Internet Appendix for

"On the High Frequency Dynamics of Hedge Fund Risk Exposures"

This internet appendix provides supplemental analyses to the main tables in "On the High Frequency Dynamics of Hedge Fund Risk Exposures"

The first section describes the process used to create the consolidated database of hedge funds that is employed in this paper from the TASS, HFR, MSCI and CISDM databases. Prior to Table IA.IV, we describe the simulation, and prior to Table IA.V, we describe the robustness checks conducted in that table. The tables and figures are as follows:

The Consolidated Hedge Fund Database

As hedge funds can report to one or more databases, the use of any single source will fail to capture the complete universe of hedge fund data. We therefore aggregate data from TASS, HFR, CISDM, BarclayHedge and Morningstar, which together have 48,508 records that comprise administrative information as well as returns and AUM data for hedge funds, fund of funds and CTAs. However this number hides the fact that there is significant duplication of information, as multiple providers often cover the same fund. To identify all unique entities, we must therefore consolidate the aggregated data.

To do so, we adopt the following steps:

1. Group the Data: Records are grouped based on reported management company names. To do so, we first create a `Fund name key' and a `Management company key' for each data record, by parsing the original fund name and management company name for punctuations, filler words (e.g., `Fund', `Class'), and spelling errors. We then combine the fund and management name keys into 4,409 management company groups.

2. De-Duplication: Within a management company group, records are compared based on returns data (converted into US dollars), and 18,130 match sets are created out of matching records, allowing for a small error tolerance limit (10% deviation) to allow for data reporting errors.

3. Selection: Once all matches within all management company groups are identified, a single record representing the unique underlying fund is created for each match set. We pick the record with the longest returns data history available is selected from the match set, and fill in any missing administrative information using the remaining records in the match set. The process thus yields 18,130 representative fund records.

Finally, we apply the criterion that 24 contiguous months of return data are available for each of the funds in the sample we use in the paper. This brings the final number of funds in the sample to 14,194. Table IA.I below shows the number of these final funds from each of the five sources (HFR, TASS, CISDM, MSCI and BarclayHedge), and the number of these funds that are alive and defunct (either liquidated or closed).

Table IA.I
Data Sources

This table shows the number of funds from each of the five sources (HFR, TASS, CISDM, MSCI and BarclayHedge), and the number of these funds that are alive and defunct (either liquidated or closed) in the consolidated universe of hedge fund data.

Source Dataset	Number of Funds	Alive	Defunct	% Defunct
TASS	5962	2738	3224	54.076
HFR	3712	2449	1263	34.025
CISDM	2782	860	1922	69.087
BarclayHedge	966	930	36	3.727
Morningstar	772	681	91	11.788
Total	14194	7658	6536	46.048

Table IA.II
Static Factor Models for Daily and Monthly Hedge Fund Style Indexes

This table shows results from a simple two-factor model applied to five hedge fund style index returns, identified in the first row of the table. In all cases a constant is included, and two factors from the set of four daily Fung-Hsieh factors are selected using the Bayesian Information Criterion. The first row presents annualized alpha. Robust t-statistics are reported below the parameter estimates, and the R2 and adjusted R2 are reported in the bottom two rows of the table.

	Equity Hedge		Macro		Directional		Merger Arbitrage		Relative Value	
	Daily	Monthly	Daily	Monthly	Daily	Monthly	Daily	Monthly	Daily	Monthly
Alpha	1.575	1.595	3.738	3.322	3.044	3.709	5.444	5.331	-1.032	-0.473
t-stat	0.781	0.880	0.985	0.985	0.995	1.453	3.376	4.208	-0.422	-0.234
SP500	0.259	0.321			0.270	0.327	0.111	0.063	0.063	0.190
t-stat	15.395	6.070			11.298	4.029	5.181	2.332	1.811	3.935
SMB			0.070	0.111						
t-stat			2.111	0.849						
TCM10Y			-0.905	-0.376						
t-stat			-2.742	-0.405						
BAAMTSY	-2.006	-2.027			-2.720	-3.861	-0.460	-0.579	-2.544	-6.080
t-stat	-4.184	-2.728			-3.847	-3.686	-0.804	-1.935	-3.724	-9.584
R2	0.549	0.681	0.014	0.007	0.454	0.664	0.290	0.182	0.090	0.784
R2adj	0.548	0.672	0.013	-0.021	0.453	0.652	0.289	0.160	0.089	0.778

Table IA.III
Comparison of results based on simple and log returns

This table reports the results from the estimation of the linear model for hedge fund risk exposures, described in Section 2.2.1, and can be compared with the results presented in Table II of the paper. The columns labeled “True” are based on daily simple returns that are compounded to compute monthly simple returns. The columns labeled “Approx” use simple returns but ignore compounding when computing monthly returns. The columns labeled “Log” are based on log returns, which cumulated to monthly returns exactly, but render the factor model only approximate.

	Equity Hedge			Macro			Directional			Merger Arbitrage			Relative Value		
	True	Approx	Log	True	Approx	Log	True	Approx	Log	True	Approx	Log	True	Approx	Log
alpha	3.127	3.187	3.064	3.586	3.452	3.200	7.702	7.778	7.612	4.494	4.462	4.450	3.891	4.006	3.980
s.e.	2.058	2.051	2.053	3.864	3.839	3.859	3.230	3.206	3.213	1.404	1.396	1.402	1.817	1.818	1.820
beta1	0.328	0.326	0.327	0.125	0.123	0.121	0.250	0.248	0.248	0.104	0.104	0.104	0.076	0.076	0.076
s.e.	0.051	0.051	0.051	0.145	0.144	0.144	0.082	0.081	0.081	0.035	0.034	0.035	0.045	0.045	0.045
beta2	-1.732	-1.761	-1.802	-0.829	-0.810	-0.766	-3.202	-3.239	-3.293	-0.617	-0.592	-0.614	-5.718	-5.750	-5.795
s.e.	0.698	0.695	0.698	1.255	1.247	1.248	1.040	1.032	1.037	0.476	0.473	0.477	0.616	0.616	0.619
gam1	0.010	0.010	0.010	0.001	0.002	0.003	0.003	0.002	0.002	0.009	0.009	0.009	-0.010	-0.010	-0.010
s.e.	0.006	0.006	0.006	0.025	0.025	0.025	0.009	0.009	0.009	0.004	0.004	0.004	0.005	0.005	0.005
gam2	0.027	0.023	0.027	0.075	0.071	0.069	-0.025	-0.028	-0.025	0.028	0.029	0.027	-0.014	-0.004	0.002
s.e.	0.113	0.112	0.113	0.150	0.149	0.148	0.170	0.169	0.170	0.077	0.076	0.077	0.100	0.100	0.100
delta1	0.031	0.033	0.034	-0.024	-0.024	-0.022	0.051	0.054	0.054	-0.003	-0.004	-0.003	0.047	0.050	0.050
s.e.	0.012	0.012	0.012	0.073	0.073	0.071	0.018	0.018	0.018	0.008	0.008	0.008	0.011	0.011	0.011
delta2	0.200	0.280	0.325	1.128	1.103	1.082	0.877	1.040	1.119	0.200	0.128	0.187	-0.084	0.007	0.099
s.e.	0.791	0.788	0.763	0.776	0.771	0.763	1.176	1.167	1.130	0.539	0.536	0.521	0.698	0.698	0.676
R2	0.723	0.730	0.734	0.049	0.049	0.049	0.696	0.710	0.715	0.262	0.250	0.262	0.834	0.843	0.846
R2adj	0.699	0.706	0.710	-0.034	-0.035	-0.035	0.661	0.677	0.683	0.197	0.183	0.197	0.820	0.829	0.832
pval	0.012	0.010	0.008	0.551	0.558	0.557	0.068	0.043	0.035	0.146	0.122	0.150	0.000	0.000	0.000

Simulation Study

We consider a simulation study designed to further investigate the accuracy of our proposed estimation method. For simplicity, we consider a one-factor model for a hypothetical hedge fund, and as in our main empirical analysis, we allow factor exposures to vary at both the daily and monthly frequencies. We simplify the notation and assume that each month contains exactly 22 trading days. This yields a process for daily hedge fund returns as:

$$r_d^* = \alpha + \beta f_d^* + \gamma f_d^* Z_{d-1} + \delta f_d^* Z_{d-1}^* + \varepsilon_{R,d}^*, \quad d = 1, 2, ..., 22 \times T, \tag{1}$$

The parameter β captures the average level of beta for this fund, γ captures variations in beta that are attributable to the monthly variable Z_d, and δ captures variations in beta that are attributable to the daily variable Z_d^*. If we aggregate this process up to the monthly frequency we obtain:

$$r_t = 22\alpha + \beta f_t + \gamma f_t Z_{t-1} + \delta \sum_{j=0}^{21} f_{22t-j}^* Z_{22t-j-1}^* + \varepsilon_{R,t}, \quad t = 1, 2, ..., T. \tag{2}$$

where $r_t \equiv \sum_{j=0}^{21} r_{22t-j}^*$, is the monthly equivalent of the daily variable in the above specification, and analogously for f_t and Z_t. The parameters α, β and γ are all estimable using only monthly data; the focus of this simulation study is our ability to estimate δ, and whether attempting to do so adversely affects our estimates of the remaining parameters.

We next specify the dynamics and distribution of the factor and the conditioning variable. To allow for autocorrelation in the conditioning variable (as found in such variables as volatility and turnover) we use an AR(1) process for Z_d^* :

$$Z_d^* = \phi_Z Z_{d-1}^* + \varepsilon_{Z,d}^*$$

The conditioning variable is de-meaned prior to estimation, and so the omission of an intercept in the above specification is without loss of generality. We also assume an AR(1) for the factor returns, to allow for the possibility that these are also autocorrelated:

$$f_d^* = \mu_F + \phi_F \left(f_{d-1}^* - \mu_F \right) + \varepsilon_{F,d}^*$$

Finally, we assume that all innovations are normally distributed, and we allow for correlation between the factor innovations and the innovations to the conditioning variable:

$$\left[\varepsilon_{R,d}^*, \varepsilon_{F,d}^*, \varepsilon_{Z,d}^* \right]' \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\varepsilon R}^2 & 0 & 0 \\ \cdot & \sigma_{\varepsilon F}^2 & \rho_{FZ} \sigma_{\varepsilon F} \sigma_{\varepsilon Z} \\ \cdot & \cdot & \sigma_{\varepsilon Z}^2 \end{bmatrix} \right)$$

To obtain realistic parameter values for the simulation we calibrate the model to the results obtained when estimating the model using daily HFR equity hedge index returns. This leads to the following parameters for our simulation:

$$\alpha = 2/(22\times 12),\ \beta = 0.4,\ \gamma = 0.002,\ \delta = -0.004$$
$$\mu_F = 10/\left(22\times 12\right),\ \sigma_F = 20/\sqrt{22\times 12},\ \sigma_Z = 10,\ \sigma_{\varepsilon R} = \sqrt{0.1}$$

Thus we assume that the fund generates 2% alpha per annum with an average beta of 0.4, and a daily beta that varies with both daily and monthly fluctuations in the conditioning variable (Z^* and Z). The factor is assumed to have an average return of 10% per annum and an annual standard deviation of 20%. The conditioning variable has daily standard deviation of 10 (similar to the VIX), and the innovation to the returns process has a daily variance of 0.1, which corresponds to an R^2 of around 0.6 in this design.

We vary the other parameters of the returns generating process in order to study the sensitivity of the method to these parameters. We consider:

$$\phi_Z \in \{0, 0.5, 0.9\},\ \phi_F \in \{-0.2, 0, 0.2\},\ \rho_{FZ} \in \{0, 0.5\},\ T \in \{24, 60, 120\}$$

Thus, we allow the conditioning variable to vary from *iid* ($\phi_Z = 0$) to persistent ($\phi_Z = 0.9$); we allow for moderate negative or positive autocorrelation in the factor returns; we allow for zero or positive correlation between the factor and the conditioning variable; and we consider three sample sizes: 24 months, 60 months or 120 months, which covers the relevant range of sample sizes in our empirical analysis (the average sample size in our empirical application is 62 months). We simulate each configuration of parameters 1000 times, and report the results in Table IA.IV.

The results show that the estimation method proposed in the paper performs very well in all the scenarios that we consider. In the "base" scenario, even with just 60 months of data we are able to reasonably accurately estimate the parameters of this model, including the parameter δ, which allows us to capture daily variation in hedge fund risk exposures. Across a range of different sample sizes, degrees of autocorrelation, and correlation with the factor return, the estimation method performs well: the 90% confidence interval of the distribution of parameter estimates contains the true parameter in all ten scenarios that we consider.

Table IA.IV
Results from a simulation study of the estimation method

This table reports the mean and standard deviation, across 1000 independent simulation replications, of estimates of the parameters of a model of time-varying factor exposures. The results for ten different simulation designs are presented. Simulation design parameters are presented in the first panel of the table, and the mean and standard deviation of the simulation distribution of parameter estimates are presented in the second and third panels. The true values of the four parameters are presented in the first column of the table. The values for alpha, gamma and delta are scaled up by a factor of 100 for ease of interpretability.

			1	2	3	4	5	6	7	8	9	10
		True values	Base scenario	Short sample	Long sample	Low autocorr in Z	High autocorr in Z	Corr b/w F, Z	Neg autocorr in F, rhoFZ=0	Pos autocorr in F, rhoFZ=0	Neg autocorr in F, rhoFZ=0.5	Pos autocorr in F, rhoFZ=0.5
	T		60	24	120	60	60	60	60	60	60	60
	rhoFZ		0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.5	0.5
	phiZ		0.5	0.5	0.5	0.0	0.9	0.5	0.5	0.5	0.5	0.5
	phiF		0.0	0.0	0.0	0.0	0.0	0.0	-0.2	0.2	-0.2	0.2
Mean	Alpha*100	0.758	0.715	0.827	0.794	0.729	0.773	0.812	0.780	0.735	0.725	0.721
Mean	Beta	0.400	0.400	0.397	0.399	0.400	0.401	0.400	0.401	0.400	0.399	0.401
Mean	Gamma*100	0.200	0.198	0.198	0.200	0.198	0.200	0.200	0.199	0.201	0.198	0.199
Mean	Delta*100	-0.400	-0.391	-0.400	-0.409	-0.399	-0.404	-0.410	-0.381	-0.392	-0.397	-0.394
St dev	Alpha*100		0.089	0.146	0.062	0.089	0.092	0.194	0.085	0.091	0.191	0.191
St dev	Beta		0.035	0.060	0.024	0.033	0.035	0.035	0.042	0.031	0.041	0.029
St dev	Gamma*100		0.005	0.009	0.003	0.008	0.003	0.005	0.006	0.004	0.005	0.004
St dev	Delta*100		0.038	0.062	0.026	0.035	0.052	0.034	0.039	0.034	0.037	0.031

Robustness Checks

Table IA.V presents robustness checks used to identify whether the method proposed in this paper performs well over different sample periods, over different samples of funds, and under different transformations of hedge fund returns. In all these robustness checks, we consider the linear model for g(Z), and use dLevel as Z. We consider a model using both daily and monthly conditioning information, as well as a model based only on monthly information.

The first robustness check that we run is to split the sample period into two halves, with the second half beginning in 2002, after the NASDAQ crash, and extending up to 2009, including the credit crisis period. We find that our method performs relatively less well in the early period, with 14.2% of funds rejecting the null of no significant interaction variables. In the second sub-period, we find almost 30% of the funds selecting interactions. This might be explained by the population of funds shifting towards funds with faster-moving trading strategies, which would suggest that our method is more appropriate to use in the contemporary setting. The model using only monthly information performs equally well in both sub-periods.

Next, we investigate our use of the Getmansky, Lo and Makarov (2004) unsmoothing of hedge fund returns. Our baseline results us 2 lags when implementing this model, and we find that our results are essentially unaffected by the choice of using more lags (4) or no lags.

We then condition on the length of available return history for the funds, and find that the both models perform better for funds with longer return histories, which is understandable given that it gives us more information with which to pick up changes in risk exposures.

Finally, we condition on the size of the fund. We find that both methods work better for larger hedge funds (measured by the average AUM over the fund's lifetime), and worse for funds in the smallest tercile of AUM.

Table IA.V
Robustness Checks

This table presents the proportion of funds with significant time variation based on our linear model (g(Z) linear in Z, allowing for both daily and monthly variation in Z; or only monthly variation in Z), using dLevel as Z. These proportions selected change when the specification is altered according to the robustness checks itemized in the rows of the first column. The first row is taken from the main table of linear model results (Table III) and is repeated here for ease of reference. The first set of robustness checks splits the sample period into two halves; the second set alters the type of "unsmoothing" done to the returns prior to testing; the third set sorts funds according to their history lengths in the consolidated database; and the fourth set sorts funds according to their average assets under management.

Robustness Test	N(Funds)	Daily and Monthly	Only Monthly
Main Results in Paper	14194	25.673	15.393
Sample Period			
Earlier period (1994:01-2001:12)	5106	14.219	14.044
Later period (2002:01-2009:06)	11386	29.832	15.168
Smoothing			
Raw returns (no GLM)	14194	24.600	13.800
Longer (4) MA lags in GLM model	14194	24.700	16.300
Fund History Length			
24<=N(Fund Observations)<36	2697	13.200	7.379
36<=N(Fund Observations)<60	4394	20.096	11.174
N(Fund Observations)>=60	7103	31.635	20.301
Fund Size			
Avg AUM <= 33rd Prctile	3963	15.821	11.380
33rd Prctile< Avg AUM <= 66th Prctile	4085	24.357	16.010
Avg AUM > 66th Prctile	3963	32.677	18.193

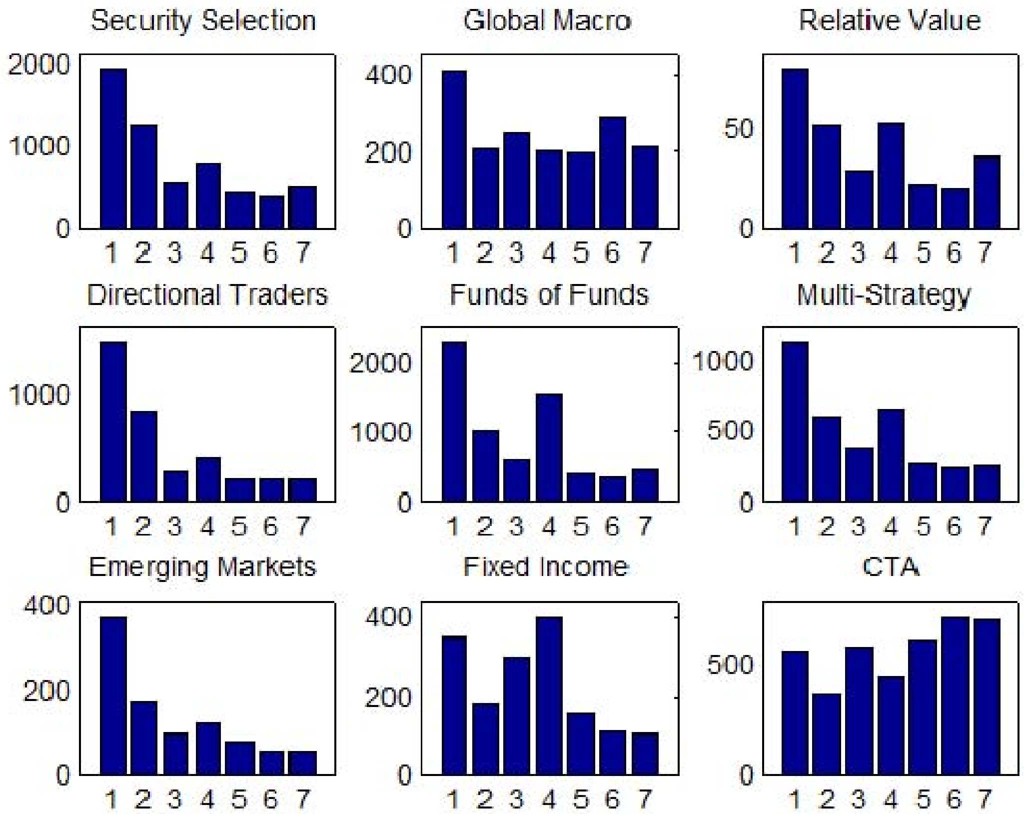

Figure IA.1: Static factors selected by strategy

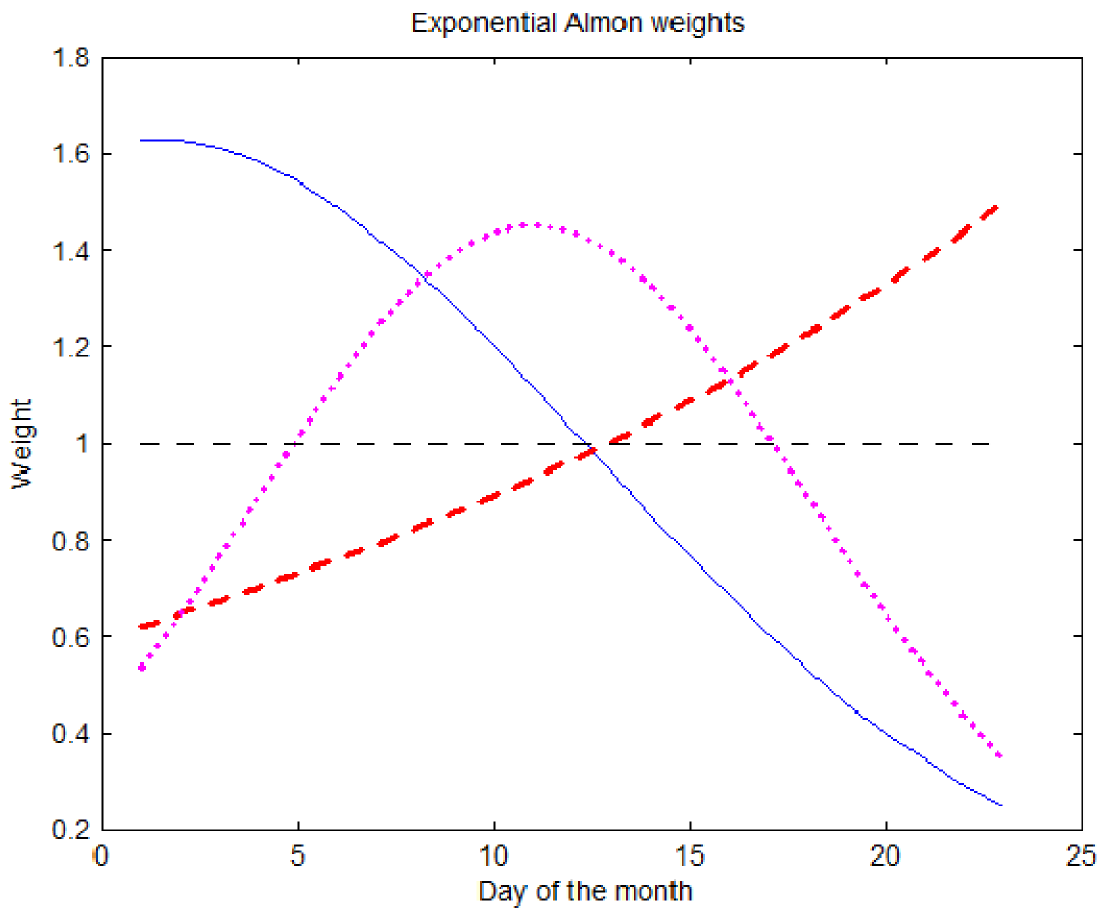

Figure IA.2: Some possible shapes with exponential-Almon weight functions