Low-High Basis Factor in the Commodity Futures Market

Daehwan Kim

Department of Economics, Konkuk University, Seoul, South Korea

+82 2 450 3937, dkim@konkuk.ac.kr

May 25, 2013

Abstract

Previous studies have shown that the profit to the "buy low-basis commodities and sell high-basis commodities" strategy is an effective pricing factor for commodity futures returns. Let us call this factor the low-high basis factor, or the LHB factor, in short. The goal of this paper is to better understand the LHB factor. We make three contributions to the literature: First, we present additional tests of the significance of the LHB factor in the factor model framework, and confirm the relevance of the LHB factor. Second, we investigate the relationship between the LHB factor and other factors of stock, FX, and commodity markets, and we show that the LHB factor is not redundant. Finally, we examine the time-variation in, and the predictability of, the LHB factor returns. We find that the LHB factor returns can be predicted by spread variables, especially by the implied volatility spread. Our findings support the view that the LHB factor is a risk premium rather than a behavioral phenomenon. We discuss implications on strategic and tactical asset allocations.

Keywords

Commodity futures returns, Low-high basis factor, Implied volatility, Strategic asset allocation, Tactical asset allocation

JEL Classification

G10, G12, G13

1. Introduction

Since Dusak (1973), economists have attempted to explain the cross-section of commodity futures returns with the capital asset pricing model (CAPM) and its variants. Whereas Dusak (1973) has estimated the standard CAPM, Breeden (1980) and Jaganathan (1985) have examined consumption and intertemporal CAPM. Carter, Rausser, and Schmitz (1983) and Bessembinder (1992) have modified these models by including a hedging pressure variable. Overall, these attempts have had only limited success.

Recently, a number of studies point toward a more promising direction: replacing the market return in the CAPM with a new factor that is based on a popular investment strategy in the commodity futures market. The strategy buys low-basis commodities and sells high-basis commodities. The profit to this strategy turns out to be an effective pricing factor for commodity futures returns. Basis is defined as the ratio of the futures price over the spot price (or a longer-term futures price over a shorter-term futures price).[1] Thus, in this strategy, we buy commodities with downward sloping term structure, and sell commodities with upward sloping term-structure.[2] We call the profit to this strategy "the low-high basis factor," or "the LHB factor," in short. Gorton and Rouwenhort (2006), Fuertes, Miffre, and Rallis (2010), and Gorton, Hayashi, and Rouwenhorst (2012) report significant premiums accruing to the LHB factor.[3] Erb and Harvey (2006) discuss a market-timing strategy based on this factor. De Roon and Szymanowska (2010) and Szymanowska, de Roon, Nijman, and Goorbergh (2013) report that a single- and a two-factor models with the LHB factor explain a significant part of the cross-section of commodity futures returns.[4]

The goal of this paper is to better understand the LHB factor. From a review of the convenience yield theory (Kaldor (1939), Working (1949), Brennan (1958), and more recently, Litzenberger and Rabinowitz (1995)) and the normal backwardation theory (Keynes (1930) and Cootner (1960)), we

---

[1] Percentage basis is a more accurate term. We use the term basis for brevity.

[2] This strategy is often called the backwardation-contango strategy. Different people assign somewhat different meanings to the term "backwardation." Some compares futures price to the spot price, whereas others compare it to the expected spot price. To minimize confusion, we avoid the name backwardation-contango strategy.

[3] Gorton, Hayashi, and Rouwenhorst (2012) define basis in the opposite way, so a low basis in our sense corresponds to a high basis in their sense. Our definition of basis corresponds to that of Fama and French (1987).

[4] Fama and French (1987) have also contributed to the development of this idea by emphasizing the fact that the basis can be decomposed into the expected change in the spot price and the return to the futures.

formulate risk-based, quantity pressure-based, and behavioral bias-based interpretations of the LHB factor return. While distinguishing among alternative interpretations is not the main focus of this paper, our analyses are more supportive of the risk-based interpretation, i.e. that the LHB factor return is a risk premium.

We carry out three sets of empirical analyses using commodity futures market data for the period between 1981 and 2010. First, we review the evidence for the effectiveness of the LHB factor. After confirming the significant premium accruing to the LHB factor, we perform formal econometric comparison between the factor models with and without the LHB factor. Our model comparison is based on the two-pass cross-sectional regression R squared and the Hansen-Jaganathan (1997) distance as suggested by Kan, Robotti, and Shanken (2013) and Kan and Robotti (2009) as well as the Newey-West (1987b) test as suggested by Cochrane (1996). These tests show that the LHB factor is indeed making a significant contribution over and above the market factor.

Second, we examine the relationship between the LHB factor and the other factors of the stock, FX, and commodity markets. In particular, we compare the LHB factor to the currency carry returns of Lustig, Roussanov, and Verdelhan (2008) and the HML returns of Fama and French (1993). These returns share one important property with the LHB factor: they are all constructed from yield differences. A currency carry trade buys high-yield (interest) currencies and sells low-yield currencies[5]. Under the covered interest parity[6], the futures of high-yield currencies have low basis whereas the futures of low-yield currencies have high basis. Thus, the currency carry strategy is essentially identical to the strategy underlying the LHB factor. This relationship has been noted by Szymanowska, de Roon, Nijman, and Goorbergh (2013) and others[7]. The HML returns are also based on yield differences; the relevant yield in this case is the dividend or earnings yield. The HML returns are based on the strategy of buying high book-to-market stocks and selling low book-to-market stocks. High (low) book-to-market stocks tend to be high (low) dividend- and earnings-yield stocks. Despite

---

[5] Pojarliev and Levich (2008) use the carry factor in analyzing the performance of currency fund managers. Kim and Song (2012) also use this factor in a multi-factor model of currency returns.

[6] When the covered interest parity (CIP) holds, basis equals the interest rate of the base currency over the interest rate of the given currency. We assume that exchange rates are quoted such that the value of the given currency is expressed in terms of the number of the base currency unit.

[7] Fama and French (1987) make it clear that basis may have similar roles in the currency market and in the commodity futures market. They decompose the basis of commodity futures using the framework that has been developed for the currency market basis by Fama (1984).

these similarities, our analysis shows that the LHB returns do not co-move with these and other factors of stock, FX, and commodity markets. To provide economic meaning to this finding, we examine the implication on the strategic asset allocation; i.e. we compare the investment possibility sets that are created with and without the LHB factor. We find that the investment possibility set with the LHB factor is meaningfully larger than the one without it.

Finally, we examine the time-variation in, and the predictability of, the LHB factor returns. The theories of convenience yield and the normal backwardation suggest a possible relationship between the LHB factor and the business cycle. Our analysis indicates no strong correlations between the LHB factor and the business cycle, though some negative correlation is detected between the LHB factor and the US industrial production. We explore the predictability of the LHB factor from spread variables. When a factor is created out of sorted portfolios, the spread in the sorting variable and other related variables can often predict the factor returns. See, for example, the papers by Asness, Friedman, Krail, and Liew (2000), by Cohen, Polk, and Vuolteenaho (2003), and by Stivers and Sun (2010)[8]. We consider three spread variables: the basis spread, the hedging pressure spread, and the implied volatility spread. Data for these three variables are easily available, and previous studies have examined these variables in various contexts. Our predictive regression shows that the volatility spread is a significant predictor of the LHB factor return. As an alternative measure of predictability, we calculate the returns to market timing strategies (a.k.a. tactical asset allocation strategies), where we follow the strategy underlying the LHB factor only if a predictor variable is high enough. This analysis indicates that all three spread variables are useful, and the volatility spread is the most effective predictor. This finding is supportive of the view that the LHB factor is a risk premium.

The remainder of this paper is organized as follows. In Section 2, we review the theories of basis and futures returns. Section 3 describes the data that is used in the analysis. Basis sorted portfolios are examined in Section 4. Factor models are estimated and compared in Section 5. Section 6 examines the relationship between the LHB factor and other factors of the stock, FX, and commodity markets, whereas Section 7 explores the predictability the LHB factor. We conclude in Section 8. Further details on data and methodology are included in the appendix.

---

[8] In a cross-country analysis, Kim (2012) shows that the spread variables help to explain the cross-country variation of the equity value premium as well.

2. Basis and Futures Returns: A Review of Theories

In this section, we review the theories of basis and futures returns. We use the following notation. $F_{t,n}$ is the end-of-period-t price of the futures contract which expires at the end of period $\mathrm{t+n}$. That is, n is the time-to-maturity of the contract. S_t is the spot price at the end of period t. r is the deposit rate, u is the cost of storage, and c is the convenience yield. r, u, and c are expressed as continuously compounded rates for one period. Their values may change over time; we do not add time subscript for the brevity of notation. With this notation, the futures price is related to the spot price in the following way:

$$F_{t,n} = S_t e^{(r+u-c)n} \tag{1}$$

This relationship is referred to the theory of convenience yield; it is also known as the theory of storage and the cost-of-carry model. See Kaldor (1939), Working (1949), and Brennan (1958) for earlier development. Needless to say, the exact content of this theory depends on how convenience yield c is defined. If there are any cash payments such as dividends or interests, they are always included in c. In addition, c may include non-cash income such as the option value of holding inventory, as in Litzenberger and Rabinowitz (1995).

Futures price $F_{t,n}$ can be compared to the spot price at the time of maturity S_{t+n}. According to the normal backwardation theory of Keynes (1930), $F_{t,n}$ should be lower than the expected value of S_{t+n} and the difference is the reward for the speculators who provide liquidity to short sellers. Thus,

$$F_{t,n} = E_t(S_{t+n}) - \pi \tag{2}$$

where π is a positive number, which could be varying over time. π is often called the risk premium. The expectation hypothesis, on the other hand, requires that $F_{t,n}$ should be identical to the expected value of S_{t+n}, in which case π should be zero. Numerous tests have been conducted of whether π is indeed zero. The majority rejects the hypothesis. Of course, the rejection of $\pi = 0$ does not necessarily support the Keynesian interpretation of π.

We define basis as the percentage difference between the nearest-maturity futures price F_{t,n_1} and the spot price S_t, adjusted for the time-to-maturity of the contract: $100\left[\left(F_{t,n_1}/S_t\right)^{1/n_1} - 1\right]$. When the spot price is not readily available (or not reliable), we calculate the basis as the percentage difference between the prices of two nearest-maturity contracts: $100\left[\left(F_{t,n_2}/F_{t,n_1}\right)^{1/(n_2-n_1)} - 1\right]$. The theory of convenience yield implies the following relationship between the basis and r, u, and c:

$$\frac{1}{n_1}\log(\frac{F_{t,n_1}}{S_t}) = r + u - c \tag{3}$$

According to the normal backwardation theory, the basis is related to the spot return $E_t(S_{t+n})/S_t$ and the risk premium π:

$$\frac{F_{t,n_1}}{S_t} = \frac{E_t(S_{t+n_1})}{S_t} - \frac{\pi}{S_t} \tag{4}$$

Fama and French (1987) examine these two equations empirically.

The LHB factor is the return to the strategy of buying low basis commodities and selling high basis commodities. From Eq. (3), one can see that low basis commodities are the commodities with high convenience yield. Thus, the LHB factor has positive exposures to high convenience-yield commodities. Positive LHB factor returns indicate that high convenience-yield commodities have higher returns than low convenience-yield commodities. This could be due to the under-valuation of high convenience-yield commodities, or due to the higher required return of high convenience-yield commodities. The higher required return, in turn, may reflect greater quantity pressure or higher risk. Due to high convenience yield, there can be strong spot demand, and the demand for the futures can be weak, suppressing the futures price. Alternatively, investing in high convenience-yield commodities may be riskier in that high convenience-yield may indicate greater price uncertainty in the future. From Eq. (4), one can see that low basis commodities are either those commodities for which market participants expect low prices or those commodities for which market participants assign high risk premium. Positive LHB factor returns suggest a possibility that when market participants expect low prices, they are too pessimistic and there is a downward bias.

From the above considerations, we can formulate five alternative interpretations of the LHB factor returns, the first three related to Eq. (3) and the last two related to Eq. (4):

(i) The LHB factor returns arise from market participants' under-valuation of high convenience-yield commodities.

(ii) The LHB factor returns reflect the high required returns of high convenience-yield commodities, which is due to quantity pressure.

(iii) The LHB factor returns reflect the high required returns of high convenience-yield commodities, which is due to risk.

(iv) The LHB factor returns arise from the downward bias in market participants' expectation when they expect a price decline.

(v) The LHB factor return is a risk premium.

The above 5 interpretations are not mutually exclusive. For example, (iii) is a special case of (v). Note also that the Keynesian idea of compensating speculators is consistent both with (ii) and (iii). While our analyses are not focused on distinguishing among alternative interpretations, our findings are more supportive of the risk-based views, i.e. (iii) and (v).

When we apply Eqs. (1) and (2) to the currency markets, we can relate our study to the international finance literature. Suppose that the exchange rate is quoted as the USD value of a non-USD currency. Then, in Eq. (1), r is the USD interest rate, c is the non-USD rate, u is zero, and we get the covered interest parity (CIP) results. On the other hand, Eq. (2) represents the forward premium puzzle, where π represents the forward bias. See the paper by Fama (1984) for further discussion of the forward premium puzzle. When Eqs. (1) and (2) are combined, the resulting formula represents the failure of the uncovered interest parity (UIP):

$$S_t e^{(r-c)n} = E_t\left(S_{t+n_1}\right) - \pi \tag{5}$$

The carry trades exploit this failure of the UIP, which is ultimately attributable to the violation of the expectation hypothesis as stated in Eq. (2). Note that the source of the LHB factor is also the violation of the expectation hypothesis. This similarity has been noted by many authors including Szymanowska, de Roon, Nijman, and Goorbergh (2013). In the international finance literature, the currency carry return has been related to the various risks of high yield countries such as the risk of de-valuation, sudden stop, and sovereign default. See, for example, Brunnermeier, Nagel, and Pedersen (2008). It is possible that the LHB return is the reward for a similar type of risk, and that the LHB returns and the carry returns are correlated. We examine this possibility later in the paper.

The return to holding a futures contract from time t to time $\mathrm{t}+1$ can be expressed as $F_{t+1,n-1}/F_{t,n}$. Szymanowska, de Roon, Nijman, and Goorbergh (2013) decompose this return into the "spot premium" and the "term premium":

$$\frac{F_{t+1,n-1}}{F_{t,n}} = \frac{\mathrm{S}_{\mathrm{t+1}}}{F_{t,n_1}} \cdot \frac{F_{t+1,n-n_1}/S_{t+1}}{F_{t,n}/F_{t,n_1}}$$

The spot premium $\mathrm{S}_{\mathrm{t+1}}/F_{t,n_1}$ arises from the "tail" of the term structure, whereas the term premium $(F_{t+1,n-n_1}/S_{t+1})/(F_{t,n}/F_{t,n_1})$ arises from the "shift" of the term structure over time. Szymanowska, de Roon, Nijman, and Goorbergh (2013) report that the term premium is relatively small. In our analysis, we calculate the returns out of short maturity futures. Thus, it is very unlikely that the term premium is important to our analysis.

3. Data

Our analysis is based on commodity futures market data for the 30-year period between January 1981 and December 2010. We have collected futures price data from the Commodity Research Bureau database. We have selected 22 commodities for which complete price series are available for a 30-year period.[9] Further details on commodity selection are provided in Appendix 1.

Out of futures price data, we calculate returns of each commodity. We adopt the "nearest-maturity contract formulation": for each day, we identify the nearest-maturity contract that has at least 5 calendar days remaining until the last trade date.[10] From the nearest-maturity contracts, we create a single price index for each commodity. The price index reflects any changes of contracts, though it does not account for transaction cost. Monthly return is calculated from this price index.

We define basis as the second nearest-maturity contract price over the nearer-maturity contract price,[11] with the adjustment for the interval between two maturity dates. More specifically, basis is defined as $100\left[\left(F_{t,n_2}/F_{t,n_1}\right)^{1/(n_2-n_1)}-1\right]$, where F_{t,n_1} and F_{t,n_2} are the nearer-maturity and the second nearest-maturity contract prices, respectively.

Table 1 shows the univariate statistics on monthly returns and end-of-the-month basis for each of 22 commodities. One can see that average monthly returns are mostly positive, between 0% and 1% except for corn, wheat, and lumber. Average basis is mostly positive as well, suggesting that commodity markets are more likely to exhibit upward-sloping term structure.

---

[9] We carried out a preliminary analysis using the price data from Bloomberg, and obtained essentially the same results. It appears that our results are quite robust to changes in the time period and the list of commodities as well.

[10] That is, our "roll-over" date is 5 calendar dates prior to the expiration date. This corresponds to traders' rollover strategy. Rollover at the last possible moment faces greater uncertainty.

[11] Alternatively, we could use the spot price in the denominator. The CRB database includes spot prices, but the quality of these data seems questionable. Also, the coverage is limited. So we do not follow this alternative. Gorton, Hayashi, Rouwenhorst (2012) also calculate the basis as we do in this paper, except for the fact that their basis is the inverse of our basis. See footnote 3.

4. Basis Sorted Portfolios

The easiest way to motivate the LHB factor is to construct basis sorted portfolios and compare their average returns. If low basis portfolios have significantly higher returns than high basis portfolios, then the strategy of buying low basis portfolios and selling high basis portfolios is profitable, and the return to this strategy has the potential to be an effective pricing factor. This is exactly the case, as shown by Gorton and Rouwenhorst (2006), Fuertes, Miffre, and Rallis (2010), and Gorton, Hayashi, and Rouwenhorst (2012). Below we reaffirm this pattern from our data.

We construct basis sorted portfolios in the following way. At the beginning of each month, we rank all 22 commodities by basis, from the lowest to the highest. Then we create four portfolios of approximately equal sizes. We compute the average return of each portfolio and also of the long-short strategy of buying the low basis portfolio and selling the high basis portfolio. We call the return to the long-short strategy the LHB factor. We repeat the analysis separately for agricultural commodities and non-agricultural commodities, and also for each 10-year sub-period.

Table 2 reports average portfolio returns and standard deviations. When all 22 commodities are included in the sort, the lowest quartile portfolio (P1) has the average return of 0.9% whereas the highest quartile portfolio (P4) has the average return of 0.12%. The difference between these two, i.e. the return to the long-short portfolio, is 0.78%, which is significant at the conventional significance level.[12] The return to the other portfolios (P2 and P3) also fit into the pattern. As basis increases, the average return falls. The significance of the long-short portfolio is observed in about half of subsets and sub-periods. The effect is stronger in the 2001-2010 sub-period and among non-agricultural commodities.

5. The LHB Factor and the Cross-Section of Commodity Futures Returns

In the previous section, we have seen that the LHB factor has significant returns. We now incorporate the LHB factor into a factor model, and document the contribution of this factor in explaining the cross-section of commodity futures returns. We consider two variants of factor model: (i) the market model which has the market return as the only factor, and (ii) the “market + LHB model” which

---

[12] This number is comparable to the annualized average return of 10.04% that Gorton and Rouwenhorst (2006) report.

includes both the market return and the LHB return as factors. We highlight the contribution of the LHB factor by showing that the second model has more explanatory power than the first model. For each model, we apply two estimation procedures: (a) “one-step estimation” where the time-series of commodity futures returns are regressed on the time-series of factors, and “beta” is obtained for each commodity, (b) “two-pass CSR” where the cross-sectional regression is run of average returns on betas, and the “price of risk” is estimated. We first describe these two procedures, and then discuss the results from each procedure accordingly.

5.1. Estimation and Test Procedures

Our factor models can be expressed as the following system of equations:

$$\mathrm{r}_t = \alpha + \beta f_t + \varepsilon_t \tag{6}$$

r_t is a list of asset returns, and f_t is a list of factor returns. When there are N assets and k factors, r_t, α, ε_t are N-vectors, β is an N-by-k matrix, and $\mathrm{f_t}$ is a k-vector. $\mathrm{f_t}$ is not de-meaned. The main restriction of the models is that the intercept terms are all zero[13]:

$$\alpha = 0 \tag{7}$$

We can test this restriction based on F statistic after OLS estimation. This test has been discussed by many authors, including Gibbons, Ross, and Shanken (1989). See Appendix 2.1. for exact formula. Alternatively, we can carry out GMM estimation after imposing the restriction, and perform specification tests. See Appendix 2.2. for exact formula and the paper by Cochrane (1996) for further discussion.

The factor models have implications on the cross-sectional relationship between expected returns $\mathrm{E}(r_{i,t})$ and beta β_i. According to the arbitrage pricing theory (APT), there exists a k-vector θ such that

$$\mathrm{E}(r_{i,t}) = \beta_i{}'\theta \tag{8}$$

θ is called the factor price or the price of risk. Fama and MacBeth (1973) and Shanken (1992) estimate θ from the “two-pass cross-sectional regression (CSR).” In two-pass CSR, average returns are regressed on estimates of β_i. Shanken (1992) notes that the standard errors of the θ estimates need to be adjusted for the errors in average returns and estimates of β_i. In our estimation, all the standard errors are duly adjusted. We estimate a slightly modified version of Eq. (8):

[13] This restriction can be motivated from the arbitrage pricing theory.

$$\mathrm{E}(r_{i,t}) = V_{rf,i}\lambda \tag{9}$$

$V_{rf,i}$ is the covariance between returns and factors, and λ is the factor price. Needless to say, the estimations of Eqs. (8) and (9) are equivalent.[14] We estimate λ by OLS and also by GMM. For the purpose of specification tests and model comparison, we follow Kan, Robotti, and Shanken (2013) and Kan and Robotti (2009). Their methodologies allow formal model comparison under possible misspecifications. See Appendix 2.3. and Appendix 2.4. for exact formula.

Note that, in our setup, the two-pass CSR has a limited theoretical appeal. When the factor returns are calculated out of zero-investment long-short portfolios, the factor price equals the mean of the factor return, i.e., $\theta = \mathrm{E}(\mathrm{f_t})$. In this case, Eq. (9) is equivalent to Eq. (7). As there is a direct way of testing Eq. (7), there is no reason to test Eq. (9) once again. Nonetheless, we carry out two-pass CSR for two reasons. First, the estimates of θ (and λ) are interesting on their own. It is unlikely to be identical to the sample mean of factors, $\hat{E}(f_t)$. Second, recent development of econometric techniques for model comparison, such as Kan and Robotti (2009) and Kan, Robotti, and Shanken (2013), are more suitable for the two-pass CSR than for the one-step estimation.

5.2. One Step Estimation Results

Before estimating the factor model, an appropriate market return measure needs to be selected. Previous research has used various measures of the market return. Dusak (1973) and Bessembinder (1992) have used the stock market return; Miffre and Rallis (2007) have used the commodity market return as well as the stock market return; Carter, Rausser, and Schmitz (1983) have used the combined stock-and-commodity-market return. Similar to Carter, Rausser, and Schmitz (1983), we define the market return as the average of the stock market return and the commodity market return. This combined return has more explanatory power.[15] The stock market return is calculated from the MSCI World Index, and we subtract the 3-month US government yield from it.[16] The commodity market is

---

[14] Eq. (9) is more popular formulation for an SDF model than Eq. (8). See Cochrane (1996).

[15] We have examined alternative specifications of the market return, and we have obtained similar results. As our goal is to indicate the marginal contribution of a new factor, it is conservative for us to use the market factor with the largest explanatory power.

[16] While we subtract the risk-free rate from the stock market return, we do not do so for the commodity market return. The commodity index is from futures contracts, and buying futures contracts does not require investment. Therefore, the return to this index is already “excess return,” and there is no need to subtract the riskfree rate.

calculated as the equally weighted portfolio return out of all the commodity futures in our database.[17] The summary statistics of the market factor and its components are presented in Table 7 together with statistics on other factors.

The results of one-step estimation are shown in Table 3.[18] In Panel A, we report the results for the market model. Our primary interest lies in the significance of the constant estimates. The constant estimate is highly significant (at the 5% level) for live cattle, and is marginally significant (at the 10% level) for corn, feeder cattle, heating oil, and lumber. In total, there are 5 commodities for which the constant is significant. The market factor is significant for all commodities, and the beta varies from 0.45 to 1.38. R squared is mostly between 5% and 20%. In Panel B, we report the results for the market + LHB model. Now the constant estimate is significant for live cattle and lumber. The constant estimates of corn, feeder cattle, and heating oil are not significant anymore. The LHB factor is significant for about half of commodities, and R squared does not change dramatically.

Simply counting the number of significant constant terms, the market model has 5 whereas the market + LHB model has 2. Is this a meaningful reduction? Does this indicate the significant contribution of the LHB factor in explaining the cross-section of commodities futures? To answer this question, we carry out specification tests and model comparison as summarized in Table 4. We perform four specification tests. There are four such tests as we have two models and we apply two different testing procedures for each model. Thus, the first two tests are the F test and the J test on the market model, and the last two tests are the F test and the J test on the market + LHB model. The F test is performed after OLS estimation, whereas the J test is performed after GMM estimation. The null hypothesis is that all intercept terms are simultaneously zero, and a high p value supports this hypothesis. Clearly, the null hypothesis is not rejected. In the F test, the p value is 0.15 for the market model and 0.31 for the market + LHB model. The increase in the p value suggests that the LHB factor improves the performance of the model. In the J test, the p value is quite high for both models, at 0.90 and 0.88. Note, however, that these two numbers (0.90 and 0.88) are not directly comparable as the two tests are based on two different GMM weighting matrices.[19] The model comparison reported at the bottom of Table 4 corrects this problem, and compares the explanatory power of the two models while keeping

---

[17] We have also experimented with the S&P Goldman Sachs Commodity index return. As has been noted by Erb and Harvey (2006), this index assigns a too large weight on energy commodities and appears less suitable than the equal weighted index.

[18] We have GMM estimates as well as OLS estimates. To save space, we only report the OLS estimates.

[19] Cochrane (1996) emphasized this aspect when discussing the Newey-West (1987b) test.

the GMM weighting matrix constant. The test statistic is quite high, indicating that the difference between the two models is highly significant.

The model comparison result can be anticipated from the fact that the constant terms are mostly zero in the market + LHB model, and that the LHB factors are significant for many commodities. As the constant terms are mostly zero, the models (especially the “unrestricted” market + LHB model) are likely to be “correct.” As the LHB factor is significant, removing this factor would be very “restrictive.”

One limitation of this model comparison is that the validity of the analysis critically depends on the correctness of the model. While the specification test suggests that the model is plausible, it is unlikely that any model is exactly correct. The analysis in the next subsection based on the two pass CSR does not suffer from this limitation.

5.3. Two Pass CSR Results

Factor price estimates from two pass CSR are reported in Table 5. Our primary interest here is whether the price of the LHB factor is significantly positive. When the estimation method is OLS, the LHB factor price is positive but not significant. When GMM is the estimation method, the LHB factor price is positive and significant. GMM uses an efficient weighting matrix, which makes it possible to obtain more precise estimates. The same is true for the market factor price as well: The GMM estimates are significant whereas the OLS estimates are not.

When the LHB factor price is significant, one may conclude that the two models are significantly different, i.e. the LHB factor makes a significant improvement. A more formal comparison can be based on the R squared and the Hansen-Jaganathan (HJ) distance. The HJ distance measures the weighted average of the pricing errors. As shown in Table 5, the R squared is 0.47 vs. 0.58, and the HJ distance is 0.09 vs. 0.07. We test the null hypothesis that the R squared is the same for the two models, and also the null hypothesis that the HJ distance is the same for the two models. As summarized in Table 6, the hypothesis of equal R squared cannot be rejected at the conventional significance level, whereas the hypothesis of equal HJ distances is rejected. Considering the fact that the GMM estimates are more efficient (and also considering the fact that not rejecting a hypothesis does not show the correctness), we take the GMM result. So we conclude that the market + LHB model is significantly superior to the market model and that the LHB factor is making a significant improvement for the factor model.

6. The LHB Factor vs. Other Factors of Stock, FX, and Commodity Markets

We have observed the significance of the LHB factor in explaining the cross-section of commodity futures returns. We now ask whether the LHB factor is really new, as compared to the other known factors of the stock, FX, and commodity markets. To put it another way, we ask whether the LHB factor expands the investment opportunity set. We carry out two analyses to answer this question. First, we examine the contemporaneous and dynamic correlations between the LHB factor and other factors, to determine whether the LHB factor can be "explained away" by other factors. Our finding is that it is not the case. Second, we analyze the investment possibility sets that are obtained from the factors, with and without the LHB factor. We find that an investor who creates a portfolio out of the factors benefits substantially by including the LHB factor in the investment universe.

6.1. Contemporaneous and Dynamic Correlations

In addition to the LHB factor, we consider three stock market factors ("stock", SMB, HML), one FX market factor ("carry"), and one commodity market factor ("commodity"). "Stock" is the stock market factor that is calculated as the MSCI World Index returns over the short-term US treasury bill returns. SMB and HML are the small-minus-big and the high-minus-low factors of Fama and French (1993); we have obtained these data from the web site of Professor Kenneth French.[20] These three are the most popular stock market factors that are used in the asset pricing literature since Fama and French (1993). "Carry" is the return to the currency carry strategy. This factor is calculated from Deutsch Bank G10 Currency Harvest Index.[21] For the "commodity" factor, we use the equal-weighted average return of the 22 commodities in our database.[22]

---

[20] The Fama-French factors can be downloaded from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french. Since our "stock" factor is a global factor, we could have used the international version of SMB and HML for the sake of consistency. The international version of HML is often used in the literature; however, the international version of SMB is not used much due to the practical problem of applying the large-vs-small distinction across countries. So we have decided against using the global HML and SMB, and have used US market-based HML and SMB.

[21] Deutsche Bank currency indexes are downloadable from index.db.com.

[22] As mentioned earlier, the S&P Goldman Sachs index is not suitable as it has a very high weight on energy

Recall the similarity between the LHB return and the carry return. In a currency carry trade, one buys high yield currencies and sells low yield currencies. Under the covered interest parity, high yield currency futures have low basis, and low yield currency futures have high basis. (For a low yield currency, the futures price is higher than the spot price to compensate for a low interest rate. For a high yield currency, the opposite is true.) Thus, buying high yield currencies and selling low yield currencies in the futures market amount to buying low basis futures and selling high basis futures—the strategy underlying the LHB factor. Given this similarity in terms of strategy, it is possible that the returns of these two factors are also similar.

The LHB factor can be also compared to the HML factor. The strategy underlying the HML factor is long in high book-to-market stocks and short in low book-to-market stocks. High book-to-market stocks tend to have high yield (high dividend or earnings relative to price), whereas low book-to-market stocks tend to have low yield. Thus, we may say that the HML factor is based on the yield difference. Note that the LHB factor is also based on the yield difference, this time the yield being the "convenience yield." The HML factor is often interpreted as the payoff to being exposed to financial distress. It is possible that the LHB factor is a payoff to similar exposure, and that the LHB factor returns co-move with the HML returns.

Table 7 reports the correlations among the factors together with univariate statistics. Note that the carry returns are available only from October 2000, so our sample period is between October 2000 and December 2010. We see some large correlations among the "stock" factor, the carry factor, and the "commodity" factor.[23] However, the LHB factor has very little correlations with other factors.

To see the dynamic correlation structure, we estimate the vector auto-regression for the 6 factors. We include up to 3-month lagged variables in the right hand side. Including more lagged variables does not affect the estimation. The coefficient estimates are reported in Table 8. We see some positive cross-serial correlations among stock market factors. For example, SMB positively influences the HML one month later. We also see positive cross-serial correlations among the two commodity market factors: "Commodity" has positive influences on LHB one month later. There are two rather unexpected patterns: (i) Carry has positive influences on "stock" with two-month lags, (ii) carry has

---

sector.

[23] Gorton and Rouwenhorst (2006) have noted that commodity futures are negatively correlated with US stocks. We find, however, that the correlation with MSCI World Index is positive rather than negative.

negative influence on LHB with two- and three-month lags. We do not have good explanations for these patterns. One thing that we can say is that the LHB return appears to be quite different from the carry. This is apparent from a simple time-series plot: the LHB factor exhibits significantly positive returns in the crisis periods, which is not the case with the carry returns. We comment further on this pattern in the next section.

Overall, while there are some significant patterns, the other factors cannot replace the LHB factor entirely. Below, we provide economic meaning to this conclusion by taking an asset allocation perspective and showing the weight of the LHB factor in an optimal allocation.

6.2. A (Strategic) Asset Allocation Perspective

Consider an investor who creates a portfolio out of the 6 factors of the stock, FX, and commodity markets. The investor considers these 6 factors because he believes that they represent the most important common risk factors of the financial markets and that adding other assets would only increase the idiosyncratic risk. If the investor is a mean-variance investor, he may draw an efficient frontier on the mean-standard deviation space and choose one point on the frontier corresponding to his risk aversion. Given the existence of the riskfree asset, we may draw the efficient frontier on the excess mean return-standard deviation space. Then the efficient frontier is a straight line passing through the origin which is tangent to the parabolic curve representing the portfolios of risky assets. In Figure 1, we show such efficient frontiers. We draw two efficient frontiers, one with the LHB factor and the other without the LHB factor. We draw the parabolic curves of risky assets as well.

The tangent portfolios (corresponding to the tangent points where the linear efficient frontiers meet the parabolic curves of risky assets) are shown in Table 9. Our main interest is whether a large weight is assigned to the LHB factor. If so, it indicates the importance of the LHB factor for the asset allocation. It turns out that the weight of the LHB factor in the tangent portfolio is 16%. While this weight is smaller than the weights assigned to three stock factors and the commodity market factor, it is greater than the weight assigned to the carry factor. When the LHB factor is included, the optimal portfolio has the monthly mean excess return of 0.78% with the standard deviation of 1.89%, resulting in the annualized Sharpe ratio of 1.44. This Sharpe ratio compares to 1.31 that is obtained without the LHB factor. Thus, for the same standard deviation, excluding the LHB factor may eliminate up to one-tenth of the average return.

We have carried out the same analysis for a longer sample period. Recall that, for this analysis, our sample starts from October 2000 because the carry returns are available only from that month. Excluding the carry returns from the analysis, we can extend the sample period back to January 1981. From that analysis, we see a much larger difference between the two Sharpe ratios, supporting the view that the LHB factor cannot be replaced with the other factors.

7. Predictability of the LHB Factor

In this section, we explore the time-variation in, and the predictability of, the LHB factor. We first discuss the relationship between the LHB factor and the US business cycle. The theories of convenience yield and normal backwardation suggest certain relationships between the LHB factor the business cycle. In our analysis, we do not find any strong relationship between the two. We then run predictive regressions where predictors include industrial production as well as the spread in the basis, in the hedging pressure, and in the implied volatility. We find that the implied volatility spread has predictive power, whereas the other two variables do not. Finally, we examine the returns to market timing strategies based on the value of the predictor variables. Once again, the implied volatility turns out to be the most powerful variable. The other predictors are shown to be somewhat useful as well. Considering that these spread variables are unlikely to affect the behavioral bias, this finding is consistent with the view that the LHB factor return is a risk premium.

7.1. The LHB Factor over US Business Cycle

The theory of convenience yield suggests a link between the LHB factor and business cycle. As we have seen in Section 2, an important component of basis is convenience yield. Convenience yield is affected by the inventory level through the option value of the inventory and the possibility of stock-out. Inventory level tends to be counter-cyclical; thus convenience yield and basis are possibly pro-cyclical. If the overall basis level affects the LHB factor returns, the LHB factor returns may fluctuate over business cycle. The theory of Keynesian normal backwardation also provides a possible link between the LHB factor and the business cycle. From this perspective, the LHB return is compensation to speculators for providing insurance to hedgers. The value of this insurance is higher when the recession is imminent. This could be due to quantity pressure or due to higher risk. Either way, the LHB return can be counter-cyclical.

Figure 2 is the plot of the LHB return index over the US business cycle. The contraction period (as determined by National Bureau of Economic Research) is indicated by shaded area. In three out of four contraction periods from the past 30 years, the LHB return index increased noticeably, suggesting a possible counter-cyclicality.[24] The counter-cyclicality, however, is either non-existent or very weak when we examine the LHB factor against the US industrial production index. The correlation between the LHB factor and the rate of change in the US industrial production is almost zero as shown in Table 10. In the regression of the LHB factor returns on the rate of change in the US industrial production, the coefficient estimate is negative but not significant. Overall, the LHB return is certainly not cyclical; there is some indication, but no statistical evidence, that the LHB return might be counter-cyclical.

7.2. Predictive Regressions

The idea that the cross-sectional spread variables can predict the factor return comes from the literature on the stock market. From the decomposition of the price-to-earnings ratio, Asness, Friedman, Krail, and Liew (2000) have suggested that the cross-sectional spreads in price-to-earnings ratio and earnings growth help to predict the value premium. Cohen, Polk, and Vuolteenaho (2003) confirmed the pattern in the context of the book-to-market ratio.

Applying the idea to our case, it is natural to consider the basis spread as a predictor of the LHB factor return. The basis is the variable by which the commodities are ranked during the construction of the LHB factor. Moreover, the basis is similar to the price-to-earnings ratio and the book-to-market ratio in that it is indicative of the intrinsic values of the asset. We consider two more spread variables: the spread in the hedging pressure and in the implied volatility. The importance of the hedging pressure in explaining commodity futures return is well-known. Cootner (1960) has interpreted normal backwardation idea as being driven by the hedging pressure. Carter, Rausser, and Schmitz (1983) and Bessembinder (1992) have shown empirically the relevance of the hedging pressure. The relevance of volatility has been discussed by many authors as well. Bessembinder (1992) has also shown the relevance of volatility. He has included volatility in his analysis in the belief that the "residual risk," not captured by CAPM-beta, may affect the futures return. The importance of volatility is also apparent in the theory of convenience yield. The convenience yield is a function of the possibility of stock-outs and price hikes. Litzenberger and Rabinowitz (1995) clarify the real option aspects of the

[24] Figure 2 also plots the equal-weighted return index of commodity futures. Gorton and Rouwenhorst (2006) have noted that commodity futures are pro-cyclical, which is apparent in Figure 2.

convenience yield, and consequently, the relevance of the volatility.

Each spread variable is calculated as the difference between the 12.5th and 87.5th percentiles. These percentiles are the median values of the top and the bottom quartile portfolios—the underlying portfolios of the LHB factor. The hedging pressure is calculated as hedgers' net long position relative to all open interest. All the information is from the Commitment of Traders reports.[25] The Commitment of Traders reports are released weekly, with more than one week's delay. The end-of-month hedging pressure is based on the last weekly report of the month, without considering the release gap. Volatility is implied volatility from the futures options. We have obtained implied volatilities of each commodity from the Commodity Research Bureau database. The volatility spread is available from February 1989, therefore the subsequent analysis is based on the period between February 1989 and December 2010.

Table 10 reports univariate statistics and correlations for the three spread variables as well as the LHB factor and the rate of change in the US industrial production. There is some negative correlation between the hedging spread and the basis spread, but it is not large. The results of predictive regressions are reported in Table 11. The first four columns of the table show the estimates when each predictor variable is included one at a time. The last column shows the regression where all three spread variables are included at the same time. The pattern is the same in both cases. The volatility spread is a significant predictor, whereas the other predictors are not. R squared is not very high, so the predictability may not be economically significant. Below we examine the economic significance of the predictability by implementing various market timing strategies.

7.3. A Market Timing Perspective

We consider the following market timing strategies (a.k.a. tactical asset allocations)[26]. At the end of each month t, we calculate the average value of predictor variable x from the past m months, i.e. $(x_{t-m+1} + \cdots + x_t)/m$. If the current value of the predictor variable is higher than a fraction of this average, i.e. $x_t \geq k(x_{t-m+1} + \cdots + x_t)/m$ for some k, we invest in the strategy underlying the LHB factor. Otherwise, we stay with the risk-free asset. Let us denote this strategy by (m, k).

[25] Following the convention, we identify 'commercial traders' as hedgers. By ignoring the positions of 'non-reportable' we are implicitly assuming that small traders as speculators.

[26] Erb and Harvey (2006) have also considered a market timing strategy based on basis. In their strategy, the investor buys the commodity market index rather than the LHB factor.

Table 12 reports average returns to various market timing strategies. We consider four predictor variables--industrial production (IP) as well as three spread variables. We allow m to be 6 months, 9 months, or 12 months. And we let k to be one of 0, 0.25, 0.5, 0.75, and 1. Note that when $k = 0$, there is no market timing, and the strategy returns are identical to the LHB returns. As it turns out, all three spread variables produce meaningful improvements, but the volatility spread is the most effective predictor in the market timing strategy. In the best case (m=6, k=1), the average excess return is 0.82% and the standard deviation is 3.54%, resulting in the annualized Share ratio of $\sqrt{12} \cdot 0.82/3.54 = 0.8$. This is about 1.5 times the Sharpe ratio of no timing (m=6, k=0).

The effectiveness of the implied volatility spread supports the view that the LHB factor is a risk premium. It is likely that risk premium moves together with the implied volatility spread. It is rather unlikely that intensity of behavior bias moves together with the volatility spread. Hedging pressure spread has some predictive power, so we cannot exclude the possibility that the LHB factor is partially driven by quantity pressure.

8. Conclusion

This study has confirmed the significance of the LHB factor in explaining the cross-section of commodity futures returns. We have shown that the LHB factor does not have strong correlations with the other factors of the stock, FX, and commodity markets, and that the investment possibility set expands meaningfully when we include the LHB factor to the investment universe. The LHB factor is mildly counter-cyclical, and can be predicted by spread variables, especially the volatility spread. These findings support the view that the LHB factor is a risk premium. Our analysis indicates a possibility of devising profitable market-timing strategies from these predictors.

Appendix

A.1. Commodity Futures Data

Our commodity futures price data are from the Commodity Research Bureau database. These data have been used by many, including Gorton and Rouwenhorst (2006), Gorton, Hayashi, Rouwenhorst (2012), and Shen, Szakmary, and Sharma (2007). These authors have based their analysis on unbalanced panels; thus, they have included a larger number of commodities. Our analysis is based on a balanced panel; so the number of commodities included in the analysis is smaller. We have excluded all the commodities whose price series start after January 1981. We have also excluded the metal commodities that are traded on the London Metal Exchange. The CRB data on these commodities are limited.[27] Moreover, the Commitment of Traders data are not available for these commodities, making these commodities less attractive for our purpose. After the selection, we have 22 commodities, as listed in Table 1.

For the return calculation, we select the nearest-maturity contract for each day. However, we exclude those contracts with very short history. The following contracts have been excluded even if there are price data in the CRB database: G, M, and X for coffee, G, J, M, Q, V, and X for copper and silver, F and X for corn, Q and U for cotton, F, H, K, N, U, and X for gold, lean hogs, and live cattle, Z for lumber, F, G, J, K, N, Q, V, and X for palladium, G, H, K, M, Q, U, X, and Z for platinum, F, J, and M for pork bellies, X for soybean meal and soybean oil, and F and U for sugar. (F, G, H, J, K, M, N, Q, U, V, X, and Z indicate the expiration month of each contract, from January to December.) For the implied volatility, we select only those options contracts for which the corresponding futures contracts are used in the return calculation. There are three exceptions to this rule: Q and V of soybean oil are excluded, and F of copper and silver are also excluded, all of them due to short history. Implied volatility for palladium is excluded again due to short history.

A.2. Estimation, Specification Tests, and Comparison of Factor Models

In this appendix, we present the formula for estimations, specification tests, and model comparison. There are four subsections: we discuss OLS and GMM estimation of “one step models” followed by OLS and GMM estimation of “two pass CSR.” Our discussion of two pass CSR is mostly based on the papers by Kan and Robotti (2009) and by Kan, Robotti, and Shanken (2013).

A.2.1. One Step Estimation – OLS

Model Consider the following system of equations:

$$\mathrm{r}_t = \alpha + \beta f_t + \varepsilon_t = B x_t + \varepsilon_t, \qquad \varepsilon_{\mathrm{t}} \sim N(0, \Sigma) \tag{10}$$

where r_t is an N-vector of *excess* returns and x_t is a (1+k)-vector whose first element is 1 and the

[27] There are no contract-by-contract data for these commodities.

remaining k elements are factor returns. Factor returns are not de-meaned. ε_t is an N-vector, B is a N-by-(1+k) matrix, Σ is an N-by-N matrix, and t runs from 1 to T.

Parameter Estimate By stacking the row vectors made from r_t, x_t, ε_t, we define $\mathrm{R}, \mathrm{X}, \text{and } \mathrm{E}$ so that

$$\mathrm{R} = XB' + E$$

Then the OLS estimate can be written as:

$$\hat{\mathrm{B}}' = (X'X)^{-1}X'R = \mathrm{B}' + (X'X)^{-1}X'E \tag{11}$$

Standard Error By vectorizing $\hat{\mathrm{B}}'$ (taking the first column, and then the second column, and so on), we may express the distribution of $\hat{\mathrm{B}}'$ as following:

$$\mathrm{vec}\left(\hat{\mathrm{B}}'\right) \sim \mathrm{N}[vec(B'), \Sigma \otimes (\mathrm{X}'\mathrm{X})^{-1}] \tag{12}$$

Specification Test The first column of X includes only 1. So we may partition X and B as follows:

$$\mathrm{X} = [1\ F]$$

$$\mathrm{B} = [\alpha\ \beta]$$

Then the distribution of $\hat{\alpha}$ is:

$$\hat{\alpha} \sim \mathrm{N}\{\alpha, \Sigma[(\mathrm{X}'\mathrm{X})^{-1}]_{11}\}$$

The estimator of Σ, $\hat{\Sigma} = E'E/T$, has the following Wishart distribution:

$$T\hat{\Sigma} \sim W(\Sigma, T-k-1)$$

Note that k does not include the constant term. From the distribution of $\hat{\alpha}$, $\hat{\alpha}\Sigma^{-1}\hat{\alpha}/[(\mathrm{X}'\mathrm{X})^{-1}]_{11}$ has $\chi^2(N)$ when $\alpha = 0$. Also, $\hat{\alpha}\Sigma^{-1}\hat{\alpha}/(\hat{\alpha}\hat{\Sigma}^{-1}\hat{\alpha})$ has $\chi^2(T-k-1-N+1)$. Note that N and 1 come from $\hat{\alpha}$. See Proposition 8.9 of Eaton (2007). From these two distributions, we get

$$\mathrm{F} \equiv \frac{\hat{\alpha}\left(T\hat{\Sigma}\right)^{-1}\hat{\alpha}}{[(\mathrm{X}'\mathrm{X})^{-1}]_{11}} \frac{T-k-1-N+1}{N} \sim F(N, T-k-1-N+1) \tag{13}$$

A.2.2. One Step Estimation – GMM

One motivation for adopting GMM is that the OLS model outlined above is too restrictive, in that ε_t is assumed to have the iid property. GMM only exploits the moment condition, so no need for the iid assumption. That is, adjusting for heteroskedasticity and autocorrelation is easier.

Model The GMM exploits the following moment conditions:

$$E(\varepsilon_t) = 0$$
$$E(\varepsilon_t f_t') = 0$$

Or more explicitly, after imposing the restriction that $\alpha = 0$,

$$\begin{aligned} E(r_t - \beta f_t) &= 0 \\ E[(r_t - \beta f_t) \otimes f_t] &= 0 \end{aligned} \tag{14}$$

There are N(K+1) moment conditions and NK parameters (β), so one may think that there are N overidentifying restrictions. (If the factors are, however, investment returns and linear combinations of r_t, then k of these conditions are redundant, and there are only N-k overidentifying restrictions.)

Parameter Estimate Let us denote the moment conditions as $\mathrm{E}[g_t(\beta)]$ where

$$g_t(\beta) = \begin{bmatrix} r_t \\ r_t \otimes f_t \end{bmatrix} - \begin{bmatrix} I \otimes \mathrm{f_t}' \\ I \otimes \mathrm{f_t f_t}' \end{bmatrix} \mathrm{vec}(\beta') = \mathrm{C_t} - \mathrm{D_t vec}(\beta')$$

Given a 2N-by-2N weighting matrix W, the GMM estimator of β, $\hat{\beta}_{GMM}$ is the solution to $\min g_T(\beta)' W g_T(\beta)$, where $g_T(\beta)$ is the sample counterpart of $\mathrm{E}[g_t(\beta)]$, i.e.

$$g_T(\beta) = \frac{1}{\mathrm{T}} \sum C_t - \frac{1}{\mathrm{T}} \sum D_t \,\mathrm{vec}(\beta') = C - D\mathrm{vec}(\beta')$$

The solution to the minimization problem is

$$\hat{\beta}_{GMM} = (\mathrm{D'WD})^{-1} D'WC \tag{15}$$

Standard Error The variance-covariance matrix of $\hat{\beta}_{GMM}$ is given as

$$\mathrm{var}(\hat{\beta}_{GMM}) = (\mathrm{D'WD})^{-1} D'W \left[\mathrm{var}\left(\frac{1}{\mathrm{T}} \sum C_t \right) \right] WD(D'WD)^{-1}$$

or in terms of the asymptotic variance,

$$\mathrm{var}(\sqrt{T} \hat{\beta}_{GMM}) = (\mathrm{D'WD})^{-1} D'W \left[\mathrm{var}\left(\frac{1}{\sqrt{T}} \sum C_t \right) \right] WD(D'WD)^{-1}$$

We choose W so that it is a consistent estimator of $\mathrm{S}^{-1} = \mathrm{var}\left(\sum C_t / \sqrt{T}\right)^{-1}$ so that the variance-covariance matrix of $\hat{\beta}_{GMM}$ can be estimated by

$$\mathrm{var}(\hat{\beta}_{GMM}) = (\mathrm{D'WD})^{-1} \tag{16}$$

Following Newey and West (1987a), we use

$$W^{-1} = \hat{S} = \hat{\Omega}_0 + \sum_{j=1}^{m} \left(1 - \frac{j}{m+1}\right) (\hat{\Omega}_j + \hat{\Omega}_j')$$

where $\hat{\Omega}_{\mathrm{j}}$ is the sample autocovariance matrix of order j calculated from $C_1, \cdots, C_T$, and we set $m = 12$. (S is often called the covariance matrix of the sample pricing errors $g_T(\theta)$, though S/T, not S, is in fact the covariance matrix of $g_T(\theta)$.)

Specification Test The J test of the overidentifying restrictions is based on the following

$$\mathrm{J} = \mathrm{T}\mathrm{g}_{\mathrm{T}}\left(\hat{\beta}_{GMM}\right)' W \mathrm{g}_{\mathrm{T}}\left(\hat{\beta}_{GMM}\right) \sim \chi^2(N) \qquad (17)$$

N is the number of over-identifying restrictions.

Model Comparison Given two models A and B where model A nests model B, how do we test the null hypothesis that model B is as good as model A? A test, proposed by Newey and West (1987b) and discussed by Cochrane (1996), is based on the following statistic:

$$\mathrm{T}\mathrm{g}_{\mathrm{T}}\left(\hat{\beta}_B\right)' W \mathrm{g}_{\mathrm{T}}\left(\hat{\beta}_B\right) - T\mathrm{g}_{\mathrm{T}}\left(\tilde{\beta}_A\right)' W \mathrm{g}_{\mathrm{T}}\left(\tilde{\beta}_A\right) \sim \chi^2(1) \qquad (18)$$

The degree of freedom 1 corresponds to the number of restricted parameters. Note that the restricted estimator $\tilde{\beta}_B$ is obtained using the weighting matrix W of the unrestricted estimation. Also, the list of instruments is kept constant across models. That is,

$$\hat{\beta}_A = \left(\mathrm{D}_A'\hat{\mathrm{S}}_A^{-1}\mathrm{D}_A\right)^{-1}\mathrm{D}_A'\hat{\mathrm{S}}_A^{-1}\mathrm{C}_A$$

$$\tilde{\beta}_B = \left(\tilde{\mathrm{D}}_B'\hat{\mathrm{S}}_A^{-1}\tilde{\mathrm{D}}_B\right)^{-1}\tilde{\mathrm{D}}_B'\hat{\mathrm{S}}_A^{-1}\tilde{\mathrm{C}}_B$$

where $\tilde{\mathrm{D}}_B$ and $\tilde{\mathrm{C}}_B$ are obtained from the following moment conditions:

$$\tilde{g}_t(\beta) = \begin{bmatrix} r_t \\ r_t \otimes f_t^A \end{bmatrix} - \begin{bmatrix} I \otimes \mathrm{f}_t^{A\prime} \\ I \otimes \mathrm{f}_t^B \mathrm{f}_t^{A\prime} \end{bmatrix} \mathrm{vec}(\beta') = \tilde{\mathrm{C}}_t - \tilde{\mathrm{D}}_t \mathrm{vec}(\beta')$$

Note that the last test assumes that model A is correct. (Under the null, both models are correct.) If model A is not correct, the test is meaningless. Later, we discuss the tests based on Kan and Robotti (2009) and Kan, Robotti, and Shanken (2013) that deal with the situations where models are allowed to be incorrect!

A.2.3. Two Pass CSR – OLS

Model In the two pass CSR, average return is related to individual asset beta. Kan, Robotti, and Shanken (2013) suggest using the covariance between returns and factors instead of beta. Let us denote the covariances by $V_{rf} = \left(V_{rf,1}, \cdots, V_{rf,N}\right)'$. Then $V_{rf} = R'MF/T$ where $M = (I - 1/T 1_T 1_T')$. Then the second stage can be written as

$$\bar{\mathrm{r}}_i = V_{rf,i}\lambda + \eta_i, \qquad \mathrm{i} = 1, \cdots, \mathrm{N} \qquad (19)$$

Alternatively, one could use "simple beta," i.e. $\tilde{\beta} = \left(\tilde{\beta}_1, \cdots, \tilde{\beta}_N\right)' = V_{rf}\left[\mathrm{diag}\left(V_{ff}\right)\right]^{-1}$, so that the second stage can be expressed as

$$\bar{\mathrm{r}}_i = \tilde{\beta}_i'\theta + \eta_i, \qquad \mathrm{i} = 1, \cdots, \mathrm{N}$$

Obviously, $\lambda = \left[\mathrm{diag}(V_{ff})\right]^{-1}\theta$. Following Kan, Robotti, and Shanken (2013), we do not include a constant term. (A constant term is allowed only in the gross return formulation.)

Parameter Estimate We estimate λ by the standard OLS, i.e.

$$\hat{\lambda} = \left(V_{rf}{}'V_{rf}\right)^{-1}V_{rf}{}'\bar{r} \tag{20}$$

where $\bar{\mathrm{r}} = R'1_T/T$.

Standard Error We calculate the standard error following Proposition D.3 of Kan, Robotti, and Shanken (2013). The asymptotic variance of $\hat{\lambda}$ (i.e. the limit of $\mathrm{var}(\sqrt{T}\hat{\lambda})$) can be estimated from the time-series of

$$\mathrm{h_t} = \left(\hat{\lambda}_t - \hat{\lambda}\right) - AG_t\hat{\lambda} + Hz_t$$

where

$$\hat{\lambda}_t = \left(V_{rf}'V_{rf}\right)^{-1}V_{rf}{}'r_{\mathrm{t}}$$

$$A = \left(V_{rf}'V_{rf}\right)^{-1}V_{rf}{}'$$

$$G_t = (r_t - \bar{r})\left(f_t - \bar{f}\right) - V_{rf}$$

$$H = \left(V_{rf}'V_{rf}\right)^{-1}$$

$$\begin{aligned} z_t &= \left(f_t - \bar{f}\right)(r_t - \bar{r})'e \\ &= \left(f_t - \bar{f}\right)(r_t - \bar{r})'\left(\bar{r} - V_{rf}\hat{\lambda}\right) \end{aligned}$$

A heteroskedasticity-autocorrelation-consistent-covariance estimator is what we need. Following Newey and West (1987a), we use

$$S = \hat{\Omega}_0 + \sum_{j=1}^{m}\left(1 - \frac{j}{m+1}\right)\left(\hat{\Omega}_j + \hat{\Omega}_j{}'\right)$$

where $\hat{\Omega}_{\mathrm{j}}$ is the sample autocovariance matrix of order j calculated from $h_1, \cdots, h_T$, and we set $m = 12$. Thus, the estimator for the variance matrix of $\hat{\lambda}$ is

$$\mathrm{var}\left(\hat{\lambda}\right) = \frac{1}{T}S \tag{21}$$

Model Comparison To test the null hypothesis that the two lists have the same explanatory power, we follow Proposition D.5 of Kan, Robotti, and Shanken (2013). Let $\hat{\theta}_A$ and $\hat{\theta}_B$ be the factor price estimates of two models A and B. We consider the case where model A nests model B. Let K_1 be the number of factors included both in A and in B ($\mathrm{K}_1 = 1$ in our case), and let K_2 be the number of factors unique in A ($\mathrm{K}_2 = 1$ in our case). In this case, the R squared of model A is greater than the R squared of model B. The R squared is defined as

$$\rho_A^2 = 1 - \frac{Q_A}{Q_0} = 1 - \frac{\left(\bar{r} - V_{rf_A}\hat{\lambda}_A\right)'\left(\bar{r} - V_{rf_A}\hat{\lambda}_A\right)}{\bar{r}'\bar{r}}$$

$$\rho_B^2 = 1 - \frac{Q_B}{Q_0} = 1 - \frac{\left(\bar{r} - V_{rf_B}\hat{\lambda}_B\right)'\left(\bar{r} - V_{rf_B}\hat{\lambda}_B\right)}{\bar{r}'\bar{r}}$$

The test statistic is the difference between the two R squared, i.e. $\rho_A^2 - \rho_B^2$. This statistic has the same asymptotic distribution as the following variable:

$$\rho_A^2 - \rho_B^2 \sim \frac{1}{\mathrm{T}}\sum_{j=1}^{K_2}\frac{\xi_j}{Q_0}\chi_j^2$$

where χ_j^2's are independent random variables having the Chi-square distribution of order 1, and ξ_j's are eigenvalues of $\mathrm{H}_{22}^{-1}S_{22}$, where H_{22} and S_{22} are the lower-right blocks of $H_A = \left(V_{rf_A}{}'V_{rf_A}\right)^{-1}$ and S_A. When $\mathrm{K}_1 = \mathrm{K}_2 = 1$,

$$\rho_A^2 - \rho_B^2 \sim \frac{1}{\mathrm{T}}\sum_{j=1}^{K_2}\frac{\xi_j}{Q_0}\chi_j^2 = \frac{1}{T}\frac{1}{\left[\left(V_{rf_A}{}'V_{rf_A}\right)^{-1}\right]_{22}}\frac{S_{22}}{Q_0}\chi^2$$

Thus,

$$T\left[\left(V_{rf_A}{}'V_{rf_A}\right)^{-1}\right]_{22}\frac{Q_0}{S_{22}}(\rho_A^2 - \rho_B^2) \sim N(0,1)^2 \qquad (22)$$

Note that this test allows both models to be misspecified.

A.2.4. Two Pass CSR – GMM

Model A justification of the GMM this time is that it is implied by an SDF model, as in Cochrane (1996). As an SDF model, the moment conditions can be expressed as:

$$E\left\{r_t\left[1 - \left(f_t - \mu_f\right)'\lambda\right]\right\} = 0 \qquad (23)$$

There are N equations, and k parameters (ignoring the fact that μ_f needs to be estimated as well). Thus, there are N-k over-identifying restrictions.

Note that Eq. (14) is different from the “one-step models” of Eqs. (1) and (5). In the one-step model, we utilize the fact that the mean factor returns are the factor price. Eq. (14) is equivalent to Eq. (10) in that we estimate the factor price without reference to the mean factor returns.

Parameter Estimate Let $V_{Rf} = R'MF/T$ and $V_{RR} = R'MR/T$ where $M = (I - 1/T1_T1_T')$. Also, let $\bar{\mathrm{r}} = R'1_T/T$. Then the GMM estimator of λ is

$$\hat{\lambda} = \left(V_{Rf}'V_{RR}^{-1}V_{Rf}\right)^{-1}V_{Rf}'V_{RR}^{-1}\bar{r} \tag{24}$$

The only difference from Eq. (11) is the weight matrix V_{RR}^{-1}.

Standard Error We follow Kan and Robotti (2009). See page 1 of the online appendix. The asymptotic variance of $\hat{\lambda}$ depends on following variable:

$$\mathrm{h_t} = HV_{Rf}'V_{RR}^{-1}(r_t-\bar{r})\left(f_t-\bar{f}\right)'\hat{\gamma} + H\left[\left(f_t-\bar{f}\right) - V_{Rf}'V_{RR}^{-1}(r_t-\bar{r})\right]u_t + \hat{\lambda}$$

where

$$H = \left(V_{Rf}'V_{RR}^{-1}V_{Rf}\right)^{-1}$$

$$u_t = e'V_{RR}^{-1}(r_t-\bar{r}) = \left(\bar{r} - V_{Rf}\hat{\lambda}\right)'V_{RR}^{-1}(r_t-\bar{r})$$

$$\bar{\mathrm{r}} = \frac{1}{T}F'1_T$$

A heteroskedasticity-autocorrelation-consistent-covariance estimator is what we need. Following Newey and West (1987a), we use

$$S = \hat{\Omega}_0 + \sum_{j=1}^{m}\left(1-\frac{j}{m+1}\right)\left(\hat{\Omega}_j + \hat{\Omega}_j{}'\right)$$

where $\widehat{\Omega}_{\mathrm{j}}$ is the sample autocovariance matrix of order j calculated from $h_1,\cdots,h_T$, and we set $m = 12$. Thus, the estimator for the variance matrix of $\hat{\lambda}$ is

$$\mathrm{var}\left(\hat{\lambda}\right) = \frac{1}{T}S \tag{25}$$

Model Comparison We follow Kan and Robotti (2009). See Proposition 1 of the online appendix. Let $\hat{\lambda}_A$ and λ_B be the factor price estimates of two models A and B. We consider the case where model A nests model B. Let K_1 be the number of factors included both in A and in B ($\mathrm{K}_1 = 1$ in our case), and let K_2 be the number of factors unique in A ($\mathrm{K}_2 = 1$ in our case). The HJ distance of the two models are determined as

$$\delta_A^2 = \left(\bar{r} - V_{Rf_A}\hat{\lambda}_A\right)'V_{RR}^{-1}\left(\bar{r} - V_{Rf_A}\hat{\lambda}_A\right)$$

$$\delta_B^2 = \left(\bar{r} - V_{Rf_B}\hat{\lambda}_B\right)'V_{RR}^{-1}\left(\bar{r} - V_{Rf_B}\hat{\lambda}_B\right)$$

Under the null hypothesis that $\delta_A^2 = \delta_B^2$, we have

$$T(\delta_A^2 - \delta_B^2) \sim \sum_{j=1}^{K_2}\xi_j\chi_j^2$$

where χ_j^2's are independent random variables having the Chi-square distribution of order 1, and ξ_j's are eigenvalues of $\mathrm{H}_{22}^{-1}S_{22}$, where H_{22} and S_{22} are the lower-right blocks of $H_A = \left(V'_{Rf_A}V_{RR}^{-1}V_{Rf_A}\right)^{-1}$ and S_A. When $\mathrm{K}_1 = \mathrm{K}_2 = 1$,

$$\left[\left(V'_{Rf_A}V_{RR}^{-1}V_{Rf_A}\right)^{-1}\right]_{22}\frac{1}{S_{22}}\,T(\delta_A^2 - \delta_B^2) \sim N(0,1)^2 \tag{26}$$

Note that this test allows both models to be misspecified.

References

Asness, Clifford S., Jacques A. Friedman, Robert J. Krail, and John M. Liew (2000), "Style Timing: Value versus Growth," *Journal of Portfolio Management*, 26(3), 50-60.

Bessembinder, Hendrik (1992), “Systematic Risk, Hedging Pressure, and Risk Premiums in Futures Markets,” *Review of Financial Studies*, 5(4), 637~667.

Breeden, Douglas T. (1980), "Consumption Risks in Futures Markets," *Journal of Finance*, 35(2), 503~520.

Brunnerrmeier, Markus K., Stefan Nagel, and Lasse H. Pedersen (2008), “Carry Trades and Currency Crashes,” NBER Working Paper No. 14473.

Carter, Colin A., Gordon C. Rausser, and Andrew Schmitz (1983), “Efficient Asset Portfolios and the Theory of Normal Backwardation,” *Journal of Political Economy*, 91(2), 319-331.

Cochrane, John H. (1996), “A Cross-Sectional Test of an Investment-Based Asset Pricing Model,” *Journal of Political Economy*, 104(3), 572-621.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho (2003), "The Value Spread," *Journal of Finance*, 58(2), 609-641.

Cootner, Paul H. (1960), "Returns to Speculators: Telser versus Keynes," *Journal of Political Economy*, 68(4), 396~404.

Dusak, Katherine (1973), "Futures Trading and Investor Returns: An Investigation of Commodity Market Risk Premium," *Journal of Political Economy*, 81(6), 1387-1406.

Eaton, Morris L. (2007), *Multivariate Statistics: A Vector Space Approach*, Institute of Mathematical Statistics Lecture Notes-Monograph Series Vol. 53.

Erb, Claude B., and Campbell R. Harvey (2006), “The Strategic and Tactical Value of Commodity Futures,” *Financial Analysts Journal*, 62(2), 69-97.

Fama, Eugene F. (1984), “Forward and spot exchange rates,” *Journal of Monetary Economics*, 14, 319-338.

Fama, Eugene F., and Kenneth R. French (1987), “Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage,” *Journal of Business*, 60(1), 1987, 55~73.

-------- (1993), “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics*, 33, 3-56.

Fama, Eugene F., and James D. MacBeth (1973), “Risk, Return, and Equilibrium: Empirical Tests,” *Journal of Political Economy*, 81(3), 607-636.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken (1989), “A Test of the Efficiency of a Given Portfolio,” *Econometrica*, 57(5), 1121-1152.

Gorton, Gary B., Fumio Hayashi, and K. Geert Rouwenhorst (2012), "The Fundamentals of Commodity Futures Returns," *Review of Finance*, forthcoming.

Gorton, Gary B., and K. Geert Rouwenhorst (2006), “Facts and Fantasies about Commodity Futures,” *Financial Analysts Journal*, 62, 47-68.

Hansen, Lars Peter, and Ravi Jaganathan (1997), “Assessing Specification Errors in Stochastic Discount Factor Models,” *Journal of Finance*, 52(2), 557-590.

Jagannathan, Ravi (1985), "An Investigation of Commodity Futures Prices Using the Consumption-based Intertemporal Capital Asset Pricing Model," *Journal of Finance*, 40(1), 175~191.

Kaldor, Nicholas (1939), “Speculation and Economic Stability,” *Review of Economic Studies*, 7, 1-27.

Kan, Raymond, and Cesare Robotti (2009), “Model Comparison Using Hansen-Jaganathan Distance,” *Review of Financial Studies*, 22(9), 3449-3490.

Kan, Raymond, Cesare Robotti, and Jay Shanken (2013), “Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology,” *Journal of Finance*, forthcoming.

Keynes, John Maynard (1930), *A Treatise on Money* Vol. 2, London: Macmillan. (Reprinted as The *Collected Writings of John Maynard Keynes* Vol. 5 by Cambridge University Press in 1978.)

Kim, Daehwan (2012), "Value Premium across Countries," *Journal of Portfolio Management*, 38(4), 75-86.

Kim, Daehwan, and Chi-Young Song (2012), "Country Fundamentals and the Cross-Section of Currency Excess Returns," working paper.

Litzenberger, Robert H., and Nir Rabionwitz (1995), "Backwardation in Oil Futures Markets: Theory and Empirical Evidence," *Journal of Finance*, 50(5), 1517-1545.

Lustig, Hanno, Nikoali Roussanov, and Adiren Verdelhan (2008), “Common Risk Factors in Currency Markets,” NBER Working Paper No. 14082.

Miffre, Joëlle, and Georgios Rallis (2007), “Momentum Strategies in Commodities Futures Markets,” *Journal of Banking and Finance*, 31, 1863-1886.

Newey, Whitney K., and Kenneth D. West (1987a), "A Simple, Positive Semi-Definite, Heteroskdasticity and Autocorrelation Constant Covariance Matrix," *Econometrica*, 55(3), 703-708.

-------- (1987b), "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777-787.

Pojarliev, Momtchil, and Richard M. Levich (2008), "Do Professional Currency Managers Beat the Benchmark?" *Financial Analyst Journal*, 64(5), 18-32.

Shanken, Jay (1992), "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5(1), 1-33.

Shen, Qian, Andrew C. Szarkmary, and Subhash C. Sharma (2007), "An Examination of Momentum Strategies in Commodities Futures Markets," *Journal of Futures Markets*, 27(3), 227-256.

Stivers, Chris, and Licheng Sun (2010), "Cross-Sectional Return Dispersion and Time-Variation in Value and Momentum Premia," *Journal of Financial and Quantitative Analysis*, 45(4), 987-1014.

Szymanowska, Marta, Frans de Roon, Theo Nijman, and Rob Van Den Goorbergh (2013), "An Anatomy of Commodity Futures Risk Premia," *Journal of Finance*, forthcoming.

Tables and Figures

Table 1. Returns and Basis by Commodity

Statistics for monthly return and end-of-month basis are reported for each commodity, for the period from January 1981 to December 2010. The monthly return is calculated as the percentage change in the price of the nearest contract, with the adjustment for the possible change of the nearest contracts during the month. The basis is calculated as the percentage difference between the prices of the second-nearest and the nearest contracts (i.e. the second-nearest-contract price over the nearest-contract price, expressed in percent). The basis is adjusted for the difference in the expiration dates of the two contracts, i.e. the resulting figure is the rate of change over one month. T is the number of observations; SD is the standard deviation.

		Return				Basis			
	T	Mean	SD	Min	Max	Mean	SD	Min	Max
Agricultural Commodities									
Cocoa	360	0.07	8.77	-25.39	36.54	0.61	0.98	-5.77	2.91
Coffee	360	0.61	11.03	-31.27	50.60	0.38	1.49	-7.15	4.98
Corn	360	-0.16	7.22	-20.31	45.82	0.86	1.53	-12.22	2.55
Feeder Cattle	360	0.48	4.08	-20.58	14.83	-0.20	1.24	-4.94	5.40
Lean Hogs	360	0.53	7.74	-26.12	30.20	0.80	4.16	-10.26	13.19
Live Cattle	360	0.74	4.47	-24.48	14.50	-0.10	1.77	-4.74	3.94
Oats	360	0.61	10.29	-26.74	93.42	0.66	2.19	-10.03	3.82
Orange Juice	360	0.61	9.02	-22.13	61.01	0.44	1.45	-4.86	3.51
Pork Bellies	360	0.67	11.35	-33.87	44.11	-0.27	1.85	-7.83	9.29
Soybean Meal	360	0.87	7.50	-24.00	27.09	-0.31	2.28	-19.27	8.66
Soybean Oil	360	0.34	7.86	-28.13	43.99	0.47	1.09	-6.15	2.26
Soybeans	360	0.48	6.97	-25.12	27.99	0.19	1.60	-12.37	1.87
Sugar	360	0.21	11.25	-29.70	67.42	0.22	1.74	-5.91	4.68
Wheat	360	-0.12	7.39	-21.14	38.50	0.60	1.73	-8.72	3.24
Non-Agricultural Commodities									
Copper	360	1.09	8.23	-36.47	39.74	-0.14	1.16	-6.79	1.66
Cotton	360	0.20	7.57	-22.70	30.53	0.39	2.21	-22.34	9.08
Gold	360	0.09	4.73	-17.98	19.02	0.43	0.27	0.03	1.43
Heating Oil	360	1.16	9.86	-28.86	44.45	-0.13	2.43	-15.27	4.36
Lumber	360	-0.48	9.39	-23.87	45.24	1.19	2.49	-7.54	7.05
Palladium	360	0.96	9.97	-33.88	46.89	0.15	0.60	-4.51	4.91
Platinum	360	0.51	6.93	-31.86	34.03	0.15	0.47	-1.81	1.59
Silver	360	0.21	8.07	-27.52	28.84	0.50	0.47	-0.06	7.35

Table 2. Returns of Basis Sorted Portfolios

At the beginning of each month, commodities are sorted by basis and, based on the sort, they are grouped into four portfolios of approximately equal size. P1 is the lowest quartile portfolio; P4 is the highest quartile portfolio. “P1-P4” is the zero investment portfolio of the long position in P1 and the short position in P4. Portfolios are created out of (i) all 22 commodities, (ii) 14 agricultural commodities, and (iii) 8 non-agricultural commodities, as listed in Table 1. Statistics are reported for all years, and also for each 10-year sub-period. The average monthly returns of the portfolios are reported together with standard deviation in the round brackets and the t-statistic in the square brackets. (The t-statistic is for the null hypothesis that the average monthly return is zero.) * and ** indicate significance at 10% and 5%, respectively.

All Commodities:					Agricultural Commodities:					Non-Agricultural Commodities:				
P1(Low)	P2	P3	P4(High)	P1-P4	P1(Low)	P2	P3	P4(High)	P1-P4	P1(Low)	P2	P3	P4(High)	P1-P4
							A. All Years							
0.90	0.45	0.25	0.12	0.78	0.62	0.43	0.47	0.18	0.44	1.26	0.40	0.10	0.12	1.14
(4.67)	(4.68)	(4.71)	(4.77)	(5.49)	(5.19)	(5.35)	(5.34)	(5.64)	(6.63)	(7.17)	(6.01)	(6.18)	(6.13)	(8.32)
[3.66]	[1.82]	[1.01]	[0.49]	[2.69]	[2.28]	[1.52]	[1.67]	[0.62]	[1.26]	[3.35]	[1.26]	[0.30]	[0.38]	[2.60]
**	*			**	**		*			**				**
							B. 1981-1990							
0.45	0.15	-0.40	0.15	0.30	0.40	-0.01	0.03	0.14	0.26	0.99	-0.49	-0.43	0.01	0.98
(4.28)	(4.67)	(4.78)	(5.52)	(6.23)	(4.29)	(5.12)	(5.59)	(6.79)	(7.05)	(7.29)	(6.19)	(5.99)	(6.00)	(8.69)
[1.16]	[0.35]	[-0.92]	[0.30]	[0.53]	[1.01]	[-0.02]	[0.06]	[0.22]	[0.40]	[1.49]	[-0.86]	[-0.79]	[0.01]	[1.24]

Table 2. Returns of Basis Sorted Portfolios [Continued]

All Commodities:					Agricultural Commodities:					Non-Agricultural Commodities:				
P1(Low)	P2	P3	P4(High)	P1-P4	P1(Low)	P2	P3	P4(High)	P1-P4	P1(Low)	P2	P3	P4(High)	P1-P4
							C. 1991-2000							
0.86	0.26	0.29	-0.22	1.08	0.61	0.21	0.32	-0.34	0.95	1.09	0.67	-0.11	0.36	0.73
(3.99)	(3.43)	(3.27)	(3.61)	(5.17)	(4.75)	(4.69)	(4.48)	(4.38)	(6.31)	(6.13)	(5.41)	(4.35)	(5.74)	(8.11)
[2.37]	[0.81]	[0.96]	[-0.67]	[2.30]	[1.41]	[0.50]	[0.79]	[-0.85]	[1.65]	[1.95]	[1.36]	[-0.27]	[0.68]	[0.99]
**				**					*	*				
							D. 2001-2010							
1.38	0.94	0.87	0.44	0.94	0.87	1.09	1.06	0.75	0.12	1.71	1.01	0.84	0.01	1.70
(5.59)	(5.67)	(5.71)	(5.00)	(4.98)	(6.34)	(6.13)	(5.85)	(5.51)	(6.52)	(7.99)	(6.35)	(7.71)	(6.64)	(8.18)
[2.71]	[1.82]	[1.67]	[0.96]	[2.07]	[1.50]	[1.94]	[1.99]	[1.49]	[0.19]	[2.35]	[1.74]	[1.19]	[0.01]	[2.28]
**	*	*		**		*	**			**	*			**

Table 3. Factor Model Estimates – One Step Estimation

Factor models are estimated for each commodity. The dependent variable is the return of each commodity. The explanatory variables are indicated in the model name. “Market” refers to the stock-and-commodity-market factor, which is the simple average of the stock market return (MSCI World Index return over the short-term US T-bill rate) and the commodity market return (equal-weighted average of commodity futures returns). “LHB” refers to the low-high basis factor. Monthly data from January 1981 to December 2010 are used for the regressions. Each row represents a separate regression. Coefficient estimates are reported first, followed by t statistics inside square brackets. * and ** indicate significance at 10% and 5%, respectively.

A. “Market” Model

	Constant			Market			R sq
Agricultural Commodities:							
Cocoa	-0.21	[-0.46]		0.66	[4.71]	**	0.06
Coffee	0.22	[0.39]		0.90	[5.16]	**	0.07
Corn	-0.57	[-1.66]	*	0.97	[9.06]	**	0.19
Feeder Cattle	0.40	[1.88]	*	0.18	[2.64]	**	0.02
Lean Hogs	0.37	[0.90]		0.37	[2.94]	**	0.02
Live Cattle	0.61	[2.62]	**	0.31	[4.26]	**	0.05
Oats	0.13	[0.26]		1.11	[6.98]	**	0.12
Orange Juice	0.35	[0.75]		0.59	[4.05]	**	0.04
Pork Bellies	0.44	[0.74]		0.52	[2.82]	**	0.02
Soybean Meal	0.46	[1.26]		0.96	[8.57]	**	0.17
Soybean Oil	-0.10	[-0.27]		1.03	[8.75]	**	0.18
Soybeans	0.03	[0.08]		1.05	[10.42]	**	0.23
Sugar	-0.14	[-0.24]		0.81	[4.51]	**	0.05
Wheat	-0.54	[-1.53]		0.99	[9.01]	**	0.18
Non-Agricultural Commodities							
Copper	0.60	[1.54]		1.12	[9.25]	**	0.19
Cotton	-0.18	[-0.49]		0.88	[7.61]	**	0.14
Gold	-0.11	[-0.44]		0.45	[6.10]	**	0.09
Heating Oil	0.90	[1.75]	*	0.60	[3.77]	**	0.04
Lumber	-0.88	[-1.86]	*	0.93	[6.39]	**	0.10
Palladium	0.37	[0.78]		1.38	[9.37]	**	0.20
Platinum	0.06	[0.18]		1.07	[10.77]	**	0.24
Silver	-0.23	[-0.60]		1.04	[8.57]	**	0.17

Table 3. Factor Model Estimates – One Step Estimation [Continued]

B. "Market + LHB" Model

	Constant			Market			LHB			R sq
Agricultural Commodities:										
Cocoa	-0.17	[-0.37]		0.65	[4.66]	**	-0.05	[-0.58]		0.06
Coffee	0.23	[0.40]		0.90	[5.14]	**	-0.01	[-0.13]		0.07
Corn	-0.40	[-1.15]		0.95	[8.96]	**	-0.21	[-3.46]	**	0.21
Feeder Cattle	0.33	[1.51]		0.19	[2.80]	**	0.09	[2.39]	**	0.03
Lean Hogs	0.32	[0.79]		0.37	[2.97]	**	0.05	[0.71]		0.02
Live Cattle	0.53	[2.28]	**	0.31	[4.41]	**	0.09	[2.20]	**	0.06
Oats	0.39	[0.75]		1.08	[6.86]	**	-0.31	[-3.33]	**	0.15
Orange Juice	0.52	[1.11]		0.57	[3.92]	**	-0.21	[-2.44]	**	0.06
Pork Bellies	0.14	[0.23]		0.56	[3.07]	**	0.37	[3.48]	**	0.05
Soybean Meal	0.48	[1.30]		0.96	[8.52]	**	-0.02	[-0.36]		0.17
Soybean Oil	-0.06	[-0.17]		1.02	[8.69]	**	-0.05	[-0.66]		0.18
Soybeans	0.04	[0.11]		1.04	[10.37]	**	-0.01	[-0.19]		0.23
Sugar	0.10	[0.17]		0.78	[4.37]	**	-0.29	[-2.75]	**	0.07
Wheat	-0.40	[-1.13]		0.97	[8.91]	**	-0.17	[-2.63]	**	0.20
Non-Agricultural Commodities										
Copper	0.59	[1.48]		1.13	[9.24]	**	0.02	[0.27]		0.19
Cotton	-0.20	[-0.52]		0.88	[7.60]	**	0.02	[0.27]		0.14
Gold	-0.19	[-0.81]		0.46	[6.29]	**	0.11	[2.48]	**	0.11
Heating Oil	0.78	[1.50]		0.61	[3.86]	**	0.14	[1.56]		0.04
Lumber	-0.82	[-1.71]	*	0.93	[6.33]	**	-0.08	[-0.88]		0.10
Palladium	0.06	[0.13]		1.42	[9.88]	**	0.38	[4.49]	**	0.24
Platinum	-0.12	[-0.37]		1.09	[11.17]	**	0.21	[3.74]	**	0.27
Silver	-0.28	[-0.71]		1.04	[8.60]	**	0.06	[0.84]		0.17

Table 4. Specification Tests and Model Comparison: One Step Estimation

The one factor model (Market) and the two factor model (Market + LHB) are tested and compared. The specification test is the test of the zero-intercept property of factor models. F test and J test are carried out. In the F test, the test statistic is calculated after the commodity-by-commodity OLS estimation as reported in Table 3. In the J test, the test statistic is calculated after the pooled GMM estimation. In both tests, a low p-value indicates that the null of zero-intercept is rejected. A high p-value supports the zero-intercept hypothesis. The model comparison is based on the chi-square test statistic, which is computed as the difference in the objective function of the GMM estimation. A low p-value indicates the similarity whereas a high p-value indicates the difference.

	Test Stat	P Value
Model: Market		
Specification Test - F (df1 = 22, df2 = 337)	1.33	0.15
Specification Test - J (df = 22)	14.02	0.90
Model: Market + LHB		
Specification Test - F (df1 = 22, df2 = 336)	1.14	0.31
Specification Test - J (df = 22)	14.46	0.88
Model Comparison		
Chi Square Test (df = 22)	191.20	<.0001

Table 5. Factor Model Estimates – Two Pass CSR

In the two pass CSR, the dependent variable is the average commodity futures return, and the right-hand-side variable is the covariance between commodity futures returns and factor returns. The coefficients on the right-hand side variables are interpreted as "factor price." Four sets of two pass CSR estimates are reported: (1) the OLS estimates of the "market" model, (2) the OLS estimates of the "market + LHB" model, (3) the GMM estimates of the "market" model, and (4) the GMM estimates of the "market + LHB" model. Coefficient estimates are reported first, followed by t statistics inside square brackets. * and ** indicate significance at 10% and 5%, respectively. R squared is reported for the OLS estimations, whereas Hansen-Jaganathan distance (HJ) is reported for the GMM estimation.

	OLS		GMM	
	(1)	(2)	(3)	(4)
Market Price	0.0458	0.0498	0.0488	0.0516
	[1.59]	[1.62]	[3.73]	[3.40]
			**	**
LHB Price		0.0366		0.0487
		[1.04]		[2.45]
				**
R Sq	0.4725	0.5774		
HJ Sq			0.0872	0.0735

Table 6. Model Comparison – Two Pass CSR

The model comparison is based on the R squared of OLS estimation and also on the HJ distance of GMM estimations. In both cases, the test statistics follow the chi square distribution with the degree of freedom 1. A low p-value indicates the similarity whereas a high p-value indicates the difference.

	Test Stat	P Value
Test of		
Equal R squared (chi square, df = 1)	1.07	0.30
Equal HJ distance (chi square, df = 1)	6.00	0.01

Table 7. Stock, FX, and Commodity Market Factors: Univariate Statistics and Correlations

Monthly returns are collected for the period from October 2000 (first month when carry is available) to December 2010. "Stock" is the stock market return; it is calculated as the MSCI World Index return over the short-term US T-bill rate. "SMB" is the return to the strategy of buying small capitalization stocks and selling large capitalization stocks, whereas "HML" is the return to the strategy of buying high book-to-market stocks and selling low book-to-market stocks. Both SMB and HML are as calculated by Kenneth French. "Carry" is the return to the FX carry strategy; it is calculated as the rate of change of the G10 Currency Harvest USD Index of Deutsche Bank, which is the return to the strategy of buying high yield currencies and selling low yield currencies. "Cmdty" is the commodity market return; it is calculated as the equal-weighted average of commodity futures returns. "LHB" refers to the low-high basis factor, i.e. the return to the strategy of buying low basis commodities and selling high basis commodities. T is the number of observations; SD is the standard deviation.

	T	Mean	SD	Min	Max	Correlation					
						(1)	(2)	(3)	(4)	(5)	(6)
(1) Stock	123	0.07	4.93	-19.00	11.21	1.00					
(2) SMB	123	0.52	2.77	-6.53	6.98	0.28	1.00				
(3) HML	123	0.58	3.28	-9.93	13.88	0.00	-0.05	1.00			
(4) Carry	123	0.37	2.78	-14.26	8.22	0.59	0.16	0.18	1.00		
(5) Cmdty	123	0.89	4.31	-15.93	12.55	0.48	-0.01	0.01	0.47	1.00	
(6) LHB	123	1.02	4.95	-13.35	14.20	-0.07	0.00	0.07	-0.11	0.06	1.00

Table 8. Stock, FX, and Commodity Market Factors: VAR

Vector Auto-Regression of order 3 is estimated for the stock, FX, and commodity market factors. See Table 7 for the description of each factor. Each column represents a separate equation. Coefficient estimates are reported first, followed by t statistics inside square brackets. * and ** indicate significance at 10% and 5%, respectively.

	Stock	SMB	HML	Carry	Cmdty	LHB
Constant	0.49	0.62	-0.26	0.38	1.34	0.63
	[0.94]	[2.03]	[-0.88]	[1.23]	[2.74]	[1.16]
		**			**	
One-Month Lagged Variables						
Stock[-1]	0.32	0.14	0.09	0.09	0.10	0.00
	[2.59]	[2.04]	[1.28]	[1.24]	[0.87]	[0.02]
	**	**				
SMB[-1]	-0.33	-0.18	0.30	-0.21	-0.15	0.32
	[-1.87]	[-1.80]	[2.94]	[-2.04]	[-0.93]	[1.73]
	*	*	**	**		*
HML[-1]	-0.20	0.02	0.11	-0.06	-0.14	0.15
	[-1.26]	[0.20]	[1.20]	[-0.65]	[-0.96]	[0.90]
Carry[-1]	0.15	0.09	0.13	0.13	0.03	-0.07
	[0.70]	[0.73]	[1.02]	[1.03]	[0.15]	[-0.32]
Cmdty[-1]	-0.21	0.00	-0.05	-0.10	-0.09	0.26
	[-1.69]	[0.00]	[-0.65]	[-1.30]	[-0.80]	[1.98]
	*					**
LHB[-1]	0.03	-0.01	0.04	0.05	0.01	-0.05
	[0.32]	[-0.27]	[0.75]	[0.94]	[0.14]	[-0.53]
Two-Month Lagged Variables						
Stock[-2]	-0.21	-0.13	-0.08	0.04	0.02	0.05
	[-1.68]	[-1.81]	[-1.09]	[0.60]	[0.17]	[0.36]
	*	*				
SMB[-2]	-0.12	0.16	0.24	-0.11	-0.26	0.01
	[-0.68]	[1.53]	[2.31]	[-1.01]	[-1.51]	[0.06]
			**			
HML[-2]	-0.10	-0.11	-0.21	-0.12	-0.25	0.02
	[-0.69]	[-1.28]	[-2.55]	[-1.36]	[-1.86]	[0.11]
			**		*	
Carry[-2]	0.17	0.12	0.03	0.01	0.23	-0.47
	[0.79]	[0.95]	[0.22]	[0.07]	[1.14]	[-2.16]
						**
Cmdty[-2]	0.08	0.00	-0.01	0.02	0.02	0.02
	[0.65]	[0.03]	[-0.10]	[0.34]	[0.17]	[0.13]
LHB[-2]	-0.01	-0.01	-0.05	0.07	0.04	0.04
	[-0.11]	[-0.11]	[-0.97]	[1.42]	[0.44]	[0.49]

Table 8. Stock, FX, and Commodity Market Factors: VAR [Continued]

	Stock	SMB	HML	Carry	Cmdty	LHB
Three-Month Lagged Variables						
Stock[-3]	0.09	-0.03	0.03	-0.01	0.22	-0.03
	[0.73]	[-0.34]	[0.40]	[-0.09]	[1.88]	[-0.20]
					*	
SMB[-3]	-0.10	-0.01	0.18	-0.06	-0.01	0.12
	[-0.56]	[-0.12]	[1.74]	[-0.58]	[-0.06]	[0.63]
			*			
HML[-3]	-0.34	0.07	0.19	-0.09	-0.12	0.19
	[-2.38]	[0.81]	[2.37]	[-1.09]	[-0.92]	[1.32]
	**		**			
Carry[-3]	0.44	0.06	0.03	0.30	0.06	-0.65
	[2.00]	[0.45]	[0.20]	[2.33]	[0.28]	[-2.88]
	**			**		**
Cmdty[-3]	0.06	-0.03	0.19	0.00	-0.14	0.16
	[0.45]	[-0.40]	[2.71]	[0.04]	[-1.17]	[1.26]
			**			
LHB[-3]	0.05	-0.06	0.07	0.10	0.06	-0.04
	[0.59]	[-1.09]	[1.37]	[1.88]	[0.76]	[-0.47]
				*		
R squared	0.24	0.18	0.35	0.21	0.14	0.20

Table 9. Optimal Allocation with and without the LHB Factor

An optimal portfolio is the portfolio with the maximum Sharpe ratio, i.e. it is the tangent portfolio in the *excess* mean return-standard deviation space. The weights of two optimal portfolios, with and without including the LHB factor, are listed together with the monthly excess mean return and standard deviation. SR refers to the Sharpe ratio, which is calculated as the annualized excess mean return over the annualized standard deviation.

	(1)	(2)
Portfolio Weights:		
Stock	-0.20	-0.25
SMB	0.40	0.49
HML	0.22	0.29
Carry	0.14	0.11
Cmdty	0.27	0.36
LHB	0.16	
Return Statistics:		
Mean (monthly)	0.78	0.77
SD (monthly)	1.89	2.03
SR (annual)	1.44	1.31

Table 10. LHB Return, IP, and Spreads: Univariate Statistics and Correlations

Statistics are calculated for the period from Feb 1989 (the first month when the volatility spread is available) to December 2010. "LHB" refers to the low-high basis factor, i.e. the return to the strategy of buying low basis commodities and selling high basis commodities. "IP" is the rate of changes of the seasonally adjusted US industrial production index. At the end of each month, for each commodity, the basis is calculated as the percentage difference between the prices of the second-nearest and the nearest contracts, and adjusted for the difference in the expiration dates of the two contracts. (See Table 1 for further description of basis.) Then the basis spread is calculated as the difference between the 12.5th and 87.5th percentiles. Hedging pressure is calculated as hedgers' net long position relative to all open interest. Hedging spread is also calculated as the difference between the 12.5th and 87.5th percentiles. Volatility is the implied volatility from the futures options. Volatility spread is also calculated as the difference between the 12.5th and 87.5th percentiles. T is the number of observations; SD is the standard deviation.

	T	Mean	SD	Min	Max	Correlation				
						(1)	(2)	(3)	(4)	(5)
(1) LHB	263	0.90	5.02	-13.35	14.20	1				
(2) IP	263	0.16	0.67	-4.21	2.12	0.02	1			
(3) Basis spread	263	3.21	1.06	1.13	7.59	0.11	0.00	1.00		
(4) Hedging spread	263	0.55	0.11	0.25	0.91	0.00	0.08	-0.13	1	
(5) Volatility spread	263	24.71	9.71	12.90	131.62	-0.05	0.04	0.04	-0.07	1

Table 11. LHB Return, IP, and Spreads: Predictive Regressions

The LHB return is regressed on IP and spread variables. See Table 10 for the description of IP and spread variables. Each column represents a separate regression. Coefficient estimates are reported first, followed by t statistics inside square brackets. * and ** indicate significance at 10% and 5%, respectively.

	(1)	(2)	(3)	(4)	(5)
Constant	0.94	1.15	1.14	-0.72	-0.39
	[2.96]	[1.16]	[0.69]	[-0.85]	[-0.18]
	**				
IP[-1]	-0.34				-0.38
	[-0.74]				[-0.82]
Basis spread[-1]		-0.08			-0.10
		[-0.27]			[-0.35]
Hedging spread[-1]			-0.45		0.03
			[-0.15]		[0.01]
Volatility spread[-1]				0.07	0.07
				[2.05]	[2.08]
				**	**
R sq.	0.0021	0.0003	0.0001	0.0159	0.0190

Table 12. Returns to Market Timing Strategies

With the market timing strategy (m, k), an investor follows the strategy underlying the LHB factor only if predictor variable x is greater than k times the m-month moving average of x, i.e., $x_t \geq k(1/m)(x_{t-m+1} + \cdots + x_t)$. Otherwise, the investor stays with the risk-free asset. We consider IP and three spread variables as predictor variable x. See Table 10 for the description of IP and spread variables. The moving average parameter m takes one of three values: 6, 9, 12. The coefficient k takes one of five values: 0, 0.25, 0.5, 0.75, 1. Note that, when k=0, the market time strategy return is identical to the LHB factor return. For each strategy, the average monthly excess return is listed first followed by standard deviation inside the round brackets and the t statistics inside the square bracket. (The t-statistic is for the null hypothesis that the average monthly return is zero.) * and ** indicate significance at 10% and 5%, respectively.

	m=6				m=9				m=12			
k=0	k=0.25	k=0.5	k=0.75	k=1	k=0.25	k=0.5	k=0.75	k=1	k=0.25	k=0.5	k=0.75	k=1
						Predictor: IP[-1]						
0.59	0.39	0.30	0.29	0.37	0.36	0.27	0.27	0.21	0.32	0.29	0.26	0.26
(4.44)	(4.03)	(3.94)	(3.87)	(3.61)	(4.02)	(3.92)	(3.81)	(3.63)	(4.10)	(3.96)	(3.81)	(3.55)
[2.11]	[1.52]	[1.19]	[1.19]	[1.62]	[1.43]	[1.10]	[1.12]	[0.90]	[1.25]	[1.17]	[1.08]	[1.17]
**												
						Predictor: Basis Spread[-1]						
0.59	0.95	0.95	0.94	0.72	0.95	1.00	0.99	0.56	0.95	0.93	0.90	0.56
(4.44)	(5.01)	(5.01)	(4.63)	(3.71)	(5.01)	(4.99)	(4.58)	(3.61)	(5.01)	(4.98)	(4.56)	(3.58)
[2.11]	[3.00]	[2.99]	[3.22]	[3.08]	[3.00]	[3.16]	[3.42]	[2.46]	[3.00]	[2.94]	[3.13]	[2.47]
**	**	**	**	**	**	**	**	**	**	**	**	**
						Predictor: Hedging Spread[-1]						
0.59	0.95	0.95	0.94	0.58	0.95	0.95	0.96	0.62	0.95	0.95	0.97	0.68
(4.44)	(5.01)	(5.01)	(4.97)	(3.65)	(5.01)	(5.01)	(4.99)	(3.56)	(5.01)	(5.01)	(4.92)	(3.48)
[2.11]	[3.00]	[3.00]	[2.97]	[2.53]	[3.00]	[3.00]	[3.06]	[2.74]	[3.00]	[3.00]	[3.12]	[3.11]
**	**	**	**	**	**	**	**	**	**	**	**	**
						Predictor: Volatility Spread[-1]						
0.59	0.95	0.95	1.00	0.82	0.95	0.95	0.92	0.65	0.95	0.95	0.90	0.68
(4.44)	(5.01)	(5.01)	(4.85)	(3.54)	(5.01)	(5.01)	(4.91)	(3.48)	(5.01)	(5.02)	(4.84)	(3.53)
[2.11]	[3.00]	[3.00]	[3.25]	[3.64]	[3.00]	[3.00]	[2.97]	[2.95]	[3.00]	[2.98]	[2.95]	[3.06]
**	**	**	**	**	**	**	**	**	**	**	**	**

Figure 1. Investment Possibility Sets with and without the LHB Factor

The investment possibility set in the excess mean return-standard deviation space is identified by the efficient frontier (the collection of the minimum standard deviation achievable for each excess mean return) and the tangent line passing through the origin. Two such sets are drawn below, one with the LHB factor and the other without the LHB factor. All the other factors described in Table 7 are used to construct both investment possibility sets.

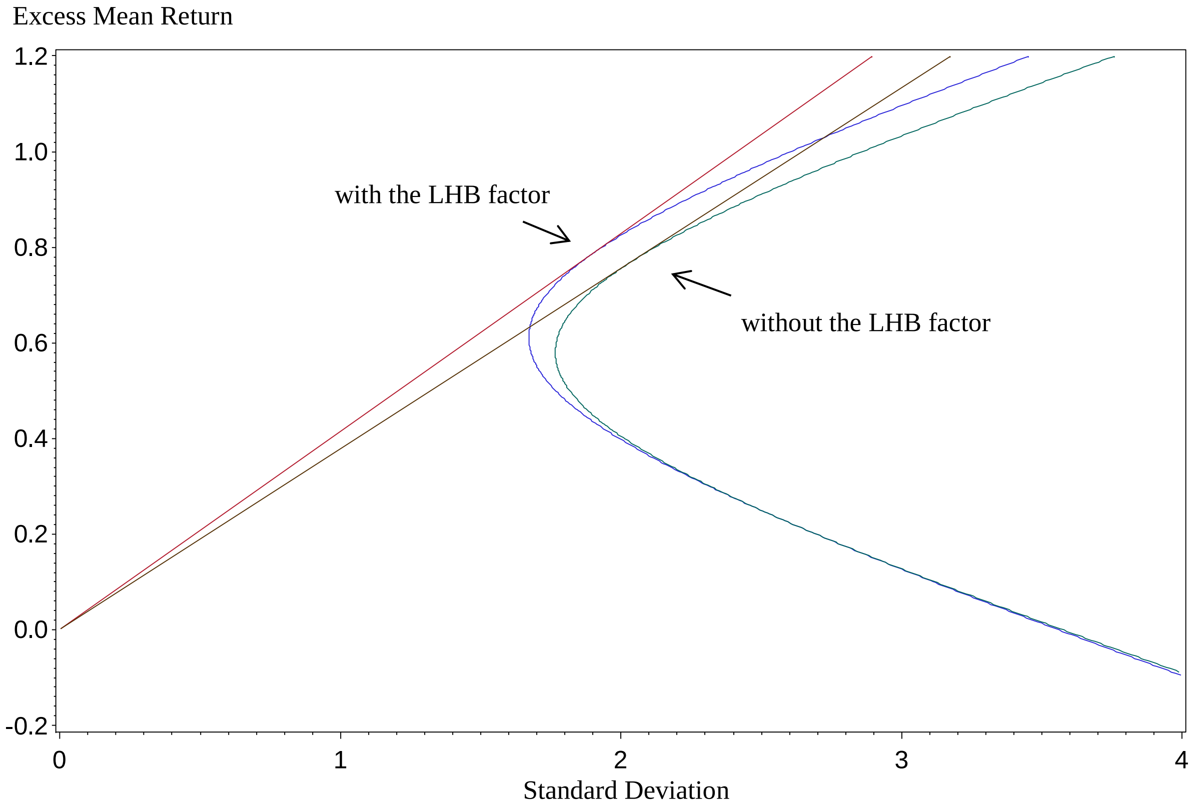

Figure 2. LHB Index and Commodity Market Index over Business Cycles

LHB Index represents the cumulative returns of the long-short strategy—buying low basis commodity futures and selling high basis commodity futures. Commodity Market Index represents the cumulative returns of the equal weighted portfolio of commodity futures. Shaded area indicates the “contraction periods” (peak to bottom) of the US economy as determined by National Bureau of Economic Research.