Asymmetry and Interdependence when Evaluating U.S. Energy Information Agency Forecasts

Garratt, Anthony and Petrella, Ivan and Zhang, Yunyi

25 August 2022

Asymmetry and Interdependence when Evaluating U.S. Energy Information Agency Forecasts*

Anthony Garratt†
University of Warwick

Ivan Petrella‡
University of Warwick

Yunyi Zhang§
Xiamen University

August 25, 2022

Abstract

We evaluate US Energy Information Agencies (EIA) forecasts of the world petroleum market, emphasising the importance of taking a multivariate perspective, considering asymmetric loss and allowing for time-variation. Forecasts for total demand, total supply, total stock withdrawals and the oil prices are biased, with biases that change over time and differ across variables. A loss function that takes into account asymmetry and interdependence can rationalise these biases. The implied asymmetric loss gives less weight to under-prediction of both demand and supply, while for oil prices, we document significant regime changes in the implied loss due to asymmetry. The EIA forecasts dominate a simple random walk benchmark when evaluated using symmetric and independent loss in the form of MSE statistical criteria. Yet, when allowing for asymmetry and interdependence that rationalize the EIA forecasts, the performance of the EIA forecasts worsens and is comparable to the random walk benchmark.

JEL codes: C32; C53; E37; Q47.

Keywords: EIA forecasts, oil market, forecast rationality, non-separable loss, asymmetric loss.

*We thank Maximilian Auffhammer, Boqiang Lin and participants at the ICCEF2022 conference for useful comments and suggestions. The authors gratefully acknowledge funding from the British Academy Small Research Grants scheme (Award number: SRG2111516).

†Warwick Business School, University of Warwick. Email: Anthony.Garratt@wbs.ac.uk

‡Warwick Business School, University of Warwick and CEPR. Email: Ivan.Petrella@wbs.ac.uk

§School of Management, China Institute for Studies in Energy Policy, Xiamen University. Email: yunyi_zhang@xmu.edu.cn

1 Introduction

The use and influence of the many types of energy forecasts produced by the U.S. Department of Energy's Energy Information Agency (EIA) is widespread. Numerous public, private sector, government organisations and analysts use EIA energy forecasts as important inputs into their environmental, energy, monetary, fiscal and investment policy decisions. Therefore developing a greater understanding and assessing the accuracy of EIA forecasts, over long sample periods and overtime, is an important and useful exercise to undertake. The forecasts performance determines their usefulness as inputs to the many decisions they are currently used for. Existing evaluations of EIA forecasts, such as those undertaken by the EIA of their annual reference case projections forecasts (EIA, 2020), and those undertaken in the academic literature have their limitations. In particular, very little emphasis is placed on how forecast performance varies over time, and it is often the case that only short samples are considered. The range of variables and forecast horizons analysed is limited, and, most importantly, evaluations are often based on the assumption that the forecasts are unrelated to one another. In fact, there are good reasons to suppose that forecasts are formulated jointly – not least because the demand and supply forecasts are conditional on a baseline scenario for future oil prices – suggesting alternative approaches to forecast evaluation allowing for joint determination and interdependence is an important aspect to be considered.

In this paper, we seek to further our understanding of the EIA's ability to produce accurate forecasts of the world petroleum market. To this end, we examine the EIA forecasts for total demand, supply, stock withdrawals in the world petroleum market and Refiner Acquisition Cost (RAC) oil prices, for the period 1983Q1-2019Q4. We consider a wide range of horizons, from a one quarter backcast, the current period or nowcast, through to forecasts for the next six quarters. The value of constructing and analysing a forecast data set with a long-time span and broad set of energy variables is twofold. First, the longer time span enables a thorough and rigorous examination of time variation. The energy market has undergone several substantial changes, in particular over the last 10/20 years, and knowing whether the predictions of the EIA are currently as reliable as they were documented to be in the past is important. As we aim to identify changing and evolving patterns/trends in the consumption of oil, the EIA forecasts play an important role potentially acting as a guide on, for example, how fast the economy is moving towards a "green economy". Second, by examining more variables, particularly the interaction between demand/supply/inventories and how these relate to prices, allows us to understand whether the overall balance between demand and supply is consistent, even when the exact forecast is imprecise.

The forecasts produced by the EIA are widely used by policymakers and the energy industry as an exogenous inputs to the decision making process. Often users assume that forecast producers use symmetric loss functions, which may not be the case. This misperception would result in a suboptimal outcome for the forecast user. A distinguishing feature of our analysis is the emphasis we place on, and the evidence we provide for, the importance of taking into account directional asymmetries and the likely correlations and interdependence of the forecast variables when evaluating forecasts (see Komunjer and Owyang, 2020). Forecasts from a directionally asymmetric loss functions will be biased, and we establish which variables the EIA systematically under-predicts or over-predicts. Separability implies that the costs of forecast errors for one variable do not depend on the forecast errors of others. Not accounting for this potential interdependence is clearly undesirable as the interdependence of the variables we examine would naturally arise in a classical theory of storage (Working, 1949; Pindyck, 1980). Uncertainty in future demand and supply fundamentals induces storage, which in turn, plays a stabilising role in consumption, production and prices. If demand and supply of crude oil are inelastic in the short run, any deviations from equilibrium storage levels could have an immediate impact on futures prices and volatility. Therefore, it makes sense to consider the inter- relationships when evaluating the forecasts.

We find evidence of bias in EIA forecasts. When time variation in the forecasting performance is not explicitly considered, there is a statistically significant tendency to under-predict total demand and supply, but forecasts for stock withdrawals and the RAC oil price are unbiased. In fact, we reject the rationality of the forecasts of total demand, supply and stock withdrawals. Forecast error distributions are characterised by outliers and fat tails and there is also a suggestion of multimodality for demand and supply. The latter may potentially reflect time-variation in the properties of the forecast errors. Indeed, we document significant degrees of time variability in the bias and in the distributions of the forecast errors. The early part of our sample (until the late 90s), as well as the last part in our sample (post-2010), are characterised by a broad under-prediction of both demand and supply of oil, and over-prediction in price forecasts. Whereas the decade running up to the Great Recession is distinguished by under-prediction in the price forecasts, a somewhat over-prediction in the forecast of demand and production and positive bias in the stock of inventories.

Motivated by this first look at the forecast error characteristics we move away from assuming symmetric quadratic loss and undertaking individual variable forecast evaluation, in favour of joint evaluation under non-separable asymmetric loss (along the lines of Komunjer and Owyang, 2020). Adopting a different loss function alters the perspective on bias. We find strong evidence in the full-sample that asymmetries are important and we can no longer reject the null of rationality, for all of our forecasts. The size of the asymmetries in the EIA implicit loss function, as with the bias, varies overtime, moreover there is strong evidence of non-separability between the forecasts when specifying the loss function throughout the sample being analysed. The implied asymmetric loss gives less weight to under-prediction of both demand and supply, while for oil prices, we document significant regime changes in the implied loss due to asymmetry. Taken together those features of the asymmetric loss allow us to rationalize the time varying biases in the EIA forecasts.

Last, we conduct a relative forecast evaluation exercise, comparing the EIA forecasts to random walk-based benchmarks. For the full-sample period, we find mean square error ratios that indicate good comparative performance of the EIA forecasts, particularly at shorter time horizons, but where the RAC oil price performs noticeably worse. Moreover, these ratios show a large degree of time variation. However, if we adopt an asymmetric and nonseparable loss function, which is consistent with rational forecasts for the EIA, we observe a worsening of the EIA forecast performance, where the loss is no longer significantly different to the random walk benchmark, with especially poor performance at short horizons.

Previous related studies that examine the accuracy of EIA forecasts are varied in terms of objective function, horizon, specific variable or variables examined and evaluation methods. For example, and among others, studies which focus on a symmetric point forecast evaluation include Winebrake and Sakva (2006), Sanders et al. (2008), Sanders et al. (2009), and Lady (2010). Overall, they suggest EIA forecasts have good performance, with relatively small mean squared prediction errors, and largely rational forecasts, particularly at short horizons, less so for long horizons. The accuracy of EIA projections has also been tested by the EIA themselves. In the Annual Energy Outlook (AEO) Dec 2020 EIA evaluation document, when comparing AEO reference annual case projections with realised outcomes from 1994 to 2019, of the 25 variables examined greater than 50 percent over estimate the actual.[1] In addition to the more conventional energy EIA based forecast evaluation, is a closely related macro-based forecasting literature, whose focus is on forecasting the oil price. Baumeister et al. (2014) examine oil price forecasts, making a direct comparison with the forecasts produced by the EIA. Baumeister and Kilian (2015) compare the forecasting performance of six econometric models for the real oil price, individually and in combination always assuming a symmetric loss function.[2] Evaluation methods assuming asymmetric loss functions are used to evaluate the rationality of EIA forecasts in for example Auffhammer (2007) and Mamatzakis and Koutsomanoli-Filippaki (2014). However, prior to this paper, methods which account for interdependence between forecasts, applied by Caunedo et al. (2012) to evaluate forecasts produced by the Federal Reserve and by Bora et al. (2021) to evaluate the forecasts of U.S. Agriculture Department, have yet to be used on EIA forecasts. Our emphasis is on modelling interdependence, as well as asymmetries, between the EIA forecasts of the key variables of world petroleum market. Our results highlight that the EIA forecasts of demand, supply, inventory and prices, together, reflect a joint view of the developments in the world petroleum market from the EIA, who when producing their forecasts do not always weight equally costs associated with over- and under-predictions of the variables of interest. Therefore,

[1] More recently, Kaach et al. (2017) have developed and evaluated EIA forecast prediction intervals (densities), based around empirical density forecasting methods (using observed point forecast errors and assuming normality).

[2] For a review of this literature see Alquist et al. (2013).

users should be careful when taking each separate forecast in isolation of the others, or as a reflection of the EIA unconditional view of the variables of interest, when using those as inputs in their decision processes.

The remainder of this paper is organised as follows. In Section 2 we define the EIA forecast variables we analyse. Section 3 takes a first look at the EIA forecasts, adopting absolute forecast evaluation approach analsying unconditional bias and testing for rationality, for both the full-sample and over time. In Section 4 we adopt alternative loss functions allowing for non-separability and asymmetry, where we test for joint rationality. Section 5 conducts a relative forecast evaluation exercise using ratio which use MSE and alternative loss function estimated in Section 4. Section 6 concludes.

2 Data

We examine quarterly EIA world petroleum market forecasts from 1983Q1 to 2019Q4 for four key variables: total demand, total supply, total stock withdrawals (inventory) and the Refiners' Acquisition Cost (RAC) oil price. We use the natural logarithm of total demand and supply, measured in millions of barrels per day. Total stock withdrawals is also defined in millions of barrels per day. The RAC oil price is defined as dollars per barrel and in the empirical analysis we use the natural logarithm of this series. See Appendix A for additional details of the variable definitions and descriptive statistics of the data.

As the EIA began forecasting these variables in 1983, the sample period represents the longest history available to examine time variation in forecast performance, which is longer than typical previous comparable research in the literature. For example, this represents ten additional years of observations compared to recent EIA forecast evaluations (AEO Dec 2020) and is longer than the sample periods from academic studies cited in the introduction. To construct the long sample of quarterly data we spliced forecasts over different reporting frequencies. Up to 1997Q1, the EIA produced quarterly reports documenting observations and forecasts at a quarterly frequency. Specifically, in each quarter, the EIA produced a backcast for the last quarter, a nowcast of the current quarter and forecasts of one through to six quarters ahead, i.e. for horizons $h = -1, 0, 1, 2, ..., 6$. In March of 1997 the EIA switched to producing monthly reports, which continued to document quarterly observations and forecasts, but for each of the start, mid and end months in any quarter. From 1997Q2 through to 2007Q3, we choose to use the quarterly forecasts from the first or start month reported in the quarter.[3] Since October 2007 only monthly observations and forecasts have been reported by the EIA for variables on the international balance sheet.

[3]Using the first or start month forecast, as opposed to mid or end month, has no significant effect on the results. As, in the appendices, we document some basic features of the forecasts and forecast errors using forecasts reported at the start, mid and end of each month and find they have very similar properties. Note that for the RAC oil price the switch to monthly reporting occurred later in July 2004.

Hence from 2007Q4 through to 2019Q4 we use the average of the monthly forecasts. We take the vintage of data available in June 2020 as the target for the forecast evaluation throughout the paper.[4]

While we evaluate the EIA forecasts as if they reflect the EIA unconditional expectations of future developments in the oil market, it is worth highlighting that they are in fact projections of what may happen given the assumptions in the underlying National Energy Modeling System of the EIA. Those assumptions include projections of oil prices and gross domestic product. Moreover those projections assume current laws and regulations, and include current views of economic and demographic trends and technology improvements (see, e.g., EIA, 2020). Nonetheless, the forecasts produced by the EIA are widely used by policymakers, industry and modelers often under the assumption of a symmetric loss function. Therefore, our analysis will allow forecast users to gain a better understanding of the implicit loss function of the EIA, where this reflects the complex process of producing projections of the oil market.

As part of our emphasis on time-variation when conducting our forecast evaluation we identify three sub-samples in the oil market, reflecting different regimes. The three periods are: 1983Q4-1999Q4, 2000Q1-2009Q4 and 2010Q1-2019Q4. The first pre 2000's period, is one which is dominated by advanced economies being the drivers of demand and normal supply, and where volatility is relatively low. The mid-2000 period is characterized by excess demand dominated by emerging markets (China) and supply constraints, where we observe high oil price growth and a large price fall in 2008 (Hamilton, 2009). Finally, the post 2010 period stands out as a time of distinct change, with episodes of supply disruption (shale revolution) and demand changes (climate related concerns to lower demand) and more generally high volatility.

3 A first look at the EIA Forecasts

In this section we document the basic univariate properties, unconditional bias and conditional bias or rationality of the EIA forecasts. We introduce the importance of taking into account how the forecast errors vary overtime highlighting their potential inter-dependencies across variables. Here, in contrast to the subsequent analysis in Section 4, we evaluate the forecast errors of our four EIA variables, assuming they are independent of each other, using a symmetric quadratic loss function, the implication of which is that a basic requirement for a rational forecast is it be unconditionally unbiased.[5] We then estimate (for the full-sample and allowing for time-varying parameters) a re-parameterised Mincer-Zarnowitz regression, allowing us to test unconditional

[4]Results are robust to using the first vintage of data available for each quarter as an alternative target.

[5]Evidence of bias in the forecasts are of obvious interest to policymakers and energy planners, who often use those forecasts either as benchmarks or as inputs to other forecasts. In addition, whilst we assess each of the forecast independently, we are conscious that those forecasts are jointly developed by the EIA using different modelling assumption as discussed in Section 2.

and conditional bias or rationality, over the full-sample and how it changes over time.

We define forecast errors, for total demand, supply and the oil price as:

$$e_{t+h|t} = 100 \times (y_{t+h} - f_{t+h|t}), \tag{1}$$

where $y_{t+h} = log(Y_{t+h})$, $f_{t+h|t} = log(F_{t+h|t})$ and Y_{t+h} and $F_{t+h|t}$ are the observed outcomes of our variables at time $t+h$ and their forecasts, for period $t+h$, made at time t, respectively. We consider forecast horizons $h = -1, 0, 1, \ldots, 6$, where $h = -1$ denotes backcasts and $h = 0$ nowcasts. The forecast errors for total stock withdraws are defined as: $Y_{t+h} - F_{t+h|t}$. A *positive* value of a forecast error implies that a forecast *under-predicts* the actual observed outcome, whereas a *negative* forecast error is associated with a forecast that *over-predicts*.

3.1 Are the EIA Forecasts Biased?

In Table 1 we report results analysing unconditional bias measured by the mean, but also document the standard deviation, mean squared error (MSE), mean absolute error (MAE) and skew. The violin plots in Figure 1 present the distributions of the forecast errors, for each variable, at each forecast horizon.[6] For total demand and supply, at all forecast horizons (with the exception of $h = 6$), we observe statistically significant positive (unconditional) bias i.e. forecasts under-predict. The size of the under predictions in total demand and supply, whilst statistically significant, are relatively small, ranging from around 0.6% to 0.8% for demand and from 0.7% to 1.0% for supply. In contrast, forecasts for stock withdrawals and the RAC oil price do not exhibit any statistically significant bias.

The violin plots in Figure 1 visualise characteristics or features of the forecast error distributions not immediately apparent from the Table 1.[7] For example, the violin plots for RAC oil price forecast errors show the largest variability across all forecast horizons (note the scaling of the axis). The narrowest range is exhibited in the nowcast ($h = 0$) distributions, increasing considerably at longer forecast horizons, with clear outlier observations associated with large oil price changes. The orders of magnitude are large with standard deviations ranging from around 8% through

[6]Using forecast errors constructed using forecasts documented at the start month. In the appendix we report the same violin plots comparing start, mid and end month quarter forecasts. These indicate that the forecast error distributions (and the bias results) are very similar for the three different forecast timings. Moreover, this observations applies to all four variables across all forecast horizons.

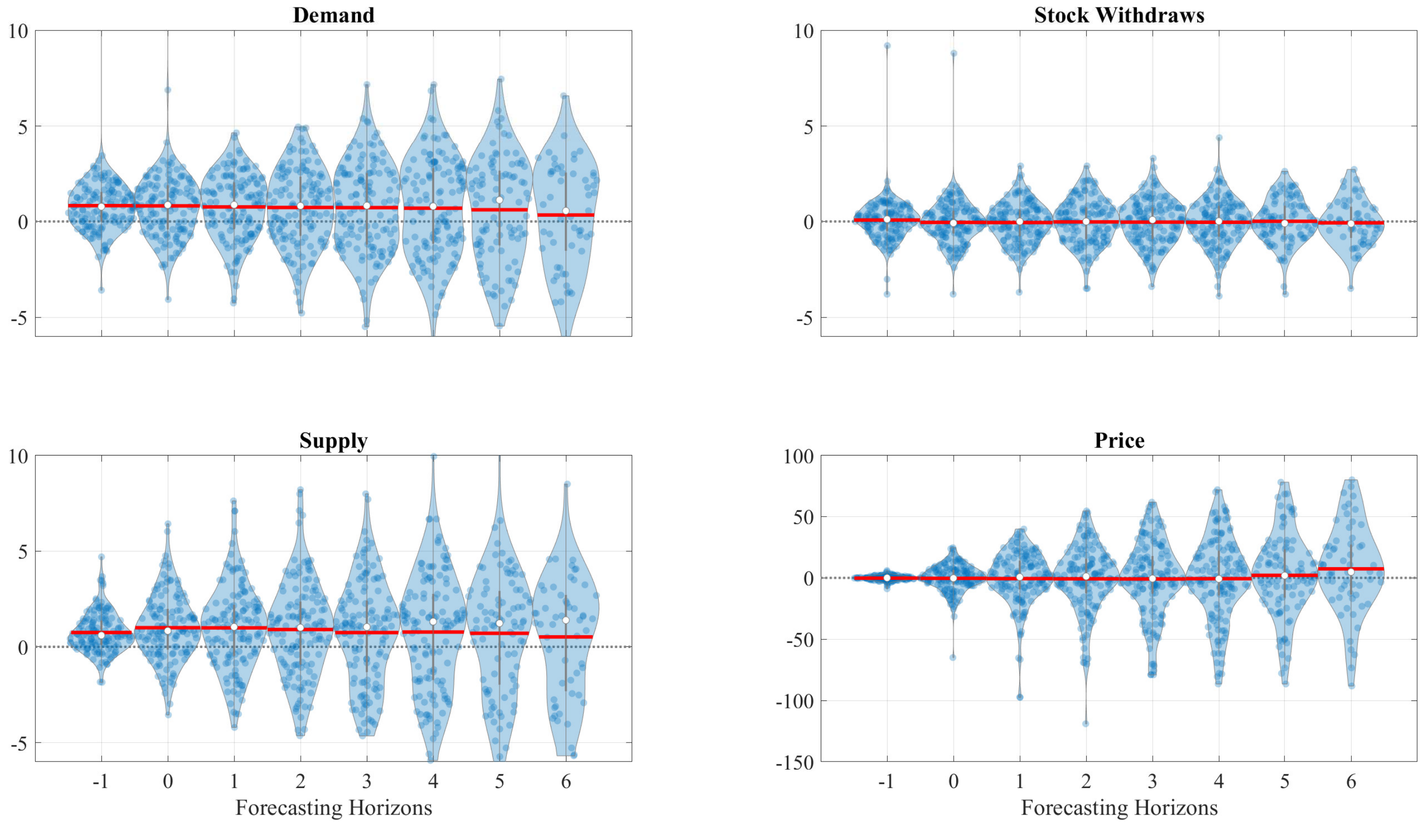
[7]The means and standard deviations, numbers relatively easily noted in the Tables, can arguably be discerned still more easily when plotted visually. We observe forecast error means (the red line) above zero and centred at zero (for all forecast horizons), indicating positive bias for demand and supply and no bias for total withdrawals and the RAC oil price. For the demand and supply forecast error distributions, we observe at the longer forecast horizons a more stretched shape, indicating increasing standard deviations (from around 1.8% for short-run forecasts, increasing to 3.0% for longer forecast horizons). This contrasts with violin plots for stock withdrawal forecast errors, which show an (approximate) similar range over the forecast horizons, indicating a constant standard deviation (of around 1.1 to 1.2 billion barrels a day).

Table 1: FORECAST ERRORS DESCRIPTIVE STATISTICS

	1-Quarter Backcasts					Nowcasts					1-Quarter Forecasts					2-Quarter Forecasts				
	Mean	Std.	MSE	MAE	Skew.	Mean	Std.	MSE	MAE	Skew.	Mean	Std.	MSE	MAE	Skew.	Mean	Std.	MSE	MAE	Skew.
Dmand	**0.82**	1.65	3.39	1.20	4.07	**0.81**	1.78	3.80	1.43	1.67	**0.76**	1.70	3.44	1.53	-0.39	**0.73**	1.98	4.42	1.72	-0.16
	(0.00)					(0.00)					(0.00)					(0.00)				
Supply	**0.75**	1.05	1.64	0.96	0.72	**1.00**	1.71	3.89	1.53	0.30	**0.99**	2.19	5.76	1.89	0.30	**0.90**	2.49	6.95	2.09	0.26
	(0.00)					(0.00)					(0.00)					(0.00)				
Stock withdraws	0.07	1.14	1.30	0.71	3.05	-0.05	1.21	1.47	0.81	2.34	-0.06	1.05	1.09	0.82	-0.13	-0.03	1.15	1.32	0.91	-0.20
	(0.22)					(0.31)					(0.25)					(0.39)				
Price	-0.19	2.01	4.04	1.44	-0.55	-0.40	11.43	129.88	8.14	-1.35	-0.60	22.13	486.81	15.67	-1.48	-0.76	27.24	737.28	19.74	-0.92
	(0.12)					(0.34)					(0.37)					(0.37)				
	3-Quarter Forecasts					4-Quarter Forecasts					5-Quarter Forecasts					6-Quarter Forecasts				
	Mean	Std.	MSE	MAE	Skew.	Mean	Std.	MSE	MAE	Skew.	Mean	Std.	MSE	MAE	Skew.	Mean	Std.	MSE	MAE	Skew.
Dmand	**0.72**	2.29	5.75	1.98	-0.04	**0.69**	2.55	6.93	2.21	-0.10	**0.61**	2.71	7.65	2.36	-0.07	0.34	2.75	7.51	2.35	-0.25
	(0.00)					(0.00)					(0.02)					(0.18)				
Supply	**0.74**	2.69	7.75	2.31	0.03	**0.77**	2.98	9.42	2.58	0.01	**0.71**	3.19	10.58	2.73	-0.10	0.52	3.11	9.73	2.70	-0.18
	(0.00)					(0.00)					(0.02)					(0.11)				
Stock withdraws	-0.04	1.17	1.35	0.94	-0.08	-0.04	1.21	1.46	0.93	-0.05	0.01	1.14	1.30	0.88	-0.38	-0.07	1.16	1.33	0.89	0.01
	(0.36)					(0.34)					(0.47)					(0.32)				
Price	-0.91	30.39	917.82	22.75	-0.48	-0.68	32.81	1069.45	24.98	-0.35	2.07	33.58	1120.58	25.33	-0.19	**7.33**	36.01	1329.11	28.06	-0.26
	(0.36)					(0.40)					(0.27)					(0.06)				

Notes: The forecasting errors are defined as, $e_{t+h|t} = 100 \times [log(Y_{t+h}) - log(F_{t+h|t})]$, for total demand, supply, RAC oil price; and as $E_{t+h|t} = Y_{t+h} - F_{t+h|t}$, for total stock withdraws, where we denote the errors at forecasting horizon h ($h = -1,\ 0,\ 1,\ \ldots,\ 6$) for the quarter t as $E_{t+h|t}$. P-values of Newey-West adjusted t-tests are reported in parenthesis, where boldface indicates the Newey–West corrected t-test for the null hypothesis that the forecasting error is significantly different from 0 at the 10% level. Further, the forecasts from EIA are Starting-month Report 1983Q1-2019Q4 (reported on Jan., Apr., Jul., Oct. since 1997:03).

Figure 1: Forecast Error Distributions



Note: For each variable the violin plots summarize the statistical properties of the forecast errors: the mean is denoted by the red line, the median by a white dot, the interquartile range by the vertical grey bar in the center of violin, the lower/upper adjacent values by the grey vertical lines stretched from the bar, defined as first quartile — 1.5 times the inter quartile range (IQR) and the third quartile + 1.5 times the IQR respectively. Wider sections of the violin plot represent a higher probability of observations taking that value, and the narrower sections correspond to a lower probability. Sample: 1983Q1–2019Q4.

to 35%. Visually the contrast is very stark, indicating the different behaviour depending on the forecast horizon. In addition the violin plots allow us to view skew more easily. For demand, supply and stock withdrawals forecast errors we observe positive skew for the backcast and nowcast horizons i.e., we observe more of the probability mass above the zero value. For the RAC oil price forecast errors, we observe a positive skew throughout, where the pattern is one of skew diminishing with longer forecast horizons. The violin plots allow us to view skew in the case where the distribution may be bi-modal (and so may not have the usual relationship between mean, median and skew). For example, both total demand and supply forecast errors suggest degrees bi-modal behaviour at higher forecasts horizons. Moreover forecast errors, in particular, for RAC oil price at short forecast horizon as well as total demand and supply display large outliers, which can effect usual moment based estimates of skewness.

We estimate standard joint (conditional) tests of unbiasedness or rationality, based on the following re-parameterised Mincer-Zarnowitz regression:

$$e_{t+h|t} = \alpha_h + \beta_h \tilde{f}_{t+h|t} + \varepsilon_{t,h}, \tag{2}$$

where $\tilde{f}_{t+h|t} = f_{t+h|t} - \mu_f$ is the demeaned forecast term, using the log level for total demand, total supply, and the RAC oil price and the level for total stock withdrawals.[8] Estimates of α_h reported in Table 2 (identical to estimates of the mean reported in Table 1) measure the forecast bias over the whole sample, whereas the interpretation of β_h is one of forecasts under predicting the outcomes if negative and over predicting if positive.[9]

In Table 2 we report the estimated coefficients $\hat{\alpha}_h$, $\hat{\beta}_h$ and their p-values (in parenthesis), testing *separate* conditional bias null hypotheses: $\alpha_h = 0$ and $\beta_h = 0$, respectively. We also report the p-value of χ^2-statistic for the *joint* test of the null hypothesis: $\alpha_h = 0 \cap \beta_h = 0$, where p-values (based on HAC robust estimates with bandwidth equal to 3) less than 0.10 are highlighted in bold characters. The results in Table 2 are consistent with the unconditional bias results reported in Table 1. They provide evidence that the backcasts, nowcasts and forecasts, for horizons $h = 1, 2, ..., 5$, for total demand, supply and stock withdrawals are biased (not rational). The joint null, $\alpha_h = 0 \cap \beta_h = 0$, is strongly rejected, with p-values of zero (in most cases) for demand, supply and stock withdrawals (where the highest p-value is 0.07). In contrast, as in the unconditional case, we find evidence suggesting RAC oil prices forecasts are rational, Where the p-values for the RAC oil price forecasts range from 0.11 to 0.82, providing strong evidence not rejecting the joint null.[10]

[8]See Appendix B defining the re-parameterisation.

[9]Here $\beta_h = b_h - 1$, where b_h is the coefficient from the conventional Mincer-Zarnowitz regression which regresses the outcome on a constant and the forecast.

[10]The exception to this pattern of results are at the longer forecast horizon of $h = 6$. The forecasts of total demand and supply are now unbiased, with p-values of 0.62 and 0.37 respectively, whilst the RAC oil price $h = 6$ forecast is biased, with a p-value of 0.01. Stock withdrawals remain biased at forecast horizon $h = 6$.

Table 2: MINCER-ZARNOWITZ FORECAST RATIONALITY TESTS

	1-Quarter Backcasts			Nowcasts			1-Quarter Forecasts			2-Quarter Forecasts		
	α	β	P (χ^2)	α	β	P (χ^2)	α	β	P (χ^2)	α	β	P (χ^2)
Dmand	**0.82**	**-0.01**	**0.00**	**0.81**	**-0.02**	**0.00**	**0.76**	**-0.01**	**0.00**	**0.73**	**-0.02**	**0.00**
	(0.00)	(0.10)		(0.00)	(0.02)		(0.00)	(0.01)		(0.00)	(0.01)	
Supply	**0.75**	**-0.01**	**0.00**	**1.00**	**-0.02**	**0.00**	**0.99**	**-0.02**	**0.00**	**0.90**	**-0.02**	**0.00**
	(0.00)	(0.03)		(0.00)	(0.00)		(0.00)	(0.01)		(0.00)	(0.01)	
Stock withdraws	0.07	**-0.46**	**0.01**	-0.05	**-0.52**	**0.00**	-0.06	**-0.40**	**0.00**	-0.03	**-0.51**	**0.00**
	(0.19)	(0.00)		(0.28)	(0.00)		(0.23)	(0.00)		(0.38)	(0.00)	
Price	-0.19	0.00	0.43	-0.40	-0.01	0.82	-0.60	-0.04	0.50	-0.76	**-0.06**	0.36
	(0.12)	(0.46)		(0.34)	(0.29)		(0.37)	(0.12)		(0.37)	(0.08)	
	3-Quarter Forecasts			4-Quarter Forecasts			5-Quarter Forecasts			6-Quarter Forecasts		
	α	β	P (χ^2)	α	$\beta_{0,V}$	P (χ^2)	α	β	P (χ^2)	α	β	P (χ^2)
Dmand	**0.72**	**-0.02**	**0.00**	**0.69**	**-0.02**	**0.00**	**0.61**	-0.02	**0.07**	0.34	-0.01	0.62
	(0.00)	(0.00)		(0.00)	(0.00)		(0.01)	(0.14)		(0.18)	(0.34)	
Supply	**0.74**	**-0.02**	**0.00**	**0.77**	**-0.03**	**0.00**	**0.71**	**-0.03**	**0.02**	0.52	-0.03	0.37
	(0.00)	(0.02)		(0.00)	(0.01)		(0.01)	(0.02)		(0.11)	(0.22)	
Stock withdraws	-0.04	**-0.54**	**0.00**	-0.04	**-0.59**	**0.00**	0.01	**-0.59**	**0.00**	-0.07	**-0.58**	**0.00**
	(0.34)	(0.00)		(0.31)	(0.00)		(0.47)	(0.00)		(0.30)	(0.00)	
Price	-0.91	**-0.07**	0.21	-0.68	**-0.09**	0.12	2.07	**-0.09**	0.11	**7.33**	**-0.16**	**0.01**
	(0.36)	(0.04)		(0.40)	(0.02)		(0.26)	(0.03)		(0.05)	(0.01)	

Notes: This table reports the estimated values from the following regressions: $e_{t+h|t} = \alpha_h + \beta_h(f_{t+h|t} - \overline{f}_{t+h|t}) + \varepsilon_{t,h}$, for variables total demand, supply, RAC; and $E_{t+h|t} = \alpha_h + \beta_h(F_{t+h|t} - \overline{F}_{t+h|t}) + \varepsilon_{t,h}$, for total stock withdraws. All parameters are estimated using Newey-West adjusted heteroscedastic-serial consistent Least-squares Regression. We report α_h, β_h and their p-values of the standard t-test statistic (in parenthesis). We also report the p-value of χ^2-statistic for the joint test of the null hypothesis: $\alpha_h = 0 \cap \beta_h = 0$. P-values less than 0.10 are marked with boldface. If the forecasts are unbiased, the α_h should be statistically insignificantly different from zero; if the forecasts are optimal, the β_h should be statistically insignificantly different from zero. Sample: 1983Q1–2019Q4.

3.2 Time-variation and interdependence in the EIA forecast errors

During the full sample period there have been different policy regimes and business cycles, and a wide range of crisis and economic events have occurred. Therefore it is useful to evaluate to what extent the properties of the forecasts have changed over time.

To examine time variation in unconditional and conditional bias we estimate equation (2) recursively, using a 10-year rolling window. Figure 2 reports estimates of $\hat{\alpha}_h$, $\hat{\beta}_h$ (and their 95% confidence bands) and the values of the χ^2-statistic for the joint test of rationality. The estimated values of the unconditional bias of the EIA forecasts, $\hat{\alpha_h}$, exhibit large degrees of time variation, both in terms of size and direction, throughout the full sample period. For example, the tendency of demand and supply forecasts to under-predict in the full sample looks to be driven by the period prior to the early 2000's and after 2010, with positive estimates. However, for the approximate period 2001 through to 2010 the forecasts over-predict. Demand and supply forecast error bias show strong positive co-movements, where the patterns are similar for $h = 1$ and $h = 4$ forecast horizons. Higher levels of negative bias or over-prediction are associated with recession periods.[11] The estimates of $\hat{\beta}_h$, plotted in the second column of Figure 2, highlight the degree of time variation of conditional bias. For total demand and supply the estimated values are negative, suggesting under-prediction, up until the period around 2012, where they become positive and hence over-predict. For stock withdrawals the estimated values are negative, but slowly become less negative over the period, suggesting under prediction which decreases in size overtime. For the RAC oil price forecast errors we observe negative estimates and under-prediction between 1992Q4 and 2005, which becomes less negative suggesting a switch towards a reduced tendency to under-predict.[12]

The p-values of the Chi-squared statistic, plotted in the third column of Figure 2, highlight the (approximate) period between 2004 and 2007 as having the strongest evidence of rational forecasts (for most variables). However, the data implies that the forecasts are *not* rational for the majority of the sample, across all variables, since the Chi-squared statistic p-values reject the joint tests of rationality.[13]

[11] There is also (a less pronounced) positive co-movement between the unconditional forecast bias for stock withdrawals and the RAC oil price. Both estimated parameters are positive, suggesting over prediction in 1992Q4, but then slowly evolve and become negative and therefore under predict from around 2000 on-wards. Notably they exhibit high negative values during the 2008 recession. However, we then observe a rise in the size of the bias (which is much larger for the RAC oil price) which drifts upwards from around 2008 on-wards towards zero, and in the case of the RAC oil price becomes positive.

[12] Examination of the time varying variance of the residuals from the re-parameterised Mincer-Zarnowitz regressions, which can be thought of as a bias adjusted forecast error volatility, reveal large fluctuations over time. For total demand and supply the movements (across both forecast horizons examined) range from around 1% to 3% and where the volatility for the RAC oil price is much larger ranging from around 15% to 35%. Stock withdrawals variability is between 0.5% and 1% (see Appendix C).

[13] Exceptions to this, other than the 2004-2007 period, are the $h = 1$ RAC oil price forecasts, which rejects rationality for a sustained period 2000-2007, but notably not for the 2008 crisis period.

Figure 2: Time Variation in Forecast Bias and Rationality

Note: The plots report the results of Mincer-Zarnowitz coefficients and forecast rationality tests of 1 and 4-quarter ahead Forecasts, 1992Q1–2019Q4 (rolling 10-year windows, the first estimation sample is 1983q1-1992q4 and the last 2010q1-2019q4.). The 90% confidence bands for the intercepts and slopes are calculated according to the Newey-West standard errors on parameters. Grey shading highlights periods of NBER designated US recessions.

To examine in more detail the change in forecast error distributions and interdependence overtime, in Figure 3 (on the diagonal) we plot the probability density functions (PDFs) of the forecast errors (for $h = 4$) for three sub-periods: 1983Q4-1999Q4 (blue), 2000Q1-2009Q4 (green) and 2010Q1-2019Q4 (purple). Overall we observe a large degree of time-variation in the distribution of forecast errors, in terms of mean values, standard deviations and shape, which shows non-normality and large amounts of skew. To highlight the interdependence of the forecast errors we also plot the *joint* probability densities for all the bi-variate pairings of our forecast errors (for forecast horizon $h = 4$). The darker the shading (for each colour) the higher is the probability mass. As an approximation, the greater the difference in interdependence over time, the more distinct each set of contour colours are. A visual examination of Figure 3 suggests that notable change over time is also a feature when considering forecast error interdependence. The joint distributions for the first two periods show greater dispersion across the range of forecast error combinations across variables, whereas the joint distributions for the post-2010 period (purple) are tighter, and where the most likely forecast error combination across variables differs from the earlier periods. This suggests interdependence has increased over time.[14]

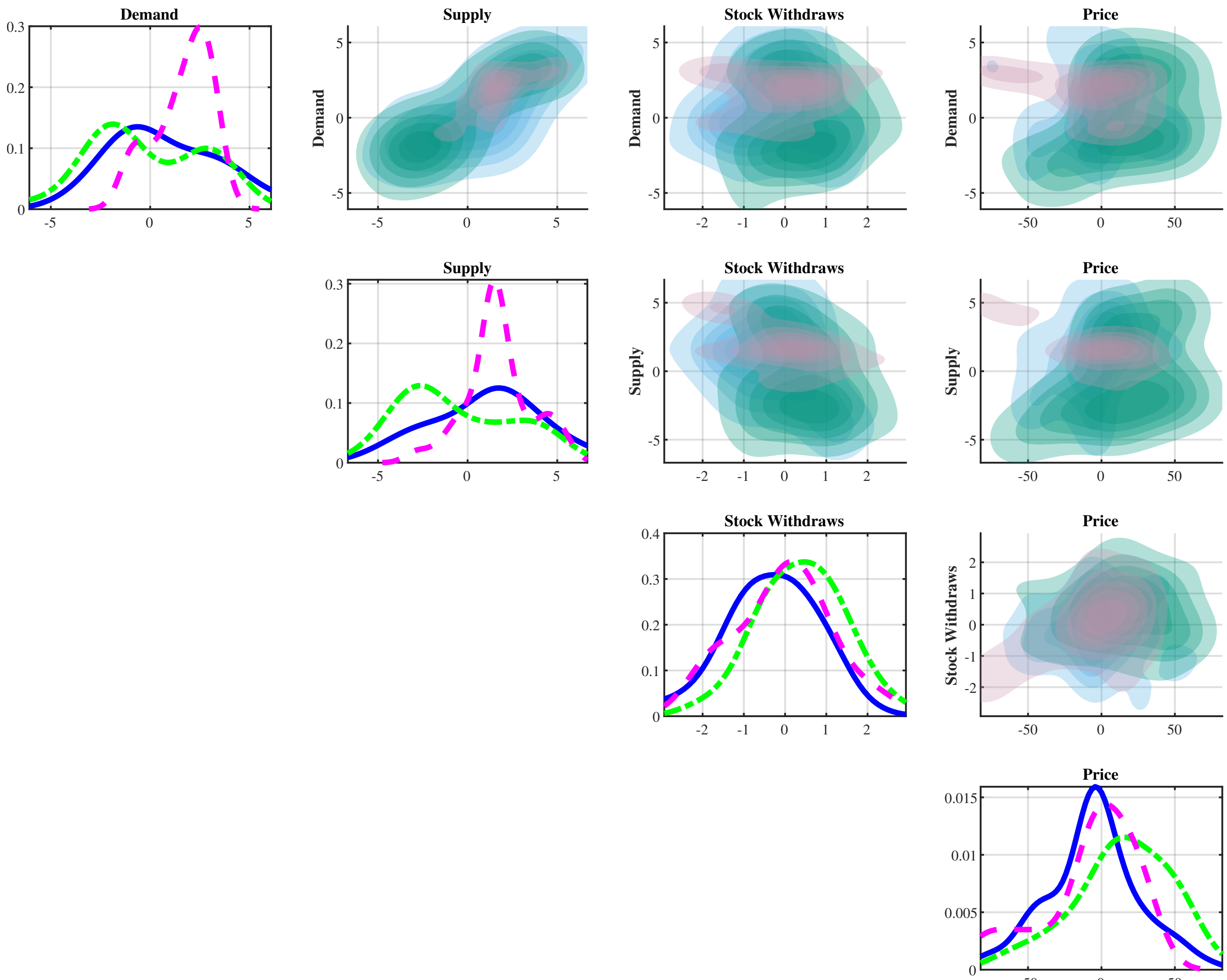
These features are illustrated by the combinations of forecast errors between demand and the other three EIA variables (see top row). Most prominent is demand's forecast errors interaction with RAC oil price forecast errors, indicating how the joint density changes over time. For supply and stock withdrawal forecast errors, small negative demand forecast error concentrations are centred around 0% for earlier time periods. However, from 2010 onwards, the joint density suggests a significant frequency of occurrence of large positive forecast error, which clearly differs from earlier periods. The joint or bi-variate distribution reveals significant multi-modalities, for instance, forecast errors for supply and demand over the past ten years reveal instances where both are positive or both are negative. This decade is characterised by risk-skewed price forecast error distributions and positive mistakes (i.e., under-prediction of both supply and demand).

Another important features of the data is the narrowing of the forecast error variance post 2010, where the variance for supply is typically larger than that of demand. This alternatively shaped joint density, toward positive RAC oil prices and right skew, is also true of supply (which, in general, is similar to demand) and, to a lesser extent, stock withdrawal forecast errors. Interactions with RAC oil price forecast errors appear to have the highest degree of time-variation, with the ability to distinguish the three sets of joint densities colours illustrating the extent of the change in interdependence over time. The joint densities involving stock withdrawals appear to be dispersed, implying weaker interdependence with other variables, whereas interactions with demand and supply forecast errors appear to be closer as the joint densities are tighter.[15]

[14]Note the pattern of the joint distribution varies across forecast horizons, for example the $h = 1$ forecast horizon joint densities show less dispersion than those at $h = 6$.

[15]Appendix C reports similar graphs highlighting the intricate relationship among the forecast errors for other forecast horizons.

Figure 3: Forecast Errors Distribution (4 quarters ahead forecasts)



Note: Probability Density Functions of the Forecast Errors (4 quarters ahead), for three sub-samples 1983Q4-1999Q4 (blue), 2000Q1-2009Q4 (green) and 2010Q1-2019Q4 (purple).

4 Can we rationalize the EIA forecasts?

Up until this point we have implicitly assumed that the EIA, when producing forecasts, is seeking to minimise a linear symmetric quadratic loss function, which is directionally symmetric, separable across variables, and is time-invariant. If that were the case the evidence is pointing to biases in the EIA forecast. However, the bias we have documented need not indicate non rational forecasts but rather they reflect an underlying asymmetric loss function, as the EIA prefers to either systematically under-predict or over-predict. In addition, when constructing the forecasts, the EIA uses the same model to produce or guide forecasts for the wide range of variables and as such they are not independent of one another. Moreover, any judgemental adjustments the EIA make to their model forecasts, reflecting off-model information for example, could take into consideration various interactions and dependencies that are known to exist between variable in the energy market.

Therefore in this section, we study the EIA's forecasting behaviour in an environment in which the loss functions, which use the forecast errors in demand, supply, stock withdrawal and the RAC oil prices are allowed to be asymmetric, and take into account inter-dependencies between the variables forecasted. To this end, we adopt the approach described in Komunjer and Owyang (2020), who propose a family of multivariate loss functions to test rationality of vector forecasts without assuming independence across variables. In Table 3, we report the results of the multivariate rationality tests of Komunjer and Owyang (2020). We test rationality conditional on three types of loss function: separable assuming symmetry (SS), separable with asymmetry (SA) and non-separable with asymmetry (NSA). The separable symmetric loss function takes the following form:

$$L_{SS} = \sum_{j=1}^{4} e_j^2 \tag{3}$$

separable asymmetric loss is defined as:

$$L_{SA} = \sum_{j=1}^{4} e_j^2 + \sum_{j=1}^{4} \tau_j sign(e_j) e_j \tag{4}$$

and non-separable asymmetric loss as:

$$L_{NSA} = \sum_{j=1}^{4} e_j^2 + \left(\sum_{j=1}^{4} \tau_j e_j\right) \left(\sum_{j=1}^{4} e_j^2\right)^{1/2} \tag{5}$$

where e_1, e_2, e_3 and e_4 are the forecast errors for demand, supply, stock withdrawals and the RAC oil price, and τ_1, τ_2, τ_3 and τ_4 are their estimated asymmetry parameters respectively. Where we assume asymmetry, we report the estimated asymmetry parameters τ (which take values between -1 and 1), and their t-statistics p-values. Joint J-tests of rationality are reported along with Wald tests testing the joint significance of the asymmetry parameters (see Komunjer and Owyang, 2020, for further details).[16]

We first consider the separable symmetric loss function L_{SS}, as this is the closest to the loss function used when assessing the results reported in Tables 1 and 2. Here the distinction is one of joint as opposed to independent evaluation, as no interdependence is allowed for. When analysed individually for the full-sample, we found clear evidence of bias and non-rationality for energy demand, supply and stock withdrawals but found no evidence of bias and non-rationality for the RAC oil price If we test the rationality assumption of the forecast errors jointly however,

[16]Following Komunjer and Owyang (2020), the instruments we use to test rationality are one lag of each of the forecasted series, available at the time the forecast is released. Note that the results are robust if the absolute values of lagged forecast errors are used as instruments.

assuming a L_{SS} loss function, we find the results on rationality are weaker (see the first row of Table 3). When evaluating the variables in the full-sample separately, the J-tests reject the null of rationality at the 10% significance level for backcasts, nowcasts and forecast horizons $h = 1$ and $h = 2$. But, using a joint test, we cannot reject the null of rational forecasts for the longer forecast horizons $h = 3$ through to $h = 6$. Hence joint evaluation makes a difference, particularly at the longer horizons.

If you allow the loss function, as in L_{SA}, to accommodate asymmetric behaviour whilst retaining separability, we observe J-tests which do not reject the null of rationality, for all forecast horizons (see seventh row of Table 3). Hence the introduction of asymmetric loss, coupled with joint evaluation, suggests that the EIA forecasts can be viewed as being rational.[17] The estimated asymmetry parameters for demand and supply, τ_1 and τ_2, are negative and significant at all forecast horizons. Values greater (less) than 0 indicate a greater loss for positive (negative) forecast errors i.e. in this instance the positive errors we observe suggest under-prediction induce *lower* loss. We observe the size of the asymmetry diminishing as the forecast horizon gets longer, with estimates of around -0.8 for short run horizons up to $h = 2$, which then fall to around -0.4 for $h = 3$ to $h = 5$, only to increase again back to around -0.8 for $h = 6$. The estimated asymmetry parameters for stock withdrawals and the RAC oil price, τ_3 and τ_4, are not significantly different from zero, implying symmetric loss (with the notable exception of forecast horizon $h = 6$). The Wald tests strongly reject the joint null of symmetry, $\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$, for all horizons, providing further evidence of the prevalence of asymmetric loss.

Finally, relaxing the assumption of separability and adopting the loss function L_{NSA}, which allows for interaction between variables and asymmetry, the evidence for rational forecasts and significant asymmetry is retained. The pattern of results, for the J-tests, Wald tests and significance of the τ parameters are very similar to those under the L_{SA} loss function. The degree of directional asymmetry is reduced, once separability is relaxed, suggesting that assuming separability leads us to infer more directional asymmetry than may actually be the case.[18] Nonetheless allowing for interactions between variables does not remove all the asymmetry.

4.1 Time-varying Asymmetry and Rationality Tests

As in the previous section it is of interest to examine whether the full-sample results using joint evaluation, examining asymmetries and allowing for interactions between forecasts vary over time. First, we examine time-variation in multi-variable rationality by constructing J-tests tests using forecast errors from a sequence of 10-year rolling-window samples. Multivariate rationality

[17]Note these results are consistent with those reported in Auffhammer (2007), who finds evidence of asymmetric loss for a range of EIA forecasts

[18]Komunjer and Owyang (2020) conduct a Monte Carlo study examining the consequences of mis-specifying the loss function, by not allowing for interdependence, and find that it can exaggerate the true degree of asymmetry

Table 3: Multivariate Rationality Test

	Backcasts	Nowcasts	Forecasts: 1-Quarter	2-Quarters	3-Quarters	4-Quarters	5-Quarters	6-Quarters
			Separable & Symmetric Loss					
J–stat	**24.37**	**24.62**	**24.15**	**23.71**	17.30	16.52	14.99	11.33
	(0.08)	(0.08)	(0.09)	(0.10)	(0.37)	(0.42)	(0.53)	(0.79)
			Separable & Asymmetric Loss					
Demand	**-0.79**	**-0.82**	**-0.78**	**-0.84**	**-0.36**	**-0.45**	**-0.43**	**-0.80**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)
Supply	**-0.91**	**-0.85**	**-0.80**	**-0.77**	**-0.34**	**-0.49**	**-0.43**	**-0.84**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.01)	(0.02)	(0.00)
Stock withdraws	-0.20	-0.00	0.11	-0.03	-0.02	0.16	0.05	**0.35**
	(0.11)	(0.98)	(0.35)	(0.79)	(0.87)	(0.24)	(0.70)	(0.02)
Price	0.18	-0.13	-0.04	-0.22	-0.23	-0.02	-0.01	**0.54**
	(0.25)	(0.29)	(0.75)	(0.18)	(0.20)	(0.90)	(0.94)	(0.03)
Wald $\chi^2(4)$	**518.13**	**319.24**	**126.43**	**183.04**	**18.81**	**23.70**	**24.58**	**93.28**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
J–stat	13.94	14.71	14.41	15.67	15.76	12.60	13.51	8.83
	(0.60)	(0.55)	(0.57)	(0.48)	(0.47)	(0.70)	(0.63)	(0.92)
			Nonseparable & Asymmetric Loss					
Demand	**-0.48**	**-0.24**	**-0.16**	**-0.15**	**-0.07**	**-0.07**	**-0.06**	**-0.09**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.04)	(0.06)	(0.01)
Supply	**-0.45**	**-0.27**	**-0.19**	**-0.15**	**-0.07**	**-0.08**	**-0.07**	**-0.11**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.04)	(0.07)	(0.00)
Stock withdraws	-0.01	-0.00	0.00	-0.01	-0.00	-0.00	-0.00	0.00
	(0.78)	(0.99)	(0.67)	(0.23)	(0.69)	(0.87)	(0.83)	(0.60)
Price	0.19	-0.03	-0.13	-0.28	-0.12	0.01	-0.19	**-0.75**
	(0.10)	(0.78)	(0.35)	(0.14)	(0.54)	(0.97)	(0.42)	(0.00)
Wald $\chi^2(4)$	**151.35**	**86.48**	**64.10**	**63.83**	**12.46**	**10.92**	**11.17**	**158.24**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.03)	(0.02)	(0.00)
J–stat	17.36	14.74	15.98	15.23	14.68	11.61	13.66	10.00
	(0.36)	(0.54)	(0.45)	(0.51)	(0.55)	(0.77)	(0.62)	(0.87)

Notes: The table reports: (i) J–stat tests of the null of rationalizability of the forecasts (see Komunjer and Owyang, 2020). Where P-Values of the J–test correspond to a χ^2 distribution with 16 degrees of freedom, (ii) for the asymmetric loss functions we report the estimated asymmetric loss parameters (τ, with p-values are shown in parentheses), and (iii) Wald Tests of the null that the asymmetric parameters are jointly equal to zero. The instruments are ([1, $\Delta_h D_t$, $\Delta_h S_t$, $\Delta_h I_t$, $\Delta_h P_t$]): $\Delta_h D_t = \frac{400}{h} \times [log(D_t) - log(D_{t-h})]$, $\Delta_h S_t = \frac{400}{h} \times [log(S_t) - log(S_{t-h})]$, $\Delta_h I_t = I_t - I_{t-h}$, and $\Delta_h P_t = \frac{400}{h} \times [log(P_t) - log(P_{t-h})]$, all calculated in real-time where we use annual rate (i.e. h=4). Boldface denotes significance at the 10% level.

is not rejected for any of the subsamples considered.[19]

Figure 4 plots the estimated asymmetry parameters, τ_1, τ_2, τ_3 and τ_4, for the forecasts of demand, supply, stock withdrawals and the oil price respectively. The estimates use the most general loss function which allows for non-separability and asymmetry, L_{NSA}, and we plot two forecast horizons, $h = 1$ and $h = 4$. Overall, we observe sizeable time-variation in the degree of asymmetric loss for the forecasts of our four variables. For example, the estimated RAC oil price asymmetry parameter, τ_4 (bottom right panel), exhibits the highest degree of time-variation, with large fluctuations, suggesting both positive and negative asymmetric loss. For $h = 1$ we observe fluctuations in τ_4 of between 1 and -1, which contrasts with the full-sample estimated value of -0.13. For $h = 4$, the asymmetric loss is volatile for the periods up to around 2000 and after 2015, but is constant with an estimated value for τ_4 of near -1 between these two periods. The fluctuating estimate of the RAC oil price asymmetry parameter, between two extreme regimes makes for a different implication regarding the loss. The positive bias in the forecast observed for the period 2000-2015 in Figure 2, implying under-prediction, reflects the presence of asymmetry in the EIA loss. Specifically, negative values of τ_4 imply greater loss penalising negative forecast errors, i.e. over this period over-predicting prices is more costly and hence under-prediction minimises the loss.

The time fluctuations in the asymmetry parameters for demand and supply also suggest that during the post 2010 period (and early 1990's) the size of the loss for a given degree of under-prediction is greater than that incurred during the mid 2000 period. Furthermore, a feature of the estimated asymmetry parameters for demand and supply forecast errors, τ_2 and τ_3, is their high level of co-movement and significant degrees of time-variation compared to the full sample estimates. Higher positive asymmetry, for both $h = 1$ and $h = 4$, are apparent between 1992-1998 and in the post 2010 periods. This is particularly true for the $h = 1$ forecast horizon, with estimated values of τ_2 and τ_3 of around 0.2/0.3, slightly higher than the full sample estimates. However, a notable feature, for both $h = 1$ and $h = 4$, is the period between 1998-2010, where the estimated asymmetry parameters are around zero suggesting symmetric loss is dominant during this period.[20]

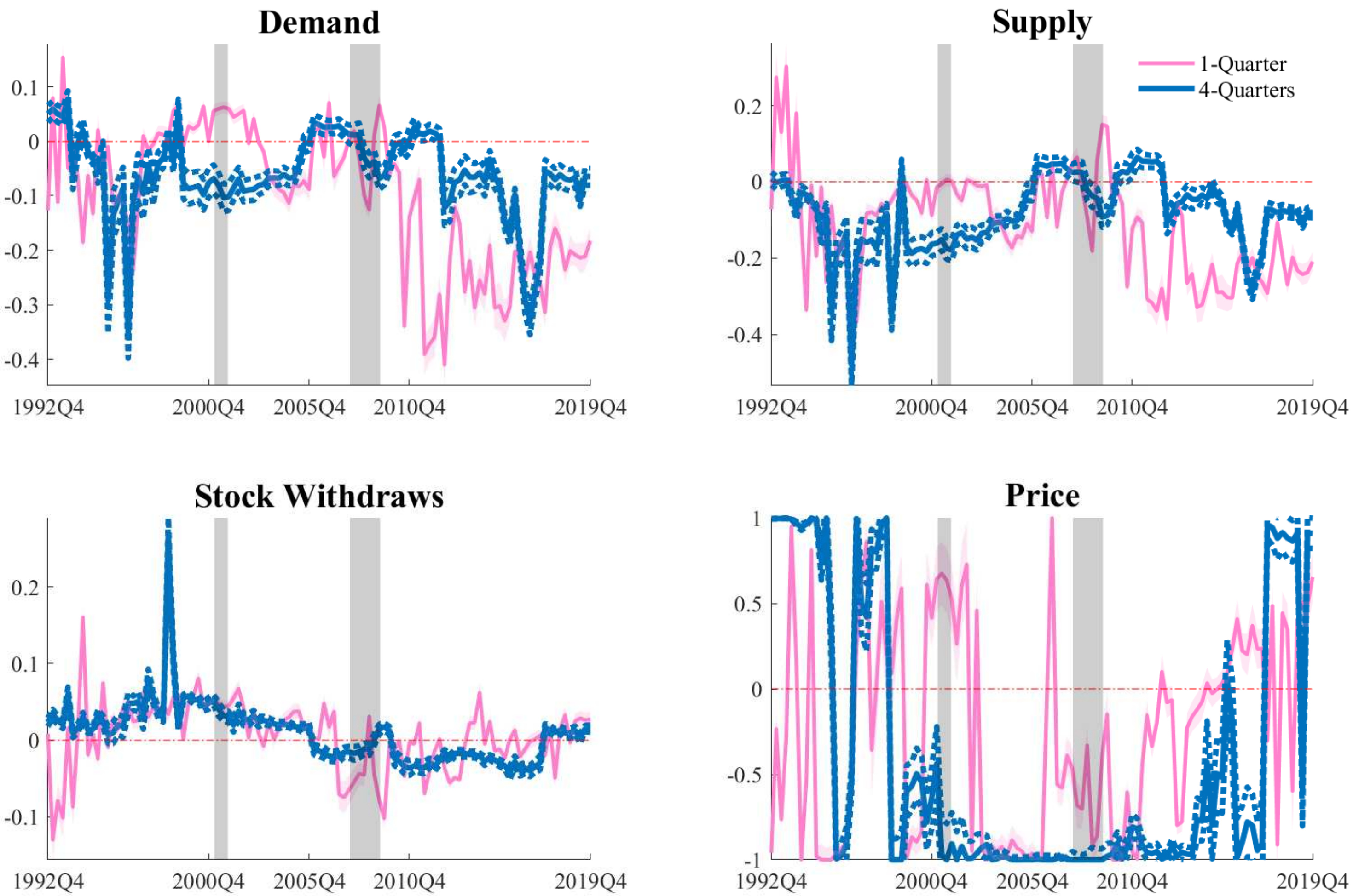
5 Can the EIA beat the Random Walk forecast?

In this section we undertake a relative evaluation exercise of the EIA point forecasts, first adopting the conventional univariate based approach assuming a symmetric quadratic loss function, as used by the EIA, and then widening the evaluation exercise by adopting loss functions which

[19] Plots of the J-tests are available in the appendix C.

[20] The estimated asymmetry parameter, τ_3, for stock withdrawals forecast errors are the least volatile and close to zero, mostly indicating symmetric loss over the period.

Figure 4: Time varying estimates of the asymmetry parameters (τ)



Notes: The estimated asymmetry parameters for the forecasts of demand, supply, stock withdrawals and the oil price, 1992Q1–2019Q4 (rolling 10 year window). The pink shaded area and dotted blue lines are the 95% confidence intervals of 1 and 4-quarter ahead Forecasts respectively.

evaluate the variables jointly, allowing for non-separability and asymmetry.

We examine the forecasts mean square errors (MSE) relative to those generated by a random walk (plus seasonal dummy) benchmark model, over the full sample period 1983Q1-2019Q4 and overtime.[21] The benchmark model differs for the RAC oil price forecasts, where we adopt the pure random walk forecasts, as used in Baumeister and Kilian (2015) and Garratt et al. (2019). See Appendix D for a more detailed description of the benchmark models.

Table 4 reports the MSE ratios relative to the benchmarks, where in parenthesis we report the p-value of the Harvey et al. (1997) small-sample adjusted Diebold and Mariano (1995) two-sided test statistic of whether the EIA forecasts are significantly different from the benchmark models. A value of less than one favours the EIA forecasts over the benchmark models.

Table 4: MSE RATIOS

	Backcast	Nowcast	1-Quarter	2-Quarters	3-Quarters	4-Quarters	5-Quarters	6-Quarters
Dmand	**0.44** (0.00)	**0.39** (0.00)	**0.64** (0.00)	1.09 (0.74)	0.68 (0.15)	**0.59** (0.07)	0.96 (0.87)	**1.26** (0.02)
Supply	**0.32** (0.00)	**0.48** (0.01)	0.68 (0.11)	0.67 (0.31)	0.59 (0.23)	0.72 (0.28)	0.82 (0.51)	**1.38** (0.02)
Stock withdraws	**0.39** (0.00)	**0.37** (0.00)	**0.41** (0.00)	**0.46** (0.01)	**0.34** (0.00)	**0.27** (0.01)	**0.32** (0.03)	**0.56** (0.02)
Price	**0.25** (0.00)	**1.89** (0.02)	1.17 (0.28)	1.05 (0.43)	0.98 (0.80)	0.89 (0.32)	0.89 (0.40)	0.92 (0.67)

Notes: The table reports Mean Squared Error (MSE) ratios, of the EIA forecasts relative to the random walk plus seasonal adjustment benchmark for demand, supply and stock withdraws and random walk for oil prices. A value of less than one represents an improvement of the EIA over the benchmark. P-values of a Harvey et al. (1997) small-sample adjusted Diebold and Mariano (1995) 2-sided test are reported in brackets after the MSE ratios. Boldface denotes significance at the 10% level.

We observe that a majority of MSE ratios are less than one, implying improved performance of EIA forecasts relative to the benchmark. The gains are large, in particular for backcasts, nowcasts and at forecast horizon $h = 1$ (ranging from around 30% to 60%), but they tend to worsen (with MSE ratios still being less than one) in the sense of becoming not statistically different (at the 10% level) from the benchmark model as the forecast horizons increases, $h = 2$ through to $h = 6$. There are notable differences in forecast performance across variables. Only for the backcasts are the RAC oil price forecasts significantly improved relative to the benchmark. Whereas for the RAC oil price nowcasts and forecasts we cannot reject the null that the EIA forecasts are not significantly different from the benchmark. In contrast, stock withdrawals show large significant

[21] The conventional benchmark model is typically a pure random-walk, but such a model is disadvantaged when considering total demand, supply, and stock withdrawals by not including seasonality, which is likely to be present in the underlying series. EIA forecasts most likely explicitly factor these seasonal effects into their forecasts, and as such would make a comparison with a model that does not take this into account uneven. We therefore adopt as a benchmark a random walk with seasonal drift coefficients, the latter are estimated using real time information. Our benchmark is deliberately chosen so as to be simple and easily replicable.

MSE ratio gains (typically around 60% and upto 73%). Statistically significant MSE gains for demand and supply, are observed for the backcasts, nowcasts and at forecast horizon $h = 1$. Beyond $h = 1$, into the longer horizon forecasts, whilst we mostly observe MSE ratios of less than one, these are not statistically significant from the benchmark (one exception being for demand at $h = 4$).

To evaluate the stability of the relative performance of the EIA forecasts with respect to our random walk with seasonal drift benchmark, we plot in Figure 5 MSE ratios calculated over rolling 10 years windows. We observe considerable time variation in relative forecast performance. In particular, the EIA forecasts show significant gains in predicting stock withdrawals, especially 4 quarters ahead, at all points in the sample. The inability of improving over the simple benchmark for RAC oil prices forecasts is confirmed for all periods, with the exception of a short period after the 2001 recession where the EIA forecasts are superior for long range forecasts but are worse for short term predictions. For the forecasts of demand and supply, we document significant superior forecasts for oil production for the decade starting around 1997, whereas the EIA forecasts for demand appear to be significantly more accurate than the simple benchmark after the Great Recession and for a small period around the 2001 recession for short term predictions.

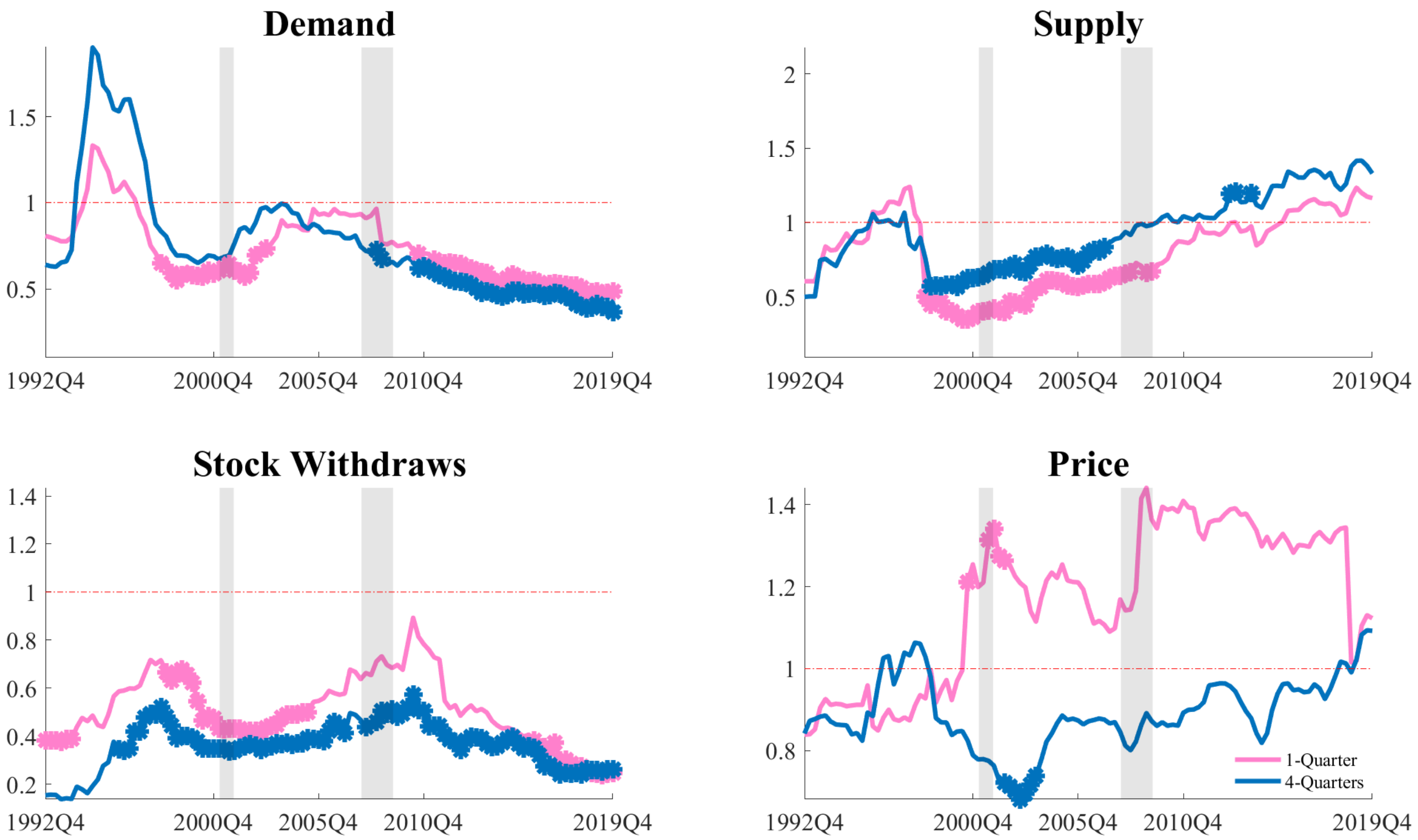
In addition, we also evaluate to what extent the EIA forecasts are successful in providing an assessment of the direction of change of the underlying variables. Looking at the annual growth/change forecasts we note a strong ability of the EIA forecasts to detect the direction of change in all the variables.[22]

The evidence presented in Section 4 suggested a role for loss functions that allow for asymmetry and non-separability, as under this type of loss, we cannot reject the null of rationality for the EIA forecasts, in contrast to when using symmetric separable loss, where we find evidence of bias or non-rationality. A natural question to ask therefore is: how does the EIA forecasts compare to benchmark models using joint or multivariate, asymmetric, non-separable loss functions? Do we form a different view of the performance of the EIA forecasts relative to standard benchmarks, if we move away from the symmetric separable quadratic loss functions? To answer these questions we construct loss ratios, using the EIA and benchmark forecast errors, defined as: L_{SS}^{EIA}/L_{SS}^{RW}, L_{SA}^{EIA}/L_{SA}^{RW} and $L_{NSA}^{EIA}/L_{NSA}^{RW}$, where RW denotes the random walk benchmark, and the loss functions are as defined in Section 4. For the τ asymmetry parameters, we use the estimated values reported in Table 3. Table 5 reports the three multivariate loss ratios, where values less than one indicate an improvement relative to the benchmark model and we test whether they are significantly different from the benchmarks using Harvey et al. (1997) small-sample adjusted Diebold and Mariano (1995) two-sided test.

For the backcasts we observe large (around 65%) and significant improvements relative to the benchmark models. However, we observe a strong reversal of this result for nowcasts and forecast

[22]Results are available in Appendix C.

Figure 5: Rolling MSE Ratios



Note: Rolling MSE Ratios of 1 and 4-quarter ahead Forecasts, 1992Q1–2019Q4 (rolling 10 year window). Marks denote p-values < 10% for the Harvey et al. (1997) small-sample adjusted Diebold and Mariano (1995) 2-sided test.

horizons $h = 1$ and $h = 2$, with loss ratios larger than one, although (with the exception of the nowcasts) they are not significantly different from the benchmarks. For the forecast horizons $h = 3$ through to $h = 6$, the loss ratios are less than one, with the lowest showing a 13% improvement. But, in all cases, we cannot reject the null of being equal to the benchmark model losses. Overall, using alternative loss functions, suggests a worsening of the EIA forecast performance relative to the benchmarks, where the major change is the poor performance at the shorter forecast horizons.

Table 5: MULTIVARIATE LOSS RATIOS (Full Sample)

	Backcast	Nowcast	1-Quarter	2-Quarters	3-Quarters	4-Quarters	5-Quarters	6-Quarters
Sep. & Symmetry	**0.31**	**1.72**	1.24	1.05	0.95	0.87	0.88	0.90
	(0.00)	(0.03)	(0.11)	(0.44)	(0.56)	(0.24)	(0.32)	(0.58)
Sep. & Asymmetry	**0.35**	**1.64**	1.23	1.03	0.92	0.87	0.87	0.97
	(0.00)	(0.03)	(0.11)	(0.71)	(0.43)	(0.24)	(0.32)	(0.82)
Nonsep. & Asymmetry	**0.35**	**1.71**	1.22	1.02	0.94	0.88	0.85	0.83
	(0.00)	(0.03)	(0.11)	(0.78)	(0.50)	(0.25)	(0.29)	(0.50)

Notes: The table reports multivariate loss ratios of the EIA forecasts relative to the RW with seasonal dummies benchmark. The loss functions are as defined in the text: separable and symmetric, separable and asymmetric and non-separable and asymmetric. A value of less than one represents an improvement of the EIA over the benchmark. p-values of a Harvey et al. (1997) small-sample adjusted Diebold and Mariano (1995) 2-sided test are reported in brackets. Boldface denotes significance at the 10% level.

Previously we identified the post 2010 period as being one where both the univariate and joint distributions of the forecast errors showed significant differences compared to those in the pre 2010 period. Therefore, in Table 6, we report the same set of multivariate loss ratios as in Table 5 but for the 2010Q1-2019Q4 period. The results are similar to the full sample results, in that EIA backcasts do well relative to the benchmarks, but at most other horizons either do worse or are not significantly different from the benchmark.[23]

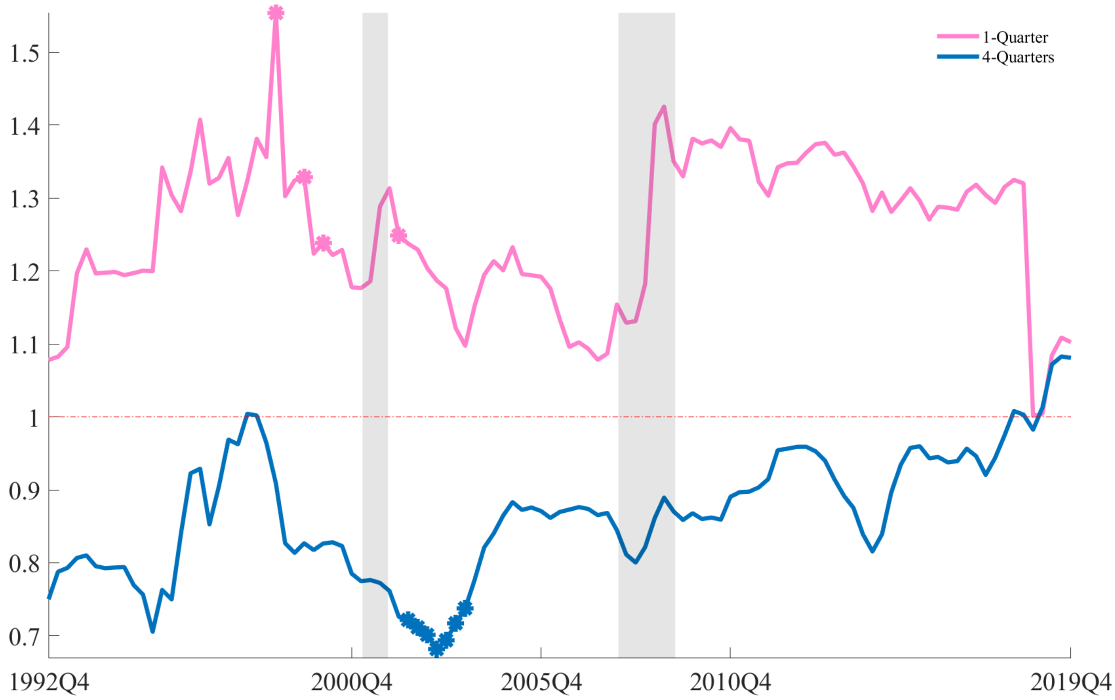
Table 6: MULTIVARIATE LOSS RATIOS (Post-2010 Sample)

	Backcast	Nowcast	1-Quarter	2-Quarters	3-Quarters	4-Quarters	5-Quarters	6-Quarters
Sep. & Symmetry	**0.51**	1.54	1.10	1.06	1.13	1.08	0.89	0.94
	(0.01)	(0.12)	(0.20)	(0.75)	(0.62)	(0.72)	(0.28)	(0.27)
Sep. & Asymmetry	**0.43**	1.38	1.09	1.10	1.29	1.26	1.08	0.39
	(0.01)	(0.19)	(0.30)	(0.67)	(0.45)	(0.41)	(0.47)	(0.11)
Nonsep. & Asymmetry	**0.58**	1.51	1.10	1.11	1.29	1.26	**0.67**	0.84
	(0.06)	(0.13)	(0.26)	(0.67)	(0.44)	(0.41)	(0.03)	(0.26)

Notes: The table reports multivariate loss ratios of the EIA forecasts relative to the RW with seasonal dummies benchmark. The loss functions are as defined in the text: separable and symmetric, separable and asymmetric and non-separable and asymmetric. A value of less than one represents an improvement of the EIA over the benchmark. p-values of a Harvey et al. (1997) small-sample adjusted Diebold and Mariano (1995) 2-sided test are reported in brackets. Boldface denotes significance at the 10% level.

[23] A notable exception to this is at forecast horizon $h = 5$, for asymmetric loss functions with separability and non-separability. Here we observe ratio of 0.67, which is significantly different from the benchmark. Hence there are, albeit limited, differences in the post 2010 performance.

Figure 6: Non-separable Asymmetric Loss Ratios



Notes: 1 and 4-quarter ahead forecasts relative multivariate loss assuming non-separable and asymmetric loss, $L_{NSA}^{EIA}/L_{NSA}^{RW}$, where RW denotes the random walk benchmark, and the loss functions are as defined in Section 4. The forecasts are evaluated on a rolling 10 year window (for the period 1992Q1-2019Q4).

Time -variation is also a feature of the multi-variate loss ratios, as illustrated in Figure 6 which plots the non-separable asymmetric loss (the most general loss function) for forecast horizons $h = 1$ and $h = 4$, for the period 1992Q1-2019Q4, calculated using a 10-year rolling window. Although there is time variation in the results, overall they confirm the full-sample results. Namely, that EIA forecasts perform worse than the the RW+SD benchmark models at short forecast horizons (in the figure $h = 1$) for all periods, and at the longer forecast horizons ($h = 4$) have ratios less than one with a tendency to increase towards one over time. However, they are (nearly) always not significantly different from the RW+SD benchmark model. The exception to this is a short period around 2002-2003, where the $h = 4$ ratio is around 0.7 and is significantly different from the benchmark.

6 Conclusions

The EIA's forecasts represent and quantify the agency's narrative of the evolving forces in the oil market, which embody a view of the strength of global demand and supply, their balance, as well as their impact on prices. When evaluating these forecasts it is important to account for the inherent inter-dependencies across the forecast variables as well as any potential asymmetries in the implicit loss function faced by the forecaster. Conventional use of separable and symmetric loss functions do not consider differing costs of over and under predicting that we observe, nor do they allow for the process which produces the forecast which intrinsically link forecasts together.

We document substantial, time varying, biases in the individual forecasts produced by the EIA. However, by using non-separable and asymmetric loss we find we can rationalise these biases. The implied asymmetric loss gives less weight to under-prediction of both demand and supply, while for oil prices, we document significant regime changes in the implied loss due to asymmetry. In particular, the period of rising oil prices from the late 90s to the oil collapse in 2007 is associated with higher cost of underpredicting prices, whereas the last 10 years, as well as the decade following the collapse of the oil price in the mid 80s, is associated with larger costs of overpredicting prices.

The EIA forecasts outperform the naive random walk forecast when evaluated using the conventional MSE loss. Yet, this loss results in large and significant biases in the EIA forecasts. Allowing for the interactions between the forecast variables and asymmetries embodies the complex relationships between key variables in the world petroleum market. When we evaluate the EIA forecasts allowing for joint evaluation using non-separable asymmetric loss functions, which rationalize the observed biases in the EIA forecasts, we observe a deterioration in the EIA's forecast performance with respect to a naive random walk benchmark, particularly at short forecast horizons. While EIA forecasts are undoubtedly informative, oil market participants and observers should take into account that those forecasts do not necessarily reflect the EIA unbiased view on the market.

References

Alquist, R., Kilian, L., and Vigfusson, R. J. (2013). Forecasting the price of oil. In *Economic Forecasting Volume 2A*, pages 427–507. North Holland.

Auffhammer, M. (2007). The rationality of eia forecasts under symmetric and asymmetric loss. *Resource and Energy Economics*, 29(4):102–121.

Baumeister, C. and Kilian, L. (2015). Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics*, 33(3):338–351.

Baumeister, C., Kilian, L., and Lee, T. K. (2014). Are there gains from pooling real-time oil price forecasts? *Energy Economics*, 46(3):S33–S43.

Bora, S. S., Katchova, A. L., and Kuethe, T. H. (2021). The Rationality of USDA Forecasts under Multivariate Asymmetric Loss. *American Journal of Agricultural Economics*, 103(3):1006–1033.

Caunedo, J., Riccardo, D., Komunjer, I., and Owyang, M. T. (2012). Asymmetry, complementarities, and state dependence in federal reserve forecasts. *Journal of Money Credit and Banking*, 52:1066–1080.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

EIA (2020). Annual Energy Outlook (AEO) Retrospective Review: Evaluation of AEO 2020 and Previous Reference Case Projections. Analysis & projections, U.S. Energy Information Administration.

Garratt, A., Vahey, S. P., and Zhang, Y. (2019). Real-time forecast combinations for the oil price. *Journal of Applied Econometrics*, 34(3):456–462.

Hamilton, J. D. (2009). Causes and Consequences of the Oil Shock of 2007-08. *Brookings Papers on Economic Activity*, 40(1 (Spring):215–283.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Kaach, L. H., Apt, J., Morgan, M. G., and McSharry, P. (2017). Empirical prediction intervals improve energy forecasting. *Proceedings of the National Academy of Sciences of the United States of America*, 114(33):8752–8757.

Komunjer, I. and Owyang, M. T. (2020). Multivariate forecast evaluation and rationality testing. *Review of Economics and Statistics*, 94(4):205–228.

Lady, G. M. (2010). Evaluating long term forecasts. *Energy Economics*, 32(3):450–457.

Mamatzakis, E. and Koutsomanoli-Filippaki, A. (2014). Testing the rationality of doe's energy price forecasts under asymmetric loss preferences. *Energy Policy*, 68:547–575.

Pesaran, M. H. and Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485):325–337.

Pindyck, R. S. (1980). Uncertainty and Exhaustible Resource Markets. *Journal of Political Economy*, 88(6):1203–1225.

Sanders, D., Manfredo, M. R., and Boris, K. (2008). Accuracy and efficiency in the u.s. department of energy's short-term supply forecasts. *Energy Economics*, 30(1):1192–1207.

Sanders, D., Manfredo, M. R., and Boris, K. (2009). Evaluating information in multiple horizon forecasts: The doe's energy price forecasts. *Energy Economics*, 31(1):189–196.

Winebrake, J. J. and Sakva, D. (2006). An evaluation of errors in us energy forecasts: 1982-2003. *Energy Policy*, 34(3):3475–3483.

Working, H. (1949). The theory of price of storage. *The American Economic Review*, 39(6):1254–1262.

Asymmetry and Interdependence when Evaluating U.S. Energy Information Agency Forecasts

– Online Appendix –

Anthony Garratt
Warwick Business School, University of Warwick
Email: Anthony.Garratt@wbs.ac.uk

Ivan Petrella
Warwick Business School, University of Warwick
Email: ivan.petrella@wbs.ac.uk

Yunyi Zhang
School of Management, China Institute for Studies in Energy Policy, Xiamen University
Email: yunyi_zhang@xmu.edu.cn

A Additional Data Details

We examine quarterly world petroleum market forecasts from 1983Q1 to 2019Q4 for four key variables: total demand, total supply, total stock withdrawals (inventory), and the oil price (Refiners' Acquisition Cost, RAC hereafter). Specifically,

- **Total Demand** is defined as millions of barrels per day. For the OECD countries this measures the consumption of petroleum and is synonymous with "petroleum product supplied", defined in the glossary of the EIA Petroleum Supply Monthly (DOE/EIA-0109). For the non-OECD countries, consumption of petroleum is "apparent consumption," which includes internal consumption, refinery fuel and loss, and bunkering.[1] In the empirical analysis when defining forecast errors we use the natural logarithm of this series.

- **Total Supply** is defined as millions of barrels per day. It includes production of crude oil (including lease condensates), natural gas plant liquids, biofuels, other liquids, and refinery processing gains.[2] In the empirical analysis when defining forecast errors we use the natural logarithm of this series.

- **Stocks Total Withdrawals** is defined as millions of barrels per day. It includes OECD and 'Other Stock Draws and Balance', whose size is larger than the size of OECD draws.[3] As this series can take negative values, in the empirical analysis when defining forecast errors, we use this series as defined and not the natural logarithm.

- **The RAC crude oil price** is defined as dollars per barrel and is the average of all EIA-14 refiners' monthly cost reports. In the empirical analysis when defining forecast errors we use the natural logarithm of this series.

Table A.1 reports the mean, standard deviation, and skew for the actual observed data that the EIA seeks to forecast, for the sample period 1983Q1-2019Q4. Column 1 reports the *levels* data of our four variables, whilst the remaining columns contain the h-period growth rates.

[1] Definitional changes have occurred overtime: (i) before 1991Q3, total demand accounted for market economies only; while after 1991Q3 (including 1991Q3) it accounts for the world petroleum demand. The EIA data "vintage" reported in 1991Q3 used the new definition for observations starting in 1990Q1 and (ii) since 2001M12, total OECD demand reported by the EIA includes all other OECD countries in addition to the U.S. (50 states), U.S. Territories, Canada, Europe, Japan, Australia and New Zealand. Under the data vintage reported in 2001M12, the EIA reports use the new definition with observations starting in 2000Q1.

[2] Before 1991Q3, total supply only accounted for the U.S. (50 states), OPEC, other non-OPEC, and net centrally planned economies exports; while after 1991Q3 (including 1991Q3) it accounts U.S. (50 states), Canada, North Sea, other OECD, OPEC, U.S.S.R., China, Mexico and other non-OECD. Under the data vintage 1991Q3, EIA reports using the new definition were started with the observations on 1990Q1.

[3] Before 1996Q3, OECD closing stocks reported by EIA were world total stocks excluding stocks held in the former CPEs, while after 1996Q3 (including 1996Q3) it is the closing stocks (millions barrels) for OECD only. In the EIA data vintage for 1996Q3 the reports started with observations in 1995Q1.

Table A.2 and Figures A.1 and A.2 describe the forecast errors using forecasts reported at the first, mid and end month of each quarter (for the period where only monthly forecasts are available). Recall that in the main text we use the first month in the quarter. The violin plots highlight the main point visually, with the numbers in the tables confirming, namely that the forecast error distributions and properties are very similar across the three first, mid and end month definitions. As such the results in the main text are robust to this choice.

In Figure A.3, we report the probability density functions for the full set of forecast horizons, $h = 0, 1, 2, 3, 4, 5, 6$, for the three sub-periods 1983Q1-1999Q4, 2000Q1-2009Q4 and 2010Q1-2019Q4. They confirm the general pattern we observed at $h = 4$ in the main text (row 5) is also true of the other forecast horizons we consider. Finally, Figure A.4 reports the multivariate density of the forecast errors for h=1, matching the equivalent plot for h=4 in Figure 3 in the main text.

Table A.1: DESCRIPTIVE STATISTICS FOR THE OBSERVED DATA

		Level			1-Quarter			2-Quarter			3-Quarter	
	Mean	Std.	Skew.	Mean	Std.	Skew.	Mean	Std.	Skew.	Mean	Std.	Skew.
Demand	75.86	16.49	-0.44	1.60	9.83	-0.58	1.66	6.40	0.16	1.65	3.39	0.78
Supply	75.92	16.49	-0.44	1.85	6.93	0.46	1.76	4.83	0.94	1.70	3.25	0.80
Stock withdraws	-0.13	1.14	0.23	-0.08	1.46	-0.19	-0.02	1.80	-0.37	-0.01	1.55	0.03
Price	40.90	28.62	1.05	1.66	62.77	-1.04	1.79	48.23	-0.97	1.90	38.17	-0.70
		4-Quarter			5-Quarter			6-Quarter				
	Mean	Std.	Skew.	Mean	Std.	Skew.	Mean	Std.	Skew.			
Demand	1.64	1.44	-0.58	1.62	2.21	-0.41	1.63	2.28	0.19			
Supply	1.64	2.45	0.55	1.66	2.18	0.37	1.67	1.94	0.33			
Stock withdraws	-0.01	1.35	-0.22	-0.03	1.75	-0.34	-0.03	1.94	-0.00			
Price	1.88	32.54	-0.35	1.88	28.91	-0.34	1.96	25.37	-0.39			

Notes: 1. We use the most resent vintage (2020:06) for actual observations denoted as A_t, and the first three columns measures the basic statistics of level total demand and supply, total stock withdraws, and dollar price of the RAC oil price. 2. The rest columns calculate the changes for variable A_t as $\Delta A_{t+Qh} = \frac{400}{h} \times [log(A_{t+h}) - log(A_t)]$, while for total stock net withdraws and the US net stock withdraws we use $\Delta A_{t+Qh} = A_{t+h} - A_t$. Sample: 1983Q1-2019Q4.

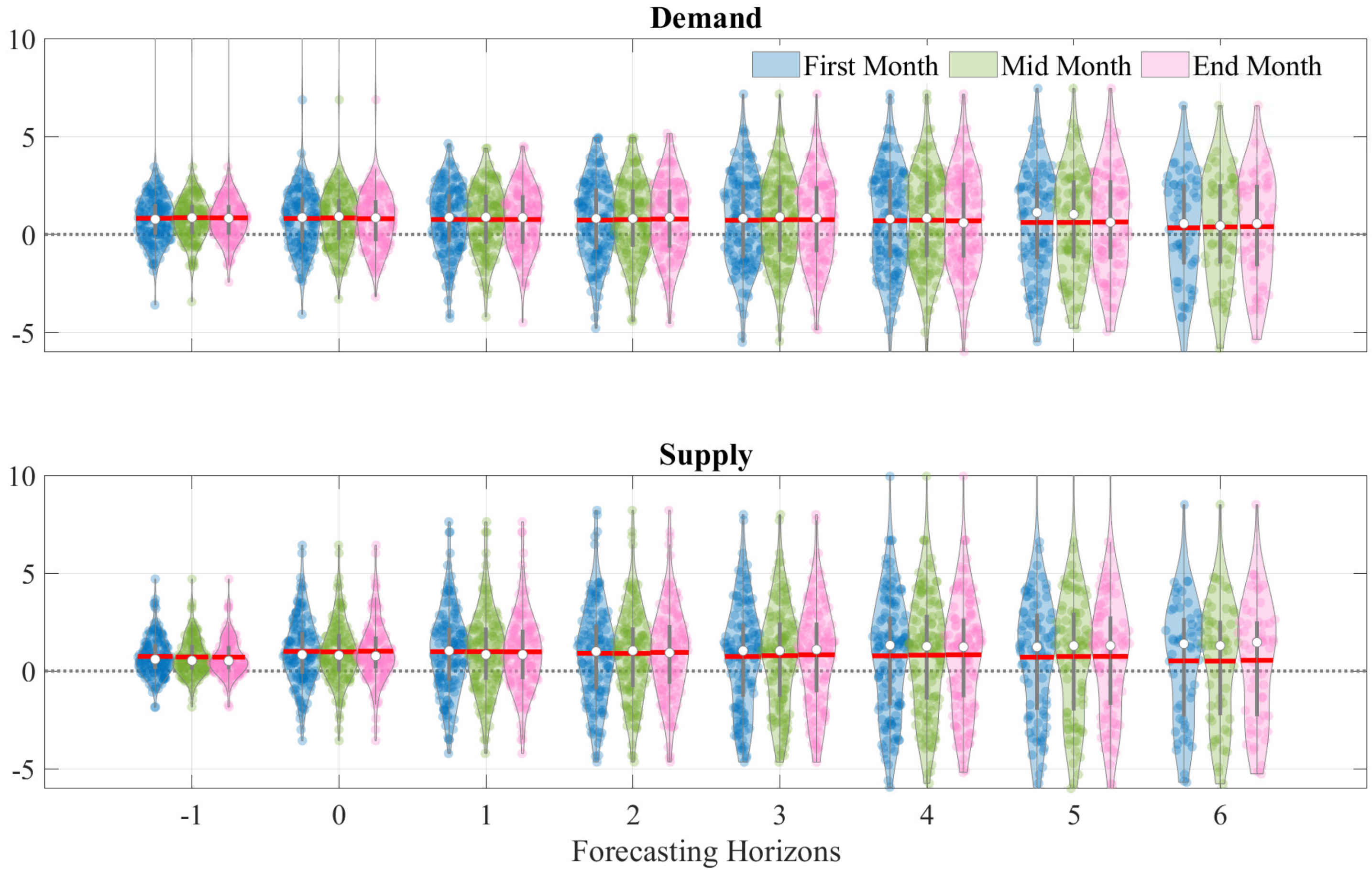
Table A.2: DESCRIPTIVE STATISTICS OF FORECAST ERRORS

	1-Quarter Backcasts					Nowcasts					1-Quarter Forecasts					2-Quarter Forecasts				
	Mean(P)	Std.	MSE	MAE	Skew.	Mean(P)	Std.	MSE	MAE	Skew.	Mean(P)	Std.	MSE	MAE	Skew.	Mean(P)	Std.	MSE	MAE	Skew.
Starting-month Report (reported on Jan., Apr., Jul., Oct. since 1997:03)																				
Demand	**-0.82***(0.00)	1.65	3.39	1.20	-4.07	**-0.81***(0.00)	1.78	3.80	1.43	-1.67	**-0.76***(0.00)	1.70	3.44	1.53	0.39	**-0.73***(0.00)	1.98	4.42	1.72	0.16
Supply	**-0.75***(0.00)	1.05	1.64	0.96	-0.72	**-1.00***(0.00)	1.71	3.89	1.53	-0.30	**-0.99***(0.00)	2.19	5.76	1.89	-0.30	**-0.90***(0.00)	2.49	6.95	2.09	-0.26
Stock withdraws	-0.07(0.22)	1.14	1.30	0.71	-3.05	0.05(0.31)	1.21	1.47	0.81	-2.34	0.06(0.25)	1.05	1.09	0.82	0.13	0.03(0.39)	1.15	1.32	0.91	0.20
Price	0.19(0.12)	2.01	4.04	1.44	0.55	0.40(0.34)	11.43	129.88	8.14	1.35	0.60(0.37)	22.13	486.81	15.67	1.48	0.76(0.37)	27.24	737.28	19.74	0.92
Mid-month Report (reported on Feb., May., Aug., Nov. since 1997:03)																				
Demand	**-0.85***(0.00)	1.60	3.27	1.17	-4.45	**-0.83***(0.00)	1.71	3.59	1.39	-1.94	**-0.75***(0.00)	1.59	3.09	1.45	0.28	**-0.76***(0.00)	1.90	4.16	1.66	0.15
Supply	**-0.72***(0.00)	0.98	1.47	0.90	-0.78	**-0.98***(0.00)	1.59	3.47	1.41	-0.43	**-0.99***(0.00)	2.12	5.43	1.81	-0.40	**-0.90***(0.00)	2.33	6.20	1.98	-0.19
Stock withdraws	-0.10(0.13)	1.10	1.22	0.66	-3.38	0.04(0.36)	1.21	1.46	0.81	-2.33	0.06(0.26)	1.05	1.10	0.83	0.21	0.03(0.35)	1.11	1.22	0.88	0.09
Price	0.06(0.29)	1.44	2.06	1.05	0.33	0.17(0.40)	8.06	64.61	6.06	0.63	0.11(0.47)	19.07	361.26	14.08	1.08	0.31(0.44)	25.32	636.64	18.53	0.97
End-month Report (reported on Mar., Jun., Sep., Dec. since 1997:03)																				
Demand	**-0.84***(0.00)	1.55	3.10	1.11	-5.00	**-0.80***(0.00)	1.70	3.52	1.37	-1.99	**-0.76***(0.00)	1.58	3.06	1.44	0.24	**-0.78***(0.00)	1.90	4.19	1.66	0.13
Supply	**-0.71***(0.00)	0.97	1.43	0.88	-0.85	**-1.01***(0.00)	1.55	3.40	1.38	-0.51	**-0.98***(0.00)	2.01	4.99	1.69	-0.51	**-0.95***(0.00)	2.33	6.32	1.98	-0.22
Stock withdraws	-0.10(0.12)	1.08	1.16	0.64	-3.70	0.08(0.21)	1.24	1.54	0.87	-2.28	0.07(0.19)	1.04	1.07	0.82	0.25	0.04(0.33)	1.09	1.19	0.86	0.09
Price	0.05(0.30)	1.13	1.27	0.63	0.97	0.43(0.22)	6.85	46.85	4.71	0.97	0.63(0.33)	17.69	311.04	12.90	0.99	1.00(0.31)	25.04	623.82	18.17	1.10

	3-Quarter Forecasts					4-Quarter Forecasts					5-Quarter Forecasts					6-Quarter Forecasts				
	Mean(P)	Std.	MSE	MAE	Skew.	Mean(P)	Std.	MSE	MAE	Skew.	Mean(P)	Std.	MSE	MAE	Skew.	Mean(P)	Std.	MSE	MAE	Skew.
Starting-month Report (reported on Jan., Apr., Jul., Oct. since 1997:03)																				
Demand	**-0.72***(0.00)	2.29	5.75	1.98	0.04	**-0.69***(0.00)	2.55	6.93	2.21	0.10	**-0.61***(0.02)	2.71	7.65	2.36	0.07	-0.34(0.18)	2.75	7.51	2.35	0.25
Supply	**-0.74***(0.00)	2.69	7.75	2.31	-0.03	**-0.77***(0.00)	2.98	9.42	2.58	-0.01	**-0.71***(0.02)	3.19	10.58	2.73	0.10	-0.52(0.11)	3.11	9.73	2.70	0.18
Stock withdraws	0.04(0.36)	1.17	1.35	0.94	0.08	0.04(0.34)	1.21	1.46	0.93	0.05	-0.01(0.47)	1.14	1.30	0.88	0.38	0.07(0.32)	1.16	1.33	0.89	-0.01
Price	0.91(0.36)	30.39	917.82	22.75	0.48	0.68(0.40)	32.81	1069.45	24.98	0.35	-2.07(0.27)	33.58	1120.58	25.33	0.19	**-7.33***(0.06)	36.01	1329.11	28.06	0.26
Mid-month Report (reported on Feb., May., Aug., Nov. since 1997:03)																				
Demand	**-0.75***(0.00)	2.22	5.46	1.92	0.01	**-0.72***(0.00)	2.47	6.55	2.12	0.03	**-0.61***(0.01)	2.61	7.12	2.26	-0.05	-0.38(0.15)	2.68	7.18	2.29	0.20
Supply	**-0.78***(0.00)	2.62	7.41	2.25	-0.03	**-0.81***(0.00)	2.96	9.36	2.56	-0.02	**-0.74***(0.01)	3.21	10.76	2.75	0.08	-0.51(0.11)	3.10	9.71	2.70	0.14
Stock withdraws	0.06(0.27)	1.12	1.24	0.88	0.17	0.05(0.32)	1.21	1.46	0.94	0.09	-0.01(0.46)	1.17	1.35	0.87	0.32	0.03(0.41)	1.12	1.24	0.86	-0.09
Price	0.70(0.39)	28.98	834.49	21.72	0.57	0.71(0.39)	31.56	989.35	23.89	0.38	-1.20(0.35)	32.49	1046.77	24.24	0.22	**-6.13***(0.09)	34.99	1241.29	26.58	0.25
End-month Report (reported on Mar., Jun., Sep., Dec. since 1997:03)																				
Demand	**-0.75***(0.00)	2.20	5.35	1.90	-0.01	**-0.69***(0.00)	2.46	6.51	2.10	-0.00	**-0.63***(0.01)	2.61	7.11	2.25	-0.06	-0.39(0.14)	2.71	7.36	2.35	0.13
Supply	**-0.83***(0.00)	2.57	7.23	2.23	-0.01	**-0.83***(0.00)	2.87	8.86	2.49	-0.06	**-0.75***(0.01)	3.14	10.28	2.67	0.01	**-0.55***(0.10)	3.10	9.74	2.71	0.07
Stock withdraws	0.07(0.23)	1.10	1.20	0.87	0.24	0.10(0.16)	1.22	1.50	0.95	0.13	0.00(0.48)	1.13	1.26	0.87	0.18	0.06(0.34)	1.14	1.28	0.88	0.02
Price	1.11(0.32)	28.95	833.67	21.37	0.65	1.07(0.34)	31.62	994.12	23.61	0.36	-0.42(0.45)	32.48	1045.00	24.14	0.19	-5.21(0.12)	34.74	1213.99	26.67	0.19

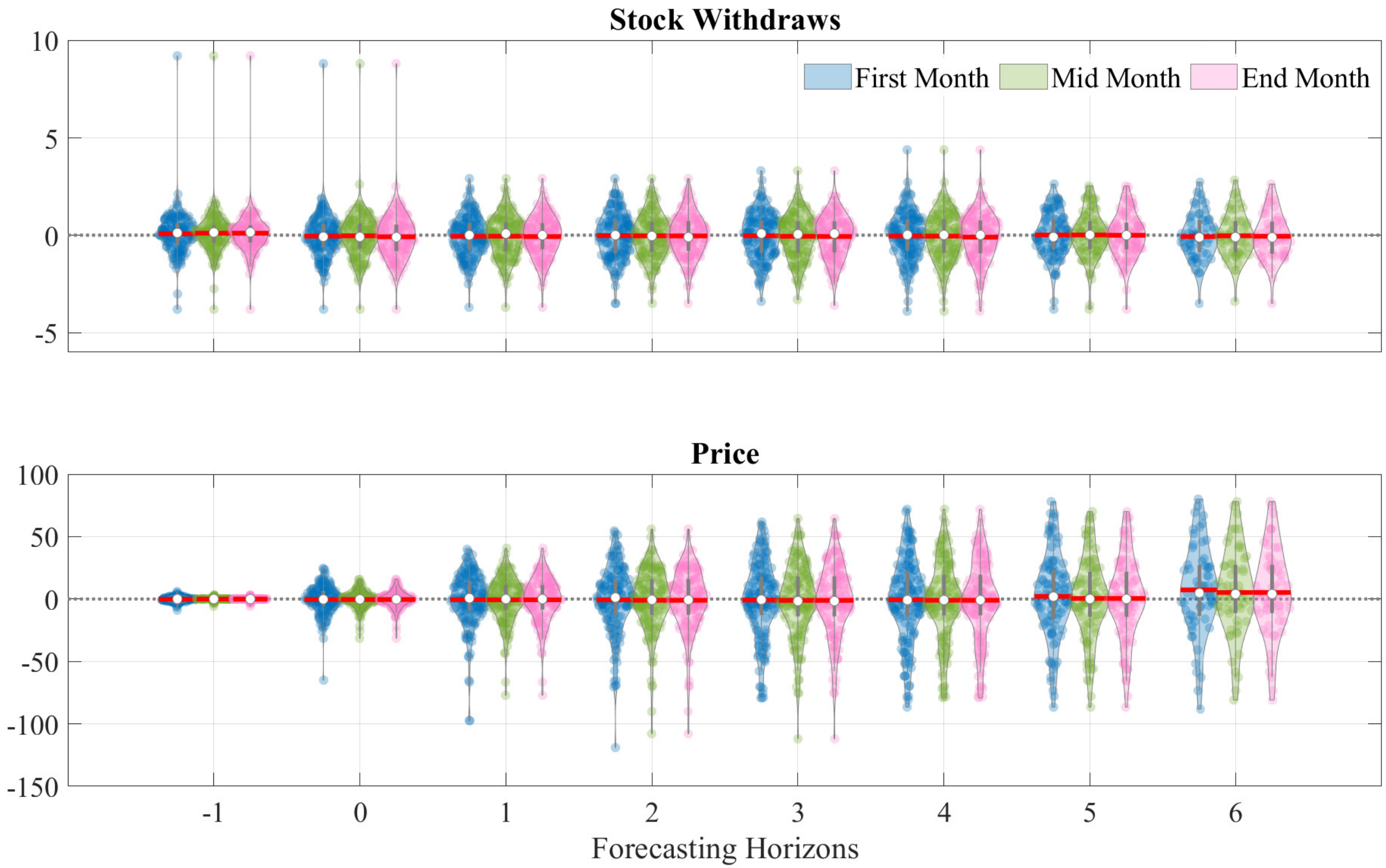
Notes: 1. The forecasting errors at (log) level for each variables, $e_{t+h} = 100 \times [log(F_{t+h|t}) - log(A_{t+h})]$, for total demand, supply, RAC; and $E_{t+h} = F_{t+h|t} - A_{t+h}$, for total stock withdraws, where we denote the errors at forecasting horizon h ($h = -1, 0, 1, \ldots, 6$) for the quarter t as E_{t+h}.
2. P-value of Newey-West adjusted t-test are reported in paranthesis where * indicates the Newey–West corrected t-test for the null hypothesis that the forecasting error is significantly different from 0 at 10% level. Sample: 1983Q1–2019Q4.

Figure A.1: Forecast Error Distributions



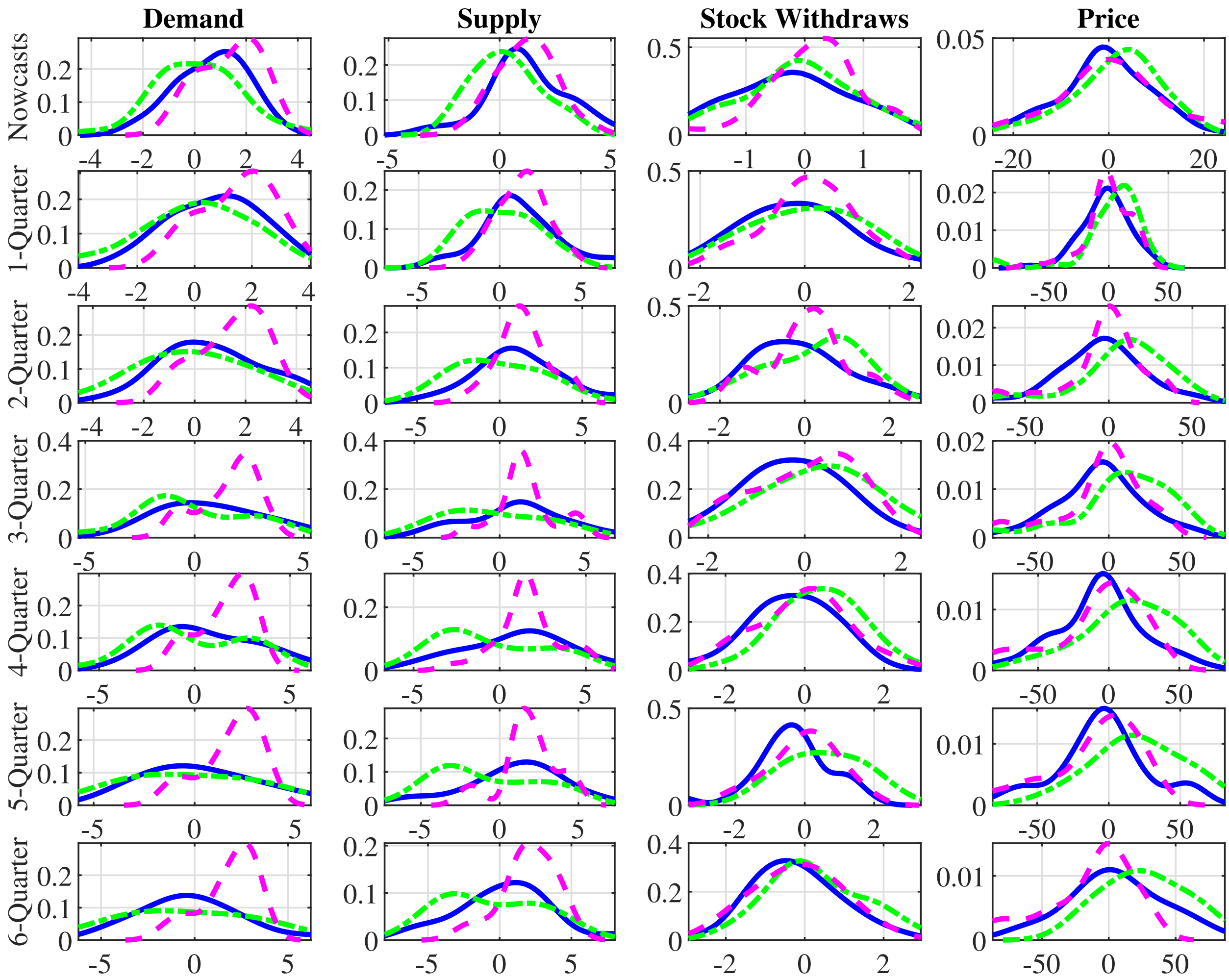
Note: For each variable the violin plots summarize the statistical properties of the forecast errors: the mean is denoted by the red line, the median by a white dot, the interquartile range by the vertical grey bar in the center of violin, the lower/upper adjacent values by the grey vertical lines stretched from the bar, defined as first quartile — 1.5 times the inter quartile range (IQR) and the third quartile + 1.5 times the IQR respectively. Wider sections of the violin plot represent a higher probability of observations taking that value, and the narrower sections correspond to a lower probability. Sample: 1983Q1–2019Q4.

Figure A.2: Forecast Error Distributions



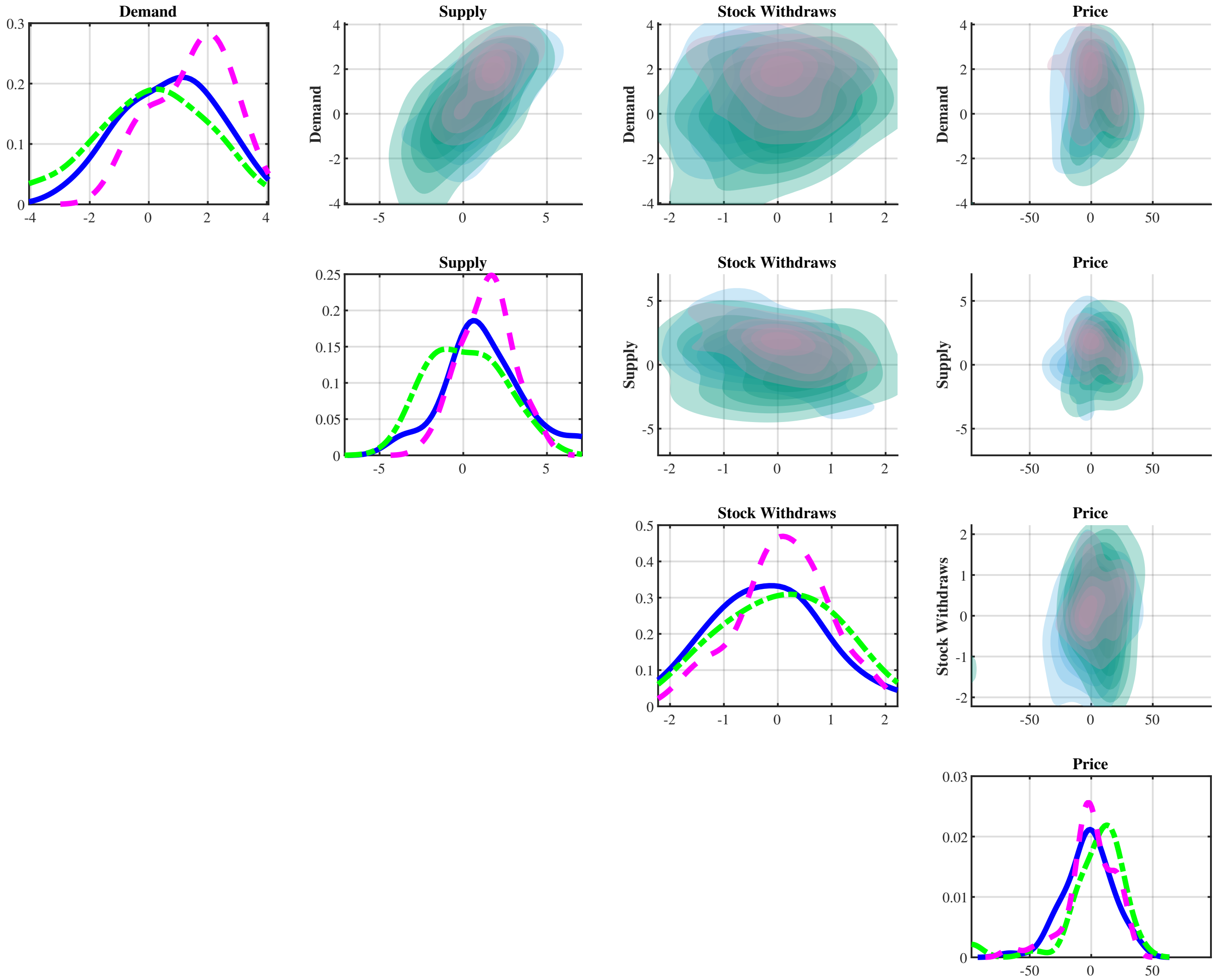
Note: For each variable the violin plots summarize the statistical properties of the forecast errors: the mean is denoted by the red line, the median by a white dot, the interquartile range by the vertical grey bar in the center of violin, the lower/upper adjacent values by the grey vertical lines stretched from the bar, defined as first quartile — 1.5 times the inter quartile range (IQR) and the third quartile + 1.5 times the IQR respectively. Wider sections of the violin plot represent a higher probability of observations taking that value, and the narrower sections correspond to a lower probability. Sample: 1983Q1–2019Q4.

Figure A.3: Forecast Errors Distribution



Note: Probability Density Functions of the Forecast Errors, for three sub-samples 1983Q4-1999Q4 (blue), 2000Q1-2009Q4 (green) and 2010Q1-2019Q4 (purple).

Figure A.4: Forecast Errors Distribution (1 quarter ahead forecasts)



Note: Probability Density Functions of the Forecast Errors (1 quarter ahead), for three sub-samples 1983Q4-1999Q4 (blue), 2000Q1-2009Q4 (green) and 2010Q1-2019Q4 (purple).

B Reparametrization of the Mincer-Zarnowitz regression

The mean values of the forecast errors reported in Table 2 measure unconditional bias and are equivalent to the estimated constant, α_h, from the following regression:

$$e_{t+h} = \alpha_h + u_{t+h},$$

where $e_{t+h} = y_{t+h} - f_{t+h|t}$ and $\alpha_h = \mu_y - \mu_f$. Conventional specifications of the Mincer-Zarnowitz (MZ) regression, testing conditional bias, take either the form: $y_{t+h} = a_h + b_h f_{t+h|t} + u_{t+h}$, where the joint null hypothesis of unbiasedness is $H_0: \ a_h = 0, b_h = 1$, or:

$$e_{t+h} = a_h + \beta_h f_{t+h|t} + u_{t+h},$$

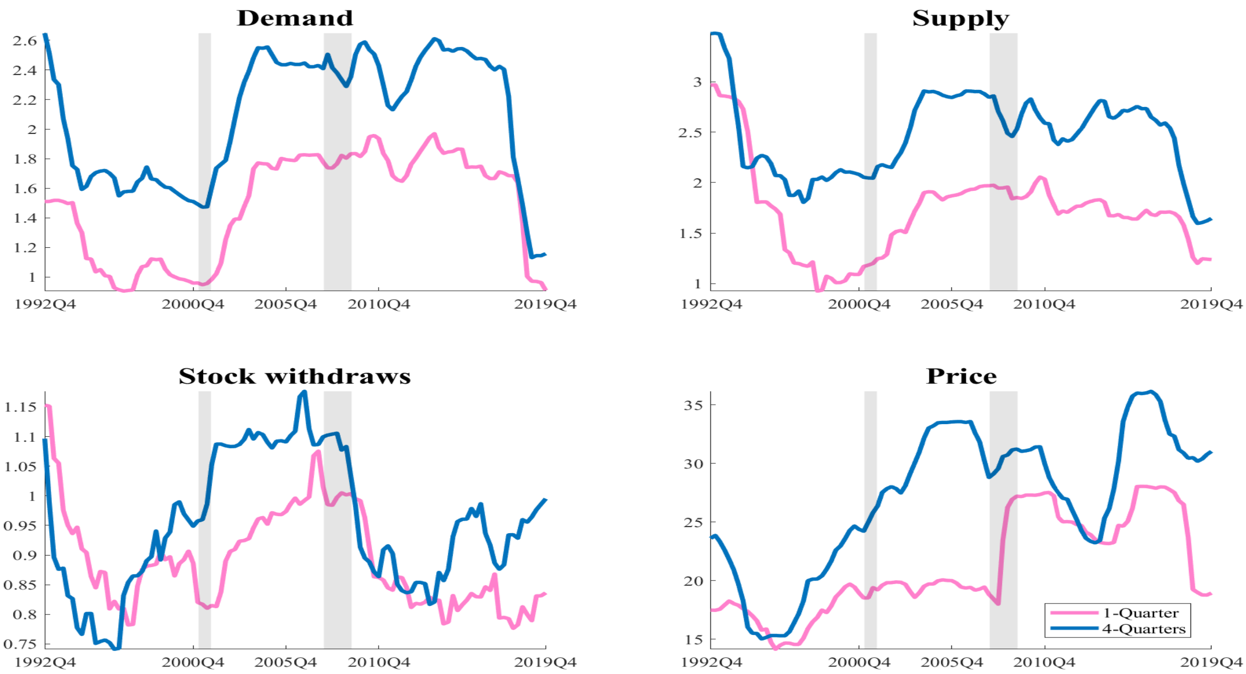
where $\beta_h = b_h - 1$, and the null hypothesis for unbiasedness is $H_0: \ a_h = 0, \beta_h = 0$. The interpretation of β_h or b_h is typically one of either under or over prediction, if $\beta_h < 0$ or $\beta_h > 0$ respectively, or equivalently $b_h < 1$ or $b_h > 1$. Furthermore, it is possible to reparametrize the regression so that the constant term parameter can be interpreted as a measure of (unconditional) bias. In particular, we choose to estimate and report (in Table 3 and Figure 3) estimates of the following *reparamerised* MZ regression:

$$e_{t+h} = \alpha_h + \beta_h \widetilde{f}_{t+h|t} + u_{t+h},$$

where $\widetilde{f}_{t+h|t} = f_{t+h|t} - \mu_f$. The reparametrization leaves the interpretation of the slope coefficient: $\beta_h = \frac{cov\left(e_{t+h}, \widetilde{f}_{t+h|t}\right)}{var\left(\widetilde{f}_{t+h|t}\right)} = \frac{cov\left(e_{t+h}, f_{t+h|t}\right)}{var\left(f_{t+h|t}\right)}$ unaffected, as $\alpha_h = \mu_e = \mu_y - \mu_f$ since $\mu_{\widetilde{f}} = 0$.

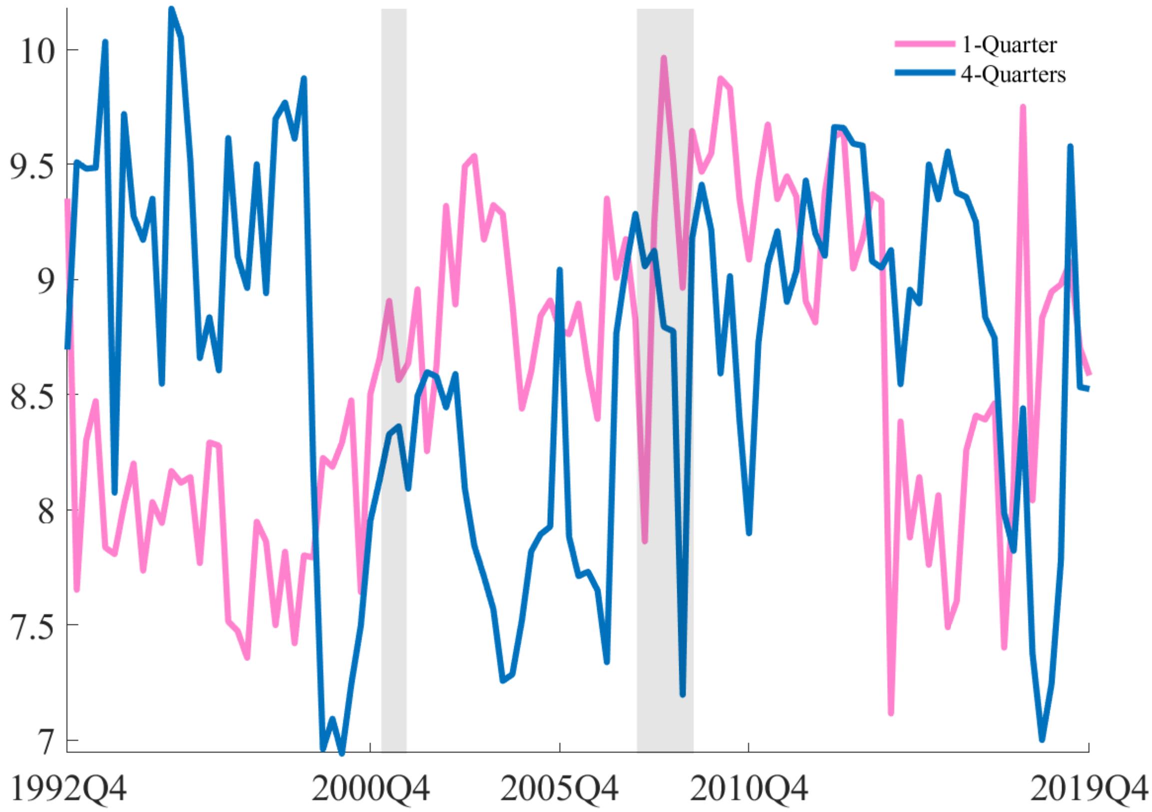
C Additional Results

Figure C.1: Time Varying Variance of The Residuals from The Mincer-Zarnowitz Regressions



Note: The plots report the results of Mincer-Zarnowitz variance (i.e. the square root of the variance) of 1 and 4-quarter ahead Forecasts, 1992Q1–2019Q4 (rolling 10-year windows, the first estimation sample is 1983q1-1992q4 and the last 2010q1-2019q4.). Grey shading highlights periods of NBER designated US recessions.

Figure C.2: Rolling J-tests of the non-separable asymmetric rationality



Note: Rolling J-statistics of 1 and 4-quarter ahead Forecasts, 1992Q1–2019Q4 (rolling 10 year window). Marks denote the p-values $< 10\%$ from the χ^2-distribution. Please see more details of the J-test in Komunjer and Owyang (2020).

Table C.1: SUCCESS RATIOS

	Backcast	Nowcast	1-Quarter	2-Quarters	3-Quarters	4-Quarters	5-Quarters	6-Quarters
Dmand	**0.82** (0.00)	**0.87** (0.00)	**0.82** (0.00)	0.86 (0.77)	**0.93** (0.00)	**0.80** (0.00)	**0.81** (0.01)	0.95 (NaN)
Supply	**0.81** (0.00)	**0.78** (0.00)	**0.74** (0.00)	0.70 (0.94)	0.78 (0.32)	0.83 (0.19)	0.84 (0.43)	0.93 (0.81)
Stock withdraws	**0.79** (0.00)	**0.90** (0.00)	**0.78** (0.00)	**0.63** (0.00)	**0.76** (0.00)	**0.85** (0.00)	**0.81** (0.00)	**0.73** (0.00)
Price	**0.92** (0.00)	**0.85** (0.00)	**0.67** (0.00)	**0.70** (0.00)	**0.69** (0.00)	**0.65** (0.00)	**0.64** (0.00)	0.60 (0.16)

Notes: The table reports Success Ratios, defined as the proportion of EIA forecasts which correctly forecast the direction of change. The P-values in brackets are for the Pesaran and Timmermann (2009) 2-sided test, where the null is no directional accuracy or a random walk i.e. the ratio is equal to 0.5. Boldface denotes significance at the 10% level. Sample 1983Q1:2019Q4.

D Construction of random walk benchmark

The demand, supply, and stock withdrawals display seasonal fluctuations. Hence, we assume as a benchmark a simple random walk with (constant) seasonal drift for these three variables ($RW + SD$). The seasonal dummies of the model are estimated in real-time, e.g. at Vintage T. Specifically, each variable we assume that

$$\triangle y_t = b_0 + b_1 D_{Q1,t} + b_2 D_{Q2,t} + b_3 D_{Q3,t} + b_4 D_{O,t} + e_t,\ e_t \sim N(0, \sigma^2),$$

where $t = 1,\ \dots,\ \bar{t}$ ($\bar{t} = T - 2$), observed at Vintage T. $\triangle y_t$ is the first difference observation, $y_t - y_{t-1}$. $D_{Q1,t}$, $D_{Q2,t}$, and $D_{Q3,t}$ are dummy variables indicating quarters 1 to 3, respectively. The dummy variable $D_{O,t}$ indicates the outlier due to the definition change, which is detailed in Appendix A. More specifically, if the observation on Vintage T is the quarter of the definition change $D_{O,t} = 1$, otherwise $D_{O,t} = 0$. Then, we forecast the variable using:

$$y_{\bar{t}+h|t} = y_{\bar{t}+h-1|t} + \hat{b}_0 + \hat{b}_1 D_{Q1,\bar{t}+h} + \hat{b}_2 D_{Q2,\bar{t}+h} + \hat{b}_3 D_{Q3,\bar{t}+h},$$

where h is the forecast horizon.

As for real-time no change forecasts of the oil price RAC, we follow Baumeister and Kilian (2015) and Garratt et al. (2019) use WTI observations as the RAC has two months delay in real time, e.g. Vintage T's last observation is $T - 2$. Therefore:

$$\begin{aligned} y_{T-1} &= y_{T-2} \times (1 + g_1), \\ y_{T+h} &= y_{t-2|t} \times (1 + g_1) \times (1 + g_2), \end{aligned}$$

where $h = 0,\ \dots,\ 6$. And

$$\begin{aligned} g_1 &= \frac{WTI_{T-1} - WTI_{T-2}}{WTI_{T-2}}, \\ g_2 &= \frac{WTI_{T} - WTI_{T-1}}{WTI_{T-1}}, \end{aligned}$$

where WTI_T is the average of daily observations of WTI spot prices on Month T. Then, the quarterly forecasts are the average of the monthly forecasts.