Web Appendix to:

Long-Term versus Short-Term Contingencies in Asset Allocation

A Numerical Implementation

Let Z_t denote the vector of time t long-term and short-term components for a specific state variable, where both components are divided by their time-series standard deviation to make them comparable. To enhance numerical stability of the non-parametric estimation and subsequent GMM optimization, we also standardize the index βZ_t by its time-series standard deviation for every value of β. To save on notation, define $z_t = z_t(\beta) = \beta Z_t$.

For each value of index z_t, we find optimal asset allocations by maximizing the weighted average utility in (4) using a kernel weighting scheme that is explained further below. The asset weights are constrained to lie between 0 and 1. If the optimal asset weights turn out to lie on the boundary, we solve for the corresponding Kuhn-Tucker multipliers that will equate the first order condition to zero as described in Equation (12). This is a linear system of equations for every z_t. The optimal asset allocations and possible Kuhn-Tucker multipliers for every z_t are used to compute the moment variables $\tilde{m}_{t+1}(z_t) \otimes g(Z_t)$ and the GMM criterion as a function of β, which is a function of ϕ through $\beta(\phi) = (\cos(\phi), \sin(\phi))$. The GMM criterion is then minimized to find the estimate of ϕ.

The derivatives of the objective function needed to compute the standard errors are obtained by numerical differentiation. The GMM criterion function can become very small

numerically, resulting in instabilities particularly for the initial choices of the weighting matrix. We avoid this problem by re-scaling the objective function using a fixed large constant. The problem generally disappears after the second round estimation of the weighting matrix. The different scales of the elements of the moment vectors are then accounted for by the weighting matrix, which captures the variance of each element of the moment vector.

A Weighting Function and Choice of the Bandwidth

For the weights $\omega_t(z) = \omega(z_t, z)$ used in (4) and (5), we take a Gaussian kernel

$$\omega(z_t, z) = (2\pi)^{-1/2} \exp\left(-0.5(z_t - z/h)^2\right). \tag{A1}$$

where h is the so-called bandwidth. The choice of the bandwidth has a direct effect on the optimal value of ϕ, and thus β. A high value of h leads to smoother weights in the conditional moment condition and therefore to smaller differences between states. A low value of h on the other hand leads to more differentiation between states, but also uses less observations in determining moment conditions. Brandt (1999) proposes the bandwidth to be chosen as

$$h = \lambda_b \sigma_z n^{(1/(K+4))}, \tag{A2}$$

where σ_z is the standard deviation of z_t, n is the number of observations, K is the dimension of z_t ($K = 1$), and λ_b is a tuning parameter that should be chosen to minimize the standard deviation of β. Examples of the resulting objective functions for different values of the bandwidth parameter are presented in Figure A1 for the term spread (TR) using the CF(12) and CF(24) filter.

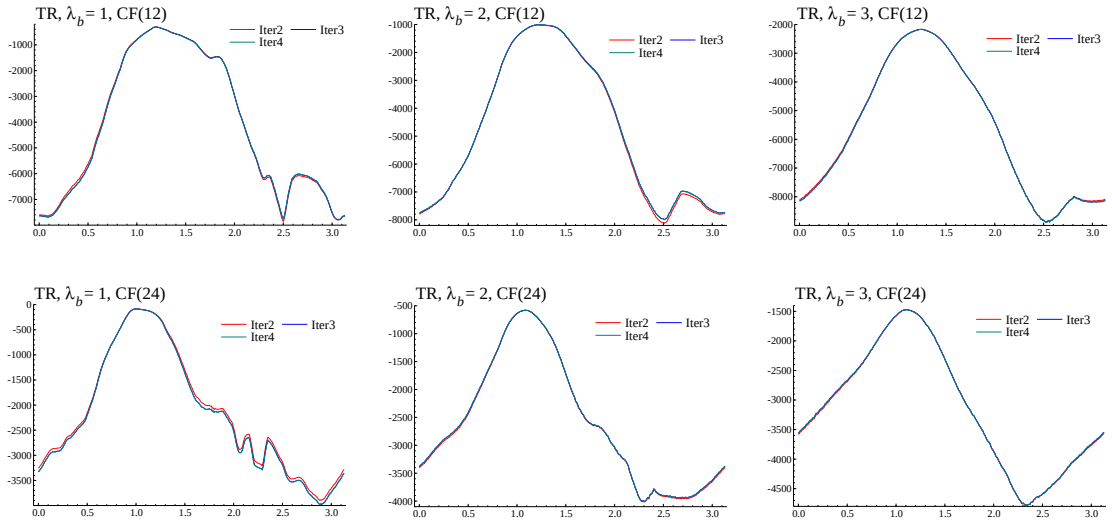

Figure A1: GMM objective functions and different bandwidths

This figure presents plots of (the negative value of) the GMM objective function as a function of the angle ϕ for stock market trend (TR) and different λ_b that determine the bandwidth parameter as described in Equation (A2). The results are presented for filters CF(12) (left) and CF(24) (right). For each filter, the upper, middle and lower graphs correspond to $\lambda_b = 1$, 2, and 3 respectively.

For $\lambda_b = 3$, the objective function is relatively smooth and clearly unimodal. The objective function becomes more irregular and multimodal for $\lambda_b = 1$. This causes a potential issue with multiple local optima of the GMM objective function. Therefore we use $\lambda_b = 2$ as a benchmark in our empirical analysis, which strikes a balance between these two. Choosing λ_b between 2 and 3 typically does not change the optimal ϕ substantially, and therefore has a minor effect on the relative importance of short-term and long-term components for asset allocation decisions.

B Adaptive Bandwidth

For some value of β and some state variable, the index z_t is fat-tailed and skewed, such that there are tail observations z_t that are far from the bulk of the data. This causes end-point problems in estimating the optimal asset allocation function $x(z_t)$ non-parametrically:

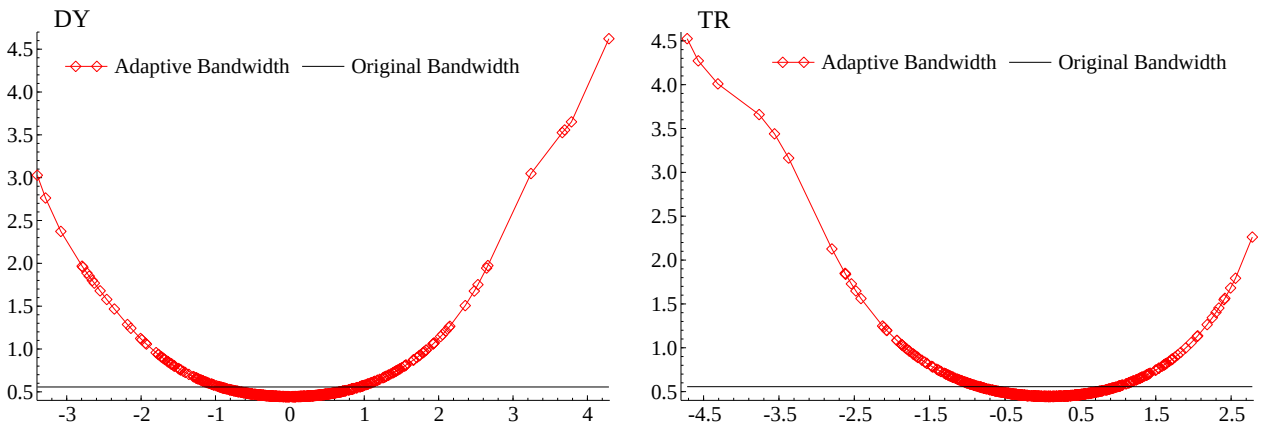

Figure A2: Original and Adaptive Bandwidth

This figure presents plots of original and adaptive bandwidth for different value of optimal index z for dividend yield (DY) and stock market trend (TR) using CF(12) for the decomposition. Here we use $\lambda_b = 2$ and original Bandwidth for DY and TR is 0.556. For each value of the index z, the adaptive bandwidth are calculated by method proposed by Abramson (1982).

the weights in Equation (A1) will be concentrated on few observations only and the asset allocation estimated for these observations will have a high standard error. To solve this issue, we use an adaptive bandwidth method proposed by Abramson (1982). This method increases the bandwidth for tail observations where neighboring observations are sparse, and conversely shrinks the bandwidth if neighboring observations are plenty. This renders the kernel weights more smooth, particularly in the z_t areas where it matters most. Figure A2 shows the value of the adaptive bandwidth for the optimal index found for the dividend yield (DY) and stock market trend (TR) using $\lambda_b = 2$. The original fixed bandwidth of Equation (A2) is also presented in each panel by a horizontal line. We see that the adaptive bandwidth is higher in the tails, where neighboring z_t values are sparser. By contrast, in the middle range, the adaptive bandwidth is slightly lower than the fixed bandwidth.

C Additional Results

Table A1: Optimal component weights for different risk aversion levels

The table is analogous to Table 2 and presents the GMM estimates of the optimal weights for different levels of the risk aversion parameter γ. The decomposition uses the (recursively implemented) CF(12) filter. The state variables are the log dividend yield (DY), stock market trend (TR), term spread (TS), default spread (DS), short rate (SR), and dividend growth rate (DG).

		DY	**TR**	**TS**	**DS**	**SR**	**DG**
$\gamma = 1$	$\hat{\phi}$	1.911	1.351	0.510	1.541	1.081	2.972
		(0.336)	(0.492)	(0.330)	(0.838)	(0.367)	(0.561)
	β_τ	-0.334	0.218	0.873	0.030	0.471	-0.986
	β_c	0.943	0.976	0.488	1.000	0.882	0.169
$\gamma = 5$	$\hat{\phi}$	1.951	1.221	0.370	1.621	0.870	2.911
		(0.271)	(0.334)	(0.352)	(0.818)	(0.339)	(0.445)
	β_τ	-0.371	0.343	0.932	-0.050	0.644	-0.974
	β_c	0.929	0.939	0.362	0.999	0.765	0.228
$\gamma = 10$	$\hat{\phi}$	1.971	1.211	0.310	1.681	0.850	2.671
		(0.239)	(0.331)	(0.309)	(0.684)	(0.407)	(0.521)
	β_τ	-0.390	0.352	0.952	-0.110	0.660	-0.891
	β_c	0.921	0.936	0.305	0.994	0.752	0.453

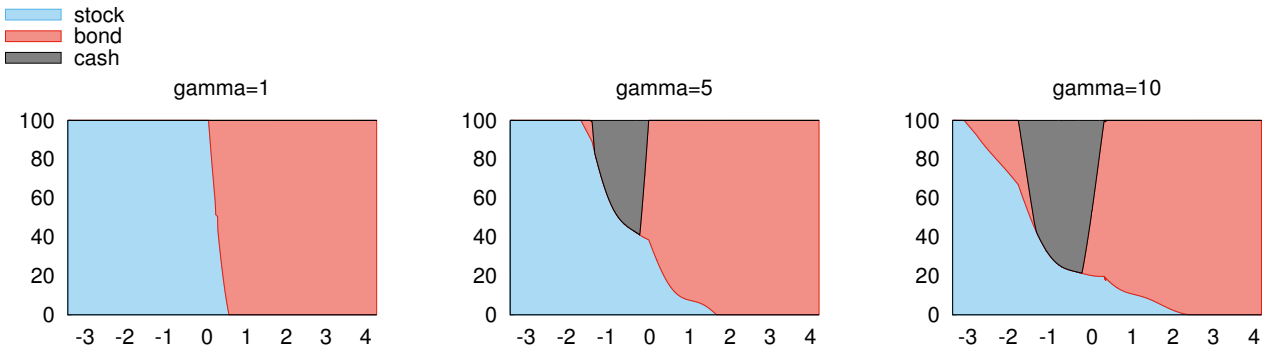

Figure A3: Optimal allocations for varying risk aversion

Each graph presents the optimal asset allocation as a function of the index βZ_t computed using the optimal β from Table A1, and Z_t holding the long-term and short-term component of the log dividend yield. The CF(12) filter is used for the decomposition. The index βZ_t is standardized its time-series standard deviation. The vertical axis gives the percentage invested in stocks, bonds, and cash. The different panels are for a CRRA utility function with different relative risk aversion parameter $\gamma = 1, 5, 10$.

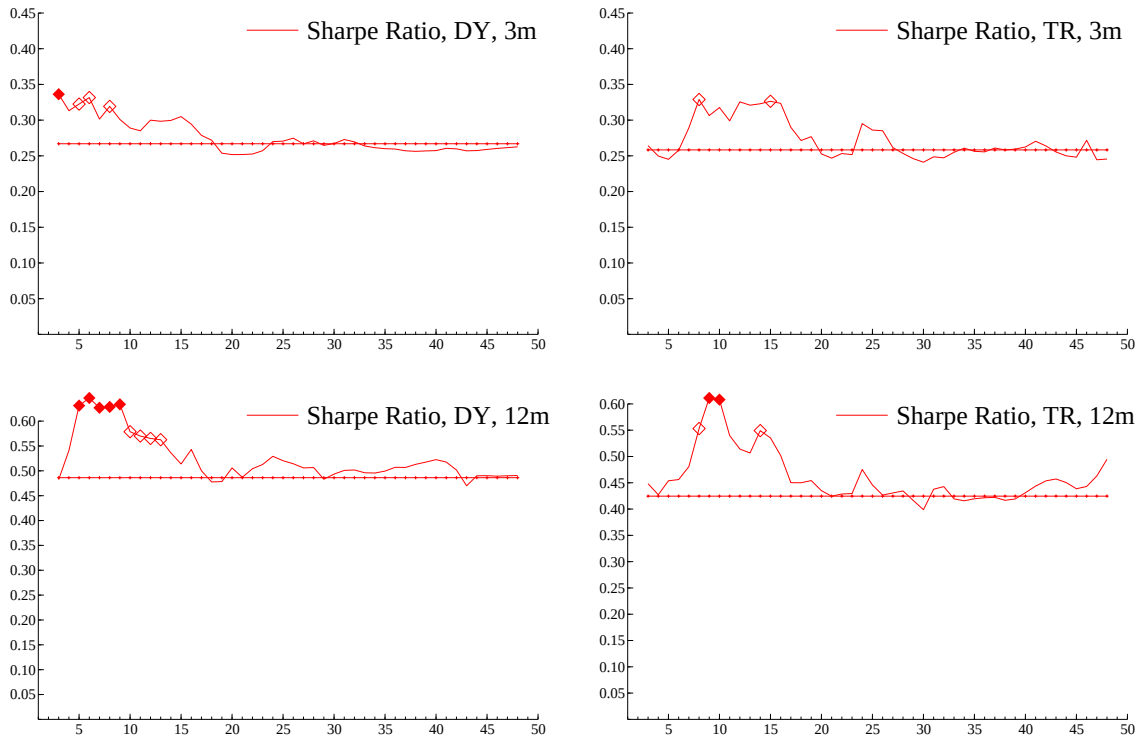

Figure A4: Sharpe ratios of investment strategies for all filters and 3m and 12m holding periods

The graphs present the Sharpe ratios for different portfolio strategies using the recursively implemented Christiano-Fitzgerald (CF(p)) filter for $p = 3, 4, \ldots, 48$ and log dividend yield (DY, left-hand side) and stock market trend (TR, right-hand side) as state variables. In each graph, the horizontal line depicts the Sharpe ratio of the investment strategy that uses the non-decomposed state variable. The filled and empty diamonds highlight CF(p) filters for which we can reject the null hypothesis that the non-decomposed state variable yields the same result as the decomposed state variable at 5% and 10% significance level, respectively. The sample period is Jan 1963 to December 2012 with non-overlapping data.

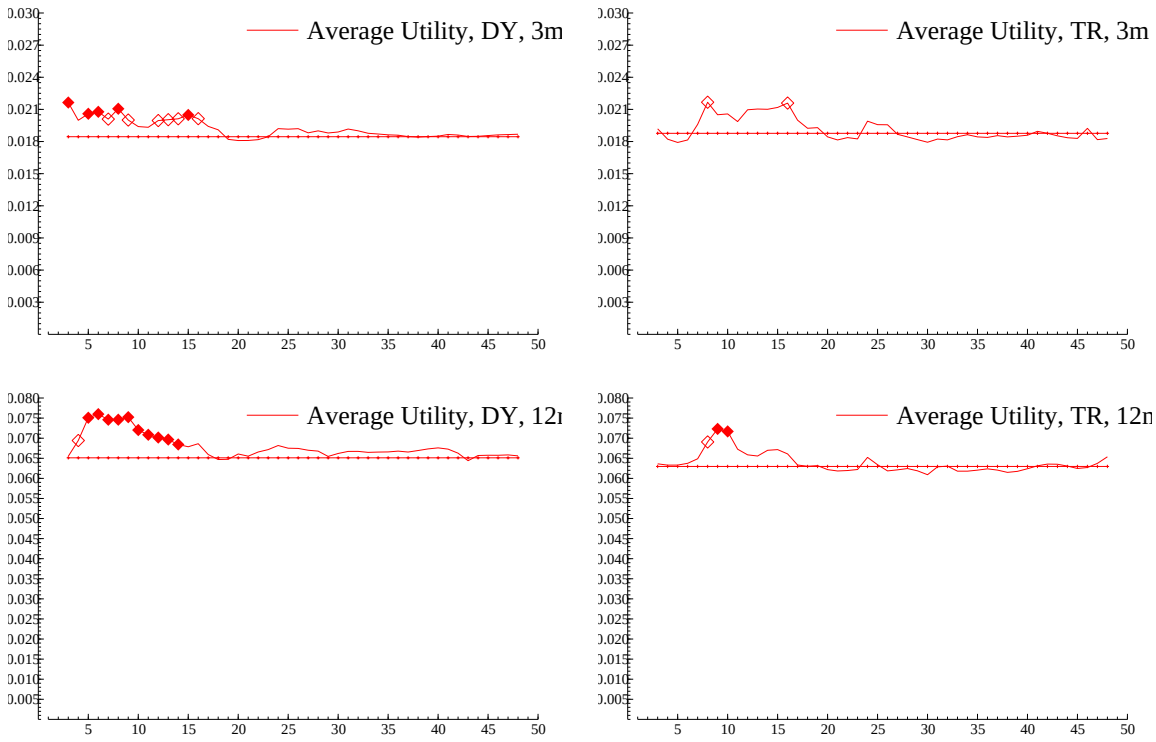

Figure A5: Average utilities of investment strategies for all filters and 3m and 12m holding periods

The graphs present the average utilities for different portfolio strategies using the recursively implemented Christiano-Fitzgerald (CF(p)) filter for $p = 3, 4, \ldots, 48$ and log dividend yield (DY, left-hand side) and stock market trend (TR, right-hand side) as state variables. In each graph, the horizontal line depicts the average utility of the investment strategy that uses the non-decomposed state variable. The filled and empty diamonds highlight CF(p) filters for which we can reject the null hypothesis that the non-decomposed state variable yields the same result as the decomposed state variable at 5% and 10% significance level, respectively. The sample period is Jan 1963 to December 2012 with non-overlapping data.