ESTIMATING THE RISK-RETURN TRADE-OFF WITH OVERLAPPING DATA
INFERENCE

Esben Hedegaard
Robert J. Hodrick

Estimating the Risk-Return Trade-off with Overlapping Data Inference
Esben Hedegaard and Robert J. Hodrick
NBER Working Paper No. 19969
March 2014
JEL No. G12

ABSTRACT

Asset pricing models such as the conditional CAPM are typically estimated with MLE using a monthly or quarterly horizon with data sampled to match the horizon even though daily data are available. We develop an overlapping data inference methodology (ODIN) that uses all of the data while maintaining the monthly or quarterly forecasting period, and we apply it to the conditional CAPM. Our approach recognizes that the first order conditions of MLE can be used as orthogonality conditions of GMM. Using historical data, we find considerable differences in the estimates from the non-overlapping samples that begin on different days.

Esben Hedegaard
W. P. Carey School of Business
Arizona State University
400 E. Lemon Street, BAC 518
Tempe, AZ 85287-3906
esben.hedegaard@asu.edu

Robert J. Hodrick
Graduate School of Business
Columbia University
3022 Broadway
New York, NY 10027
and NBER
rh169@columbia.edu

1 Introduction

Investors in theoretical financial models must hold assets for an abstract amount of time. The econometrician who desires to test the theory must decide on the time interval for which the implications of the theory are thought to hold. For example, Merton (1980, p. 336) proposed a one-month time interval as "not an unreasonable choice" for the horizon to examine the predictions of his (1973) continuous time intertemporal capital asset pricing model (CAPM). The conditional CAPM is the simplest version of that model, and it postulates a risk-return trade-off between the conditional expected return on the market portfolio and its conditional variance. To test the model, econometricians often sample the data at their chosen frequency to employ maximum likelihood estimation (MLE), whether the conditional variance of the market return is modeled as a generalized autoregressive conditional heteroskedasticity (GARCH) process or a mixed data sampling (MIDAS) model is used, even though higher frequency daily data are available.[1]

This paper demonstrates how to use all the available daily data in estimating models such as the conditional CAPM even if the specified horizon of the representative investor is greater than one day. We thus extend the analysis of Hansen and Hodrick (1980), who introduced the idea of what we call overlapping data inference (ODIN), to situations in which researchers typically sample the data to use MLE. The goals of this paper are to demonstrate this ODIN approach in the familiar setting of the risk-return trade-off for the market return implied by the conditional CAPM, to examine its potential biases, to analyze its ability to increase the power to reject the null hypothesis of no risk-return trade-off, and to provide new estimates of the risk-return trade-off in GARCH and MIDAS models. We find that the ODIN approach works well in our sample sizes and implies substantial improvements in power that translate into large increases in effective sample sizes of non-overlapping data.

---

[1]Lettau and Ludvigson (2010), and Nyberg (2012) provide extensive references to some of the vast empirical literature that investigates this conditional risk-return trade-off.

Whenever the data are sampled in non-overlapping intervals, alternative non-overlapping estimates can be generated by shifting the starting date. We compare the various estimates generated in this way to the one ODIN estimation that constrains them to be the same, and we find strikingly large variability in the non-overlapping estimates based on different starting dates.

While it would be quite difficult to formally model the conditional distributions of the errors within MLE when using all of the daily data and the one-month or one-quarter forecasting interval, our ODIN approach allows estimation with all of the data constraining the various non-overlapping samples to have the same coefficients. We do this by viewing the first order conditions of the MLE as orthogonality conditions in Hansen's (1982) Generalized Method of Moments (GMM) that should be satisfied simultaneously by the same set of parameters for each of the different starting days.

Ghysels, Santa-Clara, and Valkanov (2005) estimate the conditional CAPM risk-return trade-off with a MIDAS approach. They find positive, highly significant estimates of the trade-off in contrast to the prior empirical evidence of the trade-off, mostly employing various GARCH models, that can only be described as decidedly mixed, with some studies finding positive and significant estimates of the risk-return trade-off and others finding insignificant or negative coefficients. Intrigued by these MIDAS results and as a first step in thinking about how to extend MIDAS to conditional covariances and multivariate sources of risks, we attempted to replicate the results of Ghysels, Santa-Clara, and Valkanov (2005). When we were unable to do so, we contacted the authors and eventually discovered a bug in their program.[2] In thinking about how best to estimate these models, we discovered the ODIN approach.

Perhaps the diverse findings in the literature on the significance of the risk-return trade-

---

[2]Ghysels, Plazzi, and Valkanov (2013) provide a partial correction of the results in Ghysels, Santa-Clara, and Valkanov (2005) and extend it to include a discussion of the effects of financial crises on the risk-return trade-off.

off should be considered unsurprising in light of Lundblad's (2007) simulations that show the difficulty of finding a significant risk-return trade-off in the available sample sizes when such a trade-off actually exists. Lundblad (2007) focuses on the distribution of parameter estimates rather than on test statistics. Focusing on test statistics improves power a bit, and using ODIN improves power even more.

The plan of the paper is as follows. Section 2 motivates the study of the conditional CAPM, discusses the choice of horizon, notes that including a constant in the conditional mean is necessary to test the prediction of the model, and introduces the original MIDAS model as well as a modified version that allows simulations. Section 3 presents our ODIN methodology. Section 4 examines power issues in testing the conditional CAPM, first within the basic models and then in the ODIN versions. Section 5 examines estimation of the GARCH and MIDAS models with calendar-sampled data. Section 6 presents the estimation of the ODIN models, documents the differences between the results from the one ODIN model and those of the various non-overlapping monthly models generated by shifting the sampling start date, and examines the estimator that is the average of the non-overlapping estimates. Section 7 provides conclusions, and an Appendix provides some technical details. An Online Appendix contains additional figures and technical details.

2   The Conditional Risk-Return Trade-off

The conditional CAPM is the simplest version of Merton's (1973) model, and it implies a linear risk-return trade-off between the conditional mean of the market return, $E_t\left(R_{M,t+1}\right)$, and the conditional variance of the market return, $\sigma_{M,t}^2$, as in Merton's (1980) Model 1:

$$E_t\left(R_{M,t+1}\right) = \mu + \gamma\sigma_{M,t}^2 \tag{1}$$

We specify the model with a constant term, μ, for two reasons. Including a constant in the conditional mean is necessary to test the prediction of the conditional CAPM that the conditional mean of the market return is dynamically linked to its conditional variance, even though under the null hypothesis that the model is true, the constant is zero. Estimating without a constant simply relates the average future return to the average conditional variance, whereas if a constant is included, the estimate of γ will only be significantly different from zero if the covariance of the future return with its conditional variance is positive and significant.[3] This point is formally demonstrated in our online appendix, but it is intuitively clear from the analogous regression context.

The second reason to include a constant is that the model may be misspecified in which case the constant would capture the unconditional influences of other variables that would be present such as the conditional covariances of the return on the market with state variables. Scruggs (1998) and Guo and Whitelaw (2006) argue that one reason many studies fail to find a significant risk-return trade-off is such omitted variable bias.

2.1 The Choice of Horizon

When researchers test the risk-return trade-off or examine the holding period returns on any financial strategy, they must choose the horizon for which they think the theory holds. While daily returns are available, most of the existing literature prefers to test the conditional CAPM using longer horizon monthly or quarterly returns. Two considerations motivate this choice of horizon.

First, aspects of the trading process induced by market microstructure frictions, non-synchronous portfolio investment decisions, and individual stock illiquidity that are outside

---

[3]There is confusion in the literature on whether it is desirable to include a constant or not. Lanne and Saikkonen (2006) explicitly advocate estimating the conditional CAPM without a constant, and they find strong support for the conditional CAPM. Scruggs (1998) and Nyberg (2012) estimate both with and without a constant finding much higher significance of the risk-return trade-off without a constant.

the theory dominate the autocorrelations in short-horizon returns.[4] More importantly, when more volatile trading environments arise, theory predicts that stock returns are expected to be contemporaneously negatively correlated with the increase in volatility because prices must fall to provide an increase in expected returns, as in Campbell and Hentschel (1992). If the adjustment of expected returns to news that increases the conditional variance is not precisely contemporaneously correlated with the increase in the conditional variance because of market illiquidity or the non-synchronous trading of investors, using a short horizon for testing the conditional risk-return trade-off may problematically find a negative relation as volatility increases and asset prices fall slightly later. Thus, researchers use a longer horizon to balance the theoretical idea that there is a risk-return trade-off over a particular horizon against the loss of power that arises from sampling the data. Then, as noted above, many researchers sample the data to use MLE imposing the restrictions of the theory. ODIN modeling improves this situation by allowing the econometrician to use any forecast horizon while maximizing power from using all available data.

In what follows we use the term 'basic' to refer to models estimated with non-overlapping observations to distinguish them from ODIN models. Section 4 demonstrates that the power of the ODIN model is superior to that of the basic model in which the frequency of observations is the forecast interval. We also find that the improvement in power increases with the length of the sampling interval. Here, though, we rely on large samples in which the overlap is a small fraction of the total sample size, so this statement should not be extrapolated literally.

---

[4]Zhang, Mykland, and Aït-Sahalia (2005) discuss how to use all of the ultra high frequency intraday data that is clearly highly contaminated with market micro structure noise in estimating volatility models.

2.2 The MIDAS Framework

One of the challenges in estimating the risk-return trade-off is that the conditional variance is not observable, and a GARCH model is but one model for the conditional variance. Because GARCH models are extensively analyzed in the literature, we presume that the reader knows such specifications. MIDAS models are newer, and this section first explains the MIDAS model of the conditional variance introduced by Ghysels, Santa Clara, and Valkanov (2005) before explaining our version that we implement with the ODIN methodology.

2.2.1 The Original MIDAS Model

Ghysels, Santa Clara, and Valkanov (2005) model the monthly conditional variance as a weighted average of past squared daily excess returns on the market, where lower case r's refer to daily excess returns:[5]

$$\sigma_t^2 \equiv V_t^{MIDAS} = 22 \sum_{d=0}^{D} w_d r_{t-d}^2. \tag{2}$$

The weights on the past squared excess returns sum to one and are initially modeled as

$$w_d(\kappa_1, \kappa_2) = \frac{\exp(\kappa_1 d + \kappa_2 d^2)}{\sum_{j=0}^{D} \exp(\kappa_1 j + \kappa_2 j^2)}, d = 0, ..., D. \tag{3}$$

The number of days into the past, D, is set at 250, and multiplying by 22 in equation (2) scales the variance to a monthly value. The numerators of the weights and thus the shape of the distributed lag function are exponential functions of two parameters, κ_1 and κ_2, that are estimated. While this specification allows for considerable flexibility in the shape of the distributed lag, as long as $\kappa_2 < 0$, the weights die out eventually as the days get further into the past. Ghysels, Santa Clara, and Valkanov (2005) use MLE to estimate the four

---

[5]The mixture of monthly and daily data is the hallmark of the MIDAS approach. See Ghysels, Sinko, and Valkanov (2007) for an introduction to the MIDAS literature and examples of its use in other applications.

parameters of the model, μ, γ, κ_1 and κ_2, assuming that $R_{t+1} \sim N(E_t(R_{t+1}), V_t^{MIDAS})$, where the conditional mean and variance are given in equation (1) and (2). The approach is thus similar to a GARCH-in-mean (GARCH-M) model but with a different specification of the conditional variance.

Simulating from this MIDAS model produces unrealistic sample paths. Even though the unconditional variance of r_t is the same for all t, the conditional distribution of r_t becomes increasingly peaked as t grows large.[6] This leads to unrealistic data because most simulated paths essentially 'die out' as the distribution of returns becomes more peaked around zero until a tail-event occurs, and the variance increases dramatically.

2.2.2 An Alternative MIDAS Specification

It is easy to modify the original MIDAS specification of the conditional variance to eliminate this problem. Inspired by the usual GARCH model, let the conditional variance for a horizon of K days be

$$V_t^{MIDAS} = K \left(\omega + \phi \sum_{d=0}^{D} w_d(\theta) r_{t-d}^2 \right) \tag{4}$$

where $\omega > 0, 0 < \phi < 1$, and as above, the weights, w_d, sum to 1. The unconditional variance of the daily return process is $E(\sigma^2) = \frac{\omega}{1-\phi}$. Using this specification results in realistic simulated sample paths.

While the two-parameter exponential weight function in equation (3) provides flexibility in the weights applied to past squared returns, our experience with MIDAS indicated that

---

[6]To see this, consider the simplest example of a MIDAS model: The variance is estimated based on 1 past return and is used to describe only the next day. For simplicity, consider the case where $\mu = \gamma = 0$. With r_t denoting daily returns, the model is

$$r_{t+1} = z_{t+1} \sqrt{V_t^{MIDAS}}, \quad z_{t+1} \sim N(0, 1), \quad V_t^{MIDAS} = r_t^2.$$

Let $V_0^{MIDAS} = \sigma_0^2$ be given. We can then explicitly write out the sequence of returns and show that $r_t = z_t \prod_{s=1}^{t-1} |z_s| \sigma_0$. As the z_t's are independent, $E(r_t) = 0$ for all t, and $V(r_t) = E(r_t^2) = \sigma_0^2$. This shows that all returns have the same unconditional mean and variance. However, the distribution of r_t clearly changes with t, and a small shock will cause the conditional distribution to be very tight around zero.

estimates of κ_2 are often insignificantly different from 0 and are sometimes disturbingly positive, in which case the past weights do not die out.[7] An alternative weight function can be parameterized with the one-parameter beta-polynomial:

$$w_d(\kappa) = \frac{f(u_d, \kappa)}{\sum_{j=0}^{D} f(u_j, \kappa)}, \quad f(u_d, \kappa) = u_d^{(\kappa-1)}, \quad d = 0, \ldots, D \tag{5}$$

where $u_d = (1 - d/D)$ are the points where f is evaluated.[8] We also set $D = 500$, as we find that the choice of lag length influences the estimate of γ, but that the variation in $\widehat{\gamma}$ is very small once D is larger than 300. In this MIDAS specification, the variance process is parameterized with three parameters as in the GARCH(1,1) model.

3   Overlapping Data Inference with GMM

This section derives our ODIN estimation strategy for the conditional CAPM. Throughout, we maintain the usual assumption in the literature that monthly or quarterly returns satisfy the conditional CAPM. The Appendix describes construction of the overlapping monthly or quarterly excess returns.

Consider the ODIN-GARCH estimation first. Instead of presenting a general case, for ease of exposition and consistency with the past literature, we assume that the model holds at the monthly frequency. Thus, the GARCH-M model is

$$R_{t_m+1} = \mu + \gamma \sigma_{t_m}^2 + \varepsilon_{t_m+1}, \quad \varepsilon_{t_m+1} \sim N(0, \sigma_{t_m}^2) \tag{6}$$

where

$$\sigma_{t_m+1}^2 = \omega + \alpha \varepsilon_{t_m+1}^2 + \beta \sigma_{t_m}^2. \tag{7}$$

---

[7]One sees this problem in the new results of Ghysels, Plazzi, and Valkanov (2013).

[8]Ghysels, Sinko, and Valkanov (2007) discuss various weighting schemes including exponential Almon lags and beta functions. Our specification corresponds to a restricted version of their equation (4).

The time index $t_m = 0, \ldots, T_m - 1$ counts 22 day periods or approximately one month of trading days, and there are T_m of these monthly periods in the sample. Our notation distinguishes between monthly data indexed with t_m and daily data indexed with t. Capital R's represent monthly excess returns, and lower case r's represent daily returns. Hence, the excess monthly return, R_{t_m+1}, can also be represented as an excess return over 22 trading days, $R_{t+22,t}$.[9] Although the model in equations (6) and (7) is specified at the monthly frequency, we assume that only the number of days in the forecast matters, in which case the starting date for the month does not matter. Thus, we can write

$$R_{t+22,t} = \mu + \gamma \sigma^2_{t,t+22} + \varepsilon_{t+22,t}, \quad \varepsilon_{t+22,t} \sim N(0, \sigma^2_{t,t+22}) \tag{8}$$

where the notation indicates that the monthly model can also be written with daily subscripts, and $\varepsilon_{t+22,t}$ denotes the innovation in the monthly return realized on day $t + 22$. The conditional variance evolves as

$$\sigma^2_{t+22,t+44} = \omega + \alpha \varepsilon^2_{t+22,t} + \beta \sigma^2_{t,t+22}. \tag{9}$$

We assume that equations (8) and (9) hold *for all* $t = 0, 1, \ldots T - 1$.

For the MIDAS version of the model, monthly returns indexed by t_m are generated by

$$R_{t_m+1} = \mu + \gamma V_{t_m}^{MIDAS} + \varepsilon_{t_m+1}, \quad \varepsilon_{t_m+1} \sim N(0, V_{t_m}^{MIDAS}) \tag{10}$$

where

$$V_{t_m}^{MIDAS} = 22 \left(\omega + \phi \sum_{d=0}^{D} w_{t_m-d}(\kappa) r^2_{t_m,-d} \right) \tag{11}$$

and $r_{t_m,-d}$ is the daily return d days before the start of month t_m. Again, we assume that

---

[9]The first subscript of a variable denotes the day that the variable enters the information set, and the second subscript is a second day that indicates the period of time, either from the past for a return or into the future for a conditional variance, that is necessary to describe the variable.

equations (10) and (11) are true independently of the day we use as the starting date for the month. Thus, we have

$$R_{t+22,t} = \mu + \gamma V_{t,t+22}^{MIDAS} + \varepsilon_{t+22,t}, \quad \varepsilon_{t+22,t} \sim N(0, V_{t,t+22}^{MIDAS}) \tag{12}$$

where

$$V_{t,t+22}^{MIDAS} = 22 \left(\omega + \phi \sum_{d=0}^{D} w_{t-d}(\kappa) r_{t-d}^2 \right) \tag{13}$$

and equations (12) and (13) hold *for all* $t = 0, 1, \ldots T - 1$.

As a caveat to our analysis, it is not at all clear whether there exists a data generating process for daily returns that has the postulated return properties over the 'monthly' intervals. Our point is that if the model is viewed as an abstraction that holds at the monthly horizon better than it does at any other horizon, and if there is nothing particular about calendar months, the starting date becomes irrelevant. We thus have the opportunity to increase the power of the tests by using overlapping data.

3.1 ODIN-GARCH

This section derives the first order conditions from the MLEs of the monthly models that must hold for each starting date. We use Hansen's (1982) GMM to derive the asymptotic distribution of the parameter estimates that must satisfy the average across starting days of the monthly first order conditions. Because it is difficult to know how the serial correlation induced by the creation of overlapping observations affects small sample inference, we examine extensive simulations in the sections that follow.

The log likelihood function for the GARCH-M model with a monthly sampling interval

in equations (6) and (7) is

$$\log(L) = \sum_{t_m=0}^{T_m-1} \left(-\frac{1}{2}\log(2\pi) - \log(\sigma_{t_m}) - \frac{1}{2}\frac{\varepsilon_{t_m+1}^2}{\sigma_{t_m}^2} \right).$$

Rather than estimating ω as a free parameter, we estimate ω by variance targeting as in Engle and Kroner (1995).[10]

Cochrane (2005) notes that the first order conditions of MLE are the sample counterparts to unconditional moment conditions in GMM estimation. We think of these orthogonality conditions as holding for the 22 possible daily starting dates that index the months associated with t_m. By utilizing GMM, we can use all of the daily data with these orthogonality conditions and appropriately account for the serial correlation induced by the overlapping data in constructing the GMM standard errors.

Let the parameter vector be $\theta = (\mu, \gamma, \alpha, \beta)$. Then, using the first order conditions of the MLE gives the vector of sample orthogonality conditions with t indexing daily data:

$$G_T(R;\theta) = \frac{1}{T}\sum_{t=0}^{T-1} g_t(R_{t+1};\theta) = \begin{pmatrix} \frac{1}{T}\sum_{t=0}^{T-1}\left(\frac{\varepsilon_{t+22,t}}{\sigma_{t,t+22}^2} + \frac{1}{\sigma_{t,t+22}}\frac{\partial\sigma_{t,t+22}}{\partial\mu}\left(\frac{\varepsilon_{t+22,t}^2}{\sigma_{t,t+22}^2} + 2\gamma\varepsilon_{t+22,t} - 1 \right) \right) \\ \frac{1}{T}\sum_{t=0}^{T-1}\left(\varepsilon_{t+22,t} + \frac{1}{\sigma_{t,t+22}}\frac{\partial\sigma_{t,t+22}}{\partial\gamma}\left(\frac{\varepsilon_{t+22,t}^2}{\sigma_{t,t+22}^2} + 2\gamma\varepsilon_{t+22,t} - 1 \right) \right) \\ \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{\sigma_{t,t+22}}\frac{\partial\sigma_{t,t+22}}{\partial\alpha}\left(\frac{\varepsilon_{t+22,t}^2}{\sigma_{t,t+22}^2} + 2\gamma\varepsilon_{t+22,t} - 1 \right) \\ \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{\sigma_{t,t+22}}\frac{\partial\sigma_{t,t+22}}{\partial\beta}\left(\frac{\varepsilon_{t+22,t}^2}{\sigma_{t,t+22}^2} + 2\gamma\varepsilon_{t+22,t} - 1 \right) \end{pmatrix}$$
$$(14)$$

where $g_t(R_{t+1};\theta)$ denotes the vector of right-hand-side functions. Because the system of equations (14) is just identified, GMM chooses the parameter estimates, $\hat{\theta}$, to set $G_T(R,\hat{\theta}) =$

---

[10]Variance targeting guarantees that the unconditional estimate of the variance of a GARCH model equals the sample variance. Francq, Horvath, and Zakoian (2011) examine the econometric properties of this popular estimation strategy.

0. Intuitively, the parameters may be estimated by maximizing the average of the 22 monthly log-likelihood functions.

Let the gradient of the sample orthogonality conditions with respect to the parameters be

$$D_T(\hat{\theta}) = \nabla_\theta G_T(R; \hat{\theta}).$$

Then, the asymptotic distribution theory of Hansen's (1982) GMM implies that

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \to N\left(0, D_T(\hat{\theta})^{-1} S(\hat{\theta}) D_T^{\mathsf{T}}(\hat{\theta})^{-1}\right)$$

where

$$S(\hat{\theta}) = \sum_{j=-21}^{21} C_T\left(g_t(R_t; \hat{\theta}), g_{t-j}(R_{t-j}, \hat{\theta})^{\mathsf{T}}\right) \tag{15}$$

and the $C_T\left(g_t(R_t; \hat{\theta}), g_{t-j}(R_{t-j}; \hat{\theta})^{\mathsf{T}}\right)$ are the sample autocovariances of $g_t(R_t, \hat{\theta})$. Under the null hypothesis, these autocovariances will be non-zero until $j = 22$. Note that we estimate $S(\hat{\theta})$ by equally weighting the sample covariances, as in Hansen and Hodrick (1980).

3.2 ODIN-MIDAS

The log likelihood function for the MIDAS model specified at the monthly horizon is

$$\log(L) = \sum_{t_m=0}^{T_m-1}\left(-\frac{1}{2}\log(2\pi) - \log\left(\sqrt{V_{t_m}^{MIDAS}}\right) - \frac{1}{2}\frac{\varepsilon_{t_m+1}^2}{V_{t_m}^{MIDAS}}\right).$$

We again estimate ω by variance targeting and use the first order conditions as GMM orthogonality conditions as above. A noticeable difference is that the first order conditions for μ and γ, in contrast to the analogous equations for the GARCH model, do not involve derivatives of the conditional variances. This arises because in MIDAS the conditional variances depend on daily data, and there is no link in the MIDAS model between the conditional mean of

the low frequency returns and the conditional mean of the higher frequency returns, which is assumed to be effectively zero. Consequently, there is no indirect effect of conditional mean errors on the conditional mean parameters through their effect on the conditional variances.

4   Simulation Analysis of the Models

This section briefly discusses some of the most important findings from simulations of the models. Readers more interested in the estimation of the models can safely skip this section as our primary finding is that the asymptotic distributions of the ODIN methodology are appropriate in the sample sizes we use. Later, we nevertheless report p-values from the empirical distributions. We also refer the interested reader to the Online Appendix for a more complete discussion of the simulations. Our first simulations examine the power of the basic models before examining the increased power of the ODIN models.

4.1   Power in the Basic Monthly GARCH Model

The introduction notes that the previous evidence for a significantly positive risk-return trade-off, γ, is decidedly mixed. Lundblad (2007) demonstrates that negative estimates of γ sometimes reported in the literature are not inconsistent with the conditional CAPM being the true data-generating process in the sense that 19% of the estimates of γ fall below zero when simulating 500 months of returns from a GARCH-M model in which $\gamma = 2$ and the volatility process is calibrated to historical data.

Our first simulations address a related question. If $\gamma \neq 0$, how likely are we to reject the null of $\gamma = 0$ at a 5% marginal level of significance in particular non-overlapping sample sizes? This is a question about the power of the test of $\gamma = 0$ for different sample sizes against the alternative hypothesis that $\gamma = 1, 2$, or 3, for example.[11]   We simulate from

[11]We begin with a non-overlapping data framework to maintain consistency with the analysis in Lundblad

the GARCH-M model using the same parameters as Lundblad (2007) under the null and alternative hypotheses varying the sample length from $T = 500$ to 5,000. For each sample we do 5,000 simulations jointly estimating the parameters of the model, including a constant, μ, with MLE. With at least 500 observations, the bias in the MLE of γ is quite small, but the distribution is fairly spread out, as in Lundblad (2007). In particular, with $T = 500$, we confirm that 19% of the estimates are less than zero. We also find that the 95th-percentile of $\hat{\gamma}$ is 4.48 and that the 97.5-percentile is 6.03. Hence, with 500 observations, if one were to base a significance test strictly on the distribution of the coefficient estimate, one could argue that to reject the null hypothesis of $\gamma = 0$ at the 5% marginal level of significance one would need to observe $\hat{\gamma} > 6.03$ when using a two-sided test or $\hat{\gamma} > 4.48$ for a one-sided test, which is more relevant because we think that the true γ is positive. Increasing the sample size to 1,000 months, the 95th and 97.5-percentiles of $\hat{\gamma}$ under the null hypothesis decrease to 2.56 and 3.27, respectively. When simulating under the alternative hypothesis that $\gamma = 2$ with 500 months of data, only 9.2% of the samples have a $\hat{\gamma} > 6.03$, while with 1,000 months of data, 21.1% of the samples have a $\hat{\gamma} > 3.27$. Viewed in this way, the power of MLE in this environment is quite low.

Appropriately assessing the power of an estimation strategy, though, requires examination of a test statistic. Under the null hypothesis of $\gamma = 0$, the t-statistic is a pivot whose distribution does not depend on the actual underlying parameters and which should be asymptotically $N(0, 1)$. The Online Appendix shows QQ-plots of the t-statistics against the quantiles of a standard normal distribution, and with at least 500 observations we find that convergence of the test statistic is excellent as the QQ-plot virtually overlays the diagonal line.[12] Thus, the 95% confidence interval of the t-test is approximately $(-1.96, 1.96)$ for these

---

(2007). Then, in Section 4.3, we show that using the ODIN framework increases the power of the t-test even more.

[12]Rather than present these figures, we have placed them in an Online Appendix. The standard errors are calculated using the quasi-maximum likelihood (QMLE) approach of Bollerslev and Wooldridge (1992). We take care in evaluating the Hessian of the likelihood function. In particular, the Hessian returned from

sample sizes. Under the alternative hypotheses that $\gamma = 1, 2,$ or 3, the t-statistic follows a non-central t-distribution where the non-centrality parameter depends on the value of γ and the sample size. As the sample size increases, the distribution moves farther to the right, and the power of the test increases.

As expected, the power of the t-test is always higher than the power of the test based on coefficient estimates. For a 500 month sample, using the t-test roughly doubles the power compared to using the point estimates. For the lower power test to have the same power as the t-test requires approximately an additional 1,200 observations when $\gamma = 1$, about 950 observations when $\gamma = 2$, and about 800 observations when $\gamma = 3$. For 1,000 months, the increase in power is a little less, but it is still substantial. Because Lundblad (2007) focuses directly on the distributions of $\hat{\gamma}$, he somewhat overstates the difficulty of rejecting the null hypothesis of $\gamma = 0$. We nevertheless agree with his central point: If the true risk-return trade-off is $\gamma = 2$, with 1,000 months of data, we only have a 30% chance of rejecting the false null of $\gamma = 0$ (up from 21% if one uses coefficient estimates instead of the t-statistic). If $\gamma = 1$, power drops to 11%, while if $\gamma = 3$, power increases to 57%. Section 4.3 demonstrates that using the ODIN framework increases the power of the tests, and hence the effective sample size, even more.

4.2   Power Analysis in the Basic Monthly MIDAS Model

This section asks whether a MIDAS model provides more power than a GARCH model when estimating the risk-return trade-off. The power of the test depends on the volatility-of-volatility of the estimated volatility process, which is clear from the linear regression analogy where variation in the right-hand-side variable increases the precision of the estimation. As

the MatLab optimization routine is not reliable, as it is a so-called 'pseudo-Hessian' constructed with the purpose of choosing sensible step-sizes, not to be a high-precision estimate of the second derivatives. Instead, we use the DERIVEST suite by D'Errico (2011), an adaptive numerical differentiation toolbox that provides high-precision first-order and second-order derivatives.

above, all simulations of the MIDAS model have $\mu = 0$ in the data generating process, but μ is estimated with the other parameters using MLE. We examined the power of the test of the null hypothesis, $\gamma = 0$, against different alternatives in the MIDAS model for two sets of parameters.[13] The first set of parameters were chosen to match the volatility-of-volatility in the GARCH model and use $\phi = 0.87$ and $\kappa = 21$. In this case, the power of the MIDAS model is similar to that of the GARCH model. The second set of parameters are associated with the estimation of the value-weighted market return over 1927-2011. In this case, the volatility-of-volatility is much lower as $\phi = 0.69$ and $\kappa = 5.55$. The smaller values of ϕ and κ decrease the implied volatility-of-volatility, directly with a lower ϕ but also because a lower κ makes the beta-polynomial weight function flatter, which 'averages' over a longer history of past daily squared returns in producing monthly volatility. With these parameters, the power of the test fell to be less than 0.1 for sample sizes less than 2,000 and $\gamma = 1, 2, 3$. This shows that using the MIDAS model to estimate conditional volatility does not necessarily increase the precision of the estimation of the risk-return trade-off.

4.3   Power Analysis of ODIN Models

ODIN estimation shares a lot with the basic estimation method. Because the ODIN estimator specified at the monthly horizon maximizes the average of 22 likelihood functions based on the different starting dates, the probability limit of the ODIN estimator is the same as the probability limit of the basic estimator. The asymptotic variance of the ODIN estimator is however always smaller than the asymptotic variance of the basic estimator. The asymptotic variance of the monthly ODIN estimator in equation (15) depends on $S = \sum_{j=-21}^{21} C_T(j)$, where $C_T(j)$ is the sample autocovariances of g_t with g_{t-j}. In the Online Appendix we show that the corresponding matrix in the asymptotic variance of the basic monthly model is

---

[13]The Online Appendix contains QQ-plots of the simulated t-statistics against the quantiles of a standard normal distribution under the null of $\gamma = 0$. To calculate power under the alternative hypotheses, we calculate the percentages of the observations that fall outside the 95% confidence intervals under the null.

$S = 22C_T(0)$. Hansen and Hodrick (1980) demonstrate that $S = \sum_{j=-21}^{21} C_T(j)$ is less than $22C_T(0)$ by a positive definite matrix, and it follows that the asymptotic variance of the ODIN estimator is smaller than the asymptotic variance of the basic estimator.

While these asymptotic results are interesting, to assess the performance of ODIN estimation on historical sample sizes we simulate 22,000 days or 1,000 months of data from a continuous time GARCH model. Following Nelson (1990), Drost and Nijman (1993), Andersen and Bollerslev (1998), and Lundblad (2007) the continuous-time limit for a GARCH(1,1)-M is specified as

$$\frac{dP_t}{P_t} = \gamma \sigma_t^2 dt + \sigma_t dW_{P,t}$$

$$d\sigma_t^2 = \theta(\omega - \sigma_t^2)dt + \sqrt{2\lambda\theta}\,\sigma_t^2 dW_{\sigma,t}$$

where P_t is the market price level, σ_t^2 is the stochastic instantaneous variance process, and $W_{P,t}$ and W_{σ_t} are independent Brownian motions. Andersen and Bollerslev (1998) derive the mapping between discrete time GARCH parameters estimated on monthly data ($\omega = 0.0002, \alpha = 0.10, \beta = 0.85$), and the continuous-time parameters ($\theta = 0.0023, \omega = 1.8182 \cdot 10^{-4}, \lambda = 0.459$, assuming 22 trading days per month), as in Lundblad (2007). We then simulate from the continuous time model in 5-minute increments using a standard Euler scheme and different values of γ. Finally, we sample the process to get daily log prices and compute daily log returns. Summing these daily log returns gives log returns for any forecasting horizon, and these returns satisfy a weak GARCH model. The Online Appendix contains further details on the simulations. We then estimate basic GARCH-M models and ODIN GARCH-M models with forecasting horizons of one, five, ten, 22, 33, 44, 55, and 66 days. In the basic model, the sampling frequency is the same as the forecast horizon, whereas the ODIN-GARCH model always uses all available daily data. Although no analytical results are available for QMLE or GMM applied to weak GARCH models, Drost and Nijman (1993)

show that the asymptotic bias of the QMLE estimates is small, which we confirm is also the case for the ODIN estimator.

Across 10,000 simulations, both for the null hypothesis, $\gamma = 0$, and for three alternative hypotheses, $\gamma = 1, 2$, and 3, we find that $\hat{\gamma}$ is only slightly biased. The Online Appendix shows QQ-plots of $\hat{\gamma}$ against the quantiles of a normal distribution. For all sampling frequencies and all values of γ, the distributions of $\hat{\gamma}$ for the ODIN model are closer to a normal distribution than is the distributions of $\hat{\gamma}$ for the basic model. Further, the empirical means of $\hat{\gamma}$ based on 1,000 months of data are much closer to the large-sample means for the ODIN model, suggesting that the small-sample bias is larger for the basic model than for the ODIN model.

The simulations of the ODIN methodology indicate that decreases in standard errors and increases in power can be substantial and correspond to large increases in the sample size. For example, with a 1,000 month sample of daily data, a horizon of 22 days, and $\gamma = 3$; the average standard error for the ODIN-GARCH model is 84.8% of its basic counterpart. Because standard errors decrease linearly in the square root of the sample length, a 15.2% reduction in the standard error corresponds to a 38.9% increase in the sample length, which is effectively an additional 389 months of non-overlapping data or more than 32 years. For quarterly horizons and quarterly sampling of the data in the basic model, we find that ODIN cuts standard errors by approximately 30%, which corresponds to more than doubling the non-overlapping sample length. See the Online Appendix for more details.

4.4   A Caveat on the Sample Size

The previous discussion could leave the reader with the impression that ODIN is a free lunch. One can increase power and not suffer any ill consequences. But, we use only relatively long samples in which the overlap remains a small fraction of the sample size. We know from Richardson and Stock (1989) and Valkanov (2003) that building up highly serially correlated error processes can cause the standard asymptotic distribution theory underlying

test statistics to provide poor approximations if the sample size is not sufficiently large. We have not worked out the asymptotic distribution theory associated with the functional central limit theorem discussed in these papers, but this is a worthwhile idea. With this caveat in mind, we turn to the evidence from actual data.

5   Estimating the Basic Models

This section presents estimates for the basic GARCH and MIDAS models for three samples of monthly and quarterly non-overlapping calendar data. Table 1 presents the GARCH estimation, and Table 2 presents the MIDAS estimation as described in equations (4) and (5). As above, we estimate ω by variance targeting.

We use the same observations on the dependent variable, R_{t+1}, for the GARCH and MIDAS models. Although CRSP data begin in 1926, the full sample for monthly data is 1927:10 to 2011:12 because of the lags introduced in the MIDAS model. The corresponding full sample for quarterly data is 1927:4 to 2011:4. We also split the sample after 1952 to recognize that the Great Depression, World War II, and the lack of Federal Reserve independence prior to the Treasury-Fed Accord of 1952 may have produced data that require more complex modeling than the conditional CAPM. The second monthly sample is consequently 1927:10 to 1952:12, and its quarterly counterpart is 1927:4 to 1952:4. To avoid use of data from before 1952, the third monthly sample is 1955:1 to 2011:12, and its quarterly counterpart is 1955:1 to 2011:4.

The tables report the estimates of γ, with the QMLE asymptotic standard errors in parenthesis, and the associated p-values of the t-tests of $\gamma = 0$. We also report bootstrap p-values under the null hypothesis that the true coefficient is 0 (see Appendix B for details).

In Table 1, for the monthly GARCH model, the $\hat{\gamma}$'s range from .44 to 3.01, but only the latter estimate from the post 1952 sample has a bootstrap p-value less than .10. With

19

quarterly data, the GARCH model $\hat{\gamma}$'s range from 1.75 to 6.34, and each of the values is significantly different from zero at the 10% marginal level of significance from the bootstrapped p-values. Note, though, that for the early quarterly sample from 1927-1952, $\hat{\alpha}$ is unusually high, and $\hat{\beta}$ is unusually low.

For the MIDAS results in Table 2, the $\hat{\gamma}$'s do not approach statistical significance at traditional levels. The $\hat{\gamma}$'s for the monthly full sample and the first sub-sample are negative, and the $\hat{\gamma}$ for the second sub-sample has a bootstrap p-value that is .904. With quarterly data, only the $\hat{\gamma}$ of 3.56 in the second sub-sample approaches traditional statistical significance with a bootstrap p-value of .17.

The estimates $\kappa > 1$ for the MIDAS models imply that the weights on past squared return in the conditional variance estimate decline as the daily returns move further into the past. While monotonically declining weights make intuitive sense, nothing in the data inherently makes the weight functions decline. Using the exponential weight function in equation (3) for our samples produces U-shaped weight functions. Here, we use the beta-polynomial with one parameter which ensures monotonically declining weights. The point estimate of κ imply that the importance of the recent past is much higher in the second sub-sample, for which κ is higher, than in the whole sample or the first sub-sample.

The Online Appendix includes a figure showing the estimates of the conditional standard deviations for the GARCH and MIDAS models. For the full sample period, 1927-2011, the two estimates are very similar although the volatility-of-volatility is smaller for the MIDAS model. For the first sub-period 1927-1952, the conditional standard deviation in the MIDAS model is smoother and reacts less quickly to new innovations, due to a very flat weight function. On the other hand, for the second sub-period 1955-2011, the more steeply declining weights allow the conditional standard deviation from the MIDAS model to be much more 'choppy' than that of the GARCH model, which is seen most clearly in the 2004-2010 period.

Table 1: GARCH Estimation Results and Bootstrapped p-values

Panel A: Monthly GARCH(1,1)-M model							
Period	μ	γ	$\omega \times 10,000$	α	β	Obs	LLF
1927:10–2011:12	0.005	1.331	0.764	0.129	0.846	1011	1653.61
Standard error	(0.002)	(0.917)	(0.260)	(0.021)	(0.026)		
p-value	.029	.146	.003	.000	.000		
Bootstrap p		.176					
1927:10–1952:12	0.010	0.443	0.707	0.141	0.846	303	419.91
Standard error	(0.004)	(1.017)	(0.468)	(0.034)	(0.039)		
p-value	.012	.663	.131	.000	.000		
Bootstrap p		.725					
1955:1–2011:12	0.001	3.011	1.028	0.105	0.843	684	1188.81
Standard error	(0.004)	(1.903)	(0.404)	(0.028)	(0.037)		
p-value	.801	.114	.011	.000	.000		
Bootstrap p		.083					
Panel B: Quarterly GARCH(1,1)-M model							
Period	μ	γ	$\omega \times 10,000$	α	β	Obs	LLF
1927:12–2011:12	0.006	1.747	16.930	0.334	0.536	337	321.24
Standard error	(0.011)	(1.008)	(8.678)	(0.085)	(0.137)		
p-value	.570	.083	.051	.000	.000		
Bootstrap p		.090					
1927:12–1952:12	−0.012	1.884	58.054	0.690	0.094	101	69.98
Standard error	(0.020)	(0.401)	(19.197)	(0.072)	(0.072)		
p-value	.554	.000	.002	.000	.189		
Bootstrap p		.012					
1955:3–2011:12	−0.025	6.337	17.208	0.178	0.588	228	246.31
Standard error	(0.043)	(5.320)	(7.183)	(0.104)	(0.064)		
p-value	.565	.234	.017	.086	.000		
Bootstrap p		.098					

Note: The table shows estimation results for the Basic GARCH model, using non-overlapping monthly and quarterly returns. The bootstrapped p-values are based on 5,000 simulations with $\gamma = 0$, keeping the remaining parameters at their estimated values.

Table 2: MIDAS Estimation Results and Bootstrapped p-values

Panel A: Monthly Alternative MIDAS model with β-weights							
Period	μ	γ	$\omega \times 10{,}000$	ϕ	κ	Obs	LLF
1927:10–2011:12	0.006	−0.149	0.351	0.693	5.550	1011	1655.76
Standard error	(0.003)	(1.318)	(0.057)	(0.050)	(3.433)		
p-value	.015	.910	.000	.000	.106		
Bootstrap p		.789					
1927:10–1952:12	0.013	−1.544	0.304	0.811	2.929	303	420.14
Standard error	(0.004)	(1.693)	(0.090)	(0.056)	(1.137)		
p-value	.004	.362	.001	.000	.010		
Bootstrap p		.285					
1955:1–2011:12	0.005	0.209	0.396	0.557	43.312	684	1204.40
Standard error	(0.003)	(1.694)	(0.062)	(0.070)	(17.418)		
p-value	.102	.902	.000	.000	.013		
Bootstrap p		.904					
Panel B: Quarterly Alternative MIDAS model with β-weights							
Period	μ	γ	$\omega \times 10{,}000$	ϕ	κ	Obs	LLF
1927:12–2011:12	0.015	0.620	0.401	0.650	6.427	337	320.18
Standard error	(0.012)	(2.044)	(0.115)	(0.100)	(7.859)		
p-value	.222	.762	.000	.000	.414		
Bootstrap p		.774					
1927:12–1952:12	0.036	−1.111	0.165	0.897	3.365	101	70.41
Standard error	(0.013)	(2.017)	(0.174)	(0.108)	(1.312)		
p-value	.007	.582	.344	.000	.010		
Bootstrap p		.390					
1955:3–2011:12	−0.004	3.564	0.488	0.454	9.431	228	244.96
Standard error	(0.017)	(3.016)	(0.134)	(0.150)	(4.750)		
p-value	.816	.237	.000	.003	.047		
Bootstrap p		.172					

Note: The table presents estimation results for the alternative MIDAS specification of the conditional variance in equation (4) with beta-polynomial weights as in equation (5). The estimates of μ have been multiplied by 22 (66) for the monthly (quarterly) results to make them comparable to the GARCH estimates. The bootstrapped p-values are based on 5,000 simulations with $\gamma = 0$, keeping the remaining parameters as their estimated values.

6 Estimating the ODIN Models

The results of the ODIN-GARCH estimation are presented in Table 3 with the results for monthly and quarterly data in Panels A and B, respectively. Table 4 presents the comparable ODIN-MIDAS results. Each panel contains results for the three sample periods examined above: the full sample, 1927:10-2011:12; the first sub-sample, 1927:10-1952:12; and the second sub-sample, 1955:1-2011:12. Standard errors are presented in parenthesis below the point estimates with p-values below the standard errors.

In Table 3, the $\hat{\gamma}$'s for the monthly samples are similar to the basic GARCH model. The $\hat{\gamma}$ for the full sample ODIN model is 1.678 with a p-value of .210 compared to a $\hat{\gamma}$ of 1.331 with a p-value of .146 for the basic model. The $\hat{\gamma}$ for the first sub-sample ODIN model is .654 with a p-value of .684 compared to a $\hat{\gamma}$ of .443 with a p-value of .663 for the basic model. The $\hat{\gamma}$ for the second sub-sample ODIN model is now 3.354 with a p-value of .022 compared to 3.011 and .114 for the basic model.

For the quarterly results, the $\hat{\gamma}$ in the full sample decreases from 1.747 to 1.572, and the p-value increases from .083 to .097. The p-value for $\hat{\gamma}$ in the first sub-sample increases substantially from .012 to .666, while the estimates of α and β now have the more traditional values expected in a GARCH (1,1) model. This suggests that the calendar results from the first sub-sample are spuriously significant. The $\hat{\gamma}$ for the second sub-sample ODIN model is now 4.894 with a p-value of .213 compared to 6.337 and .234 for the basic model.

The results for the ODIN-MIDAS model show some improvements in the standard errors for the quarterly data relative to the results in Table 2, but the standard errors of the monthly models are actually larger for the two sub-samples. None of the $\hat{\gamma}$'s has a p-value smaller than .334. The first sub-sample ODIN model $\hat{\gamma}'s$ for both the monthly and quarterly samples are negative. For monthly data, the full sample $\hat{\gamma}$ is now positive, but for quarterly data, the full sample $\hat{\gamma}$ is now negative. The estimate of γ that is the most comparable to

Table 3: Estimation Results for ODIN GARCH Model.

Panel A: Monthly ODIN-GARCH						
Period	μ	γ	α	β	Obs	LLF
1927:9:30–2011:12:8	0.004	1.678	0.122	0.841	22220	35509.63
Standard error	(0.003)	(1.340)	(0.024)	(0.033)		
p-value	.159	.210	.000	.000		
1927:9:30–1952:12:12	0.008	0.654	0.109	0.869	7370	10374.01
Standard error	(0.005)	(1.608)	(0.034)	(0.044)		
p-value	.136	.684	.001	.000		
1955:1:26–2011:12:13	0.001	3.354	0.133	0.783	14300	24114.36
Standard error	(0.004)	(1.464)	(0.026)	(0.037)		
p-value	.853	.022	.000	.000		
Panel B: Quarterly ODIN-GARCH						
Period	μ	γ	α	β	Obs	LLF
1927:9:30–2011:12:8	0.011	1.572	0.220	0.655	22176	21532.11
Standard error	(0.009)	(0.947)	(0.062)	(0.099)		
p-value	.244	.097	.000	.000		
1927:9:30–1952:12:12	0.022	0.439	0.133	0.856	7326	5540.65
Standard error	(0.014)	(1.017)	(0.066)	(0.092)		
p-value	.111	.666	.043	.000		
1955:1:26–2011:12:13	-0.013	4.894	0.170	0.500	14256	15508.53
Standard error	(0.031)	(3.931)	(0.101)	(0.098)		
p-value	.687	.213	.092	.000		

Note: The table presents results from estimation of the GMM estimation of the ODIN-GARCH model specified in the orthogonality conditions of equation (14). Standard errors are in parenthesis with p-values below.

Table 4: Estimation Results for ODIN MIDAS Model.

Panel A: Monthly ODIN-MIDAS						
Period	μ	γ	ϕ	κ	Obs	LLF
1927:9:30–2011:12:8	0.005	0.483	0.656	21.520	22220	35721.81
Standard error	(0.002)	(1.230)	(0.058)	(13.577)		
p-value	.048	.695	.000	.113		
1927:9:30–1952:12:12	0.012	−1.359	0.804	3.319	7370	10562.47
Standard error	(0.005)	(1.741)	(0.065)	(1.388)		
p-value	.010	.435	.000	.017		
1955:1:26–2011:12:13	0.004	0.868	0.534	64.657	14300	24311.54
Standard error	(0.003)	(1.708)	(0.064)	(27.948)		
p-value	.219	.611	.000	.021		
Panel A: Quarterly ODIN-MIDAS						
Period	μ	γ	ϕ	κ	Obs	LLF
1927:9:30–2011:12:8	0.019	−0.011	0.620	5.035	22176	21917.19
Standard error	(0.009)	(1.502)	(0.083)	(1.383)		
p-value	.046	.994	.000	.000		
1927:9:30–1952:12:12	0.036	−1.331	0.776	3.571	7326	5811.72
Standard error	(0.015)	(1.812)	(0.115)	(1.264)		
p-value	0.016	0.463	0.000	0.005		
1955:1:26–2011:12:13	0.001	2.619	0.454	7.762	14256	15587.65
Standard error	(0.015)	(2.710)	(0.152)	(2.411)		
p-value	.954	.334	.003	.001		

Note: The table presents results from the GMM estimation of the ODIN-MIDAS model. The values of μ in the MIDAS model have been multiplied by 22 (66) to measure them in monthly (quarterly) values, making them comparable to the estimates in the GARCH model. Standard errors are in parenthesis with p-values below.

the ODIN-GARCH model is the second sub-sample in which $\hat{\gamma} = 2.619$ with a p-value of .334.

We next examine the differences in the individual estimates of the various non-overlapping samples to determine whether the reason we do not see uniform improvement in the standard errors of the ODIN models relative to their basic counterparts can be traced to instability in the parameters.[14]

6.1   Individual Estimations vs. ODIN

Realizing that the starting date of the sample does not matter raises the question of how different can the various non-overlapping estimates be. The surprising answer is, quite different. Figures 1, 2, and 3 illustrate the differences between the 22 possible basic, non-overlapping, monthly estimations and the ODIN estimations for the three sample periods: the full sample, labeled A; the first sub-sample, labeled B; and the second sub-sample, labeled C.

Figure 1 shows the results for the full sample, 1927–2011. The top four plots present the estimates of the four parameters in the GARCH model, and the bottom four plots present the results for the MIDAS model. The solid black line shows the 22 point estimates from estimating the basic GARCH model on 22 day returns, shifting the sampling start date one day at a time. The shaded gray area shows the 95% confidence intervals for the estimates. These 22 estimates are obviously not independent as they are based on the same data, which are just sampled differently, and as expected the estimate moves slowly with the sampling start date. Next, the 'Basic' estimates represent the parameter estimates from the basic models estimated from monthly calendar returns, with horizontal lines showing the 95%

---

[14]It is also possible to do ODIN estimation with alternative conditional density functions, such as Student's t-distribution. We did not find that allowing for this alternative specification improved the performance of the conditional CAPM. These results are available in the Online Appendix, which also presents results for models in which the conditional variance responds asymmetrically to the innovation in returns. When we allow for this asymmetry, we find that the risk-return trade-off is imprecisely estimated.

confidence intervals. Finally, the ODIN estimate is shown together with its 95% confidence interval.

The top right plot presents $\hat{\gamma}$ in the GARCH model, and the (3,2) plot presents $\hat{\gamma}$ in the MIDAS model. There is clearly a negative correlation between $\hat{\mu}$ and $\hat{\gamma}$ for both the GARCH and MIDAS models as the starting date varies, which arises because increasing either μ or γ increases the unconditional return in the model. The individual $\hat{\gamma}$'s vary from 1 to 3 for the GARCH model, and between -0.1 and 0.8 for the MIDAS model. For the GARCH model, one sees a negative correlation between $\hat{\alpha}$ and $\hat{\beta}$, and for the MIDAS model, one sees a negative correlation between $\hat{\phi}$ and $\hat{\kappa}$ because increasing either of these parameters gives rise to higher volatility-of-volatility. None of the 22 individual $\hat{\gamma}$'s are significantly different from zero, for either the GARCH or the MIDAS model. For the GARCH model, the ODIN standard error of $\hat{\gamma}$ is actually larger than that of the basic monthly GARCH model. This situation arises because of the variation in $\hat{\gamma}$ that comes from varying the starting date. For some starting dates, the basic non-overlapping standard error is much larger than for other starting dates. While the basic monthly model does not 'see' this variation, the ODIN model recognizes the variation that comes from changing the starting date resulting in a larger standard error. Also, note that the point estimate from the ODIN model is closer to the average value of the individual estimates than is the point estimate for the basic monthly model.

Figure 2 presents the results for the first sub-sample, 1927–1952. Again, none of the individual $\hat{\gamma}$'s are significantly different from zero, for either of the models. The $\hat{\gamma}$'s vary between -0.1 and 2.4 for the GARCH model, and between -1.5 and -0.8 for the MIDAS model. As for the full sample, the standard error of $\hat{\gamma}$ is actually larger in the ODIN GARCH model than for the basic GARCH model, again due to the variation in the $\hat{\gamma}$'s and their standard errors as the starting date changes.

Figure 3 presents results for the second sub-sample, 1955–2011. For the GARCH model,
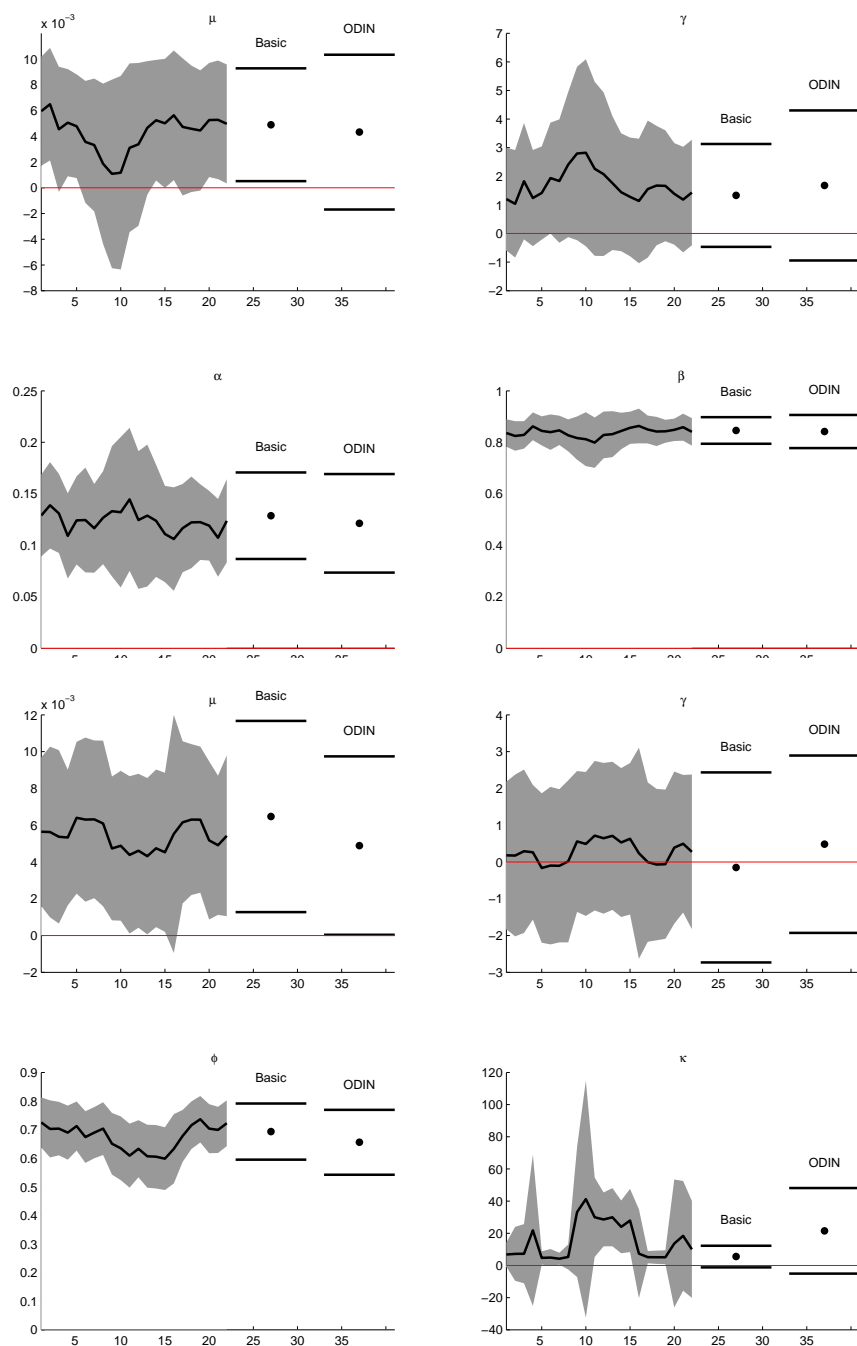
Figure 1: Monthly Estimates, Sample A: 1927–2011. The plots show the 22 individual estimates obtained by shifting the start-date and their 95% confidence interval in shaded grey. The top four plots show the GARCH estimates, the bottom four plots show the MIDAS estimates. The 'basic monthly' estimate is based on calendar months and the 95% confidence interval is indicated with horizontal lines. Finally, the ODIN estimate is shown along with its 95% confidence interval.
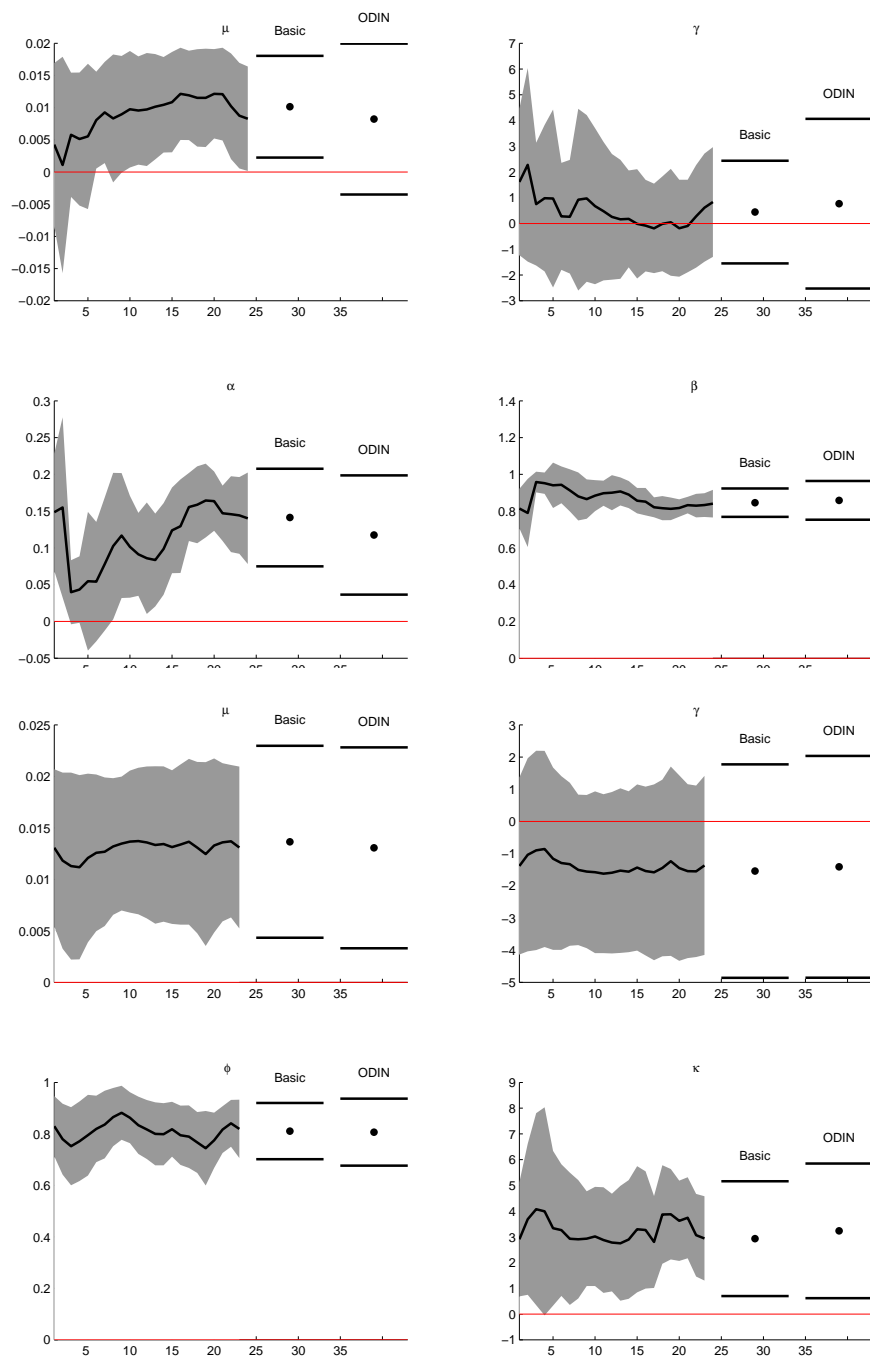
28

Figure 2: Monthly Estimates, Sample B: 1927–1952. The plots show the 22 individual estimates obtained by shifting the start-date and their 95% confidence interval in shaded grey. The top four plots show the GARCH estimates, the bottom four plots show the MIDAS estimates. The 'basic monthly' estimate is based on calendar months and the 95% confidence interval is indicated with horizontal lines. Finally, the ODIN estimate is shown along with its 95% confidence interval.

most of the individual $\hat{\gamma}$'s are significantly different from zero, even though the estimate from the basic model estimated on calendar month returns happens not to be significantly different from zero. Nevertheless, the ODIN estimation recognizes that it is highly unlikely that the true γ is zero because so many of the individual estimates are significantly different from zero. For the MIDAS model, none of the individual $\hat{\gamma}$'s are significantly different from zero.

Note that there are some similarities between the GARCH and MIDAS estimates. The $\hat{\gamma}$'s are lowest for sample B, slightly higher for sample A, and highest for sample C. The Online Appendix shows similar plots for quarterly estimations. As might be anticipated from the variation in the monthly samples, the variation in the quarterly estimates as the sample starting date changes is much larger for the quarterly estimations. For instance, for the second sub-sample, the GARCH $\hat{\gamma}$'s vary between 2 and 20, and the MIDAS $\hat{\gamma}$'s vary between 0 and 14. These plots also indicate that the basic models fail to converge for some of the starting dates.

6.2 The Average of the Individual Estimates

ODIN constrains the various non-overlapping estimates to have the same value. One could also consider the estimator that is the average of the individual estimates from the correlated non-overlapping samples. In the Online Appendix, we demonstrate that the standard error of this estimator is the same asymptotically as the standard error for the ODIN estimator. For monthly data, the averages of the $\hat{\gamma}$'s for samples A, B, and C, with the ODIN standard errors in parenthesis for the basic GARCH model are 1.699 (1.340), 0.498 (1.608), and 3.302 (1.464), respectively. For the basic MIDAS model, the averages of the $\hat{\gamma}$'s for samples A, B, and C, with the ODIN standard errors in parenthesis, are 0.277 (1.230), -1.396 (1.741), and 0.768 (1.708), respectively.
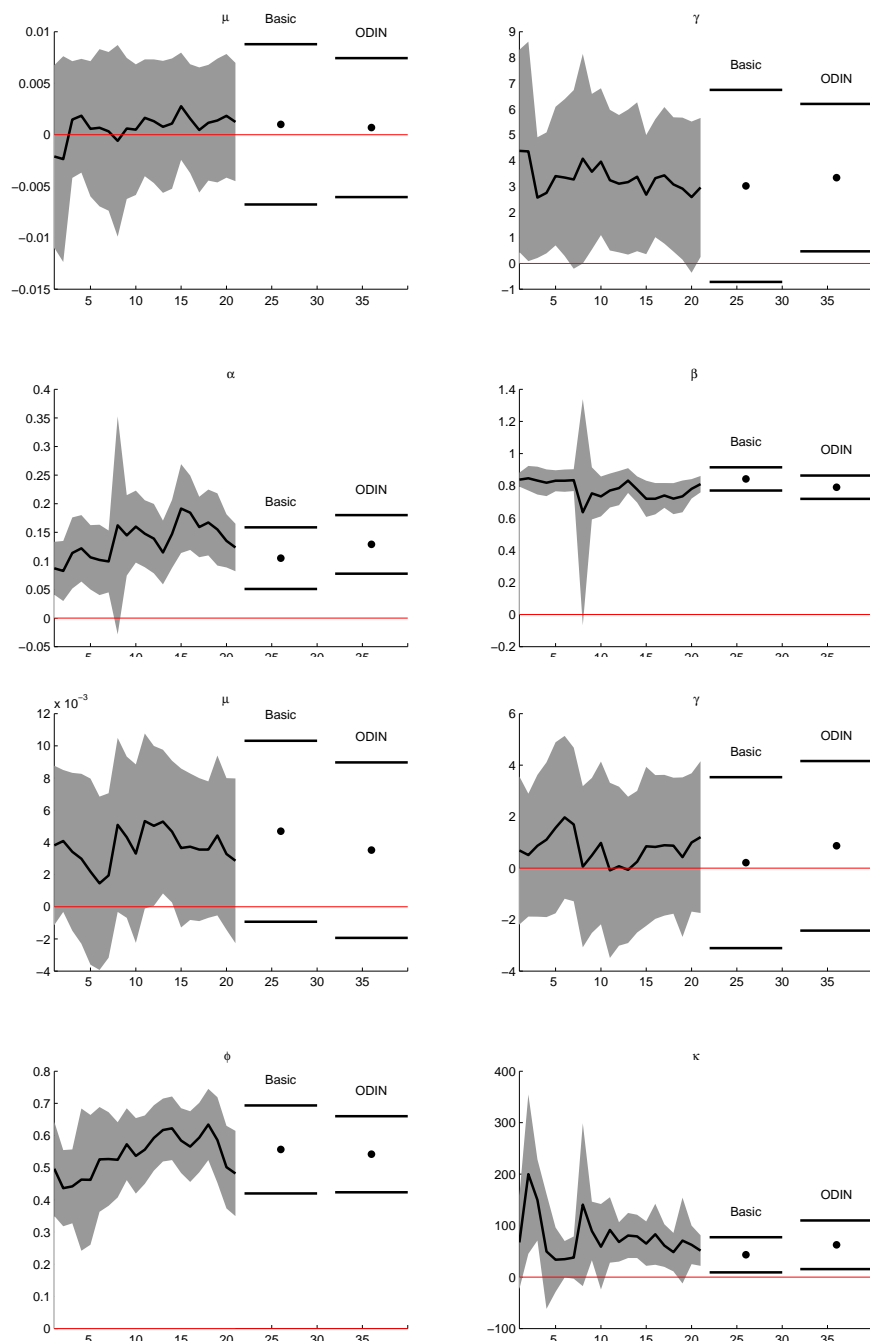
Figure 3: Monthly Estimates, Sample C: 1955–2011. The plots show the 22 individual estimates obtained by shifting the start-date and their 95% confidence interval in shaded grey. The 'basic monthly' estimate is based on calendar months and the 95% confidence interval is indicated with horizontal lines. Finally, the ODIN estimate is shown along with its 95% confidence interval.

31

7 Conclusions

When financial economists empirically investigate the predictions of their models, they must choose the horizon over which the agents in the model hold their investments. For example, Merton's (1973) ICAPM is a theoretical continuous time model, but empirical researchers usually choose a one-month or one-quarter horizon as the most appropriate test environment even though daily data are available. The most popular methods for modeling the conditional variances and covariances that are the sources of risk in these models are GARCH and MIDAS, which are usually implemented with MLE by sampling the data at the same frequency as the horizon chosen for the model. Here we demonstrate that when the data are sampled more finely than the horizon of the model, we can use all of the available data to lower the standard errors of the estimates and improve the power of the tests of the theories by using overlapping data inference (ODIN). Our insight is to use the first order conditions of MLE as orthogonality conditions of GMM. We estimate the parameters of the model from the average of the overlapping MLE samples and construct appropriate standard errors that account for the serial correlation induced by the overlapping data.

We apply this ODIN methodology to investigate the risk-return trade-off implied by the conditional CAPM using GARCH and MIDAS modeling of the conditional variance of the market return. Simulations of the ODIN methodology indicate that if the true model were the conditional CAPM with a one-month horizon, the ODIN approach would be substantially more powerful than the basic approach. When we examine actual data, the basic GARCH approach produces a positive conditional risk-return trade-off for the sample period 1955:1 to 2011:12 that has a p-value of .08. When we use the ODIN methodology, the p-value falls to .022. As with much of the literature, though, we find insignificant or even negative trade-offs in other samples and with asymmetric responses to shocks.

Yu and Yuan (2011) use Baker and Wurgler's (2006) measure of investor sentiment and

find a positive risk-return trade-off in low sentiment periods but not high ones. Their analysis could be reexamined with ODIN.

Of course, the conditional CAPM is the simplest specification of the ICAPM. Many authors, including Campbell (1996), Scruggs (1998), Guo and Whitelaw (2006), Bali and Engle (2010), and Campbell, Giglio, Polk, and Turley (2012) estimate ICAPMs that include additional state variables. Some of these papers could be done with ODIN. For example, Scruggs (1998) uses monthly data on the excess market return, the excess return on a long-term bond index, and the risk free rate with QMLE. Monthly measurements of these variables are all available at a daily frequency. Campbell, Giglio, Polk, and Turley (2012) use quarterly data and a six variable vector autoregression. The variables are the quarterly real stock return, the within-quarter realized return volatility from daily data, the price-earnings ratio measured as the price of the S&P 500 index divided by a ten-year trailing moving average of aggregate earnings of companies in the S&P 500 index, the term spread, the small-stock value spread, and the default spread. Only the aggregate earnings variable is truly only measured at the quarterly frequency, and the use of the ten-year moving average of earnings implies that the earnings part changes very slowly. Thus, one could change the stock price across days within a quarter while keeping the earnings constant throughout the quarter without much loss of content or induced measurement error. This model could therefore be estimated with ODIN, either at the quarterly frequency or the monthly frequency.

We certainly agree that additional state variables, such as the change in the interest rate, are no doubt necessary to adequately capture the changing investment environment faced by investors. We plan to include the conditional covariances of returns with such state variables in future research that investigates the conditional expected returns on multiple assets.

More generally, any study that uses only financial data that are available at the daily frequency is a candidate for the ODIN modeling strategy. For example, the Gaussian term structure model of Joslin, Singleton, and Zhu (2011) could be done with ODIN as could

estimation of the five-factor model of Fama and French (2013).

A   Data and Returns

We start with daily rates of returns, r^d_{t+1}, on the value-weighted stock index as well as monthly returns, $R^f_{t_m}$, on one-month T-bills from CRSP. As the MIDAS model uses daily returns, we proceed as follows: 1) For each month, we construct daily risk-free returns as $r^f_t = (R^f_{t_m})^{(1/N_m)} - 1$, where N_m is the number of trading days in the month. Hence, we get N_m daily risk-free returns, which are all the same within the month. 2) For the conditional variances of the MIDAS model, we construct daily excess returns as $r_{t+1} = r^d_{t+1} - r^f_t$. 3) For the monthly excess returns that are the dependent variable in the MIDAS and GARCH models, we first compute monthly stock returns and monthly risk-free returns as $R^m_{t_m} = (1 + r^d_{t+1})(1 + r^d_{t+2}) \cdots (1 + r^d_{t+N})$ and $R^f_{t_m} = (1 + r^f_t)(1 + r^f_{t+1}) \cdots (1 + r^f_{t+N-1})$ and then take the difference, $R_{t_m} = R^m_{t_m} - R^f_{t_m}$.

When we estimate the basic monthly or quarterly GARCH-M and MIDAS models, we use actual calendar periods as this has been the standard in the literature. For the ODIN models, we construct returns over 22-day periods, or 66-day periods, for any given starting date and always estimate the GARCH-M and MIDAS models on the same dependent variable excess returns.

B   Simulation and Bootstrapping

B.1   Simulation and Bootstrapping from the GARCH-M Model

Simulating from the GARCH-M model is straightforward. To bootstrap the model, we first construct standardized residuals as

$$\hat{\varepsilon}_{t_m+1} = \frac{R_{t_m+1} - \hat{\mu} - \hat{\gamma}\hat{\sigma}^2_{t_m}}{\hat{\sigma}_{t_m}}$$

where $\hat{\mu}$ and $\hat{\gamma}$ are the estimated parameters, and $\hat{\sigma}^2_{t_m}$ is the estimated conditional variance of R_{t_m+1}. Because the process of standardized residuals does not necessarily have a sample mean of zero and variance of one, we ensure the standardized residuals have mean zero and variance one by calculating $u_{t_m} = (\hat{\varepsilon}_{t_m} - \mu_{\hat{\varepsilon}_{t_m}})/\sigma_{\hat{\varepsilon}_{t_m}}$, where $\mu_{\hat{\varepsilon}_{t_m}}$ and $\sigma_{\hat{\varepsilon}_{t_m}}$ are the sample mean and standard deviation of the $\hat{\varepsilon}_{t_m}$. We then simulate from the GARCH-M model using innovations drawn with replacement from $u_1, u_2, \ldots, u_T$. Estimating the GARCH-M model based on real or bootstrapped data with non-normal innovations can be viewed as quasi-maximum-likelihood (QMLE) and is thus consistent. Note that simulating from the model using innovations that do not have mean zero and variance one would not provide consistent parameter estimates.

B.2   Simulation and Bootstrapping from the MIDAS Model

The MIDAS model is harder to simulate than the GARCH-M model because the MIDAS model is based on daily returns but is estimated on monthly or quarterly returns. When simulating using normal innovations, each daily return is drawn from a standard normal distribution and is scaled by the conditional standard deviation for that month. We proceed as follows: 1) The conditional variance is based on the previous 500 daily returns. Hence, we first draw 500 standard normal variables and scale them by the unconditional standard

deviation of the model, $\sqrt{\omega/(1-\phi)}$. 2) We next calculate the conditional variance for the daily returns over the next month, $\sigma_{t_m}^2 = V_{t_m}^{MIDAS}$ as in equation (2). 3) If there are N_m days in the next month, we draw N_m standard normal variables $u_d, d = 1, \ldots, N_m$ and calculate daily returns as $r_{m,d} = \mu + \gamma\sigma_{t_m}^2 + \sqrt{\sigma_{t_m}^2}u_d, d = 1, \ldots, N_m$, where $r_{m,d}$ is the return on day d in month m. 4) We repeat steps 2-3 for the following months. The result is a series of daily returns, and the MIDAS model is then calibrated to these daily returns.

To bootstrap the MIDAS model we need daily innovations for the simulation, even though the model is estimated based on monthly data. To obtain this, we 'interpolate' daily means and variances in the following way: 1) For each day, we use the estimated parameters to construct a daily forecast of the conditional variance of the daily return, $\widehat{\sigma}_t^2 = \hat{\omega} + \hat{\phi}\sum_{d=1}^{500} w(\hat{\kappa})r_{t-d}^2$. 2) For each day, we use the estimated parameters to construct a daily forecast of the conditional mean of the daily return: $\hat{\mu}_t = \hat{\mu} + \hat{\gamma}\widehat{\sigma}_t^2$. 3) For each day, we obtain the residual for the daily return as $\hat{\varepsilon}_{t+1} = \frac{r_{t+1}-\hat{\mu}_t}{\sqrt{\widehat{\sigma}_t^2}}$.

As above for the GARCH-M model, this process of standardized residuals does not necessarily have a sample mean of zero and a variance of one. Hence, we ensure the standardized residuals have mean zero and variance one by calculating $u_t = (\hat{\varepsilon}_t - \mu_{\hat{\varepsilon}_t})/\sigma_{\hat{\varepsilon}_t}$, and we simulate from the MIDAS model using innovations drawn with replacement from $u_1, u_2, \ldots, u_T$.

References

Andersen, T. G. and T. Bollerslev, 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* 39, 885-905.

Baker, M. and J. Wurgler, 2006. Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Bali, T. and R. F. Engle, 2010. The intertemporal capital asset pricing model with dynamic conditional correlations, *Journal of Monetary Economics* 57, 377-390.

Bollerslev, T. and J. M. Wooldridge, 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances, *Econometric Reviews* 11, 143–172.

Campbell, J. Y., 1996. Understanding risk and return, *Journal of Political Economy* 104, 298-345.

Campbell, J. Y. and L. Hentschel, 1992. No news is good news: An asymmetric model of changing volatility. *Journal of Financial Economics* 31, 281-318.

Campbell, J. Y., S. Giglio, C. Polk, and R. Turley, 2012. An Intertemporal CAPM with Stochastic Volatility, manuscript, Harvard University.

Cochrane, J. H. 2005. *Asset Pricing: Revised Edition.* Princeton, NJ: Princeton University Press.

D'Ericco, J., 2011. DERIVEST, http://www.mathworks.com/matlabcentral/fileexchange/13490-adaptive-robust-numerical-differentiation.

Drost, F.C. and T.J. Nijman, 1993, Temporal aggregation of GARCH processes, *Econometrica* 61, 909-927.

Engle, R. F. and K. F. Kroner, 1995. Multivariate simultaneous generalized ARCH, *Econometric Theory* 11, 122-150.

Fama, E. F. and K. French, 2013, A five-factor asset pricing model, manuscript, University of Chicago.

Francq, C., L. Horvath, and J. Zakoian, 2011. Merits and drawbacks of variance targeting in GARCH models, *Journal of Financial Econometrics* 9, 619-656.

Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005. There is a risk-return trade-off after all, *Journal of Financial Economics 76, 509-548.*

Ghysels, E., A. Sinko, and R. Valkanov, 2007. MIDAS regressions: Further results and new directions, *Econometric Reviews*, 26, 53–90.

Ghysels, E., A. Plazzi, and R. Valkanov, 2013. The risk-return relationship and financial

crises, manuscript, University of North Carolina.

Guo, H., and R. F. Whitelaw, 2006. Uncovering the risk-return relation in the stock market. *Journal of Finance* 61, 1433-1463.

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

Hansen, L. P., and R. J. Hodrick, 1980. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* 88, 829-853.

Joslin, S., K. J. Singleton, and H. Zhu, 2011, A new perspective on Gaussian dynamic term structure models, *Review of Financial Studies* 24, 926-970.

Lanne, M. and P. Saikkonen, 2006. Why is it so difficult to uncover the risk-return trade-off in stock returns? *Economic Letters* 92, 118-125.

Lettau, M. and S. C. Ludvigson, 2010. Measuring and modeling variation in the risk-return trade-off, Chapter 11 in Y. Ait-Shalia and L. P. Hansen, eds., *Handbook of Financial Econometrics*: *Volume 1 - Tools and Techniques*, Elsevier: Amsterdam, The Netherlands.

Lundblad, C., 2007. The risk return trade off in the long run: 1836–2003, *Journal of Financial Economics* 85, 123–150.

Merton, R. C., 1973. An intertemporal asset pricing model, *Econometrica* 41, 867–887.

Merton, R. C., 1980. On estimating the expected return on the market: an exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347-370.

Nyberg, H., 2012. Risk-return trade-off in U.S. stock returns over the business cycle, *Journal of Financial and Quantitative Analysis* 47, 137-158.

Richardson, M. and J. Stock, 1989. Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323-348.

Scruggs, J. T., 1998. Resolving the puzzling intertemporal relation between the market

risks premium and conditional market variance: A two-factor approach, *Journal of Finance* 53, 575-603.

Valkanov, R., 2003, Long-horizon regressions: Theoretical results and applications, *Journal of Financial Economics* 68, 201-232.

Yu, J. and Y. Yuan, 2011, Investor sentiment and the mean–variance relation, *Journal of Financial Economics* 100, 367–381.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394-1411.