Tactical Asset Allocation with Commodity Futures: Implications of Business Cycle and Monetary Policy*

Van Thi Tuong Nguyen†
Piet Sercu‡

Tuesday 9th November, 2010

*We gratefully acknowledge financial support from FWO, as well as very useful comments from Michael Brennan, Frans de Roon, Geert Dhaene, and Gunther Wuyts (Van's PhD committee). All remaining errors are our responsibility.

†FWO PhD Fellow at Leuven School of Business and Economics, K.U.Leuven, Naamsestraat 69, B-3000 Leuven; +32 16 32 64 62; thituongvan.nguyen@econ.kuleuven.be

‡Leuven School of Business and Economics, K.U.Leuven, Naamsestraat 69, B-3000 Leuven; +32 16 32 67 56; Piet.Sercu@econ.kuleuven.be

Abstract

While Gorton and Rouwenhorst (2005) suggest using business cycle in tactical asset allocation with commodity futures, Jensen *et al* (2002) suggest using monetary policy in guiding the timing of investment. We investigate whether it is useful to watch both. The performance of out-of-sample optimal portfolios show that the proposed strategy with commodity futures performs better than (i) any stand-alone assets (stocks, bonds, commodity futures); (ii) the optimal portfolio without commodity futures and (iii) strategies that consider only one type of information.

Diverging from Gorton and Rouwenhorst (2005), we suggest to divide a business cycle into three stages (early, middle and late). To exploit the implication of business cycle in tactical asset allocation, we use the average duration of a recession or an expansion to divide a cycle which is announced by NBER into *ex ante* stages. Our tactical strategy is to go long commodity futures: (i) with a restrictive policy in middle, late stages of booms and during the recession; and (ii) under an expansive policy: in a boom. This strategy works not only for US but also for British and Japanese investors.

JEL classification:
Key words: commodity futures, asset allocation, portfolio diversification, business cycle, monetary policy

Introduction

With respect to investing in commodity futures, Gorton and Rouwenhorst (2005; henceforth GR) and Jensen, Johnson and Mercer (2002; henceforth JJM) both recommend a flexible strategy, letting the investment decision depend on the stage of the business cycle (GR) or on monetary policy (JJM). Each of these papers documents interesting improvements in returns and risks, relative to a static buy-and-hold policy. We think there could be both more, and less, than meets the eye, in these papers. On the upside we hope for further improvement by taking into account both business-cycle and monetary-policy information at the same time. But on the downside we also note that the GR and JJM results might just be exploiting hindsight: in order to get their results, the investor should have known the conditional means given each scenario (that is, everything in GR and JJM is in-sample), as well as the exact starting date and length of each cycle. This ignores randomness in sample means, gaps between cycle turning points and their announcement by the NBER, and uncertainty about the length of each boom or bust period. In fact, neither GR nor JJM offer any significance tests on the in-sample results.

It turns out that, indeed it is useful to take into account the interaction between business cycle and monetary policy, and indeed (or not, if you wish), all the potential gains evaporate or become at least statistically insignificant when we use only public, historic information. While the partial unpredictability of conditional mean returns and of the exact length of each boom/bust episode are not so crucial, the delays in the announcements of the turning points seem to be the main killjoy. In short, the case for tactical management of commodity positions following the GR or JJM rules, or a combination, is far weaker than the hasty reader would think. Even the gains from statically investing in commodity futures are statistically unclear—although at least the sample means still point in the right direction, in line with economically logical priors.

In the rest of this introduction we flesh out the arguments, tests and findings in greater detail.

Although commodities have been considered as an investible asset class since at least 1978 (Greer, 1978), widespread inclusion of commodities in the asset allocation decision is a more recent phenomenon. Bjornson and Carter (1997), and Weiser (2003) report that expected commodity returns and commodity futures returns all depend on the business cycle. In particular, Bjornson and Carter (1997) find that commodity expected returns are lower during times of high interest rates, expected inflation and economic growth, and Weiser (2003) reports that commodity futures returns change with different stages of a business cycle. As direct investment in physical commodities is not practical because of storage costs and the perishable nature of many commodities, research on the strategic and tactical asset allocation focuses on the commodity futures.

Studies about the role of commodity futures in a diversified portfolio (Lummer and Siegel, 1993; Kaplan and Lummer, 1997; Greer, 2000; Jensen, Jonson and Mercer, 2000, 2002; Gorton and Rouwenhorst, 2005; Erb and Harvey, 2006; Ibbotson Associates, 2006; Laws and Thompson, 2007; and Roache, 2008) all bring out the same finding: under the appropriate circumstances, a diversified portfolio with commodity futures provides higher average returns and a better Sharpe ratio than the traditional portfolio of stocks, bonds and even real estate. This conclusion is also robust for international investors when taking into account the international stock market or the stock markets of other countries (GR; Laws and Thompson, 2007). In short, the most important issue in asset allocation with commodity futures may not be whether we should include commodities into a diversified portfolio at all, but under what economic circumstances we should do so.

The answer to this question is related to the above finding that the behavior of commodity futures returns varies with the business cycle and monetary policy. Thus far, the impacts of cyclical and monetary circumstances have been studied separately only. According to GR, investors should include commodity futures in their diversified portfolios in the early stage of a recession and the late stage of an expansion (henceforth 'boom' to differentiate an expansion from an expansive monetary policy). They find that commodity futures perform well in the early stage of a recession: during those episodes, stock and bond returns are negative while commodity futures returns are positive. However, in the later phase of a recession, the signs of the returns reverse: stock and bond returns become positive while commodity futures turn negative. They also show that when stock and bond returns are below their overall average, which happens in the late boom and early recession stages, commodity returns are positive and outperform both stocks and bonds.

Following another lead, JJM use the Federal Reserve's monetary policy in guiding the timing of investment in commodity futures. They conclude that investors should increase their exposure to commodity futures and reduce their weights to equities during restrictive monetary phases, *i.e.* following a rise in the interest rate. When an expansive policy is followed, in contrast, futures have insignificant weights in mean-variance efficient portfolios. JJM state that periods of restrictive monetary policy tend to coincide with periods of heightened inflationary concerns. Thus, this evidence echoes the view that commodity futures act as an inflation hedge.

There are four problems with these two conclusions. First and foremost, all the above is in-sample analysis, that is, its validity is untested outside their data base. Second, the two rules often contradict. Based on historical data from the NBER and the Fed, we find that in more than half of the recessions the Fed applied an expansionary monetary policy (by lowering the discount rate), especially during this current crisis. So there is a conflict: following GR, investors should put more weight on commodity futures in this period of recession, but according to JJM, they should usually do the reverse. One way to try and solve the conundrum is to hypothesize that the choice is not between working with either two monetary regimes or (in GR) four business-cycle substages, but with 4×2 regimes, thus letting the monetary rule vary with the cycle stages and *vv.* The third issue is that, as we know, business cycles are announced after the fact. For instance, the current recession was announced by NBER in December 2008 while it actually started in December 2007. Real-world investors, therefore, cannot use the actual beginning and end business cycle, and reacting 'late' may very well wipe out the gains. Next, GR/JJM just compare values of returns and Sharpe ratios to draw their conclusions without adding any significance test. Lastly, relating to the GR conclusion there is an extra problem: even if there were no delay in announcing a turning point, the investors would still have problems determining exactly when the current phase is half its way—the point where an early boom, for instance, turns into a late boom.

The above implies the following three research questions:

1. Is there an interaction effect between monetary policy and business cycle, and does a combined strategy of considering both business cycle and monetary policy perform better?

   Behind this idea is our finding that monetary policy varies across different stages of a boom: during a typical boom, a restrictive policy applies towards the end of the episode while during its early and middle stages an expansive policy is being administered. We therefore propose to divide a boom into three stages (early, middle, late) rather than two stages as in GR. For recessions, we entirely give up on substages: recessions typically last 12-14 months only (as opposed to 60-120 for booms), and by the time the bust phase is actually announced it is already well into its second half. In addition, and probably relatedly, during recessions the monetary stance is confusing things rather than reinforcing the cycle: money is as likely to be expansive as restrictive, and stock returns hardly differ across two or three sub-phases.[1]

2. What is the outcome if, while trying to exploit the effect of the business cycle in asset allocation, we take into account that the NBER announcement is *ex post* and that the length of the newly announced phase is unknown?

3. If, *prima facie*, the inclusion of commodity futures in a diversified portfolio creates higher returns and a Sharpe ratio, does the phenomenon persist out of sample and is the gain significant?

We study monthly data from 1/1970 to 8/2009. For the purpose of verification, we also study the GR sample, 4/1973-12/2007. In our in- versus out-of-sample analysis we work with an in-sample test period that runs from 1/1970 to 6/1990, which we use to extract the proposed tactical allocation rule. The 1990-2009 sample then is used for out-of-sample testing. Following GR and JJM, good and bad stages per asset class—stocks, bonds, and commodity futures—are identified by comparing their mean returns and Sharpe ratios in each phase—recessions, and early/middle/late booms. In order to exploit the effect of the business cycle in a tactical asset allocation, we propose to use the historic average duration of a boom to get three *ex ante* stages for a boom following NBER's announcement. For the purpose of performance evaluation, we study the assets either in a stand-alone mode or as part of a portfolio. In stand-alone tests we compare, per asset, static buy-and-hold returns to the results obtained, if we switch from 100% long to 100% short depending on what our rule suggests—a trading rule, in short. In our Markowitz-style analysis we compare mean-variance efficient portfolios that include commodities to portfolios that do not, and we compare a tactical to a static version. In the tactical version we use historical means and variance/covariance matrices given each economic condition to proxy for conditional expected mean and variance returns, updating the historical matrices each time a new NBER announcement is issued. For the static portfolio rule we use similar data except that they are not conditioned on a particular cycle stage. To test for the significance of the observed differences, lastly, we apply the Diebold and Mariano (1995) test to returns and to standardized excess returns (whose average is the Sharpe ratio).

The good news is that there is indeed an interaction effect of business cycles and monetary policies in asset allocation. The rule is most easily stated negatively, *i.e.* as a statement about

[1] The issue may also be a low number of observations. For science' sake we need more recessions :-).

when not to hold commodities. In-sample and assuming realistic information on the turning points of the cycle, the modified GR×JJM rule says to get out of commodities, stand-alone:

– when an expansive monetary policy applies: during a recession; and

– when a restrictive monetary policy applies: during the early phase of a boom.

This rule does better, in-sample, than either the GR or JJM version (where only one condition is considered) or than buy-and-hold. Making the application more realistic by working with *ex-ante* substages rather then exact *ex post* substages does not materially change our conclusions. The extra returns even keep their signs when we apply the rule out of sample. But—and here comes the bad news —when the late announcement of turning points is brought into the picture, all benefits from the clever timing in tactical asset allocation become insignificant (in portfolio tests) or seem to disappear totally (in stand-alone tests). For the optimal portfolio, it is true that the combined tactical strategy still offers a better return and Sharpe ratio than the unconditional strategy and than the rival conditional strategies of GR and JJM. However, the differences are not statistically significant. The same mixed message emerges about whether to hold commodities or not: for each rule (static, JJM, GR, and combined) the portfolio that includes commodities has better returns and Sharpe ratios than the alternative without commodities, but the gains are not significant. For a stand-alone investment, in contrast, the return differential turns from significantly better to insignificantly worse, that is, the combined rule seems to underperform JJM and even buy-and-hold, even though the difference is statistically insignificant. In short, our results are substantially less encouraging than those of previous studies, which may have relied too much on in-sample hindsight and on descriptive numbers without significance test.

The remainder of this paper is structured as follows. In Section 1, we describe our data and our way to have *ex ante* stages of a new business cycle. In Section 2, following the in-sample analysis, we suggest the combined tactical asset strategy. We check for the validity of the proposed strategy in the out-of-sample in Section 3. Section 4 concludes.

1 Data and suggested method to have *ex ante* stages of a new cycle

In this section, we first describe our choice of stock and commodity indices to be used in our analysis, and we present a summary of historical data about business cycles. We then explain our methodology for dividing a new business cycle into ex-ante stages.

1.1 Stock and commodity indices

We use monthly total returns[2] to perform our analysis. The benchmark portfolio for a US investor includes four assets: (i) US stocks: US index from Morgan Stanley Capital International (MSCI);[3] (ii) international stocks: following JJM we choose the Far East index (EAFE) from MSCI; (iii) US corporate bonds: the Corporate US Aggregate AA Long from Barlays (formerly Lehman Brothers); and (iv) 1-month T-Bill: data from Ibbotson Associates.

According to Erb and Harvey (2006), the Goldman Sachs Commodity Index (GSCI) Composite is more suitable than an equally-weighted commodity index as provided by either GR, or by the Center Research Bureau (CRB index), or the Dow Jones AIG index. Currently, the GSCI Composite (henceforth GSCI) includes nearby futures contracts for 24 commodities in five components: Energy (accounting for 71.80% dollar weights), Industrial Metals (7.41%), Precious Metals (3.09%), Agricultural (13.73%) and Livestock (3.96%). The GSCI total return index measures the pay-off from fully collateralized commodity futures, *i.e.* the zero-investment return from a long position in the futures contract plus the T-bill rate. This makes it comparable to the total return from investing in stocks or bonds. To verify Gorton and Rouwenhorst (2005)'s conclusion, we also apply our analysis to their equally-weighted commodity index without energy components, GRCI, downloaded as a total return from the NBER web site. The other data we use are all from Datastream.

Our in-sample analysis is from 1/1970 to 6/1990[4] and our out-of-sample test is from 7/1990 to 8/2009. For verification and comparison, the results for the GR sample period (4/1973 to 12/2007) are reported in Appendix I. Comparing GSCI with GRCI, we find that GSCI provides a higher average return and a better Sharpe ratio than GRCI, especially either in the recession period or during the restrictive monetary phase. This is in line with Erb and Harvey (2006), and largely reflects the trendwise (if uneven) rise of oil prices against other raw materials. One implication is that the less than stellar performance of actively trading and even buying & holding commodities would have held *a fortiori* if a rival index had been used for the analysis.

---

[2]For the stock index, the total return index includes the aggregate daily dividend into the incremental amount of the price index. For bonds, the total return index likewise includes the accrued interest into the value of bonds.

[3]In fact, we would have preferred to use S&P 500 in our analysis but we only have a total return index for S&P 500 from 1/1979 while we have such data for the MSCI index from 1/1973. In addition, we observe that in terms of average return and standard deviation the MSCI index is nearly the same as the S&P 500.

[4]We choose 6/1990 because it is the end of a business cycle and 20 years is enough for setting up an in-sample strategy.

The recession and boom periods are identified by the NBER. JJM, on the other hand, examine asset performance in two monetary policy phases (restrictive and expansive) using the Fed's discount rate to define phases. Thus, when the discount rate is raised a restrictive monetary policy is said to apply, while the reverse situation is referred to as an expansive monetary policy. A new monetary policy starts when the sign of the discount rate change shifts from negative to positive and via versa. Also following JJM (2002), we do not include in our sample the months in which monetary policy actually switches direction.

1.2 A suggested method to have *ex ante* stages of a new business cycle

As mentioned in the introduction, we have two problems in applying the implication of business cycle stages in asset allocation: (i) the actual starting date of a new business cycle is announced late and (ii) even if there were no delay in the announcement, the duration of a newly announced business cycle is still uncertain. Therefore, in order to exploit business-cycle stages for asset allocation, we need a way to define substages (like early boom and late boom) in an *ex ante* way, avoiding all hindsight, so that they can be used realistically in an investment strategy. Given that we use these *ex ante* in the out-of-sample period, the buy-or-sell rules obtained from the in-sample data should also be based on *ex ante* stages.

We propose to use the periodically updated historical average durations of each stage of booms. According to Moore (2002), there is a significant trend for recessions to become shorter and for booms to become longer. From 1857 to 1990, the average duration for a boom was 35 months. However from 1970 until 1990, this number had risen to 60 months. Because of this trend and also because the period from 1970 to 1990 is quite short, we suggest to use data from 1950 to calculate average durations for the in-sample booms. This period gives us a mean life of 50 months for a boom, implying 17, 17 and 16 months for early, middle and late stages, respectively. Investors are assumed to have these numbers in mind (until the next update, that is, when the full cycle is over). For example, if, *ex post*, a boom lasts just 25 months, *ex post*, people still expect a 50-month boom. So, when a recession is announced, they realize that what they thought to be a mid-boom is now immediately followed by a recession phase, and they never apply the allocation strategy for a late stage. If the boom lasts more than 50 months, in contrast, they continue in an unexpectedly long late-stage mode until the new business cycle starts.

Not only the duration of the cycle is a problem, though: we also have to handle the announcement lag. We do two adjustments whenever an announcement is made. First, we update the average durations of the boom substages and the recessions. Second, when there is an announcement, we subtract the delay in the announcement from the new *ex ante* stage. For example, if a boom is announced with a 10-month delay, and in this period the expected duration of an early boom is 17 months, the investor realizes that the *ex ante* early-boom stage has already lasted 10 months, implying there are only 7 months left for actual investment purposes. If the delay period is more than 17 months (for a boom) or 11 months (for a recession), the investor realizes that he/she missed an entire stage, and immediately shifts his/her portfolio to the next mode, without applying the early-boom or recession strategy.

1.3 Business cycles: perfect versus imperfect foresight

In the next section we derive the trading rule which did best in-sample. But first we look at actual announcement lags and durations, as well as at a cross-tabulation of the *ex post* and *ex ante* cyclical status of the in-sample months.

In Table 1 we report announcement dates, starting dates, announcement lag, *ex post* duration and, for the out-of-sample years, updated average duration, starting from 6/1980.[5] Remember that for the out-of-sample period, the average durations are updated when there is a new announcement, and the data used to calculate average durations start from 1950. Next to the updated average, we also show how each next boom was subdivided into its expected early, mid and late stage. In the last row, we show the maximum, minimum and standard deviation duration of booms and recessions in the period from 1950 until September 2009.

The table shows that the delay between announcement and starting dates of a cycle ranges from 5 to 20 months. Combining this fact with the short duration of a recession (11 months on average), the table illustrates the pointlessness of dividing recessions into sub-stages. This is especially true in the post-1990 test period, where each of the slowdowns is over, or essentially over, by the time it is announced. The long duration for a boom, in contrast, (with a long-run historic average of 50 months, a maximum of 120 months and a minimum of 12 months) leaves us ample room for sub-stages. The booms in the test period were unusually long, actually: following our rule, the investor would have shifted his portfolio to a late-boom mode by the time the recession is actually only half-way or not even half-way. But that, of course, is pure hindsight.

---

[5] This is the first date on which we have a record of the announcement date.

Table 1: Dates and duration (in months) of recessions and booms: 6/1980 - 9/2009

Announcement dates	Boom				Recession			
	Starting	Delay	Dur	Updated avg duration	Starting	Delay	Dur	Updated avg dur
06/1980					1/1980	5	7	
07/1981	08/1980	11	11					
01/1982					7/1981	6	17	
07/1983	12/1982	7	91					
04/1991				50 (17,17,16)	7/1990	9	9	11
12/1992	04/1991	20	119	50 (17,17,16)				11
11/2001				59 (20,20,19)	3/2001	8	9	11
07/2003	11/2001	19	73	59 (20,20,19)				10
12/2008				62 (21,21,20)	12/2007	12		10
1950-2009	Max	Min	Stdev		Max	Min	Stdev	
Duration	119	11	36.5		18	5	3.3	

Note: In this table, starting from 6/1980, we report announcement dates, starting dates, announcement lag, *ex post* duration and, for the out-of-sample years, updated average duration. For the out-of-sample period, the average durations are updated when there is a new announcement, and the data used to calculate average durations start from 1950. Next to the updates average we also show how each boom was subdivided into its expected early, mid and late stage. The last row shows the maximum, minimum and standard deviation of the durations of booms and recessions in the period from 1950 until September 2009.

In Table 2, we present a cross-tabulation of *ex post* and *ex ante* stages of booms. 122 of the 186 boom months are classified correctly—but remember that the announcement-lag issue is not yet affecting this comparison. Comparing the totals, the biggest difference occurs for the late stage. As, in this period, the booms have durations longer than the historic average duration of 50 months, the *ex post* stages occur later than the *ex ante* stages, and especially the late *ex ante* stages have far more observations than the late *ex post* stages.

In short, the joint issues of late announcements and unexpected lengths are quite material and could substantially affect the benefits that one can reap from any tactical asset allocation strategy based on the business cycle. We also conclude that, for the sake of realism, the in-sample analysis should use *ex ante* stages instead of *ex post* stages, since *ex ante* stages are what we use in the out-of-sample test. In the same logic, the in-sample analysis should take into account the announcement lags. The problem, however, is that the NBER website lists announcement dates only starting from 1980, halfway the first subperiod. Thus, we have little choice but to ignore the lag when devising the trading rule. For the more important part, the out-of-sample test, however, we do incorporate that imperfection.

Table 2: Cross table of *ex post* and *ex ante* stages of booms in the in-sample 1/1970 - 6/1990

	Ex post early	Ex post middle	Ex post late	Total
Ex ante early	46	9	3	58
Ex ante middle	14	25	9	48
Ex ante late	0	29	51	80
Total	60	63	63	186

Note: This table reports the cross comparison of *ex post* and *ex ante*s stages of booms in the in-sample 1/1970 to 6/1990. An *ex post* stage is based on pure hindsight: it lasts the *ex post* life of any recession, and one third of the *ex post* life of any boom.

2 A proposed trading rule from in-sample analysis

In this section, we use monthly data from 1/1970 to 6/1990 for the in-sample analysis to distill trading rules that can be applied to post-1990 data without any looking-ahead bias w.r.t. the length of the cycles. Relative to GR and JJM, another innovation is that we also consider a trading rule that is based on both the monetary-policy effect and the business-cycle effect: after all, these effects are not likely to be just additive.

2.1 Asset returns in the in-sample analysis

We use OLS to estimate various conditional mean returns and obtain significance tests for each asset a in different economic conditions:

$$R_{a.t} = \sum_{m=0}^{1} \sum_{s=0}^{3} \mu_{m,s,a} \, \mathbf{1}_{m,s,t}. \tag{1}$$

In the above, $R_{a.t}$ is return of asset a (including US stocks, international stocks, and GSCI, the Goldman-Sachs commodity index) in month t. As we only have data for US corporate bonds as of 1/1973, we do not include bonds at this stage. Subscript m stands for monetary policy: restrictive ('0') or expansive ('1'). Subscript s refers to a stage in the (full) cycle: we set $s = 0$ for a recession, and $s = 1$ to 3 for an early, middle and late boom. $\mathbf{1}_{m,s,t}$ is dummy variable: it is set equal to unity if month t is characterized by monetary regime m and cycle stage s. Otherwise, it is set equal to zero. There is no general intercept, so each coefficient μ is a pure mean return and each t-statistic is against a null of zero expected return.

In Table 3, we report the in-sample average returns, and their t-tests, of assets for various economic circumstances using *ex ante* stages. The 'Wald (1)' columns report the probabilities from the Wald tests on the significance of the differences among mean returns of three stages in booms. In the 'Wald (2)' row, we report the probabilities from the Wald tests on the significance

Table 3: In-sample monthly average returns (%) in different periods with *ex ante* stages: 1/1970 - 12/1985

	Restrictive policy					**Expansive policy**				
	Recession	**Boom**				**Recession**	**Boom**			
		Early	Middle	Late	Wald(1)		Early	Middle	Late	Wald(1)
Stock										
US	**-1.45**	**0.70**	**-0.27**	**0.93**	**0.62**	**2.88**	**1.66**	**0.84**	**2.93**	**0.26**
	(−1.74)	(0.52)	(−0.26)	(1.52)		(2.57)	(2.43)	(0.97)	(3.05)	
EAFE	**-1.81**	**-0.26**	**0.24**	**0.97**	**0.65**	**1.62**	**2.18**	**1.96**	**4.73**	**0.06**
	(−2.14)	(−0.19)	(0.22)	(1.56)		(1.42)	(3.14)	(2.25)	(4.81)	
Futures										
GSCI	**2.42**	**-1.51**	**3.41**	**1.85**	**0.03**	**0.00**	**0.85**	**0.65**	**1.47**	**0.84**
	(2.62)	(−1.02)	(2.89)	(2.71)		(0.00)	(1.12)	(0.68)	(1.37)	
GRCI	**1.21**	**-1.95**	**3.05**	**1.88**	**0.01**	**0.32**	**1.34**	**0.93**	**1.43**	**0.90**
	(1.53)	(−1.54)	(3.02)	(3.22)		(0.30)	(2.07)	(1.14)	(1.56)	
Wald (2)	**0.00**	**0.27**	**0.02**	**0.32**		**0.09**	**0.43**	**0.89**	**0.31**	
Obs. *ex ante*	31	12	19	57	119	17	46	29	23	115
Obs *ex post*	31	11	15	62		17	49	48	1	

Note: 1. The average return in each cell is the value of the coefficient corresponding to each condition. The numbers in brackets are t-statistics of the coefficient from the regression of each asset return on the system:

$$R_{a.t} = \sum_{m=0}^{1} \sum_{s=0}^{3} \mu_{m,s,a} \, \mathbf{1}_{m,s,t}. \tag{2}$$

In the above, $R_{a.t}$ is return of asset a (including US stocks, international stocks, commodity indices: GSCI and GRCI) in month t. As we only have data for US corporate bonds as of 1/1973, we do not include bonds in this test. m stands for monetary policy: restrictive ('0') or expansive ('1'); s refers to a stage in the (full) cycle: $s = 0$ for a recession, and $s = 1$ to 3 for an early, middle and late boom. $\mathbf{1}_{m,s,t}$ is dummy variable: $\mathbf{1}_{m,s,t} = 1$ when month t is under a monetary regime m and a cycle stage s. Otherwise, $\mathbf{1}_{m,s,t} = 0$. As we have 4 assets, and 8 economic conditions for each asset, we have 32 coefficients for this system.
2. The 'Wald (1)' columns report the probabilities from the Wald tests on the significance of the differences among mean returns of three stages in booms.
3. The 'Wald (2)' row reports the probabilities from the Wald tests on the significance of the differences among mean returns of GSCI against US stocks.

of the differences among mean returns of GSCI against US stocks. The table concludes with numbers of observations for each of the eight economic conditions ($m \times s$). Unlike in the calculation of the mean returns, here we show results for both *ex ante* and *ex post* cycle lengths.

Those bottom rows show how monetary policy varies across the stages of a boom. Restrictive episodes are about as prevalent as expansionary ones (119 against 115), but booms totally dominate slumps (243 months against 48). A restrictive policy is most often applied in a late-boom stage (62 times, out of 119 restrictive-policy months and out of 63 [actual] late-boom months). Interestingly, a restrictive policy is typically also continued in the subsequent recessions (31 times, out of 119 restrictive-policy months and out of 48 recession months). An expansive policy is followed in almost all cases of early- and middle-boom stages (97 times, out of 115 expansive-policy months and out of 123 early/midboom months).

In terms of mean returns, we first note that only 13 out of 32 t-statistics can more or less confidently confirm that the mean return in a condition is different from zero. In light of the small sub-samples, this should not be too surprising. The Wald (2) tests confirm that GSCI returns are statistically different from the US stocks returns in three of the eight scenarios: a recession regardless of monetary policy and a middle boom under a restrictive policy. Second, we also find three out of eight cases in which there are significant differences in the mean returns among the boom's sub-stages. All of these are encouraging us to try and find a way to use both monetary and business cycle stages for asset allocation.

Table 3 also shows that commodity futures indices only provide higher returns than stocks and bonds under a restrictive policy in recessions, in middle and late booms and not under an expansive policy. This result is different from GR (who only consider two stages) in two ways. First, also the middle stage is found to be attractive, under a restrictive policy, not just the late boom. Second, commodity futures do not always have higher returns than stocks in a late boom: this is patently valid for expansionary scenarios, but seems to be reversed when the Fed turns restrictive. Relative to JJM, we confirm that futures generally do better than stocks under restrictive scenarios, but early booms seem to be an exception. We also confirm that, in expansionary situations, stocks are the winners in relative terms, but we also note that futures still provide positive returns. Therefore, in such conditions, commodity futures are still attractive when the alternative is to sell, that is, in a stand-alone or trading-rule analysis.

2.2 A proposed trading rule

The general trading rule is to hold commodity when the expected commodity futures return is positive, and to sell otherwise. In light of the assets' returns that we get from the in-sample analysis, investors should go long in commodity futures (i) with a restrictive policy: in the middle and late stages of a boom and during the recession; and (ii) under an expansive policy: in a boom. Note that the latter case would have led to shorting under the JJM strategy, as futures returns, even though positive, are below stock returns. As in the introduction, we can also state the combined rule negatively: sell commodities in a recession if the Fed's policy is expansionary, and in an early boom if the Fed is still putting on the brakes.

The story makes some sense: raw materials are not much in demand when things are slow, that is, during the recession or early recovery. One might accordingly wonder how come such a simple effect is not anticipated by markets. One answer, of course, might be that we use hindsight: even the official turning points will not be known to the investor until about

Table 4: In-sample monthly returns, Sharpe ratios (%) for GSCI of three competing trading rules: 1/1970 - 6/1990

	Mean	Standard deviation	Sharpe ratio
GR strategy	1.64	5.26	0.67
JJM strategy	1.32	5.34	0.46
GR×JJM strategy	**1.68**	**5.14**	**0.72**

Note: In this table, the Sharpe ratio is the annualized excess mean return divided by the standard deviation, *i.e.* the ratio for monthly returns multiplied by $\sqrt{12}$. The trading rules are:

– GR: hold commodity futures in early recession and late boom, and short otherwise,

– JJM: hold commodity futures when a restrictive policy applies and sell when there is an expansive policy

– GR×JJM: sell commodities in early booms when the Fed is restrictive, and in recessions when the Fed is expansive; otherwise buy.

a year after the fact, so each bit of information or hint that becomes available in the mean time is a surprise to the market. In short, we are just non-randomly sampling from history, picking months that are later confirmed to be from bad times.[6] Another possible view is that the streamlined facts underlying the trading rule are just a coincidence. An out-of-sample application, including a statistical significance test, should shed some light on the latter view.

In order to illustrate the economic performance of the trading rule in-sample, we compare the mean returns and Sharpe ratios of GSCI from the combined strategy with the GR and JJM strategies. The three conditional strategies to compare are:

- the GR strategy, which only keeps track of the 2×2 business cycle condition: hold commodity futures in late-boom and early-recession episodes, and short them otherwise;

- the JJM strategy, based on the monetary policy condition regardless of the business cycle: hold commodity futures when a restrictive policy applies and sell otherwise;

- the GR×JJM strategy, which considers both the monetary policy and the business cycle conditions: sell commodities in early recoveries when the Fed is still restrictive, and in recessions when the Fed is expansive; otherwise buy.

The total return on a short position of a commodity futures index is calculated as the negative percentage change in the futures price index plus the 1-month T-bill rate. From

[6]In early recoveries, one could expect rebuilding of inventories, and hence rising prices. But remember that managers rebuild inventories only when they are pretty certain about a recovery. Here the recovery is not yet announced, and the Fed's policy is restrictive, which again is not likely to change the managers' views.

Table 4, it is clear that the GR×JJM strategy for GSCI results in the highest average return and Sharpe ratios compared with GR and JJM strategies. In addition, at 0.72, the GR×JJM rule's Sharpe ratio easily passes the 0.5 hurdle which makes traders sit up and take notice. However, the analysis assumes knowledge of the subsample means, the delay in business cycle announcement is not considered either and there is no statistical analysis. The next section takes care of that.

3 Out-of-sample results

In this section we want to find out, first and foremost, how well the rule does in realistic circumstances. A realistic scenario means that, of course, we can only adjust the investment decision as of the announcement date of a new phase (boom, slump). Also, the substages are *ex ante*, and the conditional mean returns are unknown as we are out of sample. But we would also like to know what the impact of the announcement lag is.[7] We therefore present results also under the assumption that there is no announcement lag, so that we adjust the portfolio at the actual starting date of this new cycle. We start with a cross-tabulation of months, which gives us a feel as to how many months are 'misclassified' because of late announcements. After that we turn to returns and risks.

3.1 Cross-tabulation of portfolio modes: with/without announcement lag and with/without information about length of phase.

In Panel A of Table 5, we compare the differences between actual vs. perceived business-cycle stages for the out-of-sample from 7/1990 to 8/2009 that can be traced to late announcements. In both cases, the investor subdivides the booms into substages using historical averages. We find that, predictably, the delay in announcements mostly affects the recessions and early booms. The really new information in the table is about the size of that impact. In 28 out of 38 months which are actually recessions, the investor thinks of the economy as being in a late boom. Likewise, no less than 36 out of the 37 months which are actually part of an *ex ante* early boom are still thought of as depressed. The delay in announcement has almost no effect in middle and late booms, although in three cases the investor still perceives the economy as in a recession while, if the tuning point had been announced already, these months would have

[7] The use of *ex ante* substages was also adopted in-sample, so this is not a potential cause of underperformance.

Table 5: Cross-tabulation of portfolio modes: with/without announcement lag and with/without information about length of phase. 7/1990 - 8/2009

	actual stage of business cycle				
perceived stage	recession	early boom	mid boom	late boom	Total
	Panel A: Ex ante stages for actual booms				
recession	10	36	3	0	49
early boom	0	1	0	0	1
middle boom	0	0	33	0	33
late boom	28	0	0	112	140
Total	38	37	36	112	223
	Panel B: Ex post stages for actual booms				
recession	10	39	0	0	49
early boom	0	1	0	0	1
middle boom	0	18	15	0	33
late boom	28	5	47	60	140
Total	38	63	62	60	223

Note: This table reports the cross comparison for the out-of-sample 7/1990 to 8/2009 of: (i) in Panel A, actual vs. realized recessions and actual, *ex ante* booms vs. realized, *ex ante* booms and (ii) in Panel B, actual vs. realized recessions and actual, *ex post* booms vs. realized, *ex ante* booms.

been classified as in a middle boom.

Lest there be an impression that most of the time the investor still gets it right, we remind the reader that late announcements are just part of the issue. In Panel A, the numbers we just discussed, *ex ante* rules were used to divide up booms into substages, and this in itself also creates misclassifications. In Panel B we therefore take the four stages as they are realistically perceived, and compare them to the perfect scenario, where all turning points are known beforehand and where, therefore, also the substages can be set at exact one-thirds of actual lengths. Obviously, the investor does not get it right most of the time. The sum of the diagonal numbers, the number of months that are diagnosed correctly, amounts to a mere 86 out of 223. The investor gets systematically wrong-footed by the unexpectedly long booms, so that when he/she thinks the state is mid or late boom, he/she actually is running ahead of reality more often than not (for mid booms: 18 misclassified, against 15 on the diagonal; for late booms, 28+5+47=80 misclassified, against 60 on the diagonal).

In the next two subsections, we examine whether the combined effect of the late announcements and out-of-sample averages wipes out the benefits of our flexible asset allocation. We first consider stand-alone trading rules and then mean-variance optimal portfolios.

3.2 Performance of trading rules, stand-alone

As mentioned before, the out-of-sample period is from 7/1990 to 8/2009.[8] In this subsection, we consider the performance of the three competing commodity trading rules. We also compare them to the unconditional mean returns and Sharpe ratios for stocks, bonds and GSCI, that is, the results from a buy-and-hold policy.

In Table 6, we report these numbers for US investors.[9] It is clear from the table that if, magically, the NBER could identify turning points immediately—see the numbers in the 'ignoring lags' columns—the GR×JJM strategy for GSCI would perform better than the GR and JJM trading rules in terms of higher mean returns and Sharpe ratios. Still, under this assumption the GR×JJM rule also provides higher average returns and a better Sharpe ratio than a buy-and-hold strategy for commodities and even for stocks. However, when taking into account the late announcement dates (see the columns labeled 'realistic') the benefits are wiped out: the GR×JJM rule for GSCI now provides a worse return and Sharpe ratio than JJM and than buy-and-hold strategies for stocks, bonds and even the GSCI itself. The active rule that does best, realistically, is the JJM one, where announcement dates and cycle lengths are not an issue. However, even this one hardly improves on a buy-and-hold strategy for commodities and is beaten by simply holding stocks and bonds. The GR rule, lastly, provides negative returns during the test period.

In short, due to late announcements of turning points in business cycles, the above rules do not provide any benefits in timing the commodity market and they do not make raw materials more attractive than stocks and bonds. More generally, the familiar message is that in-sample results are not necessarily a reliable basis for investments, especially if they also ignore announcement lags.

All the above findings are based on just the values of returns and Sharpe ratios, ignoring significance issues. We now apply the Diebold and Mariano (1995) (DM) statistical test to compare the raw returns from every pair of strategies, thus testing for equality of expected returns. We also apply DM to the standardized excess returns (excess returns divided by

[8] We also test for the period ending before the current crisis (7/1990 to 12/2007). We obtain the same results as with our reported period.

[9] We also did the test for British and Japanese investors and we observe the same results.

Table 6: Stand-alone trading: out-of-sample returns, standard deviations and Sharpe ratios (%): 7/1990 - 9/2009

	Unconditional				**Conditional for GSCI**				
	Stock	EFEA	Bond	GSCI	JJM	Ignoring lags		Realistic	
						GR	GR×JJM	GR	GR×JJM
Return	0.70	0.54	0.70	0.62	0.73	0.76	**1.19**	0.05	**0.33**
Stdv	4.41	5.09	2.60	6.45	6.37	6.40	6.32	6.45	6.45
Sharpe-ratio	0.31	0.16	0.52	0.17	0.23	0.25	**0.48**	−0.14	**0.01**

Note: In this table, Sharpe ratio is the annualized excess mean return divided by the standard deviation, *i.e.* the ratio for monthly returns multiplied by $\sqrt{12}$. The trading rules are:

- GR: hold commodity futures in early recession and late boom, and short otherwise,
- JJM: hold commodity futures when a restrictive policy applies and sell when there is an expansive policy
- GR×JJM: sell commodities in early booms when the Fed is restrictive, and in recessions when the Fed is expansive; otherwise buy.

standard deviation), thus testing for equality of the Sharpe ratios.[10] We report the DM test results in Table 7. As the GR strategy in the 'realistic' column performs worse than in the 'ignoring lags' column, in order to make the table simple we will not undertake the test for this strategy. The table shows that there is basically no significant difference between average returns and Sharpe ratios of different assets and of different strategies for buying and selling commodities.

Results until now show that we cannot earn significantly higher returns and Sharpe ratios for commodity futures compared to stocks and bonds when we try to time the market following monetary-policy switches or business-cycle news or both of them. Methodologically, the message is that, next to out-of-sample testing under realistic assumptions on information, also significance tests should be considered. Still, all the above is about asset-by-asset trading rules, ignoring the conditional differences of returns across assets and their variances and correlations. In the next subsection we therefore examine mean-variance optimal portfolios with and without commodities, and with or without conditioning the input data on monetary policy and/or the business cycle.

3.3 Performance of the optimal portfolio

The contending investment policies in this race differ from each other in two dimensions. First, the strategies disagree as to how investors should condition the estimated parameters on the

[10] Please see Appendix 2 for more details about the Diebold and Mariano (1995) test.

Table 7: DM test results for returns and standardized returns for buy-and-hold assets: 7/1990 - 9/2009

	Returns			**Standardized returns**		
	Coefs	t-stats	p-value	Coefs	t-stats	p-value
Stock vs. Bond	0.000	0.00	1.00	−0.06	−0.69	0.49
Stock vs. GSCI	0.001	0.15	0.88	0.04	0.41	0.68
Stock vs. JJM	0.000	−0.05	0.96	0.02	0.17	0.86
Stock vs. GR (no lag)	−0.001	−0.09	0.93	0.02	0.13	0.90
Stock vs. GR×JJM (no lag)	−0.005	−0.69	0.49	−0.05	−0.39	0.70
Stock vs. GR×JJM (realistic)	0.004	0.67	0.50	0.09	0.86	0.39
Bond vs. GSCI	0.001	0.14	0.89	0.10	0.94	0.35
Bond vs. JJM	0.000	−0.06	0.95	0.08	0.87	0.38
Bond vs. GR (no lag)	−0.001	−0.12	0.91	0.08	0.75	0.45
Bond vs. GR×JJM (no lag)	−0.005	−0.93	0.35	0.01	0.11	0.92
Bond vs. GR×JJM (realistic)	0.004	0.68	0.49	0.15	1.38	0.17
GSCI vs. JJM	−0.001	−0.13	0.90	−0.02	−0.14	0.89
GSCI vs. GR (no lag)	−0.001	−0.20	0.84	−0.02	−0.20	0.84
GSCI vs. GR×JJM (no lag)	−0.006	−0.80	0.43	−0.09	−0.81	0.42
GSCI vs. GR×JJM (realistic)	0.003	0.84	0.40	0.04	0.84	0.40
GR (no lag) vs. JJM	0.000	0.08	0.94	0.00	0.07	0.94
GR (no lag) vs. GR×JJM (no lag)	−0.004	−1.24	0.21	−0.07	−1.27	0.21
GR (no lag) vs. GR×JJM (realistic)	0.004	0.76	0.45	0.07	0.76	0.45
JJM vs. GR×JJM (no lag)	−0.005	−0.99	0.32	−0.07	−1.00	0.32
JJM vs. GR×JJM (realistic)	0.004	0.56	0.58	0.06	0.56	0.58
GR×JJM (no lag) vs. GR×JJM (realistic)	0.009	1.37	0.17	0.14	1.38	0.17

Note: This table reports the DM test for returns and standardized return of pairwise assets and strategies for GSCI. The trading rules are:

- GR: hold commodity futures in early recession and late boom, and short otherwise,
- JJM: hold commodity futures when a restrictive policy applies and sell when there is an expansive policy
- GR×JJM: sell commodities in early booms when the Fed is restrictive, and in recessions when the Fed is expansive; otherwise buy.

The coefficients (coefs), t-statistics (t-stats) and p-values are from the following regressions:

$$\mathrm{d}R_{a,b} = \alpha + \varepsilon$$

with either $\mathrm{d}R_{a,b} = R_a - R_b$, in the expected return test, or $\mathrm{d}R_{a,b} = (R_a - R_0)/\sigma_a - (R_b - R_0)/\sigma_b$, in the Sharpe ratio test, where $\sigma_.$ is the standard deviation of the returns on a or b.
In the first column, we report the asset pair that is being compared. The first asset is the asset a in above equation and the second asset is asset b. GR, JJM, 'GR×JJM (no lag)' and 'GR×JJM (realistic)' refer to the GR, JJM and GR×JJM strategies where the portfolio is switched at the actual turning point ('no lag') or at the announcement date ('realistic').

state of the economy. There are four strategies that we will compare in this respect: (i) the 'unconditional' strategy, where we consider neither business cycle nor monetary policy in setting up optimal portfolios; (ii) a modified GR strategy, similar to the original one in that it considers business-cycle information, except that it works with just one recession stage and defines the boom substage in an *ex ante* way;[11] (iii) the JJM strategy, looking at changes in interest rates; and (iv) the GR×JJM strategy, combining both.

The second way in which the contending strategies differ is w.r.t. the presence or absence of commodities. That is, beside considering the performance of different conditioning rules, we also want to compare the performance of optimal portfolios with and without commodity futures to check for the usefulness of including commodity futures in diversified portfolios *per se*.

We apply the mean-variance (Markowitz) framework to construct the optimal portfolio without any constraint.[12] The benchmark risky portfolio for investors includes the US stock index, international stocks (EAFE), and the US bond index. Familiarly, in a mean-variance framework the optimal weights are characterized by

$$\mathbf{W}_t \propto \Omega_t^{-1}\mathbf{E}_t, \tag{3}$$

with $\mathbf{W}$ a 4×1 vector of assets' weights, $\mathbf{E}$ a 4×1 vector of assets' excess returns and Ω a 4×4 variance/covariance matric. The 1-month T-bill rate is used as a risk-free asset. The expected excess returns and the variance/covariance matrices are estimated from data available at time t. We update the expected mean vectors and covariance matrices each time a turning point is announced by NBER. For the unconditional and JJM strategies, we use all data available at the new announcement date to calculate updated mean vectors and covariance matrices. For the modified GR and GR×JJM strategies, in contrast, even though we update the means vector and covariance matrix on each announcement date, we only use data until the newly announced turning point rather than until the announcement date.

To reduce the number of conditions for the GR×JJM strategy, where in principle eight sets of estimates should be used, we use the in-sample results of the pre-test period and merge some stages, so as to end with six scenarios. Specifically, for the three substages of booms under an expansive monetary policy, there are too few observations, and given an uptick in the latest interest-rate change, the trading rule did not suggest any portfolio switch during a boom anyway. So we group these three scenarios into one. Thus, our six states of the world are (i) restrictive policies in recessions; (ii-iv) restrictive policies in different stages (early, middle, late) of booms; (v) expansive policies in recessions; and (vi) expansive policies in booms. In

[11] The original GR rule, based on $2 + 2$ cycle stages, is not useful when announcement lags are taken into consideration: recessions are too short and are announced too late for subdivisions to make sense.

[12] We allow short-sales and borrowing and lending at the risk-free rate.

practice, we estimate mean vectors and covariance matrices for these six conditions, using *ex ante* stages, and we update them when a new turning point is announced. We need to estimate four mean vectors and covariance matrices at each moment in time for the modified GR strategy, i.e: recession, early boom, middle boom and late boom. For the unconditional strategy, we only need to estimate and update one mean vector and covariance matrix for the whole sample while for JJM strategy at each moment of updating, we need to estimate two vectors and covariance matrices for restrictive and expansive policies.

In principle, the right-hand side of Equation (3) describes the log-utility portfolio, and, in the absence of an estimation risk, this would be a good portfolio to study as it would incorporate the investor's reaction to any changes in risk over time. The alternative is to study the tangency portfolio, that is, to rescale all weights to get a unit sum, reflecting a portfolio fully invested in risky assets. This second solution is appealing when parameter estimates are noisy and when, as a result, many first-pass-estimated weights are even individually outside the range [0,1]: then the rescaling also shrinks the estimation errors.[13] All this is non-controversial when none of the weights are negative, as they should be in a homogenous-expectations mean-variance world. But we are not using consensus parameters, to name just one issue. The question then is what to do when some weights are negative.

DeMiguel, Nogales and Uppal (2010) rescale by the sum of the positive weights, but this sometimes leaves us with extravagant negative weights or sums of negative weights. The problem, and our proposed procedure, is illustrated in Panel A of Table 8. Most of the calculated weights have values outside the range [0,1]. The first scenario, for example, suggests big negative weights for US and foreign stocks (weights −1.7, −6.0 and −2.5, summing to −10.2) and a hefty 2.9 weight for raw materials. (Further down the table we even see numbers like −55 and 65.) Rescaling by 2.9 would still leave us with one weight well below −1, and a sum of negative weights equal to $-10.2/2.9 = -3.5$. To avoid situations such as this, we rescale the first-stage weights in such a way that either the short side or the long side sums to unity, and the other side below unity. In the above example, for instance, we would rescale by 10.2 (the sum of the negative weights, up to the sign) not 2.9 (the sum of the positive weights). In general, we first add weights of the same sign together. To get the adjusted weights, we

[13] Alternatives are to shrink the means towards the general mean, see Jorion (1986), and/or to shrink the variance-covariance matrix towards the Lintner version, the beta-based matrix sometimes used in Value at Risk, see Ledoit and Wolf (2003). These solutions are, to our knowledge, untested on assets as heterogenous as ours (stocks, bonds, commodities).

then divide each of the first-pass weights by the maximum absolute value of these two partial sums. The weight for the risk-free asset is equal to one minus the sum of the adjusted weights of all risky assets in the portfolio. The adjusted optimal weights are reported in Panel B of the same table. Considering the signs of optimal weights, we find that, for most of the conditions, they are consistent with the GR×JJM tactical allocation.[14]

In Table 9, we report the performance of out-of-sample optimal portfolios for different strategies. The results in this test show that the optimal portfolio based on GR×JJM-style conditioned estimates offers a higher return and Sharpe ratio than either the unconditional policy or the rival conditional strategies even when the late announcement is taken into account. This finding is more positive than our conclusion from the trading rules in the preceding subsection. Prima facie, therefore, redirecting funds across assets taking into account relative expected returns and risk does help, over and above what can be achieved on the basis of changes in expected return over time, and the best conditioning policy is to look at both the Fed and the NBER. The numbers also suggest that the inclusion of a commodity futures index into a portfolio brings benefits for the investors. Comparing the performances of optimal portfolios with the performances of stand-alone assets in Table 6, we find that the optimally diversified portfolio following the GR×JJM strategy has a higher return and Sharpe ratio than any stand-alone buy-and-hold policy. This is as it should be: diversification does help. But note that, under the rival flexible approaches, GR and JJM, diversified portfolios containing commodities, do not even beat a pure bond portfolio (compare Table 6 and Table 9). In short, the GR×JJM strategy does seem to add value, this time.

However, all the above ignores significance issues, so we apply the DM test to check whether or not the benefits are statistically meaningful. From Table 10 we conclude that there is no significant difference between the GR×JJM strategy and any of its rivals, whether unconditional or conditional. More soberingly, even diversified portfolios with commodity futures do not provide significantly higher returns and Sharpe ratios than portfolios without commodity futures.

In general, then, the picture seems to be that the inclusion of commodity futures into diversified portfolios, with and without conditioning of the parameters, may not really create

---

[14] We can provide the optimal weights for other strategies upon request.

Table 8: Optimal weights in the mean-variance portfolio from the GR×JJM strategy

Conditions	Panel A: Original optimal weights				Panel B: Adjusted optimal weights			
	Stock	EAFE	**Bonds**	GSCI	**Stock**	EAFE	**Bonds**	GSCI
Restrictive in recession								
01/70-04/91	−1.7	−6.0	−2.5	2.9	−0.17	−0.59	−0.24	0.29
01/70-12/92	−3.0	−3.1	−1.9	2.1	−0.37	−0.39	−0.24	0.26
01/70-11/01	−3.0	−3.1	−1.9	2.1	−0.37	−0.39	−0.24	0.26
01/70-07/03	−3.0	−3.1	−1.9	2.1	−0.37	−0.39	−0.24	0.26
01/70-12/08	−3.0	−3.1	−1.9	2.1	−0.37	−0.39	−0.24	0.26
Restrictive in early boom								
01/70-04/91	32.7	−55.6	60.1	65.9	0.21	−0.35	0.38	0.42
01/70-12/92	32.7	−55.6	60.1	65.9	0.21	−0.35	0.38	0.42
01/70-11/01	20.5	−41.9	47.0	56.1	0.17	−0.34	0.38	0.45
01/70-07/03	20.5	−41.9	47.0	56.1	0.17	−0.34	0.38	0.45
01/70-12/08	20.5	−41.9	47.0	56.1	0.17	−0.34	0.38	0.45
Restrictive in middle boom								
01/70-04/91	−12.4	4.2	−26.3	4.5	−0.32	0.11	−0.68	0.12
01/70-12/92	−12.4	4.2	−26.3	4.5	−0.32	0.11	−0.68	0.12
01/70-11/01	−1.5	1.7	10.6	3.7	−0.10	0.11	0.66	0.23
01/70-07/03	−1.5	1.7	10.6	3.7	−0.10	0.11	0.66	0.23
01/70-12/08	−2.3	6.6	13.4	2.2	−0.10	0.30	0.60	0.10
Restrictive in late boom								
01/70-04/91	4.2	0.3	−5.7	4.0	0.50	0.04	−0.66	0.46
01/70-12/92	4.2	0.3	−5.7	4.0	0.50	0.04	−0.66	0.46
01/70-11/01	1.9	0.1	−2.0	5.1	0.26	0.02	−0.29	0.72
01/70-07/03	1.9	0.1	−2.0	5.1	0.26	0.02	−0.29	0.72
01/70-12/08	2.1	1.6	−2.8	3.8	0.28	0.22	−0.37	0.50
Expansive in recession								
01/70-04/91	7.9	−3.3	4.2	−0.9	0.65	−0.27	0.35	−0.07
01/70-12/92	11.0	−3.8	3.5	−1.2	0.76	−0.27	0.24	−0.08
01/70-11/01	8.8	−4.8	4.6	−2.7	0.66	−0.36	0.34	−0.21
01/70-07/03	8.4	−6.7	6.9	−5.5	0.55	−0.44	0.45	−0.36
01/70-12/08	8.3	−6.9	7.4	−4.7	0.53	−0.44	0.47	−0.30
Expansive in boom								
01/70-04/91	4.3	10.6	−1.3	0.7	0.28	0.68	−0.08	0.04
01/70-12/92	4.3	10.6	−1.3	0.7	0.28	0.68	−0.08	0.04
01/70-11/01	4.1	5.3	0.6	−0.9	0.41	0.53	0.06	−0.09
01/70-07/03	4.1	5.3	0.6	−0.9	0.41	0.53	0.06	−0.09
01/70-12/08	2.1	4.8	4.0	1.3	0.17	0.39	0.33	0.11

Note:
1. In Panel A of this table, we report the optimal value weight in the efficient frontiers of each asset for six conditions in each period. Efficient frontiers are constructed following the mean-variance optimal portfolio (Markowitz model) without any constraint. The risky portfolio includes US stocks, international stocks (EAFE), US corporate bonds and commodity futures index (GSCI).
2. In Panel B, we report the adjusted weights which are calculated as follows:

$$w_a = \frac{w}{m} \tag{4}$$

with m the maximum absolute values between the sum of negative weights and the sum of positive weights.
3. Our anticipated business cycle stages are applied in this analysis.

Table 9: Out-of-sample optimal portfolios' returns, standard deviations and Sharpe ratios: 7/1990 - 9/2009

	Unconditional		JJM		**modified** GR		GR×JJM	
	With GSCI	Without GSCI	With GSCI	Without GSCI	With GSCI	Without GSCI	With GSCI	Without GSCI
Return	**0.58**	**0.55**	**0.62**	**0.30**	**0.56**	**0.45**	**0.80**	**0.41**
Standard deviation	3.10	3.47	4.71	3.37	4.10	4.37	4.27	3.81
Sharpe-ratio	**0.30**	**0.24**	**0.23**	**-0.01**	**0.21**	**0.11**	**0.40**	**0.09**

Note: In this table, Sharpe ratio is the annualized excess mean return divided by the standard deviation, *i.e.* the ratio for monthly returns multiplied by $\sqrt{12}$. The trading rules are:

- Unconditional: where we consider neither business cycle nor monetary policy in setting up optimal portfolios,
- JJM: looking at changes in interest rates,
- modified GR: similar to the original in that it considers business-cycle information, except that it works with just one recession stage and defines the boom 3 substage in an *ex ante* way,
- GR×JJM: combining both business cycle and monetary policy in setting up optimal portfolios.

significant benefits, even though the omens from the portfolio approach are definitely better than those from the trading rules. There is, of course, an issue of statistical power, in which case the investor's priors should also enter the picture. We return to this in the concluding section.

4 Conclusion

A relatively new branch of literature argues that the commodity futures index provides diversification benefits for investment portfolios but that expected gains from including long commodity futures into a diversified portfolio depend on either the stage of business cycle (Gorton and Rouwenhorst, 2005—GR) or monetary policy (Jensen *et al*, 2002—JJM). In practice, though, a conflict arises when one tries to apply both GR and JJM rules for asset allocation: in recessions, for instance, *i.e.* when activity is down, monetary policy often is (and should be) expansive, and then the two rules contradict each other. Another issue with these recent studies is that they are all about hindsight: they make no attempt to either test out-of-sample or to work with rules that take into account the lag in the NBER's identification of turning points and the uncertainty of the length of the ongoing boom or recession. Lastly, the above studies contain no test as to whether the returns they promise are significantly different from static alternatives (policies that ignore interest rates and business cycles) or from returns offered by other assets like stocks and bonds, whether taken separately or combined into portfolios.

First, in our point of view, the fact that both the GR and JJM rules seem to work despite the

Table 10: DM test results for returns and standardized returns for optimal portfolios: 7/1990 - 9/2009

	Returns			**Standardized returns**		
	Coefs	t-stats	p-value	Coefs	t-stats	p-value
Unconditional vs.						
without commodities	0.0003	0.13	0.90	0.0172	0.27	0.79
JJM strategy	−0.0004	−0.15	0.88	0.0221	0.36	0.72
modified GR	0.0002	0.09	0.93	0.0268	0.42	0.68
GR×JJM strategy	−0.0022	−0.95	0.34	−0.0271	−0.44	0.66
JJM **vs.**						
without commodities	0.0031	1.51	0.13	0.0669	1.38	0.17
modified GR	0.0006	0.16	0.87	0.0047	0.06	0.96
GR×JJM strategy	−0.0018	−0.99	0.32	−0.0492	−1.20	0.23
modified GR **vs.**						
without commodities	0.0011	0.51	0.61	0.0286	0.57	0.57
GR×JJM	−0.0024	−0.89	0.38	−0.0539	−0.83	0.41
GR×JJM **strategy vs.**						
without commodities	0.0039	1.49	0.14	0.0878	1.36	0.18

Note: This table reports the DM test for returns and standardized returns of pairwise assets and strategies for GSCI. The trading rules are:

- Unconditional: where we consider neither business cycle nor monetary policy in setting up optimal portfolios,
- JJM: looking at changes in interest rates,
- modified GR: similar to the original in that it considers business-cycle information, except that it works with just one recession stage and defines the boom 3 substage in an *ex ante* way,
- GR×JJM: combining both business cycle and monetary policy in setting up optimal portfolios.

The coefficients (coefs), t-statistics (t-stats) and p-values are from the following regressions:

$$\mathrm{d}R_{a,b} = \alpha + \varepsilon$$

with either $\mathrm{d}R_{a,b} = R_a - R_b$, in the expected return test, or $\mathrm{d}R_{a,b} = (R_a - R_0)/\sigma_a - (R_b - R_0)/\sigma_b$, in the Sharpe ratio test, where $\sigma_.$ is the standard deviation of the returns on a or b.
In the first column, we report the asset pair that is being compared. The first asset is asset a in the above equation and the second asset is asset b.

frequent conflict between their recommendations, this may mean that there is an interaction effect between the two effects on assets' returns: when the rules contradict each other, at least one of the two effects may be weaker than on average. Thus, the first issue in this paper is to try to sort out the interaction effect between these two economic conditions and to come up with a rule that considers both of them at the same time. Second, because the lag in the announcement of a new business cycle is quite long (at least 5 months), it may very well be that any benefit from timing one's asset allocation is wiped out when, realistically, the investor loses his/her foresight about what the NBER will proclaim in six or nine months. Third, an investor does not have the knowledge as to when exactly the current phase is half-way its total *ex post* length. Therefore, we also need a way to define *ex ante* stages of business cycles so that we can actually use them in asset allocation. Fourth, what works well in-sample almost always does more poorly out-of-sample, so the question is whether anything is left at all when data snooping is excluded. Lastly, we need to know whether any surviving gains may have arisen just because of chance, that is, whether these gains are significant.

We find that clever timing in asset allocation, taking into account both business cycle and monetary policy, does perform better than either the GR or JJM strategies. This is true in-sample, and also true out-of-sample and significant, if one ignores announcement lags. In short, there is indeed an interaction effect between the effects of the two economic variables. However, the benefits from clever timing asset allocation seem much reduced by the late announcement of business cycles. When stand-alone trading in commodity futures is evaluated, actually, the strategies even underperform buy and hold. When optimal portfolios are constructed, the news is not that negative: the combined allocation rule does offer, in the test sample, a higher mean return and Sharpe ratio than the unconditional timing strategy and the GR and JJM alternatives. Still, even then the returns or standardized excess returns are not significantly better. We even find that static diversified portfolios with commodity futures do not have significantly higher returns and Sharpe ratios than static portfolios without commodity futures.

In general, then, the tests tell us that clever timing of commodity markets, or flexible policies about including commodity futures into diversified portfolios, may not really create significant benefits. Since the statistical power may be low, one can bring up priors. The orthodox prior would be that commodity futures, being a financial product priced in open markets, should offer a reasonable return as part of a well-diversified portfolio. The portfolio tests in this paper confirm that view, albeit only weakly so (in the sense that the benefits are not statistically irrefutable). The orthodox view would also be that futures markets should take into account very obvious information like the sign of the Fed's most recent interest-rate change and the nature of, and time since, the NBER's most recent turning-point announcement. The trading rule tests confirm this. Actually, there is some evidence that timing may be counterproductive, although that conclusion is again not clear statistically. In that light, the conclusion may be that commodities, rather than being a neglected and frequently mispriced asset class, are investments like other assets rather than turnpikes to untold wealth. The orthodox view would also suggest that, as their market weight (including cash positions) is tiny, commodities should be a small part of a diversified portfolio only.

References

Bjornson B. and C.A. Carter (1997). "New Evidence on Agricultural Commodity Return Performance under Time-Varying Risk." *American Journal of Agricultural Economics.* 79: 918-930.

DeMiguel, V., F.J. Nogales and R. Uppal (2010). "Stock Return Serial Dependence and Out-of-Sample Portfolio Performance." Working paper.

Erb, C.B. and C.R. Harvey (2006). "The Tactical and Strategic Value of Commodity Futures." Duke University Working Paper, pp 61.

Errunza, V., K. Hogan and M.W. Hung (1999). "Can the Gains from International Diversification Be Achieved without Investing Abroad?" *The Journal of Finance.* 54(6): 2075-2107.

Frankel, J.A. (2006). "The Effect of Monetary Policy on Real Commodity Prices." *Asset Prices and Monetary Policy.* John Campbell, ed., University of Chicago Press.

Fama, E.F. (1982). "Inflation, Output and Money." *The Journal of Business.* 55(2): 201-231.

Greer, R.J. (1978). "Conservative Commodities: A Key Inflation Hedge." *The Journal of Portfolio Management.* Winter: 86-91.

Greer, R.J. (1994). "Methods for Institutional Investment in Commodity Futures." *The Journal of Derivatives.* Winter: 28-36.

Greer, R.J. (2000). "The Nature of Commodity Index Returns." *The Journal of Alternative Investments.* Summer: 45-53.

Gorton, G. and K.G. Rounwenhorst (2005). "Facts and Fantasies about Commodity Futures." *Financial Analysts Journal.* 62(2): 47-68

Gorton, G. and K.G. Rounwenhorst (2006). "A Note on Erb and Harvey (2005)." Yale ICF Working Paper no. 06-02, January.

Gorton, G. and K.G. Rounwenhorst (2006). "Are Commodity Futures too Risky for Your Portfolio? Hogwash!" Knowledge Wharton.

Hess, D., H. Huang and A. Niessen (2008). "How Do Commodity Futures Respond to Macroeconomic News?" *Financial Market Portfolio Management.* 22: 127-146.

Huberman G. and S. Kandel (1987). "Mean-Variance Spanning." *The Journal of Finance.* 42(4): 873-888.

Ibbotson Associates (2006). "Strategic Asset Allocation and Commodity." pp 57.

Jensen, G., R.R. Johnson and J.M. Mercer (2000). "Efficient Use of Commodity Futures in Diversified Portfolio." *The Journal of Futures Markets.* 20(5): 489-506.

Jensen, G., R.R. Johnson and J.M. Mercer (2002). "Tactical Asset Allocation and Commodity Futures." *The Journal of Portfolio Management.* Summer: 100-110.

Jorion, P. (1986). "Bayes Stein estimation for portfolio analysis." *Journal of Financial and Quantitative Analysis.* 21: 279-292.

Kaplan, P. and S.L. Lummer (1998). "Update: GSCI Collateralized Futures as a Hedging Diversification Tool for Institutional Portfolios." *The Journal of Investing.* 7(4):11-18.

Laws, J. and J. Thompson (2007). "Portfolio Diversification and Commodity Futures." Liverpool Business School Working Paper.

Ledoit, O. and M. Wolf (2003). "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Empirical Finance.* 10: 603-621.

Lummer, S.L. and L.B. Siegel (1993). "GSCI Collateralized Futures: A Hedging and Diversification Tool for Institutional Portfolio." *The Journal of Investing.* Summer: 75-82.

Markowitz, H.M. (1959). Portfolio Selection: Efficient Diversification of Investments. John Wiley, New York.

Roache, S.K. (2008). "Commodities and Market Price of Risk." IMF Working Paper WP/08/221.

Weiser S. (2003). "The Strategic Case for Commodities in Portfolio Diversification." *Commodities Now.* 7-11.

Appendix I: Results for GR sample

In this Appendix, we replicate the GR and JJM (2002) conclusions and confirm the conflict in their findings. We lastly test our suggestion for their sample.

Empirical results following Jensen et al. (2002) and Gorton et al. (2005)

In this section, the sample analyzed is the GR sample from 4/1973 to 12/2007. Following GR, we consider the performance of assets in different stages of business cycles. The recession and boom periods are identified by the NBER and are then divided into early and late phases.

The descriptive statistics 1973-2007 in Table 11 confirm both the conclusions of Gorton and Rouwenhorst (2005) and Jensen et al. (2002): (1) when considering the whole sample, the stand-alone commodity futures indices do not perform really better than stocks; (2) commodity futures indices do much better than the other assets in the restrictive monetary phase but perform worse than stocks in the expansive monetary phase; (3) commodity futures provide higher returns than either stocks and bonds in the early stage of recessions and in the late stage of booms.

Applying the Gorton and Rouwenhorst conclusions to the current crisis, which according to NBER started from 12/2007, investors should have held commodity futures in their diversified portfolio from 12/2007. However, on the other hand, the Fed has been continuously reducing its discount rate— thus applying an expansive monetary policy—since the crisis. Therefore, following JJM's findings, commodity futures should not be included in the diversified portfolio. Consequently, these recommendations create a conflict.

Dividing a business cycle into three instead of two stages:

In this subsection, we test whether our three-stage way is better than a two-stage way. In this section, to obtain stages for each cycle, we divide a business cycle duration into 3. We use OLS to estimate the following system in order to obtain the mean return and the significance tests

Table 11: Monthly returns (%), standard deviations and Sharpe ratios: 4/1973 - 12/2007

	US stock	EAFE	Bonds	GSCI	GRCI
	Whole sample				
Mean return	0.95	0.96	0.71	1.07	0.99
Standard deviation	4.45	4.84	2.61	5.63	3.71
Sharpe ratio	0.36	0.33	0.30	0.36	0.47
	Restrictive monetary policy				
Mean return	0.46	0.39	0.53	1.87	1.35
Standard deviation	4.55	4.81	2.74	6.26	4.41
Sharpe ratio	−0.09	−0.13	−0.07	0.72	0.61
	Expansive monetary policy				
Mean return	1.36	1.42	0.87	0.42	0.70
Standard deviation	4.34	4.83	2.49	4.97	2.99
Sharpe ratio	0.75	0.72	0.63	0.00	0.33
	Recession business cycle				
Mean return	0.15	−0.75	0.80	1.21	0.53
Early	−1.62	−2.21	−0.20	2.44	1.04
Late	2.51	1.19	2.12	−0.44	−0.15
Standard deviation	6.18	6.81	4.08	8.12	5.51
Sharpe ratio	−0.29	−0.73	0.10	0.23	−0.09
	Boom business cycle				
Mean return	1.08	1.23	0.70	1.05	1.07
Early	1.06	1.40	0.96	0.65	0.69
Late	1.11	1.07	0.44	1.44	1.43
Standard deviation	4.10	4.39	2.30	5.13	3.33
Sharpe ratio	0.53	0.61	0.36	0.40	0.63

Note: In this table, Sharpe ratio is the annualized excess mean return divided by the standard deviation.

for each asset in different economic conditions:

$$R_{a.t} = \sum_{b=0}^{1} \sum_{s=1}^{3} \delta_{b,s,a} \mathbf{1}_{b,s,t} \tag{5}$$

$R_{a.t}$ is return of asset a (including US stocks, international stocks, US corporate bonds, commodity indices: GSCI and GRCI) in month t; s is a stage of a boom/recession: $s = 1$ to 3 which stands for an early, middle and late stage; b stands for business cycle: recession ('0') or boom ('1'); $\mathbf{1}_{b,s,t}$ is dummy variable: it is equal to unity if month t is characterized by business cycle b and stage s; otherwise it is set equal to zero.

In Table 12, we report the average returns of assets in different stages of recessions and booms. When dividing the business cycle into three stages, we find that commodity futures indices provide higher returns than stocks and bonds in the middle and late boom and in the early and middle recession. This result is different from the findings of Gorton and Rouwenhorst

Table 12: Monthly returns (%) in different stages of business cycles: 4/1973 - 12/2007

	Recession period				Boom period			
	Early	Middle	Late	Wald(1)	Early	Middle	Late	Wald(1)
Stock								
US	**-1.56**	**-1.58**	**4.10**	**0.00**	**1.01**	**1.24**	**1.01**	**0.90**
	(−1.60)	(−1.59)	(3.88)		(2.47)	(3.01)	(2.53)	
International	**-2.41**	**-2.31**	**2.94**	**0.00**	**0.92**	**1.80**	**0.99**	**0.29**
	(−2.29)	(−2.14)	(2.57)		(2.07)	(4.07)	(2.30)	
Fixed income								
Corporate Bonds	**-1.44**	**1.81**	**2.28**	**0.00**	**0.92**	**0.78**	**0.41**	**0.28**
	(−2.53)	(3.11)	(3.70)		(3.85)	(3.28)	(1.77)	
Futures index								
GSCI Composite	**3.45**	**1.42**	**-1.67**	**0.02**	**0.46**	**1.32**	**1.35**	**0.40**
	(2.76)	(1.11)	(−1.23)		(0.88)	(2.51)	(2.64)	
GRCI	**2.02**	**-0.66**	**0.10**	**0.06**	**0.59**	**1.24**	**1.34**	**0.25**
	(2.45)	(−0.78)	(0.12)		(1.71)	(3.58)	(3.99)	
Wald	**0.00**	**0.06**	**0.00**		**0.41**	**0.90**	**0.60**	
Observations	20	19	17		113	113	120	

Note: 1. The average return in each cell is the value of the coefficient corresponding to each condition. The numbers in brackets are t-statistics of the coefficient from the regression of each asset return on the system:

$$R_{a.t} = \sum_{b=0}^{1}\sum_{s=1}^{3} \delta_{b,s,a}\, \mathbf{1}_{b,s,t} \qquad (6)$$

$R_{a.t}$ is return of asset a (including US stocks, international stocks, US corporate bonds, commodity indices: GSCI and GRCI) in month t; s is a stage of a boom/recession: $s = 1$ to 3 which stands for an early, middle and late stage; b stands for business cycle: recession ('0') or boom ('1'); $\mathbf{1}_{b,s,t}$ is dummy variable: $\mathbf{1}_{b,s,t} = 1$ when month t is under business cycle b, stage s; otherwise $\mathbf{1}_{b,s,t} = 0$. As we have 5 assets and 6 coefficients for each assets we will have 30 coefficients for this system.
2. The 'Wald (1)' columns report the probabilities from the Wald tests on the significance of the differences among mean returns of three stages in booms.
3. The 'Wald (2)' row reports the probabilities from the Wald tests on the significance of the differences among mean returns of GSCI against US stocks.

(2005) (where they only consider two stages) in that the middle stages are now found to be attractive too, not just the late boom and early recession periods.

To test where a 3-way split does better than a 2-way, we apply the Diebold and Mariano (1995) (DM) statistical test to compare the squared residuals obtained from the 3-stage and 2-stage models. The results[15] show that for all assets except corporate bonds, the 3-stage model has lower squared residuals than the 2-stage one, suggesting that the 3-way split does better than the 2-way one. While the differences are not statically significant, the consistency

[15] We can provide them upon request.

Table 13: Monthly average returns (%) in different periods with *ex post* stages: 4/1973 - 12/2007

	Restrictive policy					**Expansive policy**				
	Recession	**Boom**				**Recession**	**Boom**			
		Early	Middle	Late	Wald(1)		Early	Middle	Late	Wald(1)
Stock										
US	**-1.83**	**2.10**	**0.43**	**0.92**	**0.53**	**1.87**	**0.91**	**1.85**	**1.34**	**0.40**
	(−2.12)	(1.51)	(0.68)	(2.06)		(2.33)	(2.09)	(3.37)	(1.49)	
EAFE	**-2.32**	**-0.18**	**0.85**	**0.95**	**0.77**	**0.60**	**1.02**	**2.53**	**1.14**	**0.12**
	(−2.48)	(−0.12)	(1.25)	(1.97)		(0.69)	(2.18)	(4.26)	(1.17)	
Fixed-inc										
Bonds	**-0.17**	**1.69**	**0.82**	**0.45**	**0.30**	**1.63**	**0.85**	**0.76**	**0.27**	**0.62**
	(−0.34)	(2.06)	(2.20)	(1.68)		(3.44)	(3.30)	(2.34)	(0.51)	
Futures										
GSCI	**3.66**	**0.38**	**1.34**	**1.81**	**0.70**	**-0.91**	**0.47**	**1.30**	**-0.52**	**0.36**
	(3.34)	(0.22)	(1.69)	(3.19)		(−0.89)	(0.86)	(1.87)	(−0.46)	
GRCI	**1.25**	**-0.64**	**1.17**	**1.67**	**0.15**	**-0.10**	**0.71**	**1.30**	**0.02**	**0.32**
	(1.73)	(−0.55)	(2.23)	(4.44)		(−0.15)	(1.96)	(2.81)	(0.03)	
Wald	**0.00**	**0.45**	**0.37**	**0.22**		**0.03**	**0.53**	**0.53**	**0.20**	
Obs	26	10	49	96	155	30	103	64	24	191

Note: 1. The average return in each cell is the value of the coefficient corresponding to each condition. The numbers in brackets are t-statistics of the coefficient from the regression of each asset return on the system:

$$R_{a.t} = \sum_{m=0}^{1} \sum_{s=0}^{3} \mu_{m,s,a} \mathbf{1}_{m,s,t}. \tag{7}$$

In the above, $R_{a.t}$ is return of asset a (including US stocks, international stocks, US corporate bonds, commodity indices: GSCI and GRCI) in month t. As we only have data for US corporate bonds as of 1/1973, we do not include bonds in this test; m stands for monetary policy: restrictive ('0') or expansive ('1'); s refers to a stage in the (full) cycle: $s = 0$ for a recession, and $s = 1$ to 3 for an early, middle and late boom; $\mathbf{1}_{m,s,t}$ is dummy variable: $\mathbf{1}_{m,s,t} = 1$ when month t is under a monetary regime m and a cycle stage s; otherwise, $\mathbf{1}_{m,s,t} = 0$. As we have 5 assets, and 8 economic conditions for each asset, we have 40 coefficients for this system.
2. The 'Wald (1)' columns report the probabilities from the Wald tests on the significance of the differences among mean returns of three stages in booms.
3. The 'Wald (2)' row reports the probabilities from the Wald tests on the significance of the differences among mean returns of GSCI against US stocks.

across assets is nevertheless encouraging. Thus, we pursue what seems to be a more promising definition of stages, the threefold one.

The interaction effects of monetary policy and business cycle on asset returns:

In this subsection, we want to test whether our GR×JJM strategy is also confirmed in Gorton's sample. In Table 13, we examine the average returns of different monetary policies in different stages of business cycles as we do for our in-sample analysis. The only difference is that in this period we use *ex post* stages instead of *ex ante* stages to make our results comparable to Gorton's and Jensen's results.

From Table 13, we also find that, differently from the findings of GR (2005) and JJM (2002), under a restrictive policy commodity futures indices do not work better than stocks and bonds in early boom stages. The above facts make their conclusions weaker in the sense that (i) periods of a late recession or a middle boom still bring higher returns for commodities; (ii) a restrictive policy does not always result in a better performance for commodities.

Appendix II: The DM test

The DM method is an unconditional test of the null hypothesis of equal return.

$$H_{0,dm} : E_t[\mathrm{d}ret_t] = 0. \tag{8}$$

We define T_{out} := the number of observations used out-of-sample; $\overline{\mathrm{d}ret} := T_{out}^{-1} \sum_{t=T_{in}+1}^{T} \mathrm{d}ret_t$; $\widehat{LRRET}(\mathrm{d}ret_t)$:= the estimate of long run return of $\mathrm{d}ret_t$; and:

$$dm := \frac{\sqrt{T_{out}} \cdot \overline{\mathrm{d}ret}}{\sqrt{\widehat{LRRET}(\mathrm{d}ret_t)}}, \tag{9}$$

which is a standard t-test on a mean (of $\mathrm{d}ret$, here). Diebold and Mariano (1995) show that $dm \overset{\mathrm{d}ret}{\rightarrow} N(0,1)$. In an application, the DM statistic can be computed as the t-statistic in a regression of $\mathrm{d}ret_t$ on a constant with the Newey-West standard error. Note that the sign of $\overline{\mathrm{d}ret}$ indicates the direction of rejection of the two-sided DM test. If in our test we observe, for instance, $\overline{\mathrm{d}ret} < 0$, this implies that $ret(t^1) > ret(t^2)$, i.e. that strategy t^1 does better than t^2 and vice versa.

For the DM test for equal Sharp ratios or standardized returns, we substitute ret by $sret = ret/stdv$ with $stdv$ as the standard deviation of an asset return in the sample.