# Hedge Effectiveness Testing Revisited

## IRA G. KAWALLER AND PAUL D. KOCH

**IRA G. KAWALLER**
is the president of
Kawaller & Co. in
Brooklyn, NY.
kawaller@kawaller.com

**PAUL D. KOCH**
is a professor at the
University of Kansas
School of Business in
Lawrence, KS.
pkoch@ku.edu

*This article provides two recommendations for hedging practitioners attempting to qualify for special hedge accounting treatment. First, we propose an alternative measure to the traditional dollar offset ratio, in which we prescribe division by the starting value of the hedged item rather than division by the change in the value of the hedged item. Our proposed measure provides similar information to the dollar offset ratio, but it is less prone to the potential problem of division by zero in times of low volatility. As a result, it is less likely to exceed acceptable boundary conditions during periods of calm markets (i.e., when small price changes occur). Second, we propose an alternative to the R-squared of a regression analysis as a measure of anticipated hedge effectiveness, which documents the proportion of the total risk that is actually mitigated given the hedge ratio chosen by the hedger, as opposed to the proportion of total risk that would be mitigated if the hedger used the regression slope coefficient as the hedge ratio.*

This article addresses two problems often encountered by practitioners who endeavor to qualify for special hedge accounting treatment. The first problem has to do with the traditional dollar offset ratio (DOR), which is constructed as the ratio of gains or losses on the hedging derivative relative to gains or losses associated with the risk being hedged. During periods of low price volatility, the denominator of this ratio (i.e., the change in the value of the hedged item) is likely to be close to zero, making the DOR prone to exceed the upper bound that serves as a limiting condition for applying hedge accounting (i.e., above 1.25).

As a result of this problem, hedgers that rely on DORs for effectiveness testing may find that hedge accounting is unfairly disallowed during periods when price changes have limited economic effect. We propose an alternative measure to the DOR—the percentage offset ratio (POR), which prescribes division by the starting value of the hedged item, rather than division by the change in this value. Our proposed measure provides similar information to the DOR regarding the anticipated effectiveness of the hedge, but it is less prone to being inflated in periods of low volatility due to division by a price change near zero.

The second problem has to do with the $R^2$ measure of a linear regression analysis, which is often used by hedgers to document anticipated hedge effectiveness. The $R^2$ statistic measures the proportion of the total risk that would be eliminated by implementing a hedge for which the size of the hedge position (i.e., the hedge ratio) is dictated by the slope of the associated regression. In general, however, the actual hedge ratio implemented will differ from this slope value. We propose an alternative measure we call the

"$R^2$ analogue," which correctly reflects the proportion of the total risk being hedged (i.e., the total variation in the price change or price level of the hedged item) that is actually mitigated by the hedge ratio put in place.

## LITERATURE REVIEW ON HEDGE ACCOUNTING, HEDGE RATIOS, AND EFFECTIVENESS

Accounting Standards Codification (ASC) 815 (formerly, Financial Accounting Standard No. 133) requires hedgers to validate their expectation that a prospective hedge will be "highly" effective in offsetting a particular risk exposure, in order to qualify for special hedge accounting treatment. Companies that meet this requirement are permitted to recognize offsetting gains or losses on the hedged item in the same period as any loss or gain on the hedging instrument, potentially dampening the overall impact on earnings (e.g., see Charnes et al. [2003]).

The common measure for validating the expectation that a hedge will be "highly" effective is the $R^2$ from a regression analysis, with the regression slope coefficient providing an estimate of the minimum variance hedge ratio. While it is widely accepted that a necessary condition to qualify for special hedge accounting treatment is a regression $R^2 \geq 0.80$, the proper design of the regression model is a matter of some dispute.[1]

The textbook prescription for measuring the optimal hedge ratio and assessing hedge effectiveness relies on a simple regression on price changes. The dependent variable in this regression relates to the risk exposure (e.g., a cash market price), the independent variable relates to the hedging instrument (e.g., a futures contract price), and the time interval for measuring price changes matches the hedge horizon (Ederington [1979]; Hull [2014]).[2] An alternative practice that is common in industry is to estimate a simple regression on price levels rather than price changes.[3] Additionally, a growing body of work recommends applying greater econometric sophistication to estimate the optimal hedge ratio, such as an error correction model to account for cointegration, or variations of GARCH models to account for the dynamic behavior of volatility.[4] See Bunea–Bontas et al. [2009] and Ramlall [2009] for comprehensive reviews that cover the relevant issues in this dialogue.

A lively debate has developed relating to how these derivatives accounting rules affect the risk management activities of firms. Defenders of the rules see them as necessary to avoid the application of hedge accounting when derivatives fail to achieve the intended risk management objectives. On the other hand, critics find that the rules often preclude hedge accounting on technical grounds, thereby resulting in financial statements that fail to reflect the economic intent of the hedging activity. Critics also emphasize that the requirement to validate hedge effectiveness is difficult for practitioners to understand and costly to implement. As a result, this requirement may deter the legitimate use of derivatives for hedging.[5]

This study addresses two specific problems associated with hedgers trying to comply with hedge accounting guidance. First, we expand on existing literature relating to limitations of the DOR discussed above.[6] In general terms, the DOR metric is deficient because during periods of low volatility, it involves division by a price change that may be near zero, which tends to inflate the DOR measure to a value that exceeds the traditional 80%–125% bounds. In statistical terms, if changes in the futures and spot prices are both assumed to be normally distributed, the DOR takes the form of a Cauchy distribution, which may have no moments due to possible division by zero in a period with small price changes.

Canabarro [1999] conducts a simulation that shows that under basic assumptions about the distribution of price changes, the traditional boundaries of 80%–125% for the DOR are violated in more than one-third of all simulated hedges, even though his simulated correlation between price changes is 0.98.[7] The prior work on this issue simply warns hedgers to be wary of this problem with the DOR and recommends using other methods to test hedge effectiveness. Our study is the first to propose instead an alternative measure of hedge effectiveness—the POR. Our proposed measure maintains the spirit of the DOR while addressing its well-documented problems.

One other study focuses on the second problem addressed in this study that has to do with the use of a regression $R^2$ to measure hedge effectiveness. This statistic is deficient if the hedger chooses a hedge ratio other than the slope coefficient from the traditional textbook regression approach. Charnes et al. [2003, p. 10] observe the following:

"… the choice of an appropriate hedge ratio (*h*) enables the hedger to create a combined position with a smaller variance than that associated with the underlying unhedged item alone ... A speculator, on the other hand, will choose a value of *h* such that the variance of the combined position is greater than the variance of the unhedged position …"

Charnes et al. [2003] note that the slope coefficient from the textbook regression approach (*h\**) results in the minimum variance possible for the combined hedged position, or equivalently, the maximum $R^2$ attainable. They further observe that, if the hedger chooses a hedge ratio other than *h\**, then the variance of the combined position will necessarily increase above this minimum risk possible. They argue, however, that as long as the hedger chooses a hedge ratio (*h*) that results in a lower variance for the combined position than that for the unhedged position, risk will be reduced. The authors favor this notion of risk reduction, as opposed to any specific assessment of "offsets" as the basis for qualifying for hedge accounting.

Charnes et al. [2003] pursue this reasoning further to document that if hedgers choose a positive hedge ratio that is smaller than *h\**, then although their risk exposure is underhedged, their derivative position still represents a bona fide hedge since it results in a reduction of overall risk.[8] Alternatively, Charnes et al. [2003] show that, if hedgers choose a hedge ratio that is larger than *h\**, then they are overhedged, in the sense that the variance of the combined position begins to increase above the minimum risk attainable.

We are not aware of any work in the academic or practitioner literature, other than Charnes et al. [2003], that recognizes this second problem addressed in this study—namely, that the $R^2$ statistic is a valid measure for hedge effectiveness only if the hedger applies the regression slope coefficient as the hedge ratio. Our study attempts to fill this void by suggesting an alternative metric—the $R^2$ analogue. This proposed statistic measures the proportion of the total variation in the unhedged item (i.e., the risk exposure being hedged) that is addressed by the hedge ratio actually implemented, rather than by the slope coefficient of the regression analysis.

## BACKGROUND

For most public companies that use derivative contracts in their risk management activities, it is critical to qualify for and apply special hedge accounting. This accounting treatment results in derivatives' gains or losses being reported concurrently with the earnings recognition associated with the risks being hedged. This treatment is generally perceived to be preferred by readers of financial statements, as it reflects the economic objectives of the hedge. In contrast, without hedge accounting, these two earnings components would likely be reported in different periods. Thus, in any single period, reported earnings would essentially tell only half of the story. Moreover, without hedge accounting, reported income each period would likely be more volatile than with hedge accounting, and more volatile earnings are often associated with a lower stock price.

Given these desirable outcomes associated with hedge accounting, one might expect this treatment to be applied universally for all hedge relationships, but that is not the case. Although hedge accounting treatment is an elective, a number of prerequisite conditions must be satisfied before this treatment can be applied. Without satisfying these requirements, hedge accounting is simply not permitted by the accounting authorities.

One of the more substantive requirements to qualify for hedge accounting is that two distinct hedge effectiveness assessment tests must be performed and repeated at least once per quarter, throughout the life of the hedge. One test relates to how the hedge is expected to perform, prospectively, and the other relates to how the hedge has actually performed, retrospectively. In both assessments, the hedger must determine that the hedging derivative is "highly effective"[9] at offsetting either (a) changes in fair values attributable to the hedged risk (for fair value hedges) or (b) cash flows attributable to the hedged risk (for cash flow hedges).[10] It is left to the hedging entity to devise these testing methodologies and to articulate these procedures in the official hedge documentation.

Such tests are typically trivial when hedges can be designed to achieve this offset perfectly. In such a case, the assertion of a perfect hedge essentially satisfies the testing requirements. On the other hand, if perfectly offsetting hedges are not feasible or are not applied, more explicit objective testing methodologies are needed to

satisfy the requirements for both prospective and retrospective testing. For example, an imperfect hedge might involve an interest rate exposure or foreign exchange exposure, whereby the critical pricing dates or settlement dates for the derivative differ from those relating to the risks being hedged.

A hedging relation might also be judged as imperfect if the hedging derivative includes as a component of its value some amounts that are not explicitly related to the risk being hedged, such as an option premium or an embedded financing in a swap. Another common example of an imperfect hedge arises with commodity hedges, for which it is typical for the derivative to reference some industry-standard price, while the risk being hedged might involve terms relating to a similar but not identical commodity—for example, a commodity with either a different quality grade or a different delivery point.[11]

Since satisfying prospective and retrospective effectiveness assessments serve as prerequisites for applying hedge accounting, it should be obvious that practitioners would want to design and implement effectiveness tests that they can pass. The design of these tests must be specified as part of the official hedge documentation, along with the determination of whether those tests are satisfied. Such tests should expressly address any inherent source of ineffectiveness that can be readily identified; at present, it is the auditor who has the ultimate authority to determine whether the tests are appropriately specified.

**THE DOR**

One permissible approach to satisfy the effectiveness testing requirement is to perform a "scenario analysis." Typically, this method involves constructing a DOR, whereby the gains or losses of the hedging derivative (the numerator) are compared to the gains or losses of the item being hedged (the denominator). The hedge documentation may specify that this metric be constructed either with reference to period-by-period changes, or for a time frame relating to cumulative effects—but only one or the other.[12] In certain situations, the single-period DOR might fall within acceptable boundaries while the cumulatively calculated ratio might not, or vice versa. Whether hedge accounting is permitted would depend on which construction was specified in the hedge documentation, and whether the boundary conditions were satisfied for that particular metric.

For prospective testing, the calculation would involve either some hypothesized situation or an analysis of a comparable hedge that might have been implemented at some earlier time. Once again, the selection of which time frames should be used for this analysis or what hypothetical data are appropriate is left to the judgment of the practitioner, subject to the approval of the auditor. For the retrospective test, on the other hand, the dollar offset calculations should reflect the performance of the actual hedge under consideration. None of these data selection conditions, however, is specifically articulated by the accounting guidance. Also absent from the accounting guidance is any explicit minimum DOR that would serve as a qualifying threshold condition for passing an effectiveness test; but in this case, the auditing community has come to a consensus that DORs must fall between the extreme values of 80% and 125% as a qualifying criterion.

These boundary conditions for the DOR, however, have proven to be problematic. Both theory and practice have come to appreciate that the failure rate on this test appears to be unreasonably high, even for seemingly well-designed hedges. In particular, the test tends to fail too often in periods having little price variation because when the denominator of the offset ratio approaches zero, the ratio can become large without bound and can thus reach values well above the 1.25 maximum allowed.[13]

This article proposes an alternative metric that maintains the spirit of the DOR orientation, while mitigating the problem of potential division by zero. We suggest examining the ratio of the change in the value of the combined hedge portfolio (i.e., the numerator) to the starting value of the hedged item (i.e., the denominator), whereby the combined hedge portfolio consists of both the item at risk and the derivative used to manage that risk. We propose that this alternative metric should remain within the some maximum percentage of the starting value of the hedged item (e.g., $-0.20$ to $+0.20$), in order to qualify for special hedge accounting.[14] We leave the precise boundary thresholds for others to stipulate, but the idea should be clear. By scaling the denominator of this proposed metric by the starting value of the hedged item, rather than by *changes* in the value of the hedged item, we circumvent the problem of potential division by zero that is inherent in the traditional DOR calculation. Henceforth, we refer to our proposed metric as the POR.

Mathematically, the traditional, widely used DOR is defined as follows:

$$\text{DOR}_t = \frac{N \cdot \Delta D_t}{Q \cdot \Delta P_t}$$

where $\text{DOR}_t$ = DOR at time $t$;

$N$ = Notional size of the derivative;
$\Delta D_t$ = Change in the underlying price for the derivative contract over the testing interval;
$Q$ = Quantity of the hedged item;
$\Delta P_t$ = Change in the price of the hedged item over the testing interval.

Our proposed POR is defined as follows:

$$\text{POR}_t = \frac{(Q \cdot \Delta P_t - N \cdot \Delta D_t)}{Q \cdot P_0}$$

where $\text{POR}_t$ = POR at time $t$; and

$P_0$ = Price of the hedged item at the start of the evaluation period (time $t_0$).

This formula is expressed in a general form, allowing for consideration of a cross hedge, whereby the volume of the exposure ($Q$) and the volume of the derivative ($N$) may be measured in different units. In the more typical situation, whereby the exposure and the derivative pertain to a common underlying, $Q$ and $N$ would be expected to be identical.

In constructing either the DOR metric or the POR metric, the length of the testing interval is critical. The hedging entity would still have to satisfy the Financial Accounting Standards Board (FASB) directive to choose between assessing hedge effectiveness over a single-period hedge horizon or over a cumulative horizon. In the former case, the change interval would be a single period. In the latter case, the prospective test's testing interval would reflect the expected remaining horizon of the hedge, while the retrospective test's interval would be one that expands each successive quarter as the hedge continues throughout its life. Later in this article, we offer a case study that explicitly calculates these metrics and discusses their application to hedge effectiveness testing requirements.

For a perfect hedge, the numerator of the POR statistic (i.e., the sum of gains/losses for the combined hedged portfolio) would be zero throughout the hedge horizon. Obviously, for an imperfect hedge, this numerator would be non-zero; and the magnitude of that non-zero value would reflect the ineffectiveness of the hedge. This ineffectiveness, however, fosters an earnings impact only when the derivative's result is larger than the gain or loss associated with the hedged item. Given this asymmetric accounting consideration, some hedgers might be tempted to game the process by applying a smaller hedge than theory would dictate, but the accounting guidance generally precludes this effort. That is, by requiring the user to identify the portion of any exposure that is defined as the hedged item, any hedge construction that addresses a smaller volume necessarily biases the measure in such a way as to increase the chance of violating the lower bound while at the same time decreasing the prospect of violating the upper bound.

Although we view our proposed alternative POR as an improvement over the standard DOR, even this proposed measure may be problematic if the price of the hedged item ($P_0$) is close to zero. In that case, the denominator of the $\text{POR}_t$ (i.e., the starting value of the hedged item) would also be close to zero, resulting in a potentially large value of the $\text{POR}_t$ that may inappropriately signal a problem with the hedge, even though little or no material price effects occur. For example, short-term interest rates are currently near zero, and as a consequence, even a small change in interest rates may cause the $\text{POR}_t$ to push beyond seemingly generous boundary conditions. Clearly, interest rate hedgers who elect to use this metric for effectiveness testing would need to appreciate their vulnerability to failing these effectiveness tests during periods of extremely low rates, if an abrupt interest rate adjustment occurs.

## REGRESSION ANALYSIS

The accounting guidance pertaining to special hedge accounting allows the use of regression analysis and other statistical methods for prospective and retrospective effectiveness testing. Although no precise regression design was prescribed, the FASB explicitly authorized using the same regression design for both prospective and retrospective tests.[15] This guidance has served as a workaround to allow for the continuance of hedge accounting even if DOR tests failed, as long as the structure of the regression analysis was detailed in the official hedge documentation.

Soon after that guidance was published, the U.S. Securities and Exchange Commission (SEC) reacted to it in the form of a speech, given by John M. James, to the Thirty-First AICPA National Conference on Current SEC Developments in 2003. Despite the disclaimer that the speech did not represent official commission views, this speech—like any other given by an SEC staffer—strongly influenced practice. One relevant portion of his speech is presented below:[16]

> "What is problematic is when regression analysis is used and the statistical validity of such analysis is not adequately considered. Specifically, the staff is aware of situations where certain registrants have not fully considered the relevant outputs from the regression analysis when assessing whether the hedge is expected to be highly effective. The staff acknowledges that the assessment of whether a hedging relationship is expected to be highly effective will be determined based on the facts and circumstances of that specific relationship. However, the staff believes that, at a minimum, certain regression outputs such as the coefficient of determination (R-squared), the slope coefficient and the *t* or F-statistic should be considered when using regression analysis to assess whether a hedge is expected to be highly effective. Additionally, depending on the specifics of the hedging strategy, other regression outputs may also need to be considered. The staff expects that if registrants are utilizing statistical techniques to assess hedge effectiveness that they understand how to use and appropriately evaluate such techniques, which may necessitate the use of specialists."

For the most part, this speech resulted in hedgers routinely including the full regression output from their hedge effectiveness tests in their hedge documentation, often with additional stipulations relating to the significance levels associated with parameter estimates. This seeming remedy, however, ignored a fundamental and ubiquitous conceptual conundrum. At present, these statistical concerns expressed by Mr. James notwithstanding, the auditing community universally applies the requirement that the $R^2$ statistic in any regression used for effectiveness testing to be no lower than 0.8 as a prerequisite to allow hedge accounting to be applied.

Although this threshold condition is nowhere to be found in the FASB's official accounting guidance, it is universally accepted as a necessary qualifying condition when regressions are used for effectiveness testing. Unfortunately, depending on the design of this regression analysis, the resulting $R^2$ statistic may fail to address the central question that concerns the FASB—how well do hedge results actually offset the risk being hedged?

## THE REGRESSION $R^2$ AS A MEASURE OF HEDGE EFFECTIVENESS

Regardless of the regression design, the resulting $R^2$ statistic only applies as a reasonable measure of anticipated hedge effectiveness if the hedger chooses a hedge ratio *identical* to the slope coefficient from the regression analysis. Alternatively, the $R^2$ might reasonably be considered to be a measure relevant to *potential* risk reduction, as opposed to actual risk reduction. In this case, the $R^2$ reflects the proportion of the total risk (i.e., the total variation in the price change or price level of the hedged item) that is explained or mitigated by the price change or price level of the hedging instrument. Such a metric offers a reasonable measure of the proportion of the total risk embodied by the hedged item that is eliminated by the implementation of a hedge ratio equivalent to the slope coefficient. In contrast, if the hedger chooses a hedge ratio different from the regression-generated slope coefficient, the $R^2$ statistic will relate to something *other than* the actual hedge in place.

In many hedging applications, the price (or rate) associated with the exposure may be expected to move one-to-one with the price underlying the derivative. In such cases, a hedge would typically be constructed to equate the notional size of the derivative to the size of the exposure (i.e., a one-to-one hedge ratio). For example, when hedging oil purchases or sales with oil futures, it may be reasonable to expect the number of barrels involved in the derivative position to equal the number of barrels being hedged. In this case, implicitly, the hedger would be ignoring basis considerations and effectively expecting (hoping) that these basis effects would be small enough so as not to cause the effectiveness tests to fail.[17] Alternatively, this hedger might use the regression slope coefficient as the optimal hedge ratio, as defined above, with the goal of minimizing the variance of the dependent variable (either the price level or price changes).

In general, the slope for a "cross hedge" will *not* be one-to-one. A cross hedge involves a hedging relation in which the price underlying the hedging derivative is something other than the price associated with the exposure being hedged. Consider, for example, the case of hedging prospective jet fuel purchases (i.e., the exposure) with heating oil futures (i.e., the hedging derivative). In this situation, the hedger would have to determine how these two respective prices relate in order to properly size the optimal hedge that would minimize the risk exposure (i.e., the variance of the combined hedged position). A regression of the price change (or price level) of jet fuel on the price change (or price level) of heating oil futures would serve this purpose. The regression's slope coefficient would then represent the optimal hedge ratio. In this case, however, one would expect a regression analysis to generate a slope coefficient different from one.

In practice, hedgers likely apply hedge ratios that differ from their associated slope coefficient, if only because of rounding considerations. For example, a slope coefficient of 1.12 might likely result in matching each unit of exposure with 1.1 (rather than 1.12) units of the hedging derivative. In any case, to the extent that the actual hedge ratio applied does not *precisely* match the regression slope coefficient, the regression $R^2$ statistic would not accurately measure the effectiveness of the hedge under consideration. And in such a situation, Mr. James's comments notwithstanding, if the actual sizing of the hedge differs from that dictated by the regression's slope coefficient, the statistical significance of that slope coefficient would be irrelevant to the issue of the effectiveness of the hedge actually applied.

In light of these concerns, we propose an alternative statistic—the $R^2$ analogue—as a more appropriate metric for assessing how closely the actual hedge gains or losses have (or will) offset the designated risks being hedged. The $R^2$ analogue that is appropriate for analysis of price changes (as opposed to price levels) is defined as follows:

$$R^2 \text{ Analogue} = 1 - SSE^*/SST$$

where SSE* = the total variation in the time series of the period by period changes in the value of the combined exposure and the derivative position about its mean, *employing the actual hedge ratio used*; and SST = the total variation in the period by period changes in the value of the exposure about its mean.[18] In the general case, when the actual hedge ratio applied is $H$, SSE* and SST would be calculated as follows:

$$SSE^* = \sum (\Delta E - H \cdot \Delta D)^2$$

and

$$SST = \sum (\Delta E - mean(\Delta E))^2$$

where $\Delta E$ is the change in the price of the exposure and $\Delta D$ is the change in the price of the derivative. Furthermore, in the special case when the actual hedge ratio is unity (i.e., for a one-for-one hedge, where $H = 1$), SSE* would be:

$$SSE^* = \sum (\Delta E - \Delta D)^2$$

Critically, when constructing the $R^2$ analogue, a specific time interval must be chosen to measure price changes. In this regard, the choice of a three-month interval, while arbitrary, would seem to be consistent with the FASB's directives that: (a) hedge effectiveness must be assessed no less frequently than quarterly, and (b) the assessment may be made on a period-by-period basis. The $R^2$ analogue constructed with three-month price changes (and re-calculated every three months with updated data) would appear to satisfy this FASB requirement.

## A CASE STUDY

To illustrate our concerns, we examine the performance of a hypothetical hedge of gasoline to be purchased in Montgomery, Alabama, where the hedging derivative used is the actively traded reformulated blendstock for oxygenate blending (RBOB) gasoline futures contract at the New York Mercantile Exchange (NYMEX, a division of the CME Group), which references gasoline prices in New York and New Jersey. Again, in any period, both the prospective and retrospective tests have to be satisfied (no less frequently than quarterly) in order to qualify for hedge accounting in that period; and both the specific metric and the qualifying criteria must be explicitly stated in the hedge documentation. This case examines two alternative approaches—one relying on scenario analysis and the other relying on regression analysis.

Exhibit 1 plots the respective histories of the cash price of gasoline delivered at Montgomery, Alabama, and the nearby futures price of gasoline (RBOB). Clearly, the two prices are highly correlated ($\rho = 0.982$).[19] This plot, along with the high correlation, should foster

## EXHIBIT 1
**Gasoline Cash and Futures Prices ($/gal.)**



the expectation that the contract could reasonably be expected to serve well as a hedging derivative.
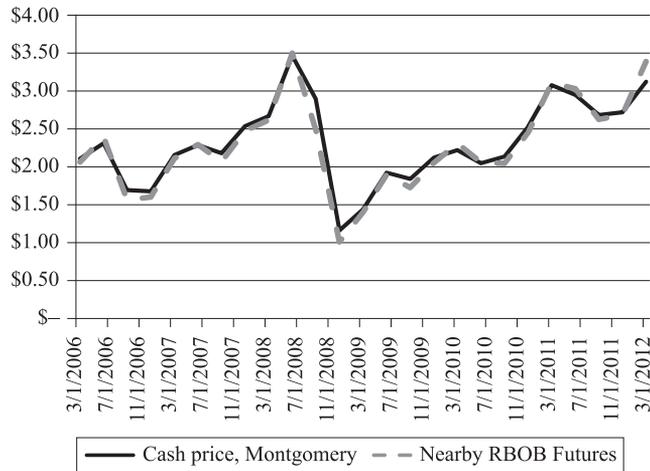
### Scenario Analysis

This intuition, however, is not validated by traditional DORs. In Exhibit 2, we list the change in the end-of-quarter cash price for gasoline at Montgomery ($\Delta P$) and the analogous quarterly change in the nearby RBOB price ($\Delta D$), from the second quarter of 2006 through the first quarter of 2012. We also present the quarterly values of the DOR and POR measures.

In this exhibit, we assume that the hedge is initially implemented during the second quarter of 2006, when the price for gasoline, represented by the Y variable, was $2.11 per gallon. We calculate current and cumulative values for DORs and PORs and assume limits of 80%–125% for the DORs and +/–20% for the PORs as boundary conditions required to qualify for hedge accounting. We show values in bold, italicized print to highlight periods in which the metrics fall outside of these acceptable bounds. The current DOR falls outside the acceptable range 16 times (out of 24);[20] the cumula-

## EXHIBIT 2
**Effectiveness Testing Metrics, the DOR and POR**

(Y starting value = $2.11; and Q = N)

| | ΔP na | ΔD na | Cum ΔP na | Cum ΔD na | Cur DOR na | Cum DOR na | Cur POR 0% | Cum POR 0% |
|---|---|---|---|---|---|---|---|---|
| 6/30/2006 | 0.20 | 0.33 | 0.20 | 0.33 | *162%* | *162%* | −6% | −6% |
| 9/29/2006 | (0.62) | (0.83) | (0.42) | (0.50) | *134%* | 121% | 9% | 4% |
| 12/29/2006 | (0.02) | 0.04 | (0.43) | (0.46) | *−249%* | 107% | −3% | 1% |
| 3/30/2007 | 0.48 | 0.51 | 0.05 | 0.05 | 107% | 100% | −2% | 0% |
| 6/29/2007 | 0.13 | 0.18 | 0.18 | 0.23 | *138%* | *128%* | −2% | −2% |
| 9/28/2007 | (0.11) | (0.23) | 0.07 | 0.00 | *208%* | *5%* | 5% | 3% |
| 12/31/2007 | 0.36 | 0.41 | 0.43 | 0.41 | 114% | 96% | −2% | 1% |
| 3/31/2008 | 0.13 | 0.14 | 0.56 | 0.55 | 105% | 98% | 0% | 1% |
| 6/30/2008 | 0.80 | 0.89 | 1.36 | 1.44 | 111% | 106% | −3% | −4% |
| 9/30/2008 | (0.57) | (1.02) | 0.79 | 0.42 | *178%* | *53%* | 13% | 17% |
| 12/31/2008 | (1.73) | (1.48) | (0.95) | (1.06) | 85% | 111% | −9% | 5% |
| 3/31/2009 | 0.28 | 0.39 | (0.67) | (0.66) | *141%* | 99% | −10% | 0% |
| 6/30/2009 | 0.48 | 0.50 | (0.19) | (0.17) | 103% | 90% | −1% | −1% |
| 9/30/2009 | (0.08) | (0.17) | (0.27) | (0.34) | *211%* | *127%* | 5% | 3% |
| 12/31/2009 | 0.28 | 0.33 | 0.01 | (0.01) | 117% | *−90%* | −3% | 1% |
| 3/31/2010 | 0.10 | 0.26 | 0.11 | 0.25 | *262%* | *220%* | −7% | −6% |
| 6/30/2010 | (0.17) | (0.25) | (0.06) | (0.00) | *145%* | *6%* | 3% | −3% |
| 9/30/2010 | 0.09 | (0.02) | 0.03 | (0.02) | *−18%* | *−72%* | 5% | 2% |
| 12/31/2010 | 0.39 | 0.41 | 0.42 | 0.39 | 104% | 92% | −1% | 2% |
| 3/31/2011 | 0.55 | 0.65 | 0.97 | 1.04 | 120% | 108% | −4% | −4% |
| 6/30/2011 | (0.13) | (0.08) | 0.84 | 0.97 | *60%* | 115% | −2% | −6% |
| 9/30/2011 | (0.26) | (0.41) | 0.58 | 0.56 | *153%* | 97% | 5% | 1% |
| 12/30/2011 | 0.03 | 0.06 | 0.61 | 0.62 | *175%* | 102% | −1% | −1% |
| 3/30/2012 | 0.40 | 0.70 | 1.01 | 1.33 | *175%* | *131%* | −11% | −15% |

tive DOR falls outside 10 times; both current cumulative POR measures stay in their acceptable bounds in all periods.

Although offset ratios can be used for hedge effectiveness testing, the manner in which they are used requires some further explanation. With the "scenario analysis" orientation prescribed in the current accounting guidance, each row in Exhibit 2 would serve as a specific effectiveness test that is applied at the end of each quarterly period throughout the hedge horizon. The very first prospective test, however, would have to be constructed using *earlier* data (i.e., collected prior to the hedge implementation), but following the same methodology. This requirement thus allows for some discretion as to exactly which period the hedger chooses to reference for the first prospective test.

More likely than not, the analysis would use data for the period immediately prior to the hedge implementation—unless that metric failed. Clearly, in that case, the desire to pass the test would motivate selecting an alternative reference period. In any case, irrespective of the period chosen for the first prospective test, after that first prospective test, all the following offset ratios would serve double duty, functioning both as the retrospective test for the period just passed and as the prospective test for the next period to come. This recursive process would be repeated, period by period, throughout the life of the hedge.

Although the approach described above is the most common application of offset ratios for effectiveness testing in practice, the FASB also allows for the use of regression analysis or "other statistical analysis," without being specific about what other kinds of tests would be acceptable.[21] We suggest that an analysis of the POR over a series of past periods that reveals a low frequency of falling out of acceptable bounds should similarly satisfy the FASB's testing requirements.[22] We are unaware, however, of any hedging entity that currently employs an analogous application of the DOR to provide such a frequency test of hedge effectiveness. In any case, it should be clear from Exhibit 2 that a frequency test performed using either POR metric would be far more likely to permit hedge accounting treatment than would a frequency measure based on either of the DOR metrics.

### Regression Analysis

Using the same data on gasoline cash and futures prices that are analyzed in Exhibit 2, we now turn to the regression approach. Exhibit 3 shows the results from a simple regression of quarterly changes in the Montgomery cash price of gasoline on quarterly changes in the RBOB futures price. In this example the $R^2$ is comfortably above 0.8, so a typical conclusion would be that the effectiveness assessment is satisfied.[23]

## E X H I B I T   3
**Regression Results**

| Regression Statistics | |
|---|---|
| Multiple R | 0.963 |
| R-Squared | 0.928 |
| Adjusted R-Squared | 0.925 |
| Standard Error | 0.139 |
| Observations | 24 |

**ANOVA**

| | df | SS | MS | F | Significance F |
|---|---|---|---|---|---|
| Regression | 1 | 5.475 | 5.475 | 282.686 | 0.000 |
| Residual | 22 | 0.426 | 0.019 | | |
| Total | 23 | 5.901 | | | |

| | Coefficients | Standard Error | *t*-stat | P-value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|
| Intercept | −0.006 | 0.029 | −0.209 | 0.836 | −0.065 | 0.053 |
| X Variable 1 | 0.873 | 0.052 | 16.813 | 0.000 | 0.765 | 0.981 |

Again, the $R^2$ statistic from this regression is an appropriate measure of anticipated hedge effectiveness only if the hedge in question is sized consistently with the value of the slope coefficient (in this analysis, 0.873). That is, the notional size of the derivative applied must be precisely 0.873 times the notional size of the exposure, in order for this $R^2$ to be a relevant measure of the anticipated effectiveness of the hedge actually put in place.[24]

In fact, a large portion of hedgers does not structure their hedges that way. Rather, they match the two respective notional sizes—that is, they apply a hedge ratio of unity. In that case, the $R^2$ analogue with a value of $H = 1$ would offer an accurate assessment of the anticipated effectiveness of a one-to-one hedge. In contrast, the traditional $R^2$ statistic would be an exaggerated measure that overstates the actual hedge's effectiveness. In this instance, the regression's $R^2$ is 0.928, while the $R^2$ analogue for a hedge ratio of 1.0 is 0.908—both values are comfortably above 0.8.

Such a situation will not always hold, however. As a rule, the $R^2$ analogue is a reliable measure of hedge effectiveness, while the traditional $R^2$ statistic is reliable only if the slope coefficient is used as the hedge ratio. To the extent that a qualifying criterion for hedge accounting is deemed to be necessary, regulators and practitioners alike should prefer to rely on an appropriately constructed metric, such as our proposed $R^2$ analogue.

## CONCLUSIONS

For too long, on a widespread basis, practice has either relied on a discredited methodology for assessing hedge effectiveness (i.e., DORs), or it has appealed to an inappropriately constructed $R^2$ statistic to serve as the basis for satisfying the hedge effectiveness assessments. This article addresses concerns associated with both of these metrics.

First, the traditional DOR suffers from the possibility of division by a number close to zero, which results in a tendency to inflate this measure and preclude hedge accounting during periods of low volatility. Our own sensibilities are that it is inappropriate to disallow hedge accounting on the basis of measured values that arise out of inconsequential or immaterial price changes during periods of relatively calm markets. Our proposed POR is consistent with the dollar offset orientation, but it is less vulnerable to this fatal flaw. We also propose an alternative statistical approach that analyzes the relative

frequency for which the POR metric falls beyond acceptable limits, over some specifically defined time span. In our judgment, a low frequency of failure over a recent span of history should serve to satisfy the FASB's hedge effectiveness testing requirements.

A second problem addressed in this article deals with the typical use of regression for assessing hedge effectiveness. Specifically, an $R^2$ statistic is a relevant measure of hedge effectiveness *only* when the hedger uses the precise regression slope coefficient as the actual hedge ratio in the hedge under consideration. In the general case, these two values (i.e., the slope coefficient and the actual hedge ratio applied) will *not* be the same. We propose an alternative metric—the $R^2$ analogue—which measures the portion of total risk mitigated by the actual hedge ratio employed. This metric provides relevant information about the actual hedge put in place, as opposed to some optimal hedge construction that is not actually implemented. Transitioning to any of our proposed measures or methodologies would represent a marked improvement over current accounting practice.

## ENDNOTES

[1]The Financial Accounting Standards Board has released an exposure draft that, if enacted, would lower the requirement from "highly" effective to "reasonably" effective. It is not clear whether or how this change would affect the current practice of requiring the regression $R^2$ statistic to be at least 0.80 to qualify for hedge accounting.

[2]Prior work has shown that the results of this regression analysis are sensitive to the hedge horizon selected. See Geppert [1995], Howard and D'Antonio [1991], Hill and Schneeweis [1981, 1982], Juhl et al. [2012], and Malliaris and Urrutia [1991a,b].

[3]Levy and Sarnat [1984] and Marmer [1986] discuss various drawbacks to the textbook approach. Dale [1981], Hill and Schneeweis [1981, 1982], and Kawaller and Koch [2012] discuss the use of a simple regression on price levels versus price changes.

[4]For example, see Benz and Hengelbrock [2009], Hsu et al. [2008], Juhl et al. [2012], and Kenourgios et al. [2008].

[5]For further discussion, see Bunea-Bontas et al. [2009], Charnes et al. [2003], DeMarzo and Duffie [1995], Eckstein et al. [2007], Finnerty and Grant [2006, 2007], Juhl et al. [2012], Kawaller and Koch [2000], and Ramlall [2009].

[6]For example, see Canabarro [1999], Charnes et al. [2003], and Finnerty and Grant [2006, 2007].

[7]In this article, we provide a case study that similarly shows that the DOR violates the boundary conditions in 60%

of the quarterly periods analyzed, for a hedging relation with a correlation of 0.98 between the relevant prices.

[8]As long as the hedge ratio remains positive but smaller than $h^*$ (i.e., $0 < h < h^*$), the derivative position offsets the underlying exposure and thereby reduces risk, although it does not achieve the minimum risk possible (at $h^*$).

[9]As discussed above, in an exposure draft currently outstanding, the FASB has proposed changing the criterion from "highly effective," to "reasonably effective," suggesting that the FASB would like to lower the barrier for companies seeking to qualify for hedge accounting. Precisely how the entity would calibrate these assessments, however, is not well defined in either the current or proposed accounting guidance.

[10]Fair value hedges pertain to hedges of recognized assets, recognized liabilities, or firm commitments; cash flow hedges pertain to hedges of forecasted uncertain cash flows.

[11]The Financial Accounting Standards Board (FASB) explicitly requires the hedging entity to identify the change of the *entire* price of a commodity as the risk being hedged, such that any variance between the invoice price and the price underlying the derivative contract would foster deviation from a perfect offset.

[12]The period-by-period approach involves comparing the changes in the hedging instrument's fair values (or cash flows) that have occurred during the period being assessed to the changes in the hedged item's fair value (or hedged transaction's cash flows) attributable to the risk hedged that have occurred during the same period. In contrast, the cumulative approach involves comparing the analogous cumulative changes (to date from inception of the hedge). See ASC 815-20-35-5 for details.

[13]In statistical terms, if changes in the futures and spot prices are both assumed to be normally distributed, then the DOR takes the form of a Cauchy distribution, which may have no moments due to possible division by zero in a period with small price changes. See Charnes et al. [2003].

[14]Our proposed bounds for the POR, between −20% and +20%, are generally consistent with the traditional bounds of 80% and 125% that are applied to the DOR. This proposed range is, admittedly, somewhat arbitrary, but no more so than the original 80%–125% boundary conditions that have been relied upon.

[15]This refinement originally was put out as a Derivatives Implementation Group issue (DIG Issue E7), "Methodologies to Assess Effectiveness of Fair Value and Cash Flow Hedges." It was subsequently codified in 815-20-25, 815-20-35, and 815-20-55.

[16]At the time of this speech, John James was Professional Accounting Fellow, Office of the Chief Accountant, SEC.

[17]In commodity markets, the basis is understood to be the difference between the cash price paid or received for some physical commodity and, typically, the price of the nearby futures contract for the same (or a closely related) commodity.

[18]The $R^2$ analogue is not an original construct in this study. Juhl et al. [2012] propose using this statistic in connection with a discussion of hedge effectiveness assessments when the prices of the hedged item and the hedging derivative are cointegrated and hedge ratios are determined with the aid of an error correction model (ECM). The $R^2$ analogue is appropriate in their context, because their ECM has many right-hand-side variables. As a result, the $R^2$ from the ECM does not reflect the proportion of the total variation in the dependent variable (i.e., the total risk) that is explained or offset by the hedging instrument alone.

[19]This calculation uses quarterly data for the period from March 31, 2006, through March 30, 2012.

[20]Canabarro [1999] reaches a similar conclusion using an interest rate example.

[21]See ASC 815-20-55-69.

[22]ASC 805-20-35-3 allows a common statistical analysis to satisfy both the prospective and retrospective testing requirements, subject to a revision of the dataset to reflect the latest information.

[23]While having an $R^2$ statistic at or above 0.8 is generally considered a necessary condition to qualify for special hedge accounting, it may not be sufficient. Many auditing firms have required their clients to satisfy additional statistical requirements that are not explicitly detailed in the FASB's guidance.

[24]An intuitive justification for applying such a hedge ratio that is less than one-to-one can be found by closely examining Exhibit 1. Note that throughout the history shown, RBOB price changes have tended to be slightly larger in magnitude than Montgomery price changes. Thus, a smaller volume of RBOB-based derivatives should be paired with any exposure to achieve the desired offset. See Hull [2014] for further intuition on this justification.

## REFERENCES

Accounting Standards Codification (ASC), The Financial Accounting Standards Board, asc.fasb.org.

Benz, E.A., and J. Hengelbrock. "Liquidity and Price-Discovery in the European $CO_2$ Futures Market: An Intraday Analysis." Working paper, University of Bonn, 2009.

Bunea-Bontas, C., M. Petre, and G. Culita. "Issues on Hedge Effectiveness Testing." Working paper, Constantin Brancoveanu University, 2009.

Canabarro, E. "A Note on the Assessment of Hedge Effectiveness Using the Dollar Offset Ratio Under FAS 133." Working paper, Goldman Sachs Fixed Income Research, 1999.

Charnes, J., H. Berkman, and P. Koch. "Measuring Hedge Effectiveness for FAS 133 Compliance." *Journal of Applied Corporate Finance*, 15 (2003), pp. 95-103.

Dale, C. "The Hedging Effectiveness of Currency Futures Market." *Journal of Futures Markets*, 1 (1981), pp. 77-88.

DeMarzo, P., and D. Duffie. "Corporate Incentives for Hedging and Hedge Accounting." *Review of Financial Studies*, 8 (1995), pp. 743-772.

Derivatives Implementation Group (DIG) Issue E7, YEAR? "Methodologies to Assess Effectiveness of Fair Value and Cash Flow Hedges."

Eckstein, C., A.J. Markelevich, and A. Reinstein. "Accounting for Derivative Instruments and Hedging Activities (SFAS No. 133): Implications for Profitability Measures and Stock Prices." Working paper, Columbia University, 2007.

Ederington, L. "The Hedging Performance of the New Futures Markets." *Journal of Finance*, 34 (1979), pp. 157-170.

Finnerty, J.D., and D. Grant."Testing Hedge Effectiveness under SFAS 133." *The CPA Journal*, Vol. 73, No. 4 (2006), pp. 40-45, www.nysscpa.org/cpajournal/2003/0403/features/f044003.htm.

——. "Alternative Approaches to Testing Hedge Effectiveness under SFAS No. 133." *Accounting Horizons*, Vol. 16, No. 2 (2007), pp. 95-108.

Geppert, J.M. "A Statistical Model for the Relationship between Futures Contract Hedging Effectiveness and Investment Horizon Length." *Journal of Futures Markets*, Vol. 15, No. 5 (1995), pp. 507-536.

Hill, J., and T. Schneeweis."A Note on the Hedging Effectiveness of Foreign Currency Futures." *Journal of Futures Markets*, 1 (1981), pp. 659-664.

——. "The Hedging Effectiveness of Foreign Currency Futures." *Journal of Financial Research*, 5 (1982), pp. 95-104.

Howard, C.T., and L.J. D'Antonio. "Multiperiod Hedging Using Futures: A Risk Minimization Approach in the Presence of Autocorrelation." *Journal of Futures Markets*, 11 (1991), pp. 697-710.

Hsu, C.C., C.P. Tseng, and Y.H. Wang. "Dynamic Hedging with Futures: A Copula-Based GARCH Model." Working paper, National Taiwan University, 2008.

Hull, J.C. *Options, Futures, and Other Derivatives*, 8th Edition. New Jersey: Pearson-Prentice Hall, 2014.

Juhl, T., I. Kawaller, and P. Koch. "The Effect of the Hedge Horizon on Optimal Hedge Size and Effectiveness When Prices Are Cointegrated." *Journal of Futures Markets*, Vol. 32, No. 9 (2012), pp. 837-876.

Kawaller, I., and P.D. Koch. "Meeting the 'Highly Effective Expectation' Criterion for Hedge Accounting." *The Journal of Derivatives*, Vol. 7, No. 4 (2000), pp. 79-87.

——."Designing the Proper Hedge: Theory versus Practice," Working paper, University of Kansas, 2012.

Kenourgios, D., A. Samitas, and P. Drosos. "Hedge Ratio Estimation and Hedging Effectiveness: The Case of the S&P 500 Stock Index Futures Contract." *International Journal of Risk Assessment and Management*, Vol. 9, Nos. ½ (2008), pp. 121-134.

Levy, H., and M. Sarnat. *Portfolio and Investment Selection: Theory and Practice*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

Malliaris, A.G.B., and J.L. Urrutia. "Tests of Random Walk of Hedge Ratios and Measures of Hedging Effectiveness for Stock Indexes and Foreign Currencies." *Journal of Futures Markets*, 11 (1991a), pp. 55-68.

——. "The Impact of the Lengths of Estimation Periods and Hedging Horizons on the Effectiveness of a Hedge: Evidence from Foreign Currency Futures." *Journal of Futures Markets*, 11 (1991b), pp. 271-289.

Marmer, H.S. "Portfolio Model Hedging with Canadian Dollar Futures: A Framework for Analysis." *Journal of Futures Markets*, 6 (1986), pp. 84-92.

Ramlall, I. "Striving for the Optimal Hedge Ratio." Working paper, University of Technology, Mauritius, 2009.

Zhang, H. "Effect of Derivative Accounting Rules on Corporate Risk-Management Behavior." *Journal of Accounting and Economics*, 47 (2009), pp. 244-264.

*To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.*