimation results are shown in the first column of Panels B and C in Table 3a. CAPM residual correlations are slightly less persistent than the no factor case, with $\hat{\alpha} + \hat{\beta} = .992$. Note, $\alpha$ becomes statistically insignificant. Adding the market factor substantially increases the log likelihood, even after accounting for its additional parameters (as seen by the decreased AIC versus column 1 of Panel A). CAPM residual correlations are plotted against S&P 500 Index price level in Figure 2b. On average, the residual correlation is very low, dropping to less than 2% for most of the sample from a time average of over 20% with no factor (Figure 2a). The most striking feature of this plot is the large increase in residual correlations from 1999 through
late 2001, corresponding to the rise and fall of the technology bubble. It appears that, before
and after the tech boom, the CAPM does a very good job of describing return correlations.
During the tech episode, an additional factor seems to surface. The impact of this factor on
dependence among assets is not captured by the CAPM, but is picked up by residual DECO.

Estimates for residual correlations using the Fama-French three-factor model show much
weaker dynamics among residual correlations, as persistence drops to $\hat{\alpha} + \hat{\beta} = .814$. Including
three factors further improves the AIC. Figure 2c shows that residual correlations are almost
always about 1.5% and flat. The only exceptions are brief spikes to 3% during the peak of
the tech bubble and the crisis of late 2008.

### 2.4.4 Comparing DECO and DCC Correlations

The previous subsections have explored DECO fits using no factor structure, the one-factor
CAPM, and the Fama-French three-factor model. We now examine the fits of DCC for each
of these three first-stage models to compare with DECO. DCC is estimated using composite
likelihood with sub-models that are pairs of stocks; $n$ of the possible $n(n-1)/2$ pairs are
randomly selected as sub-models, where $n = 466$ for our sample. The use of a large subset
of all pairs reduces computation time while negligibly degrading the performance of the
estimator (as suggested by Engle, Shephard and Sheppard 2008). As a check of this point,
we use all pairs for estimating composite likelihood DCC in a smaller cross section of 36
Dow Jones constituents (a subset of our S&P sample) and find that the results are nearly
identical to the results when only 36 randomly selected pairs are used.

DCC parameter estimates are shown in the third column of Table 3a. When no factor
is used, parameter estimates for DECO and DCC are similar and within two standard errors
of each other. DECO achieves a lower AIC, making it the better model according to this
criterion. Note, DECO and DCC use the same number of parameters, so the lower AIC for
DECO is due solely to its better likelihood fit. We next evaluate how much pairwise DCC
correlations deviate from the equicorrelation series of DECO. Figure 3a plots DECO against the 25th, 50th and 75th percentile of pairwise DCC correlations when the first stage model has no factor. These quartiles give a sense of the dispersion of pairwise correlations. As the figure shows, the upper and lower quartiles are almost always within 5% of the median, and the dynamic pattern of the quartiles closely track the equicorrelation. The similar correlation dynamics for pairs of stocks and for equicorrelation is consistent with DECO’s ability to achieve a superior fit.

When the first-stage model includes the CAPM market factor, DCC $\alpha$ and $\beta$ estimates again are very close to those of DECO. In this case, DCC fits the data better according to the Akaike criterion. To understand how DCC might provide a better fit, consider the DCC residual correlation quartiles shown in Figure 3b. We see first that the dispersion of correlations has increased relative to the average correlation. Residual equicorrelation is roughly 2-3% over time, while the 75th and 25th DCC percentiles are around 6% and -3% on average. Furthermore, other than during the technology bubble, there appears to be no systematic relationship between the time series pattern in equicorrelation and the pattern of pairwise correlations. This picture therefore suggests that the ability of DECO to describe residual CAPM correlations is limited, consistent with the AIC values we find.

Using the Fama-French model reinforces the notion that DCC is a more apt descriptor of factor model residuals due to the tendency for residual pairwise correlations to exhibit idiosyncratic dynamics. Table 3a, Panel C shows that DCC continues to find stronger dynamics in correlations than DECO in terms of $\alpha$ and $\beta$ estimates, and pairwise DCC correlations in Figure 3c are quite distinct from the equicorrelation in their time series behavior.

In summary, our results elucidate the conditions under which DECO can provide a good description of the data. When comovement among all pairs shows broadly similar time series dynamics, DECO fits well and outperforms DCC. Conversely, when dynamics in
pairwise correlations are dissimilar, DCC may be a more appropriate model. This analysis fails to emphasize one of the greatest strengths of DECO, its ability to describe correlations among more opaque systems, such as among pools of credit securities. In such a setting, factor models may be difficult to apply, and the attributes of DECO become most apparent. The comparison here has been used to show the strengths and weaknesses of DECO in the laboratory of equity returns, whose dependence behavior is among the best understood of all assets studied in the finance literature.

2.4.5 Block DECO

In our last description of correlations among S&P constituents we repeat the above analyses using 10-Block DECO as the correlation estimator. Stocks are assigned to blocks based on SIC codes according to the industry classification scheme for Ken French’s 10 industry portfolios. We estimate 10-Block DECO using Gaussian composite likelihood with sub-models that are pairs of blocks. Due to the low number of blocks, all $10(10 - 1)/2$ pairs of industries are used to form the Block DECO composite likelihood. In particular, when formulating the likelihood contribution of industry pair $i, j$, a total of $n_i + n_j$ stocks are used in the sub-model.

When no factors are used in the first-stage GARCH regressions, Block DECO achieves a better AIC than both DCC and DECO, and finds similar parameter estimates. To get a sense of the flexibility Block DECO adds to the correlation structure, Figure 4 plots within-industry correlations for energy, telecom and health stocks. We choose only three of the 10 sectors to keep the plot legible while illustrating the richness a block structure can add to the cross section of correlations. A few interesting patterns emerge. First, the correlation among energy stocks has slowly trended upward over the entire sample. While correlations were low for the market as a whole over 2004-2007, energy correlations remained high and continued to climb. Telecom stocks, meanwhile, had the sharpest rise in correlations in the
market downturn following the technology boom. Health stocks maintained relatively low correlations throughout the sample. All three groups, however, experience drastic increases in correlations during late 2008, at which time all groups saw their highest level of comovement.

We also estimate Block DECO on residuals from the CAPM and Fama-French model. While Block DECO achieves a better fit than DECO in these factor models, DCC maintains the superior AIC. Block DECO, like DCC, finds more persistent dynamics in correlations for Fama-French residuals than DECO.

2.4.6 Equicorrelation and Implied Correlations, Dow Jones Index

Options traded on an index and its members provide an opportunity to validate fits from correlation models against forward-looking implied correlations that are based solely on options prices. We briefly compare fitted correlations from DECO and DCC to option-implied correlations.

Since options do not exist for all members of the S&P 500, we instead examine the Dow Jones Index, for which liquid options are traded on all constituents. Our sample of options data for the Dow Jones and its members begins in October 1997 (when Dow Jones Index options were introduced) through September 2008. Implied correlation is calculated from implied volatilities of the index and its constituents as in Equation 2.1. We use implied volatilities on call options standardized to have one month to maturity, available from OptionMetrics. We also estimate DECO, 10-Block DECO, and DCC using daily returns on Dow Jones stocks from 1995-2008. The first-stage model in all cases has no factors. Estimates are reported in Table 3b. Figure 5 plots the implied correlation against the average fitted pairwise correlation of each model. All three models broadly match the time series pattern of implied correlation. DECO seems to adjust more quickly and more dramatically during periods of sharp movements in the implied series. Implied correlations are almost always higher than model-based correlations, representing the correlation risk premium documented
2.4.7 Out-of-Sample Hedging Performance

One way to evaluate the performance of DECO and DCC in an economically meaningful way is to use out-of-sample covariance forecasts to form minimum variance portfolios. A superior forecasting model should provide portfolios with lower variance than portfolios formed based on competing models. This type of comparison is motivated by the well-known mean-variance optimization setting of Markowitz (1952). Consider a collection of \( n \) stocks with expected return vector \( \mu \) and covariance matrix \( \Sigma \). Two hedge portfolios of interest are the global minimum variance (GMV) portfolio and the minimum variance portfolio subject to achieving an expected return of at least \( q \). The GMV portfolio weights are the solution to the problem

\[
\min \omega^\prime \Sigma \omega \quad \text{s.t.} \quad \omega^\prime \iota = 1.
\]

The MV portfolio is found by solving this problem subject to the additional constraint \( \omega^\prime \mu \geq q \). The expressions for optimal weights are

\[
\omega_{GMV} = \frac{1}{A} \Sigma^{-1} \iota \\
\omega_{MV} = \frac{C - qB}{AC - B^2} \Sigma^{-1} \iota + \frac{qA - B}{AC - B^2} \Sigma^{-1} \mu,
\]

where \( A = \iota^\prime \Sigma^{-1} \iota \), \( B = \iota^\prime \Sigma^{-1} \mu \) and \( C = \mu^\prime \Sigma^{-1} \mu \).

We focus on two forecasting questions. The first is motivated by Elton and Gruber (1973), who demonstrate that minimum variance portfolio choices can be improved by averaging pairwise correlations within groups. Our question extends this idea to the conditional setting, and is linked to the question of best correlation fit structure to employ with ex post averaging. Once DECO is estimated, it can be used to form out-of-sample unrestricted
pairwise correlation forecasts (as in DCC). These pairwise can then be used to form different fitted correlation structures by averaging pairwise correlation forecasts within blocks as discussed in Section 2.2.4 and outlined in Figure 1. Ultimately, DECO estimates can be used to construct correlation forecasts that are equicorrelated, block equicorrelated, or unrestricted. By varying the choice of correlation structure in our forecasts we can evaluate the portfolio choice benefit of averaging pairwise correlations in a conditional setting (while keeping the estimation structure fixed as basic DECO).

Our experiment proceeds as follows. Using daily returns of the S&P cross section for the five-year estimation window beginning in January 1995 and ending December 1999, we

1. Estimate first-stage factor volatility models for each stock
2. Use estimates of regression/volatility models to form one-step ahead volatility forecasts for each stock
3. Using de-volatized residuals from the first stage, estimate the second-stage correlation model
4. Use correlation model parameter estimates to forecast unrestricted pairwise correlations one step ahead
5. Conduct ex post averaging of pairwise correlations to achieve each of the following correlation forecast fit structures
   - Unrestricted
   - 30 industry blocks
   - 10 industry blocks
   - A single block
6. Combine correlation forecasts for each fit structure with regression/volatility model estimates and forecasts to construct the full covariance matrix forecast.

7. Plug the resulting covariance forecast into Equations 2.14 and 2.15 to find optimal portfolio weights.

- This step also requires an estimate of mean return $\mu$. We set $\mu$ equal to the historical mean since 1995, and choose $q = 10\%$ annually.

8. Record realized returns for portfolios based on forecasts.

One-step ahead forecasts and portfolio choices are made in this manner for the next 22 days. After 22 days, the second-stage model is reestimated and the new parameters are used to generate the one-day ahead forecasts for the next 22 periods and new out-of-sample portfolio returns are calculated. This is repeated until all data through December 2008 have been used. The result is a set of 2,263 out-of-sample GMV and MV portfolio returns for each model.

After completing the forecasting procedure and recording portfolio returns, we calculate the realized daily variance for each ex post correlation fit structure. A superior model will produce optimal portfolios with lower variance realizations. We can test the significance of differences between portfolio variances for different correlation fit structures with a Diebold-Mariano test between the vectors of squared returns for each method.

Out-of-sample GMV portfolio standard deviations when the first-stage has no factors model are reported in the first column of Table 4. A 30-block fit structure generates the lowest variance GMV portfolio with a standard deviation of 0.0093. This improves over the next best ex post structure, which used 10 blocks and achieves a standard deviation of 0.0096. The difference is significant at the 0.1\% one-sided significance level. The same result is found for MV portfolios.
We repeat the hedging experiment using the CAPM and Fama-French factor structures. When the CAPM is used, ex post averaging over 10 blocks takes over the lowest variance position, significantly outperforming the second best (unrestricted) structure at the 2.5% one-sided significance level. The 10 block fit structure also achieves the lowest variance MV portfolio, though its improvement is not statistically significant over the next best model. For the Fama-French model, the unrestricted fit becomes the superior structure and significantly outperforms (block) equicorrelated structures.

DECO’s minimum variance portfolio results so far suggest that there can be hedging benefits to varying the block structure of correlations after estimation. These results, however, do not speak to the estimation ability of different correlation models. If the same exercises were repeated using different correlation models than DECO to estimate $\alpha$ and $\beta$, how would portfolio choices fare?

Table 5 reports the standard deviations of GMV and MV portfolios when 10-Block DECO and DCC are used for estimation. For ease of reference and to conduct a new set of hypothesis tests, we reproduce DECO’s results from Table 4. The table has a column for each first-stage factor model. Within a column, we report portfolio standard deviations achieved across different estimation and fit structures for correlations. Then, within columns (that is, for a given factor structure), Diebold-Mariano tests are performed to compare the variances achieved by DECO and Block DECO against the base case of DCC, holding the correlation fit structure fixed. For instance, the variance of DECO with an unrestricted fit structure is tested against DCC with an unrestricted structure, and 10-Block DECO with a 30-block fit is compared against DCC with a 30-block fit. In this way we can test whether (Block) DECO chooses significantly superior portfolios, holding the correlation fit structures fixed.

When the first-stage model includes no factors, the best overall GMV portfolio performance is achieved by estimating DECO, then using parameter estimates to construct
30-block correlation matrices ex post. This superior performance is statistically significant compared to DCC with 30-block fit structure. Block DECO also manages to significantly outperform DCC when the ex post structure has 30 blocks or is unrestricted. When a market factor or Fama-French factors are included, DECO continues to be the best estimation model, though its outperformance loses significance in the Fama-French case. The results are essentially the same for MV portfolios, with the exception that block-DECO becomes the best estimator for the first-stage model without a factor.

What qualitative assessments can we make from this analysis? First of all, the results highlight that tailoring the ex post block structure, regardless of the correlation estimator, can provide substantial improvement in hedging performance. Our evidence in a conditional setting corroborates findings of a long strand of literature on unconditional correlations and portfolio choice. Second, it appears that (Block) DECO, besides offering a relatively good in-sample fit of the data as shown in Table 3a, provides statistically superior out-of-sample correlation forecasts compared with DCC. We interpret this as evidence that DECO can be a valuable way of dealing with noisy pairwise correlations during estimation.

2.5 Conclusion

DECO represents a major simplification in modeling time varying conditional covariance matrices for returns of an arbitrary number of assets. The equicorrelation assumption can be used to reduce noise and improve portfolio selection procedures, and it is a simplifying assumption that arises naturally in a variety of financial contexts, such as valuation of derivative securities ranging from CDO’s to correlation swaps. We extend the model to accommodate equicorrelated blocks which can be used to ease the restrictiveness of DECO while maintaining its simplicity and robustness. DECO and Block DECO are valuable models in the presence of non-equicorrelated variables. We prove quasi-maximum likelihood
results that ensure (Block) DECO is a consistent estimator even when the true correlation process follows DCC. Analytical complexity for Block DECO can be overcome by using two-block sub-models and appealing to the method of composite likelihood.

The theoretical properties of DECO are confirmed in an experiment estimating simulated systems. When an equicorrelated generating process is used, DECO is superior in performance to DCC at all parameter values and sample sizes, as suggested by maximum likelihood theory. In the case of non-equicorrelated return simulations, DECO is consistent, but can suffer a downward bias in small samples. When the ratio of sample length to cross section size is large, this bias disappears. Finally, DECO is estimated with data for constituents of the S&P 500 Index. Estimates show that when pairwise correlations share a large common component so that correlations possess broadly the same time series pattern, DECO provides the best fit in the sense of information criteria. A test of forecasting performance shows that DECO can be used to construct out-of-sample hedge portfolios with significantly lower variance than those based on DCC.
Bibliography


**Figure 1. Procedure for Generating Fitted Correlation Structures.**
The schematic diagram summarizes how a correlation structure used as part of the maximum likelihood estimation procedure, what we call the “estimation structure,” can differ from the “fit structure” of the correlation series eventually generated from the estimated model.

![Diagram](image)

**Step 1**
Choose Estimation Structure: $ES = DECO$, Block DECO, or DCC

**Step 2**
Find $a^{ES} \beta^{ES}$ Estimates

**Step 3**
Fit $R^{DCC}$ route
Process by Plugging
$a^{ES} \beta^{ES}$ into Equation 6

**Step 4**
Average $R^{DCC}$
Within Blocks to Achieve Fitted Correlation Structure

---

**Figure 2. Fitted Dynamic Equicorrelation by Factor Model, S&P 500 Constituents, 1995-2008.**
The figure shows fitted residual equicorrelations of S&P 500 constituents estimated with the DECO model (black line) and the S&P 500 Index level (grey area). Equicorrelation fits are based on model estimates in the first column of Table 3. The graphs correspond to the following factor schemes: a) no factor, b) the Sharpe-Lintner CAPM and c) the Fama-French (1993) three-factor model.

*Figure 2a. S&P 500 Index Level vs. DECO with First-Stage Model with No Factor*
Figure 2. Continued.

Figure 2b. S&P 500 Index Level vs. DECO with First-Stage CAPM One-Factor Model

Figure 2c. S&P 500 Index Level vs. DECO with First-Stage Fama-French Three-Factor Model
Figure 3. Fitted DECO and DCC Correlations, S&P 500 Constituents, 1995-2008.
The figure shows fitted correlations of S&P 500 constituents estimated with DECO and DCC. Correlation fits are based on model estimates in Table 3. The graphs correspond to the following first-stage factor schemes: a) no factor, b) the Sharpe-Lintner CAPM and c) the Fama-French (1993) three-factor model. Each plot shows the fitted one-block equicorrelation and the 25th, 50th and 75th percentile of pairwise DCC correlations in each period.

Figure 3a. DECO and DCC with First-Stage Model with No Factor

Figure 3b. DECO and DCC with First-Stage CAPM One-Factor Model
Figure 3. Continued

Figure 3c. DECO and DCC with First-Stage Fama-French Three-Factor Model

Figure 4. Selected Block Equicorrelations for the S&P 500, 1995-2008.
The figure shows within-block equicorrelations of energy, telecom and health stocks (according to the 10-industry assignments on Ken French’s website) in our S&P 500 sample. Estimates are made with Block DECO composite likelihood using a first-stage model with no factor. Correlation fits correspond to model estimates in Table 3.
Figure 5. Dow Jones Index Option-Implied Correlations and Model-Based Average Correlations.  
The figure shows the option implied correlation of Dow Jones stocks and the average pairwise correlation estimated using DECO, 10-Block DECO and DCC. Model-based correlations correspond to parameter estimates shown in Table 3b. The options sample horizon covers October 1997 to September 2008, and the correlation model fits are estimated using data from January 1995 to December 2008.
Table 1. Monte Carlo with Equicorrelated Generating Process.

Using the DECO model of Equations 6-9, return data for 10, 30 or 100 assets are simulated over 1,000 or 5,000 periods using a range of values for $\alpha$ and $\beta$. Then, DECO is estimated with maximum likelihood and DCC is estimated using the (pairwise) composite likelihood of Engle, Shephard and Sheppard (2008). Simulations are repeated 2,500 times and summary statistics are calculated. The table reports the mean, median and standard deviation of $\alpha$ and $\beta$ estimates, as well as their mean quasi-maximum likelihood asymptotic standard error estimates. Asymptotic standard errors are calculated using the “sandwich” covariance estimator of Bollerslev and Wooldridge (1992). We also calculate the root mean squared error (RMSE) for the true versus fitted average pairwise correlation process and report the average RMSE over all simulations. Correlation targeting is used in all cases, thus the intercept is the same for both models and not reported.

<table>
<thead>
<tr>
<th></th>
<th>DECO</th>
<th>DCC</th>
<th>DECO</th>
<th>DCC</th>
<th>DECO</th>
<th>DCC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>RMSE</td>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>RMSE</td>
</tr>
<tr>
<td>$T=1000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n=10$</td>
<td>Mean</td>
<td>0.100</td>
<td>0.785</td>
<td>0.015</td>
<td>0.016</td>
<td>0.685</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.099</td>
<td>0.793</td>
<td>0.015</td>
<td>0.824</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.025</td>
<td>0.061</td>
<td>0.007</td>
<td>0.073</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.025</td>
<td>0.064</td>
<td>0.009</td>
<td>0.308</td>
<td>0.011</td>
</tr>
<tr>
<td>$n=30$</td>
<td>Mean</td>
<td>0.098</td>
<td>0.793</td>
<td>0.010</td>
<td>0.009</td>
<td>0.684</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.098</td>
<td>0.797</td>
<td>0.009</td>
<td>0.810</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.019</td>
<td>0.047</td>
<td>0.004</td>
<td>0.070</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.020</td>
<td>0.049</td>
<td>0.005</td>
<td>0.295</td>
<td>0.009</td>
</tr>
<tr>
<td>$n=100$</td>
<td>Mean</td>
<td>0.099</td>
<td>0.796</td>
<td>0.008</td>
<td>0.007</td>
<td>0.718</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.099</td>
<td>0.797</td>
<td>0.007</td>
<td>0.811</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.013</td>
<td>0.030</td>
<td>0.002</td>
<td>0.053</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.013</td>
<td>0.030</td>
<td>0.004</td>
<td>0.249</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Table 1. Continued.

<table>
<thead>
<tr>
<th></th>
<th>DECO</th>
<th>RMSE</th>
<th>DCC</th>
<th>RMSE</th>
<th>DECO</th>
<th>RMSE</th>
<th>DCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T=1000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n=10$</td>
<td>Mean</td>
<td>0.050</td>
<td>0.919</td>
<td>0.015</td>
<td>0.012</td>
<td>0.709</td>
<td>0.053</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.049</td>
<td>0.926</td>
<td>0.011</td>
<td>0.927</td>
<td>0.050</td>
<td>0.929</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.014</td>
<td>0.026</td>
<td>0.006</td>
<td>0.044</td>
<td>0.006</td>
<td>0.009</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.014</td>
<td>0.050</td>
<td>0.007</td>
<td>0.371</td>
<td>0.006</td>
<td>0.009</td>
<td>0.003</td>
</tr>
<tr>
<td>$n=30$</td>
<td>Mean</td>
<td>0.049</td>
<td>0.924</td>
<td>0.010</td>
<td>0.007</td>
<td>0.706</td>
<td>0.045</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.049</td>
<td>0.927</td>
<td>0.007</td>
<td>0.922</td>
<td>0.049</td>
<td>0.930</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.011</td>
<td>0.019</td>
<td>0.004</td>
<td>0.038</td>
<td>0.005</td>
<td>0.007</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.011</td>
<td>0.022</td>
<td>0.004</td>
<td>0.364</td>
<td>0.005</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td>$n=100$</td>
<td>Mean</td>
<td>0.049</td>
<td>0.928</td>
<td>0.008</td>
<td>0.005</td>
<td>0.741</td>
<td>0.042</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.049</td>
<td>0.928</td>
<td>0.005</td>
<td>0.926</td>
<td>0.049</td>
<td>0.930</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.007</td>
<td>0.012</td>
<td>0.002</td>
<td>0.026</td>
<td>0.003</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.007</td>
<td>0.013</td>
<td>0.003</td>
<td>0.344</td>
<td>0.003</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T=5000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n=10$</td>
<td>Mean</td>
<td>0.022</td>
<td>0.928</td>
<td>0.015</td>
<td>0.008</td>
<td>0.548</td>
<td>0.035</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.020</td>
<td>0.963</td>
<td>0.006</td>
<td>0.704</td>
<td>0.020</td>
<td>0.969</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.010</td>
<td>0.029</td>
<td>0.007</td>
<td>0.060</td>
<td>0.004</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.011</td>
<td>0.142</td>
<td>0.007</td>
<td>0.409</td>
<td>0.004</td>
<td>0.007</td>
<td>0.003</td>
</tr>
<tr>
<td>$n=30$</td>
<td>Mean</td>
<td>0.021</td>
<td>0.953</td>
<td>0.011</td>
<td>0.004</td>
<td>0.490</td>
<td>0.030</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.020</td>
<td>0.966</td>
<td>0.003</td>
<td>0.545</td>
<td>0.020</td>
<td>0.970</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.008</td>
<td>0.020</td>
<td>0.005</td>
<td>0.050</td>
<td>0.003</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.008</td>
<td>0.082</td>
<td>0.004</td>
<td>0.406</td>
<td>0.003</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td>$n=100$</td>
<td>Mean</td>
<td>0.020</td>
<td>0.966</td>
<td>0.008</td>
<td>0.002</td>
<td>0.486</td>
<td>0.027</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.020</td>
<td>0.968</td>
<td>0.002</td>
<td>0.547</td>
<td>0.020</td>
<td>0.970</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.005</td>
<td>0.010</td>
<td>0.002</td>
<td>0.033</td>
<td>0.002</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.005</td>
<td>0.031</td>
<td>0.002</td>
<td>0.400</td>
<td>0.002</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Table 2. Monte Carlo with Non-Equicorrelated Generating Process.

Using the DCC model of Equations 6 and 7, return data for 10, 30 or 100 assets are simulated over 1,000 or 5,000 periods using a range of values for \( \alpha \) and \( \beta \). Then, DECO is estimated with maximum likelihood and DCC is estimated using the (pairwise) composite likelihood of Engle, Shephard and Sheppard (2008). Simulations are repeated 2,500 times and summary statistics are calculated. The table reports the mean, median and standard deviation of \( \alpha \) and \( \beta \) estimates, as well as their mean quasi-maximum likelihood asymptotic standard error estimates. Asymptotic standard errors are calculated using the “sandwich” covariance estimator of Bollerslev and Wooldridge (1992). We also calculate the root mean squared error (RMSE) for the true versus fitted average pairwise correlation process and report the average RMSE over all simulations. Correlation targeting is used in all cases, thus the intercept is the same for both models and not reported.

<table>
<thead>
<tr>
<th></th>
<th>DECO</th>
<th>DCC</th>
<th>DECO</th>
<th>DCC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha )</td>
<td>( \beta )</td>
<td>RMSE</td>
<td>( \alpha )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( T=1000 )</td>
<td></td>
</tr>
<tr>
<td>( n=10 )</td>
<td>Mean</td>
<td>0.104</td>
<td>0.762</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.100</td>
<td>0.786</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.038</td>
<td>0.098</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.038</td>
<td>0.128</td>
<td>0.012</td>
</tr>
<tr>
<td>( n=30 )</td>
<td>Mean</td>
<td>0.108</td>
<td>0.754</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.101</td>
<td>0.790</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.047</td>
<td>0.123</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.049</td>
<td>0.154</td>
<td>0.007</td>
</tr>
<tr>
<td>( n=100 )</td>
<td>Mean</td>
<td>0.115</td>
<td>0.741</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.104</td>
<td>0.793</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.056</td>
<td>0.133</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.059</td>
<td>0.184</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th></th>
<th>DECO</th>
<th></th>
<th>DCC</th>
<th></th>
<th>DECO</th>
<th></th>
<th>DCC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\alpha)</td>
<td>(\beta)</td>
<td>RMSE</td>
<td>(\alpha)</td>
<td>(\beta)</td>
<td>RMSE</td>
<td>(\alpha)</td>
<td>(\beta)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T=1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=10)</td>
<td>Mean</td>
<td>0.052</td>
<td>0.905</td>
<td>0.015</td>
<td>0.051</td>
<td>0.922</td>
<td>0.009</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.050</td>
<td>0.923</td>
<td>0.051</td>
<td>0.922</td>
<td>0.050</td>
<td>0.929</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.021</td>
<td>0.042</td>
<td>0.005</td>
<td>0.008</td>
<td>0.009</td>
<td>0.014</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.022</td>
<td>0.095</td>
<td>0.006</td>
<td>0.010</td>
<td>0.008</td>
<td>0.014</td>
<td>0.003</td>
</tr>
<tr>
<td>(n=30)</td>
<td>Mean</td>
<td>0.056</td>
<td>0.889</td>
<td>0.013</td>
<td>0.054</td>
<td>0.918</td>
<td>0.007</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.051</td>
<td>0.921</td>
<td>0.054</td>
<td>0.918</td>
<td>0.050</td>
<td>0.929</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.028</td>
<td>0.057</td>
<td>0.003</td>
<td>0.004</td>
<td>0.011</td>
<td>0.017</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.030</td>
<td>0.131</td>
<td>0.004</td>
<td>0.007</td>
<td>0.011</td>
<td>0.019</td>
<td>0.002</td>
</tr>
<tr>
<td>(n=100)</td>
<td>Mean</td>
<td>0.065</td>
<td>0.877</td>
<td>0.016</td>
<td>0.054</td>
<td>0.918</td>
<td>0.007</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.055</td>
<td>0.919</td>
<td>0.054</td>
<td>0.918</td>
<td>0.051</td>
<td>0.929</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.036</td>
<td>0.073</td>
<td>0.001</td>
<td>0.002</td>
<td>0.014</td>
<td>0.022</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.041</td>
<td>0.148</td>
<td>0.003</td>
<td>0.005</td>
<td>0.014</td>
<td>0.026</td>
<td>0.001</td>
</tr>
<tr>
<td>(n=10)</td>
<td>Mean</td>
<td>0.026</td>
<td>0.888</td>
<td>0.014</td>
<td>0.023</td>
<td>0.954</td>
<td>0.011</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.021</td>
<td>0.959</td>
<td>0.022</td>
<td>0.956</td>
<td>0.020</td>
<td>0.968</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.017</td>
<td>0.093</td>
<td>0.003</td>
<td>0.009</td>
<td>0.005</td>
<td>0.011</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.019</td>
<td>0.209</td>
<td>0.004</td>
<td>0.013</td>
<td>0.005</td>
<td>0.014</td>
<td>0.002</td>
</tr>
<tr>
<td>(n=30)</td>
<td>Mean</td>
<td>0.029</td>
<td>0.878</td>
<td>0.012</td>
<td>0.027</td>
<td>0.942</td>
<td>0.009</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.023</td>
<td>0.957</td>
<td>0.027</td>
<td>0.943</td>
<td>0.020</td>
<td>0.969</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.020</td>
<td>0.065</td>
<td>0.002</td>
<td>0.006</td>
<td>0.007</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.025</td>
<td>0.220</td>
<td>0.003</td>
<td>0.010</td>
<td>0.007</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>(n=100)</td>
<td>Mean</td>
<td>0.037</td>
<td>0.860</td>
<td>0.014</td>
<td>0.027</td>
<td>0.942</td>
<td>0.010</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.027</td>
<td>0.951</td>
<td>0.027</td>
<td>0.943</td>
<td>0.021</td>
<td>0.969</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>Mean ASE</td>
<td>0.027</td>
<td>0.072</td>
<td>0.001</td>
<td>0.003</td>
<td>0.009</td>
<td>0.017</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>0.035</td>
<td>0.230</td>
<td>0.002</td>
<td>0.007</td>
<td>0.009</td>
<td>0.025</td>
<td>0.001</td>
</tr>
</tbody>
</table>
The table presents estimation results for nine dynamic covariance models. Each model is a two-stage quasi-
maximum likelihood estimator and is a combination of one of three first-stage models with one of three second-
stage models. The first-stage models are GARCH regression models imposing a factor structure for the cross
section of returns, in which the structures are no factor (Panel A), the Sharpe-Lintner one-factor CAPM (Panel B)
and the Fama-French (1993) three-factor model (Panel C). The second-stage correlation models, estimated on
standardized residuals from the first stage, are one- and 10-Block DECO and composite likelihood DCC. Below
each estimate we report quasi-maximum likelihood asymptotic standard errors in italics. Asymptotic standard
errors are calculated using the “sandwich” covariance estimator of Bollerslev and Wooldridge (1992). For each
model we report the Akaike information criterion calculated using the sum of the first- and second-stage log
likelihoods penalized for the number of parameters in both stages. The analysis is performed on the S&P 500 data
set described in Section 4.

<table>
<thead>
<tr>
<th></th>
<th>DECO</th>
<th>10-Block</th>
<th>DCC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: No Factor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.021</td>
<td>0.023</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>0.007</td>
<td>0.005</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.979</td>
<td>0.975</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>0.008</td>
<td>0.034</td>
</tr>
<tr>
<td>AIC</td>
<td>-6,802.0</td>
<td>-6,861.6</td>
<td>-6,735.0</td>
</tr>
<tr>
<td><strong>Panel B: CAPM Market Factor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.006</td>
<td>0.008</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.986</td>
<td>0.982</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>0.018</td>
<td>0.028</td>
<td>0.019</td>
</tr>
<tr>
<td>AIC</td>
<td>-6,851.4</td>
<td>-6,921.9</td>
<td>-7,182.2</td>
</tr>
<tr>
<td><strong>Panel C: Fama-French 3 Factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.010</td>
<td>0.002</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.804</td>
<td>0.998</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>0.114</td>
<td>0.001</td>
<td>0.029</td>
</tr>
<tr>
<td>AIC</td>
<td>-6,899.2</td>
<td>-6,955.2</td>
<td>-7,270.2</td>
</tr>
</tbody>
</table>

Table 3b. Full Sample Correlation Estimates for Dow Jones Constituents, 1995-2008.
This table repeats the analysis of Table 3a, Panel A (no factor) for the subsample of 36 Dow Jones constituents.

<table>
<thead>
<tr>
<th></th>
<th>DECO</th>
<th>10-Block</th>
<th>DCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.034</td>
<td>0.023</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.010</td>
<td>0.005</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.964</td>
<td>0.971</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.013</td>
<td>0.007</td>
</tr>
<tr>
<td>AIC</td>
<td>-572.7</td>
<td>-581.3</td>
<td>-577.4</td>
</tr>
</tbody>
</table>
Table 4. DECO Out-of-Sample Minimum Variance Portfolio Comparison (by Ex Post Fit Structure).
The table presents the results of an out-of-sample portfolio formation experiment to test covariance forecasting ability. The following procedure is used to create sequential, non-overlapping covariance forecasts, which are then used to form portfolios. First, the covariance model is estimated using a cross section of daily returns for the five-year estimation window beginning January 1995 and ending December 1999, and one-step ahead covariance forecasts for the next 22 days are formed. Using each day’s forecast, we construct the global minimum variance (GMV) portfolio and the minimum variance (MV) portfolio subject to a 10% required annual return (see Section 4.6) and record the return for each portfolio that day. Data from the 22-day forecast period is then added to the estimation sample and the model is re-estimated. The new estimates are used to generate covariance forecasts for the subsequent 22 days and new out-of-sample portfolio returns are calculated. This is repeated until all data through December 2008 has been used. The result is a set of 2,263 out-of-sample GMV and MV portfolio returns. The table reports standard deviations of the portfolio return time series. Three different first-stage models, which are factor structures for the cross section of returns, are considered: 1) no factor, 2) the Sharpe-Lintner CAPM and 3) the Fama-French (1993) three-factor model. The second-stage model is DECO, estimated on standardized residuals from the first stage. Portfolios are formed with four different ex post correlation fit structures based on industry assignment, as described in Section 4.6. Within each column, the correlation forecasting method that achieves the lowest standard deviation portfolio is shown in bold. A Diebold-Mariano test is calculated for the significance of the best model against the second best model in the same column. The best model is accompanied by *, **, or *** if it achieves a lower standard deviation than the next best model at the 2.5%, 1.0% or 0.1% (one-sided) level, respectively. The analysis is performed on the S&P 500 data set described in Section 4.1.

<table>
<thead>
<tr>
<th>Estimation Structure</th>
<th>Fit Structure</th>
<th>GMV No Factor</th>
<th>GMV CAPM</th>
<th>GMV FF3</th>
<th>MV No Factor</th>
<th>MV CAPM</th>
<th>MV FF3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECO</td>
<td>Unrestricted</td>
<td>0.0117</td>
<td>0.0090</td>
<td><strong>0.0073</strong></td>
<td>0.0106</td>
<td>0.0086</td>
<td><strong>0.0073</strong></td>
</tr>
<tr>
<td>30-Block</td>
<td></td>
<td><strong>0.0093</strong></td>
<td>0.0093</td>
<td>0.0091</td>
<td><strong>0.0088</strong></td>
<td>0.0090</td>
<td>0.0091</td>
</tr>
<tr>
<td>10-Block</td>
<td></td>
<td>0.0096</td>
<td><strong>0.0090</strong></td>
<td>0.0087</td>
<td>0.0090</td>
<td><strong>0.0086</strong></td>
<td>0.0086</td>
</tr>
<tr>
<td>1-Block</td>
<td></td>
<td>0.0101</td>
<td>0.0100</td>
<td>0.0091</td>
<td>0.0096</td>
<td>0.0098</td>
<td>0.0091</td>
</tr>
</tbody>
</table>
Table 5. Out-of-Sample Minimum Variance Portfolio Comparison (by Correlation Model and Factor Model).

The table presents the results of an out-of-sample portfolio formation experiment to test covariance forecasting ability. The following procedure is used to create sequential, non-overlapping covariance forecasts, which are then used to form portfolios. First, a covariance model is estimated using a cross section of daily returns for the five-year estimation window beginning January 1995 and ending December 1999, and one-step ahead covariance forecasts for the next 22 days are formed. Using each day’s forecast, we construct the global minimum variance (GMV) portfolio and the minimum variance (MV) portfolio subject to a 10% required annual return (see Section 4.6) and record the return for each portfolio that day. Data from the 22-day forecast period is then added to the estimation sample and the model is re-estimated. The new estimates are used to generate covariance forecasts for the subsequent 22 days and new out-of-sample portfolio returns are calculated. This is repeated until all data through December 2008 has been used. The result is a set of 2,263 out-of-sample GMV and MV portfolio returns. The table reports standard deviations of the portfolio return time series. Three different first-stage models, which are factor structures for the cross section of returns, are considered: 1) no factor, 2) the Sharpe-Lintner CAPM and 3) the Fama-French (1993) three-factor model. Three second-stage models, DECO, 10-Block DECO and DCC, are estimated on standardized residuals from the first stage. Portfolios are formed with four different ex post correlation fit structures based on industry assignment, as described in Section 4.6. Within each column, the correlation forecasting method that achieves the lowest standard deviation portfolio is shown in bold. A Diebold-Mariano test is calculated for the significance of difference between each structure/fit pair (for DECO and Block DECO) against the same fit structure based on DCC estimates. Values accompanied by *, **, or *** achieve a lower standard deviation than their DCC counterpart at the 2.5%, 1.0% or 0.1% (one-sided) level, respectively. The analysis is performed on the S&P 500 data set described in Section 4.1.
Table 5. Continued.

<table>
<thead>
<tr>
<th>Estimation Structure</th>
<th>Fit Structure</th>
<th>GMV</th>
<th></th>
<th></th>
<th></th>
<th>MV</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No Factor</td>
<td>CAPM</td>
<td>FF3</td>
<td>No Factor</td>
<td>CAPM</td>
<td>FF3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DECO</td>
<td>Unrestricted</td>
<td>0.0117</td>
<td>0.0090</td>
<td><strong>0.0073</strong></td>
<td>0.0106</td>
<td>0.0086</td>
<td><strong>0.0073</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30-Block</td>
<td><strong>0.0093</strong></td>
<td>0.0093</td>
<td>0.0091</td>
<td>0.0088</td>
<td>0.0090</td>
<td>0.0091</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>***</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>**</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10-Block</td>
<td>0.0096</td>
<td><strong>0.0090</strong></td>
<td>0.0087</td>
<td>0.0090</td>
<td><strong>0.0086</strong></td>
<td>0.0086</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td></td>
<td>***</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-Block</td>
<td>0.0101</td>
<td>0.0100</td>
<td>0.0091</td>
<td>0.0096</td>
<td>0.0098</td>
<td>0.0091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-Block DECO</td>
<td>Unrestricted</td>
<td>0.0112</td>
<td>0.0097</td>
<td>0.0077</td>
<td>0.0102</td>
<td>0.0091</td>
<td>0.0077</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30-Block</td>
<td>0.0094</td>
<td>0.0093</td>
<td>0.0091</td>
<td><strong>0.0088</strong></td>
<td>0.0090</td>
<td>0.0092</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10-Block</td>
<td>0.0096</td>
<td>0.0091</td>
<td>0.0087</td>
<td>0.0090</td>
<td>0.0087</td>
<td>0.0088</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-Block</td>
<td>0.0101</td>
<td>0.0101</td>
<td>0.0091</td>
<td>0.0096</td>
<td>0.0099</td>
<td>0.0091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCC</td>
<td>Unrestricted</td>
<td>0.0118</td>
<td>0.0091</td>
<td>0.0073</td>
<td>0.0105</td>
<td>0.0086</td>
<td>0.0073</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30-Block</td>
<td>0.0094</td>
<td>0.0093</td>
<td>0.0091</td>
<td>0.0088</td>
<td>0.0090</td>
<td>0.0091</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10-Block</td>
<td>0.0096</td>
<td>0.0091</td>
<td>0.0087</td>
<td>0.0089</td>
<td>0.0087</td>
<td>0.0087</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-Block</td>
<td>0.0101</td>
<td>0.0100</td>
<td>0.0092</td>
<td>0.0096</td>
<td>0.0098</td>
<td>0.0091</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3

Testing Asymmetric-Information Asset Pricing Models (with Alexander Ljungqvist)

Abstract: Modern asset pricing theory is based on the assumption that investors have heterogeneous information. We provide direct evidence of the importance of information asymmetry for asset prices and investor demands using a natural experiment. The experiment captures plausibly exogenous variation in information asymmetry on a stock-by-stock basis for a large set of U.S. companies. Consistent with predictions derived from a Grossman and Stiglitz-type model, we find that prices and uninformed investors’ demands fall as information asymmetry increases. In the cross-section, these falls are larger, the more investors are uninformed, the larger and more variable is stock turnover, the more uncertain is the asset’s payoff, and the noisier is the better-informed investors’ signal. We show that at least part of the fall in prices is due to expected returns becoming more sensitive to liquidity risk. Our results confirm that information asymmetry has a substantial effect on asset prices and imply that a primary channel linking asymmetry to prices is liquidity.

1We thank Anat Admati, Yakov Amihud, Brad Barber, Xavier Gabaix, Charles Jones, Marcin Kacperczyk, Ambrus Kecskés, and Bryan Routledge (our WFA discussant) as well as audiences at Columbia, Berkeley, and the 2009 WFA meeting for helpful suggestions. Address for correspondence: Stern School of Business, New York University, 44 West Fourth Street, New York NY 10012-1126. Kelly: Phone 212-998-0368; e-mail bkelly@stern.nyu.edu. Ljungqvist: Phone 212-998-0304; fax 212-995-4220; e-mail al75@nyu.edu.
3.1 Introduction

Theoretical asset pricing models routinely assume that investors have heterogeneous information. The goal of this paper is to establish the empirical relevance of this assumption for equilibrium asset prices and investor demands. To do so, we exploit a novel identification strategy which allows us to infer changes in the distribution of information among investors and hence to quantify the effect of information asymmetry on prices and demands. Our results suggest that information asymmetry has a substantial effect on prices and demands and that it affects assets through a liquidity channel.

Asymmetric-information asset pricing models typically rely on noisy rational expectations equilibria in which prices reveal the better-informed investors’ information only partially due to randomness in the supply of the risky asset. Random supply might reflect the presence of ‘noise traders’ whose demands are independent of information. Prominent examples of such models include Grossman and Stiglitz (1980), Hellwig (1980), Admati (1985), Wang (1993), and Easley and O’Hara (2004).

To derive empirical predictions representative of these models, Section I adapts the Grossman and Stiglitz model to show that an increase in information asymmetry leads to a fall in share price and a reduction in uninformed investors’ demand for the risky asset. Simple expressions for equilibrium prices reveal that an important channel linking information asymmetry and price is liquidity risk: Price falls because greater information asymmetry exposes uninformed investors to more liquidity risk. The model’s comparative statics relate the changes in price and demand to the fraction of investors who are better informed, the supply of the asset, uncertainty about the asset’s payoff, and the noisiness of the better-informed investors’ signal.

Testing whether and how information asymmetry affects asset pricing poses a tricky identification challenge, and so it is perhaps not surprising that despite the fundamental
nature of our research question, the existing empirical evidence is indirect. To appreciate the nature of the challenge, imagine regressing the change in investor demand on the change in a proxy for information asymmetry, such as the stock’s bid-ask spread. Interpreting the coefficient would be hard if, as is likely, there are omitted variables (such as changes in the riskiness of the company’s cash flows) which simultaneously affect bid-ask spreads and demand. Another problem is reverse causality. Spreads may increase precisely because demand is expected to fall. To overcome such simultaneity problems, we need a source of exogenous variation in the degree of information asymmetry in the asset market.

The Grossman and Stiglitz model and its descendants suggest what we should be looking for: An exogenous change in the cost of information about an asset’s future payoffs. These models demonstrate how an increase in information cost will increase information asymmetry by inducing fewer investors to purchase information. Exogeneity means, in this setting, that the change in information cost must affect prices and demands only via its effect on the asymmetry of information. The cost change would have to be independent of all other underlying determinants of asset prices and demands, particularly the asset’s future payoffs.

Equity research analysts are among the most influential information producers in financial markets. We argue that their presence or absence affects the extent of information asymmetry. Suppose investors are heterogeneous in their information costs. All those whose costs exceed the price at which an analyst sells his research will purchase it. If the analyst were to stop selling research, investors would have to fall back on alternative information sources, and for some, the cost of becoming informed would exceed the benefit. (For instance, hedge funds or mutual funds likely have relatively low-cost substitute information sources,

---

2 A large literature provides evidence that analyst research is an important channel through which information is impounded in stock prices; see, among others, Womack (1996), Barber et al. (2001), Gleason and Lee (2003), Jegadeesh et al. (2004), and Kelly and Ljungqvist (2008). Some scholars take the opposite view, regarding analysts mainly as cheerleaders for companies who produce biased research in the hope of currying favor with executives (see, for instance, James and Karceski (2006)). While ultimately an empirical question, the skeptics’ viewpoint should bias us against finding a relation between the presence of analysts and equilibrium prices and demands.
but retail investors presumably do not.) Thus, information asymmetry would increase following an analyst’s departure.

Consider the extreme case of a market in which each stock is covered by exactly one analyst. The ideal experiment would be to randomly ban analysts from researching some stocks. Because the cost of information substitutes for buyers of analyst research must (weakly) exceed the analyst’s asking price, the average cost of information would increase for the stocks that randomly lose research, and information asymmetry would increase as a result. An econometrician could then measure the effect of information asymmetry on equilibrium asset prices and investor demands.

Our empirical approach is to identify a quasi-experiment of this nature. We exploit reductions in the number of analysts covering a stock that resulted from 43 brokerage firms in the U.S. closing down their research departments between Q1, 2000 and Q1, 2008. For identification purposes, we view brokerage closures as a source of exogenous variation in the extent of analyst coverage and examine changes in prices and investor demands for affected stocks around the closure announcements.

Our identification strategy makes two assumptions. First, the coverage terminations must correlate with an increase in information asymmetry. This will be the case if information costs are heterogeneous, as we have argued, and Section II shows that standard proxies for information asymmetry do indeed change as required. Second, they must affect price and demand only through their effect on information asymmetry. In particular, they must be uninformative about the affected stocks’ future prospects. Unless a brokerage firm closed down its research department because its analysts had foreknowledge of imminent falls in some companies’ share prices (which seems unlikely), brokerage closures are plausibly exogenous. We discuss in Section II the reasons why so many brokerage firms have quit research since 2000. We also show empirically that our sample terminations are uninformative about
the future prospects of the affected companies, as required for identification.3

The list of brokerage firms that exited equity research over our sample period includes both large firms (e.g., Wells Fargo) and smaller outfits (e.g., Schwab’s Soundview Capital Markets division) and encompasses both retail-oriented firms (e.g., First Montauk) and institutional brokerage houses (e.g., Emerald Research) and brokers with either generalist (e.g., ABN Amro) or specialist coverage (e.g., Conning & Co.). In total, the 43 closures are associated with 4,429 coverage terminations. The closures were well publicized in the media and affected clients were notified directly. Other investors would learn which stocks were affected by a closure, and why, from so called termination notices or through various financial websites, such as Dow Jones’ marketwatch.com.

We test the model’s predictions in Section III. As predicted, both price and uninformed demand fall following a coverage termination, even though the termination conveys no information about the company’s future prospects. The estimated price effects are fairly large and they are quite precisely estimated. Using the market model, cumulative abnormal returns average −112 basis points on the day of an exogenous termination, increasing to −212 basis points on average by day 3, with similar results for alternative benchmarks. We find no evidence of reversal even one month later. Institutional investors—which are more likely to be better informed—increase their holdings of affected stocks by around 1.4% while retail investors—who are more likely to be uninformed—sell. We confirm these findings using an alternative and independent source of exogenous variation in analyst coverage, due not to brokerage closures but to the terrorist attacks of September 11, 2001.

Our cross-sectional tests support the comparative statics. Prices and, to a noisier extent,

---

3Focusing on exogenous coverage terminations is essential if we want to causally link price changes around coverage terminations to changes in information asymmetry and estimate economic magnitudes. In contrast to our sample of closure-related terminations, the vast majority of coverage terminations instead reflect the analyst’s private information and so are selective. As McNichols and O’Brien (1997) note, analysts usually terminate coverage to avoid having to issue a sell recommendation when they expect future performance declines. Reasons to avoid issuing a sell include not wanting to upset an investment banking client or jeopardize access to the company’s management in future.
uninformed demand experience larger declines, the more investors are uninformed, the larger and more variable is turnover, the more uncertain is the asset’s payoff, and the noisier is the signal.

Finally, using the Acharya-Pedersen (2005) model of expected returns, we show that affected stocks become more exposed to liquidity risk following exogenous coverage terminations and, as a result, expected returns increase. We interpret this finding as support for the notion that prices fall on announcement of a brokerage closure in anticipation of greater information asymmetry which in turn increases uninformed investors’ liquidity-risk exposure and therefore their required returns.

Our tests contribute to our understanding of asset pricing by providing direct evidence of the role of asymmetric information, the channel through which it operates, and the magnitude of its effects on prices and demands. While the fact that our findings support the key mechanism of asymmetric-information asset pricing models is perhaps not unexpected, such direct evidence is in fact quite rare, due to the identification problems we noted earlier. Thus, our second contribution is to provide a clean identification strategy with which to capture exogenous changes in asymmetric information. The extant literature has instead relied on a variety of proxies for information asymmetry, employed either in the cross-section or as within-firm changes. Both are potentially problematic: In the cross-section due to omitted variables and in differences because changes are unlikely to be exogenous.

Arguably the most notable proxy for information asymmetry in the empirical literature is Easley and O’Hara’s (1992) PIN measure. PIN is based on the idea that the presence of privately informed traders can be noisily inferred from order flow imbalances. Easley et al. (1996) show that PIN correlates with measures of liquidity such as bid-ask spreads, while Easley, Hvidkjaer, and O’Hara (2002) find that PIN affects expected returns. While we offer no opinion on the matter, PIN has proven controversial. Mohanram and Rajgopal (2006), for example, find that PIN is not priced if one extends Easley et al.’s (2002) sample period.
Similarly, Duarte and Young (2008) show that PIN is only priced to the extent that it proxies for illiquidity rather than information asymmetry. The advantage of our approach relative to this literature is that it exploits exogenous variation in the supply of information, thus side-stepping the need for a proxy for information asymmetry whose potential correlations with unobserved variables are unknown and contentious.

The bulk of the literature has bypassed the question of asymmetric information by instead focusing on the role of liquidity in asset pricing—perhaps because liquidity is considerably easier to measure than information asymmetry. Relevant empirical studies include Pastor and Stambaugh (2003), Amihud and Mendelson (1986), Amihud (2002), Hasbrouck and Seppi (2001), Bekker, Harvey, and Lundblad (2007), Jones (2002), and Eleswarapu (1997). Following theoretical models such as Amihud and Mendelson (1986), Acharya and Pedersen (2005), and Huang (2003), these studies treat liquidity as exogenous. Our findings suggest that liquidity varies with information asymmetry, consistent with microstructure models such as Kyle (1985) and Glosten and Milgrom (1985). Thus, one fundamental driver of asset prices appears to be information asymmetry, consistent with the models of Grossman and Stiglitz (1980), Hellwig (1980), Admati (1985), Wang (1993), and Easley and O’Hara (2004).

Our paper is also loosely related to the analyst literature. While we are interested in coverage terminations purely as a source of exogenous variation with which to identify changes in the supply of information, there is an active literature focusing on the causes and consequences of coverage terminations. For the most part, these papers study selective (as opposed to exogenous) terminations; see Khorana, Mola, and Rau (2007), Scherbina (2008), Kecskes and Womack (2009), Ellul and Panayides (2009), and Madureira and Underwood (2008). The exception is a contemporaneous paper by Hong and Kacperczyk (2010) who use a similar identification strategy to ours though for a different research question. Specifically, Hong and Kacperczyk exploit exogenous variation in analyst coverage due to mergers among
brokerage firms with overlapping coverage universes. This is similar to our focus on brokerage closures. However, while we are interested in estimating the pricing and demand effects of information asymmetry among investors, their focus is on the effect of competition among analysts on the extent of bias in analyst forecasts. Their results show that competition keeps analysts on their toes (in the sense that forecast bias increases following reductions in coverage). Combined with our results, this suggests that terminations not only lower the level of information production about a stock but also its quality. Either would increase information asymmetry. Our paper sheds light on the questions whether this is in turn priced, whether it affects investor demands, and what channel links information asymmetry to prices.

3.2 The Model

The purpose of formalizing the effect of information asymmetry on asset prices and investor demands is to guide our empirical tests. Thus, we adapt the simplest and perhaps best-known model of asset pricing with asymmetric information, due to Grossman and Stiglitz (1980). As we will see, despite its simplicity, this model has all the features we need to discipline our empirical analysis.

The set-up is as follows. There is a unit mass of investors who have identical initial wealth, \( W_0 \), and are risk-averse with CARA utility of consumption \( -e^{-\gamma C} \). Investors trade in period 1 and consume in period 2. There is a risk-free asset with gross return \( R \) and a single risky asset with aggregate supply \( X \sim N(\bar{X}, \sigma_x^2) \) and payoff \( u = \theta + \eta \). Investors know \( \theta \) and can engage in research at cost \( c > 0 \) which results in a noisy signal \( s \) about the risky asset’s payoff innovation, \( \eta \sim N(0, \sigma_\eta^2) \): \( s = \eta + v \), with \( v \sim N(0, \sigma_v^2) \). When investor \( i \) observes the signal, his information set, \( \mathcal{F}^i \), includes both \( s \) and the equilibrium price, \( P \), though

\footnote{The risk aversion coefficient is assumed equal to unity for simplicity and without loss of generality.}
price information is then redundant. When \( s \) is not observed, \( P \) contains useful conditioning information for payoff \( u \). As we will show later, the assumptions of a single risky asset and a single signal are not restrictive.

In addition to the investors, there may or may not be an analyst, working for a brokerage firm, who can also produce signal \( s \). For simplicity, we assume the analyst disseminates \( s \) (in the form of earnings forecasts, research reports, or trading recommendations) for free. This mirrors institutional practice: Investors are not charged for each analyst report they receive, so at the margin, the cost of observing the analyst’s signal is zero. Brokers recoup the cost of producing research through account fees, trading commissions, or cross-subsidies from market-making or investment banking.\(^5\)

In the analyst’s presence, \( s \) is public information, so information is symmetric. In his absence, we are in the Grossman-Stiglitz (1980) world where investors must decide whether to produce the signal privately. Grossman and Stiglitz show that in equilibrium, some fraction \( 0 < \delta < 1 \) of the investors pay to acquire the signal. Thus, in the analyst’s absence, information is asymmetric.

Investors do not observe aggregate supply \( X \). The three random variables of the model \((X, \eta, \text{and } v)\) are assumed to be independent.

### 3.2.1 Equilibrium Effects

To determine the equilibrium effects of a coverage termination, we compare prices and demands for the risky asset in the symmetric information case, which results from an analyst’s presence, to the asymmetric information case that occurs in the analyst’s absence. The following proposition summarizes the equilibrium changes when the analyst ceases to produce research on the asset.

---

\(^5\)We leave the brokerage firm’s incentive to disclose the analyst’s signal unmodeled. For models that endogenize this decision, see Admati and Pfleiderer (1986), Fishman and Hagerty (1995), or Veldkamp (2006a).
**Proposition 1**: Following a coverage termination, a strictly positive fraction \(1 - \delta\) of investors choose not to produce the signal themselves and so become uninformed, while a measure \(\delta > 0\) of investors become privately informed. The price of the risky asset falls by

\[
\Delta EP \equiv E[P_{asymm} - P_{symm}] = \frac{-\sigma_u^4\sigma_v^4\bar{X}\sigma_x^2(1 - \delta)}{R\left(\sigma_u^2 + \sigma_v^2\right)\left(\sigma_x^2\sigma_u^4 + \sigma_v^2\delta\sigma_u^2\sigma_x^2 + \sigma_u^2\delta^2 + \sigma_v^2\delta^2\right)} < 0 \tag{3.1}
\]

as uninformed investors sell. Privately informed investors increase their demand for the risky asset by the amount

\[
\Delta EID = \frac{(1 - \delta)\sigma_u^2\sigma_v^2\bar{X}\sigma_x^2}{\sigma_x^2\sigma_u^4 + \sigma_v^2\delta\sigma_u^2\sigma_x^2 + \sigma_u^2\delta^2 + \sigma_v^2\delta^2} > 0. \tag{3.2}
\]

**Proof**: See Appendix A.

### 3.2.2 Discussion

CARA utility combined with Gaussian random variables implies that investors optimize their demands by trading off mean and variance conditional on their information set. (For details, see Appendix A.) A coverage termination affects only the information set of investors who decide not to produce the signal themselves. From their perspective, a coverage termination increases the conditional payoff variance while having a mean zero effect on their expected payoff. This lowers uninformed demand and thus equilibrium price. At a lower price, privately informed investors, whose payoff beliefs are unchanged, are induced to increase their demand until the market-clearing condition is satisfied.

Why does payoff uncertainty increase from the perspective of the uninformed? In the analyst’s absence, the uninformed do not observe \(s\) and so base their demand solely on the observed price, \(P\). However, \(P\) is not fully revealing, because investors do not observe aggregate supply, \(X\). As a result, the uninformed cannot simply back out the informed...
investors’ signal from observed prices; they cannot tell whether a price change reflects a change in aggregate supply or a change in the signal. Instead, they noisily infer the signal by forming an expectation of payoff $u$ conditional on observed price $P$. Thus, a coverage termination exposes the uninformed to aggregate supply risk.

More formally, Appendix A shows that, under rational expectations, both the symmetric and asymmetric-information equilibrium prices are linear in the signal and in aggregate supply:

$$P_{\text{symm}} = \frac{\theta}{R} + \frac{\sigma_u^2}{R(\sigma_u^2 + \sigma_v^2)} s - \frac{\sigma_u^2}{R} (1 - \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}) X$$

$$P_{\text{asymm}} = c_1 + c_2 s + c_3 X$$

where $c_2 > 0$ and $c_3 < 0$ (the precise expressions can be found in Appendix A). Following Brennan and Subrahmanyam (1996) and Amihud (2002), we can interpret $X$ as trade volume and the coefficient on $X$ as the price impact of trade. Note that trades have a price impact even if information is symmetric, as long as the payoff is uncertain. This follows from risk aversion.

As the following corollary shows, the effect of a coverage termination is to increase the price impact of trade:

**Corollary 1:** Let the price impact of trade be defined as $\frac{\partial P}{\partial X}$. Then the change in price impact of trade following a coverage termination is given by the quantity

$$\Delta(\frac{\partial P}{\partial X}) = c_3 + \frac{\sigma_u^2}{R} (1 - \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2})$$

$$= - \frac{1}{R} \left( \frac{\sigma_v^2 \sigma_u^2 (\delta + \sigma_u^2 \sigma_x^2)}{\sigma_x^2 \sigma_v^4 + (\delta^2 + \delta \sigma_u^2 \sigma_x^2) \sigma_v^2 + \sigma_u^2 \delta^2} + \sigma_u^2 (1 - \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}) \right) < 0.$$
**Corollary 2:** Following a coverage termination, the asset’s expected return increases as the asset’s exposure to liquidity risk increases.

### 3.2.3 Robustness

Grossman and Stiglitz (1980) make two simplifying assumptions: There is a single risky asset, and there is a single signal $s$ which, in our setting, can be produced by both the analyst and investors. Neither assumption is restrictive.

Easley and O’Hara (2004) extend the Grossman-Stiglitz model to multiple assets and show that information risk cannot be diversified away even in large portfolios. Intuitively, the uninformed are at an informational disadvantage for every asset in their portfolio and so information risk is priced in equilibrium in the way we have modeled it.

Multiple independent signals also do not change the conclusions. Suppose there are $N$ analysts publishing public signals and $M$ other signals that may be purchased at cost $c_j > 0$. An analyst’s departure corresponds to a (weakly) positive shift in the cost schedule for the menu of possible signal combinations. In equilibrium, investors consume all the free signals and a Grossman-Stiglitz solution determines consumption of the costly signals. Any signal combination an investor consumed before the analyst’s departure is more expensive now, so fewer signals are consumed and the fraction of informed investors $\delta$ falls. In the new equilibrium, prices drop due to the decrease in the supply of information and the accompanying increase in overall payoff uncertainty following a coverage termination.

Finally, we note that it is straightforward to recast the analysis within the Kyle (1985) setup. This would leave Proposition 1, the corollaries, and the comparative statics discussed in the next section qualitatively unchanged. What we call the change in the price impact of trade in Corollary 1 would be analogous to the change in Kyle’s $\lambda$. We prefer the Grossman and Stiglitz (1980) setting because a coverage termination effectively increases the cost of information, which in their model is explicitly linked to the fraction of informed investors.
and thus to equilibrium prices and demand. In Kyle, on the other hand, the distribution of information is taken as exogenous.

### 3.2.4 Further Testable Implications

Proposition 1 contains precise expressions for the magnitudes of the price and demand changes. However, absent precise empirical counterparts for the model quantities $\delta, \bar{X}, \sigma_x^2, \sigma_u^2,$ and $\sigma_v^2$, we have to content ourselves with testing whether price and demand do change following coverage terminations, and if so, whether the changes have the predicted signs and are economically meaningful. Then, using empirical proxies for the five model quantities, we will test the following comparative statics.

First, the more investors become privately informed following a coverage termination, the fewer investors are affected by the loss of analyst information. Thus, we have:

**Implication 1:** The larger the fraction of privately informed investors among the company’s shareholders, the smaller is the negative price impact of a coverage termination and the smaller is the resulting increase in informed demand for the company’s stock:

\[
\frac{\partial \Delta EP}{\partial \delta} = \frac{2 \delta (1 - \delta) + \sigma_v^2 \sigma_x^2 \sigma_u^4 \sigma_v^4 \bar{X} \sigma_x^2}{(\sigma_x^2 \sigma_v^4 + \sigma_v^2 \delta^2 + \sigma_v^2 \delta \sigma_x^2 \sigma_u^2 + \sigma_u^2 \delta^2)^2 R} > 0
\]

\[
\frac{\partial \Delta EID}{\partial \delta} = -\sigma_v^2 \sigma_u^2 \bar{X} \sigma_x^2 (\sigma_x^2 \sigma_v^4 + \sigma_v^2 \sigma_u^2 \sigma_x^2 + \delta(\sigma_v^2 + \sigma_u^2)(2 - \delta)) \frac{(\sigma_x^2 \sigma_v^4 + \sigma_v^2 \delta^2 + \sigma_v^2 \delta \sigma_x^2 \sigma_u^2 + \sigma_u^2 \delta^2)^2}{(\sigma_x^2 \sigma_u^2 \delta^2 + \sigma_u^2 \delta \sigma_x^2 \sigma_u^2 + \sigma_u^2 \delta^2)^2} < 0.
\]

Second, corollary 1 states that a coverage termination increases the (negative) sensitivity of price to aggregate supply because the uninformed have a harder time filtering the signal from the price. Thus, stocks with larger aggregate supply experience larger price and demand changes:
Implication 2: The greater is mean aggregate supply, the larger is the negative price impact of a coverage termination and the greater is the resulting increase in informed demand for the stock:

\[
\frac{\partial \Delta EP}{\partial X} = \left( \frac{-\sigma_x^2 \sigma_v^4 \sigma_u^4 (1 - \delta)}{(\sigma_u^2 + \sigma_v^2) R \left( \sigma_x^2 \sigma_v^4 + (\delta^2 + \delta \sigma_u^2 \sigma_x^2) \sigma_u^2 \sigma_v^2 + \sigma_u^2 \delta^2 \right)} \right) < 0
\]

\[
\frac{\partial \Delta EID}{\partial X} = \left( \frac{(1 - \delta) \sigma_v^2 \sigma_u^2 \sigma_u^2 \sigma_x^2}{\sigma_x^2 \sigma_v^4 \sigma_u^2 \sigma_x^2 + \sigma_u^2 \delta^2} \right) > 0.
\]

Next, in the analyst’s absence, the uninformed infer the signal from the observed price. This inference problem is harder the more variable are the aggregate supply and payoff, which in turn increases the value of analyst research to the uninformed. Thus, we have the following implications:

Implication 3a: The more variable is aggregate supply, the larger is the negative price impact of a coverage termination and the greater is the resulting increase in informed demand for the stock:

\[
\frac{\partial \Delta EP}{\partial \sigma_x^2} = \frac{-\sigma_u^4 \sigma_v^4 \bar{X} \bar{\delta}^2 (1 - \delta)}{(\sigma_u^2 + \sigma_v^2 \delta^2 + \sigma_v^2 \delta \sigma_u^2 \sigma_x^2 + \sigma_u^2 \delta^2)^2 R} < 0
\]

\[
\frac{\partial \Delta EID}{\partial \sigma_x^2} = \frac{(1 - \delta) \sigma_v^2 \sigma_u^2 \sigma_x^2 \bar{X} \delta^2 (\sigma_u^2 + \sigma_u^2 \sigma_v^2)}{(\sigma_x^2 \sigma_v^4 + \sigma_u^2 \delta^2 + \sigma_v^2 \delta \sigma_u^2 \sigma_x^2 + \sigma_u^2 \delta^2)^2} > 0.
\]

Implication 3b: The more variable is the asset’s payoff, the larger is the negative price impact of a coverage termination and the greater is the resulting increase in informed demand for the stock:

\[
\frac{\partial \Delta EP}{\partial \sigma_u^2} = \frac{-\sigma_u^2 \sigma_v^6 \bar{X} \bar{\delta}^2 \sigma_x^2 (1 - \delta)}{R \left( \sigma_u^2 + \sigma_v^2 \delta^2 \right)^2 \left( \sigma_x^2 \sigma_v^4 + \sigma_u^2 \delta^2 + \sigma_v^2 \delta \sigma_u^2 \sigma_x^2 + \sigma_u^2 \delta^2 \sigma_u^2 \sigma_x^2 + \sigma_u^2 \delta^2 \right)^2} < 0
\]

130
\[
\frac{\partial \Delta EID}{\partial \sigma^2_u} = \frac{(1 - \delta) \sigma^4_u \bar{X} \sigma^2_x (\sigma^2_v \sigma^2_u + \delta^2)}{(\sigma^2_x \sigma^4_v + \sigma^2_v \delta^2 + \sigma^2_v \delta \sigma^2_u \sigma^2_x + \sigma^2_u \delta^2)^2} > 0.
\]

Finally, signal noise has a more complicated effect. Consider the two extreme cases. If the signal is very precise, the uninformed can filter the signal from the observed price very effectively, so losing the signal has a negligible impact on demand and prices. If the signal is so noisy as to be essentially uninformative, the informed have little informational advantage over the uninformed. As a result, a coverage termination results in negligible information asymmetry among investors and so has little impact on demand and prices. In between these two extremes, losing a noisy but informative analyst signal increases the uninformed investors’ inference problem and so leads to a price fall and a decrease in uninformed demand. Thus, we have:

**Implication 4:** The effect of signal noise on the price impact of a coverage termination is U-shaped, while its effect on the change in expected informed demand is hump-shaped:

\[
\frac{\partial \Delta EID}{\partial \sigma^2_v} = \frac{(1 - \delta) \sigma^4_u \bar{X} \sigma^2_x (\sigma^2_u \delta^2 - \sigma^2_v \sigma^4_x)}{(\sigma^2_x \sigma^4_v + \sigma^2_v \delta^2 + \sigma^2_v \delta \sigma^2_u \sigma^2_x + \sigma^2_u \delta^2)^2} > 0 \text{ iff } \frac{\sigma^2_u \delta^2}{\sigma^2_x \sigma^4_v} > 1.
\]
3.3 Identification and Sample

3.3.1 Identification Strategy

Our identification strategy is straightforward. To examine the effects of an increase in information asymmetry on prices and demands, we need a source of exogenous variation in information asymmetry. Rather than relying on observed changes in various proxies for information asymmetry—which may vary for endogenous reasons—we identify an exogenous shock to the information production about a stock. Specifically, we focus on terminations of analyst coverage that are the result of brokerage firms closing down their research departments. Identification requires that coverage terminations correlate with an increase in information asymmetry but do not otherwise correlate with price or demand.

As McNichols and O’Brien (1997) observe, coverage changes are usually endogenous. Coverage terminations, in particular, are often viewed as implicit sell recommendations (Scherbina (2008)). The resulting share price fall may hence reflect the revelation of an analyst’s negative view of a firm’s prospects rather than the effects of reduced research coverage. Similarly, an analyst may drop a stock because institutions have lost interest in it (Xu (2006)). If institutional interest correlates with price, price may fall following a termination for reasons unrelated to changes in information asymmetry.

We avoid these biases by focusing on closures of research operations, rather than selective terminations of individual stocks’ coverage. The last decade has seen substantial exit from research in the wake of adverse changes to the economics of producing research.6

The fundamental challenge for equity research is a public goods problem: Because research cannot be kept private, clients are reluctant to pay for it, and hence it is provided for free. In the words of one observer, “Equity research is a commodity and it is difficult for firms to

6See, for example, “Following Wells Fargo, Others May Exit Equities Trading”, Dow Jones, 3 August 2005.
remain profitable because research is a cost center.\textsuperscript{7}

Traditionally, brokerage firms have subsidized research with revenue from trading ("soft dollar commissions"), market-making, and investment banking. Each of these revenue streams has diminished since the early 2000s. The prolonged decline in trading volumes that accompanied the bear market of 2000-2003 along with increased competition for order flow has reduced trading commissions and income from market-making activities. Soft dollar commissions have come under attack both from the S.E.C.\textsuperscript{8} and institutional clients.\textsuperscript{9} Finally, concerns that analysts publish biased research to please investment banking clients (Michaely and Womack (1999), Dugar and Nathan (1995), Lin and McNichols (1998)) have led to new regulations, such as the 2003 Global Settlement, which have made it harder for brokerage firms to use investment banking revenue to cross-subsidize research.

Brokerage firms have responded to these adverse changes by downsizing or closing their research operations. We ignore those that downsized (because it is hard to convincingly show the absence of selection bias in firing decisions) and focus on a comprehensive set of 43 closures that were announced between 2000 and the first quarter of 2008.\textsuperscript{10} These include investment banks such as Wells Fargo Securities, retail brokerage firms such as Charles Schwab and First Montauk, institutional brokerage houses such as Emerald Research, and foreign banks such as ABN Amro and ING.

\textsuperscript{7}David Easthope, analyst with Celent, a strategy consultancy focused on financial services, quoted in Papini (2005).


\textsuperscript{9}According to the TABB Group, a consultancy, nearly 90\% of all larger institutional investors stopped or decreased use of soft dollars between 2004 and 2005 (BusinessWire, June 1, 2005).

\textsuperscript{10}Our sample excludes Bear Stearns and Lehman Brothers, both of which failed in 2008 and were integrated into J.P. Morgan and Barclays Capital, respectively. Lehman in particular is not a suitable shock for identification purposes, because Barclays, which had no U.S. equities business of its own, took over Lehman’s entire U.S. research department.
3.3.2 Sample

The 43 brokerage closures are associated with 4,429 coverage terminations, after filtering out securities with CRSP share codes > 12 (including REITs, ADRs, non-common stocks, and closed-end funds), companies without share price data in CRSP, and 51 companies due to be delisted within the next 60 days. We identify the affected stocks using the coverage table of Reuters Estimates, which lists, for each stock, the dates during which each broker and analyst in the Reuters database actively cover a stock; the I/B/E/S stop file, which has similar (albeit less reliably dated) information; and the termination notices which were sent to brokerage clients and can be retrieved from Investext.

The average brokerage closure involves 103 stocks, with a maximum of 480 stocks. There are 2,181 unique stocks in the sample, so the average company is hit by two terminations over our sample period, with a maximum of 12. The 43 brokerage firms employed 557 analysts (not counting junior analysts without coverage responsibilities), so the average closure involved 13 analysts and the average analyst covered eight stocks. Sample terminations span all six-digit GICS industries (except REITs).

The top graph in Figure 1 shows the quarterly number of brokerage closures. At least one brokerage firm exited research in 24 of the 33 quarters between Q1, 2000 and Q1, 2008, and there is no obvious evidence of the closures clustering in time.

The quarterly number of coverage terminations is shown in the bottom graph in Figure 1. Here, there is somewhat more evidence of clustering. Roughly half the terminations in the sample occurred in the first 2.5 years of the sample period, while there were virtually no terminations in the two years from Q4, 2002, a period during which many brokerage firms downsized their research operations without closing them completely (see Kelly and Ljungqvist (2008)). The largest number of terminations occurred in Q4, 2000, when four brokerage closures resulted in a total of 968 terminations.

Table I compares summary statistics for sample stocks, the CRSP universe of publicly
traded U.S. stocks, and the universe of U.S. stocks covered by at least one analyst according to I/B/E/S. Sample stocks—like I/B/E/S stocks more generally—are on average larger than CRSP stocks. They are also somewhat more liquid and volatile than the average CRSP stock. On average, 10.6 analysts covered a sample stock before a coverage termination, compared to 6.4 in the I/B/E/S database, suggesting that brokerage firms that exited research over our sample period disproportionately covered stocks with above-average analyst following. To the extent that less-covered stocks are proportionately more affected by a termination, this biases our tests against finding any effects\[11]\.

### 3.3.3 Announcement Dates

The relevant announcement date is the date investors learn that a brokerage has closed down its research department. We obtain these dates from Factiva. Both press releases and media reports are time-stamped, so we can pinpoint announcements with great precision.

It is important to note that information about a closure is freely available to all investors in a timely manner\[12]. Thus, both clients and other market participants can assess the implications, in terms of information asymmetry, for affected stocks and change their demands accordingly.

NYSE Rule 472(f)(5) and NASD Rule 2711(f)(5) require that clients are sent a termination notice every time a brokerage firm terminates coverage (including when it exits research completely). Termination notices must include the rationale for the termination, which removes any doubt, in our setting, that investors might misinterpret why coverage

---

\[11\] Presumably, the increase in information asymmetry is greatest when a stock loses coverage entirely. Unfortunately, our sample contains only seven stocks that were ‘orphaned’ as a result of a closure-induced coverage termination.

\[12\] We find that the Dow Jones marketwatch.com service, which organizes its contents by ticker, typically reports news of a coverage termination the same day. Moreover, marketwatch.com provides a synopsis of the broker’s reason for the termination, allowing market participants to infer whether a stock was terminated for endogenous or exogenous reasons. In our case, the typical synopsis would relate a termination to the closure of the research operation.
was terminated. The most common statement in our cases is simply that research has been closed down “based on our decision to exit the research business” (Wells Fargo termination notices, dated August 1, 2005).

3.3.4 Are Sample Terminations Exogenous?

Closure-related terminations constitute a suitably exogenous shock to the information environment unless brokerage firms quit research because their analysts possessed negative private signals about the stocks they covered. Press reports around the time of the closure announcements make it clear that the closures are unlikely to have been motivated by negative information about individual stocks and so are plausibly exogenous. The following three examples illustrate.

Commenting on his decision to close down IRG Research, CEO Thomas Clarke blamed regulation:

“With the brokerage industry facing some of the most far-reaching regulatory changes within the last 30 years, including S.E.C. rulings regarding the use of soft dollars and possibly the unbundling of research and trading costs, we could not see the economics working in our favor without substantial additional investment.” (Dow Jones, June 28, 2005)

Dutch bank ABN Amro closed its loss-making U.S. equities and corporate finance business, which included research, in March 2002. Among the 950 redundancies were 28 senior analysts who, along with junior team members, covered nearly 500 U.S. stocks. Board member Sergia Rial made it clear that the decision reflects competitive pressure and strategic considerations:

“We’re withdrawing from businesses in which we’re strategically ill-positioned and cannot create a sustainable profit stream, whether the market turns around or not.” (Chicago Tribune, March 26, 2002)
Regional broker George K. Baum closed its capital markets division, which included its research and investment banking operations, in October 2000, blaming a lack of profitability even during the booming late 1990s:

“Neither the retail brokerage nor equity capital markets divisions made money in the past two years.” (Knight Ridder Tribune Business News, Oct. 13, 2000)

These quotes suggest that brokerage firms quit research for strategic and financial reasons (often in the context of withdrawing from investment banking more generally) rather than to mask negative news about companies their analysts covered. Nonetheless, it is useful to test formally whether sample terminations are uninformative about the affected stocks’ future prospects.

**Exogeneity Test**

Suppose a broker terminates coverage at time $t$. If this signals private information about the stock, we should be able to predict its future performance from the fact that coverage was terminated. Testing this requires an assumption about the nature of the signal. Assume that it is negative information about $t + 1$ earnings that is not yet reflected in the consensus earnings forecast dated $t - 1$ (i.e., before other analysts knew of the coverage termination). When earnings are eventually announced, they will fall short of the $t - 1$ consensus, resulting in a negative earnings surprise. If, on the other hand, the termination is exogenous, earnings will not differ systematically from consensus.

This test is more powerful than simply checking whether terminations forecast changes in operating performance (such as return on assets, ROA). By conditioning on consensus, the test exploits differences in the analyst’s and the market’s information sets under the null that the analyst has negative private information. Earnings surprises are ideal in this context because they are based on an observable forecast: $\text{surprise}_{t+1} = EPS_{t+1} - E[EPS_{t+1} | \text{consensus info}_{t-1}]$. For other measures of operating performance, such as ROA, the market’s forecast
is unobservable and must be estimated by the econometrician, which introduces noise and decreases the test’s power.

We implement the test in a large panel of CRSP companies over the period 2000 to 2008. Mirroring the construction of the terminations sample, we filter out companies with CRSP share codes greater than 12, leaving 5,823 CRSP firms and 96,455 firm-quarters. The dependent variable is the scaled earnings surprise, defined as $(EPS_{t+1} - \text{consensus forecast}_{t-1})/(\text{book value of assets per share}_{t+1})$. Earnings and forecast data come from I/B/E/S and book value data from Compustat.

The main variables of interest are two indicators. The first identifies the 4,429 terminations caused by the 43 brokerage closures. Its coefficient should be statistically zero if closure-induced terminations are uninformative about future earnings, and negative if our identifying assumption is invalid. The second identifies 10,349 firm-quarters in which the I/B/E/S stop file records that one or more brokers terminated coverage for reasons unrelated to a brokerage closure. We will refer to these as ‘endogenous’ terminations and expect a significantly negative coefficient. We control for lagged returns and return volatility measured over the prior 12 months, the log number of brokers covering the stock, and year and fiscal-quarter effects. We also add CRSP firm fixed effects to control for omitted firm characteristics and allow for auto-correlation in quarterly earnings surprises using Baltagi and Wu’s (1999) AR(1)-estimator for unbalanced panels. This yields the following estimates:

$$\text{earnings surprise} = 0.607 \text{return} - 0.276 \text{volatility} - 0.220 \log \text{# brokers}$$

$$-0.267 \text{endogenous termination} + 0.008 \text{sample termination} + \text{year effects}$$

The firm fixed effects are jointly significant with an $F$-statistic of 4.2, and earnings surprises are significantly auto-correlated, with $\rho = 0.37$. Standard errors are shown below the coefficients.
As expected, earnings surprises are significantly more negative following what we have labelled endogenous terminations ($p = 0.022$). The effect is large economically. On average, an endogenous termination is followed by a quarterly earnings surprise that is 16.5% more negative than the sample average over this period. Closure-induced terminations, by contrast, are neither economically nor statistically related to subsequent earnings surprises ($p = 0.965$). This is consistent with closure-induced terminations being uninformative about future performance and thus exogenous.

### 3.3.5 Do Coverage Terminations Increase Information Asymmetry?

Identification also requires that coverage terminations increase information asymmetry among investors. To test whether they do, we examine changes around coverage terminations in three popular empirical proxies for information asymmetry: Bid-ask spreads, Amihud’s (2002) illiquidity measure, and Lesmond et al.’s (1999) illiquidity measure based on stale returns. We also examine changes in earnings surprises and return volatilities at future earnings releases, as these news events resolve a greater deal of uncertainty when information asymmetry is larger.

Throughout, we use difference-in-difference tests to help remove biases due to omitted variables or secular trends affecting similar companies at the same time (Ashenfelter and Card (1985)). To remove common influences, difference-in-difference tests compare the change in a variable of interest for treated firms to the contemporaneous change for a set of control firms matched to have similar characteristics but which are themselves unaffected by the treatment.

Given our focus on asset pricing, we follow Hong and Kacperczyk (2010) and match firms on the Fama-French (1993) pricing factors using the Daniel et al. (1997) algorithm.
Specifically, we choose as controls for firm $i$ five unique firms in the same size and book-to-market quintile in the month of June prior to a coverage termination, subject to the condition that control firms did not themselves experience a termination in the quarter before and after the event. In view of the evidence from Table I that firms with coverage terminations are larger and more liquid than CRSP firms in general, we also require that controls be covered by one or more analysts in the three months before the event.

**Results**

Table II, Panel A reports changes in bid-ask spreads around coverage terminations. If exogenous coverage terminations increase information asymmetry, we expect spreads to increase. For each stock, we compute the average bid-ask spread (normalized by the mid-quote) from daily data over three-, six-, or 12-month estimation windows ending 10 days before and starting 10 days after a termination announcement. Net of the change among control firms, spreads increase on average by 1.8%, 2.1%, and 1.4% over the three estimation windows, consistent with an increase in information asymmetry. For the first two windows, these changes are quite precisely estimated, with bootstrapped $p$-values of 0.010 and 0.012.\(^{13}\)

Wang (1994) predicts that the correlation between absolute return and dollar volume increases in information asymmetry. This makes Amihud’s (2002) illiquidity measure ($AIM$) a natural proxy for information asymmetry, as it is defined as the log ratio of absolute return and dollar volume. Panel B reports changes in $AIM$. Relative to control stocks, $AIM$ increases after a coverage termination for each estimation window, implying that the correlation between absolute return and volume increases as predicted. The increases are again economically meaningful. They average 1.4%, 1.8%, and 2.2% relative to the pre-termination means estimated for the three windows. They are also highly statistically significant.

\(^{13}\)By construction, terminations cluster in time by broker. This poses a problem for standard cross-sectional tests, so we bootstrap the standard errors following Politis and Romano (1994). We report $p$-values based on a block bootstrap with 10,000 replications and a block length of 100, the approximate average number of terminations per brokerage closure event.
An alternative measure of illiquidity is due to Lesmond et al. (1999). A large number of zero-return or missing-return days may indicate that a stock is illiquid. Panel C shows that net of the control-firm change, this number increases significantly following a coverage termination, by 25.4%, 10.9%, and 9.1% on average, measured over three-, six-, and 12-month windows, respectively.\footnote{Note that the number of days with zero or missing returns actually declines for both treatment and control firms (presumably because markets are generally becoming more liquid in the wake of decimalization and competition from ECNs). However, the number declines relatively faster for control firms, so that the difference-in-difference statistics show a relative increase in illiquidity.}

Panel D focuses on earnings announcements. Following coverage terminations, we expect returns to be more volatile around earnings announcements as more uncertainty is left unresolved (West (1988)). We also expect greater absolute earnings surprises, to the extent that post-termination consensus forecasts reflect coarser information sets. Consistent with these predictions, we find that log daily return volatility in the three days around earnings announcements increases by 14.3% net of the change among control firms ($p < 0.001$), while the average magnitude of absolute earnings surprises increases by 13.8% ($p < 0.001$).

The evidence in Table II is consistent with the interpretation that information asymmetry increases following coverage terminations. In conclusion, sample terminations appear to be both uninformative about firms’ future prospects and correlated with changes in information asymmetry, as required for identification.

### 3.4 The Effect of Information Asymmetry on Price and Demand

#### 3.4.1 Change in Price

To test the first part of Proposition 1, we compute cumulative abnormal returns (CARs) from the close on the day before a termination announcement to the close on the announcement.
day $[-1,0]$, one day later $[-1,+1]$, or three days later $[-1,+3]$. Abnormal returns are computed using three benchmarks: The market model, the Fama-French three-factor model, and an industry benchmark. For the first two, factor loadings are estimated in a one-year pre-event window ending 10 days before the termination date. For the industry benchmark, abnormal returns are calculated as the simple difference between event-window cumulative returns for the sample stock and its corresponding Fama-French 30-industry portfolio.\footnote{Industry portfolio assignments are based on SIC codes; industry definitions, portfolio returns, and SIC mappings are taken from Ken French’s website.}

Table III, Panel A reports the results for the overall sample. Consistent with Proposition 1, price falls following the announcement of a coverage termination. On the day of announcement, CARs average $-112$, $-78$, and $-45$ basis points for the market model, the Fama-French model, and the industry benchmark, respectively (not annualized).\footnote{The fact that CARs are smaller when we use industry benchmarks suggests that there may be within-industry information spillover effects. Specifically, if stocks A and B suffer a coverage termination, the associated demand changes appear to affect stocks C and D in the same industry as well. This is a natural consequence of higher payoff correlations among stocks in the same industry (Veldkamp (2006b)).} If we measure announcement returns over slightly longer windows through day +1 or day +3, CARs become more negative still, averaging up to $-212$ basis points. For the median sample firm, these price falls imply a fall in market value of between $4.6$ and $21.6$ million. In all cases, the point estimates are highly statistically significant using either a block bootstrap or a standard event-study test.

Figure 2 plots daily abnormal returns for trading days $-10$ through $+10$ relative to the day a brokerage firm closure was announced, for each of the three benchmarks. There is no evidence of abnormal returns before the announcements, consistent with the notion that the brokerage closures were unanticipated.\footnote{New reports confirm this. For example, when Wells Fargo closed its research department, one of the fired analysts was reported as saying, “We were told in a firm-wide announcement. I had no idea [it was coming].” (Wall Street Letter, August 5, 2005)} On the announcement day, prices fall dramatically for each benchmark, and there is no evidence of a subsequent reversal over the next ten days. Though not shown in the figure, CARs continue to be negative even after one month,
averaging $-238$, $-220$, and $-129$ basis points relative to the market model, the Fama-French model, and the industry benchmark, respectively. Each of these point estimates is significant at the 0.003 level or better. Thus, we find no evidence that the price falls are temporary.

Could confounding events be causing the price falls? If our terminations are truly exogenous, we effectively have a randomized trial and so controls for confounding events are superfluous. Analysis of one particular confounding event, namely coincident negative earnings surprises, illustrates. First, we find that there are about as few coincident negative earnings surprises (1.09%) as chance alone would predict (1.26%). When these are excluded, average CARs are essentially unchanged: On the announcement day, for example, they are only two or three basis points higher (less negative) than in the full sample, ranging from $-109$ to $-43$ basis points for the three benchmarks.

As a reality check on our identification strategy, we turn to two alternative samples. The first comes from Keefe, Bruyette & Woods and Sandler O’Neill & Partners, two brokerage firms that suffered horrendous casualties in the September 11, 2001 terrorist attacks. Over the following three weeks, these two firms announced a total of 356 coverage terminations as the names of nearly 30 analysts killed in the World Trade Center were confirmed. Unlike the voluntary brokerage closures in our main sample, these terminations are clearly involuntary, and moreover, at the level of each analyst or affected stock, they are unquestionably random. Thus, if the 9/11 sample behaves similarly to the main sample, it is unlikely that brokerage firms timed their closure announcements in a way that would spuriously lead us to find a price fall.

Panel C reports the associated CARs. For the market model and the Fama-French model, CARs average $-87$ and $-77$ basis points on the announcement day ($p < 0.001$),

\[18\] In the I/B/E/S universe, 39.72% of earnings surprises are negative over our sample period. With four earnings announcements, 252 trading days per year, and a two-day window, there is thus a $2 \cdot 4/252 \cdot 0.3972 \cdot 0.3972 = 1.26\%$ chance that a coverage termination randomly coincides with a negative earnings announcement or the day before.
consistent with the results in our main model. Relative to the industry benchmark, CARs are also negative, but except for the \([-1,+3]\) window, they are not statistically significant. We interpret these findings as support for Proposition 1. Overall, therefore, we find price falls following coverage terminations using two independent sources of identification, namely brokerage closures and 9/11.

Our second hold-out sample is designed to validate our identification strategy by contradiction. We return to brokerage closure-induced terminations but now focus on the 51 stocks—excluded from the main sample—that face imminent delisting within 60 days of a closure. For these stocks, a termination should be of little consequence because the asymmetric information increase over their remaining publicly traded lives is minimal in present value terms. Panel D reports the CARs. In contrast to the results shown in Panels A through C, CARs are generally small and positive rather than large and negative. The lack of statistical significance suggests that investors are unconcerned about a stock known to soon delist losing analyst coverage. This provides indirect support for our identification strategy.

### 3.4.2 Changes in Demand

According to Proposition 1, investors who choose to become privately informed following a coverage termination increase their holdings of the stock while those who become uninformed sell. Although we cannot identify directly who becomes privately informed, it seems plausible that retail investors are less likely to produce their own research than are institutional investors, which not only likely have a cost advantage in producing research themselves but often also maintain trading accounts with multiple brokerage firms and so face a relatively smaller loss of analyst information to begin with. We thus assume that a larger fraction of institutions than of retail investors becomes privately informed.

Unfortunately, we have no high-frequency trading data to estimate changes in institu-
tional and retail demand\textsuperscript{19} Instead, we use the quarterly CDA/Spectrum data to compute the change in the fraction of a sample company’s outstanding stock held by institutions required to file 13f reports\textsuperscript{20} Clearly, use of quarterly data will generate coarser and noisier results than the pricing results discussed in the previous section. As in Section II.E, we report difference-in-difference tests.

Panel A of Table IV shows that 13f filers as a group increase their holdings from 61.2% to 62.1% of shares outstanding following the average termination, with no contemporaneous change among control stocks. The average difference-in-differences is 0.9 percentage points, or 1.4% relative to the pre-termination average, with a bootstrapped $p$-value of 0.043. Thus, 13f institutions are unusually large net buyers following terminations. If institutions are more likely to have low-cost substitute information sources than do retail investors, this result is consistent with Proposition 1\textsuperscript{21}

For robustness, Panel B provides similar evidence after excluding coincident negative earnings surprises. Panel C focuses on the 9/11 sample. Here, institutional holdings of affected stocks increase by 3.6% following a coverage terminations, though this point estimate is not statistically significant (possibly due to the small sample). Panel D shows that institutional holdings of stocks facing imminent delisting do not increase (in fact, they fall significantly) when coverage is terminated. As in the previous section, this provides indirect support for our identification strategy.

\textsuperscript{19} Trade size is sometimes used to infer retail trades, but decimalization in January 2001 and the growth in algorithmic trading mean that small trades are no longer viewed as a good proxy for retail trades.

\textsuperscript{20} Investment companies and professional money managers with over $100 million under management are required to file quarterly 13f reports. Reports may omit holdings of fewer than 10,000 shares or $200,000 in market value.

\textsuperscript{21} By contrast, Xu (2006) finds that institutions reduce their ownership following endogenous terminations. Xu’s finding is consistent with the view that unlike our exogenous terminations, his endogenous ones are implicit sells.
3.4.3 Testing the Comparative Statics

To test the comparative statics in Implications 1 through 4, we regress \( \Delta EP \) and \( \Delta EID \) on proxies for the five parameters of the model: \( \delta \) (the fraction of informed investors), \( \bar{X} \) (mean aggregate supply), \( \sigma^2_x \) (aggregate supply uncertainty), \( \sigma^2_u \) (payoff uncertainty), and \( \sigma_v^2 \) (signal noise). We also control for whether the termination coincides with a negative earnings surprise and for unobserved brokerage-firm specific effects using fixed effects. We use Fama-French announcement-day CARs to proxy for \( \Delta EP \) and the difference-in-difference change in institutional holdings to proxy for \( \Delta EID \).

Our main proxy for \( \delta \) is the fraction of the company’s stock held by institutional investors; call it \( \phi \). We do not claim that every institution will choose to become informed (i.e., that \( \delta = \phi \)). For the proxy to work, we only require that \( \delta \) correlate positively with \( \phi \). This will be the case if institutions are more likely to become informed than retail investors, as argued earlier. We use the first two moments of the distribution of log daily share turnover to proxy for \( \bar{X} \) and \( \sigma^2_x \), respectively.\(^{23}\) The proxy for payoff uncertainty \( \sigma^2_u \) is the standard deviation of quarterly earnings per share. We parameterize signal noise \( \sigma^2_v \) as a function of the number of other analysts covering the stock and the quality of the analyst whose coverage is lost. Stocks covered by fewer analysts should have noisier signals, while high-quality analysts presumably produce more informative signals, so their coverage terminations should lead to larger price and demand changes.

Table V shows that the results generally support the comparative statics of the model. Columns 1 and 2 focus on changes in price. The adjusted \( R^2 \) of around 10% are reasonably large for cross-sectional return regressions. All coefficients have the predicted sign and all but one are statistically significant at the 5% level or better using bootstrapped standard errors.

\(^{22}\)Results are robust to using market-model or industry-adjusted CARs and alternative estimation windows.

\(^{23}\)An alternative proxy for \( X \) may be log free float. We prefer turnover because float captures the maximum possible supply while mean turnover better captures the typical supply actually observed.
to reduce the impact of clustering. We find that price falls are significantly larger in retail stocks (i.e., in stocks with smaller institutional holdings); the larger and more variable is turnover; the more volatile are earnings; and the more experienced is the analyst. The extent of coverage by other analysts and coincident negative earnings surprises are not significantly related to the magnitude of price falls. Economically, \( \partial \Delta EP/\partial \delta \) has the largest effect. A one standard deviation increase in our proxy for \( \delta \) is associated with an additional price fall of 32.5 basis points, all else equal. The smallest effect comes from \( \partial \Delta EP/\partial \sigma_u^2 \) (−5.6 basis points).

Recall that Implication 4 predicts a U-shaped relation between price changes and signal noise. When we include squares of the number of analysts covering the stock, or of the analyst’s experience, we find that neither is statistically significant. In column 2, we add an alternative proxy for signal noise, namely analyst forecast dispersion. We include both the level and square and find some support for a U-shaped relation: The coefficient for the level of forecast dispersion is negative \( (p = 0.017) \) while that for the squared term is positive \( (p = 0.049) \). However, the implied minimum is in the far right tail of the empirical distribution of forecast dispersion, so as a practical matter, in our data, \( \Delta EP \) appears to decrease in all our proxies for signal noise.

Columns 3 and 4 repeat this analysis for changes in institutional holdings. The results are considerably noisier than those for changes in price—not surprisingly, given the quarterly nature of the 13f institutional ownership data we use to measure changes in informed demand. The signs for \( \partial \Delta EID/\partial \delta \), \( \partial \Delta EID/\partial \bar{X} \), and \( \partial \Delta EID/\partial \sigma^2_u \) are all as predicted and generally statistically significant. (Recall that the expected signs of the informed-demand effects are exactly opposite to those for the price effects.) None one of our proxies for signal noise is significant. We view these results as encouraging, though given the obvious data limitations, they should be interpreted with caution.
3.4.4 Testing Corollary 2: Change in Expected Returns

Falling prices following an increase in information asymmetry suggest that investors’ expected returns have increased. Corollary 2 states that expected returns increase because affected stocks become more sensitive to liquidity risk. To test this prediction, we estimate how a stock’s exposure to systematic liquidity risk changes following a coverage termination and relative to matched controls. Our empirical specification follows Acharya and Pedersen (2005) who propose an equilibrium model describing how exposure to aggregate liquidity risk affects a firm’s expected returns. Their pricing equation is

\[ E(r_{i,t} - r_{f,t}) = E(c_{i,t}) + \lambda(\beta_{1i} + \beta_{2i} - \beta_{3i} - \beta_{4i}) \quad (3.6) \]

where \( r_{i,t} \) is stock \( i \)’s month-\( t \) return; \( r_{f,t} \) is the riskfree rate; \( c_{i,t} \) is a measure of stock \( i \)’s illiquidity; \( \lambda \) is the price of risk; and the four betas measure exposure to aggregate risks, as embodied by the co-movement between: Stock returns and the market return (\( \beta_{1} \)); stock illiquidity and aggregate illiquidity (\( \beta_{2} \)); stock returns and aggregate illiquidity (\( \beta_{3} \)); and stock illiquidity and the market return (\( \beta_{4} \)).

The Acharya-Pedersen model captures the intuition that stocks should have lower prices (i.e., higher expected returns) when i) the stock return covaries more with the market return; ii) the stock’s illiquidity covaries more with aggregate illiquidity; iii) the stock return covaries less with aggregate illiquidity; and iv) the stock’s illiquidity covaries less with the market return. The first point is the CAPM rationale. The second assumes that an investor prefers stocks that are liquid when their portfolio is illiquid, all else equal. The third and forth points also capture this idea, noting that stocks with high returns in times of illiquidity, and

\(^{24}\)In a previous draft, we estimated factor models that included a liquidity factor and, like here, found that coverage terminations are followed by increased exposure to liquidity risk. We prefer the Acharya-Pedersen approach because it can be implemented over our entire sample period, 2000-2008. By contrast, data for the standard liquidity factors are available only through 2004 or 2005 (Pastor and Stambaugh (2003)) or 2006 (Sadka (2006)).
stocks that are liquid when portfolio returns are low, should command high prices.

We estimate regression (3.6) separately for each stock $i$ using weekly data over 12-month, 18-month, and 24-month windows ending two weeks prior to the termination announcement or starting two weeks after the announcement date. We focus on the weekly frequency to alleviate potential non-synchronicity concerns while providing enough data points in reasonably short windows to obtain reliable beta estimates.\footnote{A monthly frequency provides too few data points if we want to keep the estimation windows reasonably short to avoid confounding events. A daily frequency introduces a non-synchronicity problem. A common method to adjust for non-synchronous data is to include lags of independent variables and sum the loadings over all lags (see, for example, Scholes and Williams (1977)). However, construction of the Acharya-Pedersen illiquidity measure involves an adjustment for lagged aggregate market value. Including lagged independent variables therefore produces mechanical covariation between stock illiquidity and lagged market returns, used in calculating $\beta_4$, so this non-synchronicity adjustment cannot be used. Using weekly data is therefore preferable.} Aggregating over sample firms, we obtain mean betas before and after a coverage termination. We repeat this procedure for control firms, selected as described previously, and compute difference-in-differences for each of the four betas. Equation 3.6 demonstrates that all betas are multiplied by the same price of risk. As a result, the effect of a termination on expected returns depends on the change in the total risk loading, $(\beta_1 + \beta_2 - \beta_3 - \beta_4)$. We provide a test of the mean total risk loading difference-in-difference, as well as for the individual beta changes.

The results, shown in Table VI, are consistent with Corollary 2. Regardless of estimation window, terminated stocks experience an increase in total risk loading of between 0.164 and 0.191, or 11.9% and 14.3% over their pre-termination means of 1.34 to 1.38. These point estimates are significantly different from zero using bootstrap tests to adjust for potential cross-sectional dependence due to time clustering of observations. The overall effect of these increased loadings is to increase the expected returns of stocks experiencing coverage terminations.

The increase in total risk loading is dominated by a significant drop in $\beta_4$, representing an increased tendency for sample firms to experience poor returns in times of market
illiquidity. Also, the co-movement between market illiquidity and the illiquidity of sample stocks ($\beta_2$) nearly doubles, with statistical significance. While unrelated to liquidity, we also find that post-termination returns load less strongly on the market factor (i.e., $\Delta \beta_1 < 0$). Contrary to the model’s predictions, $\Delta \beta_3$ has a sign indicating decreased liquidity risk, in some cases significantly so. This effect, however, is economically small (averaging between 0.001 and 0.003) and is swamped by the liquidity risk increases captured by $\Delta \beta_2$ and $\Delta \beta_4$.

### 3.5 Conclusion

Asset pricing models routinely assume that investors have heterogeneous information, but it is an open empirical question how important information asymmetry really is for asset prices. This paper provides direct evidence of the effect of information asymmetry on asset prices and investor demands using plausibly exogenous variation in the supply of information.

The variation we exploit is caused by the closure of 43 brokerage firms’ research operations between 2000 and the first quarter of 2008, which led to analyst coverage of thousands of stocks being terminated for reasons that we argue were unrelated to the stocks’ future prospects. Following such exogenous coverage terminations, information asymmetry increased (proxied by a range of standard measures such as bid-ask spreads), while share prices and uninformed (i.e., retail) investors’ demands fell. Consistent with the comparative statics of a Grossman and Stiglitz-type model, we find that the falls in price and uninformed demand are larger, the more investors are uninformed; the larger and more variable is turnover; the more uncertain is the asset’s payoff; and the noisier is the better-informed investors’ signal. We show theoretically that prices fall because expected returns become more sensitive to liquidity risk and provide empirical support for this prediction.

In sum, our results imply that information asymmetry has a substantial effect on asset prices and that a primary channel linking asymmetry to prices is liquidity. We believe that
our quasi-experiment can serve as a useful source of exogenous variation for empirical work in other applications that examine the effects of information asymmetry in financial markets. In this spirit, we intend to make our set of exogenous coverage terminations available to fellow researchers.
References


153


James, Christopher, and Jason Karceski, 2006, Strength of analyst coverage following IPOs, Journal of Financial Economics 82, 1-34.


Veldkamp, Laura, 2006a, Media frenzies in markets for financial information, American Economic Review 96, 577-601.


Xu, Li, 2006, Institutional investing activities and firms’ information environments before and after sell-side analyst coverage initiation and termination, Unpublished working paper, Duke University.
Figure 1. Quarterly Breakdown of Brokerage Closures and Coverage Terminations.
Our sample of coverage terminations is derived from 43 brokerage closures over the period Q1, 2000 through Q1, 2008. The top graph shows the number of brokerage closures per calendar quarter while the bottom graph shows the resulting number of coverage terminations.
Figure 2. Daily Abnormal Returns Around Closure-induced Coverage Terminations.

The sample consists of 4,429 coverage terminations for 2,181 unique firms between Q1, 2000 and Q1, 2008. We compute abnormal returns for trading days -10 through +10 relative to the day a brokerage firm closure was announced (day 0). We use three separate benchmarks: The market model, the Fama-French three-factor model, and an industry benchmark. For the market and Fama-French models, factor loadings are estimated in a one-year pre-event window ending 10 days before the termination date, and abnormal returns during the event window are calculated using this estimated model as a benchmark. (Results are nearly identical if we include a Carhart (1997) momentum factor in the Fama-French model.) For the industry benchmark, abnormal returns are calculated as the simple difference between event-window returns for the sample stock and its corresponding Fama-French 30-industry portfolio. Industry portfolio assignments are based on SIC codes; industry definitions, portfolio returns, and SIC mappings are taken from Ken French’s website.
Table I. Coverage Terminations: Summary Statistics.
The sample consists of 4,429 coverage terminations for 2,181 unique firms between Q1, 2000 and Q1, 2008. This table reports summary statistics for the market value of equity, share turnover (monthly volume divided by shares outstanding), daily return volatility, and the extent of coverage for each stock in the terminations sample; the CRSP universe (share codes 10 and 11); and the universe of U.S. stocks with analyst coverage in the I/B/E/S database. For each firm in the terminations sample, we calculate equity value and turnover in the month prior to the first termination date. For the universes of CRSP stocks and covered stocks, these variables are computed for December 2004 (the midpoint of our sample period). Annualized volatility for terminations is the standard deviation of daily log returns in the 12-month period ending one month prior to a termination, times $\sqrt{252}$. For the universes of CRSP stocks and covered stocks, volatilities are the annualized daily standard deviations for firms in these samples during calendar year 2004. (In each column, a firm is omitted from this calculation if it has fewer than 200 days of non-missing returns.) The number of brokers covering a stock in the terminations sample in the month prior to a termination is taken from the I/B/E/S forecast summary file. (Each month, I/B/E/S reports the number of outstanding EPS forecasts for the coming fiscal year.) The broker count for the universe of covered stocks represents the I/B/E/S forecast count in December 2004.

<table>
<thead>
<tr>
<th></th>
<th>Terminations sample</th>
<th>CRSP universe in 2004</th>
<th>Universe of covered stocks in 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equity market value ($m)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>7,146.7</td>
<td>2,943.3</td>
<td>6,211.0</td>
</tr>
<tr>
<td>Median</td>
<td>1,017.8</td>
<td>319.6</td>
<td>397.1</td>
</tr>
<tr>
<td>Range</td>
<td>3.5 – 512,833</td>
<td>0.7 – 385,883</td>
<td>2.9 – 385,883</td>
</tr>
<tr>
<td>Monthly turnover</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.20</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Median</td>
<td>0.12</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>Range</td>
<td>0 – 3.0</td>
<td>0 – 17.9</td>
<td>0 – 9.4</td>
</tr>
<tr>
<td>Daily return volatility (annualized %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>52.1</td>
<td>45.1</td>
<td>37.0</td>
</tr>
<tr>
<td>Median</td>
<td>60.0</td>
<td>39.1</td>
<td>41.6</td>
</tr>
<tr>
<td>Range</td>
<td>11.6 – 214.3</td>
<td>8.1 – 277.4</td>
<td>8.1 – 175.5</td>
</tr>
<tr>
<td>Number of brokers covering stock</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>10.6</td>
<td></td>
<td>6.4</td>
</tr>
<tr>
<td>Median</td>
<td>9</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Range</td>
<td>1 – 40</td>
<td></td>
<td>1 – 42</td>
</tr>
<tr>
<td>Number of unique firms</td>
<td>2,181</td>
<td>5,298</td>
<td>4,305</td>
</tr>
</tbody>
</table>
Table II. The Effect of Coverage Terminations on Information Asymmetry.
The table reports cross-sectional means of proxies for information asymmetry before and after a termination for sample stocks and their matched controls. For each sample termination, a control group is formed by selecting stocks with the same Daniel et al. (1997) size and book-to-market quintile assignment in the month of June prior to a termination, subject to the conditions that control firms a) were covered by one or more analysts in the three months before the event and b) they were not themselves subject to a coverage termination in the quarter before and after the event. When more than five matches exist, we choose the five stocks closest to the sample stock in terms of the relevant pre-event information asymmetry measure. (We lose 158 observations involving sample stocks that have no viable controls, and varying numbers of observations due to missing data necessary for calculating a particular measure. The number of observations ranges from around 4,100 to around 4,300.) For each sample stock i, we construct a difference-in-difference test as DiD = (post − pre) − (post − preControl Group). The table reports the cross-sectional mean of the DiD statistic along with the mean percentage change (DiD/mean before − 1). The latter illustrates the economic magnitude of the change in information asymmetry. We test the null hypothesis that information asymmetry is unchanged around a coverage termination using bootstrapped p-values. These adjust for potential cross-sectional dependence due to overlapping estimation windows caused by time clustering as multiple stocks are terminated in each brokerage firm closure. They are calculated using a block bootstrap with a block length of 100, the approximate average number of terminations per brokerage closure event. The first three proxies for information asymmetry are computed over three-month, six-month and 12-month windows ending 10 days prior to the termination announcement or starting 10 days after the announcement date. Panel A reports changes in bid-ask spreads. Spreads are computed as 100*(ask−bid)/(ask+bid)/2 using daily closing bid and ask data from CRSP. Panel B reports changes in the log Amihud illiquidity measure. This is defined as the natural log of the ratio of the absolute stock return to the dollar trading volume and scaled by 10^4; see Amihud (2002, p. 43). Panel C reports the changes in the number of days with zero or missing returns in CRSP (using missing return codes -66, -77, -88, and -99). In Panel D, we report the effects of terminations on quarterly earnings announcements. The first measure is the annualized daily return volatility in a three-day window around earnings announcements for all earnings announcements occurring in a one-year window before or after the drop. The second measure is the mean absolute value of quarterly earnings surprises in a one-year window before or after the drop. A surprise is defined as the absolute value of actual quarterly earnings minus the latest I/B/E/S consensus estimate before the earnings announcement, scaled by book value of equity per share, and multiplied by 100 for expositional purposes.

<table>
<thead>
<tr>
<th>Panel A: Bid-ask spreads</th>
<th>Terminations Before</th>
<th>Terminations After</th>
<th>Matched controls Before</th>
<th>Matched controls After</th>
<th>Mean DiD</th>
<th>p-value DiD = 0</th>
<th>Economic magnitude (percentage change)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-month window</td>
<td>1.123</td>
<td>1.215</td>
<td>1.085</td>
<td>1.157</td>
<td>0.020</td>
<td>0.010</td>
<td>1.8%</td>
</tr>
<tr>
<td>6-month window</td>
<td>1.117</td>
<td>1.196</td>
<td>1.083</td>
<td>1.139</td>
<td>0.023</td>
<td>0.012</td>
<td>2.1%</td>
</tr>
<tr>
<td>12-month window</td>
<td>1.116</td>
<td>1.166</td>
<td>1.083</td>
<td>1.118</td>
<td>0.015</td>
<td>0.099</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Amihud illiquidity measure (AIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-month window</td>
</tr>
<tr>
<td>6-month window</td>
</tr>
<tr>
<td>12-month window</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Missing and zero-return days (in % of days in the relevant window)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-month window</td>
</tr>
<tr>
<td>6-month window</td>
</tr>
<tr>
<td>12-month window</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D: Earnings announcements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volatility (% p.a.)</td>
</tr>
<tr>
<td>Earnings surprise</td>
</tr>
</tbody>
</table>

160
Table III. Changes in Price Around Coverage Terminations.

We compute cumulative abnormal returns over three different windows using three separate benchmarks: The market model, the Fama-French three-factor model, and an industry benchmark. For the market and Fama-French models, factor loadings are estimated in a one-year pre-event window ending 10 days before the termination date, and cumulative abnormal returns during the event window are calculated using this estimated model as a benchmark. (Results are nearly identical if we include a Carhart (1997) momentum factor in the Fama-French model.) For the industry benchmark, abnormal returns are calculated as the simple difference between event-window cumulative returns for the sample stock and its corresponding Fama-French 30-industry portfolio. (Industry portfolio assignments are based on SIC codes; industry definitions, portfolio returns, and SIC mappings are taken from Ken French’s website.) We report these abnormal return metrics for the closures sample (Panel A); the subsample of closure events that do not coincide with a negative earnings surprise on the termination day or the day before (Panel B); a separate sample of terminations at two brokerage firms located in the World Trade Center that were devastated in the 9/11 attacks (Panel C); and a separate sample of closure-induced terminations of stocks facing imminent delisting, for which a termination should be of little consequence (Panel D). To correct for after-hours announcements, we use time stamps to determine the first trading day when investors could react to a closure-induced termination. Abnormal returns are reported in percent. We report significance levels using a block bootstrap to control for dependence among events as well as the Patell $t$-test. The two test statistics are separated by “/”. In Panels A and B, we use a block length of 100 for the bootstrap, corresponding to the approximate number of terminations per brokerage closure event. In Panel C and D, given the much smaller sample sizes, we use block lengths of 10 and 5, respectively. We use ***, **, and * to denote statistical significance at the 0.1%, 1%, and 5% levels (two-sided), respectively. The sample falls short of 4,429 because we require a minimum of 50 trading days of pre-event stock prices to estimate model parameters.

<table>
<thead>
<tr>
<th>Estimation window:</th>
<th>No. of obs.</th>
<th>Market model</th>
<th>Fama-French three-factor model</th>
<th>Industry benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close on day before termination to …</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: Closure-induced terminations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… close on day of termination [-1,0]</td>
<td>4,414</td>
<td>-1.12</td>
<td><em><strong>/</strong></em></td>
<td>-0.78</td>
</tr>
<tr>
<td>… close on day +1 [-1,+1]</td>
<td>4,414</td>
<td>-1.61</td>
<td><strong>/</strong>*</td>
<td>-1.26</td>
</tr>
<tr>
<td>… close on day +3 [-1,+3]</td>
<td>4,414</td>
<td>-2.12</td>
<td><em><strong>/</strong></em></td>
<td>-1.60</td>
</tr>
<tr>
<td>Panel B: Closure-induced terminations excluding coincident negative earnings surprises</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… close on day of termination [-1,0]</td>
<td>4,366</td>
<td>-1.09</td>
<td><strong>/</strong>*</td>
<td>-0.76</td>
</tr>
<tr>
<td>… close on day +1 [-1,+1]</td>
<td>4,366</td>
<td>-1.56</td>
<td><strong>/</strong>*</td>
<td>-1.22</td>
</tr>
<tr>
<td>… close on day +3 [-1,+3]</td>
<td>4,366</td>
<td>-2.04</td>
<td><strong>/</strong>*</td>
<td>-1.53</td>
</tr>
<tr>
<td>Panel C: Terminations due to 9/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… close on day of termination [-1,0]</td>
<td>356</td>
<td>-0.87</td>
<td><strong>/</strong>*</td>
<td>-0.77</td>
</tr>
<tr>
<td>… close on day +1 [-1,+1]</td>
<td>356</td>
<td>-0.90</td>
<td><strong>/</strong>*</td>
<td>-0.65</td>
</tr>
<tr>
<td>… close on day +3 [-1,+3]</td>
<td>356</td>
<td>-0.76</td>
<td><strong>/</strong>*</td>
<td>-0.44</td>
</tr>
<tr>
<td>Panel D: Imminent delistings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… close on day of termination [-1,0]</td>
<td>51</td>
<td>0.53</td>
<td>0.76</td>
<td>0.85</td>
</tr>
<tr>
<td>… close on day +1 [-1,+1]</td>
<td>51</td>
<td>-0.32</td>
<td>-0.04</td>
<td>0.49</td>
</tr>
<tr>
<td>… close on day +3 [-1,+3]</td>
<td>51</td>
<td>0.57</td>
<td>0.71</td>
<td>1.33</td>
</tr>
</tbody>
</table>
Table IV. Changes in Institutional Holdings Around Coverage Terminations.
The table reports the quarterly change in institutional investors’ holdings of stocks that experience coverage terminations. We report the mean fraction of total stock outstanding that is held in aggregate by institutional investors filing 13f reports in the quarter before and the quarter after a termination. We then calculate a difference-in-difference test, DiD \(= (\text{post}_i - \text{pre}_i) - (\text{post}_{\text{Control Group}_i} - \text{pre}_{\text{Control Group}_i}) \), that is, the difference between the pre- and post-termination change for sample stock \(i\) less the average change for control stocks. We also report percentage changes (DiD/mean before – 1). Control groups are formed as described in Table II. We report these statistics for the closures sample (Panel A); the subsample of closure events that do not coincide with a negative earnings surprise on the termination day or the day before (Panel B); a separate sample of terminations at two brokerage firms located in the World Trade Center that were devastated in the 9/11 attacks (Panel C); and a separate sample of closure-induced terminations of stocks facing imminent delisting, for which a termination should be of little consequence (Panel D). The 13f data are taken from Thomson Reuters’ CDA/Spectrum database. In Panels A and B, we lose 158 observations involving sample stocks that have no viable controls and seven observations with missing 13f data. In Panel C we similarly lose four observations without DGTW matches and one observation without 13f data. In Panel D, we lose five observations without DGTW matches. In addition, because these are stocks that are about to be delisted, we lose 30 observations for which no 13f data are available in the quarter after the termination. We report significance levels using a block bootstrap to control for dependence among events. In Panels A and B, we use a block length of 100 for the bootstrap, corresponding to the approximate number of terminations per brokerage closure event. In Panel C, given the much smaller sample sizes, we use a block length of 10. In Panel D, where we have only 17 observations due to the fact that most companies in this sample are delisted before the next 13f filing date, we compute a simple \(t\)-test. We use *, **, and *** to denote statistical significance at the 0.1%, 1%, and 5% levels (two-sided), respectively.

<table>
<thead>
<tr>
<th>Panel</th>
<th>Terminations (in %)</th>
<th>Matched controls</th>
<th>Economic magnitude (percentage change)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of obs.</td>
<td>Before drop</td>
<td>After drop</td>
</tr>
<tr>
<td>Panel A: Closure-induced terminations</td>
<td>4,264</td>
<td>61.2</td>
<td>62.1</td>
</tr>
<tr>
<td>Panel B: Closure-induced terminations excluding coincident negative earnings surprises</td>
<td>4,216</td>
<td>61.3</td>
<td>62.2</td>
</tr>
<tr>
<td>Panel C: Terminations due to 9/11</td>
<td>351</td>
<td>37.8</td>
<td>39.5</td>
</tr>
<tr>
<td>Panel D: Imminent delistings</td>
<td>17</td>
<td>45.4</td>
<td>15.0</td>
</tr>
</tbody>
</table>
Table V. Cross-sectional Determinants of Changes in Price and in Informed Demand.

We test the cross-sectional predictions of the model presented in Section I by regressing proxies for changes in price and in informed demand around coverage terminations on proxies for \( \delta \) (the fraction of informed investors); \( \bar{X} \) (mean aggregate supply); \( \sigma_{X}^2 \) (aggregate supply uncertainty); \( \sigma_{\delta}^2 \) (payoff uncertainty); and \( \sigma_{v}^2 \) (signal noise). The proxy for changes in price is the Fama-French-adjusted cumulative abnormal return from the day before the announcement of a brokerage-firm closure to the end of the announcement day; see Table III. The proxy for the change in informed demand is the difference-in-difference (DiD) change in the fraction of the company’s stock held by institutional investors from the quarter before to the quarter after a coverage termination, net of the mean contemporaneous change in institutional holdings in matched control firms; see Table IV for details of the construction. The independent covariates are defined as follows. The proxy for \( \delta \) is the fraction of the company’s stock held by institutional investors as of the quarter-end prior to the termination, estimated from 13f data. Mean and variance of aggregate supply are based on the first two moments of the distribution of log daily turnover, estimated over the six months ending one month prior to the termination. Our proxy for payoff variance is the standard deviation of quarterly earnings per share, using up to 20 quarters of data prior to the termination. We parameterize signal noise as a function of the number of remaining analysts covering the stock (estimated from the I/B/E/S forecast summary file); the experience of the analyst whose coverage is lost (estimated as the log number of quarters since the analyst first appeared in the I/B/E/S databases or, if missing, when he or she obtained their first license according to FINRA’s “Broker Check” service); and (in columns 2 and 4) the level and square of analyst forecast dispersion (defined as the time-series mean of the standard deviation of analyst EPS forecasts in the year prior to the termination). Where necessary, variables are winsorized to reduce noise and avoid outliers driving the results. We control for whether the termination coincides with a negative earnings surprise and for unobserved brokerage-firm specific effects using fixed effects. Because each brokerage-firm closes down only once, we cannot also include time-specific effects. We also cannot include any brokerage-firm characteristics, such as its size, because such variables would be perfectly collinear with the brokerage-firm fixed effects. We report significance levels using a bootstrap with 1,000 replications to control for dependence among events. Bootstrapped standard errors are reported in italics beneath the coefficient estimates. We use ***, **, and * to denote significance at the 0.1%, 1%, 5%, and 10% levels (two-sided), respectively.

<table>
<thead>
<tr>
<th></th>
<th>Fama-French CAR [-1,0], %</th>
<th>DiD change in 13f holdings</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta )</td>
<td>1.325***</td>
<td>-0.058***</td>
</tr>
<tr>
<td>( \bar{X} )</td>
<td>-0.640***</td>
<td>0.003</td>
</tr>
<tr>
<td>( \sigma_{\delta}^2 )</td>
<td>-1.160***</td>
<td>-0.007</td>
</tr>
<tr>
<td>( \sigma_{\delta}^2 )</td>
<td>-0.433**</td>
<td>0.047***</td>
</tr>
<tr>
<td>( \sigma_{v}^2 )</td>
<td>-0.026</td>
<td>0.001</td>
</tr>
<tr>
<td>( \sigma_{v}^2 )</td>
<td>-0.226†</td>
<td>0.000</td>
</tr>
<tr>
<td>( \sigma_{v}^2 )</td>
<td>-1.928†</td>
<td>-0.005</td>
</tr>
<tr>
<td>( \sigma_{v}^2 )</td>
<td>0.733†</td>
<td>-0.008</td>
</tr>
<tr>
<td>( =1 ) if coincides w/ neg. earnings surprise</td>
<td>-1.021</td>
<td>0.004</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>9.8 %</td>
<td>3.7 %</td>
</tr>
<tr>
<td>Wald test: all coef. = 0</td>
<td>14.7***</td>
<td>14.7***</td>
</tr>
<tr>
<td>No. of observations</td>
<td>4,358</td>
<td>4,321</td>
</tr>
</tbody>
</table>
Table VI. The Effect of Coverage Terminations on Liquidity Risk Exposure.
The table reports cross-sectional means of exposure to systematic liquidity risk before and after a termination for sample stocks and their matched controls. Changes in liquidity risk are assessed within the framework of Acharya and Pedersen (2005), who propose an equilibrium model describing how individual firms’ exposure to aggregate liquidity risk affects expected returns. The resulting pricing equation, \( E(r_i - r_j) = E(c_{i,j}) + \delta \beta_i + \beta_2 - \beta_3 - \beta_4 \), is described in Section III.D. The four betas capture exposures to aggregate risks embodied by the co-movement between: Stock returns and the market return (\( \beta_1 \)); stock illiquidity and aggregate illiquidity (\( \beta_2 \)); stock returns and aggregate illiquidity (\( \beta_3 \)); and stock illiquidity and the market return (\( \beta_4 \)). Increases in expected returns due to liquidity risk are associated with an increase in \( \beta_2 \) and decreases in \( \beta_3 \) and \( \beta_4 \). Our empirical methodology closely follows Acharya and Pedersen, adapted to a weekly frequency (Wednesday close to Wednesday close). For each stock, illiquidity is calculated as the adjusted Amihud measure, defined as \( \min(0.25 + 0.30 Iliq_i P_{i,t}^M, 30.00) \). where \( Iliq_i^t \) is the ratio of absolute stock return to the dollar trading volume (scaled by \( 10^6 \)) for stock \( i \) averaged over the days in week \( t \), and \( P_{i,t}^M \) is the ratio of the capitalizations of NYSE and AMEX stocks at the end of week \( t - 1 \) to the value at the end of the first week of 1998. Aggregate illiquidity is the value-weighted average illiquidity of NYSE and AMEX stocks each week. Because of the persistence in illiquidity levels, betas are calculated using innovations in stock illiquidity and aggregate illiquidity defined as residuals from AR(2) models. As in previous tables, for each sample termination, a control group is formed by selecting stocks with the same Daniel et al. (1997) size and book-to-market benchmark assignment in the month of June prior to a termination, subject to the conditions that control firms a) were covered by one or more analysts in the three months before the event and b) did not themselves experience a coverage termination in the quarter before and after the event. When more than five matches exist, we choose the five stocks closest to the sample stock in terms of market beta. (We lose 158 observations involving stocks that have no viable controls.) Betas are winsorized 2.5% in each tail to mitigate the effect of extreme outliers and reduce noise. We then calculate a difference-in-diifference test for each sample stock \( i \), \( \text{DiD} = (\text{post}_{i,t} - \text{pre}_{e}) - (\text{post}_{i,t} - \text{pre}_{e,\text{Control Group} i}) \), and report the cross-sectional mean. Betas are calculated using weekly data over 12-month, 18-month, and 24-month windows ending two weeks prior to the termination announcement or starting two weeks after the announcement date. We test the null hypothesis that exposure to systematic liquidity risk is unchanged around a coverage termination using bootstrapped \( p \)-values. These adjust for potential cross-sectional dependence due to overlapping estimation windows caused by time clustering as multiple stocks are terminated in each brokerage-firm closure event. The bootstraps use a block length of 100, correcting to the approximate number of terminations per closure event.

<table>
<thead>
<tr>
<th>Terminations</th>
<th>Matched controls</th>
<th>Mean</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before drop</td>
<td>After drop</td>
<td>Before drop</td>
</tr>
<tr>
<td><strong>Panel A: 12-month window</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net DiD: ( \Delta(\beta_1 + \beta_2 - \beta_3 - \beta_4) )</td>
<td>0.168</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>( \beta_1 ) (stock return vs. market return)</td>
<td>1.177</td>
<td>1.143</td>
<td>1.122</td>
</tr>
<tr>
<td>( \beta_2 ) (stock illiquidity vs. mkt illiquidity)</td>
<td>0.016</td>
<td>0.026</td>
<td>0.046</td>
</tr>
<tr>
<td>( \beta_3 ) (stock return vs. market illiquidity)</td>
<td>-0.032</td>
<td>-0.023</td>
<td>-0.026</td>
</tr>
<tr>
<td>( \beta_4 ) (stock illiquidity vs. market return)</td>
<td>-0.112</td>
<td>-0.174</td>
<td>-0.233</td>
</tr>
<tr>
<td><strong>Panel B: 18-month window</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net DiD: ( \Delta(\beta_1 + \beta_2 - \beta_3 - \beta_4) )</td>
<td>0.191</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>( \beta_1 ) (stock return vs. market return)</td>
<td>1.176</td>
<td>1.162</td>
<td>1.114</td>
</tr>
<tr>
<td>( \beta_2 ) (stock illiquidity vs. mkt illiquidity)</td>
<td>0.016</td>
<td>0.031</td>
<td>0.042</td>
</tr>
<tr>
<td>( \beta_3 ) (stock return vs. market illiquidity)</td>
<td>-0.027</td>
<td>-0.024</td>
<td>-0.024</td>
</tr>
<tr>
<td>( \beta_4 ) (stock illiquidity vs. market return)</td>
<td>-0.146</td>
<td>-0.330</td>
<td>-0.219</td>
</tr>
<tr>
<td><strong>Panel C: 24-month window</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net DiD: ( \Delta(\beta_1 + \beta_2 - \beta_3 - \beta_4) )</td>
<td>0.164</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>( \beta_1 ) (stock return vs. market return)</td>
<td>1.167</td>
<td>1.174</td>
<td>1.101</td>
</tr>
<tr>
<td>( \beta_2 ) (stock illiquidity vs. mkt illiquidity)</td>
<td>0.019</td>
<td>0.035</td>
<td>0.045</td>
</tr>
<tr>
<td>( \beta_3 ) (stock return vs. market illiquidity)</td>
<td>-0.027</td>
<td>-0.024</td>
<td>-0.023</td>
</tr>
<tr>
<td>( \beta_4 ) (stock illiquidity vs. market return)</td>
<td>-0.165</td>
<td>-0.337</td>
<td>-0.267</td>
</tr>
</tbody>
</table>
Chapter 4

A Practical Guide to Volatility Forecasting (with Christian Brownlees and Robert Engle)

Abstract: We present a volatility forecasting comparative study based on the methodology and financial data from Vlab, an econometric software application for automated real time volatility analysis. Our goal is to identify successful predictive models over multiple horizons and to investigate how predictive ability is influenced by choices for estimation window length, innovation distribution, and frequency of parameter re-estimation. Test assets include a range of domestic and international equity indices and exchange rates. We find that model rankings are insensitive to forecast horizon and suggestions for estimation best practices emerge. While our main sample spans 1990-2008, we take advantage of the near-record surge in volatility during the last half of 2008 to ask if forecasting models or best practices break down during periods of turmoil. We find that volatility during the 2008 crisis was well approximated by predictions one day ahead, and should have been within risk managers’ 1% confidence intervals up to one month ahead.
4.1 Introduction

The crash of 2008 has led practitioners and academics alike to reassess the adequacy of our financial models. Soaring volatilities across asset classes have made it especially important to know how well our standard tools forecast volatility, especially amid episodes of turmoil that pervade all corners of the economy. Volatility prediction is a critical task in asset valuation and risk management for investors and financial intermediaries. The price of essentially every derivative security is affected by swings in volatility. Risk management models used by financial institutions and required by regulators take time-varying volatility as a key input. Poor appraisal of the risks to come can leave investors excessively exposed to market fluctuations or institutions hanging on a precipice of inadequate capital. For these purposes, the virtue of a volatility model lies in its ability to predict out-of-sample.

The research questions pursued here are motivated by development of the Volatility Laboratory (Vlab) project, part of the Volatility Institute at New York University’s Stern School of Business. Vlab is an econometric software application providing daily volatility forecasts over multiple horizons for hundreds of securities across a spectrum of asset classes. For each asset, forecasts are based on a variety of models. While forecasts are updated daily, model parameters are re-estimated each week using all data available for an asset, starting from 1990, via Gaussian maximum likelihood. Several forecasting strategy questions naturally arise as part of the Vlab effort. These questions are the thesis of this article, and may be stated as follows: When constructing volatility forecasts for a range of assets over multiple horizons, what are the best choices for i) volatility model, ii) estimation window length, iii) frequency of parameter re-estimation and iv) innovation distribution? In addition, we take advantage of the near-record surge in volatility during the last half of 2008 to ask if our conclusions regarding forecasting models or estimation strategies break down during

\footnote{http://vlab.stern.nyu.edu. Currently, a subset of the site’s functions are publicly available. Full access may be obtained through a free password application process.}
periods of turmoil.

The set of volatility models considered are those implemented in Vlab. Included are
the basic GARCH specification and four more elaborate models in the GARCH family
that embed various degrees of asymmetry and nonlinearity. The set includes the thresh-
old (T)GARCH (?; ?), exponential (E)GARCH (?), asymmetric power (AP)ARCH ? and
the nonlinear (N)GARCH (Engle, 1990). These are chosen because of their popularity among
practitioners and academics and because of their potential for capturing long-memory be-
havior in variances as well as the so-called “leverage effect”, or tendency for volatility to
respond differently to negative and positive past returns.

Estimation window lengths range from the full post-1990 sample as in Vlab, to windows
as small as four years. The possibility of breaks in volatility dynamics has been studied by
?, ? and ?, among others. If parameters are unstable, data from the distant past can bias
estimates and pollute forecasts. In this situation, it may be practical to re-estimate using a
shorter, rolling window. We also consider whether daily re-estimation improves over Vlab’s
weekly re-estimation, and similarly whether there is any significant loss to re-estimating only
monthly. This question trades off computing cost against forecast accuracy, and is relevant
when the collection of assets and models becomes large.

The prevalence of fat-tailed returns, even after adjusting for heteroskedasticity, is well
known and was first documented soon after the first ARCH papers were written (e.g. ?). GARCH maximum likelihood estimates will be consistent when a Gaussian likelihood is
maximized (even in the presence of fat tails), assuming that the conditional variance is cor-
rectly specified. However, taking non-Gaussian aspects of the data into account by using,
for instance, a Student $t$ likelihood can potentially improve efficiency by correcting the spec-
ification, thus improving forecasting performance. Using a heavy-tailed likelihood, however,
introduces a new parameter. Considering both Gaussian and Student $t$ innovations allows
us to assess the role played by heavy tailed innovations in the forecast optimization effort.
Our test assets are eighteen domestic and international equity indices, nine sectoral equity indices, and ten exchange rates. We forecast from one day to one month ahead. Our data begins in 1990 and we use an out-of-sample interval spanning 2001 to 2008. This period contains both high volatility episodes (corresponding roughly to the NBER recessions of March 2001 to November 2001 and December 2007 through the present), as well as the protracted interval of low volatility from approximately 2003 until 2007. We devote special attention to the period of financial crisis from September 2008 to December 2008 and characterize the extremity of the crisis compared to historical standards.

Our overall assessment proceeds in two stages. We first perform exhaustive comparisons of models and estimation strategies using the S&P 500 index as the test asset. This provides the skeleton for our evaluation procedure that is then applied to all other assets. Due to the sheer volume of output, we report summarized results for the full set of test assets and detailed results only for the S&P 500.

To evaluate volatility forecast accuracy we rely on ex post proxies for the true, latent volatility process. The two standard proxies are squared returns and the more precisely estimated realized variance, calculated from ultra high frequency data. Our forecast accuracy comparisons are based on squared returns for our full set of test assets. For the S&P 500 index we also use realized volatility to make our assessments - this allows us to evaluate if and how conclusions from our experiments might change when different proxies are used. Forecast accuracy is measured with robust forecast loss functions in the sense of ?.

We find that across assets and volatility regimes, the simplest asymmetric GARCH specification, the threshold GARCH model of ?, is most often the best forecaster. Its relative outperformance is strongest at short horizons of a few days. At the one month horizon, it is often only insignificantly better than the simple GARCH(1,1).

Results show that using the longest possible estimation window (dating back to 1990 in our data, and in Vlab) gives the best results. However, estimates reveal slowly varying
movements in model parameters. As a result, we find that a short four-year rolling estimation window is competitive with the long, growing window, while a medium length eight-year window has the poorest performance. This is consistent with some parametric instability that causes longer windows to contain stale data, while the small samples from short windows result in excessively noisy estimates. Ultimately, the econometrician will prefer using a very long window, giving low parameter standard errors, or a short window that accommodates parameter variation. For window lengths in between, efficiency is lost and little flexibility is gained, resulting in poor forecasts.

We also find that while updating estimates daily leads to more accurate forecasts, the improvement over weekly re-estimation is rarely statistically significant. On the other hand, re-estimating at the monthly frequency can lead to serious forecast deterioration compared to weekly updating.

The S&P 500 results lend insight to the ability of robust loss functions to successfully select the best forecasting models. As noted by ?, quasi-likelihood-based loss and mean squared error (discussed in detail in Section 4.2.2) deliver similar conclusions; however, the quasi-likelihood loss has much better power properties. We also conclude that while realized volatility is a far less noisy proxy of true volatility, squared returns paired with quasi-likelihood loss and a sufficiently long out-of-sample window provide sensible inference.

The literature on volatility forecasting and forecast evaluation is surveyed in ?, ?, ?. Volatility forecasting assessments are commonly structured to hold the test asset and estimation strategy fixed, focusing on model choice. We take a more pragmatic approach and consider the how much data should be used for estimation, how frequently a model should be re-estimated, and what innovation distributions should be used. This is done for each model we consider. Furthermore, we do not rely on a single asset or asset class to draw our conclusions. We look beyond volatility forecasting meta-studies, in particular the Poon and Granger papers, which focus almost exclusively on one day ahead forecasts. Our
work draws attention to the relevance of multi-step ahead forecast performance for model evaluation, especially in crisis periods when volatility levels can skyrocket in a matter of days. The issues of multiple step ahead forecasting has also been addressed by ? and ?. The forecast evaluation methodology employed builds upon the recent contributions on robust forecasting assessment developed in ? and ?, which consistently rank volatility forecasts despite the fact volatility is not observable. An important implication of these results is that the conflicting evidence reported by some previously published empirical studies on volatility forecasting is due to the use of non robust losses, and this calls for a reassessment of previous findings in this field. Other approaches to volatility forecast evaluation are based on assessing the value of predictions from economic (e.g. ?) or risk management (?, ?) perspectives.

The rest of the paper is organized as follows. Section 4.2 describes the volatility prediction strategy and evaluation methodology. Section 4.3 presents the results of a detailed prediction exercise of S&P 500 volatility as well as summary results for other national and sectoral equity indices and exchange rates. Section 4.4 investigates how well standard volatility models describe episodes of severe volatility turmoil as seen during late 2008. Concluding remarks follow in Section 4.5.

4.2 Volatility Forecasting Methodology

4.2.1 Recursive Forecast Procedure

A time series of continuously compounded returns is denoted \( \{r_t\}_{t=1}^{T} \), and \( \mathcal{F}_t \) denotes the information set available at \( t \). The unobserved variance of returns conditional on \( \mathcal{F}_t \) is \( \sigma_{t+i|t}^2 \equiv \text{Var}[r_{t+i}|\mathcal{F}_t] \). Variance predictions are obtained from a set of volatility models \( \mathcal{M} \equiv \)
\{m_1, m_2, \ldots, m_M\}. Model \(m\) can generically be represented as

\[ r_{t+1} = \epsilon_{t+1} \sqrt{h_{t+1}^{(m)}} \]

(4.1)

where \(h_{t+1}^{(m)}\) is an \(\mathcal{F}_t\)-measurable function and \(\epsilon_{t+1}\) is an iid zero mean/unit variance innovation. The specification of \(h_{t+1}^{(m)}\) determines the conditional variance evolution and is typically a function of the history of returns as well as a vector of unknown parameters to be estimated from the data. The \(i\)-step ahead volatility forecasts obtained by model \(m\) conditional on \(\mathcal{F}_t\) is denoted \(h_{t+i|t}^{(m)}\).

The real time volatility forecasting procedure is implemented as follows. For each day \(t\) in the forecasting sample, we estimate model \(m\) using data ending at or before \(t\), depending on the frequency of parameter re-estimation. We use the fitted model to then predict volatility at select horizons (1, 5, 10, 15 and 22 days ahead), resulting in a daily volatility forecast path \(\{h_{t+i|t}^{(m)}\}\). This procedure generates a sequence of overlapping forecast paths, each path formulated from different conditioning information.

The estimation strategy adopted in Vlab is to use all available returns (beginning with 1990) and update parameter estimates once per week by maximizing a Gaussian likelihood. We work with the Vlab methodology as a base case, and perturb that approach to determine if alternative estimation strategies can improve forecasting performance. In particular, we consider using four and eight year rolling estimation windows, rather than a growing window that uses the full post-1990 sample. We also explore re-estimating parameters daily or monthly, in addition to weekly. Finally, maximum likelihood estimation is performed using both Gaussian and Student \(t\) likelihoods.
4.2.2 Volatility Models

The five models we consider for $h_{t+1}^{(m)}$ in Equation 4.1 are chosen from the vast literature on GARCH modeling for their simplicity and demonstrated ability to forecast volatility over alternatives.\footnote{A recent survey of the GARCH universe can be found in Bollerslev (2008).}

The first, GARCH(1,1) (\(\alpha; \beta\)), is a natural starting point for model comparison due to its ubiquity and progenesis of alternative models. Furthermore, \(\alpha; \beta\) demonstrate that GARCH exchange rate volatility forecasts are not significantly outperformed by more sophisticated models using data from the early 1990s. GARCH describes the volatility process as

\[
h_{t+1} = \omega + \alpha r_t^2 + \beta h_t.
\]

Key features of this process are its mean reversion (imposed by the restriction \(\alpha + \beta < 1\)) and its symmetry - future variance responds as much to past positive returns as it does to negative returns.

Our next two models are the most common asymmetric GARCH models, designed to capture the tendency for volatilities to increase more when past returns are negative. Threshold ARCH (TGARCH) (\(\gamma; \beta\)) appends a linear asymmetry adjustment,

\[
h_{t+1} = \omega + (\alpha + \gamma I_{r_t < c})r_t^2 + \beta h_t
\]

where \(I\) is an indicator equaling one when the previous period’s return is below some threshold \(c\) (most commonly, \(c = 0\)). The inclination of equity volatilities to rise more when past returns are negative leads to \(\gamma > 0\).

Exponential GARCH (EGARCH) (Nelson (1991) models the log of variance,

\[
\ln(h_{t+1}) = \omega + \alpha(|\epsilon_t| - E[|\epsilon_t|]) + \gamma \epsilon_t + \beta \ln(h_t)
\]
where $\epsilon_t = r_t / \sqrt{h_t}$. The leverage effect is manifested in EGARCH as $\gamma < 0$.

Engle’s (1990) Nonlinear GARCH (NGARCH) specification models asymmetry in the spirit of previous specifications using a different functional device. When $\gamma < 0$ the impact of negative news is amplified relative to positive news,

$$h_{t+1} = \omega + \alpha (r_t + \gamma)^2 + \beta h_t.$$

The last model considered, asymmetric power ARCH (APARCH) (Ding, Granger and Engle (1993)), nests at least seven other GARCH specifications (Hentschel (1995)), and has been shown to capture long-memory behavior in volatility. The APARCH evolution is

$$h_{t+1}^{\delta/2} = \omega + \alpha (|r_t| - \gamma r_t)^\delta + \beta h_t^{\delta/2}.$$

Raising the left hand side to $2/\delta$ delivers the variance series. Since serial correlation of absolute returns is stronger than squared returns (Ding et al., 1993), allowing $\delta$ to be a free parameter can capture volatility dynamics more flexibly than other specifications. Asymmetries are incorporated via $\gamma$.

### 4.2.3 Forecast Evaluation

Our measure of predictive accuracy is based on the average forecast loss achieved by a model/strategy/proxy triplet. A model that provides a smaller average loss is more accurate and thus preferred. Choices for loss functions are extensive, and their properties vary widely. \cite{hansen2006} identifies a class of loss functions that are attractively robust in a sense defined by Hansen and Lunde (2006), which we heuristically described here. Volatility forecast comparison can be tricky due to the latent nature of volatility - forecasted values must be compared against an ex post proxy of volatility, rather than its true value. Loss functions of the Patton class
asymptotically generate the same ranking of models regardless of the proxy being used. This rank preservation holds as long as the proxy is unbiased and minimal regularity conditions are met. Most importantly, ranks are preserved in the hypothetical case of observable true volatility, ensuring that model rankings achieved with squared return, high/low range or realized volatility as proxy correspond to the ranking that would be achieved if forecasts were compared against the true volatility.

The Patton class is comprised of a continuum of loss functions indexed by a parameter on the real line. It rules out all losses traditionally used in the volatility forecast literature but two. We focus on these two functions.

Let \( \hat{\sigma}_t^2 \) be an unbiased ex post proxy of conditional variance, say realized volatility or squared returns, and let \( h_{t|t-k} \) be a volatility forecast based on \( t-k \) information \((k > 0)\). The robust losses we use are

\[
\text{QL} : L(\hat{\sigma}_t^2, h_{t|t-k}) = \frac{\hat{\sigma}_t^2}{h_{t|t-k}} - \log \frac{\hat{\sigma}_t^2}{h_{t|t-k}}
\]

\[
\text{MSE} : L(\hat{\sigma}_t^2, h_{t|t-k}) = (\hat{\sigma}_t^2 - h_{t|t-k})^2.
\]

The quasi-likelihood-based (QL) loss, named for its close relation to the Gaussian likelihood, depends only on the standardized residual, \( \frac{\hat{\sigma}_t^2}{h_{t|t-k}} \). The mean squared error (MSE) loss depends solely on the additive forecast error, \( \hat{\sigma}_t^2 - h_{t|t-k} \). Both QL and MSE are used for forecast evaluation in Section 4.3. There are a few reasons, however, why we prefer QL for forecast comparison. First, as a result of the fact that QL depends on the standardized residual, the loss series is iid under the null hypothesis that the forecasting model is correctly specified. MSE, which depends on additive errors, scales with the square of variance, thus contains high levels of serial dependence even under the null.\(^3\) While loss functions are not required to be iid in order to identify successful forecasting models, this trait makes it easier

\[{}^3\text{To see this, divide } \text{MSE} \text{ by } \hat{\sigma}_t^4 \text{ and note that the resulting quantity is iid when under the null. } \text{MSE} \text{ is therefore an iid process times the square of a highly serially correlated process.} \]

174
to identify a model’s failure to adequately capture predictable movements in volatility. Second, suppose that the volatility proxy $\hat{\sigma}_t^2$ can be expressed as $\hat{\sigma}_t^2 = h_0 t \eta_t$, where $h_0 t$ is the latent true variance and $\eta_t$ is a measurement error with unit expected value and variance $\tau^2$.

The expected value of $\text{MSE}$ is then

$$
\mathbb{E}\{ \text{MSE}(\hat{\sigma}_t^2, h_{t|t-k}) \} = \mathbb{E}(\hat{\sigma}_t^2 - h_{t|t-k})^2 \\
= \mathbb{E}(\hat{\sigma}_t^2 - h_0 t + h_0 t - h_{t|t-k})^2 \\
= \mathbb{E}((\eta_t - 1)h_0 t + h_0 t - h_{t|t-k})^2 \\
= \text{MSE}(h_0 t, h_{t|t-k}) + \tau^2 h_0^2 t,
$$

while the expected value of $\text{QL}$ is

$$
\mathbb{E}\{ \text{QL}(\hat{\sigma}_t^2, h_{t|t-k}) \} = \mathbb{E}\left\{ \frac{\hat{\sigma}_t^2}{h_{t|t-k}} - \log \frac{\hat{\sigma}_t^2}{h_{t|t-k}} \right\} \\
= \mathbb{E}\left\{ \frac{h_0 t}{h_{t|t-k}} \eta_t - \log \frac{h_0 t}{h_{t|t-k}} \eta_t \right\} \\
\approx \text{QL}(h_0 t, h_{t|t-k}) + \frac{\tau^2}{2}
$$

where the last line uses a standard Taylor expansion for moments of a random variable.

$\text{MSE}$ has a bias that is proportional to the square of the true variance, while the bias of $\text{QL}$ is independent of the volatility level. Amid financial distress, large $\text{MSE}$ losses will be a consequence of high volatility without necessarily corresponding to deterioration of forecasting ability. $\text{QL}$ avoids this ambiguity, making it easier to compare losses across volatility regimes.
Table 4.1: Asset list. For each asset class, the table reports the list of assets used in the forecasting application and the first date of the sample.

### Data

Daily return data on the S&P 500 index for 1990 to 2008 is from Datastream. S&P 500 realized variance is computed using intra-daily data return data on the S&P 500 SPDR exchange traded fund from NYSE-TAQ over the same period. The expanded data set for our large scale forecasting comparison includes the assets listed in Table 4.1. We use ten exchange rates versus the US dollar, nine domestic sectoral equity indices, and eighteen international equity indices. The sector index data are returns on S&P 500 industry sector SPDR exchange traded funds. International index data are returns on iShares exchange traded funds that track the MSCI country equity index. Exchange rates and index returns are from Datastream and span from the late 1990’s to December 2008, with the exact initial dates varying according to data availability.

The out-of-sample forecast horizon covers 2001 to 2008 and contains periods of both very low volatility and severe distress. Figure 4.1 shows the time series plot of daily realized volatility for the S&P 500 index alongside one day ahead predictions of a TGARCH model (expressed in annualized terms). US equity volatility reached its peak during the financial
Figure 4.1: S&P 500 TGARCH one step ahead volatility forecasts and realized volatility. Volatilities are expressed in annualized terms.
Table 4.2: GARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by * = 10%, ** = 5%, *** = 1%).

<table>
<thead>
<tr>
<th>Loss</th>
<th>Est. Strategy</th>
<th>Horizon</th>
<th>1d</th>
<th>1w</th>
<th>2w</th>
<th>3w</th>
<th>1m</th>
</tr>
</thead>
<tbody>
<tr>
<td>QLL $\sigma^2$</td>
<td>base</td>
<td>1.460</td>
<td>1.481</td>
<td>1.520</td>
<td>1.574</td>
<td>1.645</td>
<td></td>
</tr>
<tr>
<td>Student t news</td>
<td>-0.16</td>
<td>-0.10</td>
<td>0.08</td>
<td>0.41</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium window</td>
<td>-0.71</td>
<td>-1.05</td>
<td>-2.12</td>
<td>-3.07</td>
<td>-4.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long window</td>
<td>-0.71</td>
<td>-1.06</td>
<td>-1.63</td>
<td>-2.28</td>
<td>-3.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monthly update</td>
<td>-0.02</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.03</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily update</td>
<td>0.01</td>
<td>0.01</td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QLL $rv$</td>
<td>base</td>
<td>0.273</td>
<td>0.310</td>
<td>0.343</td>
<td>0.373</td>
<td>0.414</td>
<td></td>
</tr>
<tr>
<td>Student t news</td>
<td>-3.22</td>
<td>-1.96</td>
<td>-0.99</td>
<td>-0.37</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium window</td>
<td>0.97</td>
<td>-2.36</td>
<td>-5.30</td>
<td>-7.23</td>
<td>-9.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long window</td>
<td>-8.95</td>
<td>-12.55</td>
<td>-16.06</td>
<td>-18.58</td>
<td>-20.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monthly update</td>
<td>-0.05</td>
<td>-0.02</td>
<td>-0.07</td>
<td>-0.05</td>
<td>-0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily update</td>
<td>0.11</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE $\sigma^2$</td>
<td>base</td>
<td>27.533</td>
<td>30.050</td>
<td>31.980</td>
<td>33.828</td>
<td>36.347</td>
<td></td>
</tr>
<tr>
<td>Student t news</td>
<td>-0.28</td>
<td>-0.20</td>
<td>-0.14</td>
<td>-0.26</td>
<td>-0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium window</td>
<td>0.18</td>
<td>0.81</td>
<td>-0.08</td>
<td>-0.05</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long window</td>
<td>0.37</td>
<td>0.66</td>
<td>0.044</td>
<td>-0.23</td>
<td>-0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monthly update</td>
<td>-0.10</td>
<td>-0.09</td>
<td>-0.01</td>
<td>-0.08</td>
<td>-0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily update</td>
<td>-0.04</td>
<td>0.09</td>
<td>0.22</td>
<td>0.08</td>
<td>-0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE $rv$</td>
<td>base</td>
<td>4.357</td>
<td>4.998</td>
<td>5.984</td>
<td>6.901</td>
<td>7.950</td>
<td></td>
</tr>
<tr>
<td>Student t news</td>
<td>-3.71</td>
<td>-4.25</td>
<td>-5.06</td>
<td>-6.33</td>
<td>-9.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium window</td>
<td>-0.34</td>
<td>-0.39</td>
<td>-0.67</td>
<td>-0.01</td>
<td>1.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long window</td>
<td>-4.93</td>
<td>-6.19</td>
<td>-6.87</td>
<td>-7.07</td>
<td>-6.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monthly update</td>
<td>1.04</td>
<td>1.36</td>
<td>1.22</td>
<td>0.97</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily update</td>
<td>0.16</td>
<td>0.10</td>
<td>0.23</td>
<td>-0.08</td>
<td>-0.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

turmoil of Fall 2008 with levels of realized volatilities exceeding 100%. This period is also characterized by high volatility of volatility. As of early September 2008, realized volatility was near 20%, and more than quadrupled in less than three months.
<table>
<thead>
<tr>
<th>Loss</th>
<th>Est. Strategy</th>
<th>Horizon</th>
<th>1 d</th>
<th>1 w</th>
<th>2 w</th>
<th>3 w</th>
<th>1 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>QLL</td>
<td>base</td>
<td>1.415</td>
<td>1.442</td>
<td>1.478</td>
<td>1.547</td>
<td>1.624</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student t news</td>
<td>0.01</td>
<td>0.02</td>
<td>0.30</td>
<td>0.59</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium window</td>
<td>-0.86</td>
<td>-1.07</td>
<td>-1.79</td>
<td>-3.06</td>
<td>-4.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>long window</td>
<td>-1.06</td>
<td>-1.46</td>
<td>-2.06</td>
<td>-2.65</td>
<td>-3.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>monthly update</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>daily update</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>QLL</td>
<td>base</td>
<td>0.243</td>
<td>0.289</td>
<td>0.328</td>
<td>0.368</td>
<td>0.415</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student t news</td>
<td>-2.84</td>
<td>-2.20</td>
<td>-1.45</td>
<td>-0.913</td>
<td>-0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium window</td>
<td>5.80</td>
<td>2.79</td>
<td>1.03</td>
<td>-1.052</td>
<td>-3.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>monthly update</td>
<td>-0.12</td>
<td>-0.06</td>
<td>-0.08</td>
<td>-0.05</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>daily update</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.010</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>base</td>
<td>25.583</td>
<td>28.874</td>
<td>31.151</td>
<td>33.197</td>
<td>36.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student t news</td>
<td>0.31</td>
<td>0.57</td>
<td>0.37</td>
<td>0.07</td>
<td>-0.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium window</td>
<td>0.16</td>
<td>-0.60</td>
<td>-1.41</td>
<td>-1.07</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>long window</td>
<td>0.31</td>
<td>0.05</td>
<td>-0.22</td>
<td>-0.34</td>
<td>-0.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>monthly update</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>daily update</td>
<td>-0.05</td>
<td>0.05</td>
<td>0.14</td>
<td>0.05</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>base</td>
<td>3.647</td>
<td>4.550</td>
<td>5.687</td>
<td>6.474</td>
<td>7.312</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student t news</td>
<td>-10.2</td>
<td>-10.50</td>
<td>-11.38</td>
<td>-12.78</td>
<td>-15.209</td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium window</td>
<td>5.84</td>
<td>4.05</td>
<td>4.32</td>
<td>5.87</td>
<td>8.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>long window</td>
<td>-2.71</td>
<td>-4.49</td>
<td>-4.69</td>
<td>-4.85</td>
<td>-4.84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>monthly update</td>
<td>0.70</td>
<td>0.66</td>
<td>0.16</td>
<td>0.18</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>daily update</td>
<td>-0.18</td>
<td>0.75</td>
<td>-0.11</td>
<td>-0.21</td>
<td>-0.13</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: TGARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by *=10%, **=5%, ***=1%).

179
Table 4.4: EGARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by *=10%, **=5%, ***=1%).

<table>
<thead>
<tr>
<th>Loss</th>
<th>Est. Strategy</th>
<th>1 d</th>
<th>1 w</th>
<th>2 w</th>
<th>3 w</th>
<th>1 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>QLL $r^2$</td>
<td>base</td>
<td>-0.87</td>
<td>1.45</td>
<td>2.73</td>
<td>4.23</td>
<td>1.684</td>
</tr>
<tr>
<td></td>
<td>Student t news</td>
<td>-2.31</td>
<td>-3.76</td>
<td>-5.14</td>
<td>-7.64</td>
<td>-10.31</td>
</tr>
<tr>
<td></td>
<td>medium window</td>
<td>-1.29</td>
<td>-1.75</td>
<td>-1.93</td>
<td>-2.53</td>
<td>-2.88</td>
</tr>
<tr>
<td></td>
<td>long window</td>
<td>-0.22</td>
<td>-0.05</td>
<td>0.02</td>
<td>-0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>monthly update</td>
<td>-0.17</td>
<td>0.07</td>
<td>0.14</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>daily update</td>
<td>0.234</td>
<td>0.282</td>
<td>0.320</td>
<td>0.365</td>
<td>0.413</td>
</tr>
<tr>
<td>QLL $rv$</td>
<td>base</td>
<td>-26.74</td>
<td>-15.22</td>
<td>-8.31</td>
<td>-3.50</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>Student t news</td>
<td>-1.66</td>
<td>-3.47</td>
<td>-6.69</td>
<td>-10.18</td>
<td>-14.93</td>
</tr>
<tr>
<td></td>
<td>long window</td>
<td>0.50</td>
<td>0.68</td>
<td>0.56</td>
<td>0.42</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>monthly update</td>
<td>-4.75</td>
<td>-2.76</td>
<td>-2.13</td>
<td>-1.52</td>
<td>-1.11</td>
</tr>
<tr>
<td></td>
<td>daily update</td>
<td>0.234</td>
<td>0.282</td>
<td>0.320</td>
<td>0.365</td>
<td>0.413</td>
</tr>
<tr>
<td>MSE $r^2$</td>
<td>base</td>
<td>26.746</td>
<td>31.328</td>
<td>33.984</td>
<td>35.912</td>
<td>37.828</td>
</tr>
<tr>
<td></td>
<td>Student t news</td>
<td>3.36</td>
<td>5.63</td>
<td>5.17</td>
<td>4.79</td>
<td>3.65</td>
</tr>
<tr>
<td></td>
<td>medium window</td>
<td>2.471</td>
<td>-2.31</td>
<td>-3.70</td>
<td>-3.825</td>
<td>-3.224</td>
</tr>
<tr>
<td></td>
<td>long window</td>
<td>1.284</td>
<td>-0.31</td>
<td>-0.28</td>
<td>-0.26</td>
<td>-0.18</td>
</tr>
<tr>
<td></td>
<td>monthly update</td>
<td>-0.34</td>
<td>-0.23</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>daily update</td>
<td>0.10</td>
<td>0.30</td>
<td>0.02</td>
<td>0.18</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>daily update</td>
<td>0.53</td>
<td>3.859</td>
<td>4.758</td>
<td>5.428</td>
<td>6.085</td>
</tr>
<tr>
<td>MSE $rv$</td>
<td>base</td>
<td>-37.92</td>
<td>-14.12</td>
<td>-4.37</td>
<td>-0.81</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>Student t news</td>
<td>13.62</td>
<td>-2.19</td>
<td>-6.73</td>
<td>-8.27</td>
<td>-7.59</td>
</tr>
<tr>
<td></td>
<td>medium window</td>
<td>2.17</td>
<td>-3.27</td>
<td>-2.30</td>
<td>-2.17</td>
<td>-1.68</td>
</tr>
<tr>
<td></td>
<td>long window</td>
<td>0.51</td>
<td>0.36</td>
<td>0.19</td>
<td>0.03</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>monthly update</td>
<td>-0.53</td>
<td>-0.14</td>
<td>0.13</td>
<td>0.02</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Table 4.5: APARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by * = 10%, ** = 5%, *** = 1%).

<table>
<thead>
<tr>
<th>Loss</th>
<th>Est. Strategy</th>
<th>Horizon</th>
<th>1 d</th>
<th>1 w</th>
<th>2 w</th>
<th>3 w</th>
<th>1 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>QLL</td>
<td>$\sigma^2$</td>
<td>Student t news</td>
<td>1.447</td>
<td>1.446</td>
<td>1.485</td>
<td>1.557</td>
<td>1.633</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>base</td>
<td>0.04</td>
<td>0.12</td>
<td>0.49</td>
<td>0.67</td>
<td>1.06</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>medium window</td>
<td>-1.07</td>
<td>-0.76</td>
<td>-1.23</td>
<td>-2.25</td>
<td>-3.65</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>long window</td>
<td>-1.14</td>
<td>-1.61</td>
<td>-2.24</td>
<td>-3.00</td>
<td>-3.84</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>monthly update</td>
<td>-0.03</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.07</td>
<td>-0.06</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>daily update</td>
<td>0.06</td>
<td>0.02</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>Student t news</td>
<td>-2.96</td>
<td>-0.73</td>
<td>1.27</td>
<td>2.59</td>
<td>4.03</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>medium window</td>
<td>7.49</td>
<td>5.87</td>
<td>5.83</td>
<td>4.81</td>
<td>2.98</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>long window</td>
<td>-7.09</td>
<td>-0.56</td>
<td>11.09</td>
<td>12.66</td>
<td>14.08</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>monthly update</td>
<td>-0.22</td>
<td>-0.17</td>
<td>-0.30</td>
<td>-0.27</td>
<td>-0.27</td>
</tr>
<tr>
<td>QLL</td>
<td>$r^2$</td>
<td>daily update</td>
<td>0.24</td>
<td>0.07</td>
<td>0.07</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>base</td>
<td>25.678</td>
<td>29.421</td>
<td>32.098</td>
<td>33.949</td>
<td>36.408</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>Student t news</td>
<td>0.63</td>
<td>1.02</td>
<td>0.91</td>
<td>0.60</td>
<td>-0.08</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>medium window</td>
<td>0.75</td>
<td>1.09</td>
<td>0.69</td>
<td>0.50</td>
<td>0.36</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>long window</td>
<td>0.27</td>
<td>-0.01</td>
<td>-0.54</td>
<td>-0.50</td>
<td>-0.64</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>monthly update</td>
<td>-0.15</td>
<td>-0.08</td>
<td>-0.11</td>
<td>-0.18</td>
<td>-0.21</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>daily update</td>
<td>0.01</td>
<td>0.18</td>
<td>0.35</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>Student t news</td>
<td>-14.62</td>
<td>-10.38</td>
<td>-9.21</td>
<td>-8.72</td>
<td>-8.86</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>medium window</td>
<td>0.05</td>
<td>4.12</td>
<td>2.42</td>
<td>2.64</td>
<td>4.19</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>long window</td>
<td>-3.32</td>
<td>-3.62</td>
<td>-4.81</td>
<td>-5.37</td>
<td>-5.15</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>monthly update</td>
<td>1.00</td>
<td>0.83</td>
<td>0.46</td>
<td>0.24</td>
<td>-0.32</td>
</tr>
<tr>
<td>MSE</td>
<td>$r^2$</td>
<td>daily update</td>
<td>0.08</td>
<td>0.286</td>
<td>0.36</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>Loss $\sigma^2$</td>
<td>Est. Strategy</td>
<td>1 d</td>
<td>1 w</td>
<td>2 w</td>
<td>3 w</td>
<td>1 m</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td><strong>QLL $\sigma^2$</strong></td>
<td><strong>base</strong></td>
<td>1.422</td>
<td>1.459</td>
<td>1.498</td>
<td>1.574</td>
<td>1.659</td>
<td></td>
</tr>
<tr>
<td>Student t news</td>
<td>-0.02</td>
<td>-0.06</td>
<td>0.39</td>
<td>0.71</td>
<td><strong>1.27</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium window</td>
<td>-0.80</td>
<td>-1.21</td>
<td>-1.99</td>
<td>-3.26</td>
<td>-4.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long window</td>
<td>-0.97</td>
<td>-1.98</td>
<td>-2.72</td>
<td>-3.44</td>
<td>-4.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monthly update</td>
<td>-0.00</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily update</td>
<td>0.02</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>QLL $\rho_v$</strong></td>
<td><strong>base</strong></td>
<td>0.244</td>
<td>0.296</td>
<td>0.337</td>
<td>0.380</td>
<td>0.432</td>
<td></td>
</tr>
<tr>
<td>Student t news</td>
<td>-2.19</td>
<td>-1.48</td>
<td>-0.73</td>
<td>-0.18</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium window</td>
<td>3.30</td>
<td>0.015</td>
<td>-2.28</td>
<td>-4.02</td>
<td>-6.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long window</td>
<td>-12.93</td>
<td>-16.54</td>
<td>-19.14</td>
<td>-20.35</td>
<td>-20.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monthly update</td>
<td>-0.10</td>
<td>-0.064</td>
<td>-0.138</td>
<td>-0.139</td>
<td>-0.140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily update</td>
<td>0.132</td>
<td>0.02</td>
<td>0.036</td>
<td>0.034</td>
<td>0.068</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MSE $\sigma^2$</strong></td>
<td><strong>base</strong></td>
<td>27.060</td>
<td>30.027</td>
<td>32.054</td>
<td>33.948</td>
<td>36.171</td>
<td></td>
</tr>
<tr>
<td>Student t news</td>
<td>0.26</td>
<td>0.58</td>
<td>0.66</td>
<td>0.60</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium window</td>
<td>0.28</td>
<td>-0.47</td>
<td>-1.27</td>
<td>-1.37</td>
<td>-0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long window</td>
<td>0.40</td>
<td>-0.25</td>
<td>-0.53</td>
<td>-0.66</td>
<td>-0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monthly update</td>
<td>-0.06</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.16</td>
<td>-0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily update</td>
<td>-0.00</td>
<td>0.08</td>
<td>0.15</td>
<td>0.08</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MSE $\rho_v$</strong></td>
<td><strong>base</strong></td>
<td>3.667</td>
<td>4.337</td>
<td>5.146</td>
<td>5.856</td>
<td>6.525</td>
<td></td>
</tr>
<tr>
<td>Student t news</td>
<td>-5.42</td>
<td>-5.32</td>
<td>-5.58</td>
<td>-6.05</td>
<td>-7.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium window</td>
<td>7.56</td>
<td>5.42</td>
<td>5.04</td>
<td>5.05</td>
<td>6.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long window</td>
<td>2.04</td>
<td>-0.29</td>
<td>-0.68</td>
<td>-0.93</td>
<td>-0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monthly update</td>
<td>0.66</td>
<td>0.75</td>
<td>0.49</td>
<td>0.25</td>
<td>-0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily update</td>
<td>0.09</td>
<td>0.05</td>
<td>0.14</td>
<td>-0.08</td>
<td>0.016</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6: NGARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by * = 10%, ** = 5%, *** = 1%).
4.3 Empirical Volatility Forecasting Results

4.3.1 Forecasting S&P 500 Volatility

We begin evaluating estimation strategies by assessing how out-of-sample forecast losses of S&P 500 index volatility change when we modify the base Vlab estimation procedure. Variations of the base strategy include: shorter estimation window size (rolling four or eight year samples rather than all data available), different re-estimation frequency (daily or monthly updates rather than weekly), and different innovation distribution (Student $t$ rather than Gaussian).

Tables 4.2 to 4.6 report the forecasting results for each of our five GARCH specifications. Each table has four panels that use different criteria to evaluate forecasts: the first panel uses the $\text{QL}$ penalty function based on the squared return ($r^2$) proxy, the second uses $\text{QL}$ with realized variance ($rv$) as proxy, the third $\text{MSE}$ with $r^2$, and the fourth $\text{MSE}$ with $rv$. Within each panel, a row corresponds to a particular estimation strategy. The first row is the Vlab methodology; the second row alters the base strategy to use the heavier-tailed, Student $t$ likelihood; the third and fourth rows use the base strategy but replace the full sample estimation window with windows of 4 years (denoted ”short”) and 8 years (denoted ”medium”), respectively; the fifth and sixth rows perturb the base strategy by re-estimating parameters daily or monthly, rather than weekly. Columns represent forecast horizons of one day to one month.

The first row of each panel shows average forecast loss for the base Vlab specification. For modified strategies (listed in rows two through six of each panel), we report the percentage over- or under-performance relative to the base case in the same panel. We test whether a modified strategy significantly outperforms the Vlab base case (for the same panel) using the (?) test for comparing predictive ability.

The comprehensive conclusion we draw from Tables 4.2 to 4.6 is that there are no
systematic large gains to be had by modifying the Vlab procedure along the alternatives considered. There are exceptions, as expected from a comparison with a vast number of permutations, however results suggest that Vlab’s performance is satisfactory for all models.

Adoption of a Student $t$ likelihood does not significantly improve performance at short horizons, though at longer horizons significantly positive improvements are possible. The Student $t$ down-weights extremes with respect to the Gaussian, thus it can provide a more robust estimate of the long run variance. As we see in these tables and discuss in more detail later, short horizon volatility and return realizations do not appear to violate a Gaussian assumption, though at longer horizons we see substantially more tail events than a normal curve would predict. The possibility that a Student $t$ provides a better description of volatility at long horizons is consistent with these observations.

Using a shorter, rolling estimation window tends to weaken forecasting accuracy. In some cases, the performance decreases by as much as 20%. The window length results are non-monotonic. While the full sample dominates, we often see that the medium estimation window does worse than the short window. To understand how this may arise, Figure 4.2 displays the time series plot of parameter estimates for one of our models (TGARCH) and the implied persistence ($\alpha + \gamma/2 + \beta$) using the Vlab procedure. Estimates are expressed as absolute differences with respect to the estimates obtained at the beginning of the out-of-sample period. While there is no evidence of clear breaks in the parameters, the series do exhibit a degree of variation: $\alpha$ systematically declines during this period, ending up insignificantly different from zero by the end of 2008. Movements in $\beta$ and $\gamma$ appear quite large and negatively correlated, transferring weight in the evolution equation away from past conditional variance toward past squared negative returns. The graph also suggests that periods of more severe financial distress are associated with higher $\gamma$, thus a bigger weight on past squared negative returns.

This suggests that the short window has some ability to capture variation in parameters,
Figure 4.2: TGARCH parameter estimates series. The graph plots the series of the $\omega$, $\alpha$, $\beta$, $\gamma$ and persistence ($\alpha + \gamma/2 + \beta$) parameter estimates obtained by the Vlab estimation strategy using the TGARCH model from 2001 to 2008. Parameters are expressed as absolute differences with respect to the estimates obtained on January 2, 2001.

though at the cost of less precise estimates. Ultimately, the noisiness overwhelms the gains from parameter variation, and the net effect is slightly worse performance by the short window. The medium window, on the other hand, uses such a long history that it misses much of the time variation in parameters and at the same time loses accuracy compared to the full sample window, resulting in the worst forecasting performance of the three windows. We also see more deterioration in short and medium window performance at longer forecast horizons.

Lastly, the update frequency results show that more frequent updating tends to modestly improve performance, but the difference is insignificant in most cases.
Figure 4.3: TGARCH out-of-sample QL Losses at different horizons. The graph plots the series of one month ahead (top), one week (middle) and one day (bottom) ahead QL losses using squared returns as the volatility proxy.

### 4.3.2 Direct Comparison of GARCH Models

Section 4.3.1 compared various forecasting methodologies for each GARCH model separately. In this section we make direct comparisons between GARCH models in the full sample and during the turmoil of Fall 2008.

Figure 4.3 displays the time series plot of the one day, one week and one month ahead QL (using squared returns) for TGARCH forecasts. An interesting aspect of multi-step volatility forecasting is that the loss times series exhibit different patterns depending on the horizon. For instance, February 27, 2007, the day of the so called ”Chinese Correction”, corresponds to the day with the highest one day ahead loss over the whole sample. In the monthly series, the Chinese Correction is less than half as severe as monthly losses in the fall
Table 4.7: S&P 500 volatility prediction comparison of GARCH models from 2001 to 2008. For each loss and volatility proxy the table reports the out-of-sample loss at multiple horizons of the GARCH models as well as 60 days Historical Variance (HIS). Asterisks beneath GARCH models losses denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the GARCH model (level of statistical significance denoted by *=10%, **=5%, ***=1%). The best forecasting performance for each loss and proxy pair is highlighted in bold.
Table 4.8: S&P 500 volatility prediction comparison of GARCH models in Fall 2008. For each loss and volatility proxy the table reports the out-of-sample loss at multiple horizons of the GARCH models as well as 60 days Historical Variance (HIS). The best forecasting performance for each loss and proxy pair is highlighted in bold.

<table>
<thead>
<tr>
<th>Loss</th>
<th>Model</th>
<th>1 d</th>
<th>1 w</th>
<th>2 w</th>
<th>3 w</th>
<th>1 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>QLL</td>
<td>GARCH</td>
<td>1.664</td>
<td>1.788</td>
<td>2.304</td>
<td>2.593</td>
<td>3.324</td>
</tr>
<tr>
<td></td>
<td>TGARCH</td>
<td><em>1.461</em></td>
<td><em>1.560</em></td>
<td><em>1.985</em></td>
<td><em>2.311</em></td>
<td><em>2.875</em></td>
</tr>
<tr>
<td></td>
<td>EGARCH</td>
<td>1.701</td>
<td>1.992</td>
<td>2.809</td>
<td>3.453</td>
<td>4.386</td>
</tr>
<tr>
<td></td>
<td>APARCH</td>
<td>1.565</td>
<td>1.695</td>
<td>2.233</td>
<td>2.631</td>
<td>3.306</td>
</tr>
<tr>
<td></td>
<td>NGARCH</td>
<td>1.653</td>
<td>1.815</td>
<td>2.376</td>
<td>2.781</td>
<td>3.561</td>
</tr>
<tr>
<td></td>
<td>HIS</td>
<td>2.267</td>
<td>2.522</td>
<td>3.091</td>
<td>3.426</td>
<td>4.001</td>
</tr>
<tr>
<td>QLL</td>
<td>GARCH</td>
<td>0.344</td>
<td>0.417</td>
<td>0.676</td>
<td>0.798</td>
<td>1.630</td>
</tr>
<tr>
<td></td>
<td>TGARCH</td>
<td>0.304</td>
<td><em>0.353</em></td>
<td><em>0.590</em></td>
<td><em>0.672</em></td>
<td><em>1.380</em></td>
</tr>
<tr>
<td></td>
<td>EGARCH</td>
<td><em>0.255</em></td>
<td>0.405</td>
<td>0.798</td>
<td>1.058</td>
<td>1.954</td>
</tr>
<tr>
<td></td>
<td>APARCH</td>
<td>0.293</td>
<td>0.381</td>
<td>0.652</td>
<td>0.810</td>
<td>1.552</td>
</tr>
<tr>
<td></td>
<td>NGARCH</td>
<td>0.314</td>
<td>0.398</td>
<td>0.682</td>
<td>0.832</td>
<td>1.697</td>
</tr>
<tr>
<td></td>
<td>HIS</td>
<td>0.485</td>
<td>0.624</td>
<td>0.946</td>
<td>1.170</td>
<td>1.877</td>
</tr>
<tr>
<td>MSE</td>
<td>GARCH</td>
<td>704.130</td>
<td>786.109</td>
<td>839.163</td>
<td>869.790</td>
<td>823.869</td>
</tr>
<tr>
<td></td>
<td>TGARCH</td>
<td><em>653.564</em></td>
<td><em>758.170</em></td>
<td><em>821.906</em></td>
<td><em>856.299</em></td>
<td><em>816.990</em></td>
</tr>
<tr>
<td></td>
<td>EGARCH</td>
<td>692.028</td>
<td>836.393</td>
<td>911.788</td>
<td>942.909</td>
<td>874.657</td>
</tr>
<tr>
<td></td>
<td>AGARCH</td>
<td>658.838</td>
<td>776.873</td>
<td>852.660</td>
<td>880.727</td>
<td>830.446</td>
</tr>
<tr>
<td></td>
<td>NGARCH</td>
<td>696.678</td>
<td>792.012</td>
<td>848.020</td>
<td>878.589</td>
<td>821.554</td>
</tr>
<tr>
<td></td>
<td>HIS</td>
<td>771.160</td>
<td>869.098</td>
<td>907.807</td>
<td>928.700</td>
<td>867.009</td>
</tr>
<tr>
<td>MSE</td>
<td>GARCH</td>
<td>102.628</td>
<td>116.548</td>
<td>144.042</td>
<td>166.307</td>
<td>193.407</td>
</tr>
<tr>
<td></td>
<td>TGARCH</td>
<td>88.820</td>
<td>108.576</td>
<td>140.737</td>
<td>158.336</td>
<td>178.265</td>
</tr>
<tr>
<td></td>
<td>EGARCH</td>
<td><em>60.960</em></td>
<td><em>91.652</em></td>
<td><em>118.044</em></td>
<td><em>133.112</em></td>
<td><em>147.591</em></td>
</tr>
<tr>
<td></td>
<td>AGARCH</td>
<td>79.553</td>
<td>109.399</td>
<td>135.607</td>
<td>150.671</td>
<td>164.647</td>
</tr>
<tr>
<td></td>
<td>NGARCH</td>
<td>88.126</td>
<td>102.324</td>
<td>124.960</td>
<td>140.879</td>
<td>156.213</td>
</tr>
<tr>
<td></td>
<td>HIS</td>
<td>103.100</td>
<td>114.192</td>
<td>133.902</td>
<td>150.646</td>
<td>176.237</td>
</tr>
</tbody>
</table>
of 2008. Conversely, the late 2008 crash, which resulted in harrowing losses at the monthly horizon, was comparatively mild according to one day forecasts. Large isolated shocks affect short horizons but get diluted as the horizon increases. On the other hand, transitions across volatility regimes lead to poor long horizon forecasts but have very little effect on short horizon forecastability.

Table 4.7 shows losses for each GARCH model over the full out-of-sample period 2001 to 2008. We report results for both QL and MSE loss functions with $r^2$ and $rv$ proxies. To compare accuracy across models, we use the Diebold-Mariano test to detect if a given model provides significantly lower average losses compared to the GARCH model. For comparison purposes the table also includes the naïve 60 days historical variance forecast. Significant outperformance at the 10%, 5% and 1% significance level are denoted by (\textasteriskcentered), (\textasteriskcentered\textasteriskcentered), and (\textasteriskcentered\textasteriskcentered\textasteriskcentered), respectively.

Asymmetric specifications provide lower out-of-sample losses, especially over one day and one week. At longer horizons, recent negative returns are less useful for predicting future volatility. At the one month horizon, the mean reversion effect begins to dominate as the difference between asymmetric and symmetric GARCH becomes insignificant, and historical variance becomes competitive.

The choice of loss function does not change rankings, but MSE loss seems to provide more mixed evidence than QL. There is some discrepancy in the rankings however when using different proxies. Squared returns favor TGARCH while realized volatility selects EGARCH. The discrepancy should not be overstated, however, as the methods do not significantly outperform each other. Results do suggest model rankings are stable over various forecasting horizons.

Table 4.8 repeats the direct GARCH comparison from Table 4.7 but focuses on the extreme volatility interval from September 2008 through December 2008. During this time, forecasting losses at all horizons are systematically larger than in the overall sample. Recall
that QL is unaffected by changes in the level of volatility, so that changes in average losses purely represent differences in forecasting accuracy. We find that one step ahead losses during fall 2008 are only modestly higher than those registered in the full sample. At one month, however, QL losses are twice as big as the full sample using squared returns and four times as big using realized volatility. An important feature of this table is that our conclusions about model ranking remain largely unchanged during the crisis. TGARCH tends to be most accurate, though differences with symmetric GARCH and historical variance start to lose significance at long horizons. MSE gives a much more confused picture of volatility in Fall 2008. Most noticeably, the MSE level during the crisis is an order of magnitude higher than the in full sample. GARCH specifications systematically outperform historical variance only at very short horizons, though at such horizons we again find that the asymmetric versions are superior.

4.3.3 Volatility Forecasting Across Asset Classes

We repeat the preceding experiments using a broader collection of assets: international equities indices, S&P 500 sectoral indices and exchange rates, detailed in Table 4.1. For each asset we recursively forecast volatility using the methodology described in Section 4.2.

Tables 4.9 and 4.10 contain the summary forecasting results under QL and MSE losses, respectively. In each table, there are three panels representing exchange rates, sector indices and international indices, respectively. The table reports losses from each model averaged not only across time, but also averaged over all assets in the same class. Asterisks denote that a model significantly outperformed GARCH based on the Diebold-Mariano test. We also report the relative winning frequency for each model, defined as the number of assets in a class for which a given model provided the best out-of-sample forecasts.

Of all asset classes, exchange rates appear to be most forecastable as they give the smallest losses according to QL. For several cases, the average loss point estimate is lower
Table 4.9: QL loss volatility prediction comparison of GARCH models from 2001 to 2008 across asset classes. For each asset class the table reports the out-of-sample average QL loss at multiple horizons as well as the relative frequency of cases in which a model achieved the best performance in a given asset class. Asterisks beneath losses denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the GARCH model (level of statistical significance denoted by *=10%, **=5%, ***=1%). The best average forecasting performance for each loss and proxy pair is highlighted in bold.
Table 4.10: MSE loss volatility prediction comparison of GARCH models from 2001 to 2008 across asset classes. For each asset class the table reports the out-of-sample average MSE loss at multiple horizons as the relative frequency of cases in which a model achieved the best performance in a given asset class. Asterisks beneath losses denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the GARCH model (level of statistical significance denoted by *=10%, **=5%, ***=1%). The best average forecasting performance for each loss and proxy pair is highlighted in bold.
Table 4.11: QL loss at multiple horizons volatility prediction comparison of GARCH models in Fall 2008 across asset classes. For each asset class the table reports the out-of-sample average. The best average forecasting performance for each loss and proxy pair is highlighted in bold.

<table>
<thead>
<tr>
<th>Model</th>
<th>Horizon</th>
<th>1 d</th>
<th>1 w</th>
<th>2 w</th>
<th>3 w</th>
<th>1 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>1.950</td>
<td>2.102</td>
<td>2.459</td>
<td>2.638</td>
<td>3.250</td>
<td></td>
</tr>
<tr>
<td>TARCH</td>
<td><strong>1.925</strong></td>
<td><strong>2.081</strong></td>
<td><strong>2.454</strong></td>
<td><strong>2.636</strong></td>
<td><strong>3.244</strong></td>
<td></td>
</tr>
<tr>
<td>EGARCH</td>
<td>2.050</td>
<td>2.314</td>
<td>2.898</td>
<td>3.309</td>
<td>4.410</td>
<td></td>
</tr>
<tr>
<td>APARCH</td>
<td>1.928</td>
<td>2.100</td>
<td>2.489</td>
<td>2.690</td>
<td>3.366</td>
<td></td>
</tr>
<tr>
<td>NGARCH</td>
<td>1.937</td>
<td>2.097</td>
<td>2.456</td>
<td>2.648</td>
<td>3.264</td>
<td></td>
</tr>
<tr>
<td>Equity Sectors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>2.246</td>
<td>2.305</td>
<td>2.723</td>
<td>3.056</td>
<td>4.057</td>
<td></td>
</tr>
<tr>
<td>TARCH</td>
<td><strong>2.217</strong></td>
<td><strong>2.120</strong></td>
<td><strong>2.523</strong></td>
<td><strong>2.834</strong></td>
<td><strong>3.870</strong></td>
<td></td>
</tr>
<tr>
<td>EGARCH</td>
<td>2.314</td>
<td>2.264</td>
<td>2.710</td>
<td>3.036</td>
<td>4.007</td>
<td></td>
</tr>
<tr>
<td>APARCH</td>
<td>2.239</td>
<td>2.158</td>
<td>2.579</td>
<td>2.908</td>
<td>3.934</td>
<td></td>
</tr>
<tr>
<td>NGARCH</td>
<td>2.310</td>
<td>2.258</td>
<td>2.705</td>
<td>3.057</td>
<td>4.085</td>
<td></td>
</tr>
<tr>
<td>Equity Sectors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>2.234</td>
<td>2.146</td>
<td>2.749</td>
<td>3.469</td>
<td>4.447</td>
<td></td>
</tr>
<tr>
<td>TARCH</td>
<td><strong>2.153</strong></td>
<td><strong>2.054</strong></td>
<td><strong>2.574</strong></td>
<td><strong>3.349</strong></td>
<td><strong>4.250</strong></td>
<td></td>
</tr>
<tr>
<td>EGARCH</td>
<td>2.259</td>
<td>2.321</td>
<td>3.094</td>
<td>4.020</td>
<td>4.997</td>
<td></td>
</tr>
<tr>
<td>APARCH</td>
<td>2.166</td>
<td>2.105</td>
<td>2.652</td>
<td>3.405</td>
<td>4.280</td>
<td></td>
</tr>
<tr>
<td>NGARCH</td>
<td>2.179</td>
<td>2.098</td>
<td>2.659</td>
<td>3.407</td>
<td>4.372</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.12: MSE loss at multiple horizons volatility prediction comparison of GARCH models in Fall 2008 across asset classes. For each asset class the table reports the out-of-sample average. The best average forecasting performance for each loss and proxy pair is highlighted in bold.

<table>
<thead>
<tr>
<th>Model</th>
<th>Horizon</th>
<th>1 d</th>
<th>1 w</th>
<th>2 w</th>
<th>3 w</th>
<th>1 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>110.044</td>
<td>114.499</td>
<td>116.987</td>
<td>127.002</td>
<td>132.491</td>
<td></td>
</tr>
<tr>
<td>TARCH</td>
<td>108.383</td>
<td>113.066</td>
<td>115.490</td>
<td>124.790</td>
<td>129.967</td>
<td></td>
</tr>
<tr>
<td>EGARCH</td>
<td>109.409</td>
<td>113.732</td>
<td>118.584</td>
<td>123.752</td>
<td>126.294</td>
<td></td>
</tr>
<tr>
<td>APARCH</td>
<td>107.159</td>
<td>111.600</td>
<td>114.081</td>
<td>121.754</td>
<td>125.944</td>
<td></td>
</tr>
<tr>
<td>NGARCH</td>
<td>109.882</td>
<td>114.285</td>
<td>116.648</td>
<td>126.482</td>
<td>131.721</td>
<td></td>
</tr>
<tr>
<td>Equity Sectors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>1405.665</td>
<td>1426.406</td>
<td>1534.140</td>
<td>1570.666</td>
<td>1531.705</td>
<td></td>
</tr>
<tr>
<td>TARCH</td>
<td>1368.083</td>
<td>1414.462</td>
<td>1530.184</td>
<td>1564.682</td>
<td>1535.560</td>
<td></td>
</tr>
<tr>
<td>EGARCH</td>
<td>1397.247</td>
<td>1445.153</td>
<td>1558.041</td>
<td>1595.051</td>
<td>1542.011</td>
<td></td>
</tr>
<tr>
<td>APARCH</td>
<td>1365.171</td>
<td>1423.885</td>
<td>1542.536</td>
<td>1577.000</td>
<td>1542.832</td>
<td></td>
</tr>
<tr>
<td>NGARCH</td>
<td>1400.709</td>
<td>1434.603</td>
<td>1539.774</td>
<td>1573.340</td>
<td>1518.103</td>
<td></td>
</tr>
<tr>
<td>International Equities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>1746.705</td>
<td>1760.143</td>
<td>1919.864</td>
<td>2060.783</td>
<td>1970.426</td>
<td></td>
</tr>
<tr>
<td>TARCH</td>
<td><strong>1680.130</strong></td>
<td><strong>1730.626</strong></td>
<td><strong>1899.445</strong></td>
<td><strong>2027.356</strong></td>
<td><strong>1935.291</strong></td>
<td></td>
</tr>
<tr>
<td>EGARCH</td>
<td>1742.464</td>
<td>1861.786</td>
<td>2031.493</td>
<td>2124.680</td>
<td>1966.910</td>
<td></td>
</tr>
<tr>
<td>APARCH</td>
<td>1688.682</td>
<td>1762.531</td>
<td>1932.974</td>
<td>2048.238</td>
<td>1936.920</td>
<td></td>
</tr>
<tr>
<td>NGARCH</td>
<td>1730.617</td>
<td>1762.325</td>
<td>1920.797</td>
<td>2048.917</td>
<td>1932.315</td>
<td></td>
</tr>
</tbody>
</table>
for asymmetric models, despite the fact that leverage effects for exchange rates are not well defined. While we find that no specification obtains significantly lower average losses than GARCH according to the QL loss, though there is some significant evidence in favor of asymmetric models based on MSE. Also, asymmetric models demonstrate success in terms of winning frequency.

Tables 4.9 and 4.10 are strong evidence in favor of using asymmetric models for sectoral equity indices. TGARCH emerges as the best performer, closely followed by APARCH. This is clearest from the QL results, which show that all asymmetric specifications (other than EGARCH) outperform GARCH over all horizons. MSE results are similar, but less statistically significant.

International equities deliver similar results. Asymmetric specifications perform better than GARCH, with TGARCH the most frequent top model according to both QL and MSE losses. Most evidence of outperformance, however, is limited to shorter horizons - at long horizons the winning frequency becomes more uniform across models.

Table 4.11 and 4.12 report average losses during the fall 2008 crisis. Results confirm our findings in the S&P500 case. One-day ahead losses are virtually unchanged from those during the full sample, while one month losses are magnified by a factor of about two. TGARCH appears to be the best performer at all horizons for all asset classes, although the margin appears to remain small for exchange rates.

4.4 Did GARCH predict the Crisis of 2008?

Results from the previous section suggest that relative performance of various volatility models and forecasting methods did not change substantially during the recent crisis. In this section, we turn to the question of whether volatility forecasts significantly deteriorated in absolute terms during the crisis.
On November 1, 2008, the New York Times\textsuperscript{4} declared October to be “the most wild month in the 80-year history of the S&P 500.... In normal times, the market goes years without having even one [4% move]. There were none, for instance, from 2003 through 2007. There were three such days throughout the 1950s and two in the 1960s. In October, there were nine such days.” Headlines like these and anecdotes of practitioners reporting “15 sigma” moves in certain asset classes during the fall of 2008 practically shout that our risk measures must be broken.

Did volatility models genuinely fail during the crisis? We first address this question with a simple thought experiment: “How often would we observe forecast errors as large as those observed during the crisis if the world obeys a GARCH model?” We try to answer this with a simple approach. To begin with, we estimate a TGARCH model using the full sample of daily market returns\textsuperscript{5} dating from 1926-2008, and calculate multiple horizon forecast errors

---


\textsuperscript{5}For this portion of the analysis our forecasts are done in-sample to provide a simple, apples-to-apples comparison of losses and standardized residuals over different subsamples. The market return in this section
and corresponding standardized residuals within the sample. Table 4.4 presents average daily QL losses during the full sample, during the low volatility 2003-2007 subsample, and during the fall 2008 crisis sample. Average losses at all horizons in the full sample hover around 1.7, and are very similar to the losses experienced in the low volatility interval. As we turn to average crisis losses, we see that one step ahead losses are virtually the same as the rest of the sample. The severity of the crisis only becomes noticable at longer forecast horizons. The 22 day ahead forecast loss appears to double during the crisis.

From the historical distribution of losses, we next calculate the probability of observing losses at least as large as those seen during the crisis. To do this, we calculate a four month rolling average of daily forward forecast losses. We then calculate the number of days prior to September 2008 for which the average loss was larger than the average loss over the four month crisis sample. These are reported in Table 4.14. The historical probability of observing a one day loss at least as large as that observed during the crisis is 54.5%. That is, the average one step crisis loss falls very close to the center of the empirical distribution. For longer horizons, the historical exceedence probabilities decrease quickly. At a 22 day horizon, losses at least as large as those observed during the crisis occurred only 1.3% of time between 1926 and August 2008.

Before commenting on this outcome, we take a second approach to the question. We simulate data using the TGARCH model with parameters estimated from the full sample. In each simulation we generate a return process for 80 years and estimate the correctly specified model. Using estimated parameters, we construct in-sample forecasts at multiple horizons and their corresponding average losses. Next, we average the daily losses in the last four months of each simulated sample. Simulations are repeated 5,000 times and produce a simulated distribution of average daily losses. Finally, we count the number of simulations is defined as the CRSP value-weighted index to utilize the longest possible daily time series.

6 This step in the procedure builds estimation uncertainty into the exercise to more accurately match that historical comparison.
in which losses meet or exceed crisis losses in the data. Results are reported in the last row of Table 4.14. Under the null model, the probability of observing one step ahead losses greater than the 1.4 value in the crisis is 53.8%. This exceedence probability drops to 2.0% at the 22-day horizon. The cumulative loss probabilities implied by the historical record and simulations under the null model tell the exact same story. In terms of one-step ahead forecasts, the crisis sample was a typical season in a GARCH world. In contrast, one-month forecasts indeed suffered during the crisis, but do not fall outside a 99% confidence interval.\footnote{We repeat the crisis forecast calculations and analogous simulations as a true out-of-sample exercise, and find that the conclusions from this analysis are unchanged.}

The nature of volatility during the crisis seems to be captured by the facts that (i) crisis forecasts deteriorated only at long horizons, and (ii) over one day, errors were no larger than a typical day in the full 80 year sample. On a given date during the crisis, conditioning on poor returns up until that day resulted in quite well-informed forecasts for the next day, and thus mild average one day losses. However, this conditioning provided little help in predicting abnormally long strings of consecutive negative return days that occurred during the crisis.

\section*{4.5 Conclusion}

Our study seeks to comprehensively describe the forecasting ability of different volatility models over horizons of up to one month. We examine how different forecasting methods impact forecasting performance, including i) estimation sample sizes, ii) model re-estimation frequency and iii) Gaussian and Student $t$ innovation likelihoods. Our study is unique in the breadth of securities we use to compare methods, including a variety of national equity indices, industry sector ETFs and foreign exchange rates. Taking advantage of evidence from the extreme volatility period in the latter half of 2008, we are able to compare our results from the representative 1990-2008 sample, which included multiple high and low volatility
regimes, to the performance of the same methods during crisis.

We find that asymmetric models, especially TGARCH, perform well across methods, assets and subsamples. These two models perform well using the longest available data series. Updating parameter estimates as frequently as possible provides the best forecasting performance, but does not statistically dominate updating at slightly less frequent intervals such as one week. We find no evidence that the Student $t$ likelihood improves forecasting ability in comparison to the Normal, despite its potentially more realistic description of return tails. Finally, preferred methods do not change when forecasting multiple periods ahead or in time of increased market stress.

An exploration into the degree of extremity in volatility during the 2008 crisis reveals some interesting features. First and foremost, soaring volatility during that period was well described by short horizon forecasts, as seen by mean forecast losses commensurate with historical losses and expected losses under the null. At longer horizons, observed losses have historical and simulated $p$-values of 1% to 2%. We conclude that, although multi-step forecast losses are large and in the tail of the distribution, they cannot be interpreted as a rejection of GARCH models, and would have fallen within 99% predicted confidence intervals.
Bibliography


Appendix A

Appendix to Chapter 1

A.1 Model Derivations

Proof of Proposition [1]

Proof. The proof proceeds by the method of undetermined coefficients. The equilibrium solution is obtained by demonstrating that the Euler equation $E_t[\exp(m_{t+1} + r_{c,t+1})] = 1$ is always satisfied for the consumption claim. First, substitute the Campbell-Shiller return approximation,

$$r_{c,t+1} = \kappa_0 + \kappa_1 w_{c,t+1} - w_c + \Delta c_{t+1},$$

into the pricing kernel in the Euler equation to obtain

$$E_t \left[ \exp \left( \theta \ln \beta + \theta \left( 1 - \frac{1}{\psi} \right) \Delta c_{t+1} + \theta [\kappa_0 + \kappa_1 w_{c,t+1} - w_c] \right) \right] = 1.$$

Next, substitute for the log wealth-consumption ratio using the hypothesized form in (1.3). Evaluating the expectation involves computing the cumulant generating function of Laplace variables. The following property of the Laplace distribution is informative for this end.
Lemma 5. The cumulant generating function of a unit Laplace variable $W_{t+1}$ evaluated at $s$ is

$$\ln E_t[\exp(sW_{t+1})] = \ln \left( \frac{1}{1 - s^2} \right).$$

The cumulant generating function of the cash flow growth shock is therefore

$$\ln E_t\left[ \exp\left( s\sqrt{\Lambda_t}W_{t+1} \right) \right] = \ln \left( \frac{1}{1 - s^2\Lambda_t} \right).$$

(A.1)

To obtain log prices that are linear in $\Lambda_t$, I use a first order Taylor expansion of (A.1) around zero: $\ln E_t\left[ \exp\left( s\sqrt{\Lambda_t}W_{t+1} \right) \right] \approx s^2\Lambda_t$.

Based on this, the equilibrium restriction used to determine state variable coefficients $A_x, A_\sigma$ and $A_\Lambda$ is

$$1 = \exp\left\{ \theta \ln \beta + (1 - \gamma)\mu + \theta \left( \kappa_0 + A_0[\kappa_1 - 1] + \kappa_1[A_\sigma\sigma^2(1 - \rho_\sigma) + A_\Lambda\Lambda(1 - \rho_\Lambda)] \right) \\
+ \frac{(\theta\kappa_1)^2}{2} [A^2_\sigma\sigma^2 + A^2_\Lambda\Lambda^2] + x_t\left( \theta A_x[\kappa_1\rho_x - 1] + 1 - \gamma \right) \\
+ \sigma^2_t\left( \theta A_\sigma[\kappa_1\rho_\sigma - 1] + \frac{(1 - \gamma)^2\sigma^2_c}{2} + \frac{(\theta\kappa_1)^2}{2} A^2_x\sigma^2_x \right) \\
+ \Lambda_t\left( \theta A_\Lambda[\kappa_1\rho_\Lambda - 1] + (1 - \gamma)^2 \right) \right\}. $$

In equilibrium, this restriction must hold for any realization of state variables. This is satisfied when coefficients on state variables and the constant term are exactly zero, yielding implicit solutions for $A_x, A_\sigma$ and $A_\Lambda$.

The stated wealth-consumption ratio may be used within the Campbell-Shiller approximation to deduce returns on the consumption asset and the stochastic discount factor. This, in turn, is used to derive the log price-dividend ratio for each asset in the economy, which is also linear in tail risk. The proof of the log price-dividend ratio function proceeds in the same manner. The following equilibrium restriction delivers the result and determines state...
variable coefficients:

\[ 1 = \exp \left\{ \theta \ln \beta + (\phi_i - \gamma) \mu + \mu_i + (\theta - 1) \left( \kappa_0 + A_0[\kappa_1 - 1] + \kappa_1 [A_{\sigma} \sigma^2 (1 - \rho_{\sigma}) + A_{\Lambda} \Lambda (1 - \rho_{\Lambda})] \right) \\
+ \kappa_{i,1} [A_{i,\sigma} \sigma^2 (1 - \rho_{\sigma}) + \Lambda A_{i,\Lambda} (1 - \rho_{\Lambda})] + \kappa_{i,0} + A_{i,0}[\kappa_{i,1} - 1] \\
+ \frac{\sigma_{\sigma}^2}{2} (\theta - 1) A_{\sigma} + \kappa_{i,1} A_{i,\sigma}^2 + \frac{\sigma_{\Lambda}^2}{2} (\theta - 1) A_{\Lambda} + \kappa_{i,1} A_{i,\Lambda}^2 \right\} \\
+ x_t \left[ (\theta - 1) A_x (\kappa_1 \rho_x - 1) + A_{i,x} (\kappa_{i,1} \rho_x - 1) + \phi_i - \gamma \right] \\
+ \sigma_x^2 \left[ (\theta - 1) A_x (\kappa_1 \rho_x - 1) + A_{i,x} (\kappa_{i,1} \rho_x - 1) + \frac{\sigma_x^2}{2} (\theta - 1) A_x + \kappa_{i,1} A_{i,x} \right]^2 + \frac{\sigma_x^2}{2} (\phi_i - \gamma)^2 + \frac{\sigma_x^2}{2} \\
+ \Lambda_t \left[ (\theta - 1) A_{\Lambda} (\kappa_1 \rho_{\Lambda} - 1) + A_{i,\Lambda} (\kappa_{i,1} \rho_{\Lambda} - 1) + (\phi_i - \gamma)^2 + q_i^2 \right] \right\}. \]

Of particular interest to this paper are the coefficients on tail risk. For the wealth-consumption ratio, this is\footnote{While these expressions are written in an implicit form, the characterization is sufficient for signing \( A_{\Lambda} \) since the Campbell-Shiller linearization ensures \( \kappa_1 \), which is a function of \( A \) coefficients, falls in the interval \((0, 1)\).}

\[ A_{\Lambda} = \frac{(1 - \gamma)(1 - \frac{1}{\psi})}{1 - \kappa_1 \rho_{\Lambda}} < 0 \]

while for the price-dividend ratio it is

\[ A_{i,\Lambda} = \frac{1}{1 - \kappa_{i,1} \rho_{\Lambda}} \left[ A_{\Lambda} (\theta - 1) (\kappa_1 \rho_{\Lambda} - 1) + (\phi_i - \gamma)^2 + q_i^2 \right]. \]

For future reference, I also present the following expressions:

\[ A_x = \frac{1 - \frac{1}{\psi}}{1 - \kappa_1 \rho_x} \quad \text{and} \quad A_{i,x} = \frac{\phi_i - \frac{1}{\psi}}{1 - \kappa_{i,1} \rho_x}. \]

\[ \square \]

**Proof of Proposition 2**

**Proof.** Based on the wealth-consumption ratio derived in Proposition 1 and the Campbell-
Shiller identity, shocks to the discount factor are

\[ m_{t+1} - E_t[m_{t+1}] = -\gamma \sigma_c \sigma_t z_{c,t+1} - \gamma \sqrt{\Lambda_t} W_{c,t+1} - (1 - \theta) \kappa_1 A_x \sigma_x \sigma_t z_{x,t+1} \]

\[ - (1 - \theta) \kappa_1 A_\sigma \sigma_t z_{\sigma,t+1} - (1 - \theta) \kappa_1 A_\Lambda \sigma_\Lambda z_{\Lambda,t+1}. \]

The prices of risk in the economy are the coefficients on the shocks. I use \( \lambda_j \) to denote the price of risk for shock \( j \), so that \( \lambda_c = \gamma \), \( \lambda_x = (1 - \theta) \kappa_1 A_x \), \( \lambda_\sigma = (1 - \theta) \kappa_1 A_\sigma \) and \( \lambda_\Lambda = (1 - \theta) \kappa_1 A_\Lambda \) are the prices for exposure to transitory consumption growth, long run consumption growth, variance and tail risks, respectively.

Shocks to log returns have a similar representation,

\[ r_{i,t+1} - E_t[r_{i,t+1}] = \beta_{i,c} (\sigma_c \sigma_t z_{c,t+1} + \sqrt{\Lambda_t} W_{c,t+1}) + \beta_{i,x} \sigma_x \sigma_t z_{x,t+1} \]

\[ + \beta_{i,\sigma} \sigma_\sigma z_{\sigma,t+1} + \beta_{i,\Lambda} \sigma_\Lambda z_{\Lambda,t+1} + \sigma_i \sigma_t z_{i,t+1} + q_i \sqrt{\Lambda_t} W_{i,t+1}. \]

where \( \beta_{i,c} = \phi_i \), \( \beta_{i,x} = \kappa_{i,1} A_{i,x} \), \( \beta_{i,\sigma} = \kappa_{i,1} A_{i,\sigma} \) and \( \beta_{i,\Lambda} = \kappa_{i,1} A_{i,\Lambda} \).

Evaluating the Euler condition shows that

\[ 0 = E_t[m_{t+1}] + E_t[r_{i,t+1}] + \ln \left\{ E_t \left[ \exp (m_{t+1} - E_t[m_{t+1}] + r_{i,t+1} - E_t[r_{i,t+1}]) \right] \right\} \]

\[ = E_t[r_{i,t+1}] - r_{f,t} + \frac{1}{2} \text{Var}(r_{i,t+1}) - \beta_{i,c} \lambda_c (\sigma_c^2 \sigma_t^2 + 2 \Lambda_t) - \beta_{i,x} \lambda_x \sigma_x^2 \sigma_t^2 - \beta_{i,\sigma} \lambda_\sigma \sigma_\sigma^2 - \beta_{i,\Lambda} \lambda_\Lambda \sigma_\Lambda^2. \]

Proof of Proposition 3

Proof. From the proof of the previous proposition, return shocks are

\[ r_{i,t+1} - E_t[r_{i,t+1}] = \beta_{i,c} (\sigma_c \sigma_t z_{c,t+1} + \sqrt{\Lambda_t} W_{c,t+1}) + \beta_{i,x} \sigma_x \sigma_t z_{x,t+1} \]

\[ + \beta_{i,\sigma} \sigma_\sigma z_{\sigma,t+1} + \beta_{i,\Lambda} \sigma_\Lambda z_{\Lambda,t+1} + \sigma_i \sigma_t z_{i,t+1} + q_i \sqrt{\Lambda_t} W_{i,t+1}. \]
Using these shocks I derive the tail distribution of returns.

Let $S = \phi_i \sqrt{\Lambda_t} W_{c,t+1}$ and let $Y = \exp(S)$. The density of $S$ is

$$g_S(s) = \frac{1}{2\phi_i \sqrt{\Lambda_t}} \exp\left(\frac{-|s|}{\phi_i \sqrt{\Lambda_t}}\right).$$

The derivative of $S$ with respect to $Y$ is $1/Y$. The conservation of probability law, $g_S(s)ds = g_Y(y)dy$, therefore implies that the density of $Y$ is

$$G_Y(y) = \begin{cases} 
  y^{1/(\phi_i \sqrt{\Lambda_t})} & \text{if } y < 1 \\
  y^{-1/(\phi_i \sqrt{\Lambda_t})} & \text{if } y > 1,
\end{cases}$$

showing that $Y$ obeys a power law in both tails. Repeating this argument for $\exp(\phi_i \sqrt{\Lambda_t} W_{i,t+1})$ shows that it also obeys a power law in both tails with exponents $\pm 1/(\phi_i \sqrt{\Lambda_t})$. The asymptotic aggregation properties of power law distributions (i.e., in the limit as $u$ approaches the end of the support, in this case $0$ for the lower tail and $\infty$ for the upper tail) dictate that if two variables have power law tails with exponents $\xi_1$ and $\xi_2$, their product also behaves as a power law in its tail with exponent $\min(\xi_1, \xi_2)$; that is, the heavier-tailed power law dominates.\(^2\) Since the remaining shocks are Gaussian (or lognormal when exponentiated), they have no asymptotic effect on the tail distribution and the result follows. \(\square\)

**Proof of Proposition 4**

**Proof.** The proof follows the same logic as in Proposition 1. The following property of exponential random variables is used to evaluate the expectation in the Euler condition.

**Lemma 6.** The cumulant generating function of a unit exponential variable $V_{t+1}$ evaluated

\(^2\)These properties are summarized in the appendix of Gabaix et. al (2006). It is this asymptotic equivalence that gives rise to the $\sim$ notation shown in the proposition. Rigorous proofs of asymptotic aggregation rules may be found in Gnedenko and Kolmogorov (1968).
The cumulant generating function of the disaster shock is therefore

$$
\ln E_t[\exp(sV_{t+1})] = \ln \left( 1 - \delta + \frac{\delta}{1 - s\Lambda_t} \right).
$$

(A.2)

Because $\tau_{c,t+1}$ is a Bernoulli($\delta$) variable,

$$
E_t[\exp(s\tau_{c,t+1}\Lambda_t V_{t+1})] = (1 - \delta)(1) + \delta E_t[\exp(\Lambda_t V_{t+1})].
$$

Employing the cumulant generating function to the second term proves the lemma. To obtain log prices that are linear in $\Lambda_t$, I use a first order Taylor expansion of (A.2) around $\Lambda_t = \bar{\Lambda}$:

$$
\ln \left( 1 - \delta + \frac{\delta}{1 - s\Lambda_t} \right) \approx d(s) + c(s)\Lambda_t,
$$

where $d(s) = \ln\left(1 + \frac{\delta s\bar{\Lambda}}{1 - s\bar{\Lambda}}\right) - c(s)\bar{\Lambda}$ and $c(s) = \delta s ((1 - s\bar{\Lambda})^2 + \delta s\bar{\Lambda}(1 - s\bar{\Lambda}))^{-1}$. Finally, evaluating the expectation results in the condition

$$
1 = \exp\left( \theta \ln \beta + (1 - \gamma)\mu + \theta(\kappa_0 + A_0[\kappa_1 - 1] + \kappa_1[A_\sigma\sigma^2 + A_\Lambda\bar{\Lambda}(1 - \rho_\Lambda)]) + d(\gamma - 1) + \frac{1}{2}(\theta\kappa_1)^2(A_\sigma^2 + A_\Lambda^2) + \sigma^2[\theta A_\sigma(\kappa_1\rho_\sigma - 1) + \frac{1}{2}(1 - \gamma)^2\sigma^2_c] + \Lambda_t[\theta A_\Lambda(\kappa_1\rho_\Lambda - 1) + c(\gamma - 1)] \right).
$$

This equation is an equilibrium restriction yielding implicit solutions for $A_\sigma$ and $A_\Lambda$. Of particular interest is the coefficient on $\Lambda_t$,

$$
A_\Lambda = -\frac{\delta c(\gamma - 1)}{\theta(1 - \kappa_1\rho_\Lambda)} < 0.
$$

210
(The sign of \(A_\Lambda\) may be established numerically.)

In the case of the price-dividend ratio for asset \(i\), the result follows from the next equilibrium condition.

\[
1 = \exp \left\{ \theta \ln \beta + (\phi_i - \gamma) \mu + \mu_i + (\theta - 1) \left[ \kappa_0 + A_0(\kappa_1 - 1) + \kappa_1 A_\sigma \sigma^2(1 - \rho_\sigma) + A_\Lambda A_\Lambda (1 - \rho_\Lambda) \right] \\
+ \kappa_{i,0} + A_{i,0}[\kappa_{i,1} - 1] + \kappa_{i,1}[A_{i,\sigma} \sigma^2(1 - \rho_\sigma) + A_{i,\Lambda} A_\Lambda (1 - \rho_\Lambda)] + d(\gamma - \phi_i) \\
+ d(-q_i) + \frac{1}{2}((\theta - 1)\kappa_1 A_\sigma + \kappa_{i,1} A_{i,\sigma})^2 \sigma_\sigma^2 + \frac{1}{2}([\theta - 1]A_\Lambda A_\sigma + \kappa_{i,1} A_{i,\Lambda})^2 \sigma_\Lambda^2 \\
+ \sigma_i^2 \left[ (\theta - 1)A_\sigma(\kappa_1 \rho_\sigma - 1) + A_{i,\sigma}(\kappa_{i,1} \rho_\sigma - 1) + \frac{1}{2}(\phi_i - \gamma)^2 \sigma_\sigma^2 + \frac{1}{2} \sigma_i^2 \right] \\
+ \Lambda_t \left[ (\theta - 1)A_\Lambda(\kappa_1 \rho_\Lambda - 1) + A_{i,\Lambda}(\kappa_{i,1} \rho_\Lambda - 1) + c(\gamma - \phi_i) + c(-q_i) \right] \},
\]

which delivers the result. \(\square\)

**Proof of Proposition 5**

**Proof.** The expected return on asset \(i\) is

\[
E_t[r_{i,t+1}] = \kappa_{i,0} + \phi_i \mu + \mu_i + A_{i,0}(\kappa_{i,1} - 1) + A_{i,\sigma}(\kappa_{i,1} \rho_\sigma - 1) \sigma_i^2 + [A_{i,\Lambda}(\kappa_{i,1} \rho_\Lambda - 1) - \delta(\phi_i + q_i)]] A_t.
\]

Substituting the wealth-consumption ratio from Proposition 4 into the stochastic discount factor (via the Campbell-Shiller identity) and evaluating the expectation gives

\[
r_{f,t} = r_{f,0} + b_{f,\sigma} \sigma_t^2 + b_{f,\Lambda} \Lambda_t
\]

where \(b_{f,\Lambda} = A_\Lambda(1 - \theta)(\kappa_1 \rho_\Lambda - 1) - c(-\gamma)\). Assembling the preceding expressions produces the result. \(\square\)

**Proof of Proposition 6**

211
Proof. The proof proceeds as in Proposition 3. Let \( S = -\phi_i \Lambda_t V_{c,t+1} \) and let \( Y = \exp(S) \).

The density of \( S \) is

\[
g_S(s) = \frac{1}{\phi_i \Lambda_t} \exp \left( \frac{s}{\phi_i \Lambda_t} \right), \quad s \leq 0.
\]

The derivative of \( S \) with respect to \( Y \) is \( 1/Y \), which implies that the density of \( Y \) is

\[
g_Y(y) = g_S(s) \frac{ds}{dy} = \frac{1}{\phi_i \Lambda_t} y^{1/(\phi_i \Lambda_t) - 1}
\]

with corresponding cumulative distribution function

\[
G_Y(y) = y^{1/(\phi_i \Lambda_t)}.
\]

Accounting for the interaction between \( S \) and the Bernoulli variable \( \iota_{c,t+1} \) amounts to deriving the distribution of \( Y^{\iota_{c,t+1}} \), which is

\[
G(Y^{\iota})(y) = \delta 1_{y=1} + (1 - \delta) y^{1/(\phi_i \Lambda_t)} \sim (1 - \delta) y^{1/(\phi_i \Lambda_t)}.
\]

Repeating this argument for \( \exp(-q_i \iota_{i,t+1} \Lambda_t V_{i,t+1}) \) shows that its lower tail is also power law-distributed with exponent \( 1/(q_i \Lambda_t) \). Applying the asymptotic tail aggregation properties referenced in Proposition 3 and noting that the remaining shocks to \( r_{i,t+1} \) are Gaussian (thus they have no asymptotic contribution to the tail distribution) delivers the result.

A.2 Monte Carlo Evidence

A.2.1 Correct Specification of Tail Parameter Evolution

The first Monte Carlo experiment I run is designed to assess the finite sample properties of the dynamic power law quasi-maximum likelihood estimator under different dependence
and heterogeneity conditions. In all cases, the evolution of the parameter governing tail risk follows Equation 1.7 and therefore the statistical model’s specification of this process is correct. I allow for mis-specification in terms of dependence and in the level of the tail exponent across stock. In particular, data is generated by the following process:

\[ R_{i,t} = b_i R_{m,t} + e_{i,t} \]

where \( R_{m,t} \) and \( e_{i,t} \), \( i = 1, \ldots, n \), are independent Student \( t \) variates with \( a_i \zeta_t \) degrees of freedom. A well-known property of the Student \( t \) is that its tail distribution is asymptotically equivalent to a power law with tail exponent equal to (minus) the degrees of freedom. In generating data I therefore set the degrees of freedom equal to \( \zeta_t \), whose transition is described by Equation 1.7. The \( b_i \) coefficients control cross section dependence and heterogeneity in volatility. The \( a_i \) coefficients control the tail risk heterogeneity across observations. I consider four cases:

1. Independent and identically distributed observations: \( b_i = 0 \) and \( a_i = 1 \) for all \( i \),

2. Dependent and identically distributed observations: \( b_i \sim N(1, .5^2) \) and \( a_i = 1 \) for all \( i \),

3. Independent and heterogeneously distributed observations: \( b_i = 0 \) and \( a_i \sim N(1, .2^2) \) for all \( i \),

4. Dependent and heterogeneously distributed observations: \( b_i \sim N(1, .5^2) \) and \( a_i \sim N(1, .2^2) \) for all \( i \).

The cross section size is \( n=1000 \) or \( 2500 \), and the time series length is \( T=1000 \) or \( 5000 \). Parameters used to generate data are fixed at \( \pi_1 = 0.05 \) and \( \pi_2 = 0.93 \), with an intercept.

---

\[ ^3 \text{Observations in this case are identical only in terms of their tail exponent. Differences in } b_i \text{ across stocks introduces volatility and dependence heterogeneity.} \]
that ensures the mean value of $\zeta_t$ is three.

In each simulation, the quasi-maximum likelihood procedure described in Section 1.3 is used to estimate the model and its asymptotic standard errors. Summary statistics for parameter and asymptotic standard error estimates are reported in Table 11. Also reported is the time series correlation and mean absolute deviation between the fitted tail series and the true $\zeta_t$ series, averaged across simulations.

The general conclusion of the experiment is that the asymptotic theory of Section 1.3 is a good approximation for the finite sample behavior of the dynamic power law estimator. This is true not only when data are i.i.d., but also when observations are dependent and heterogenous. In all cases, the fitted tail series achieves a correlation of at least 97% with the true tail series.

A.2.2 Mis-specification of Tail Parameter Evolution

The next experiment proceeds as in the i.i.d. case above, but the true tail diverges from that assumed in the statistical model. In particular, the true degrees of freedom parameter $\zeta_t$ is conditionally stochastic and follows a first order Gaussian autoregression,

$$
\zeta_{t+1} = \zeta(1 - \rho) + \rho \zeta_t + \sigma \eta_{t+1}, \; \eta_{t+1} \sim N(0, 1).
$$

I fix $\rho = 0.99$, $n=1000$ and $T=1000$, and let $\sigma = 0.005$ or 0.010. The parameter $\sigma$ governs the variability of the process, and thus the range of tail risk values that the data can experience.

In each simulation, the quasi-maximum likelihood procedure described in Section 1.3 is used to estimate the model and its asymptotic standard errors. Summary statistics for the true process and the fitted process are reported in Table 12, as well as summary statistics for $\pi$ parameter estimates and their asymptotic standard errors.

The deterministic tail process provides accurate estimates even when the true tail pa-
rameter is stochastic. The mean absolute error between the fitted and true series ranges from 0.173 to 0.307, and their correlation ranges from 81.8% to 87.5%.
Appendix B

Appendix to Chapter 2

B.1 Assumptions for Quasi-Maximum Likelihood Results

Before stating technical assumptions, we note the more fundamental assumption of an underlying probability environment such that data \( \{\tilde{r}_t\} \) are realizations of a stochastic process on a complete probability space, and that the functions \( f_{1,t} \) and \( f_{2,t} \) in (2.4) and (2.5) are measurable for all \( t \). All stated convergence results are implicitly written with respect to the measure of this probability space.

**Assumption 3.** (a) For all \( \theta \in \Theta \) and \( \phi \in \Phi \), \( E[\log f_{1,t}(\tilde{r}_t, \theta)] \) and \( E[\log f_{2,t}(\tilde{r}_t, \hat{\theta}, \phi)] \) exist and are finite \( \forall t \);

(b) \( E[\log f_{1,t}(\tilde{r}_t, \theta)] \) and \( E[\log f_{2,t}(\tilde{r}_t, \hat{\theta}, \phi)] \) are continuous on \( \Theta \) and \( \Phi \) \( \forall t \);

(c) \( \{\log f_{1,t}(\tilde{r}_t, \theta)\} \) and \( \{\log f_{2,t}(\tilde{r}_t, \hat{\theta}, \phi)\} \) each obey the strong uniform law of large numbers.

**Assumption 4.** \( f_{1,t} \) and \( f_{2,t} \) are each twice continuously differentiable on \( \Theta \) and \( \Phi \) \( \forall t \).

**Assumption 5.** For all \( \theta \in \Theta \) and \( \phi \in \Phi \), \( E[\nabla_\theta L_1(\{\tilde{r}_t\}, \theta)] < \infty \) and \( E[\nabla_\phi L_2(\{\tilde{r}_t\}, \hat{\theta}, \phi)] < \infty \).
Assumption 6. (a) For all $\theta \in \Theta$ and $\phi \in \Phi$, $E[\nabla_\theta L_1(\{\tilde{r}_t\}, \theta)] < \infty$ and $E[\nabla_\phi L_2(\{\tilde{r}_t\}, \hat{\theta}, \phi)] < \infty$;
(b) $E[\nabla_\theta L_1(\{\tilde{r}_t\}, \theta)]$ and $E[\nabla_\phi L_2(\{\tilde{r}_t\}, \hat{\theta}, \phi)]$ are continuous on $\Theta$ and $\Phi$;
(c) $\{\nabla_\theta' s_{1,t}(\tilde{r}_t) = \nabla_\theta \log f_{1,t}(\tilde{r}_t, \theta)\}$ and $\{\nabla_\phi' s_{2,t}(\tilde{r}_t) = \nabla_\phi \log f_{2,t}(\tilde{r}_t, \hat{\theta}, \phi)\}$ each obey the strong uniform law of large numbers.

Assumption 7. $E[L_1(\tilde{r}_t, \theta)]$ is uniquely maximized by $\theta^*$ interior to $\Theta$ and $E[L_2(\tilde{r}_t, \theta^*, \phi)]$ is uniquely maximized by $\phi^*$ interior to $\Phi$.

Assumption 8. The double array

$$\{(T^{-1/2}s_{1,t}', T^{-1/2}s_{2,t}')\} \equiv \{(T^{-1/2}\nabla_\theta' \log f_{1,t}(\tilde{r}_t, \theta^*), T^{-1/2}\nabla_\phi' \log f_{2,t}(\tilde{r}_t, \theta^*, \phi^*)\}$$

obeys the central limit theorem.
Appendix C

Appendix to Chapter 3

Proof of Proposition 1

In all cases, investor \( i \) solves

\[
\max_{D_i} E[-e^{-C_i}|\mathcal{F}^i}] \text{ s.t. } C_i = R(W_0 - \tilde{D}_i) + \tilde{D}_i(u - RP), \quad i \in \{\text{informed, uninformed}\} \quad (C.1)
\]

where \( \tilde{D}_i \) is investor \( i \)'s demand for the risky asset, and subscript \( i \) denotes whether or not the investor observes the signal. Optimal demand is therefore given by

\[
D_i = \frac{E[u|\mathcal{F}^i] - RP}{V[u|\mathcal{F}^i]} \quad (C.2)
\]

where \( V[u|\mathcal{F}^i] \) denotes the variance of payoff \( u \) conditional on information set \( \mathcal{F}^i \).

Case 1: Symmetric Information

An informed investor’s expectation for the risky asset payoff is

\[
E[u|\mathcal{F}^{\text{informed}}] = E[u|s] = \theta + \frac{\sigma_u^2}{\sigma_s^2 + \sigma_v^2}s \quad (C.3)
\]
while the variance of $u$ conditional on the signal is

$$V[u|s] = \sigma_u^2 \left( 1 - \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \right). \tag{C.4}$$

The current case assumes symmetry with all investors informed. The market-clearing condition equates total optimal demand under symmetric information with total supply, $D_{informed} = X$, implying

$$P_{symm} = \frac{1}{R} \left( E[u|s] - V[u|s]X \right) \tag{C.5}$$

$$= \frac{1}{R} \left( \theta + \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} s - \sigma_u^2 \left( 1 - \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \right) X \right). \tag{C.6}$$

**Case 2: Asymmetric Information**

In the absence of an analyst, a fraction $\delta$ of the investors pay $c$ to obtain the payoff signal $s$. The proof for $0 < \delta < 1$ is immediate from Grossman and Stiglitz (1980). Uninformed investors base their demand solely on the observed price. To do this, they must filter from the price a noisy inference of the informed investors’ signal. Under rational expectations, the uninformed know the price is linear in the informed investors’ signal and the asset’s supply, that is,

$$P = c_1 + c_2 s + c_3 X. \tag{C.7}$$

This price is a conjecture to be verified in equilibrium, at which point we also solve for the constants $c_1$, $c_2$, and $c_3$.

The conditional moments of $u$ given an uninformed investor’s information set are:

$$E[u|\mathcal{F}^{uninformed}] = E[u|P] = \theta + \frac{c_2 \sigma_u^2}{c_2^2 (\sigma_u^2 + \sigma_v^2) + \sigma_v^2} (c_2 s + c_3 [X - \bar{X}]). \tag{C.8}$$
\[ V[u|P] = \sigma_u^2 \left( 1 - \frac{c_2 \sigma_u^2}{c_2^2 (\sigma_u^2 + \sigma_v^2) + c_3 \sigma_x^2} \right). \]  

(C.9)

When a fraction \( \delta \) of investors are informed, the equilibrium price is found from the market-clearing condition \( \delta D_{\text{informed}} + (1 - \delta) D_{\text{uninformed}} = X \), giving

\[
P_{\text{asymm}} = \frac{1}{R} \left( \frac{\delta}{V[u|s]} + \frac{(1 - \delta)}{V[u|P]} \right)^{-1} \left( \frac{\delta E[u|s]}{V[u|s]} + \frac{(1 - \delta) E[u|P]}{V[u|P]} - X \right) \]  

(C.10)

We can use conditional moments (C.8) and (C.9) to show that \( P_{\text{asymm}} \) is indeed linear in the informed investors’ signal and the asset’s supply. Matching coefficients in (C.7) and (C.10) gives

\[
c_2 = \frac{\delta \sigma_u^2 (\delta + \sigma_v^2 \sigma_x^2)}{R (\sigma_x^2 \sigma_v^4 + (\delta^2 + \delta \sigma_u^2 \sigma_x^2) \sigma_v^2 + \sigma_u^2 \delta^2)} \]  

(C.11)

\[
c_3 = \frac{-\sigma_v^2 \sigma_u^2 (\delta + \sigma_v^2 \sigma_x^2)}{R (\sigma_x^2 \sigma_v^4 + (\delta^2 + \delta \sigma_u^2 \sigma_x^2) \sigma_v^2 + \sigma_u^2 \delta^2)} \]  

(C.12)

Finally, we can write \( c_1 \) as a function of \( c_2 \) and \( c_3 \):

\[
c_1 = \frac{\theta}{R} - \frac{1 - \delta}{R} c_2 c_3 X \left( c_2^2 \left( \sigma_u^2 + \sigma_v^2 \right) + c_3^2 \sigma_x^2 \right)^{-1} \left( 1 - \frac{c_2^2 \sigma_u^2}{c_2^2 \left( \sigma_u^2 + \sigma_v^2 \right) + c_3^2 \sigma_x^2} \right)^{-1} \times \left( \frac{\delta}{\sigma_u^2} \left( 1 - \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \right)^{-1} + \frac{1 - \delta}{\sigma_u^2} \left( 1 - \frac{c_2^2 \sigma_u^2}{c_2^2 \left( \sigma_u^2 + \sigma_v^2 \right) + c_3^2 \sigma_x^2} \right)^{-1} \right)^{-1}. \]  

(C.13)

The expressions shown in Proposition 1 follow directly from using these solutions in the price and demand expressions, and noting that the unconditional expectations of the signal and supply are zero and \( \bar{X} \), respectively.