A Novel Risk Management Framework for Natural Gas Markets

Panos K. Pouliasis *
Cass Business School, City, University of London, United Kingdom

Ilias D. Visvikis
School of Business Administration, American University of Sharjah, United Arab Emirates

Nikos C. Papapostolou
Cass Business School, City, University of London, United Kingdom

Alexander A. Kryukov
Cass Business School, City, University of London, United Kingdom

Abstract

This paper examines dynamic hedges in the natural gas futures markets for different horizons and explores the gains from devising risk management strategies. Despite the substantial progress made in developing hedging models, forecast combinations have not been tested. We fill this gap by proposing a framework for combining hedge-ratio predictions. Composite hedge-ratios lead to significant reduction in portfolio risk, whether spot prices are partially predictable or not. We offer insights on hedging effectiveness across seasons, backwardation-contango conditions and the asymmetric profiles of long-short hedgers. We conclude that forecast combinations better reconcile realized performance with the hedging process, mitigating model instability.

Keywords: Natural gas; Dynamic futures hedging; Forecast combination

JEL Classification: G32, C32, C53, G11, L95

Acknowledgements
The authors would like to thank the editor and one anonymous reviewer for their comments and recommendations on an earlier version of the paper. This study was also helped by the comments of the participants of the Annual Conference of the International Association of Maritime Economists (IAME), 25-28 June 2019, Athens, Greece. Ilias Visvikis would like to thank the American University of Sharjah for the Faculty Research Grant 2019 (FRG19-S-B89). The responsibility for any remaining errors rests with the authors.

* Correspondence: Dr Panos K. Pouliasis, Cass Business School, City, University of London; The Costas Grammenos Centre for Shipping, Trade and Finance; 106 Bunhill Row, London, EC1Y 8TZ, United Kingdom. Telephone: +44(0)2070408807, email: p_pouliasis@city.ac.uk.

1. Introduction

Adverse price trends and sharp fluctuations not only affect profit margins, but also impact a company's probability of default or even alter the incentives of investing (e.g., infrastructure and transportation) – reducing investment in favor of lower risk projects. Such business challenges that are directly linked to, *inter alia*, production/ purchasing costs, earnings and credit availability, create the need for coherent risk management practices. For oil and gas projects, where cash flows are almost entirely generated by oil and gas sales, price volatility increases the incentive to mitigate such effects. Effective natural gas hedging strategies are relevant in reducing price volatility for investors, traders, producers and commercial users in the sector. Moreover, hedging policies constitute a key theme for policy-makers and regulators to consider alternative reforms and mitigate deficiencies (e.g., transaction costs, poor liquidity and transparency) in the current market design. To add, with the Paris Agreement in 2015 and its predecessor the Kyoto Protocol in 1997, there is increasing interest in energy investments with low emissions, such as natural gas. Therefore, given the broad economic and financial impact of natural gas volatility, it is vital to study natural gas risk management strategies.

One crucial parameter of futures-based hedging is the hedge ratio, i.e., the number of futures contracts to buy or sell for each unit of the underlying asset on which the hedger bears risk. Earlier studies (e.g., Ederington, 1979) derive hedge ratios that minimize the variance of the spot/future portfolio based on the principles of portfolio theory. The Optimum Hedge Ratio (OHR) is typically found by regressing the returns to holding the physical asset on the returns to holding the hedging instrument. However, the regression approach has several shortcomings. For example, it omits cointegration between futures and spot prices which might lead to biased OHR forecasts, particularly in the long run (Lien, 1996). Moreover, Bollerslev (1990) and Kroner and Sultan (1993), among others argue that this approach implicitly assumes constant risk throughout time as new market information arrives.

Therefore, a number of hedging models have been developed. Kroner and Sultan (1993) and Engle and Kroner (1995) apply multivariate Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models and derive time-varying hedge ratios from the conditional second moments. GARCH models are popular due to their ability to capture some of the salient features of financial time-series, such as volatility clustering, non-linear dependence and thick tails (see Pouliasis *et al*., 2018). A popular alternative is the Markov Regime Switching (MRS) modeling framework, introduced by Hamilton (1989); see also Alizadeh *et al*. (2008) for an extension to regime switching in a cointegrated GARCH process for hedging energy commodities. Switching models overcome the limitation of constant parameters, offering better model fit and, thus, are able to improve on the hedging ability.

The consensus from the literature is that while dynamic OHRs tend to outperform static hedges, the alleged gains are market specific, though occasionally, the benefits are minimal (Lien and Tse, 2002). Ghoddusi and Emamzadehfard (2017) examine the hedging effectiveness of the Henry Hub natural gas future contract for different physical positions and find that cointegration and time-varying volatilities only marginally effect hedging ability. Overall, studies on the hedging efficiency of natural gas futures are scant, while results indicate that gas futures are the least effective contacts compared to other commodities, thus offering a rich experimental context for our tests (see Cotter and Hanly, 2012; Hanly, 2017).[1]

In this paper, we argue that a more effective hedge may be available in the form of a "*composite hedge*", by exploiting the information content of various models. Model combinations have been used extensively. Yet, most of the works focus on price (Baumeister and Kilian, 2015) or volatility forecasting (Patton and Sheppard, 2009). We fill this gap by

[1] Natural gas markets exhibit properties that distinguish them from other markets. Such are, for example, the region-specific nature of natural gas with restricted access to export markets, as well as the difficulty in storing and transporting gas which creates high basis risks (see Hanly, 2017). Moreover, natural gas price trajectories and the performance of hedges differ not only from other traditional assets, such as stocks and bonds, but also from most commodities, a result of the inherent relatively high volatility.

considering a variety of models which feature prominently in the hedging literature, and their combination in the decision-making process to minimize operating cash flow variability. To this end, this paper investigates the hedging effectiveness of New York Mercantile Exchange (NYMEX) Henry Hub and Intercontinental Exchange (ICE) National Balancing Point (NBP – virtual trading hub) contracts; the most liquid and mature futures markets in the sector.

Our contributions extend in several directions. We first build on a diverse set of models, each targeting a specific feature of natural gas price movements (seasonality, cointegration, time-varying volatility/correlation and regime shifts). From the estimated pool of models – given their theoretical pros and cons and mixed results of their empirical performance (e.g. see Lien and Tse, 2002) – this paper considers a model combination approach in hedging decisions. This way, decision-makers do not rely on some particular econometric specification when estimating the OHR which might not fully reflect the risk inherent in price movements.[2] As a number of studies on the predictive power of individual hedging models report mixed results,[3] implementing the proposed framework acts also as an insurance tool mitigating the undesirable effects of structural breaks and model misspecification, thus, leading to improved forecasting (see, *inter alios*, Pesaran and Timmermann, 2007; Patton and Sheppard, 2009; Baumeister and Kilian, 2015). Therefore, we provide a flexible framework in the hedging process under model uncertainty. Based on the work of Caldeira *et al*. (2017), we put forward an approach to combine OHR forecasts from candidate models. The combination weights are directly linked to the decision-making problem of an investor who wishes to minimize portfolio risk; each model is given

[2] For individual models it is reasonable to discard some modelling features of market dynamics to warrant a parsimonious structure. However, depending on the source of market shocks, ignoring relevant information in the formulation process might be costly.

[3] Gagnon and Lypny (1995) provide evidence in support of GARCH models. In contrast, Lien and Tse (2002) support the traditional regression approach. GARCH models exhibit few limitations. For example, the observed non-normalities in return distributions are more pronounced than those implied by GARCH; the model fails to reproduce time variability in higher moments unless explicitly modelled, and a strong degree of persistence is imputed to volatility which may be due to structural breaks (Lamoureux and Lastrapes, 1990).

importance proportional to its actual performance.

In addition, we evaluate the hedging ability of different OHR prediction models in terms of variance reduction. The analysis is executed both in- and out-of-sample and the results are validated on a statistical basis using Hansen’s (2005) reality check and Politis and Romano (1994) bootstrap simulation methods. This way we provide robust evidence on the potential gains of the proposed forecast combination hedging strategies taking into account transaction costs and utility performance fees. We also address the issue of downside risk by examining whether the effects of mean-variance OHRs differ for different hedging horizons (weekly and monthly), between long and short hedges, during seasons (winter and summer) and across market conditions (backwardation and contango).

Finally, we also assess hedging performance under the supposition of partially predictable spot prices. Ederington and Salas (2008) postulate that, when spot prices are partially predictable, traditional regression estimates the OHR inefficiently, leading to upward-biased riskiness levels for both hedged and unhedged positions and downward-biased risk reduction levels. Martínez and Torró (2015) consider seasonality when OHRs are computed in European gas markets. Their results indicate that hedging effectiveness is much higher when the seasonal pattern in spot price changes is approximated with the basis (futures-spot spread). Fama and French (1987) and Viswanath (1993) also find that the basis is a useful predictor of future changes in spot commodity prices.

The structure of this paper is as follows. Section 2 presents the OHR methodology and demonstrates the forecast combination procedure. In Section 3, the data and their properties are described. Section 4 discusses the empirical results and evaluates the hedging effectiveness of the proposed strategies. The reality check for superior hedging ability is also discussed, while this section presents information on weight and hedge ratio stability, transaction costs, performance fees, downside risk and segmentation to market conditions.

Last section concludes.

2. Methodology

Denote ΔS_t and ΔF_t the returns on the cash and futures respectively, and γ the hedge ratio. A hedger who desires to hedge a future sale (or purchase) price in the future will sell (buy) γ units to eliminate or alleviate the risk associated with being long (short) one unit of the spot position. Based on Ederington (1979), among others, the minimum variance OHR is:

$$\gamma = \frac{Covar(\Delta S_t, \Delta F_t)}{Var(\Delta F_t)} \tag{1}$$

which, is equivalent to the slope coefficient, γ, of the regression:

$$\Delta S_t = \mu + \gamma \Delta F_t + u_t;\ u_t \sim iid(0, \sigma^2) \tag{2}$$

It follows that the payoff of the hedged portfolio is given by $\Delta H_t = \Delta S_t - \gamma \Delta F_t$. The variance of the unhedged position in this case is $Var(\Delta S_t)$ and the hedged position riskiness $Var(\Delta S_t - \gamma \Delta F_t)$. The *traditional* Hedging Effectiveness (HE) measure is the percentage reduction in risk achievable through hedging:

$$HE_1 = 1 - \frac{Var(\Delta S_t)}{Var(\Delta S_t - \gamma \Delta F_t)} \tag{3}$$

In many markets, changes in the spot price are partially predictable. In this respect, traditional estimates of γ are inefficient (yet unbiased), riskiness of hedged and unhedged positions are overestimated and HE is underestimated (see Ederington and Salas, 2008). To this end, the hedger's payoff function can be modified to $\Delta H_t = (\Delta S_t - E_{t-1}[\Delta S_t]) - \gamma \Delta F_t$

where $E_{t-1}[\Delta S_t] = \lambda Y_{t-1}$, is the expected change in the spot price based on information variable(s)/predictors Y_{t-1}. Following Ederington and Salas (2008) and Martínez and Torró (2015) we set Y_t equal to the basis, i.e., $F_t - S_t$, as this spread indicates expected future price changes for the spot commodity, if markets are efficient. Setting $\lambda = 0$ leads to the hedger's traditional payoff function and the OHR of Eq. (1); this is equivalent to the assumption of no predictability which implies that spot prices evolve according to the random walk, i.e., $E_{t-1}[\Delta S_t] = 0$. However, if $E_{t-1}[\Delta S_t] \neq 0$, the OHR and HE may be expressed as:

$$\gamma = \frac{Covar(\Delta S_t - \lambda Y_{t-1}, \Delta F_t)}{Var(\Delta F_t)} \quad (4)$$

$$HE_2 = 1 - \frac{Var(\Delta S_t - \lambda Y_{t-1})}{Var(\Delta S_t - \gamma \Delta F_t - \lambda Y_{t-1})} \quad (5)$$

Finally, let $\sigma_{sf,t}$ be either $Covar_t(\Delta S_t, \Delta F_t)$ or $Covar_t(\Delta S_t - E_{t-1}[\Delta S_t], \Delta F_t)$ and $\sigma_{f,t}^2$ the $Var(\Delta F_t)$. Eqs. (1) and (5) can be extended to accommodate the conditional OHR which is regarded to be more efficient in reducing the risk of a hedged position; because it is updated as it responds to the arrival of new information in the market (see for example, Kroner and Sultan, 1993). The conditional OHR is then:

$$\gamma_t = \frac{\sigma_{sf,t}}{\sigma_{f,t}^2} \quad (6)$$

The search for alternative futures-based risk management strategies has so far focused on finding the "right" modelling framework. However, specifying the hedge ratio to be dependent upon some model that targets a large set of the stylized features of price movements (regime shifts, cointegration, volatility clustering, seasonality, etc.), might not be parsimonious and thus lead to instability and misspecification. To reduce such impact at the forecasting stage, we argue that using model combination methods, one may obtain more efficient hedge ratios.

2.1. Hedge Ratio Forecast Combination Framework

Suppose that we have *N* candidate hedge ratio predictions, with the first *T* observations used as training data. Therefore, the hedger has a toolbox of *N* different models with which to predict the hedge ratio dynamics. Let $\gamma_{i,t}$ be the forecast of the OHR from candidate model i. Accordingly, let $w_{i,t}$ be the combination weight of candidate forecast *i* for γ_t that satisfies $\sum_{i=1}^{N} w_{i,t} = 1$. Then for any *t*, the hedge ratio forecast after weighting is:

$$\gamma_{FC,t} = \sum_{i=1}^{N} w_{i,t}\gamma_{i,t} \quad (7)$$

We specify the combination weights, $w_{i,t}$, so that they take into account the economic decision in which the hedge ratio forecast will be used. That is, the ex-ante $w_{i,t}$ is determined based on the past performance of each model in the minimum variance portfolio optimization, i.e., the lower the hedged portfolio historical variance implied by a model at date *t*, the higher the weight that this model receives in the combination forecast. In addition, we follow Genre *et al*. (2013) and Caldeira *et al*. (2017) and consider a more general scheme for combining OHR forecasts, introducing a discount parameter $\xi \geq 0$. In this setting, portfolio average returns, $c_{i,H}$ and portfolio variance $\sigma_{i,H}^2$ are computed as:

$$\sigma_{i,H}^2 = T^{-1}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\xi}\left(\Delta S_t - \gamma_{i,t}\Delta F_t - c_{i,H}\right)^2$$

$$c_{i,H} = T^{-1}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\xi}\left(\Delta S_t - \gamma_{i,t}\Delta F_t\right) \quad (8)$$

The above formulation explicitly accounts for the stylized facts in financial time series, such as time-varying conditional heteroscedasticity and structural breaks (see Diebold and Pauly, 1987). Examples of the discount factor $(t/T)^{\xi}$ effect are plotted in Figure 1, as ξ ranges from 0 to 20. The case where $\xi = 0$ corresponds to no discounting and is equivalent to the common expression of the sample mean and sample variance, i.e., all observations have an equal influence. If $\xi = 1$ the importance assigned to older data declines linearly. Values of ξ which are below (above) unity produce structures that decrease the importance of older observations at a decreasing (increasing) rate.

Moreover, to obtain the weights, $w_{i,t}$, we introduce tuning parameter $\eta \geq 0$ which controls how aggressively we adjust the mixing weights in response to changes in the realized hedged portfolio risk; η could improve HE via reducing the importance of poor forecast models. Large values of η shrink the weights toward the best performing models. As $\eta \to \infty$ the weight on the model that yields the lowest hedged portfolio risk approaches 1. In particular:

$$w_{i,t} = \frac{\sigma_{i,H}^{-\eta}}{\sum_{j=1}^{N} \sigma_{j,H}^{-\eta}} \tag{9}$$

The advantage of inverse weights is that they reflect the recent hedging model capacity. In our empirical analysis we also consider as a benchmark the simpler strategy of assigning equal weights (*1/N*) to each of the *N* hedge ratio forecasts. This approach still provides insurance against model misspecification and instability but does not allow for structural changes as the weights are constant (Kilian and Baumeister, 2015); still, this issue is mitigated as our estimates of $\sigma_{i,H}^2$ are based on rolling estimates of fixed window length *T*.

To implement the forecast combination procedure outlined above we consider ten candidate models that can be classified in three categories: regression, regime switching and GARCH based hedges. Our pool of models, includes:

(I-II): *Naïve* 1:1 benchmark hedge ($\gamma = 1$) and standard *OLS* (γ of Eq. 2)

(III-IV): *OLS w seas* and *OLS w basis* models which condition the hedge ratio to seasonal fluctuations and the lagged basis, respectively:

$$\Delta S_t = \mu + \left(\gamma_0 + \gamma_1 \sin\left(\frac{2\pi t}{freq}\right) + +\gamma_2 \cos\left(\frac{2\pi t}{freq}\right)\right)\Delta F_t + u_t;\ u_t \sim iid(0, \sigma^2) \quad (10)$$

$$\Delta S_t = \mu + (\gamma_0 + \gamma_1 Basis_{t-1})\Delta F_t + u_t;\ u_t \sim iid(0, \sigma^2) \quad (11)$$

(V-VI): The *Vector Error Correction Model* (*VECM*) (see Johansen, 1988). Mathematically:

$$\Delta \mathbf{X}_t = \sum_{i=1}^{p} \mathbf{\Gamma} \Delta \mathbf{X}_{t-1} + \mathbf{\Pi} \mathbf{X}_t + \mathbf{u}_t;\ \mathbf{u}_t = \begin{pmatrix} u_{s,t} \\ u_{f,t} \end{pmatrix} | \Omega_{t-1} \sim \mathrm{IN}(0, \boldsymbol{\Sigma}) \quad (12)$$

where, $\mathbf{X}_t = (S_t, F_t)'$ the vector of spot and futures prices at time *t*; $\mathbf{\Gamma}$ and $\mathbf{\Pi}$ 2x2 coefficient matrices measuring the short- and long-run adjustment of the system to changes in $\mathbf{X}_t$ respectively; $\mathbf{u}_t = (u_{s,t},\ u_{f,t})'$ Gaussian white noise processes with covariance matrix $\boldsymbol{\Sigma}$. The OHR is obtained directly from the second moments (see Eq. 1). Restrictions of the VECM to *Vector Auto-Regressive process* (*VAR*, no error-correction term) is also considered.

(VII-VIII): The Markov Regime Switching (MRS) model (γ_{st}; state dependent). This class of models treats regimes as latent variables and are estimated together with model parameters by maximum likelihood. Let the unobserved state variable $s_t = \{1, 2\}$ follow a two-state first order Markov process. Transition probabilities are assumed either constant $Pr(s_t = i|s_{t-1} = j) = p_{ij}$, (*MRS*) or a logistic function of the basis (*MRS w basis*), i.e., $p_{ij,t} = \left[1 + \exp\left(\varphi_{0,\mathrm{st}} + \varphi_{1,\mathrm{st}}Basis_{t-1}\right)\right]^{-1}$. Allowing Eq. (2) to switch stochastically yields:

$$\Delta S_t = \mu_{s_t} + \gamma_{s_t}\Delta F_t + u_t;\ u_t \sim iid(0, \sigma_{s_t}^2) \tag{13}$$

The regime switching model generates two state-dependent hedge ratios (γ_1 and γ_2). The OHR, at any point, in time is $\gamma_t = \pi_{1,t}\gamma_1 + (1 - \pi_{1,t})\gamma_2$. $\pi_{1,t}$ is the conditional probability that the process will be in a given state at any point in time.

(IX)-(X): The Constant (CC; Bollerslev, 1990) and Dynamic Conditional Correlation (DC; Engle, 2002) models. The conditional variances of spot and futures returns $\Delta X_{i,t}$ follow a GARCH (1,1) as $\sigma_{i,t}^2 = \omega_i + \alpha_i(\Delta X_{i,t-1} - \mu_{i,t-1})^2 + \beta_i\sigma_{i,t-1}^2$; with $\mu_{i,t} = (\mu_{\Delta St}, \mu_{\Delta Ft})$ the conditional means. Based on the partition of the of spot and futures variance-covariance matrix $\boldsymbol{\Sigma}_t$ (see Bollerslev, 1990), $\boldsymbol{\Sigma_t} = \boldsymbol{D_t P_t D_t}$, with $\boldsymbol{D_t} = \mathrm{diag}\left(\sigma_{i,t}\right)$ the $2x2$ diagonal matrix of volatilities and $\boldsymbol{P_t} = \left[\rho_{ij,t}\right]$ a positive definite correlation matrix with $\rho_{ii,t} = 1$, for *i* = 1 (spot; S), 2 (futures; F) for every *t*. The *DC-GARCH* model correlation structure can be expressed as:

$$\boldsymbol{P_t} = (diag\{\boldsymbol{Q_t}\})^{-1/2}Q_t(diag\{\boldsymbol{Q_t}\})^{-1/2},$$
$$\boldsymbol{Q_t} = (1 - a - \beta)\overline{\boldsymbol{P}} + a\boldsymbol{z_{t-1}z'_{t-1}} + \beta\mathbf{Q_{t-1}}, \tag{14}$$

where, $\boldsymbol{Q_t}$ a $2x2$ symmetric positive-definite covariance matrix; $\boldsymbol{z_t} = \boldsymbol{D_t^{-1}u_t} \sim N(0, \boldsymbol{P_t})$ the standardized residuals. Then, correlation is simply expressed as $\rho_{sf,t} = q_{sf,t}/\sqrt{q_{s,t}q_{f,t}}$ with

$q_{SF,t} = (1 - a - \beta)\bar{\rho}_{sf} + a(z_{s,t-1}z_{f,t-1}) + \beta(q_{sf,t-1})$. For $a = \beta = 0$ the model is equivalent to the *CC-GARCH* model. Parameters are estimated via maximum likelihood.

3. Data Description and Preliminary Analysis

Our empirical work employs data from the two most liquid and mature natural gas markets, Henry Hub in the U.S. and NBP in the U.K. In North America, gas markets started to liberalize in the 1970s; almost 50 years later, Henry Hub serves as a very successful benchmark hub. The U.S. natural gas (NG) contract was introduced in 1990. It is the third-largest physical commodity futures contract in the world by volume. For the European continent, we limit our analysis to the NBP. The British hub was at the forefront of European gas market development, with a liberalized and mature traded market (Heather, 2012); it also provides a relatively data rich environment.[4] The U.K. natural gas futures (NBP) contract, launched in 1997. Futures negotiated at the ICE represent more than one-third of gas negotiated at NBP. Our dataset consists of daily futures and physical prices from December 30, 1998 to May 16, 2018, which are then converted to 1,012 weekly (mid-week, i.e., Wednesday prices) and 234 monthly (mid-month) observations. The futures prices are for the contract which is closer to maturity where volume and liquidity is higher. All series are obtained from Datastream.

Panel A of Table 1 reports summary statistics for the (log) spot and futures returns. The unconditional weekly and monthly return of distributions are non-normal, as evidenced by the positive skewness and high excess kurtosis. The Ljung and Box (1978) *Q* statistic on the first six-lags of the sample returns' autocorrelation function is overall significant. Same holds for the *Q* statistic on squared returns for all weekly series', but on monthly data the evidence is weak (ARCH test). Tests for unit root (Phillips and Perron, 1988) reveal that the

[4] Traded volumes of gas on the main European hubs – such as TTF (established in 2003) and Zeebrugge (ZEE in 1998), have increased substantially just after 2005 and were developed broadly from 2009 (Heather, 2012).

spot and futures prices are integrated of order 1, *I(1)*. Next, in Panel B cointegration is examined through the λ_{max} and λ_{trace} statistics which test for the rank of Π (Eq. 12) to determine the number of cointegrating relationships (Johansen, 1988). Both the maximum eigenvalue and the trace test statistics confirm one cointegrating relationship between the spot and futures prices. Finally, likelihood ratio tests on the hypothesis that there is a one-to-one relationship between spot and futures prices, i.e., restrictions $(\beta_1 = -1)$ and $(\beta_0 = 0$ and $\beta_1 = -1)$ invariably give that cointegrating vectors can be reduced to (1, -1, β_0), i.e., the basis.

Panels A and B of Table 2 summarize the dynamic hedge ratio model estimates, for both weekly and monthly frequencies: OLS w seas, OLS w basis, MRS, MRS w basis, CC- and DC-GARCH. First, the coefficients of OLS w seas indicate strong seasonal patterns in the estimated hedge ratios. Second, basis variations (OLS w basis) are significant only in the NG market. Third, for MRS models, hedge ratios in state 2 are relatively closer to one, apart from the monthly variations of NBP where they appear less distinct. Transition probabilities are relatively low indicating that regimes are persistent; time-variation in transition probabilities (MRS w basis) is significant only in the low volatility state (with the exception of NPB monthly variations). Finally, all ARCH and GARCH parameters are significant with $\alpha_i + \beta_i$ very close to unity. For the correlation processes (DC-GARCH), $a + \beta$, are also significant and correlations are also persistent with $a + \beta > 0.82$. Average weekly-monthly correlations (CC-GARCH) are 62-86% and 68-88% for NG and NBP, respectively.

4. Empirical results

This section discusses the empirical results for both in-sample (IS) and out-of-sample (OS) analysis. Note that, the postulated method for combining the hedge ratios of models (Section 2.1) is restricted to the OS exercise as the implementation procedure implies that the computation of the weighting structure is based on the past IS performance of each model.

4.1. In-Sample (IS) Hedging Performance

The setup of our IS analysis is as follows. We use data covering the period December 30 1998 to May 16, 2018; this results in 1,011 weekly (233 monthly) return observations. We then construct the minimum-variance portfolios of spot and futures and, given the optimized hedge ratios, we calculate returns on the portfolio for a holding period of one week (month). Table 3 presents the NG (Panel A) and NBP (Panel B) IS variance reduction against the unhedged position for weekly and monthly horizons using either the traditional measure of hedging effectiveness HE_1 (Eq. 3) or the Edernigton and Salas (2008) HE_2 (Eq. 5). Ten single-model hedging strategies are considered: naïve, OLS, OLS w seas, OLS w basis, VAR, VECM, MRS, MRS w basis, CC- and DC-GARCH.

For NG, the weekly-horizon HE_1 of the individual models ranges between 34.55 - 40.20% while for HE_2 the figures improve to 45.26 - 50.97%. Monthly hedging horizon yields improved HE consistent with the existing literature (e.g. see Hanly, 2017); this can relate to the finding that low frequencies are associated with higher correlations and smaller deviations from the normal distribution. In particular, for monthly variations, HE_1 lies within 73.28 - 76.14% while HE_2 figures slightly improve to 76.49 - 78.91%. Therefore, the traditional HE underestimates the risk reduction potential of hedging; in line with Martínez and Torró (2015) or Edenrigton and Salas (2008). Similar results can be obtained for NBP. The weekly-horizon single-model HE_1 is within 41.99 - 47.21% while for HE_2 this is 47.94 - 51.67%. Moreover, switching to monthly horizon yields better HE, i.e., 72.41 - 78.9%.

Across our experiments, no model consistently ranks as best or worst, which confirms the conjecture that a single model does not work equally well at all times, frequencies and performance measures. Examples are: (i) OLS w seas ranks first in the NBP but not the NG market, (ii) despite DC-GARCH generates the second highest HE_1 for NG weekly horizon, the model is not even in the best five strategies for NBP, (iii) DC-GARCH outperforms naive hedges in terms of NBP HE_1 but not HE_2, and (iv) VAR outperforms weekly naïve hedges but not monthly. The above results are not surprising since various studies report mixed consensus about the superiority of different models. This can be attributed to estimation and misspecification errors, among others, as more complex strategies do not necessarily perform better than simpler ones, as any overfitting may be penalized. Moreover, a particular formulation, may not be sufficient to provide consistent gains across different time periods and market conditions (see, e.g., Baumeister and Kilian, 2015). As such, it is not uncommon for a more sophisticated approach to prove inferior in terms of, e.g., HE.[5]

We next examine risk reduction when combining hedged portfolios. The equal weight scheme ($1/N$) produces very similar results to the best performing single-model, having a HE of just 40 basis points (bps) lower, on average. Yet, there might be opportunities to better exploit the information content of the estimated hedge ratios using combinations that optimize the weight structure. For this reason, we calculate the global minimum-variance portfolio with nonnegative weights (GMV $w \geq 0$), which can reveal the potential that model OHR combination schemes with $w \geq 0$ might achieve. Collectively, this leads to an average improvement in excess of 180 basis points (bps); 3-660 bps higher compared to the highest achieved HE. When allowing for negative combination weights (GMV_{uncon}), the gains over the GMV $w \geq 0$, deviate from less than 1 to almost 100 bps. For comparison, we also report the performance of a model that combines VECM with MRS dynamics in a DC-GARCH specification (henceforth, MRS-GARCH[6]). Though MRS-GARCH outperforms the single

[5] For instance, occasionally MRS are found sensitive in forecasting frameworks, attributed to parameter instability between IS-OS periods as well as uncertainty regarding the unobserved regime (Engle, 1994).

[6] Estimation results for the MRS-GARCH are available from the authors upon request. For technical details, we refer to Lee and Yoder (2007) and Alizadeh *et al.* (2008). Note also that, the latter model is not included the calculation of neither *1/N* nor *GMV* or the forecast combination methods in the OS experiments because first, it encompasses market features already accounted for, and second, computational simplicity is lost, reducing thus the practicality of our approach. This model involves estimation of the VECM, plus 16 parameters for the

regime GARCH models, it fails to improve on the simple MRS specification in the majority of cases. Moreover, *1/N* and *GMV* hedging strategies do a better job than MRS-GARCH.

For illustration purposes, Figure 2 plots the OHRs from the OLS w seas, OLS w basis, MRS w basis, DC-GARCH models (left panel); together with the OHR obtained from the OLS of Eq. (2) and the conditional composite OHRs derived using equal weights (*1/N*) and GMV$(w \geq 0)$ (right panel). *1/N* OHRs are smoother than the time-varying hedges - indicative of relative stability – while GMV$(w \geq 0)$ OHRs are more volatile. Evidently, models that do well IS are allocated more weight. For instance, the strong seasonal pattern of NBP gas OHR is clearly depicted in both weekly and monthly hedges; note that OLS w seas produces the best IS HE among the individual models for NBP (see Table 3). Therefore, we can see the potential of forecast combination methods to produce hedge ratios that follow diverse dynamics; which are nevertheless obtained from the individual models. As hedge ratio time-variation depends on model performance, in a rolling estimation scheme there might be periods of seasonal hedge ratios and periods that dynamics may be driven mostly by GARCH or MRS, etc.

So far, findings suggest that hedge ratio combinations lead to reliable and consistent hedging decisions. Besides, different HE measures, frequencies and markets, rank a given set of models differently as a hedging model is only able to capture distinct aspects of the market. The above strengthens our argument for utilizing forecast combination methods.

4.2. Out-Sample (OS) Hedging Performance

IS testing does not reflect real time forecasting power. Since market participants are more concerned with how well strategies perform ex-ante (OS) testing is a more realistic evaluation of actual HE. The setup of our OS experiments is as follows. Initially, data

---

variance-correlation dynamics, plus 2 parameters for the transition probabilities; in addition to implementing a *collapsing procedure* to solve the *path-dependency problem* (Lee and Yoder, 2007) covering the period from December 1998 to May 2013 are used to estimate the parameters of the hedging models; 751 weekly (173 monthly) return observations. We then forecast the OHR one-step ahead. Given the 1-week (month) forecasts we compute realized returns on the portfolio for a holding period of one week (month). Using a rolling window forecasting scheme, this exercise produces 260 weekly (60 monthly) test observations that cover a period of 5 years, from May 29 2013 (June 15, 2013) to May 16, 2018 (May 15, 2018).

Tables 4 and 5 summarize the OS HE of the forecast models and forecast combinations (FC) across different frequencies (weekly and monthly in Tables 4 and 5, respectively). Similar to IS results, judged on consistency across Tables 4 and 5, single-model forecasts produce fairly unstable results and ranking is indistinguishable among the models. In contrast, the performance of the *ad hoc 1/N* OHR is genuinely stable, generating a HE close to the highest performer (lower by 40 bps, on average), while in 2 cases it performs even better.

Although the *1/N* strategy has been sometimes shown to provide consistently more accurate forecasts compared to more sophisticated techniques (Baumeister and Kilian, 2015), equal weights throughout time are unlikely to offer the best available option. Tables 4 and 5 show that all combination schemes are able to produce higher HE relative to *1/N*. Several events and market dynamics could give rise to time variations in the combination weights; such as economic fundamentals (demand or supply shocks, shifts from contango to backwardation periods and vice versa), seasonality effects and differences in the speed of some models to respond to changing market conditions. The forecast combination potential gain is depicted in the last row of each table. In the NG (NBP) market OHR forecast combination can offer benefits in excess of 92 (2) bps; 251 (36) bps on average. Noticeably, in more than 80% of the 248 (= 31 comb. x 2 HE x 2 markets x 2 horizons) cases, combinations deliver higher HE than the single-models' maximum HE. In the remaining cases, forecast combinations typically deliver the second or third highest HE; with a downside potential no lower than the top-four (only two entries in Table 5; NBP monthly HE_2). So even if some forecast combinations do not generate the best hedge ratio forecasts, they tend to not deliver poor performance and thus, represent a risk averse choice in terms of model uncertainty. As for the MRS-GARCH, we can see that, although in some cases it outperforms the pool of candidate models, there is no improvement over the combination approaches.

To discount the possibility that HE of some model may be due to data snooping bias we employ Hansen's (2005) reality check. Different strategies may produce satisfactory results purely due to chance or due to the use of posterior information rather than the superior ability of the competing strategies (Sullivan *et al*., 1999). We construct a loss function (*Pf*) given as:

$$Pf_i = \left(\Delta S_t - \gamma_{i,t}\Delta F_t\right)^2 - \left(\Delta S_t - \gamma_{benchmark,t}\Delta F_t\right)^2 \quad (15)$$

If the i^{th} forecast combination outperforms the benchmark(s) single-model(s), the expected value of the performance measure will be negative, i.e., $H_0: E(Pf_i) \leq 0$. The loss function of Eq. (15) is constructed using the simulated portfolio returns generated by the stationary bootstrap of Politis and Romano (1994).[7] Tables 4 and 5 report two summary information measures for pairwise comparisons: (i) the average *p*-value (given ξ and η) with each of the ten benchmark models, and (ii) whether this combination method is able to significantly outperform at least seven out of the ten benchmarks.

In aggregate, the majority of combinations generate significant improvement in the

[7] Politis and Romano (1994) re-samples blocks of varying length from the original data, where the block length follows a geometric distribution; for more technical details, we refer to Politis and Romano (1994) and Sullivan *et al*. (1999, Appendix C). Our tests use 10,000 bootstrap simulations.

HE at conventional significance levels; HE significance is achieved in more than 75% of all cases considered; either in terms of average p-value (83% for U.S. and 67% for U.K.), or by outperforming at least 7 out of the 10 benchmarks (79% for U.S. and 71% for U.K.). Also, *1/N* strategy achieves significant improvement in the weekly U.S. and (HE_1) and U.K. (HE_2) cases, at the 10% significance level, as well as for U.S. monthly (HE_1), at 5% level. Therefore, combination methods result in superior forecasts, i.e., model instabilities of unknown form can be mitigated by diversification gains across forecast methods, which act as insurance against instability, misspecification or structural changes (Pesaran and Timmermann, 2007).

Finally, Figure 3 portrays the excess HE (vs. the average single-model performance) achieved by combination methods (in terms of bps) when perturbing the discount factor (ξ) and the tuning parameter (η) beyond the assumed cases in Tables 4 and 5, i.e., within the interval of (0,250) and (0,50), respectively. Both parameters have a diverse impact, depending on the market and horizon considered. In general, decisions based on the forecast combinations provide reasonable benefits; when considering a single hedge ratio model, the decision maker forgoes HE benefits in excess of 100-500 bps. Most importantly, we see that, for the most part, benefits are consistent, irrespective of the choice on ξ or η. Therefore, market participants should not rely on some particular econometric specification when estimating the OHR which might not fully reflect the risk inherent in price movements.

4.2.1.Stability in Combination Weights and Hedge Ratios

Several authors have documented that a simple equally weighted pooling of forecasts is often found to outperform more sophisticated approaches due to the associated estimation error of both weights and model parameters (see, for example, De Menezes *et al*., 2000; Genre *et al*., 2013). Thus, imposing equal weights increases robustness with respect to model uncertainty, parameter instability, and estimation errors (see also Palm and Zellner, 1992; Diebold, 1989; Caldeira *et al*., 2017, among others). Despite this does not hold in our case based on our results (see Tables 4 and 5), further investigation into the evolution and stability of the combination weights is necessary to better evince the obtained results.

To provide some estimates, the standard deviation of weekly combination weights, with $\xi = \{0,0.5,1,3,10,20\}$ and $\eta = \{0,1,2,3,6,10\}$ ranges between 0.02% - 13.6% (0.06% - 8%) for the U.S. (U.K.) weekly horizon. For monthly figures the corresponding interval lies in the region 0.03% - 24.5% (0.04% - 25.2%). Weekly (monthly) weight standard deviation of "Best FC" strategies is in the range of 2% - 22.7% and 0.3% - 3.7% (0.01% - 4.4% and 0.02% - 0.25%) for the U.S. and U.K. markets, respectively.

Figure 4 presents some additional information on the forecast combination weights. For brevity, we present the time-variation in the "Best FC" weights only (see also Tables 4 and 5). The subplots show that weights exhibit variations of mostly transient nature, reverting back to a long-run mean; this is also confirmed via unit root testing. In particular, across both natural gas markets, irrespective of ξ and η, out of the 2,480 estimated time series for weights (31 comb. of ξ and η x 2 HE x 2 markets x 2 horizons x 10 models) 88.5% (97% for weekly and 80% for monthly data) of the time series are integrated of order 0 at 5% significance level, and 91% (98% for weekly and 84% for monthly data) at the 10% level.

To further understand the structure of combination weights, we appeal to panel regression techniques. We first fix a hedging model N and let $\Delta w_{Ni,t}$ denote the weight change on day t of N^{th} hedging model (e.g. Naïve, …, DC-GARCH) using forecast combination parameters $i = (\xi, \eta)$ reported in Tables 4 and 5 incl. Best FC (excl. 1/N, as for the latter weights are constant) for any t and as extracted for both HE measures, HE_1 and HE_2. That is, i = 1, 2, …, 62 stacked time series of weights. We include as exogenous variables changes in the S&P 500 implied volatility index (VIX) and crude oil implied volatility index (OVX); variables which are common across U.S. and U.K. and are assumed to affect the two markets homogeneously. We also employ the news-based economic policy uncertainty (EPU) indices for U.S. and U.K. (see Baker *et al*., 2016) and the magnitude of the futures-spot spreads (|Basis|) which vary across U.S. and U.K. reflecting country-specific (EPU) and gas contract-specific (|Basis|) effects. All time series are collected from Datastream. Finally, we also include a dummy variable which identifies elevated general uncertainty, i.e., when all variables VIX, OVX, EPU, |Basis| are in the upper tertile of their empirical distributions. Mathematically, the econometric equation is a fixed-effects balanced panel equation, for each hedging model N:

$$\Delta w_{Ni,t} = \sum_{p=1}^{2} a_{1N,i,p} \Delta w_{N,i,t-p} + \sum_{q=0}^{2} a_{2N,i,p} \Delta VIX_{i,t-q} + \sum_{q=0}^{2} a_{3N,i,p} \Delta OVX_{i,t-q} + \sum_{q=0}^{2} a_{4N,i,p} \Delta EPU_{i,t-q} + \sum_{q=0}^{2} a_{5N,i,p} |Basis|_{i,t-q} + \sum_{q=0}^{2} a_{6N,i,p} Dummy_{i,t-q} + \omega_{N,i} + e_{N,i,t} \tag{16}$$

Results obtained under the least squares dummy variable (LSDV) estimator for fixed-effects models. The fixed effect for N^{th} model and combination parameters i is represented by $\omega_{N,i}$. The multiplier form of the model can be written more compactly by inverting the above equation as $\Delta w_{N,i,t} = \Phi(L)^{-1}\Theta(L)X_{N,i,t} + \Phi(L)^{-1}e_{N,i,t}$; L denotes the lag operator. The mean responses from the variables are therefore captured by the lag polynomial $\Psi(L) = \Phi(L)^{-1}\Theta(L)$. We measure the responses up to three periods and their sum (cumulative).

Results of the above augmented autoregressive model (ARX), are presented in Figure 5. The dynamic effects on combination weight changes, estimated separately for each model N. Note that the figure displays only responses/cum. responses that are significant at least at 5% level. Independent variables are all standardised for ease of interpretation. Overall, models with positive responses (either point responses or cumulative) are those that perform better (positive impact on combination weight; which in turn is computed based on IS HE) when the independent variables increase. For example, an increase in the magnitude of the absolute basis is associated with overall negative impact on the weight assigned to all models apart from GARCH (3rd row of the Figure), indicating that GARCH can potentially perform better when futures and spot prices deviate from each other; since weights are directly estimated from the variances of hedged portfolios and are, thus, directly linked to HE. In particular, an increase in the magnitude of the NG basis by one SD for weekly (monthly) variations is associated with a cumulative response on the weight assigned to DC-GARCH by 80 basis points or 0.8% (1.5%) while the maximum noted response is close to 1.5% per week (1.2% per month). For NBP weekly, the cumulative response is statistically zero; that is, basis shocks are expected to die out fast, yet, there is a statistically significant response of 0.04% per week. Concerning the monthly figures, an increase in the magnitude of the NBP basis by one SD is associated with a cumulative response on the weight assigned to DC-GARCH by 0.4% while the maximum noted response is slightly higher than 0.5% per month.

It must be noted that, as the resulting model combination weights lead to an actual hedge ratio, even if aggressive rebalancing of the hedged positions tends to improve, HE it may yield portfolios that require expensive trading due to transaction costs (TC). To study the cost of rebalancing strategies implied by the different prediction models, under the assumption of the same linear relationship between TC and the size of the rebalancing required for hedging instruments, we use the absolute changes in the dynamic hedge ratios as a proxy for TC (see Chen and Sutcliffe, 2012) throughout the OS period. We calculate $\overline{\mathrm{PT}} = \mathrm{T}^{-1}\sum_{t=1}^{\mathrm{T}}|\gamma_{i,t+1} - \gamma_{i,t}|$ which reflects the impact of weight instability on the composite hedge ratio (portfolio turnover); the fraction (in percentage terms) of the futures position that need to be liquidated/reallocated (Greyserman *et al*., 2006).

Results are presented in Table 6. In Panel A, it appears that the GARCH specifications require a higher proportion of the hedged portfolio to be restructured at each rebalancing point which imposes higher TC; their mean PT values lie between 7% - 12% per week and 8% - 22% per month depending on the hedging horizon. On the other hand, OLS, VAR and VECM are the least expensive strategies with mean values less than 0.5%, irrespective of the horizon. We find that the mean PT of 1/*N* strategy is always less than GARCH models, as are also both FC_{avg} and Best FC. Given the nature of the problem (trading in futures only) transaction costs as a percentage of futures contracts that need to be liquidated/reallocated appear reasonable.

To understand whether TCs are compensated for in forecast combination strategies we also compute, in Panel B of Table 6, the ratio of the incremental HE of HE_{BestFC} [HE_{FCavg}] relative to the increase in PT, if any; as a measure of benefit-cost trade-off for the hedged portfolios. Note that $HE_{BestFC} > HE_{FCavg} > HE_{model_i}$. As such, incremental HE is always positive and negative figures simply imply that TC are actually reduced compared to the benchmark. With respect to the Naïve zero-PT strategy improvement is a 0.2% to 1.7% (0.2% to 2.5%) increase in $HE_{BestFC}(HE_{FCavg})$ for every 1% increase in PT. Concerning the volatile GARCH-based hedge ratios, combination strategies not only improve HE but reduce PT as well. Therefore, the proposed hedging approaches act not only as a hedge to unstable hedge ratios but offer reasonable benefits versus more stable strategies. It is worth noting that, in some cases, the resulting combination hedge ratio is even more stable than the one implied by the 1/*N* strategy.

Furthermore, to quantify the value of model timing in a combination setting, we follow Fleming *et al*. (2001) and compare the forecast combination dynamic strategies to that of individual hedging-models. Denote $r^{*}_{p,t+1}$ the gross hedged portfolio return constructed using the realized returns from a certain model and $r_{p,t+1}$ a benchmark's return. Our evaluation focuses on the fee, Φ, an investor with quadratic utility and degree of relative-risk aversion δ is willing to pay[8] for switching from some modelling strategy to the forecast combination approaches. This is equivalent to finding the value of Φ that satisfies:

$$\sum_{t=0}^{T}\left\{\left(r_{p,t+1}^{*}-\Phi\right)-\frac{\delta}{2(1+\delta)}\left(r_{p,t+1}^{*}-\Phi\right)^{2}\right\}=\sum_{t=0}^{T}\left\{r_{p,t+1}-\frac{\delta}{2(1+\delta)}\left(r_{p,t+1}\right)^{2}\right\} \quad (17)$$

Performance fees are reported in Table 6, Panel C in annual basis points (bps). We set δ = 6, which reflects an investor with moderate risk aversion. Forecast combination switching fees range, on average, between 486 to 4,828 bps per annum. Combination approaches achieve on average a minimum fee of more than 7.6% p.a. with the potential to be in excess of 30% p.a. That is, an investor should be willing to pay more than 7.6% overall to switch to this strategy. In the absence of TC this can be interpreted as a management fee. If TCs are present, it implies that, in comparing the FC strategies with model i, an investor who pays (additional) TC lower than the performance fee will prefer the FC strategy.

Assuming proportional TC paid every time the futures position is rebalanced, if the (extra) total TC was less than 0.14% (=7.6/52) per week or 0.63% per month (=7.6/12), the investor would still be better off using composite hedge ratios. Note that only for the Naïve strategy, the hedge ratio is constant. In this case, if the total TC was less than 0.6% (0.1%) per week for the NG (NBP), the investor would be better off, with the potential for this figure to go as high as 1.14% = 5914/52 (0.27% = 1414/52). For monthly hedging horizon this figure changes to less than 0.38% (0.06%) per month for the NG (NBP), with the potential for this figure to go as high as 3.11% (0.25%). In commodity markets Locke and Venkatesh (1997) estimate that futures trading costs range between 0.0004% and 0.033% of notional value, while Fuertes *et al*., (2015) take a more conservative view, imposing TCs of 0.033% -

[8] Let W be the investor's wealth, r_p the gross portfolio return and λ an absolute relative risk aversion coefficient. The investor's realized utility in period $t+1$ can be written as $U(W_{t+1}) = W_t r_{p,t+1} - 0.5\lambda W_t^2 \left(r_{p,t+1}\right)^2$. To estimate expected utility, $\overline{U}(\cdot)$, generated by a given level of the initial wealth W_0, we hold $\delta_t = \lambda W_t/(1-\lambda W_t)$ equal to a fixed value δ, so that $\overline{U}(\cdot) = W_0\left(\sum_{t=1}^{T} r_{p,t+1} - 0.5\delta(1+\delta)^{-1}\left(r_{p,t+1}\right)^2\right)$.

0.066% per trade.

Therefore, in addition to the economic value associated with higher HE, combination strategies yield, on average, positive fees (gains in terms of utility improvement), concluding that composite hedge ratios work well vs. both individual models and the *1/N* scheme. This implies that, even after accounting for TCs, and even though the percentage variance reduction might not be large, the composite hedge would evince utility gains. Similar conclusions are found when δ increases to 10 or decreases to 4 (results are available upon request).

4.2.2. Performance Under Different Market Conditions

The risk-return profile of energy prices changes fundamentally across periods. For example, it has been well recognized that in periods of low stocks, positive demand shocks cannot be absorbed by storage and spot prices are likely to exceed the futures prices (backwardation); conversely, abundant inventories provide a buffering effect against shifts in demand, and the basis is likely to be positive (contango). Another typical feature of natural gas prices is that spot price changes are partially predictable; due to demand cycles, weather and storage seasonal patterns. Martínez and Torró (2015) find significant differences during winter and summer seasons in the natural gas basis, spot and futures returns mean and volatility. In the winter, demand inelasticity and higher marginal cost of production make active storage management less flexible to absorb demand shocks. In this context, model performance may vary when market conditions change, either in terms of market volatility and disequilibria (Nomikos and Pouliasis, 2015), volatility forecast accuracy (Nomikos and Pouliasis, 2011) or HE (see Chang *et al*., 2010).

Based on the above arguments, we divide the hedging horizon into backwardation/contango and winter/summer periods. Our definition of backwardation (contango) market is short-term; when the nearest to expiry futures contract is less (greater) than the spot price. The sample identifies 149 (110) backwardation, and 111 (150) contango weekly periods in the NG (NBP) market. For the monthly frequency, this corresponds to 32 and 28 (24 and 36), respectively. All HE results are presented in Table 7. For brevity, we report the minimum, maximum and average HE value of the 30 combinations for $\xi = \{0, 0.5, 1, 3, 10, 20\}$ and $\eta = \{1, 2, 3, 6, 10\}$.

For NBP, HE is higher in periods of backwardation and during winter (October – March), irrespective of the hedging horizon. On average, across models, backwardation HE is more than 880 bps in excess of contango HE, while winter HE is more than 370 bps in excess of summer (April – September) HE. Asymmetries are more pronounced for NG, yet, results are mixed. For monthly horizon, HE is higher in backwardation and during winter (as for NBP) by an average of 790 bps and 150 bps in excess. However, for weekly hedges, contango and summer HE figures exceed the corresponding backwardation and winter HE by 4,000 bps. Overall, the risk-return profile of natural gas does not only vary across market states but is also market-specific.[9] An important practical implication from Table 7 is that economic agents would benefit more by timing the market and dynamically altering their strategies. For instance, hedgers in the U.S. market with weekly horizon (HE_1), may adjust the hedges according to the OHRs derived from the OLS w basis model in backwardation, MRS model in winter and DC-GARCH model in both, summer and contango.

1/N hedges, perform better than the single-model hedges in more than 60% of the table entries. For NBP, *1/N* is not significantly better than the benchmarks whereas for NG, the recorded *p*-values provide some evidence of significance. The *1/N* strategy generates an average HE of 120 bps over the average HE of the individual models. The other forecast

[9] For example, a weekly naïve hedger of NG (NBP), could have achieved, an out-of-sample HE of to 21.79% (36.71%). However, depending on whether the market was in backwardation or contango, the HE figure for NG (NBP) varies from 16.14% (26.63%) to 53.01% (21.03%). Also, depending on whether the market was in the winter or summer, the HE figure varies from 17% (36.45%) to 63.74% (36.74%).

combinations achieve statistically higher HE (FC_{maxHE}) in almost all cases. For the considered ξ and η, strategies have the potential to increase HE by 100 bps compared to the best single-models (FC_{maxHE}), i.e., more than 320 bps over the average HE of the individual models.

We also report the average achieved HE of forecast combination methods (FC_{avg}). In this respect, FC_{avg} still produces a better HE than the single-model top performers, i.e., 20 bps more on average. This is 245 bps over the average HE of the individual models; with this improvement being statistically significant in most cases. Interestingly, even the worst of the combinations (FC_{minHE}) yields a HE of approx. more than 130 bps over the average HE of the single-model strategies. Still, the improvement is significant in almost half of the cases shown whereas in few instances this is supported by the average *p*-value of the reality checks.

Our findings corroborate that in the majority of cases, regardless of the market conditions and/or seasons, such strategies lead to significant improvements in HE. However, it is worth emphasizing that the core advantage of the proposed framework is not necessarily to guarantee HE gains over the best individual hedging model. Instead, the key practicality of such methods is that they provide insurance by circumventing breakdowns and model risk.[10] As with all insurance, there might be in some cases a rational cost, that is, the potential loss compared with the single-most effective model (see also Baumeister and Kilian, 2015).

4.2.3. Additional Results: Short vs. Long Hedgers

The analysis so far presents a complete picture of what forecast combinations can achieve in terms of OS HE and whether a hedging strategy is consistent across market conditions. Nonetheless, the HE measures employed do not distinguish between long and short positions (variance assigns the same weight to positive and negative outcomes). To

[10] For example, in the NG U.S. (Table 7), although the DC-GARCH model provides the best HE_1, under the supposition that spot prices can be partially predicted, the same model provides the worst HE_2. Likewise, for NBP U.K., although the 1:1 hedge ratio provides the best HE1 for weekly horizon under contango, the same model provides the worst performance for monthly horizon during summer.

overcome this impediment, tail risk metrics serve as desirable alternatives of how we think about risk (see Alizadeh *et al*., 2008; Hanly, 2017). In particular, downside risk considers the relevant distribution of returns below some specific target return, θ. To remove the effect of upside gains in the estimation of risk, we formulate HE based on Lower Partial Moments (LPMs). Let $F(\cdot)$ denote the distribution function of realized portfolio returns ΔH_t. The k^{th} order LPM is:

$$LPM_k(\theta; \Delta H_t) = E\{(max[0, \theta - \Delta H_t])^k\} = \int_{-\infty}^{\theta} (\theta - \Delta H_t)^k dF(\Delta H_t) \quad (18)$$

From a downside risk management perspective, hedging should aim at avoiding negative outcomes, i.e., to distinguish between positive and negative ΔH_t, θ is set to zero. We consider $k = \{2, 3\}$. k reflects the weight attached to the shortfalls, i.e., higher k reflects higher risk-aversion.[11] Compared to variance, LPMs do not require restrictive assumptions about distributional properties or investor preferences, and reveals information on the asymmetry of the joint distribution of spot and futures returns (separating short to long hedgers).

Results for NG (Panel A) and NBP (Panel B) are summarized in Table 8. A short (long) hedging position is equivalent to selling (purchasing) futures contracts against the purchase of natural gas. Evidently, there are marked asymmetries for short and long hedgers, in terms of adjusted HE (HE using LPMs, henceforth HE^*). For NBP, it appears that HE^* is higher for consumers (buyers) of gas, irrespective of the hedging horizon and LPM order. Results are mixed for NG, with monthly (weekly) hedging horizon long (short) hedges producing higher HE^*. Clearly, hedgers have to alter their strategy depending on their profile

[11] See also Lien and Tse (2002). k depends on the degree of investor's risk aversion: risk seeking ($0 < k < 1$), risk neutral ($k = 1$), risk averse ($k > 1$). For $k = 2$ and $\theta = \overline{R_p}$, the LPM is the semi-variance. For normal or symmetric distributions, this would be exactly proportional to variance.

(buyers vs sellers) and based on their degree of risk aversion. For example, short hedger in the U.S. market with weekly horizon (HE^*_1), may use the MRS model which provides the highest HE^* (of the individual models; LPM_2). Yet, for more risk averse hedgers (LPM_3) this is by far not the best strategy, resulting in a HE^* measure of close to 19%, while GARCH hedges' HE^* is more than 30%. Finally, results are market-specific, e.g., regardless of LPM order, the OLS w seas weekly horizon long hedge is the best (worst) alternative, in terms of HE^*, in the NG (NBP) market.

We can see that *1/N* hedges perform better than the single-model strategies in almost all cases, yet, in only few cases *1/N* results significantly better HE^*. On average, a *1/N* strategy will produce a HE^* of approx. 130 bps over the average HE^* of the individual models. Concerning the other forecast combinations, these achieve higher HE^* (FC_{maxHE}) which is also statistically significant in the majority of cases, that is, either against of at least 7 benchmarks or in terms of average *p*-value. As shown by FC_{maxHE}, forecast combinations have the potential to produce a HE of approx. 64 bps more than the best single-model performer, while this is more than 340 bps over the average HE^* of the benchmarks. On average, combination strategies (FC_{avg}) produce a HE^* of 260 bps over the average HE^* of the individual models. Still, improvements offered in excess of single hedge ratio models are found statistically significant in half of the cases considered. Lastly, note that, even the worst of the combination strategies (FC_{minHE}) still leads to a HE of approx. 110 bps over the average HE of the individual models.

5. Conclusion

This paper postulates a novel futures-based risk management framework capable of accommodating alternative hedging strategies, while it can be rectified to fine-tune the importance assigned to the best forecast models. To this end, we develop a parsimonious hedging model based on combination of forecasts that admits cointegration, seasonality, conditional second moments and regime-switching. We note that this is the first systematic empirical study on the economic value of forecast combinations from the perspective of hedging effectiveness. To assess the performance of the combined predictors and the related candidate models, we focus on weekly and monthly commodity risk exposures for the U.S. (NG) and U.K. (NBP) natural gas markets.

Results show that different hedging effectiveness measures, frequencies and markets, lead to diverse model ranking as a hedging model is only able to capture certain market aspects. Not only the combination methods offer significant improvement over the single-model specifications, but in many cases, even the worst of the combination strategies yield more effective hedges. We also investigate several other important issues. First, hedgers of different horizons have dissimilar risk profiles. Second related conjecture is that shocks in spot prices can be partially predicted using the information contained in the basis and omitting this information will lead to underestimated risk reduction. Third, there is asymmetric hedging performance during backwardation and contango periods, as well as winter and summer; suggesting that market agents should not only modify their hedge ratios as implied by the inherent model dynamics, but also switch across models. Finally, using lower partial moments, we note dissimilar hedging performance for short and long hedgers which is market-specific and depends on risk aversion. We provide strong evidence that our forecast combination approach leads to stable and more efficient hedge ratios. In- and out-of-sample tests verify that finding, economically and statistically. The core advantage of the proposed framework is not necessarily to guarantee hedging gains over the best individual hedging model, although this holds in most cases. Rather, such an approach provides insurance against model instability, over-parameterization, breakdowns and structural changes. Overall, we can conclude that our framework generates robust forecasts due to the resulting diversification gains, unlikely to be replicated by individual models. Even if a single model generates a higher hedging effectiveness, this better performance will entail higher likelihood to result in extreme losses.

As this is the first study to comprehensively assess the economic value of forecast combinations in the hedging decision, there is scope to potentially extend our analysis. For example, various studies attempt to incorporate the higher moments (conditional skewness and conditional kurtosis) in asset pricing and portfolio analysis; see, for example, Gao and Nardari (2018). Given the increasing emphasis on risk management, there is a proliferation of measures capturing different types of risk. Creating hedging strategies using alternative risk objectives in the optimization procedure (rather than mean-variance), albeit an important research question, is left for future research.

Data Availability Statement

All data that support the findings of this study are available from Refinitiv (formerly Thomson Reuters) Datastream database. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors with the permission of Refinitiv.

References

Alizadeh, A.H., Nomikos, N.K., Pouliasis, P.K., 2008. A Markov regime switching approach for hedging energy commodities. Journal of Banking and Finance 32, 1970-1983.

Baker, S., Bloom, N., Davis, S., 2016. Measuring economic policy uncertainty. Quarterly Journal of Economics 131, 1598-1636.

Baumeister, C., Kilian, L., 2015. Forecasting the real price of oil in a changing world: A forecast combination approach. Journal of Business and Economic Statistics 33, 338–

351.

Bollerslev, T.P., 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. Review of Economics and Statistics 72, 498-505.

Caldeira, J.F., Moura, G.V., Nogales, F.J., Santos, A.A.P., 2017. Combining multivariate volatility forecasts: An economic-based approach. Journal of Financial Econometrics 15, 247-285.

Chang, C.Y., Lai, J.Y., Chuang, I-Y., 2010. Futures hedging effectiveness under the segmentation of bear/bull energy markets. Energy Economics 32, 442-449.

Chen, F., Sutcliffe, C., 2012. Better cross hedges with composite hedging? Hedging equity portfolios using financial and commodity futures. European Journal of Finance 18, 575-595.

Cotter, J., Hanly, J. 2012. A utility based approach to energy hedging. Energy Economics 34, 817-827.

De Menezes, L. M., Bunn, D. W., Taylor, J. W., 2000. Review of guidelines for the use of combined forecasts. European Journal of Operational Research 120, 190-204.

Diebold. F.X, Pauly, P., 1987. Structural change and the combination of forecasts. Journal of Forecasting 6, 21-40.

Diebold, F. X., 1989. Forecast combination and encompassing: Reconciling two divergent literatures. International Journal of Forecasting 5, 589-592.

Ederington, L.H., 1979. The hedging performance of the new futures markets. Journal of Finance 34, 157-170.

Ederington, L.H., Salas, J.M., 2008. Minimum variance hedging when spot price changes are partially predictable. Journal of Banking and Finance 32, 654–663.

Engle, C., 1994. Can the Markov switching model forecast exchange rates? Journal of

International Economics 36, 151–165.

Engle, R.F., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business and Economic Statistics 20, 339-350.

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122-150.

Fama, E. F., French, K., 1987. Commodity futures prices: Some evidence on forecast power, premiums and the theory of storage. The Journal of Business 60, 55-73.

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56, 329–352.

Fuertes, A.M., Miffre, J., Fernandez-Perez, A., 2015. Commodity strategies based on momentum, term-structure and idiosyncratic volatility. Journal of Futures Markets 35, 274-297.

Gao, X., Nardari, F., 2018. Do commodities add economic value in asset allocation? New evidence from time-varying moments. Journal of Financial and Quantitative Analysis 53, 365-393.

Gagnon, L., Lypny, G., 1995. Hedging short-term interest rate risk under time-varying distributions. Journal of Futures Markets 15, 767-783.

Genre, V., Kenny, G., Meyler, A. Timmermann, A., 2013. Combining expert forecasts: Can anything beat the simple average? International Journal of Forecasting 29, 108–121.

Ghoddusi, H., Emamzadehfard, S., 2017. Optimal hedging in the US natural gas market: The effect of maturity and cointegration. Energy Economics 63, 92-105.

Greyserman, A., Jones, D.H., Strawderman, W.E., 2006. Portfolio selection using hierarchical Bayesian analysis and MCMC methods. Journal of Banking and Finance 30, 669-678.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and business cycle. Econometrica 57, 357-384.

Hanly, J., 2017. Managing energy price risk using futures contracts: A comparative analysis. Energy Journal 38, 93-112.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365-380.

Heather, P., 2012. Continental European gas hubs: are they fit for purpose? NG 63, Oxford Institute for Energy Studies: Oxford.

Johansen, S., 1988. Statistical analysis of cointegrating vectors. Journal of Economic Dynamics and Control 12, 231-254.

Kroner, K., Sultan, J., 1993. Time-varying distributions and dynamic hedging with foreign currency futures. Journal of Financial and Quantitative Analysis 28, 535–551.

Lamoureux, C.G., Lastrapes, W.D., 1990. Persistence in variance, structural change, and the GARCH model. Journal of Business and Economic Statistics 8, 225–234.

Lee, H.-T., Yoder, J., 2007. Optimal hedging with a regime-switching time-varying correlation GARCH model. Journal of Futures markets 27, 495-516.

Lien, D., 1996. The effect of the cointegration relationship on futures hedging: a note. Journal of Futures Markets 16, 773-780.

Lien, D., Tse, Y.K., 2002. Some recent developments in futures hedging. Journal of Economic Surveys 16, 357-396.

Ljung, M., Box, G., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297-303.

Locke, P., Venkatesh, P., 1997. Futures market transaction costs. Journal of Futures Markets 17, 229-245.

Martínez, B., Torró, H., 2015. European natural gas seasonal effects on futures hedging.

Energy Economics 50, 154-168.

Nomikos, N.K., Pouliasis, P.K., 2011. Forecasting petroleum market volatility: The role of regimes and market conditions. Energy Economics 33, 321-337.

Nomikos, N.K., Pouliasis, P.K., 2015. Petroleum term structure dynamics and the role of regimes. Journal of Futures Markets 35, 163-185.

Palm, F. C., Zellner, A., 1992. To combine or not to combine? Issues of combining forecasts. Journal of Forecasting 11, 687–701.

Patton, A.J., Sheppard, K., 2009. Optimal combinations of realised volatility estimators. International Journal of Forecasting 25, 218–238.

Pesaran, M.H., Timmermann, A., 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics 137, 134-161.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regressions. Biometrica 75, 335-346.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Pouliasis, P.K., Papapostolou, N.C., Kyriakou, I., Visvikis, I.D., 2018. Shipping equity risk behavior and portfolio management. Transportation Research Part A: Policy and Practice 116, 178-200.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance 54, 1647–1691.

Viswanath, M., 1993. Efficient use of information, convergence adjustments, and regression estimates of hedge ratios. Journal of Futures Markets 13, 43-53.

Table 1: Summary statistics for spot and futures prices

Panel A: Descriptive Statistics

	NG (U.S.)				NBP (U.K.)			
	Weekly		*Monthly*		*Weekly*		*Monthly*	
	Spot	*Futures*	*Spot*	*Futures*	*Spot*	*Futures*	*Spot*	*Futures*
Mean	0.0415	0.0401	0.1793	0.1453	0.1700	0.1606	0.5348	0.4741
Vol	9.1706	7.4788	16.337	13.324	7.5764	5.8784	15.358	14.131
Skew	0.7336*	0.2368*	0.4081*	-0.0377	0.9392*	1.3396*	0.1984	-0.1292
Kurt (Exc)	11.658*	2.0626*	3.3382*	0.5923	4.6701*	12.434*	1.5890*	1.4742*
$Q(6)$	36.180*	19.463*	14.779*	4.319	24.434*	72.430*	17.705*	31.522*
$Q^2(6)$	171.39*	115.81*	9.514	4.387	121.11*	59.763*	45.256*	9.076
PP (levels)	-2.6652	-2.5488	-2.6975	-2.6494	-2.5396	-2.4897	-2.4967	-2.4329
PP (returns)	-38.030*	-34.868*	-18.781*	-15.326*	-33.835*	-29.703*	-14.021*	-11.538*

Panel B: Cointegration Tests

	H_0:	λ_{max} test	λ_{trace} test	Normalized CV $(1\ \beta_1\ \beta_0)$	LR tests $H_0: \beta_1 = 1$	LR tests $H_0: \beta_1 = 1, \beta_0 = 0$
				Weekly variations		
NG (U.S.)	$r = 0$	204.49	211.82	(1 -0.994 0.001)	0.5466	7.912
	$r = 1$	7.3273	7.3273		[0.460]	[0.019]
NBP (U.K.)	$r = 0$	386.91	394.45	(1 -0.996 -0.008)	1.735	17.039
	r = 1	7.543	7.543		[0.188]	[0.000]
				Monthly variations		
NG (U.S.)	$r = 0$	89.477	96.718	(1 -1.002 0.023)	0.0421	13.787
	$r = 1$	7.241	7.241		[0.837]	[0.001]
NBP (U.K.)	$r = 0$	89.096	89.096	(1 -0.999 0.007)	0.0012	5.737
	$r = 1$	8.8651	8.8651		[0.972]	[0.057]

The table reports summary statistics, unit root and cointegration tests for the NG (U.S.) and NBP (U.K.) markets. Sample period is from December 30, 1998 to May 16, 2018; corresponding to 1,012 weekly observations and 234 monthly observations. Panel A reports the descriptive statistics for log-returns and unit root tests for both log-prices (levels) and log-price changes (returns). Mean and Vol are the mean return and standard deviation of the series. Skew and Kurt are the estimated centralized third and fourth moments of the data. $Q(6)$ and $Q^2(6)$ are Ljung-Box (1978) tests for 6th order autocorrelation in the level and squared series, respectively. PP is the Phillips and Perron (1988) unit root test which tests the null hypothesis that the variable is non-stationary, $I(1)$, against the alternative that the variable is stationary, $I(0)$.. An asterisk * denotes significance at 5% level. Panel B presents the results from cointegration tests. r represents the number of cointegrating vectors (CVs). Based on Johansen (1988), we use the λ_{max} *and* λ_{trace} statistics. λ_{max} tests the null hypothesis of r CVs against the alternative of $r+1$. The 5% critical values for H_0: r=0 and H_0: r=1 are 15.67 and 9.24, respectively. λ_{trace} tests the null hypothesis that there are at most r CVs against the alternative that the number of CVs is greater than r. The 5% critical values for H_0: r=0 and H_0: r=1 are 19.96 and 9.24, respectively. The LR statistic is a likelihood ratio test on the coefficients of the CV (β_2 β_1 β_0). This is $-T[\ln(1-\hat{\lambda}_1^*) - \ln(1-\hat{\lambda}_1)]$ where $\hat{\lambda}_1^*$ and $\hat{\lambda}_1$ denote the largest eigenvalues of the restricted and the unrestricted models, respectively. The statistic follows a χ^2 distribution with degrees of freedom equal to the total number of restrictions minus the number of the just identifying restrictions, which equals the number of restrictions placed on the CV. Exact significance levels are in square brackets [·].

Table 2: Dynamic hedging models

Panel A: NG (U.S.)

	Weekly variations					*Monthly variations*				
	μ	γ_0	γ_1	γ_2		μ	γ_0	γ_1	γ_2	
	w					m				
OLS w seas	0.0002	0.7821*	0.1087*	-0.0716		0.0006	1.0320*	0.1032	0.1439*	
	(0.002)	(0.032)	(0.043)	(0.045)		(0.005)	(0.042)	(0.059)	(0.058)	
OLS w basis	-0.0006	0.7909*	-2.1230*			-0.0022	0.9950*	1.7663*		
	(0.001)	(0.054)	(1.051)			(0.005)	(0.042)	(0.415)		
	μ_1	γ_1	σ_1	$\varphi_{0,1}$	$\varphi_{1,1}$	μ_2	γ_2	σ_2	$\varphi_{0,2}$	$\varphi_{1,2}$
					Weekly variations					
MRS	0.0027	0.6737*	0.1443*	0.1620*		-0.0003	0.8276*	0.0330*	0.0417*	
	(0.004)	(0.110)	(0.033)	(0.037)		(0.001)	(0.018)	(0.004)	(0.014)	
MRS w basis	0.0063	0.6763*	0.1433*	1.2532*	0.9680	-0.0013	0.8240*	0.0323*	3.3243*	34.754*
	(0.007)	(0.109)	(0.031)	(0.275)	(2.737)	(0.001)	(0.028)	(0.004)	(0.432)	(9.091)
					Monthly variations					
MRS	0.0065	1.2449*	0.1737*	0.3949*		-0.0007	0.9750*	0.0409*	0.0810	
	(0.010)	(0.200)	(0.045)	(0.066)		(0.001)	(0.024)	(0.006)	(0.044)	
MRS w basis	0.0096	1.2523*	0.1762*	-0.0510	-2.8941	-0.0017	0.9776*	0.0411*	2.6231*	17.698
	(0.015)	(0.143)	(0.043)	(0.411)	(4.108)	(0.002)	(0.021)	(0.005)	(0.578)	(10.20)
	ω_1	α_1	β_1	ω_2	α_2	β_2	α	β	$\bar{\rho}$	
					Weekly variations					
GARCH	0.0020*	0.1727*	0.6026*	0.0005*	0.1609*	0.7603*	0.1476*	0.6760*	0.6194*	
	(5.6E-04)	(0.071)	(0.082)	(1.3E-04)	(0.039)	(0.051)	(0.020)	(0.027)	(0.029)	
					Monthly variations					
GARCH	0.0001	0.0880*	0.9107*	0.0014	0.0333	0.8882*	0.1231*	0.8757*	0.8617*	
	(0.001)	(0.014)	(0.015)	(0.001)	(0.027)	(0.068)	(0.018)	(0.012)	(0.026)	

Panel A: NBP (U.K.)

	Weekly variations					*Monthly variations*				
	μ	γ_0	γ_1	γ_2		μ	γ_0	γ_1	γ_2	
	w					m				
OLS w seas	-0.0011	0.8654*	-0.1165*	-0.1717*		-0.0033	1.0027*	-0.0904	-0.1734*	
	(0.002)	(0.033)	(0.042)	(0.051)		(0.005)	(0.037)	(0.047)	(0.057)	
OLS w basis	-0.0001	0.8673*	-0.0675			0.0005	0.9598*	-0.0928		
	(0.002)	(0.031)	(0.242)			(0.005)	(0.036)	(0.397)		
	μ_1	γ_1	σ_1	$\varphi_{0,1}$	$\varphi_{1,1}$	μ_2	γ_2	σ_2	$\varphi_{0,2}$	$\varphi_{1,2}$
					Weekly variations					
MRS	0.0009	0.8159*	0.0838*	0.1173*		-0.0003	0.9724*	0.0267*	0.0698*	
	(0.002)	(0.069)	(0.006)	(0.031)		(0.001)	(0.040)	(0.002)	(0.017)	
MRS w basis	0.0033	0.8198*	0.0843*	1.8902*	5.3416	-0.0018	0.9665*	0.0271*	3.1609*	50.219*
	(0.003)	(0.089)	(0.008)	(0.307)	(5.074)	(0.001)	(0.043)	(0.003)	(0.526)	(17.51)
					Monthly variations					
MRS	-0.0001	0.9759*	0.0953*	0.0833		0.0018	0.8988*	0.0231*	0.1016*	
	(0.002)	(0.046)	(0.012)	(0.050)		(0.001)	(0.050)	(0.004)	(0.039)	
MRS w basis	-0.0003	0.9705*	0.0968*	2.2927	-22.35	0.0016	0.9248*	0.0240*	2.1221*	-29.52
	(0.006)	(0.061)	(0.021)	(1.246)	(16.13)	(0.002)	(0.189)	(0.011)	(0.377)	(33.72)
	ω_1	α_1	β_1	ω_2	α_2	β_2	α	β	$\bar{\rho}$	
					Weekly variations					
GARCH	0.0001*	0.1592*	0.8385*	0.0003*	0.0673*	0.8484*	0.1555*	0.8271*	0.6773*	
	(4.8E-05)	(0.035)	(0.029)	(1.0E-04)	(0.023)	(0.031)	(0.040)	(0.029)	(0.031)	
					Monthly variations					
GARCH	0.0004	0.2241*	0.7737*	0.0004	0.1777*	0.8166*	0.0966*	0.8815*	0.8809*	
	(3.1E-04)	(0.072)	(0.054)	(2.7E-04)	(0.064)	(0.051)	(0.016)	(0.025)	(0.030)	

The table presents estimation results of the dynamic hedge ratio models. Sample period is from December 30, 1998 to May 16, 2018; corresponding to 1,012 weekly and 234 monthly observations. Panel A reports the results for NG (U.S.) and Panel B for NBP (U.K.). Note that GARCH model parameters differ only in the correlation dynamics as a two-stage estimation procedure (Engle, 2002) is implemented; the first step involves the estimation of univariate models for conditional variances and, in the second step, we estimate the correlation dynamics. An asterisk * denotes significance at 5% level.

Table 3: In-sample hedging effectiveness

	Weekly variations				*Monthly variations*			
	Traditional		Predictable spot (w basis)		Traditional		Predictable spot (w basis)	
	HE_1	vs. GMV	HE_2	vs. GMV	HE_1	vs. GMV	HE_2	vs. GMV
Panel A: NG (U.S.)								
Individual models								
Naïve 1:1	34.551	6.591	47.048	4.425	74.043	2.492	78.613	0.331
OLS	38.377	2.765	50.408	1.066	74.254	2.281	78.675	0.269
OLS w seas	38.892	2.250	50.864	0.609	75.360	1.175	**78.913**	0.031
OLS w basis	**40.201**	0.942	50.856	0.617	**76.136**	0.400	78.745	0.199
VAR	38.270	2.873	50.389	1.084	74.170	2.365	78.641	0.304
VECM	38.222	2.920	50.387	1.086	74.016	2.519	78.626	0.318
MRS	38.830	2.313	**50.967**	0.506	75.895	0.640	78.681	0.264
MRS w basis	38.796	2.346	50.883	0.591	75.859	0.676	78.683	0.261
CC-GARCH	37.388	3.755	45.261	6.213	73.284	3.251	76.493	2.451
DC-GARCH	38.999	2.143	46.764	4.709	73.736	2.799	77.133	1.812
MRS-GARCH	39.142	2.000	48.603	2.287	75.139	1.396	78.259	0.685
Model combinations								
1/N	39.528	1.614	51.125	0.349	75.386	1.149	78.782	0.162
(GMV_{uncon})	41.150	-0.008	52.349	-0.876	77.427	-0.892	78.961	-0.017
GMV ($w \geq 0$)	41.142		51.473		76.535		78.944	
Panel B: NBP (U.K.)								
Individual models								
Naïve 1:1	44.640	2.883	50.060	2.669	75.210	0.751	77.450	1.471
OLS	45.873	1.650	50.679	2.050	75.481	0.481	77.606	1.316
OLS w seas	**47.212**	0.310	**51.666**	1.063	**75.918**	0.044	**78.885**	0.037
OLS w basis	45.933	1.589	50.743	1.986	75.493	0.468	77.611	1.311
VAR	45.411	2.112	50.204	2.525	73.924	2.037	76.127	2.795
VECM	41.989	5.533	47.942	4.787	72.411	3.550	74.955	3.967
MRS	46.622	0.901	51.250	1.479	75.549	0.412	77.689	1.233
MRS w basis	46.619	0.903	51.280	1.450	75.504	0.457	77.631	1.291
CC-GARCH	46.316	1.207	50.683	2.046	75.414	0.548	75.361	3.561
DC-GARCH	45.662	1.860	49.852	2.877	75.462	0.499	76.738	2.184
MRS-GARCH	47.835	-0.313	52.570	0.159	75.471	0.490	77.397	1.525
Model combinations								
1/N	46.994	0.529	51.631	1.098	75.616	0.346	77.720	1.202
(GMV_{uncon})	48.503	-0.980	52.939	-0.210	76.080	-0.119	79.184	-0.263
GMV ($w \geq 0$)	47.522		52.729		75.961		78.922	

The table shows the degree of in-sample hedging effectiveness (HE) achieved using different models in the NG (U.S.) (Panel A) and NBP (U.K.) natural gas markets for both, weekly and monthly, hedging horizons. The in-sample period is from December 30, 1998 to May 16, 2018; corresponding to 1,012 weekly observations and 234 monthly observations. The column(s) "Traditional" of refer to HE_1 (Eq. 3), where the variance of the unhedged position is calculated as $Var(\Delta S_t)$. The column(s) "Predictable spot (w basis)" refer to HE_2 (Eq.5), i.e., Ederington and Salas (ES, 2008) approach for calculating HE; the variance of the unhedged position is calculated as $Var(\Delta S_t - \lambda Basis_{t-1})$. Numbers in bold indicate the best individual model performance in terms of HE. We construct three alternative hedges, namely *1/N*, GMV and GMV ($w \geq 0$). *1/N* corresponds to the hedging portfolio after weighting equally the hedge ratios derived from the individual models. GMV assigns weights to the hedge ratios derived from the individual models under the objective to minimize overall portfolio variance. GMV ($w \geq 0$) is the GMV with the additional restriction that the weights assigned to the individual models cannot be negative. Columns "vs. GMV" denote the improvement in the HE when GMV ($w \geq 0$) model is used. For model estimation results, see Table 2. Note that, MRS-GARCH is not used in forecast combinations. All figures are in % terms.

Table 4: Out-of-sample hedging performance for weekly hedging horizon

	NG (U.S.)				*NBP (U.K.)*			
	Traditional		Predictable spot (w basis)		Traditional		Predictable spot (w basis)	
	HE_1	vs. max of *i* models	HE_2	vs. max of *i* models	HE_1	vs. max of *i* models	HE_2	vs. max of *i* models
Individual models								
Naïve 1:1	21.794		30.060		36.713		45.515	
OLS	22.373		30.962		39.258		46.492	
OLS w seas	20.591		29.113		**39.472**		**46.561**	
OLS w basis	20.903		30.552		39.182		46.443	
VAR	22.520		30.968		38.985		45.388	
VECM	22.540		30.969		37.970		42.467	
MRS	**22.602**		31.401		38.863		46.251	
MRS w basis	22.433		**31.779**		39.180		46.335	
CC-GARCH	21.075		30.629		34.391		42.268	
DC-GARCH	21.093		31.046		35.677		44.055	
MRS -GARCH	23.087		31.810		39.556		47.038	
Forecast combinations								
1/N	23.437[c]	0.836	31.371	-0.408	38.893	-0.580	45.864c	-0.697
$\xi = 0, \eta = 1$	23.587[c]	0.986	31.437	-0.342†	40.064[c,&]	0.592	47.097[c,&]	0.537
$\xi = 0, \eta = 2$	23.624[c,&]	1.022	31.512	-0.266†	40.066[c,&]	0.594	47.099[c,&]	0.538
$\xi = 0, \eta = 3$	23.711[c,&]	1.110	31.597	-0.182†	40.068[c,&]	0.596	47.100[c,&]	0.540
$\xi = 0, \eta = 6$	23.915[c,&]	1.313	31.937	0.159	40.075[c,&]	0.603	47.105[c,&]	0.544
$\xi = 0, \eta = 10$	23.957[c,&]	1.355	32.371[c]	0.593	40.084[c,&]	0.611	47.111[c,&]	0.550
$\xi = 0.5, \eta = 1$	23.604[c,&]	1.002	31.461	-0.318†	40.060[c,&]	0.588	47.091[c,&]	0.530
$\xi = 0.5, \eta = 2$	23.688[c,&]	1.087	31.567	-0.211†	40.062[c,&]	0.590	47.092[c,&]	0.532
$\xi = 0.5, \eta = 3$	23.801[c,&]	1.200	31.689	-0.089†	40.064[c,&]	0.592	47.093[c,&]	0.533
$\xi = 0.5, \eta = 6$	24.005[c,&]	1.404	32.132	0.353	40.070[c,&]	0.597	47.097[c,&]	0.536
$\xi = 0.5, \eta = 10$	23.878[c,&]	1.277	32.726[c]	0.948	40.077[c,&]	0.605	47.101[c,&]	0.541
$\xi = 1, \eta = 1$	23.598[c]	0.997	31.485	-0.294†	40.029[c,&]	0.557	47.101[c,&]	0.540
$\xi = 1, \eta = 2$	23.753[c,&]	1.151	31.643	-0.135†	40.030[c,&]	0.558	47.102[c,&]	0.541
$\xi = 1, \eta = 3$	23.886[c,&]	1.285	31.838	0.059	40.032[c,&]	0.560	47.102[c,&]	0.542
$\xi = 1, \eta = 6$	24.058[c,&]	1.457	32.360[c]	0.582	40.036[c,&]	0.564	47.105[c,&]	0.544
$\xi = 1, \eta = 10$	23.785[c,&]	1.184	33.002[c,&]	1.223	40.042[c,&]	0.570	47.108[c,&]	0.547
$\xi = 3, \eta = 1$	23.731[c,&]	1.130	31.599	-0.180†	39.895[c,&]	0.422	47.110[c,&]	0.550
$\xi = 3, \eta = 2$	23.993[c,&]	1.392	31.920	0.142	39.893[c,&]	0.421	47.110[c,&]	0.549
$\xi = 3, \eta = 3$	24.169[c,&]	1.568	32.233[c]	0.454	39.892[c,&]	0.420	47.109[c,&]	0.548
$\xi = 3, \eta = 6$	24.096[c,&]	1.495	33.077[c,&]	1.298	39.888[c,&]	0.415	47.107[c,&]	0.547
$\xi = 3, \eta = 10$	23.576[c,&]	0.974	33.540[c,&]	1.762	39.882[c]	0.410	47.105[c,&]	0.544
$\xi = 10, \eta = 1$	24.145[c,&]	1.543	32.160[c]	0.381	39.619	0.146	47.032[c,&]	0.471
$\xi = 10, \eta = 2$	24.570[b,&]	1.969	32.786[c,&]	1.008	39.616	0.144	47.025[c,&]	0.464
$\xi = 10, \eta = 3$	24.575[b,&]	1.973	33.243[c,&]	1.464	39.613	0.141	47.019[c,&]	0.458
$\xi = 10, \eta = 6$	23.951[c,&]	1.350	33.876[c,&]	2.097	39.603	0.131	47.009[c,&]	0.448
$\xi = 10, \eta = 10$	23.144[c]	0.543	33.517[c,&]	1.739	39.575	0.103	47.001[c,&]	0.440
$\xi = 20, \eta = 1$	24.490[b,&]	1.888	32.460[c]	0.682	39.592	0.120	47.029[c,&]	0.468
$\xi = 20, \eta = 2$	24.845[b,&]	2.243	33.213[c,&]	1.435	39.583	0.111	47.020[c,&]	0.460
$\xi = 20, \eta = 3$	24.688[b,&]	2.087	33.847[c,&]	2.068	39.576	0.103	47.014[c,&]	0.454
$\xi = 20, \eta = 6$	23.925[c,&]	1.323	33.953[c,&]	2.174	39.561	0.089	47.002[c,&]	0.441
$\xi = 20, \eta = 10$	23.138[c]	0.536	33.468[c,&]	1.690	39.525	0.052	46.982[c,&]	0.421
Best FC (ξ, η)	25.493[a,&]	2.892	34.733[b,&]	2.954	40.166[b,&]	0.694	47.170[b,&]	0.609
	(250,2)		(250,4)		(0,50)		(0,50)	

The table shows the of out-sample hedging effectiveness (HE) achieved using different models in the NG (U.S.) (Panel A) and NBP (U.K.) natural gas markets for weekly hedges. The out-sample period is from May 29 2013 to May 16, 2018; corresponding to 260 weekly observations. To forecast the hedge ratios we use rolling windows of equal length to estimate the parameters of the hedging models starting with data covering the period from December 1998 to May 2013. ξ and η are the discount factor and tuning forecast combination parameters; see Eq. (8) and (9). "Best FC" (Forecast Combination) reports the best hedging outcome (in terms of HE), when perturbing forecast combination parameters with, $\xi \in [0,250]$ and $\eta \in [0,50]$ (see also Figure 3). Numbers in bold indicate the best individual model performance. a, b and c, in the square brackets [·], correspond to significance levels of 1%, 5% and 10%, respectively, indicating whether the combination approach is significantly better than the individual models' performance, using Hansen's (2005) reality check and 5,000 simulations of the Politis and Romano (1994) stationary bootstrap method. Note that a, b and c represent the average *p*-value of the test (pair-wise comparisons of combination forecasts vs. i^{th} model). & denotes whether the combination significantly outperforms at least 7 out of the 10 individual models. Column "*vs. max of i models*" computes the difference in the HE between a combination procedure and the best performer of the individual models. In the case of negative differential, † denotes whether the combination ranks second best. For comparison, we also report the hedging performance of the MRS-GARCH, but this model is not included in any of the forecast combination strategies.

Table 5: Out-of-sample hedging performance for monthly hedging horizon

	NG (U.S.)				*NBP (U.K.)*			
	Traditional		Predictable spot (w basis)		Traditional		Predictable spot (w basis)	
	HE_1	vs. max of *i* models	HE_2	vs. max of *i* models	HE_1	vs. max of *i* models	HE_2	vs. max of *i* models
Individual models								
Naïve 1:1	78.419		**84.876**		77.993		78.193	
OLS	80.044		83.556		78.447		78.606	
OLS w seas	77.766		82.371		77.604		77.722	
OLS w basis	68.990		83.416		**78.503**		**78.674**	
VAR	79.182		84.232		77.890		75.719	
VECM	77.974		84.334		75.492		73.571	
MRS	78.979		84.033		78.119		78.269	
MRS w basis	78.804		84.156		78.040		77.980	
CC-GARCH	**81.917**		82.799		75.581		75.788	
DC-GARCH	81.360		80.992		75.191		75.380	
MRS -GARCH	83.230		83.000		78.193		78.615	
Forecast combinations								
1/N	82.134[b,&]	0.217	84.116	-0.760	77.718	-0.785	77.687	-0.986
ξ = 0, η = 1	81.997[b,&]	0.080	85.803[b,&]	0.927	78.597[c,&]	0.093	78.701[c,&]	0.027
ξ = 0, η = 2	81.853[b,&]	-0.064†	85.802[b,&]	0.926	78.594[c,&]	0.091	78.699[c,&]	0.025
ξ = 0, η = 3	81.702[c,&]	-0.215†	85.800[b,&]	0.925	78.592[c,&]	0.088	78.697[c,&]	0.023
ξ = 0, η = 6	81.197[&]	-0.720	85.797[b,&]	0.921	78.584[c,&]	0.080	78.691[c,&]	0.017
ξ = 0, η = 10	80.383[&]	-1.534	85.792[b,&]	0.916	78.572[c,&]	0.069	78.681[c,&]	0.007
ξ = 0.5, η = 1	82.043[b,&]	0.126	85.581[b,&]	0.705	78.597[c,&]	0.093	78.699[c,&]	0.025
ξ = 0.5, η = 2	81.945[b,&]	0.028	85.582[b,&]	0.706	78.594[c,&]	0.091	78.698[c,&]	0.024
ξ = 0.5, η = 3	81.839[b,&]	-0.078†	85.582[b,&]	0.707	78.591[c,&]	0.088	78.696[c,&]	0.022
ξ = 0.5, η = 6	81.467[&]	-0.450†	85.585[b,&]	0.709	78.583[c,&]	0.080	78.692[c,&]	0.018
ξ = 0.5, η = 10	80.826[&]	-1.091	85.589[b,&]	0.713	78.572[c,&]	0.069	78.685[c,&]	0.011
ξ = 1, η = 1	82.952[b,&]	1.035	85.642[b,&]	0.766	78.595[c,&]	0.092	78.630[&]	-0.044†
ξ = 1, η = 2	82.870[b,&]	0.953	85.642[b,&]	0.766	78.593[c,&]	0.090	78.630[&]	-0.044†
ξ = 1, η = 3	82.781[b,&]	0.864	85.643[b,&]	0.767	78.590[c,&]	0.087	78.629[&]	-0.045†
ξ = 1, η = 6	82.467[b,&]	0.551	85.646[b,&]	0.770	78.582[c,&]	0.079	78.628[&]	-0.046†
ξ = 1, η = 10	81.930[c,&]	0.013	85.651[b,&]	0.775	78.572[c,&]	0.069	78.625[&]	-0.049†
ξ = 3, η = 1	83.353[b,&]	1.436	85.537[b,&]	0.662	78.514[c]	0.011	78.679[&]	0.005
ξ = 3, η = 2	83.485[b,&]	1.568	85.523[b,&]	0.647	78.514[c]	0.011	78.679[&]	0.005
ξ = 3, η = 3	83.596[b,&]	1.679	85.508[b,&]	0.632	78.514[c]	0.011	78.679[&]	0.005
ξ = 3, η = 6	83.795[b,&]	1.878	85.461[b,&]	0.585	78.513[c]	0.010	78.680[&]	0.006
ξ = 3, η = 10	83.863[b,&]	1.946	85.404[b,&]	0.528	78.512[c]	0.008	78.680[&]	0.006
ξ = 10, η = 1	84.408[b,&]	2.491	85.495[b,&]	0.619	78.365	-0.138	78.493	-0.181
ξ = 10, η = 2	84.473[b,&]	2.556	85.420[b,&]	0.545	78.359	-0.144	78.488	-0.185
ξ = 10, η = 3	84.320[b,&]	2.403	85.326[b,&]	0.451	78.354	-0.149	78.484	-0.190
ξ = 10, η = 6	84.195[b,&]	2.278	84.968[c,&]	0.092	78.337	-0.166	78.471	-0.203
ξ = 10, η = 10	82.409[b,&]	0.492	84.631[&]	-0.245†	78.315	-0.188	78.454	-0.220
ξ = 20, η = 1	84.736[b,&]	2.819	85.436[b,&]	0.560	78.324	-0.179	78.315	-0.359
ξ = 20, η = 2	84.405[b,&]	2.488	85.251[b,&]	0.375	78.309	-0.194	78.298	-0.375
ξ = 20, η = 3	84.372[b,&]	2.455	85.009[c,&]	0.133	78.293	-0.210	78.282	-0.392
ξ = 20, η = 6	82.308[c,&]	0.391	84.434	-0.442†	78.244	-0.259	78.231	-0.443
ξ = 20, η = 10	80.497	-1.419	84.256	-0.620	78.172	-0.331	78.198	-0.476
Best FC (ξ,η):	85.199[a,&]	3.282	85.803[b,&]	0.927	78.599[c,&]	0.096	78.702[c,&]	0.028
	(120,0.5)		(0,0.5)		(0.75,0.5)		(0,0.5)	

The table shows the degree of out-sample hedging effectiveness (HE) achieved using different models in the NG (U.S.) (Panel A) and NBP (U.K.) natural gas markets for monthly hedges. The out-sample period is from June 15, 2013 to May 15, 2018; corresponding to 60 monthly observations. See also notes in Table 4.

Table 6: Out-of-sample hedge ratio stability and performance fees

	Weekly variations				*Monthly variations*			
	NG (U.S.)		*NBP (U.K.)*		*NG (U.S.)*		*NBP (U.K.)*	
	Trad.	Pred. spot (w basis)	Trad.	Pred. spot (w basis)	Trad.	Pred. spot (w basis)	Trad.	Pred. spot (w basis)
Panel A: Hedged portfolio turnover								
OLS	0.14	0.12	0.08	0.09	0.32	0.26	0.19	0.30
OLS w seas	1.30	1.08	1.83	1.55	1.25	0.74	0.96	0.99
OLS w basis	7.85	4.04	0.24	0.21	12.26	1.74	0.27	0.38
VAR	0.13	0.11	0.08	0.09	0.27	0.22	0.20	0.21
VECM	0.11	0.11	0.07	0.05	0.22	0.22	0.15	0.18
MRS	1.40	1.31	1.15	2.28	0.93	0.28	1.31	1.33
MRS w basis	1.69	1.92	1.36	1.27	2.16	0.67	3.74	2.49
CC-GARCH	9.57	7.14	7.36	8.06	21.61	8.43	8.05	8.05
DC-GARCH	9.83	8.35	11.61	11.67	21.76	9.91	8.83	8.83
1/N	5.15	4.45	0.62	0.59	6.11	1.10	1.75	1.79
Best FC	5.07	3.40	2.08	1.65	12.38	2.74	1.54	3.29
FC_{avg}	3.26	1.68	1.29	1.02	5.71	1.32	1.41	1.57
Panel B: FC increase in HE relative to increase in Turnover								
Naïve 1:1	0.7 [0.7]	1.4 [1.4]	1.7 [2.5]	1.0 [1.5]	0.6 [0.8]	0.3 [0.4]	0.4 [0.4]	0.2 [0.2]
OLS	0.6 [0.5]	1.2 [1.0]	0.5 [0.5]	0.4 [0.6]	0.4 [0.5]	0.9 [1.8]	0.1 [+0.0]	+0. [-0.]
OLS w seas	1.3 [1.7]	2.4 [5.6]	2.7 [-0.7]	6.4 [-1.0]	0.7 [1.1]	1.7 [5.3]	1.7 [2.0]	0.4 [1.5]
OLS w basis	-1.6 [-0.7]	-6.5 [-0.8]	0.5 [0.7]	0.5 [0.8]	137 [-2.1]	2.4 [-4.7]	0.1 [-0.]	+0 [-0.1]
VAR	0.6 [0.5]	1.2 [1.0]	0.6 [0.7]	1.1 [1.8]	0.5 [0.7]	0.6 [1.1]	0.5 [0.5]	1.0 [2.1]
VECM	0.6 [0.4]	1.2 [1.0]	1.1 [1.6]	3.0 [4.8]	0.6 [0.9]	0.6 [1.0]	2.2 [2.4]	1.7 [3.6]
MRS	0.8 [0.7]	1.6 [2.8]	1.4 [7.4]	-1.5 [-0.7]	0.5 [0.8]	0.7 [1.3]	2.1 [3.7]	0.2 [1.3]
MRS w basis	0.9 [1.0]	2.0 [-2.7]	1.4 [-9.7]	2.2 [-0.8]	0.6 [1.11]	0.8 [2.0]	-0.3 [-0.2]	0.9 [-0.6]
CC-GARCH	-1.0 [-0.5]	-1.1 [-0.3]	-1.1 [-0.9]	-0.8 [-0.7]	-0.4 [-0.1]	-0.5 [-0.4]	-0.5 [-0.4]	-0.6 [-0.4]
DC-GARCH	-0.9 [-0.3]	-0.7 [-0.2]	-0.5 [-0.4]	-0.3 [-0.3]	-0.4 [-0.1]	-0.7 [-0.5]	-0.5 [-0.4]	-0.6 [-0.4]
1/N	-25 [-0.3]	-3.2 [-0.4]	0.9 [1.5]	1.2 [2.8]	0.5 [-1.5]	1.0 [5.9]	-4.2 [-2.2]	0.7 [-3.9]
Panel C: FC Performance fees								
Naïve 1:1	**5135**	**5914**	**1414**	668	**3728**	**591**	298	252
	[3144]	**[3429]**	[1260]	[593]	**[2888]**	**[453]**	[242]	[67]
OLS	**4570**	**5108**	351	253	**3158**	**1242**	91	61
	[2449]	**[2326]**	[177]	[175]	**[2132]**	**[1169]**	[26]	[-155]
OLS w seas	**6653**	**6898**	722	572	**3896**	**1677**	**424**	**414**
	[4938]	**[4690]**	[556]	[497]	**[3096]**	**[1629]**	**[372]**	[255]
OLS w basis	**6360**	**5558**	358	260	**6293**	**1265**	70	32
	[4598]	**[2952]**	[184]	[183]	**[5820]**	**[1193]**	[4]	[-189]
VAR	**4364**	**5093**	468	724	**3480**	**935**	**354**	**1150**
	[2191]	**[2304]**	[297]	[649]	**[2570]**	**[837]**	**[301]**	**[1041]**
VECM	**4323**	**5093**	904	**1858**	**3877**	**885**	**1186**	**1725**
	[2140]	**[2305]**	[741]	**[1789]**	**[3074]**	**[783]**	**[1151]**	**[1636]**
MRS	**4250**	**4589**	450	279	**3547**	**1013**	**342**	**322**
	[2048]	**[1559]**	[278]	[202]	**[2657]**	**[923]**	[287]	[150]
MRS w basis	**4446**	**4117**	358	296	**3615**	**959**	**318**	**439**
	[2294]	[802]	[183]	[219]	**[2744]**	**[863]**	[263]	[282]
CC-GARCH	**6593**	**5460**	**2784**	**1932**	**2159**	**1560**	**1165**	**1134**
	[4869]	**[2818]**	**[2648]**	**[1864]**	**[443]**	**[1507]**	**[1129]**	**[1025]**
DC-GARCH	**6579**	**5422**	**2301**	**1800**	**2413**	**2068**	**1267**	**1243**
	[4852]	**[2766]**	**[2160]**	**[1731]**	**[917]**	**[2037]**	**[1233]**	**[1138]**
1/N	**3246**	**4690**	654	594	**2250**	**966**	**426**	**483**
	[732]	**[1714]**	[486]	[519]	**[634]**	**[872]**	**[375]**	**[331]**
Average	**4710**	**4828**	897	770	**3201**	**1097**	**495**	**605**
	[3114]	**[2515]**	[815]	[766]	**[2452]**	**[1115]**	**[489]**	**[507]**

Panel A of this table reports the mean out-sample portfolio turnover in % terms. "FC_{avg}" corresponds to the average figure when perturbing forecast combination parameters with, ξ, η as in Tables 4 and 5. Panel B, shows the ratio of the incremental HE of HE_{BestFC} relative to increase in portfolio turnover; numbers [.] perform the same comparison but with HE_{FCavg}. Note that $HE_{\text{BestFC}} > HE_{\text{FCavg}} > HE_{model\ i}$, therefore, *incremental HE* is always a positive number. Panel C displays the FC performance fees in annualised basis points; numbers [.] perform the same comparison but with FC_{avg}. Results are reported for NG U.S. and NBP U.K. (Panel B) natural gas markets, weekly and monthly hedging horizons, using the traditional HE measure, HE_1, and the Ederington and Salas (2008) (HE_2) approach (see also notes in Tables 4 and 5).

Table 7: Out-of-sample hedging performance under market segmentation

	Naïve	OLS	OLS seas	OLS basis	VAR	VECM	MRS	MRS basis	CC – GARCH	DC – GARCH	1/N	$FC_{\min HE}$	FC_{maxHE}	FC_{avgHE}
Panel A: NG (U.S.)														
								Backwardation						
HE_1(w)	16.14	14.91	13.42	**17.08**	15.21	15.27	15.17	15.18	13.46	13.60	16.44[&]	14.80	17.11[c,&]†	16.46
HE_2(w)	24.50	23.05	21.34	24.11	23.14	23.15	23.75	24.03	22.56	**25.15**	23.94	24.05	26.97[b,&]†	25.31[c,&]
HE_1(m)	86.26	87.60	85.39	87.56	86.73	85.57	86.73	86.78	89.11	**89.20**	88.05[&]	87.65	89.37[&]†	88.18[&]
HE_2(m)	**89.00**	87.77	85.76	88.15	88.51	88.63	88.32	88.44	87.26	88.91	88.49	89.90[c,&]	89.98[c,&]†	89.95[c,&]
								Contango						
HE_1(w)	53.01	56.24	54.45	41.84	56.47	56.47	57.11	56.39	57.37	**57.57**	57.05	57.16	60.53[b,&]†	58.57[c,&]
HE_2(w)	71.18	73.29	72.28	70.04	73.48	73.49	74.15	**74.19**	72.72	69.58	73.60	73.47	74.21[&]†	73.71
HE_1(m)	75.48	**77.06**	75.73	50.93	76.47	75.50	76.14	75.69	73.26	69.71	80.95[b,&]	75.78[c]	85.55[b,&]†	81.20[c,&]
HE_2(m)	87.27	86.34	**87.29**	84.96	86.62	86.64	86.67	86.70	85.18	71.95	86.14	81.36	87.37[b,&]†	85.91[&]
								Winter						
HE_1(w)	17.00	17.70	15.54	16.34	17.76	17.76	**17.86**	17.73	16.14	16.13	18.74[c,&]	18.44[c,&]	20.33[b,&]†	19.31[&]
HE_2(w)	24.42	25.35	23.23	24.91	25.32	25.32	25.77	**26.23**	25.06	25.57	25.73	25.81	28.68[b,&]†	26.99[c]
HE_1(m)	77.77	79.87	77.10	65.79	78.74	77.24	78.45	78.03	**83.92**	83.33	82.16[c,&]	80.01[&]	85.44[b,&]†	82.95[c,&]
HE_2(m)	**85.93**	84.06	82.88	83.90	84.97	85.12	84.71	84.82	83.06	82.01	84.89	85.92[c,&]	87.70[a,&]†	87.20b[&]
								Summer						
HE_1(w)	63.74	63.42	64.90	61.07	64.30	64.46	64.21	63.69	65.02	**65.25**	64.72	64.66	65.30[&]†	64.84
HE_2(w)	69.88	70.58	70.67	70.37	70.88	70.89	**71.16**	70.96	69.95	69.78	71.20[b,&]	71.08[&]	71.24[b,&]†	71.16[c,&]
HE_1(m)	80.16	79.38	79.60	**81.57**	79.88	80.16	80.12	81.23	75.55	75.13	80.76[&]	80.60[&]	81.94[&]†	81.14[&]
HE_2(m)	81.20	81.99	81.17	**82.22**†	81.79	81.74	81.83	82.06	82.19	78.54	81.63	79.67	80.19	79.91
Panel B: NPB (U.K.)														
								Backwardation						
HE_1(w)	26.63	33.73	**35.13**	33.62	32.86	29.90	32.89	33.36	25.00	23.52	32.10	34.09[&]	36.18[c,&]†	35.54[&]
HE_2(w)	39.52	42.97	**43.92**	42.86	39.28	32.11	42.32	42.33	31.62	36.09	40.33	44.69[&]	46.46[c,&]†	46.00[&]
HE_1(m)	83.60	84.84	84.55	84.85	83.37	78.44	**85.19**	84.72	79.98	80.35	83.60	85.15[&]	85.74[b,&]†	85.61[c,&]
HE_2(m)	83.91	85.10	84.80	85.13	78.79	74.88	**85.41**	84.95	80.28	80.59	83.33	83.57	86.45[b,&]†	85.87[&]
								Contango						
HE_1(w)	21.03	26.34	**26.70**	26.35	25.73	23.51	24.31	24.34	23.14	22.59	25.54	27.14[&]	27.89[b,&]†	27.59[c,&]
HE_2(w)	32.52	35.33	**35.50**	35.36	32.48	26.44	34.00	33.86	26.11	33.02	33.35	37.52[c,&]	38.12[b,&]†	37.91b&
HE_1(m)	72.35	73.46	70.48	73.50	72.10	67.36	**74.01**	73.99	73.16	72.31	72.78	73.85	74.13†	73.94
HE_2(m)	72.51	73.48	70.44	73.54	67.52	64.06	74.10	**74.28**	73.35	72.52	72.40	73.44	74.35[&]†	74.19
								Winter						
HE_1(w)	36.45	40.06	**41.65**	39.87	39.63	38.16	39.94	40.67	34.79	35.18	39.84	40.69[&]	41.79[c,&]†	41.36[&]
HE_2(w)	46.34	47.79	**49.14**	47.67	46.12	42.30	47.65	48.25	42.03	44.67	47.09	48.61[&]	49.15[b,&]†	49.00[c,&]
HE_1(m)	79.50	80.35	80.07	**80.50**	79.46	76.03	80.00	80.00	79.32	78.59	79.75	80.83[c,&]	80.96[b,&]†	80.89[b,&]
HE_2(m)	79.52	80.37	80.03	**80.57**	76.11	73.21	80.01	79.71	79.32	78.57	79.41	79.64	81.00[b,&]†	80.79[c,&]
								Summer						
HE_1(w)	36.74	38.05	35.71	**38.10**	37.97	37.48	37.30	37.13	34.09	36.22	37.56	37.85	38.27[b,&]†	37.97
HE_2(w)	44.24	44.68	43.09	**44.70**	44.24	42.37	44.27	43.72	42.26	43.31	44.13	44.61	44.81†	44.66
HE_1(m)	**75.71**	75.59	73.91	75.50	75.51	74.62	75.29	75.09	70.07	70.23	74.66	74.12	75.73†	75.00
HE_2(m)	**75.98**†	75.94	74.26	75.82	75.08	74.04	75.64	75.37	70.58	70.74	75.08	73.82	75.94	75.20

The table shows the degree of out-sample hedging effectiveness (HE) in the NG U.S. (Panel A) and NBP U.K. (Panel B) natural gas markets for weekly (w) and monthly (m) hedging horizons, using the traditional HE measure, HE_1, and the Ederington and Salas (2008) (HE_2) approach (see also notes in Tables 4 and 5). Results of HE are provided under the market segmentation to backwardation/contango, winter/summer periods. Backwardation (contango) relates to periods when the spot price is above (below) the futures price. Winter (summer) periods are from October to March (April to September). a, b and c, correspond to significance levels of 1%, 5% and 10%, respectively. Columns $FC_{\min HE}$ and $FC_{\max HE}$, respectively represent the worst and best forecast combination for $\xi = \{0{,}0.5{,}1{,}3{,}10{,}20\}$ and $\eta = \{1{,}2{,}3{,}6{,}10\}$. a, b and c indicate whether the combination approach is significantly better than the individual model performance using bootstrap simulations (see notes in Table 3 for more details). Columns denoted as FC_{avgHE} correspond to the average performance of all considered combination methods when varying ξ and η. In this case a, b and c are derived as the average *p*-value of the test (average of all 300 pair-wise comparisons of combination forecasts vs. i^{th} model; 30 combinations – see Tables 4 and 5 – tested against 10 models). & denotes whether the FC significantly outperforms at least 7 out of the 10 individual models, on average. Numbers in bold denote the best performing individual model; † denotes the best model overall.

Table 8: Out-of-sample hedging performance of short/long positions

	Naïve	OLS	OLS seas	OLS basis	VAR	VECM	MRS	MRS basis	CC – GARCH	DC – GARCH	1/N	$FC_{min\,HE}$	FC_{maxHE}	FC_{avgHE}
Panel A: NG (U.S.)														
							Lower partial moment of order 2: LPM_2							
Short hedgers														
HE^*_1(w)	23.77	26.19	25.13	24.32	26.18	26.17	**26.48**	26.23	22.66	23.62	27.21[c,&]	26.05[c,&]	29.95[b,&]†	28.11[c,&]
HE^*_2(w)	48.40	51.81	50.64	50.48	51.84	51.83	52.23	**52.78**	51.59	48.97	52.16	52.30	56.12[b,&]†	53.54[c]
HE^*_1(m)	75.67	77.31	76.42	57.20	76.38	75.38	76.13	75.65	**87.10**	87.02	81.78[&]	78.67[&]	87.43[b,&]†	84.47[&]
HE^*_2(m)	**84.99**	83.79	84.08	83.34	84.26	84.33	84.22	84.30	82.77	83.59	84.23	85.55[b,&]	85.80[b,&]†	85.64[b,&]
Long hedgers														
HE^*_1(w)	**20.24**	19.33	16.96	18.17	19.60	19.64	19.51	19.40	19.73	19.00	20.42[c]	20.48	20.78[b,&]†	20.65[c]
HE^*_2(w)	21.28	21.05	18.88	21.08	21.04	21.04	21.47	21.75	20.66	**22.61**	21.48	21.56	23.51[b,&]†	22.49[c,&]
HE^*_1(m)	81.16	**82.46**	79.13	80.42	81.97	80.56	81.82	81.94	76.44	75.37	82.52	73.37	82.49†	81.63
HE^*_2(m)	**84.83**	83.39	80.79	83.58	84.26	84.40	83.92	84.09	82.88	78.58	84.08	83.01	86.07[c,&]†	85.32[&]
							Lower partial moment of order 3: LPM_3							
Short hedgers														
HE^*_1(w)	19.10	18.84	18.04	22.23	19.09	19.11	19.31	19.18	30.14	**32.68**	24.19	28.37	34.99[c,&]†	31.41[c]
HE^*_2(w)	54.35	57.42	56.60	58.16	57.43	57.44	57.58	57.85	57.22	**63.17**	59.32	60.96	70.03[c,&]†	64.49[c]
HE^*_1(m)	84.05	86.26	85.06	53.09	85.23	83.87	84.82	84.11	**94.76**	94.70	91.30[&]	87.99	94.86[c,&]†	93.05[&]
HE^*_2(m)	**92.98**	92.00	92.46	91.56	92.39	92.44	92.34	92.38	91.13	92.60	92.38	93.48[b,&]	93.64[b,&]†	93.57[b,&]
Long hedgers														
HE^*_1(w)	**18.79**†	15.87	12.30	15.93	16.25	16.33	16.15	16.38	16.04	15.31	16.57	16.59	16.84	16.65
HE^*_2(w)	**18.44**†	15.89	12.97	16.54	15.91	15.92	16.87	17.34	15.22	14.91	16.23	16.25	16.93	16.55
HE^*_1(m)	90.31	**91.47**	89.03	89.33	91.06	89.74	90.92	90.59	88.67	87.55	91.58[&]	85.31	91.55[&]†	90.96[&]
HE^*_2(m)	**90.30**	88.87	85.93	89.25	89.75	89.90	89.46	89.65	88.50	87.82	89.74	90.72[c]	92.31[b,&]†	91.79[c]
Panel B: NPB (U.K.)														
							Lower partial moment of order 2: LPM_2							
Short hedgers														
HE^*_1(w)	33.19	35.01	31.16	**35.24**†	34.89	34.25	35.00	34.73	23.47	27.55	33.68	33.94	34.80	34.46
HE^*_2(w)	41.51	42.04	39.11	**42.17**†	41.39	38.83	42.15	41.89	38.61	35.91	41.18	41.55	41.95	41.83
HE^*_1(m)	75.77	**76.00**	75.53	75.99	75.63	73.49	75.02	75.34	72.62	71.87	75.12	75.30	76.17[c]†	75.96
HE^*_2(m)	75.81	**75.91**	75.46	75.91	73.51	71.55	74.97	74.81	72.74	72.00	74.94	73.70	75.93[c]†	75.55
Long hedgers														
HE^*_1(w)	39.91	43.11	**44.88**	42.75	42.69	41.34	42.36	43.21	44.09	42.88	43.59[&]	44.55[&]	44.93[c,&]†	44.80[&]
HE^*_2(w)	49.14	50.50	**52.14**	50.30	49.00	45.77	49.97	50.35	45.59	51.16	50.08	52.09[&]	52.35[c,&]†	52.16[c,&]
HE^*_1(m)	79.60	80.22	79.10	**80.33**	79.53	76.94	80.32	80.00	77.72	77.58	79.60	80.23	80.35[c]†	80.31
HE^*_2(m)	79.91	80.51	79.34	80.62	77.31	75.03	**80.64**	80.26	77.97	77.80	79.66	79.73	80.73[c]†	80.53
							Lower partial moment of order 3: LPM_3							
Short hedgers														
HE^*_1(w)	45.45	47.37	43.90	47.59	**47.61**	46.84	47.51	47.31	31.89	38.14	46.57	46.85	47.63†	47.32
HE^*_2(w)	56.04	56.47	53.91	**56.59**	55.90	52.29	56.43	56.40	51.95	49.61	55.91	56.32	56.63†	56.54
HE^*_1(m)	81.86	**82.19**	81.97	82.19	81.74	79.53	81.62	81.58	78.59	77.79	81.34	81.83	82.49[b,&]†	82.31[c,&]
HE^*_2(m)	81.99	82.23	81.96	**82.25**	79.67	77.69	81.66	81.50	78.83	78.05	81.22	80.44	82.41[c,&]†	82.08
Long hedgers														
HE^*_1(w)	58.13	61.86	**63.82**	61.36	61.39	59.83	61.46	62.55	62.94	61.49	62.34[&]	63.37[&]	63.91[c,&]†	63.75[c,&]
HE^*_2(w)	68.65	69.93	**71.40**	69.67	68.45	65.00	69.77	69.96	64.79	70.40	69.55	71.33[&]	71.53[c,&]†	71.39[c,&]
HE^*_1(m)	89.87	90.47	90.03	**90.57**	89.79	87.32	90.53	90.31	88.85	89.08	89.97	90.72[c]	90.87[c,&]†	90.77[c,&]
HE^*_2(m)	90.24	90.86	90.42	**90.96**	87.78	85.48	90.90	90.67	89.21	89.43	90.12	90.90	91.28[c,&]†	91.22[c]

The table shows the degree of out-sample hedging effectiveness (HE^*) achieved in the NG U.S. (Panel A) and NBP U.K. (Panel B) natural gas markets for weekly (w) and monthly (m) hedging horizons, using the traditional measure (HE_1^*) and the Ederington and Salas (2008) (HE_2^*) approach (see notes in Tables 4, 5 and 6). Results are provided under the segmentation to short/long hedgers using HE^* measure as the reduction in the relevant lower partial moment (see Eq. 15), rather than variance reduction; These are calculated as $HE_1^* = 1 - LPM_k(0;\Delta S_t)/LPM_k(0;\Delta S_t - \gamma\Delta F_t)$ and $HE_2^* = 1 - LPM_k(0;\Delta S_t - \lambda Y_{t-1})/LPM_k(0;\Delta S_t - \gamma\Delta F_t - \lambda Y_{t-1})$, for $k = \{2,3\}$ and Y_t the futures-spot spread (basis).

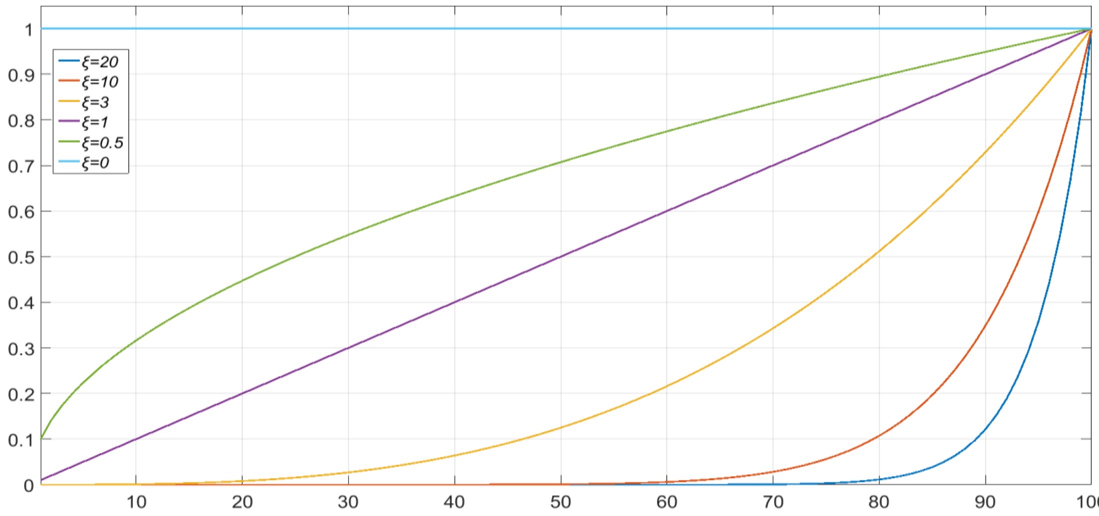

Figure 1: Discount factor and weighting scheme. The figure illustrates the effect of changing ξ on the importance given to most recent vs. distant observations. This is measured by the discount factor $(t/T)^{\xi}$ for a sample size of T = 100 observations. The discount factor is plotted for $\xi = \{0,0.5,1,3,10,20\}$; as ξ increases, the weight assigned to distant observations decreases.

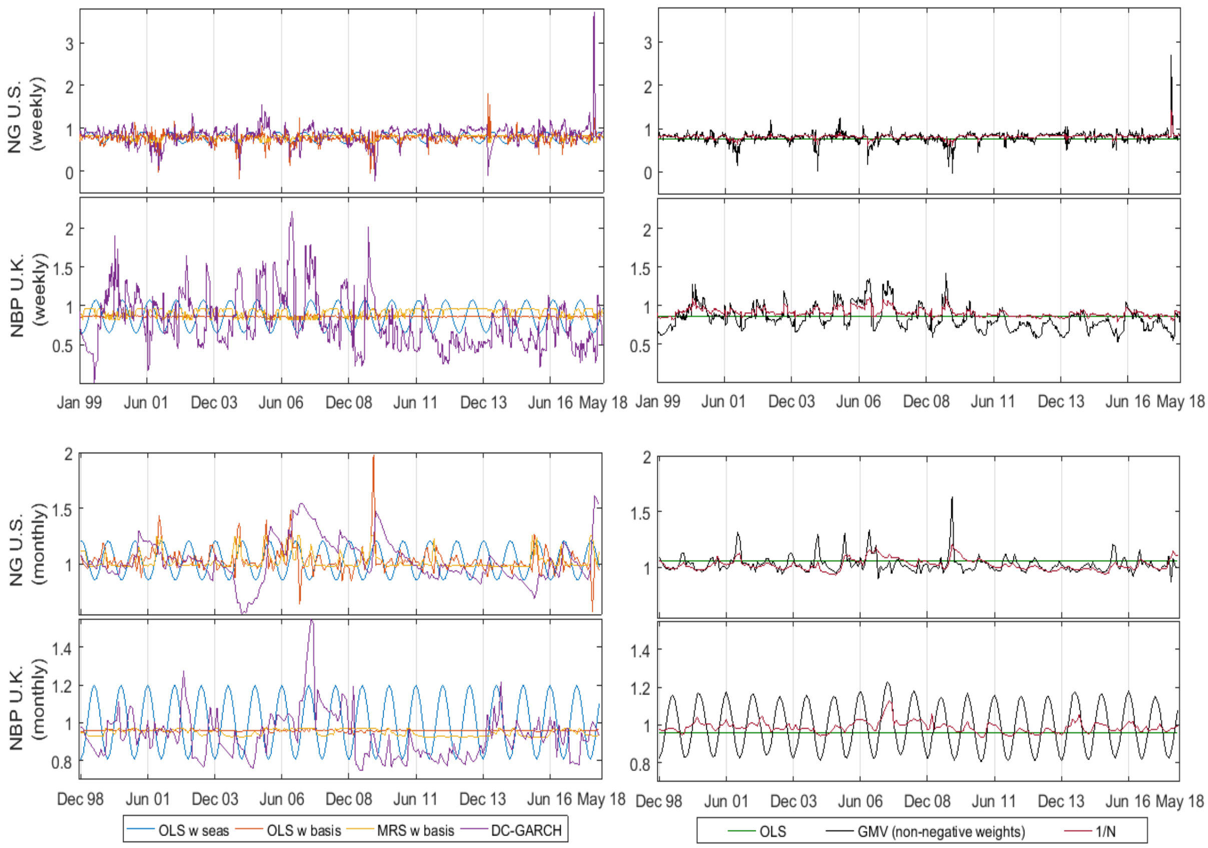

Figure 2: Dynamic hedge ratios. The subplots to the left plot the regression-based, MRS (with basis) and GARCH (with dynamic correlation) in-sample (December 1998 to May 2018) hedge ratios for the NG (U.S.) and NBP (U.K.) natural gas prices using weekly and monthly hedging horizons. The subplots to the right display two combinatory (GMV and *1/N*) hedge ratios vs. the constant OLS hedge ratio. The first (GMV) is considering a hedge ratio combination by weighting all ten individual hedge ratios based on the performance of each hedging portfolio (minimum variance portfolio with $w \geq 0$). The second (1/N) is an equally

weighted hedging portfolio based on the ten considered cases (Naïve, OLS, OLS w seas, OLS w basis, VAR, VECM, MRS, MRS w basis, CC-GARCH and DC-GARCH).

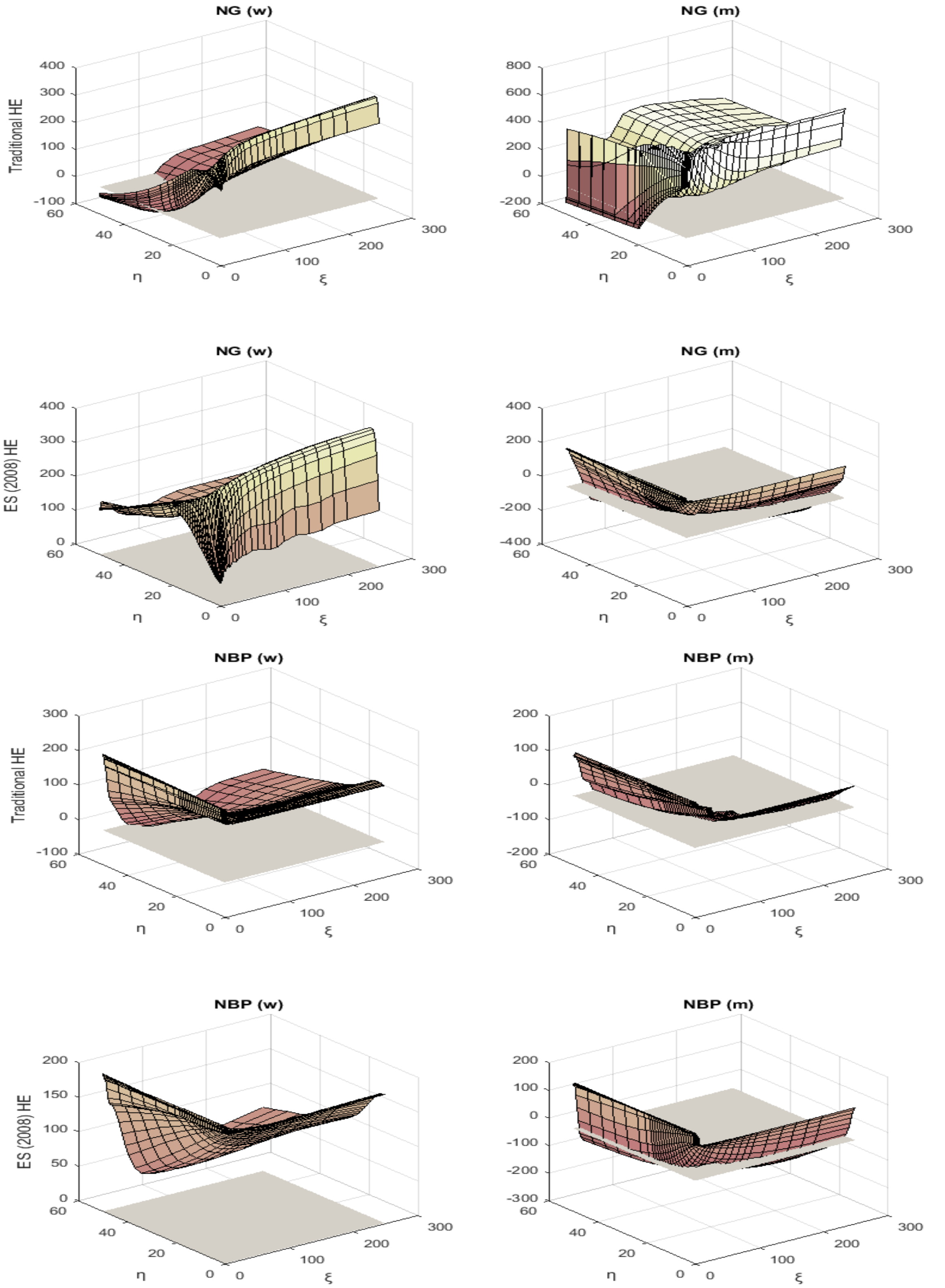

Figure 3: Effects of forecast combination parameters ξ (discount factor parameter; see Eq. 8) and η (tuning parameter; see Eq. 9) on relative hedging performance. The figure illustrates the performance of forecast combination approaches in the NG (U.S.) and NPB (U.K.) markets, in excess of the average individual model HE (in terms of basis points). Results are shown under the traditional (HE_1; Eq. 3) and Ederington and Salas (2008) (HE_2; Eq. 5) HE measures for weekly (w) and monthly (m) hedge horizons. The zero plane (gray surface) is also plotted for reference.

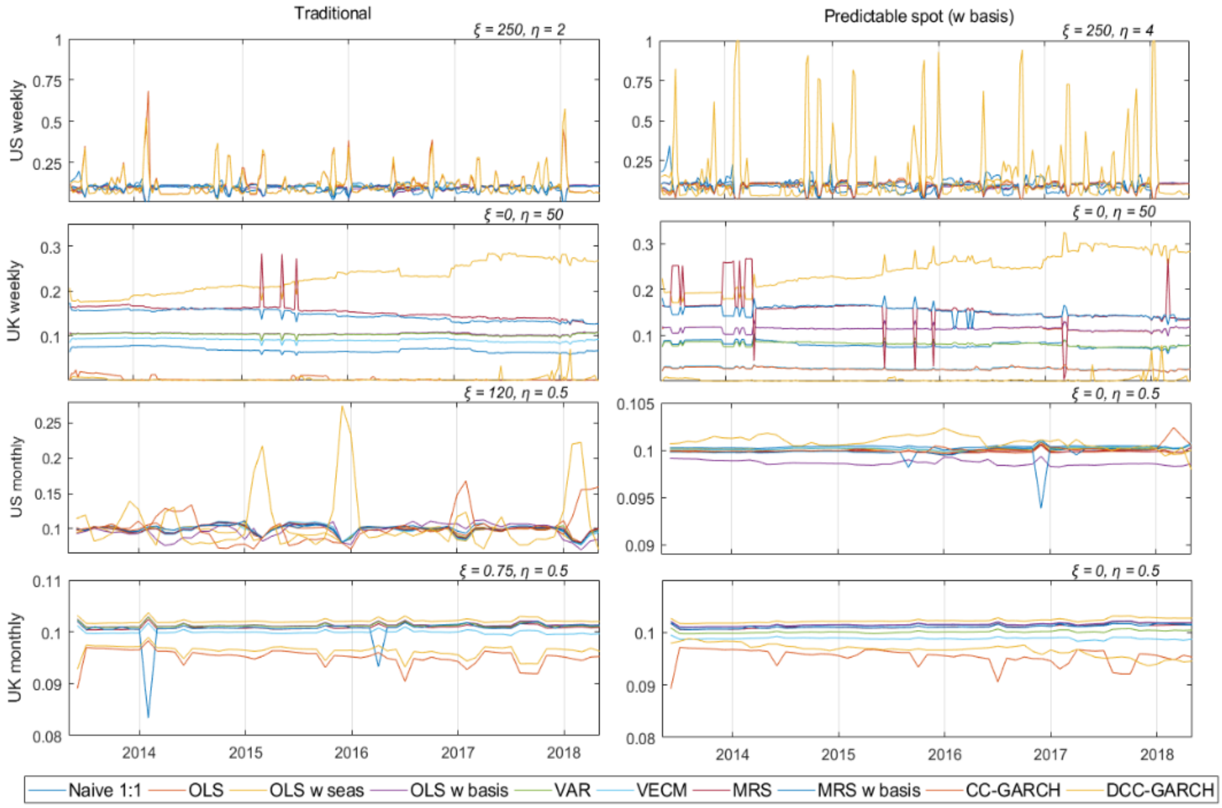

Figure 4: Forecast combination weight structure. The subplots to the left (right) plot the weights assigned to each model (Best FC) throughout the out-of-sample period (May 2013 to May 2018) for the NG (U.S.) and NBP (U.K.) natural gas prices for weekly and monthly hedging horizons using the traditional HE measure (the Ederington and Salas, 2008, HE_2 approach; see also notes in Tables 4 and 5).

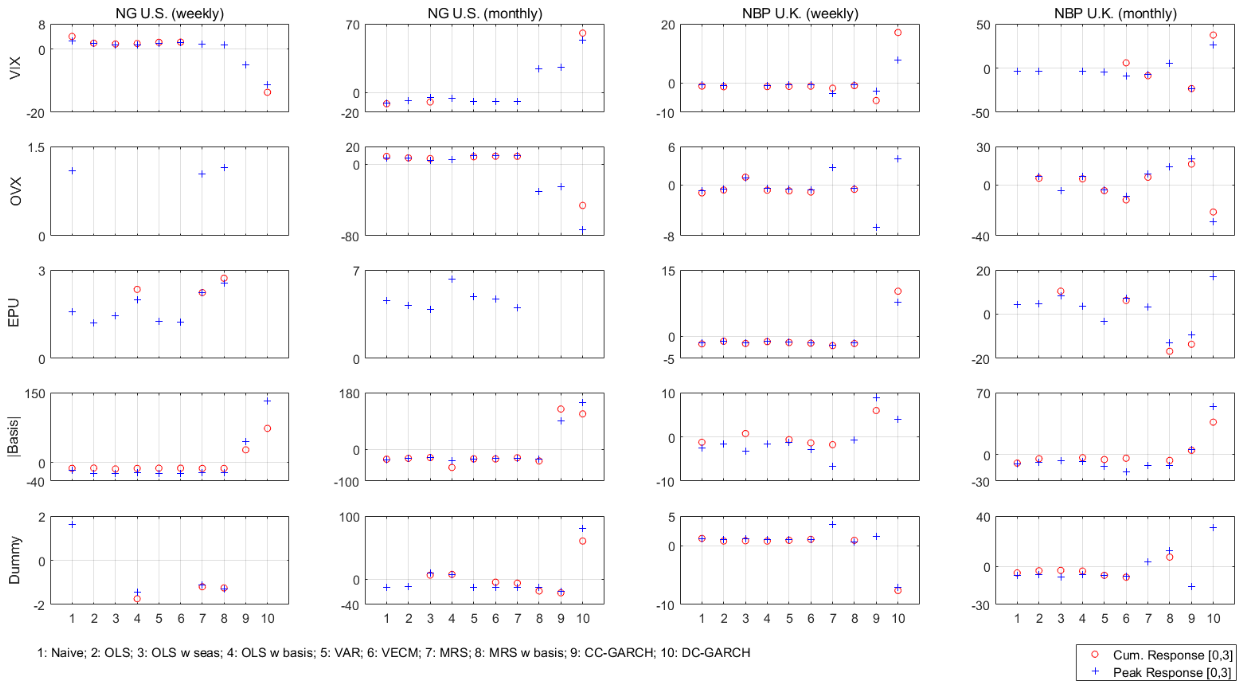

Figure 5: Impulse responses of combination weight changes. Estimates of the impulse responses obtained from a panel ARX regression of the changes in weights (in basis points) based on the least-squares dummy variable (*LSDV*) estimator. S&P500 implied volatility index (*VIX*), crude oil implied volatility index (*OVX*), Economic Policy Uncertainty (*EPU*), absolute spread between spot and futures $|Basis|$ and a *Dummy* which takes the value of one if the all above variables are simultaneously at the upper tertile of their empirical distributions are assumed to be exogenous. All variables are in changes apart from the $|Basis|$. Weekly (monthly) data comprise panels of 62 cross sections x 260 (60) obs, where 62 = 31 comb x 2 HE method, i.e., Traditional (HE_1) and Predictable spot w basis (HE_2). Numbers reported in the Figure are all significant at 5% level.