Estimating the Risk-Return Trade-off with Overlapping Data Inference

Esben Hedegaard*
AQR Capital Management

Robert J. Hodrick†
Graduate School of Business
Columbia University
and NBER

January 18, 2016

We thank Jules van Binsbergen, Eric Ghysels, Christian Lundblad, Alberto Plazzi, Rossen Valkanov, and seminar participants at the Stanford Graduate School of Business and Purdue University's Krannert School of Management for helpful comments; and we thank Norman White for research computing assistance. This research was supported by a grant from the Network for Study on Pensions, Aging, and Retirement to the Columbia Business School.

*esben.hedegaard@stern.nyu.edu.

†rh169@columbia.edu.

Abstract

Investigations of the basic risk-return trade-off for the market return typically use maximum likelihood estimation (MLE) with a monthly or quarterly horizon and data sampled to match the horizon even though daily data are available. We develop an overlapping data inference methodology for such models that uses all of the data while maintaining the monthly or quarterly forecasting period. Our approach recognizes that the first order conditions of MLE can be used as orthogonality conditions of the generalized method of moments (GMM). While parameter estimates from the different non-overlapping monthly samples that start on different days vary substantively, a formal test does not reject parameter equality and constrained estimation of the risk-return trade-off produces a statistically significant value of 3.35 in post-1955 data.

1 Introduction

When the conditional variance of the market return increases, does the conditional mean of the market return also increase, and if so, by how much? These basic questions involving the trade-off between risk and expected return have been studied empirically in various ways, but the profession has yet to settle on definitive answers.[1] Part of the empirical problem stems from the abstract nature of the theory. The fundamental theory for a risk-return trade-off traces back to Merton's (1973) continuous time intertemporal capital asset pricing model (ICAPM). Although investors in such theoretical models hold assets for an abstract amount of time, an instant in Merton's model, 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) model. The purpose of this paper is to provide estimates of the risk-return trade-off that use all of the available daily data while retaining the monthly or quarterly holding period used in previous empirical analyses.

The simplest version of the ICAPM postulates a risk-return trade-off between the conditional expected excess return on the market portfolio and its conditional variance. Since neither the conditional mean nor the conditional variance of the excess return on the market is observable, inference about the validity of the theory is complicated. Econometric tests invariably use the assumption of rational expectations to break the realized future return into its expected value plus an error term that is orthogonal to the conditioning information set. Then, some econometric model of the conditional variance of the error term is used. These models can generally be classified as either generalized autoregressive conditionally heteroskedastic (GARCH) models; mixed data sampling (MIDAS) models, which use higher

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

frequency data that the holding period horizon; or latent variable models.

Beginning with the seminal analyses of French, Schwert, and Stambaugh (1987) and Glosten, Jagannathan, and Runkle (1993) and continuing with the models of Scruggs (1998), Brandt and Kang (2004), Ghysels, Santa-Clara, and Valkanov (2005), and Nyberg (2012), econometricians have sampled the data at their chosen frequency of a calendar month or a calendar quarter and employed maximum likelihood estimation (MLE) even though higher frequency daily data are available.[2]

In this paper we first argue that if the econometrician thinks that a one-month time interval is the appropriate holding period to use in testing an asset pricing model, but nothing about calendar months is critical in the development of the theoretical model or the availability of data, then the econometric analysis can be done using any day in the month as the starting day and using a 22 day holding period as the 'one-month' interval. [3]

Thus, there are 22 possible samples of non-overlapping data that can be used to generate 22 sets of parameter estimates corresponding to the theory. Of course, if the theory is true, these alternative sets of estimates should only differ because of sampling error.

Although we find variation in the estimates from the alternative non-overlapping samples, we are unable to reject equality of the sets of parameters using a specification test based on Hansen's (1982) generalized method of moments (GMM). Because the specification test is passed, we then impose the constraint that the same set of parameters simultaneously satisfies each of the correlated non-overlapping samples. In doing so, we extend the analysis

[2]During our research on this project we discovered that the results in Ghysels, Santa-Clara, and Valkanov (2005) are incorrect as we found a bug in their MIDAS program when we were unable to replicate their results. The authors have been notified and agree with this statement. Ghysels, Plazzi, and Valkanov (2013) provide a partial correction of the earlier results and extends the model of the risk-return trade-off to include effects of the financial crisis.

[3]Because our methods our general, we can work with any horizon, and we have performed similar analyses for horizons from two days to 66 days. See Hedegaard and Hodrick (2014b). In our full sample, we use a 22 day interval because many finance studies use a one-month holding period and the average number of trading days in a month is 22 over the full sample. When we split the sample in 1952, the average number of trading days in a calendar month is 24 before 1952 and 21 after 1952, and we adjust the number of days in our monthly samples to correspond to the average number of days in a calendar month.

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. While it is now commonplace and even expected to see ODIN employed in regression forecasting situations in which the data are sampled more finely than the forecasting interval, we have never seen ODIN employed in tests of asset pricing models using MLE. Also, because we derive the distribution of the individual parameter estimates, we consider their sample average, which is also an estimator.

The plan of the paper is as follows. Section 2 presents the standard risk-return trade-off, notes that including a constant in the conditional mean is necessary to appropriately test the prediction of the model that an increase in the conditional variance increases the conditional mean, and discusses the choice of horizon. Section 3 presents the results of the basic estimations using calendar months as well as the individual estimations corresponding to different starting days. Section 4 discusses our ODIN methodology and the specification test of equality of the parameters from the different non-overlapping samples and presents the ODIN estimation. Section 5 briefly examines the small sample distributions of the estimators and discusses power issues. Section 6 provides conclusions, and an Appendix provides some technical details including the equations describing the estimators and their asymptotic distribution. An Online Appendix contains additional figures and more technical details.

2 The Conditional Risk-Return Trade-off

The simplest version of Merton's (1973) model 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, $\sigma^2_{M,t}$, as in Merton's (1980) Model 1:

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

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 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.[4] 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 of any financial strategy, they must choose the horizon for which they think the theory holds. While daily or even higher frequency returns are available, most of the existing literature prefers to examine the risk-return trade-off using longer horizon monthly or quarterly returns. Several

[4]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.

considerations motivate this choice of horizon.

The first reason is simply data availability. Fisher and Lorie (1964) describe the process that CRSP researchers went through to construct the first CRSP stock return database. Limited computing power, costly data storage, and the physical requirements of coding data on punch cards no doubt entered into their decision to report only monthly returns. While CRSP now also produces daily returns, prominent researchers, such as Fama and French (2015), continue to use a monthly time interval as the appropriate holding period for the empirical evaluation of financial theories, and they sample the data to correspond to calendar months.

Why would a monthly time interval be the most appropriate one for testing financial theories? We can think of at least two reasons. First, aspects of the trading process induced by market microstructure frictions, non-synchronous portfolio investment decisions, and individual stock illiquidity that are outside the theory dominate the autocorrelations of short-horizon returns. 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 estimate 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.

The regression forecasting situation is not the only place that overlapping data have been used in finance. Zhang, Mykland, and Aït-Sahalia (2005) demonstrate that the optimal way to estimate integrated volatility is by averaging realized quadratic variation on overlapping sampling intervals and bias-adjusting for the market microstructure noise. In a related paper, Aït-Sahalia, Mykland, and Zhang (2005) demonstrate that if one is willing to assume an explicit parametric structure for the market microstructure noise in financial markets, using all of the ultra high frequency data with maximum likelihood is optimal. We are less formal in our arguments as we are more agnostic about the choice of horizon that is appropriate for testing economic theories and the specification of trading frictions that would be necessary to apply MLE to short-horizon data. We think that at this point, it is too difficult to formally model the conditional distributions of the error processes within MLE when using all of the daily data and the one-month or one-quarter forecasting interval that previous researchers have argued is appropriate to test the implications of the theory.

Ideally, one would have a unified theory of price determination that would incorporate all aspects of the trading process and that would allow for MLE estimation as in Aït-Sahalia, Mykland, and Zhang (2005). Alternatively, one might have a good idea about the nature of the noise and could apply the approach of Zhang, Mykland, and Aït-Sahalia (2005). While these approaches work well when thinking about estimating the variance of a continuous time process, we think that the idea of a conditional risk-return trade-off makes good sense, but not as a continuous time model. We wish to remain agnostic about the horizon over which the theory might apply, and we are merely providing a better method for researchers interested in testing asset pricing theories in which the holding period is best thought of as monthly or even quarterly.

In what follows we use a monthly interval as the forecast horizon for our estimation to coincide with much of the literature, and we use the term 'basic' to refer to models estimated

with non-overlapping observations to distinguish them from ODIN models. After reviewing the estimation, Section 5 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. In these simulations and in our actual estimation, though, we rely on large samples in which the overlap in the data is a small fraction of the total sample size, so this statement should not be extrapolated literally.

3 The Basic GARCH-M Model

This section first describes the basic estimation of the conditional risk-return trade-off with non-overlapping monthly data, as is standard in the literature. Then, we interpret the notion of a month in our full sample as a 22 day interval. Hence, there are 22 possible non-overlapping specifications with different starting days, and we first estimate them individually. We denote the time index for the j-th one of these samples as $t_{m_j} = 1_{m_j}, \ldots, T_{m_j}$, which counts 22 day periods beginning with the first observation, 1_{m_j}, which is the j-th day of the full sample of daily data. There are T_{m_j} of these monthly periods in the sample. For simplicity of notation, we drop the M subscript on the market return, and we denote $R_{t_{m_j}+1}$ as the one-month excess market return from the j-th sample. The GARCH-M model for the j-th sample is therefore

$$R_{t_{m_j}+1} = \mu_j + \gamma_j \sigma^2_{t_{m_j}} + \varepsilon_{t_{m_j}+1}, \quad \varepsilon_{t_{m_j}+1} \sim N(0, \sigma^2_{t_{mj}}) \tag{2}$$

where the conditional variance of the innovation is modeled as a GARCH(1,1) process:

$$\sigma^2_{t_{m_j}+1} = \omega_j + \alpha_j \varepsilon^2_{t_{m_j}+1} + \beta_j \sigma^2_{t_{m_j}}. \tag{3}$$

We subscript the parameters in equations (2) and (3) with a j to indicate that they are sample specific, but Merton's (1980) empirical approach to his theory implies that the parameters should be equal. The log likelihood function for the GARCH-M model for the monthly sampling interval in equations (2) and (3) is

$$\log(L) = \sum_{t_{m_j}=1_{m_j}}^{T_{m_j}} \left(-\frac{1}{2}\log(2\pi) - \log\left(\sigma_{t_{m_j}}\right) - \frac{1}{2}\frac{\varepsilon^2_{t_{m_j}+1}}{\sigma^2_{t_{m_j}}} \right). \tag{4}$$

Rather than estimating ω_j as a free parameter, we estimate ω_j by variance targeting as in Engle and Kroner (1995).[5]

3.1 Results for the Basic Model with Calendar Month Data

Table 1 presents estimates of the basic GARCH-M models for three samples of monthly non-overlapping calendar data, as is usually done in the literature.[6] The next subsection presents the results for the individual non-overlapping samples using the same number of days as the average number of trading days in the calendar months of the sample. The full sample for monthly data is 1927:10 to 2011:12. We also split the sample after 1952, as do French, Schwert, and Stambaugh (1987) and others, 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 of the risk-return trade-off. The second monthly sample is consequently 1927:10 to 1952:12 with 24 observations per month, and the third monthly sample is 1955:1 to 2011:12 with 21 observations per month.[7]

[5] 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.

[6] The Online Appendix contains quarterly results, as well as the results of using the MIDAS model to estimate the conditional variance. The Online Appendix also presents extensive simulation evidence that the t-statistics on which we base the asymptotic inference are very well behaved in our sample sizes.

[7] Although the CRSP data start in 1926:01, our sample starts in 1927:10 to allow for comparisons with MIDAS models that require use of lagged data. We have chosen not to present the MIDAS results to shorten the paper. The results are available in Hedegaard and Hodrick (2014a) or in the Online Appendix.

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					

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

Table 1 reports the parameters estimates with QMLE asymptotic standard errors in parenthesis (see Bollerslev and Woolridge (1992)), and the associated p-values of the t-tests of the null-hypothesis that the parameter is zero. [8]

We also report bootstrap p-values under the null hypothesis that $\gamma = 0$ (see Appendix C for details). The estimates of the risk-return trade-off, $\hat{\gamma}$, range from 0.44 to 3.01, but only the latter estimate from the post 1952 sample has a bootstrap p-value less than .10.

3.2 The Individual Estimations

Once one realizes that the starting date of the monthly sample does not matter, one wonders how different can the various correlated non-overlapping estimates be? The surprising

[8] We take care in evaluating the Hessian of the likelihood function. In particular, the Hessian returned from the MatLab optimization routine is not reliable, as it is a so-called 'pseudo-Hessian' constructed for 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.

answer in this case is, quite different. Figures 1, 2, and 3 present the the basic, non-overlapping estimates along with the calendar month estimates for the three sample periods: the full sample, labeled A; the first sub-sample, labeled B; and the second sub-sample, labeled C. In each Figure, the four plots present the estimates of the four parameters in the GARCH-M model. The solid black line shows the point estimates from the basic GARCH-M models obtained by shifting the start date of the sample one day at a time. The shaded gray area shows the 95% confidence intervals for the estimates. These individual estimates are obviously not independent because they are based on the same data, which are just sampled differently. Thus, as one would expect, the estimates move slowly with the sampling start date, but there is substantive variation. We also report the calendar month estimates as a dot with horizontal lines showing the 95% confidence interval.

Consider the full sample results in Figure 1. The top right plot presents the estimates of the different γ_j. None of the 22 individual $\hat{\gamma}_j$'s is significantly different from zero, and the estimates range from 1 to 3. We also see a clear negative correlation between the estimates of $\hat{\mu}_j$ and $\hat{\gamma}_j$ as the starting date varies. This correlation arises because increasing either μ or γ increases the unconditional expected return in the model. One can also see a negative correlation between $\hat{\alpha}_j$ and $\hat{\beta}_j$ because increasing either of these parameters gives rise to higher volatility-of-volatility.

Figure 2 presents the results for the first sub-sample, 1927–1952. The $\hat{\gamma}_j$'s vary between -0.15 and 1.2, and as in the full sample, none of the individual $\hat{\gamma}_j$'s is significantly different from zero.

The results for the second sub-sample, 1955–2011, tell a different story in Figure 3. Now, altough the individual $\hat{\gamma}_j$'s vary between 2.4 and 4.1, most of these individual estimates are significantly different from zero at the 5% marginal level of significance. Recall that the calendar month estimate of γ from the basic model is 3.01 with a p-value of .08. The difference in implication between a value of 2.4 and a value of 4.1 for the expected return

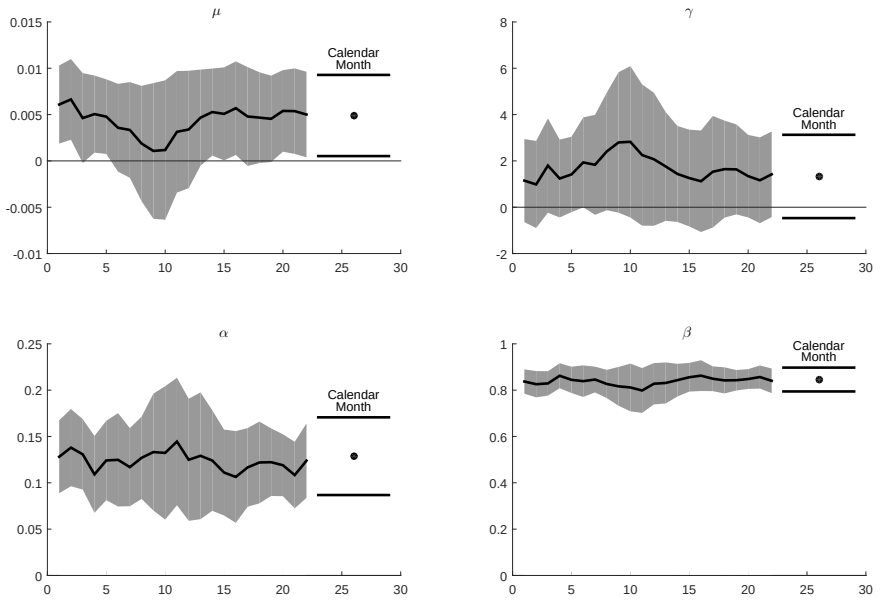


Figure 1: Monthly Estimates, Sample A: 1927–2011. The plots show the individual estimates obtained by shifting the start-date on the horizontal axis and their 95% confidence interval in shaded grey. The 'basic monthly' estimate uses calendar months and the 95% confidence interval is indicated with horizontal lines.

implied by the model is quite substantive. If we take 0.15 as a typical standard deviation of annualized market returns, the expected return implied by the model ranges from 5.4% to 9.2% for the different parameter estimates. If the parameters are statistically significantly different, the model should be rejected. If we cannot reject equality of the parameters, then constraining them to have the same value should result in a better estimate of the risk-return trade-off. Hence, the next section tests whether these different estimates are significantly different from each other. We find that we are unable to reject the hypothesis of equality of the coefficients, and we consequently estimate under the constraint that they are equal.

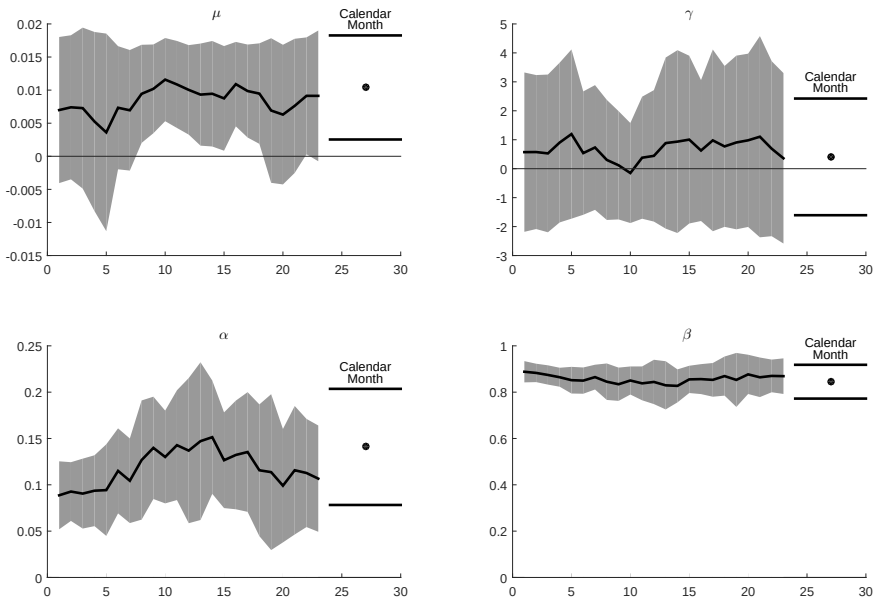


Figure 2: Monthly Estimates, Sample B: 1927–1952. The plots show the individual estimates obtained by shifting the start-date on the horizontal axis and their 95% confidence interval in shaded grey. The 'basic monthly' estimate uses calendar months and the 95% confidence interval is indicated with horizontal lines.

4 Overlapping Data Inference with GMM

This section discusses the logic of our ODIN estimation strategy for the risk-return trade-off and the specification test that we use to test the null hypothesis that the individual correlated estimates in the previous section are equal to each other. We relegate most of the equations describing the estimators and their asymptotic distributions to the Appendix.

Although the model in equations (2) and (3) is specified at the monthly frequency, we have also assumed 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 the model for the full sample, for example, using a 22 day forecasting period as an example, as

$$R_{t+22,t} = \mu + \gamma\sigma^2_{t,t+22} + \varepsilon_{t+22,t}, \quad \varepsilon_{t+22,t} \sim N(0, \sigma^2_{t,t+22}) \tag{5}$$

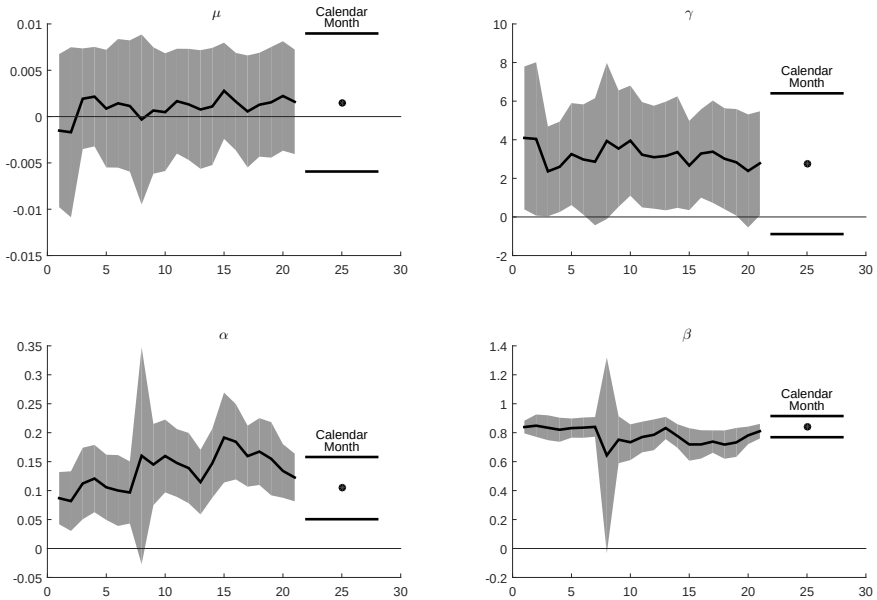


Figure 3: Monthly Estimates, Sample C: 1955–2011. The plots show the individual estimates obtained by shifting the start-date on the horizontal axis and their 95% confidence interval in shaded grey. The 'basic monthly' estimate uses calendar months and the 95% confidence interval is indicated with horizontal lines.

where the subscript t denotes a day and 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 between days t and $t+22$. Thus, the conditional variance evolves as

$$\sigma^2_{t+22,t+44} = \omega + \alpha\varepsilon^2_{t+22,t} + \beta\sigma^2_{t,t+22}. \tag{6}$$

We assume that equations (5) and (6) 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 the theory is silent about the role of calendar months, the starting date becomes irrelevant. We thus have the opportunity to

increase the sample size by using overlapping data, which should reduce standard errors, if the model is correct, and increase the power of the tests, if the model is false.

4.1 ODIN-GARCH-M

Estimation of the ODIN-GARCH-M model employs the first order conditions from the MLEs of the monthly models in equation (4) that must hold for each starting date. As Cochrane (2005) notes, first order conditions from MLE are equivalent to unconditional orthogonality conditions in Hansen's (1982) GMM estimation. Consequently, if the theory is correct, the sample orthogonality conditions should hold for the 22 possible daily starting dates that index the different months associated with the t_{m_j} indexes when evaluated at a common parameter vector. We then use GMM to derive the asymptotic distribution of these common parameter estimates that satisfy the average across starting days of the monthly first order conditions. Although the ODIN methodology induces serial correlation by the creation of overlapping observations, which could affect small sample inference, the extensive simulations discussed briefly in the following section demonstrate that our asymptotic inference is appropriate in our sample sizes. [9]

In addition to constraining the parameters to be the same across all the possible starting dates, by viewing the individual estimates corresponding to the different starting dates as being from a correlated set of GMM estimators, we derive a test of the equality of the parameters, which we label the H statistic. The econometric equations for this joint test are also presented in the Appendix. Under the null hypothesis in our analysis the H statistic has a chi-square distribution with 84 degrees of freedom.

We perform the test of equality of the individual parameters for the three sub-samples and fail to reject in each sub-sample. For the full sample, we obtain $H = 44.7$, which corresponds to a p-value of .999. Thus, while the parameter estimates in Figure 1 show

[9] We present most of the analysis of our simulations in the Online Appendix.

economically important variation across the sampling start dates, we cannot reject the null that the parameters are equal across the starting dates given the standard errors of the parameters. For the first sub-sample, the test-statistic is $H = 104.85$ with a p-value of .11, and for the second sub-sample the test statistic is $H = 55.2$ with a p-value .98. Again, in both cases we fail to reject that the parameters are equal across the different starting dates at usual marginal levels of significance.

4.2 Estimating the ODIN-GARCH-M Models

Table 2 presents the results of the constrained ODIN-GARCH-M estimation. 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.[10] Standard errors are presented in parenthesis below the point estimates with p-values below the standard errors.

In Table 2, the constrained $\hat{\gamma}$'s 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 0.654 with a p-value of .684 compared to a $\hat{\gamma}$ of 0.443 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.

Only in the post-1955 sub-sample does the ODIN model achieve a substantive reduction in the standard error which results in a statistically significant risk-return trade-off. In the full sample and the first sub-sample, the ODIN standard errors increase relative to those of the basic calendar month estimates.[11] Variation in the parameter estimates of the various

[10] The table shows the results of the estimation with a 'monthly' forecasting interval, which is set to 22 days for the full sample, 24 days for the first sub-sample, and 21 days for the second sub-sample. The Online Appendix shows the results of the quarterly estimation, as well as the ODIN MIDAS estimations.

[11] We also tried estimating the model with innovations from the Students t-distribution, but we found no improvement in the results. Estimation with an asymmetric response as in Glosten, Jagannathan, and

Table 2: 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		

Note: The table presents results from estimation of the GMM estimation of the ODIN-GARCH model specified in the orthogonality conditions of equation (7). Standard errors are in parenthesis with p-values below.

non-overlapping samples is presumably the reason. In the simulation section below, we find that standard errors with ODIN estimation are substantively smaller, is expected with the use of additional data, if the model is well specified. One interpretation of the data is that only for the post-1955 sub-sample is the model sufficiently well specified that we find the theoretical trade-off between risk and return.

4.3 The Average of the Individual Estimates

ODIN constrains the various non-overlapping estimates to have the same value. As noted above, one can also consider the estimator that is the average of the individual estimates from the correlated non-overlapping samples. The standard errors for this estimator are also presented in the Appendix. For the monthly sampling interval, the averages of the $\hat{\gamma}$'s for samples A, B, and C, with the appropriate standard errors in parenthesis, for the basic GARCH-M model are 1.68 (1.35), 0.67 (1.67), and 3.182 (1.46), respectively. As one might

Runkle (1993) is reported in the Online Appendix. When we allow for this asymmetry, we find that the risk-return trade-off is quite imprecisely estimated.

expect, these values are quite similar to the constrained ODIN values. In particular, the p-value of $\hat{\gamma}$ for sample C is reduced from .11 in the basic monthly data to .03 in the average estimator and .02 in the ODIN estimator.

5 Simulation Analysis of the Models

This section briefly discusses some of the most important findings from simulations of the models. 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.

We simulate from a GARCH-M model as in Lundblad (2007) without a particular model of market microstructure noise or other short-horizon deviations from the null model. As noted above, we are intentionally agnostic about the appropriate horizon over which it can be argued that the economic models hold. Obviously, in the simulations, because the true model holds at the shortest horizons, using all of the data in a short-horizon, non-overlapping estimation would be most powerful. Our goal here is simply to compare the power of ODIN analysis to the power of the basic sampled estimation that throws away information. We also want to check that our procedure does not induce size distortions in the sample sizes that are available for testing the theory.

5.1 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, as

in the original analysis of Hansen and Hodrick (1980). The same logic applies here.

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. We follow Nelson (1991), Drost and Nijman (1993), Andersen and Bollerslev (1998), and Lundblad (2007) and specify the continuous-time limit for a GARCH-M(1,1) 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.

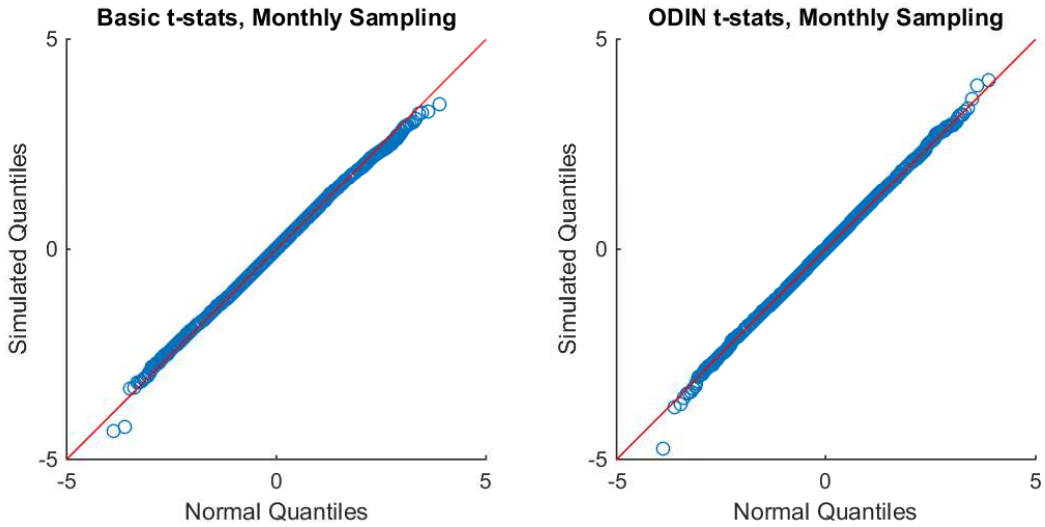
Across 10,000 simulations, both for the null hypothesis, $\gamma = 0$, and for three alternative hypotheses, $\gamma = 1, 2$, and 3, we find that $\hat{\gamma}$ is only slightly biased (see the Online Appendix for details). The Online Appendix also shows QQ-plots of $\hat{\gamma}$ against the quantiles 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.

Rather than focusing strictly on the distributions of $\hat{\gamma}$, as in Lundblad (2007), we also examine the distributions of the test statistics under both the null and the alternative hypotheses to examine the powers of the tests. Figure 4 shows QQ-plots of the t-statistics for the Basic and ODIN models. We simulate the data under the null of no risk-return relation and sample the data monthly (every 22 days). The distribution of the t-statistics are in both cases very close to a normal distribution. The QQ-plot of the t-statistics for the Basic model shows a slight asymmetry indicating that the tail of the distribution is slightly thinner than the tails of a normal distribution. The Online Appendix demonstrates that this asymmetry becomes more pronounced for longer sampling frequencies, and the distribution of the t-statistics for the ODIN model is generally better behaved.

Because Lundblad (2007) focuses directly on the distributions of $\hat{\gamma}$, rather than on the t-test, he somewhat overstates the difficulty of rejecting the null hypothesis of $\gamma = 0$. We nevertheless agree with his central point: In the basic model with non-overlapping data, 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 (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%.

The simulations of the ODIN methodology indicate that decreases in standard errors and increases in power from using overlapping data can be substantial and correspond to large

Figure 4: QQ-Plots of t-Statistics from the Basic and ODIN model



This figure shows QQ-plots of the t-statistics for monthly sampling based on the Basic model (left) and the ODIN model (right). In both cases, the distribution of the t-statistics is close to a normal distribution. The QQ-plot of the t-statistics for the Basic model shows a slight asymmetry indicating that the tail of the distribution is slightly thinner than the tails of a normal distribution.

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.

5.2 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. We also have not explored the properties of ODIN in smaller sample sizes. As in most asset pricing models, it is sensible to simulate the asymptotic distributions of the estimators as we have done here. We know that when the overlap is too big, using ODIN with the standard asymptotic theory is not advisable. Nevertheless, the k different individual estimators are available, and each of these should be investigated.

6 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. One way of modeling the conditional variances and covariances that are the sources of risk in these models is GARCH, which is 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 GMM standard errors that account for the serial correlation induced by the use of overlapping data.

We apply this ODIN methodology to investigate the risk-return trade-off using GARCH-M modeling of the conditional variance of the market return. Simulations of the ODIN methodology indicate that if this were the true model with a one-month horizon, the ODIN approach would be substantially more powerful than the basic non-overlapping data approach. Estimating the basic GARCH-M model with non-overlapping monthly data for the sample period 1955:1 to 2011:12 produces a conditional risk-return trade-off of 3.011 that has a bootstrapped p-value of .08. When we use the ODIN methodology on the same sample, the risk-return trade-off is 3.354 and the p-value falls to .022. As with much of the literature, though, we find insignificant or even negative trade-offs in samples that include the Great Depression and with asymmetric responses to shocks.

We have only investigated the simplest version of the ICAPM in which the return on the market is the only state-variable, but our methods could be used to investigate more general versions. Many authors, including Campbell (1996), Scruggs (1998), Guo and Whitelaw (2006), Bali and Engle (2010), and Campbell, Giglio, Polk, and Turley (2015) estimate ICAPMs that include additional state variables. Some of these papers could be done with ODIN, as we do in Hedegaard and Hodrick (2014b). 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 (2015) 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. Yu and Yuan (2011) use Baker and Wurgler's (2006) measure of investor sentiment and find a positive risk-return trade-off in the conditional CAPM during low sentiment periods but not in high ones. Their analysis could be reexamined with ODIN.

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 asset pricing 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 analysis of risk exposures of equities to the five-factor model of Fama and French (2015).

A Data and Returns

We start with CRSP value-weighted market daily rates of returns, r_t^d, as well as one-month returns on Treasury bills, $R_{t_m}^f$, also from CRSP. For each month, we construct daily risk-free rates of returns as $r_t^f = (R_{t_m}^f)^{(1/N_m)} - 1$, where N_m is the number of trading days in the month. Hence, we get N_m daily risk-free rates of returns, which are all the same within the month. When we estimate the basic monthly GARCH-M models, we use actual calendar periods as this has been the standard in the literature. For the various non-overlapping models and the ODIN model, we construct returns over 22-day periods for any given starting date. For any given day, we first compute 22 day stock returns and 22 day risk-free returns as $R_{t_m}^m = (1 + r_{t+1}^d)(1 + r_{t+2}^d) \cdots (1 + r_{t+22}^d)$ and $R_{t_m}^f = (1 + r_t^f)(1 + r_{t+1}^f) \cdots (1 + r_{t+21}^f)$ and then take the difference, $R_{t_m} = R_{t_m}^m - R_{t_m}^f$ to get the dependent variables in the GARCH-M models.

B Econometric Analysis

Let the parameter vector associated with equation ((5) 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:

$$G_T(R;\theta) = \frac{1}{T}\sum_{t=0}^{T-1} g_t(R_{t+1};\theta) = \begin{pmatrix} \frac{1}{T}\sum_{t=0}^{T-1}\left(\frac{\varepsilon_{t+22,t}}{\sigma^2_{t,t+22}} + \frac{1}{\sigma_{t,t+22}}\frac{\partial\sigma_{t,t+22}}{\partial\mu}\left(\frac{\varepsilon^2_{t+22,t}}{\sigma^2_{t,t+22}} + 2\gamma\varepsilon_{t+22,t} - 1\right)\right) \\ \frac{1}{T}\sum_{t=0}^{T-1}\left(\varepsilon_{t+22,t} + \frac{1}{\sigma_{t,t+22}}\frac{\partial\sigma_{t,t+22}}{\partial\gamma}\left(\frac{\varepsilon^2_{t+22,t}}{\sigma^2_{t,t+22}} + 2\gamma\varepsilon_{t+22,t} - 1\right)\right) \\ \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{\sigma_{t,t+22}}\frac{\partial\sigma_{t,t+22}}{\partial\alpha}\left(\frac{\varepsilon^2_{t+22,t}}{\sigma^2_{t,t+22}} + 2\gamma\varepsilon_{t+22,t} - 1\right) \\ \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{\sigma_{t,t+22}}\frac{\partial\sigma_{t,t+22}}{\partial\beta}\left(\frac{\varepsilon^2_{t+22,t}}{\sigma^2_{t,t+22}} + 2\gamma\varepsilon_{t+22,t} - 1\right) \end{pmatrix} \tag{7}$$

where $g_t(R_{t+1};\theta)$ denotes the vector of right-hand-side functions. Because the system of equations (7) is just identified, GMM chooses the parameter estimates, $\hat{\theta}$, to set $G_T(R,\hat{\theta}) = 0$. 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

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \to N\left(0, \Omega(\theta_0)\right)$$

and the estimate of the asymptotic variance is

$$\Omega(\hat{\theta}) = \left[D_T(\hat{\theta})^{-1} S(\hat{\theta}) D_T(\hat{\theta})^{-1}\right] \tag{8}$$

where

$$S(\hat{\theta}) = \sum_{j=-21}^{21} C_T\left(g_t(R_t;\hat{\theta}), g_{t-j}(R_{t-j},\hat{\theta})'\right) \tag{9}$$

and the $C_T\left(g_t(R_t;\hat{\theta}), g_{t-j}(R_{t-j};\hat{\theta})'\right)$ matrixes 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).

B.1 The Joint Distribution of the Sampled Estimators

Now, consider the joint distribution of the individual basic GARCH-M estimators from the respective non-overlapping samples. We use their joint distribution to test the null hypothesis that the individual estimates are equal to each other and to consider the average of the individual sampled estimates as an estimator.

Let $E(g_{t_j}(\theta_0)) = 0$ be the orthogonality conditions of the GARCH-M model for the j-th non-overlapping sample, corresponding to equations (7). The sample orthogonality conditions for the j-th data set are a function of the parameter vector θ_j:

$$G_{T_j}(\theta_j) = \frac{1}{T_j}\sum_{\{t_j\}} g_{t_j}(\theta_j) \tag{10}$$

These four sample orthogonality conditions are set to zero by the choice of the four elements of θ_j.

We assume that the overlap in the data is k, in which case $j = 1, ..., k$. To derive the joint distribution of all of the parameters, define the $(k \times 4)$-dimensional vectors $\Theta = (\theta_1', ..., \theta_k')'$ and $\widehat{\Theta} = \left(\hat{\theta}_1', ..., \hat{\theta}_k'\right)'$. Then, stack the $g_{t_j}(\theta_j)$ functions into the $(k \times 4)$-dimensional vector $g_t(\Theta) = (g_{t_1}(\theta_1)', ..., g_{t_k}(\theta_k)')'$, and let the sample orthogonality conditions evaluated at the individual parameter estimates be $G\left(\widehat{\Theta}\right) = \left(G_{T_1}\left(\widehat{\theta}_1\right)', ..., G_{T_k}\left(\widehat{\theta}_k\right)'\right)'$.

Let a be a $k-$dimensional vector of ones, and let $\Theta_0 \equiv (a \otimes \theta_0)$. The asymptotic distri-

bution of $\widehat{\Theta}$ can now be derived by recognizing that as T goes to infinity with k fixed,

$$\sqrt{T_k} G\left(\Theta_0\right) \rightarrow N\left(0, S^*\right) \tag{11}$$

As in equation (9), S^* involves the autocovariances of $g_t\left(\Theta_0\right)$, and given the structure of this vector, we know that

$$S^* = \sum_{h=-1}^{1} E\left(g_t\left(\Theta_0\right) g_{t+h}\left(\Theta_0\right)'\right) \tag{12}$$

where

$$E\left(g_t\left(\Theta_0\right) g_t\left(\Theta_0\right)'\right) = \begin{bmatrix} C(0) & C(-1) & \cdot & \cdot & \cdot & C(-k+1) \\ C(1) & C(0) & C(-1) & \cdot & \cdot & C(-k+2) \\ \cdot & C(1) & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & C(-1) \\ C(k-1) & C(k-2) & \cdot & \cdot & C(1) & C(0) \end{bmatrix}. \tag{13}$$

$$E\left(g_t\left(\Theta_0\right) g_{t-1}\left(\Theta_0\right)'\right) = \begin{bmatrix} 0 & C(k-1) & \cdot & \cdot & \cdot & C(1) \\ 0 & 0 & C(k-1) & \cdot & \cdot & C(2) \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & C(k-1) \\ 0 & 0 & \cdot & \cdot & 0 & 0 \end{bmatrix} \tag{14}$$

$$E\left(g_t\left(\Theta_0\right) g_{t+1}\left(\Theta_0\right)'\right) = \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & 0 \\ C(-k+1) & 0 & 0 & \cdot & \cdot & 0 \\ \cdot & C(-k+1) & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ C(-1) & C(-2) & \cdot & \cdot & C(-k+1) & 0 \end{bmatrix} \tag{15}$$

The derivation of the asymptotic distribution of the parameters requires the individual

gradients of the orthogonality conditions. Place the k individual gradient matrixes, $D_{T_j}\left(\widehat{\theta}_j\right)$, onto the diagonal of a $(k \times 4)$ by $(k \times 4)$ matrix with zeros in the off-diagonal entries:

$$D\left(\widehat{\Theta}\right) = \begin{bmatrix} D_{T_1}\left(\widehat{\theta}_1\right) & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & D_{T_k}\left(\widehat{\theta}_k\right) \end{bmatrix}$$

After setting equation (10) equal to zero by the choice of the parameter estimates and by taking a first-order Taylor's series expansion, we obtain the asymptotic distribution:

$$\sqrt{T_k}\left(\widehat{\Theta} - \Theta_0\right) \rightarrow N\left(0, \Omega(\Theta_0)\right) \tag{16}$$

where the estimate of the asymptotic variance is

$$\widehat{\Omega} = D^{-1} S^* (D^{-1})' \tag{17}$$

and we have simplified the notation by defining $D = D\left(\widehat{\Theta}\right)$.

The sample counterpart of S^* depends on the autocovariances of $g_t\left(\theta_0\right)$ as indicated in equation (12), and these autocovariances can be estimated from the data in the k individual samples simply by taking the sample average counterparts of equations (13) to (15).

B.2 The Specification Test of Equality of the Individual Estimates

Given the joint distribution of the individual estimates in equation (16), it is straightforward to derive a specification test of the model. Under the null hypothesis, each of the individual estimates, $\widehat{\theta}_j$, should converge to the same value, θ_0. Hence, testing that the individual estimates are equal to each other forms a specification test of the model. To perform this

test, define the $[(k-1)\times 4]$ by $(k\times 4)$ matrix

$$R = \begin{bmatrix} I & -I & 0 & \cdot & \cdot \\ 0 & I & -I & 0 & \cdot \\ \cdot & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & 0 & I & -I \end{bmatrix}$$

where the identity matrixes are of order 4. Then, the null hypothesis is

$$R\widehat{\Theta} = 0_{(k-1)4}$$

where $0_{(k-1)\times 4}$ is a $[(k-1)\times 4]-$dimensional vector of zeros. From the joint distribution, we find that the test statistic is

$$H = (T/k)\widehat{\Theta}' R' \left(R\widehat{\Omega}R'\right)^{-1} R\widehat{\Theta}$$

which should be distributed in large samples as a chi-squared statistic with $[(k-1)\times 4]$ degrees of freedom.

B.3 The Average Estimator

Once the joint distribution of $\widehat{\Theta}$ is calculated, it is straightforward to consider the average of the estimates as an estimator of θ_0. To do this, vertically stack k identity matrixes of dimension 4 to define a matrix $A = (I, ..., I)'$. Then, the average estimator is

$$\overline{\theta} \equiv \frac{1}{k}\sum_{j=1}^{k}\widehat{\theta}_j = \frac{1}{k}A'\widehat{\Theta} \tag{18}$$

Hence, the asymptotic distribution of the average estimator is

$$\sqrt{T_k}\left(\overline{\theta}-\theta_0\right) \to N\left(0, \frac{1}{k^2}A'\widehat{\Omega}A\right)$$

or recognizing that $T_k = T/k$, gives

$$\sqrt{T}\left(\overline{\theta}-\theta_0\right) \to N\left(0, \frac{1}{k}A'\widehat{\Omega}A\right) \tag{19}$$

C Simulation and Bootstrapping

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.

References

Ait-Sahalia, Y., Mykland, P.A., and L. Zhang, 2005, How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* 18, 351-416.

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.

Brandt, M. W., and Q. Kang, 2004. On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach. *Journal of Financial Economics* 72, 217-257.

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, 2015. 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, 2015, A five-factor asset pricing model, *Journal of Financial Economics 116, 1-22.*

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

French, K. R., G. W. Schwert, and R. F. Stambaugh, 1987. Expected stock returns and volatility, *Journal of Financial Economics* 19, 3-29.

Ghysels, Eric, A. Plazzi, and R. Valkanov, 2013. The risk-return relationship and financial crises, manuscript, University of California at San Diego.

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.

Glosten, L. R., R. Jagannathan, and D. E. Runkle, 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779-1801.

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.

Hedegaard, E., and R. J. Hodrick, 2014a, Estimating the risk-return trade-off with overlapping data inference, National Bureau of Economic Research Working Paper No. 19969.

Hedegaard, E., and R. J. Hodrick, 2014b, Measuring the risk-return tradeoff with time-varying conditional covariances, National Bureau of Economic Research Working Paper No.20245.

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.