ONLINE APPENDIX
for
Quantifying time-varying forecast uncertainty and risk for the real price of oil

Knut Are Aastveit*
Norges Bank & BI Norwegian Business School
Jamie L. Cross
BI Norwegian Business School
Herman K. van Dijk
Erasmus University & Tinbergen Institute & Norges Bank

November 9, 2021

Abstract

This appendix provides additional details and supporting evidence for "Quantifying time-varying forecast uncertainty and risk for the real price of oil". Section A1 provides results from the PITs test of Knüppel (2015). Sections A2 and A3 provide additional BPS analysis and additional figures, respectively. Section A4 provides robustness results for using alternative oil price measures, an alternative BPS specification and alternative predictive models.

Keywords: Oil price, Forecast density combination, Bayesian forecasting, Instabilities, Model uncertainty

*This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of the Norges Bank. We gratefully acknowledge comments and suggestions from the editor Christian Hansen, the associate editor, two anonymous referees, Lennart Hoogerheide, Rob Luginbuhl, Gary Koop, Kenichiro McAlinn James Mitchell, Francesco Ravazzolo, Mike West, an anonymous referee for our Norges Bank Working Paper and seminar and conference participants at Norges Bank, University of Strathclyde, University of Rome Tor Vergata, CPB Data Science Mini-Symposium, IAAE 2021 Annual Conference, and ISBA 2021 World Meeting. This paper is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School.

A1 PITs Test

Table A1 presents tail probabilities (p-values) for the probability integral transforms (PITs) test. The tail probabilities are associated with the test in Knüppel (2015). The null hypothesis is that the PITs are uniformly distributed over the interval $(0, 1)$, and the two-sided alternative is that they are not uniformly distributed. Bold numbers indicate a rejection of the null hypothesis at the 95% credible level. Overall, we find that BPS is better calibrated than the other models.

Table A1: Tail probabilities (p-values) for the probability integral transforms (PITs) test in Knüppel (2015). The null hypothesis is that the PITs are uniformly distributed over the interval $(0, 1)$. Bold numbers indicate a rejection of the null hypothesis at the 95% credible level.

IRAC										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	**0.02**	**0.03**	**0.03**	**0.04**	0.05	0.33	**0.00**	**0.00**	**0.00**	0.10
6	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
12	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.02**	**0.03**	0.69
24	**0.04**	**0.04**	**0.04**	0.06	0.05	**0.04**	0.05	0.06	**0.04**	0.18

A2 Additional BPS Analysis

We determine the usefulness of the proposed BPS combination method in providing an accurate approximation to the observed data distribution by plotting a histogram of the data over the forecast evaluation period together with the combined forecast densities from the BPS model in Figure A1. This is useful because economic decision makers, such as central bankers or portfolio managers, take forecasts of the price of oil into account when making their policy or investment decisions. As discussed in the section 4.1 in the main text, we see substantial time-variation in the shape of the data distributions, with asymmetry, fat tails and bi-modality all being important data features. The additional insight from Figure A1 is that the proposed BPS combination method is able to provide a good approximation of the data distribution at each of the forecast horizons. To further investigate how good this approximation is from a statistical perspective, we present in Table A2 results from a two-sample Kolmogorov-Smirnov test between the empirical cumulative distribution function of the data and the cumulative distribution function of the BPS forecasts at each forecast horizon. The results show that we are unable to reject the null hypothesis that the data and forecast distributions differ at the 5% credible level. This suggests that the BPS forecast density provides an accurate proxy for the empirical data distribution, and may consequently be useful to economic decision makers in practice.

Table A2: P-values from a two-sample Kolmogorov-Smirnov test between the empirical cumulative distribution function of the data and cumulative distribution function of the BPS forecasts.

Forecast horizon	1	6	12	24
1998-2018	0.98	0.57	0.17	0.43
1998-2007	0.95	0.67	0.22	0.09
2008-2018	0.95	0.88	0.68	0.13

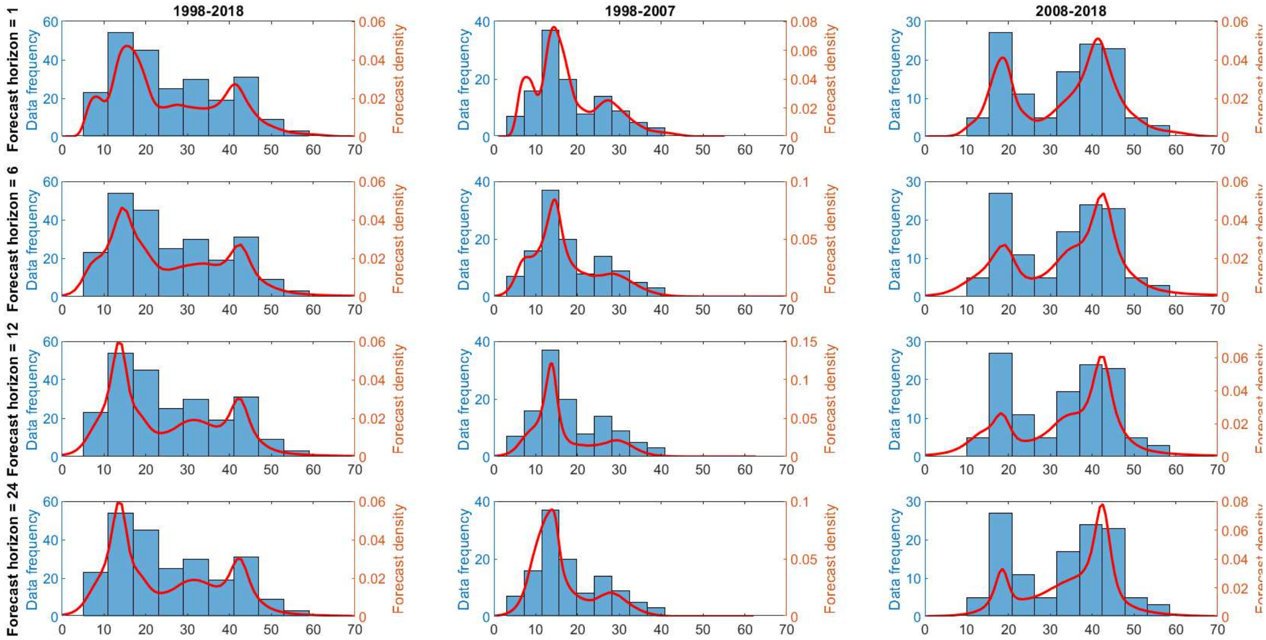

Figure A1: Forecast density combinations and data distributions of the real IRAC price of oil at monthly frequency over the forecast evaluation period: 1998:03-2017:12, and notable sub-periods. The blue histograms depict the data distributions of the real price of oil in levels over specified periods. The horizontal axis represents the level of the real price of oil in USD pooled into nine bins. The left vertical axis represents the pooled count of observations over the respective periods. The red curves depict a kernel estimate of the FDC using pooled draws from the FDC distribution across each of the periods. The right vertical axis shows the associated density values.

A3 Additional Figures

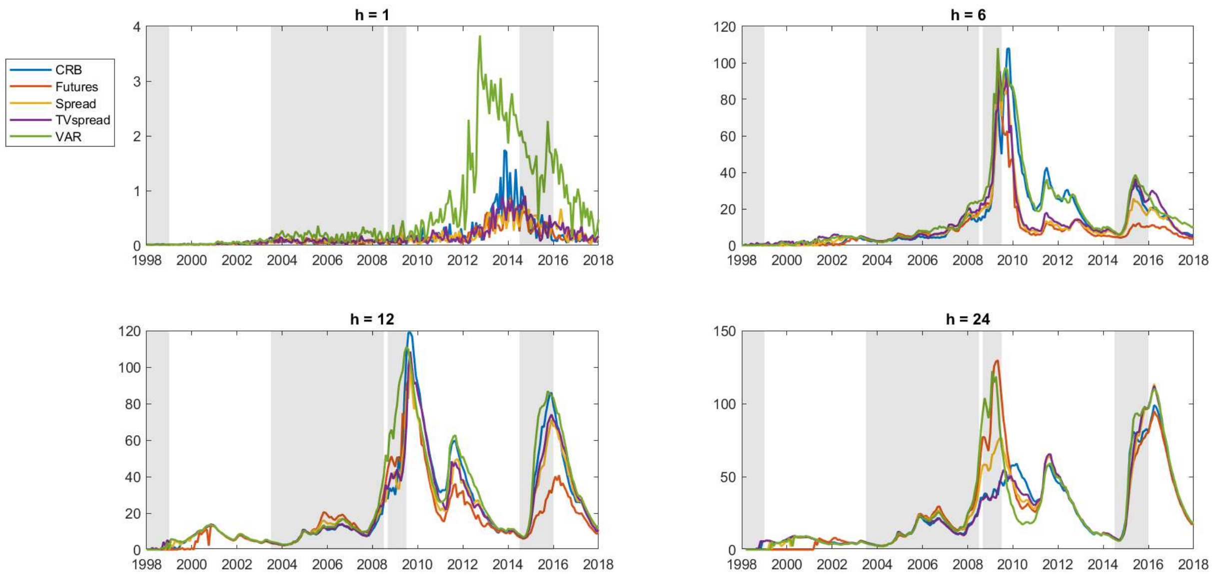

Figure A2: Time-varying variance, measured as the posterior predictive mean of the measurement variance, for individual models in the BPS model ($\sigma^2_{i,t}$), sequentially computed at each point in time over the forecast evaluation period 1998:03-2017:12.

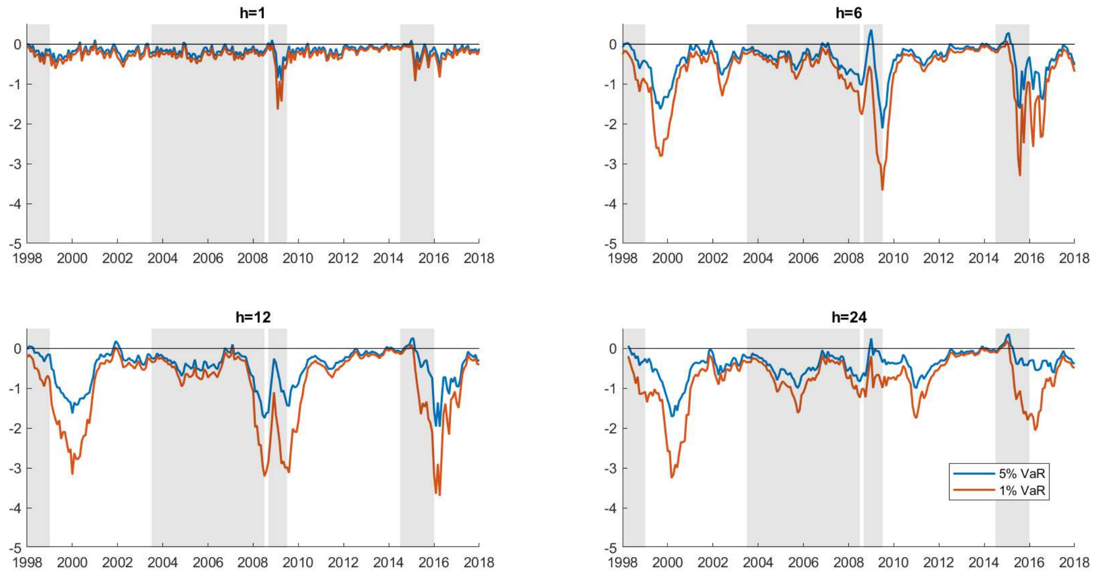

Figure A3: Value at Risk (VaR) of the profit-and-loss distributions from the BPS model, sequentially computed at each point in time over the forecast evaluation period 1998:03-2017:12.

A4 Robustness Checks

A4.1 Alternative oil price series

In our main analysis we focused on forecasting the IRAC price of crude oil, which is commonly viewed as a proxy for the global price of oil. Two alternative series that are frequently cited in the press are the Brent and West Texas Intermediate (WTI) prices of crude oil. As a robustness check, we repeated the main forecasting exercise using both of these alternative oil price series.

The series are shown together in Figure A4 over the forecast evaluation period 1998:03-2017:12. We note that their dynamics are quite similar throughout the period with minor deviations during 2011-14 due to binding storage capacity constraints in Cushing, Oklahoma that could not be eliminated by arbitrage.

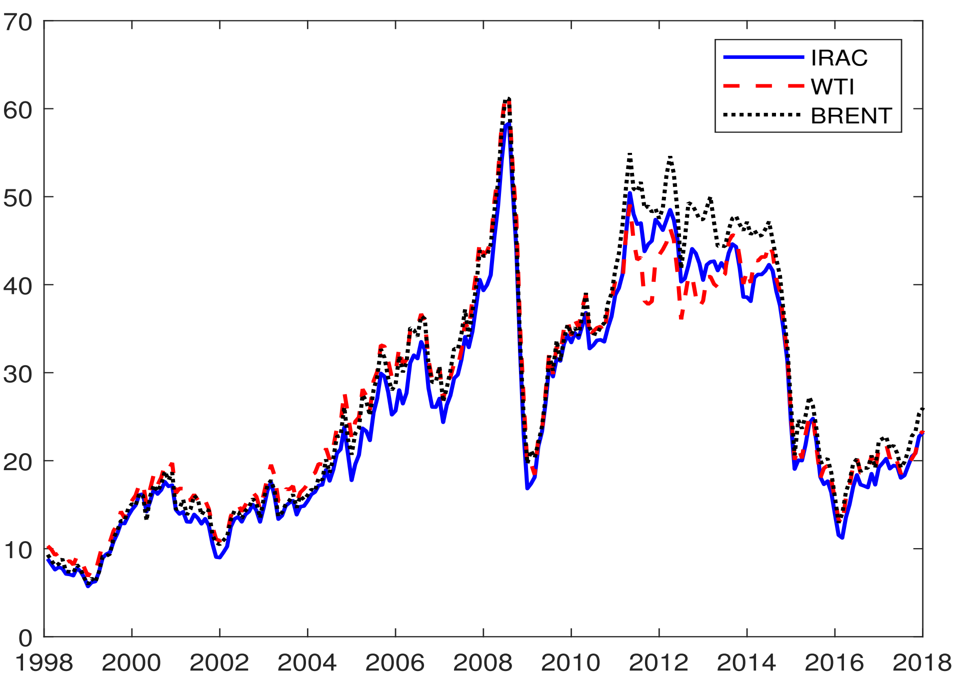

Figure A4: Real oil price series at monthly frequency over the forecast evaluation period: 1998:03-2017:12.

Forecasting results for the real WTI and Brent oil price series are shown in Table A3 and

Table A4, respectively. The results show that while some quantitative differences emerge, our qualitative conclusion that BPS provides the best forecast results at all but the one-step-ahead horizon remains robust to the choice of oil price series. We also computed the MCS (Tables A5-A6) and PITs (Table A7) for both series. Results from the MCS and PITs tests for both series and also found to be broadly consistent with those from the IRAC.

Table A3: Density (Log Score) and point (RMSFE) forecast results relative to a no-change benchmark: real WTI price of crude oil. Bold numbers indicate the best forecast performance at each horizon. One or two asterisks indicate that differences are, respectively, credibly different from zero according to the Diebold-Mariano test using 95% and 99% credible intervals.

Log Score									
Hor	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	**8.80**	-2.87	-0.26	-3.40	-21.86	-272.81**	-271.68**	-55.97**	-12.90
6	-3.67	1.05	1.32	-1.27	-27.27**	-149.57**	-146.63**	0.76*	**19.06****
12	-17.10*	9.53	0.15	-9.77	-28.89**	-124.31**	-122.29**	19.90**	**50.27** **
24	-32.77**	18.62*	-13.62**	-41.23**	-8.96	-137.08**	-123.97**	57.14**	**100.09****
RMSFE									
Hor	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0.91	0.99	1.00	1.02	1.00	0.95**	0.95**	**0.90****	0.96
6	1.08*	0.99	1.01	1.03	1.03	0.99	0.99	0.97**	**0.88****
12	1.08**	0.92**	1.00	1.01	1.04	0.97**	0.97**	0.94**	**0.72****
24	1.17**	0.91**	1.05**	1.18**	1.01	0.97*	0.97**	0.84**	**0.60****

Table A4: Density (Log Score) and point (RMSFE) forecast results relative to a no-change benchmark: real Brent price of crude oil. Bold numbers indicate the best forecast performance at each horizon. One or two asterisks indicate that differences are, respectively, credibly different from zero according to the Diebold-Mariano test using 95% and 99% credible intervals.

Log Score									
Hor	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	**9.60**	-9.42	-1.79	-4.48	-20.36	-287.40**	-291.81**	-59.30**	-12.72
6	-5.55	7.13	-9.98**	-9.95	-36.89**	-145.58**	-141.07**	10.97	**11.19***
12	-16.57	-6.33	-32.86**	-34.75**	-31.60**	-146.00**	-144.41**	-4.04	**36.99****
24	-28.30**	12.26	-24.95**	-36.92**	4.96	-154.73**	-142.82**	48.66**	**112.71****

RMSFE									
Hor	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	**0.92***	1.05	1.01	1.03	1.01	0.96**	0.95**	0.92**	0.98
6	1.08*	1.02*	1.02	1.03*	1.05	1.00	1.00	0.97**	**0.90****
12	1.04	0.96**	1.02	1.02	1.04	0.97**	0.97**	0.92**	**0.73****
24	1.15**	0.92**	1.09**	1.15**	1.02	0.99	0.98*	0.87**	**0.57****

Table A5: Model credible set (MCS) tail probabilities (p-values) for density (Log Score) and point (RMSFE) forecasts: real WTI price of crude oil. The MCS tail probabilities are computed with 100,000 block bootstrap replications using a block size of 10. Bold numbers indicate the highest ranked model at each horizon. One or two asterisks indicate that differences are in the 95% and 99% MCS, respectively.

Log Score										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0.35**	**1.00****	0.35**	0.35**	0.35**	0.00	0.00	0.00	0.00	0.00
6	0.12**	0.12**	0.46**	0.22**	0.46**	0.01*	0.00	0.00	0.46**	**1.00****
12	0.00	0.00	0.06**	0.00	0.00	0.00	0.00	0.00	0.07**	**1.00****
24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06**	**1.00****

RMSFE										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0.05**	0.55**	0.05**	0.05**	0.05**	0.05**	0.27**	0.27**	**1.00****	0.05*
6	0.22**	0.14**	0.22**	0.19**	0.14**	0.22**	0.22**	0.22**	0.22**	**1.00****
12	0.01*	0.09**	0.12**	0.01*	0.09**	0.09**	0.12**	0.12**	0.12**	**1.00****
24	0.02*	0.02*	0.02*	0.02*	0.02*	0.02*	0.02*	0.02*	0.02*	**1.00****

Table A6: Model credible set (MCS) tail probabilities (p-values) for density (Log Score) and point (RMSFE) forecasts: real Brent price of crude oil. The MCS tail probabilities are computed with 100,000 block bootstrap replications using a block size of 10. Bold numbers indicate the highest ranked model at each horizon. One or two asterisks indicate that differences are in the 95% and 99% MCS, respectively.

Log Score										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0.11	**1.00**	0.09	0.11	0.11	0.00**	0.00**	0.00**	0.00**	0.00**
6	0.44	0.44	0.95	0.15	0.44	0.01*	0.00**	0.00**	0.99	**1.00**
12	0.01*	0.01*	0.01*	0.00**	0.00**	0.00**	0.00**	0.00**	0.01*	**1.00**
24	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.01*	**1.00**

RMSFE										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0.12	0.92	0.12	0.12	0.12	0.12	0.19	0.19	**1.00**	0.12
6	0.15	0.07	0.15	0.07	0.07	0.15	0.15	0.15	0.15	**1.00**
12	0.03*	0.04*	0.04*	0.02*	0.04*	0.04*	0.04*	0.04*	0.04*	**1.00**
24	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.02*	**1.00**

Table A7: Tail probabilities (p-values) for the probability integral transforms (PITs) test in Knüppel (2015). The null hypothesis is that the PITs are uniformly distributed over the interval $(0, 1)$. Bold numbers indicate a rejection of the null hypothesis at the 95% credible level.

WTI										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	**0.00**	**0.01**	**0.01**	**0.00**	**0.00**	0.07	**0.00**	**0.00**	**0.00**	**0.03**
6	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
12	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.01**	**0.01**	0.92
24	0.06	**0.04**	0.06	0.06	0.05	**0.04**	0.05	0.05	0.07	0.10
Brent										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.08	**0.00**	**0.00**	**0.00**	**0.01**
6	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
12	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.02**	**0.00**	0.74
24	0.05	0.06	0.06	0.06	**0.04**	**0.04**	**0.04**	**0.04**	0.06	0.13

A4.2 Alternative BPS specification

In our baseline BPS specification we allow the combination weights to follow a latent process, where the BPS coefficients are able to simultaneously change over time and learn from previous performance. In Table A8 we report results from an alternative specification in which we maintain the same stochastic volatility structure as in the BPS model, however the random walk component is shut off and the combination weights and intercept are instead estimated with standard linear regression techniques. Given the recursive nature of any forecasting exercise, this enables the combination weights to change over time, however there will be significantly less flexibility in the learning process. The main insight from Table A8 is that this specification provides comparable results to the main BPS specification at the shorter horizons, however the main specification with random walk weights and intercept provide superior results at longer horizons. This result highlights the importance of allowing for both flexible combination weights and intercept at longer forecast horizons.

Table A8: Robustness using BPS specification with regression combination weights and intercept. Density (Log Score) and point (RMFSE) forecast results relative to a no-change benchmark

Hor	Log Score	RMSFE
1	-3.69	0.92
6	11.38	0.96
12	7.95	0.92
24	29.72	0.84

A4.3 Alternative Models

A4.3.1 Alternative Regression Models

In our main analysis we have expanded on the empirical results in Baumeister and Kilian (2012, 2015) by investigating whether a combination forecast using BPS can outperform their six individual models and conventional combination methods. Extensive analysis in Alquist et al. (2013) also suggests that these models tend to produce better point forecasts that simple univariate time series models such as AR and ARMA models. That being said, forecasters may not necessarily use such six models in practice, and may opt for simpler regressions with alternative predictors such as exchange rates or interest rates.

With this in mind, we estimate an additional seven predictive regressions for which the selection of variables is motivated by the tests of Granger causality in Table 8.1 of Alquist et al. (2013). Each of these specifications take the same form as in equation (12) in the main text, where we have replaced the CRB commodity price index with one of the following series (series ID from FRED are in parenthesis): (1) log of Canada / U.S. Foreign Exchange Rate (DEXCAUS), (2) log of Trade Weighted U.S. Dollar Index: Broad, Goods and Services (DTWEXBGS), (3) 3-Month Treasury Constant Maturity Rate (GS3M), (4) 10-Year Treasury Constant Maturity Rate (GS10), (5) 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity (T10Y3MM), (6) log of M1 Money Stock (M1SL), (7) A vector of zeros, in which case we have a pure inflation model.

Table A9 shows results using extra regression models (with noted data transformations). The main insight is that none of them outperform the no-change benchmark. The interest rate models are notably poor. We also ran robustness checks with different data transformations (levels and growth rates), and found none of them improved either. The results in Table A9 are the best of these transformations.

Table A9: Robustness with different regression specifications. Density (panel a) and point (panel b) forecast results relative to a no-change benchmark: real WTI price of crude oil

(a)

Hor	CAD/USD ER - log	TW ER - log	TBILL3M - level	M1 - log	INF	TBILL10M - level	TBILLSpread - level
1	-6.66	-4.60	-109.06	-3.73	-0.06	-108.53	-251.14
6	-27.87	-27.49	-157.65	3.10	-1.61	-157.82	-295.13
12	-28.55	-27.76	-212.13	-5.01	-7.93	-211.66	-311.89
24	-12.30	5.10	-256.35	-1.05	5.33	-257.31	-280.10

(b)

Hor	CAD/USD ER - log	TW ER - log	TBILL3M - level	M1 - log	INF	TBILL10M - level	TBILLSpread - level
1	1.06	1.04	2.16	1.02	1.00	2.15	2.79
6	1.07	1.04	2.73	1.02	0.99	2.74	3.19
12	1.03	1.00	3.67	1.05	0.99	3.67	3.89
24	1.09	1.02	4.78	1.14	0.99	4.78	4.05

A4.3.2 Alternative TVP Regression Models

Table A10 shows results using extra regression models with time-varying parameters via a random walk state equation. Overall, the results improve upon those without TVP in Table A9 in terms of both density and point forecasts. The TVP regression models also outperform the no-change benchmark in terms of density forecasts beyond the one-step-ahead horizon. That being said, they generally fail to outperform the benchmark in terms of point forecasts at any horizon. This result suggests that specifying time-varying parameter models is somewhat useful when forecasting the price of oil, but not as important as allowing for time-varying combination weights as in BPS.

Table A10: Robustness with different TVP regression specifications. Density (panel a) and point (panel b) forecast results relative to a no-change benchmark: real WTI price of crude oil

(a)

Hor	CAD/USD ER - log	TW ER - log	TBILL3M - level	M1 - log	INF	TBILL10M - level	TBILLSpread - level
1	-8.00	-21.74	-17.63	-11.89	-20.14	-16.90	-18.44
6	5.35	4.42	4.64	1.80	3.94	4.27	4.17
12	4.39	3.94	1.46	1.35	0.89	1.41	1.79
24	2.47	6.41	7.30	3.80	8.76	7.30	5.91

(b)

Hor	CAD/USD ER - log	TW ER - log	TBILL3M - level	M1 - log	INF	TBILL10M - level	TBILLSpread - level
1	1.07	1.12	1.18	1.12	1.20	1.18	1.18
6	1.00	1.00	1.01	1.01	1.00	1.01	1.01
12	1.00	1.00	1.00	1.00	0.99	1.00	1.00
24	1.00	1.00	1.00	1.00	1.00	1.00	1.00

A4.3.3 Alternative Combination Models

Table A11 shows forecast results for combination models that include the six models used in the main specification, plus four additional predictive regressions: CAD/USD ER, M1, INF and TBILL3M. Since the TBILL3M model is quite bad we also consider a specification where we exclude this model from the combination set. We find that including these regression models in the combination set have very little effect on the BPS forecasting performance. In contrast, the alternative combination methods generally yield worse results. This is particularly the case for density forecasts for equal weights and BMA combinations.

Table A12 shows forecast results for combination models that include the six models used in the main specification, plus the four additional predictive regressions with TVP. We find that the forecasting accuracy from BPS with individual TVP regeression models are very similar to the ones of BPS with constant coefficient individual models. For the other combinations models the point forecasts are quite similar to those in Table A11, while we find that there are some differences for the density forecasts. One particularly notable difference is that including the TVP TBILL3M model leads to a severe deterioration of the density forecasts of the equal and BMA combination models. Overall, our results indicates that it is more important to account for time-varying combination weights than individual time-varying parameters when forecasting the real price of oil.

Table A11: Robustness with different combinations of regression specifications. Density (panel a) and point (panel b) forecast results relative to a no-change benchmark: real WTI price of crude oil

	TWER, M1, Inf & TBILL3M				**TWER, M1 & Inf**			
				(a)				
Hor	Equal	BMA	BMA2	BPS	Equal	BMA	BMA2	BPS
1	-647.53	-750.36	-326.40	-22.85	-756.87	-752.60	-314.82	-20.20
6	-346.21	-419.78	-38.84	**7.58**	-452.92	-453.99	-35.56	4.83
12	-287.90	-321.44	16.38	**40.99**	-349.10	-331.26	13.27	39.13
24	-245.44	-319.30	64.22	98.36	-326.98	-322.14	64.88	**99.32**
				(b)				
Hor	Equal	BMA	BMA2	BPS	Equal	BMA	BMA2	BPS
1	0.97	0.97	**0.91**	0.98	0.98	0.98	**0.91**	0.98
6	1.00	0.99	0.96	**0.86**	0.99	0.99	0.96	0.89
12	0.91	0.93	0.88	**0.72**	0.97	0.97	0.89	0.73
24	1.13	1.00	0.79	**0.60**	1.00	0.99	0.78	0.61

Table A12: Robustness with different combinations of TVP regression specifications. Density (panel a) and point (panel b) forecast results relative to a no-change benchmark: real WTI price of crude oil

	TWER, M1, Inf & TBILL3M				**TWER, M1 & Inf**			
				(a)				
Hor	Equal	BMA	BMA2	BPS	Equal	BMA	BMA2	BPS
1	-961.56	-903.41	-178.58	-22.00	-736.25	-686.87	-155.36	-19.85
6	-717.47	-442.34	-50.68	1.11	-535.99	-306.97	-33.47	**4.20**
12	-528.23	-333.10	6.19	36.72	-426.96	-266.40	17.87	**40.58**
24	-508.88	-336.31	53.62	100,67	-410.08	-259.44	57.14	**105.50**
				(b)				
Hor	Equal	BMA	BMA2	BPS	Equal	BMA	BMA2	BPS
1	1.03	1.02	**0.90**	0.98	1.01	1.00	**0.90**	0.97
6	0.99	0.99	0.96	0.91	0.99	0.99	0.96	**0.90**
12	0.97	0.97	0.90	0.73	0.97	0.97	0.89	**0.72**
24	0.99	0.98	0.78	0,61	0.99	0.98	0.78	**0.59**

References

Alquist, R., L. Kilian, and R. J. Vigfusson (2013). Forecasting the price of oil. In *Handbook of economic forecasting*, Volume 2, pp. 427–507. Elsevier.

Baumeister, C. and L. Kilian (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics 30*(2), 326–336.

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

Knüppel, M. (2015). Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments. *Journal of Business & Economic Statistics 33*(2), 270–281.