Oil-Price Density Forecasts of U.S. GDP

Francesco Ravazzolo*

Norges Bank and BI Norwegian Business School

Philip Rothman†

East Carolina University‡

November 22, 2015

Abstract

We carry out a pseudo out-of-sample density forecasting study for U.S. GDP with an autoregressive benchmark and alternatives to the benchmark than include both oil prices and stochastic volatility. The alternatives to the benchmark produce superior density forecasts. This comparative density performance appears to be driven more by stochastic volatility than by oil prices. We use our density forecasts to compute a recession risk indicator around the Great Recession. The alternative model that includes the real price of oil generates the earliest strong signal of a recession; but it also shows increased recession risk after the Great Recession.

JEL Categories: C22, C53, E32, E37

*Contact: Norges Bank, Bankplassen 2, P.O. Box 1179 Sentrum, 0107 Oslo, Norway, Phone No: +47 22 31 61 72, e-mail: Francesco.ravazzolo@norges-bank.no

†Corresponding author: Brewster A-424, Department of Economics East, Carolina University, Greenville, NC 27858-4353, USA, Phone No: (252) 328-6151, e-mail: rothmanp@ecu.edu

‡This working 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 Norges Bank. We thank seminar and conference participants at the BI CAMP Workshop on "Forecasting and Analysing Oil Prices," the 22nd Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, and the 2nd International Workshop on "Financial Markets and Nonlinear Dynamics."

1 Introduction

In this paper we carry out a pseudo out-of-sample (OOS) density forecasting exercise to examine the predictive content of oil prices for U.S. real GDP growth. Our point of departure is the seminal paper of Hamilton (1983), who shows that large crude oil price increases systematically Granger-caused U.S. recessions from the early post-World War II period to the beginning of the 1980s. Important work in the subsequent literature on oil prices and output includes, among others, Hamilton (1996), Hooker (1996), Bernanke et al. (1997), Barsky and Kilian (2002), Hamilton (2003), Barsky and Kilian (2004), Hamilton and Herrera (2004), Kilian (2008), Edelstein and Kilian (2009), Hamilton (2009), Kilian (2009), Kilian (2010), and Baumeister and Peersman (2013). These papers primarily focus on the in-sample predictive power of oil prices for real output.

A recent branch of the literature extends the study of the oil price-GDP relationship to an OOS framework. Key papers include Bachmeier et al. (2008), Kilian and Vigfusson (2011), Alquist et al. (2013), Kilian and Vigfusson (2013), and Ravazzolo and Rothman (2013).[1] The OOS experiments in these papers evaluate point forecasts, with and without oil prices, of output growth.

Complete probability distributions over outcomes provide information helpful for making economic decisions; see, for example, Tay and Wallis (2000), Garratt et al. (2003), Gneiting (2011), and Clark (2011). In particular, density forecasts provide a characterization of forecast uncertainty. Such information about forecast uncertainty is particularly useful to central banks. For example, as part of the U.K.'s shift to an inflation-targeting regime, the Bank of England began publishing its inflation forecast as a probability distribution in the form of a fan chart in 1996; see Britton et al. (1998). Similarly, Alessi et al. (2014) explain how the Federal Reserve Bank of New York produced measures of macroeconomic risk during the Global Financial Crisis via density forecasts. Since the pioneering work of Diebold et al. (1998), there has been increasing interest in evaluating density forecasts of economic and financial data. In this paper we evaluate density forecasts, with and without oil prices, of GDP. We carry out this evaluation using statistical criteria as well as an integral-based risk measure employed by Kilian and Manganelli (2008), which they show is quite general and includes as special cases many measures of risk developed in the economic risk management literature. We use this to compute a risk of recession measure around the Great Recession.

Building upon the burgeoning literature documenting significant evidence of time-varying

[1]Kilian and Vigfusson (2011) do both in-sample and OOS analysis on this question.

volatility in macroeconomic times series of many advanced economies, Clark and Ravazzolo (2015) study the OOS forecasting implications of incorporating models of such volatility in autoregressive (AR) and vector autoregressive models. Their results favor use of a stochastic volatility (SV) models to capture time-varying volatility, especially with respect to density forecasting. Accordingly, we condition the analysis in this paper by adding an SV component to our forecasting models with and without oil prices.

Our main results are as follows. The AR benchmark without SV dominates in point forecasting. But the alternatives to the benchmark that include both oil prices and an SV component often produce superior density forecasts. Two models that generate particularly accurate density forecasts relative to this benchmark are those that include the real price of oil and the "net oil-price increase" measure of Hamilton (1996). It appears that SV plays a bigger role than oil prices in this comparative density forecast performance. The relative performance of these alternative density forecasts generally improves between the 1990-1991 recession and the Great Recession, and after the Great Recession. Further, our model which includes the real price of crude oil and SV produces, at the shorter forecast horizon, the earliest strong signal of recession risk during the Great Recession.

In Section 2 we present our forecasting models and the OOS evaluation criteria. The OOS results are presented in Section 3 and Section 4 concludes.

2 Forecasting GDP with Oil Prices

We generate and evaluate forecasts using both ex-post revised and real-time data. We use data for U.S. real GDP and the consumer price index (CPI) downloaded from the Philadelphia Federal Reserve Bank's real-time database. From past issues of the U.S. Energy Information Agency's *Petroleum Marketing Monthly* (*PMM*) available in electronic form, we constructed vintages of real-time data for the imported refiner's acquisition cost of crude oil (RAC); we use the value of the imported RAC in the third month of the quarter as the quarterly value.[2]

We generate $h-$step ahead OOS point and density forecasts, for $h = 1$ and $h = 5$, of quarterly U.S. real GDP growth rates. Our $h = 1$ forecast is a "nowcast" of the quarter $t + 1$ real GDP growth rate using real-time data vintage $t + 1$. The real-time OOS forecasts are evaluated with the actual data realization of real GDP given by the last vintage release

[2]The date of the first issue of the *PMM* available in this form is 1998M1. Issues of the *PMM* include RAC data for at most three years, so that we backcasted by approximating pre-1995M1 data with ex-post revised data; a similar approach is used by Baumeister and Kilian (2011).

available at the time our computational work was carried out, i.e., 2013Q1. For all the models we use direct forecasting for the 5−step ahead forecasts, such that we do not employ multi-equation systems to produce these forecasts.

2.1 Predictive Regressions

A standard benchmark to forecast real GDP growth at horizon h is an autoregressive model of order p (AR):

$$\Delta y_{t+h} = \alpha + \sum_{i=0}^{p-1} \beta_i \Delta y_{t-i} + \nu_{t+h}, \tag{1}$$

where $\Delta y_t = \log GDP_t - \log GDP_{t-1}$, GDP_t = real GDP for observation t, and $\nu_{t+h} \sim WN(0, \sigma^2)$. In the oil and the macroeconomy literature, the lag order p is often set equal to 4 with quarterly data; see, for example, Hamilton (2003). We follow this practice. To facilitate our density forecasts, we also assume $\nu_{t+h} \sim N(0, \sigma^2)$. Bayesian inference is applied with weak informative conjugate priors to restrict regression coefficients to zero. We use a normal-inverse-gamma prior with means for α and the β_i equal to zero and variances equal to 100. For the variance σ^2 we use an inverse-gamma with degrees of freedom equal to the number of regressors including the intercept. The predictive densities are Student−t distributed and the means (which are the same as the medians in this case) of the densities are used as point forecasts; see, for example, Koop (2003) for details.[3]

We also use an autoregressive benchmark with stochastic volatility (AR-SV):

$$\Delta y_{t+h} = \alpha + \sum_{i=0}^{p-1} \beta_i \Delta y_{t-i} + \nu_{t+h}, \tag{2}$$

where $\nu_{t+h} = \lambda_{t+h}^{0.5} \epsilon_{t+h}$, $\epsilon_{t+h} \sim N(0,1)$, $\log(\lambda_{t+h}) = \log(\lambda_{t+h-1}) + \nu_{t+h}$, $\nu_{t+h} \sim N(0, \sigma_{t+h}^2)$. Clark and Ravazzolo (2015) report that this random walk process for log volatility generates superior forecasts for U.S. GDP relative to a stationary AR(1) specification; in particular, see Table II of their paper. The random walk specification has the benefit of eliminating the need to estimate two parameters in the latent equation and allows us to avoid possibly large associated estimation errors.

In our alternatives to the benchmarks we add an oil price measure and also allow for

[3]The degrees of freedom of these Student−t distributions equals the sample size of the vintage used to produce the forecast plus the prior degrees of freedom.

random walk log volatility:

$$\Delta y_{t+h} = \alpha + \sum_{i=0}^{3} \beta_i \Delta y_{t-i} + \sum_{i=0}^{3} \delta_i oil_{t-i} + \nu_{t+h}, \tag{3}$$

where ν_{t+h} follows the specification in (2) and oil_t is the oil-price measure at time t. We refer to models given by (3) as autoregressive distributed lag models with stochastic volatility (ADL-SV). Point forecasts for the AR-SV and ADL-SV models are equal to the medians of the associated density forecasts. The oil-price measures we use are listed in Table 1. They are based on Kilian and Vigfusson (2013).

The models are estimated and forecasts are produced via a sequence of recursive windows. The first recursive window in-sample period is 1975Q1-1989Q4. For $h = 1$ and $h = 5$, the last in-sample periods are 1975Q1-2012Q3 and 1975Q1-2011Q3. The 1975Q1 initial observation is dictated by the lags we allow for and the availability of the RAC data.

A note is in order about how we estimate our models with real-time data. As explained by Clark and McCracken (2009), difficulties arise when comparing real-time OOS forecasts due to different degrees of data revision across forecast origins. One solution is to use Koenig et al.'s (2003) "strategy 1" for estimation of the predictive regressions: first-release data are used for the left-side variables; at each point in the sample, the latest available data at that date are used for right-side variables. Under this estimation approach, predictability tests developed for the case of non-revised data can be applied; see Clark and McCracken (2011). As a result, we implement Koenig et al.'s (2003) strategy 1 for estimation of our models.

2.2 Forecast Evaluation

The accuracy of point forecasts is measured with the mean squared prediction error (MSPE) metric. The density forecasts are evaluated via the average log score and average continuous ranked probability score (CRPS). The log score is considered the most comprehensive measure of density forecast accuracy. But the CRPS is thought to have advantages over the log score in that it is less sensitive to outliers and more sensitive to predictions that are close to but not equal to the outcome. Useful references on these density forecast measures include Mitchell and Hall (2005), Gneiting and Raftery (2007), Geweke and Amisano (2010), Gneiting and Ranjan (2011), and Ravazzolo and Vahey (2014).

The average log score is negative, and a higher average log score for the alternative model indicates that it performs better than the benchmark. The average CRPS is positive, and a lower average CRPS for the alternative model indicates that it performs better than the benchmark. In our tables we report MSPE, average log score, and average CRPS ratios relative to our benchmarks. A ratio less than 1 indicates superior forecast performance for the alternative to the benchmark, i.e., the model which includes oil prices. To assess whether differences in forecast accuracy are significant, we apply Diebold and Mariano (1995) $t-$tests; the associated $t-$statistics are computed with serial correlation-robust standard errors.

3 Out-of-Sample Results

We organize our discussion as follows. First we focus on MSPEs of the point forecasts. Then we analyze the log scores and CRPS values of our density forecasts. The OOS MSPE, average log score, and average CRPS results are reported in Tables 2 and 3. We conclude by using our density forecasts to compute a GDP growth rate variant of Kilian and Manganelli's (2008) risk of a negative gap measure around the most recent recession.

3.1 MSPE Comparisons

At forecast horizon $h = 1$, the point forecasts from the ADL-SV alternatives all perform worse relative to the AR(4) benchmark using both ex-post revised and real-time data. Against the AR-SV benchmark at this forecast step, the relative performance of the ADL-SV alternatives improves considerably, since the AR-SV MSPEs are a good deal higher than the AR(4) MSPEs. Using ex-post revised data the ADL-SV MSPEs are lower than the AR-SV MSPE in roughly half of the cases, and using real-time data they are lower in all but one case. None of these MSPE reductions against the AR-SV benchmark, however, are significant at conventional levels.

At forecast horizon $h = 5$, the point forecasts from the ADL-SV alternatives perform much better against the AR(4) benchmark relative to the $h = 1$ case. Using ex-post revised data the ADL-SV MSPEs are lower than the AR MSPE in roughly three-quarters of the cases, and using real-time data they are lower in all but one case. Only a few of these MSPE reductions against the AR benchmark are significant. The ADL-SV versus AR-SV MSPE results are quite similar, with the exception that the $p-$value for the equal MSPE null hypothesis is below 0.10 in five cases.

3.2 Log Score and CRPS Comparisons

At forecast horizon $h = 1$, the ADL-SV average log score and average CRPS ratios against the AR(4) benchmark are all less than 1, using both ex-post revised and real-time data.

In all but three cases, the equal average log score null hypothesis is rejected at the 10% significance level in favor of the ADL-SV models. In only one case is the $p-$value less than 0.10 for the null that the ADL-SV and AR(4) average CRPSs are equal. Against the AR-SV benchmark, the density forecast improvement obtained with the ADL-SV alternatives is considerably weaker. The ADL-SV average log score is higher than that of the AR-SV benchmark in twelve out of eighteen cases, and the ADL-SV average CRPS is higher than the AR-SV CRPS in roughly half of the cases. In only two out of eighteen cases is the ADL-SV average log score significantly higher than the AR-SV average log score, and in only three out of eighteen cases is the ADL-SV average CRPS significantly lower than the AR-SV average CRPS.

At forecast horizon $h = 5$, the average log score and average CRPS comparisons against the AR(4) benchmark are roughly the reverse of what occurs at $h = 1$. More specifically, at $h = 5$ there is much stronger evidence of density forecast improvement over the AR(4) benchmark with the ADL-SV alternatives via the average CRPS metric, while at $h = 1$ the ADL-SV density forecasts perform much better via the average log score criterion. In seventeen out of eighteen cases, the average CRPS ratios are less than 1. That said, at $h = 5$ significant ADL-SV average CRPS reductions over the AR(4) benchmark are more common with use of ex-post revised data; $p-$values for tests of the equal average CRPS null below 0.10 in seven out of nine cases versus three out of nine cases with use of real-time data. With the AR(4) model as the benchmark, in no case is the equal average log score null rejected at the 10% significance level at $h = 5$. In only three out of eighteen cases is the ADL-SV average log score significantly higher than the AR-SV average log score, and the ADL-SV average CRPS is significantly lower than the AR-SV average CRPS in only four out of eighteen cases.

To study the performance of the density forecasts across the OOS period, in Figures 1 and 2 we track the cumulative sums of the log score and CRPS for several models relative to the log score and CRPS of the AR(4) model. The cumulative sum of the relative log score at observation t is given by:

$$cusum_t^{ls} = \sum_{N+1}^{t} \log \mathrm{S}_t^a - \log \mathrm{S}_t^b, t = N+1, ..., T, \tag{4}$$

N = 1989Q4, T = 2012Q4, S = score, a = alternative model, and b = AR(4) benchmark model. Similarly, the cumulative sum of the relative CRPS at observation t is:

$$cusum_t^{crps} = \sum_{N+1}^{t} \mathrm{CRPS}_t^b - \mathrm{CRPS}_t^a, t = N+1, ..., T. \tag{5}$$

Increases in a $cusum_t$ measure indicate improvement in the alternative model's density forecast relative to the AR(4) benchmark. Likewise, if $cusum_t^{ls} > 0$ ($cusum_t^{crps} > 0$), then the average log score (average CRPS) for the alternative is higher (lower) than that of the AR(4) model when calculated over observations $N+1, ..., t$. We consider the following three alternatives to the AR(4) model in these graphs: the AR-SV, ADL-SV^{rrac}, and ADL-SV^{net+} models. These comparisons are particularly interesting since they allow us, respectively, to focus on: (i) the all else equal effect of adding a stochastic volatility component to the AR(4) model; (ii) the oil-price measure that leads to generally strong density forecasts via both the average log score and average CRPS; and (iii) the well-known oil-price measure introduced in Hamilton (1996).

Our $cusum_t$ results are presented in Figures 1 and 2, which show that the relative performance of the alternatives to the AR(4) benchmark are qualitatively similar across the $cusum_t^{ls}$ and $cusum_t^{crps}$ graphs for both ex-post revised and real-time data. At $h = 1$, the AR(4) benchmark dominates through the 1990-1991 recession. But after that downturn, the alternatives dominate and steadily improve, up to the beginning of the Great Recession. During the Great Recession, as in the 1990-1991 recession, the AR(4) model improves relative to its alternatives. The alternatives dominate after the Great Recession. This pattern shows that the $h = 1$ average log score and CRPS results in Table 2 are driven primarily by the dominant behavior of the alternatives to the AR(4) benchmark between 1991 and 2008.

The ex-post revised and real-time $cusum_t$ results at $h = 5$ differ somewhat. The behavior of the real-time $h = 5$ $cusum_t$ measures mirror pretty well what occurs at $h = 1$. In contrast, the ex-post revised $cusum_t^{ls}$ and $cusum_t^{crps}$ graphs show a worsening of the alternatives' relative performance beginning a few years before the 2001 recession. Also, the ex-post revised $h = 5$ $cusum_t^{ls}$ plots show a particularly pronounced improvement in the AR(4) model's relative density forecast performance during the Great Recession, especially against the ADL-SV^{net+} model; this improvement in the Great Recession is strong enough to push the associated average log score ratio for this case in Table 2 above one.

3.3 Risk of Recession

In their generalization of the Taylor rule, Kilian and Manganelli (2008) define the risk of a negative gap (NGR) as:

$$NGR_\gamma = -\int_{\infty}^{\underline{x}} (\underline{x} - x)^\gamma dF_x(x),\ \gamma \geq 0, \tag{6}$$

where the parameters x is the deviation of output from potential, $\underline{x}$ is the central bank's lower threshold for x, and γ is a measures of risk aversion. NGR measures the probability-weighted average loss when $x < \underline{x}$. Shifting from the output gap to the GDP growth rate, Δy_t, and setting the lower threshold for GDP growth at zero, we define the risk of recession (RR) as:

$$RR_\gamma = -\int_{\infty}^{0} (\Delta y_t)^\gamma dF_{\Delta y_t}(\Delta y_t), \tag{7}$$

which is the probability-weight average loss when output contracts, i.e., $\Delta y_t < 0$.

Via equation 7 under quadratic preferences, i.e, with $\gamma = 2$, we compute RR around the Great Recession using the AR(4), AR-SV, ADL-SV^{rrac}, and ADL-SV^{net+} density forecasts at $h = 1$ and $h =$ with both ex-post revised and real-time data. Our results are shown in Figure 3. Examination of RR under these difference density forecasts allows us to compare the extent to which these different density forecasts signaled the arrival of this extremely deep recession.

At $h = 1$ using ex-post revised data, the ADL-SV^{rrac} density forecast delivers the earliest strong signal of the recession. The eventual decline in RR under the AR(4) density forecast is larger, but it is also later, peaking near the recession's end. RR under the AR-SV and ADL-SV^{net+} forecast densities begins to decline at roughly the same point as under the ADL-SV^{rrac} forecast, but the decline is considerably smaller. Using real-time data, the relative behavior of RR under the AR(4) and ADL-SV^{rrac} density forecasts is similar during the recession. However, the ADL-SV^{rrac} forecast density sends a strong false signal of increased recession risk in 2011.

At $h = 5$, RR declines strongly during the recession only for the ADL-SV^{rrac} density forecast. But this occurs when the recession is close to ending. With both ex-post revised and real-time data, this density forecast also generates a strong false signal of increased recession risk late in the OOS period.

4 Conclusions

Motivated by the recent out-of-sample focus in the oil and the macroeconomy literature opened by Hamilton (1983), and by recent work which has provided increasing evidence the time-varying volatility in macroeconomic time series is well captured by SV modeling, we study the density forecasts of models which include both oil prices and SV. There is a sharp contrast between the OOS forecasting performance of the ADL-SV models relative to the AR benchmark across point and density forecasts. The AR benchmark dominates in point forecasting and the ADL-SV models dominate in density forecasting. The relative strength of the ADL-SV density forecasts appears to be accounted more so by SV than by oil prices.

At the shorter forecast horizon considered, the density forecast of the ADL-SV model that includes the real price of oil provides the strongest early signal of recession risk during the Great Recession. However, this density forecast also generates ex-post false signals of increased recession risk after the Great Recession.

It would be interesting to extend our analysis to a VAR-SV framework. This would allow our risk of recession measure to depend upon more than the behavior of GDP growth. Adding density forecasts of, for example, employment growth into this measure might better approximate the factors that go into the decisions of the NBER's Business Cycle Dating Committee.

References

Alessi, L., E. Ghysels, L. Onorante, R. Peach, and S. Potter (2014): "Central Bank Macroeconomic Forecasting During the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences," *Journal of Business & Economic Statistics*, 32, 483–500.

Alquist, R., L. Kilian, and R. J. Vigfusson (2013): *Graham Elliott and Allan Timmermann (eds.), Handbook of Economic Forecasting*, Amsterdam: North Holland, volume 2, chapter Forecasting the price of oil, 427–507.

Bachmeier, L., Q. Li, and D. Liu (2008): "Should oil prices receive so much attention? An evaluation of the predictive power of oil prices for the u.s. economy," *Economic Inquiry*, 46, 528–539.

Barsky, R. B. and L. Kilian (2002): *B.S. Bernanke and K. Rogoff (eds.), NBER Macroeconomics Annual 2001*, Cambridge, MA: MIT Press, chapter Do we really know that oil caused the Great Stagflation? A monetary alternative.

Barsky, R. B. and L. Kilian (2004): "Oil and the macroeconomy since the 1970s," *Journal of Economic Perspectives*, 18, 115–134.

Baumeister, C. and L. Kilian (2011): "Real-Time Forecasts of the Real Price of Oil," *Journal of Business & Economic Statistics*, 30, 326–336.

Baumeister, C. and G. Peersman (2013): "Time-Varying Effects of Oil Supply Shocks on the US Economy," *American Economic Journal: Macroeconomics*, 5, 1–28.

Bernanke, B. S., M. Gertler, and M. W. Watson (1997): "Systematic monetary policy and the effects of oil price shocks," *Brookings Papers on Economic Activity*, 91–142.

Britton, E., P. Fisher, and J. Whitley (1998): "The Inflation Report projections: understanding the fan chart," *Bank of England Quarterly Bulletin*, 38, 30–37.

Clark, T. E. (2011): "Real-time density forecasts from bayesian vector autoregressions with stochastic volatility," *Journal of Business & Economic Statistics*, 29, 327–341.

Clark, T. E. and M. W. McCracken (2009): "Tests of equal predictive ability with real-time data," *Journal of Business & Economic Statistics*, 27, 441–454.

Clark, T. E. and M. W. McCracken (2011): "Advances in forecast evaluation," Working Papers 2011-025, Federal Reserve Bank of St. Louis.

Clark, T. E. and F. Ravazzolo (2015): "Macroeconomic forecasting performance under alternative specifications of time-varying volatility," *Journal of Applied Econometrics*, 30, 551–575.

Diebold, F., A. Gunther, and K. Tay (1998): "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, 39, 863–883.

Diebold, F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–63.

Edelstein, P. and L. Kilian (2009): "How sensitive are consumer expenditures to retail energy prices?" *Journal of Monetary Economics*, 56, 766–779.

Garratt, A., K. Lee, M. H. Pesaran, and Y. Shin (2003): "Forecast uncertainties in macroeconomic modeling: An application to the UK economy," *Journal of the American Statistical Association*, 98, 829–38.

Geweke, J. and G. Amisano (2010): "Comparing and evaluating Bayesian predictive distributions of asset returns," *International Journal of Forecasting*, 26, 216–230.

Gneiting, T. (2011): "Making and evaluating point forecasts," *Journal of the American Statistical Association*, 106, 746–762.

Gneiting, T. and A. Raftery (2007): "Strictly proper score rules, prediction, and estimation," *Journal of the American Statistical Association*, 102.

Gneiting, T. and R. Ranjan (2011): "Comparing density forecasts using threshold and quantile weighted proper scoring rules," *Journal of Business & Economic Statistics*, 29.

Hamilton, J. D. (1983): "Oil and the Macroeconomy since World War II," *Journal of Political Economy*, 91, 228–248.

Hamilton, J. D. (1996): "This is what happened to the oil price-macroeconomy relationship," *Journal of Monetary Economics*, 38, 225–230.

Hamilton, J. D. (2003): "What is an oil shock?" *Journal of Econometrics*, 113, 363–398.

Hamilton, J. D. (2009): "Causes and consequences of the oil shock of 2007-08," *Brookings Papers on Economic Activity*, 215–259.

Hamilton, J. D. and A. M. Herrera (2004): "Oil shocks and aggregate macroeconomic behavior: The role of monetary policy," *Journal of Money, Credit, and Banking*, 36, 265–286.

Hooker, M. (1996): "What happened to the oil price-macroeconomy relationship?" *Journal of Monetary Economics*, 38, 195–213.

Kilian, L. (2008): "The economic effects of energy price shocks," *Journal of Economic Literature*, 46, 871–909.

Kilian, L. (2009): "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 99, 1053–1069.

Kilian, L. (2010): "Oil price shocks, monetary policy and stagflation," in R. Fry, C. Jones, and C. Kent, eds., *Inflation in an Era of Relative Price Shocks*, RBA Annual Conference Volume, Reserve Bank of Australia.

Kilian, L. and S. Manganelli (2008): "The Central Banker as a Risk Manager: Estimating the Federal Reserve's Preferences under Greenspan," *Journal of Money, Credit and Banking*, 40, 1103–1129.

Kilian, L. and R. Vigfusson (2011): "Nonlinearities in the oil price-output relationship," *Macroeconomic Dynamics*, 15, 337–363.

Kilian, L. and R. J. Vigfusson (2013): "Do oil prices help forecast U.S. real GDP? the role of nonlinearities and asymmetries," *Journal of Business & Economic Statistics*, 31, 78–93.

Koenig, E. F., S. Dolmas, and J. Piger (2003): "The use and abuse of real-time data in economic forecasting," *The Review of Economics and Statistics*, 85, 618–628.

Koop, G. (2003): *Bayesian Econometrics*, John Wiley and Sons.

Mitchell, J. and S. Hall (2005): "Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESER "fan" charts of inflation," *Oxford Bulletin of Economics and Statistics*, 67.

Ravazzolo, F. and P. Rothman (2013): "Oil and U.S. GDP: A real-time out-of-sample examination," *Journal of Money, Credit, and Banking*, 45, 449–463.

Ravazzolo, F. and S. Vahey (2014): "Forecast densities for economic aggregates from disaggregate ensembles," *Studies of Nonlinear Dynamics and Econometrics*, 18, 367–381.

Tay, A. and K. F. Wallis (2000): "Density Forecasting: A Survey," *Journal of Forecasting*, 19, 235–254.

Table 1: Definitions of Oil-Price Measures

Label	Oil-Price Measure
rrac	$oil_t = \log$ real RAC (r_t)
gr	$oil_t = \Delta r_t$
net+	$oil_t = \max(0, r_t - r^*), r^* = \max$ of r_t over preceding 3 years
net−	$oil_t = \min(0, r_t - r^{**}), r^{**} = \min$ of r_t over preceding 3 years
net	$oil_t = \max(0, r_t - r^*) + \min(0, r_t - r^{**})$
anet	$oil_t = [\max(0, r_t - r^*), \min(0, r_t - r^{**})]'$
gap	$oil_t = r_t - r^*$
large	$oil_t = \Delta r_t I(\lvert\Delta r_t\rvert > \mathrm{std}(\Delta r_t))$, std = sample standard deviation
large+	$oil_t = \Delta r_t I(\Delta r_t > \mathrm{std}(\Delta r_t))$

Notes: This table gives the definitions of the oil-price measures used in (3). See Kilian and Vigfusson (2013) for the motivation behind each; the *net+* case is the "net oil-price increase" measure introduced in Hamilton (1996). For the *anet* (short for *asymmetric net change*) oil-price measure, the δ_i parameters in (3) are 2-element vectors, allowing the coefficients on lags of net^+ and net^- to differ; with the *net* oil-price measure, these lags are constrained to be equal.

Table 2: AR Benchmark vs. ADL-SV Alternatives Out-of-Sample Forecasting Results, 1990Q1-2012Q4

	Ex-Post Revised			Real-Time		
	MSPE	Avg Log Score	Avg CRPS	MSPE	Avg Log Score	Avg CRPS
			Forecast horizon $h = 1$			
AR	0.325	-1.095	0.346	0.271	-1.082	0.325
ADL-SV^{rrac}	1.101	0.816	0.918	1.141	0.784	0.932
	(0.795)	(**0.028**)	(**0.098**)	(0.843)	(**0.008**)	(0.116)
ADL-SV^{gr}	1.205	0.841	0.959	1.272	0.881	0.981
	(0.910)	(**0.020**)	(0.283)	(0.901)	(0.200)	(0.421)
ADL-SV^{net+}	1.161	0.823	0.940	1.249	0.817	0.966
	(0.831)	(**0.022**)	(0.220)	(0.850)	(**0.040**)	(0.337)
ADL-SV^{net-}	1.167	0.888	0.944	1.259	0.794	0.967
	(0.826)	(0.215)	(0.246)	(0.868)	(**0.009**)	(0.334)
ADL-SV^{net}	1.178	0.831	0.942	1.263	0.832	0.972
	(0.821)	(**0.032**)	(0.249)	(0.863)	(**0.046**)	(0.365)
ADL-SV^{anet}	1.116	0.799	0.923	1.167	0.768	0.938
	(0.794)	(**0.007**)	(0.125)	(0.838)	(**0.002**)	(0.161)
ADL-SV^{gap}	1.297	0.866	0.986	1.324	0.815	0.985
	(0.937)	(**0.059**)	(0.463)	(0.872)	(**0.031**)	(0.460)
ADL-SV^{large}	1.478	0.884	1.033	1.403	0.790	0.998
	(0.984)	(0.118)	(0.719)	(0.912)	(**0.009**)	(0.500)
ADL-SV^{large+}	1.189	0.819	0.942	1.254	0.795	0.968
	(0.811)	(**0.023**)	(0.239)	(0.870)	(**0.009**)	(0.338)
			Forecast horizon $h = 5$			
AR	0.467	-1.173	0.388	0.376	-1.173	0.347
ADL-SV^{rrac}	0.949	0.970	0.928	0.934	0.902	0.942
	(0.152)	0.371)	(**0.021**)	(**0.049**)	(0.113)	(**0.009**)
ADL-SV^{gr}	0.981	0.997	0.945	0.974	0.964	0.967
	(0.270)	(0.383)	(**0.062**)	(0.222)	(0.288)	(0.103)
ADL-SV^{net+}	0.969	1.049	0.947	0.987	0.930	0.972
	(0.182)	(0.498)	(**0.065**)	(0.248)	(0.240)	(0.130)
ADL-SV^{net-}	0.975	1.048	0.944	0.985	0.986	0.969
	(0.225)	(0.508)	(**0.063**)	(0.240)	(0.304)	(0.113)
ADL-SV^{net}	0.964	1.053	0.942	0.980	0.911	0.964
	(0.158)	(0.532)	(**0.060**)	(0.180)	(0.164)	(**0.083**)
ADL-SV^{anet}	0.949	0.985	0.938	0.969	0.931	0.961
	(0.135)	(0.383)	**0.034**)	(0.103)	(0.162)	(**0.058**)
ADL-SV^{gap}	1.005	1.046	0.964	0.981	0.898	0.960
	(0.411)	(0.513)	(0.153)	(0.337)	(0.131)	(0.115)
ADL-SV^{large}	1.160	1.084	1.024	1.051	0.898	0.996
	(0.878)	(0.609)	(0.607)	(0.661)	(**0.077**)	(0.300)
ADL-SV^{large+}	0.960	1.026	0.942	0.991	0.992	0.971
	(0.145)	(0.472)	(**0.051**)	(0.319)	(0.312)	(0.130)

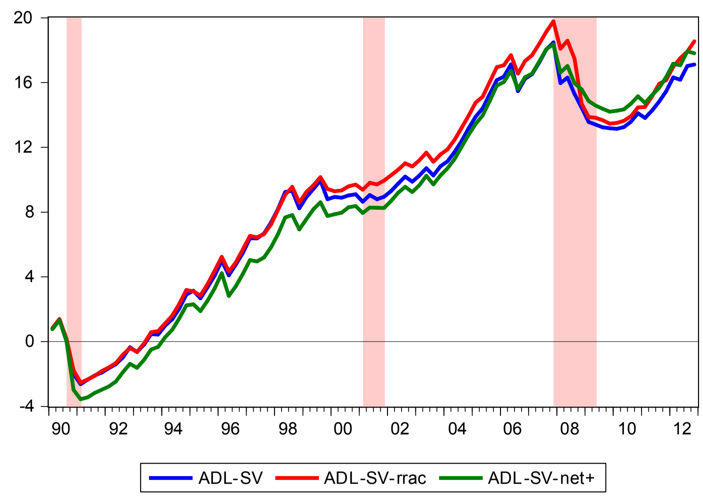
Notes: Table reports results for out-of-sample tests of equal predictability for models of US GDP growth at two forecasting horizons, $h = 1$ and $h = 5$ steps ahead. The models were estimated using recursive windows of data; the first in-sample window is 1974Q1-1989Q4. The panel labeled "Ex-Post Revised Data" reports results using the latest vintage of data for both estimation and forecasting. The panel labeled "Real-Time" reports results using vintages of real-time data via "strategy 1" of Koenig et al. (2003), with OOS forecast errors computed using the first available real-time vintages of data. For the AR(4) benchmark models, MSPEs, average log scores, and average CRPS values reported; for alternatives to the benchmark, the ratio of the alternative model's MSPE, average log score, and average CRPS relative to those of the benchmark reported. In parentheses under these ratios are reported $p-$values for the Diebold and Mariano (1995) $t-$test for equal forecast accuracy. See Table 1 for the oil-price measures associated with the ADL-SV models.

Table 3: AR-SV Benchmark vs. ADL-SV Alternatives Out-of-Sample Forecasting Results, 1990Q1-2012Q4

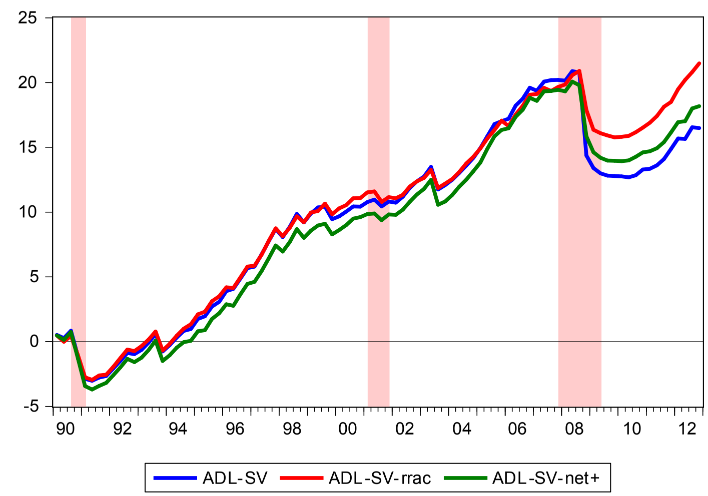
	Ex-Post Revised			Real-Time		
	MSPE	Avg Log Score	Avg CRPS	MSPE	Avg Log Score	Avg CRPS
			Forecast horizon $h = 1$			
AR-SV	0.384	-0.909	0.327	0.343	-0.902	0.314
ADL-SV^{rrac}	0.931 (0.143)	0.983 (0.294)	0.975 (**0.068**)	0.903 (0.153)	0.940 (**0.087**)	0.965 (**0.032**)
ADL-SV^{gr}	1.019 (0.657)	1.013 (0.613)	1.012 (0.687)	1.007 (0.562)	1.056 (0.871)	1.022 (0.862)
ADL-SV^{net+}	0.982 (0.148)	0.992 (0.326)	0.998 (0.326)	0.989 (0.149)	0.980 (0.235)	1.004 (0.709)
ADL-SV^{net-}	0.987 (0.128)	1.070 (0.786)	0.999 (0.405)	0.997 (0.301)	0.952 (0.139)	1.000 (0.446)
ADL-SV^{net}	0.996 (0.258)	1.002 (0.539)	0.999 (0.431)	1.000 (0.467)	0.997 (0.463)	1.002 (0.619)
ADL-SV^{anet}	0.944 (0.091)	0.963 (**0.088**)	0.976 (**0.068**)	0.924 (0.116)	0.921 (0.111)	0.971 (**0.096**)
ADL-SV^{gap}	1.097 (0.943)	1.044 (0.916)	1.045 (0.964)	1.048 (0.790)	0.977 (0.356)	1.016 (0.796)
ADL-SV^{large}	1.250 (0.983)	1.065 (0.928)	1.086 (0.974)	1.110 (0.947)	0.947 (0.178)	1.037 (0.898)
ADL-SV^{large+}	1.006 (0.605)	0.987 (0.297)	0.993 (0.122)	0.993 (0.273)	0.953 (0.158)	1.001 (0.578)
			Forecast horizon $h = 5$			
AR-SV	0.459	-1.130	0.368	0.374	-1.055	0.338
ADL-SV^{rrac}	0.966 (0.155)	1.008 (0.678)	0.983 (0.133)	0.934 (**0.075**)	0.902 (0.499)	0.943 (**0.031**)
ADL-SV^{gr}	0.998 (0.370)	1.036 (0.833)	1.000 (0.463)	0.974 (0.267)	0.964 (0.784)	0.968 (0.410)
ADL-SV^{net+}	0.985 (0.113)	1.089 (0.850)	0.997 (0.251)	0.987 (0.170)	0.930 (0.824)	0.970 (0.478)
ADL-SV^{net-}	0.992 (0.156)	1.088 (0.867)	0.997 (0.271)	0.985 (0.126)	0.986 (0.845)	0.973 (0.337)
ADL-SV^{net}	0.981 (**0.026**)	1.094 (0.846)	0.990 (**0.038**)	0.980 (**0.055**)	0.911 (0.732)	0.962 (**0.014**)
ADL-SV^{anet}	0.965 (0.165)	1.023 (0.704)	0.990 (0.147)	0.969 (**0.073**)	0.931 (0.728)	0.962 (0.106)
ADL-SV^{gap}	1.022 (0.656)	1.087 (0.864)	1.016 (0.792)	0.981 (0.396)	0.898 (0.480)	0.961 (0.317)
ADL-SV^{large}	1.180 (0.940)	1.126 (0.957)	1.079 (0.973)	1.051 (0.678)	0.898 (0.467)	0.999 (0.816)
ADL-SV^{large+}	0.976 (**0.047**)	1.065 (0.814)	0.992 (**0.047**)	0.991 (0.387)	0.992 (0.857)	0.967 (0.197)

Notes: Notes: Table reports results for out-of-sample tests of equal predictability for models of US GDP growth at two forecasting horizons, $h = 1$ and $h = 5$ steps ahead. The models were estimated using recursive windows of data; the first in-sample window is 1974Q1-1989Q4. The panel labeled "Ex-Post Revised Data" reports results using the latest vintage of data for both estimation and forecasting. The panel labeled "Real-Time" reports results using vintages of real-time data via "strategy 1" of Koenig et al. (2003), with OOS forecast errors computed using the first available real-time vintages of data. For the AR(4)-SV benchmark models, MSPEs, average log scores, and average CRPS values reported; for alternatives to the benchmark, the ratio of the alternative model's MSPE, average log score, and average CRPS relative to those of the benchmark reported. In parentheses under these ratios are reported $p-$values for the Diebold and Mariano (1995) $t-$test for equal forecast accuracy. See Table 1 for the oil-price measures associated with the ADL-SV models.

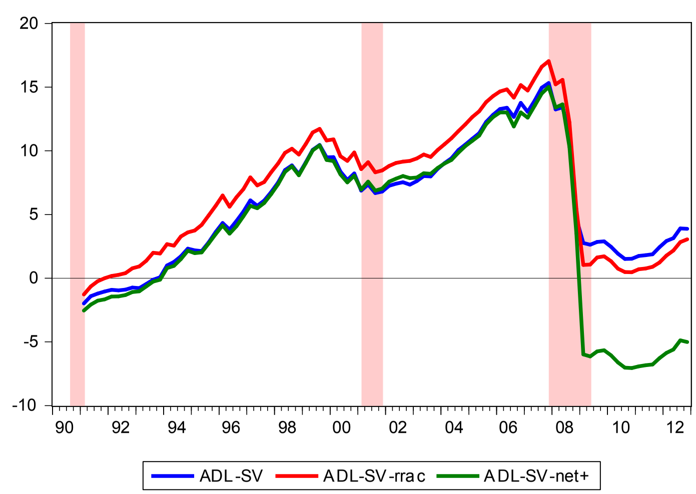
Figure 1: Cumulative Sums of Relative Log Score



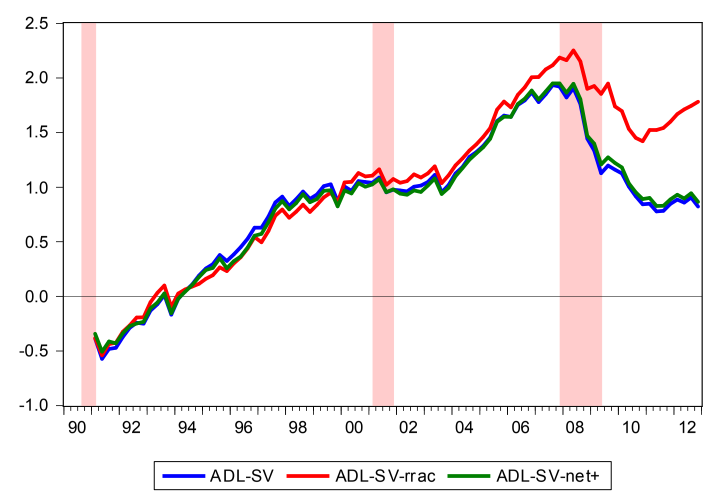
(a) Ex-Post Revised: Forecast horizon $h = 1$



(b) Real-Time: Forecast horizon $h = 1$



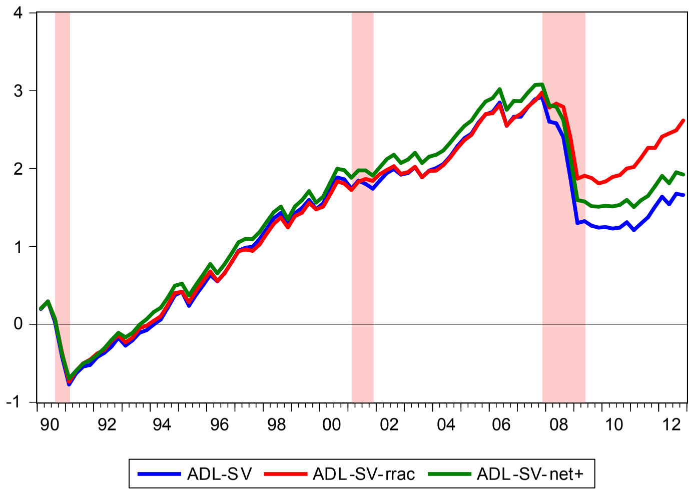
(c) Ex-Post Revised: Forecast horizon $h = 5$



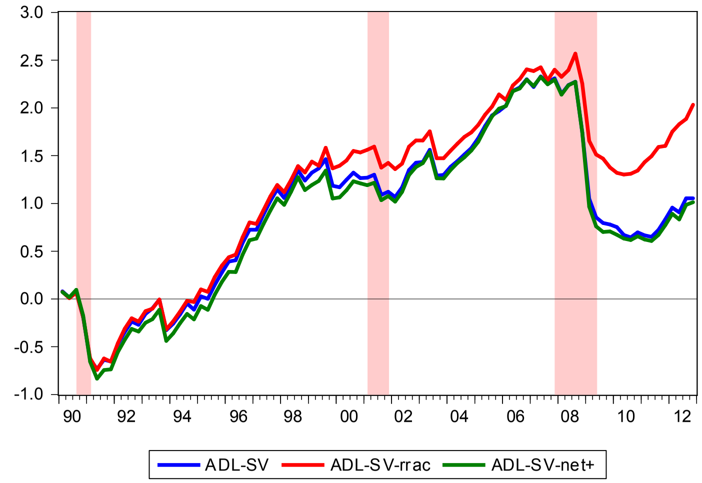
(d) Real-Time: Forecast horizon $h = 5$

Notes: $cusum_t^{ls} = \sum_{N+1}^{t} \log \mathrm{S}_t^a - \log \mathrm{S}_t^b, t = N+1, ..., T$, N = 1989Q4, T = 2012Q4, S = score, a = alternative model, b = AR benchmark model.

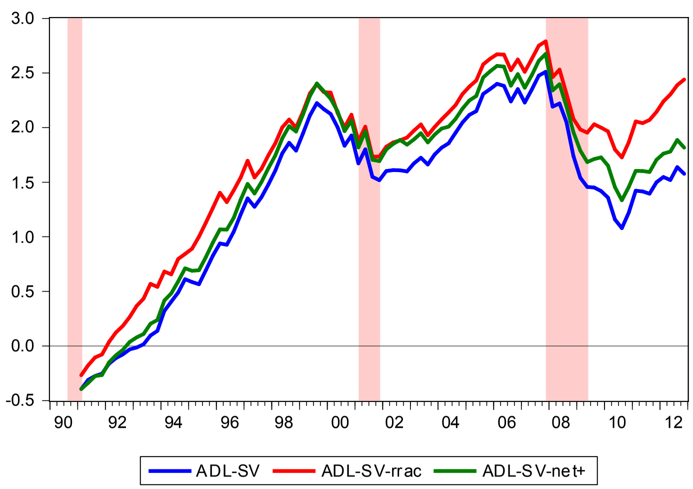
Figure 2: Cumulative Sums of Relative CRPS



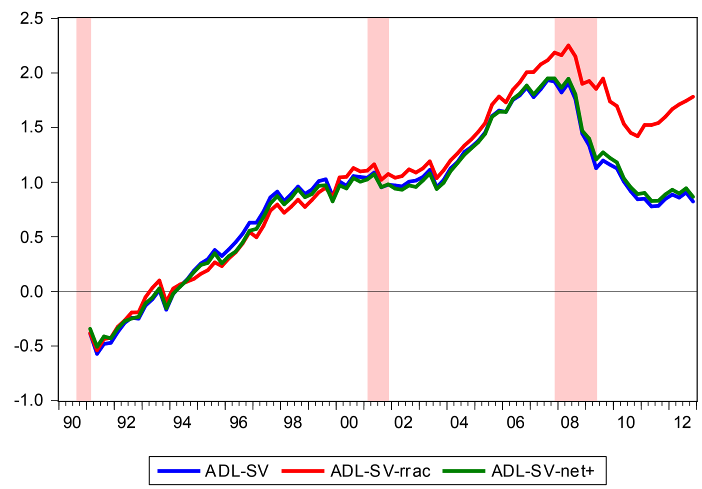
(a) Ex-Post Revised: Forecast horizon $h = 1$



(b) Real-Time: Forecast horizon $h = 1$



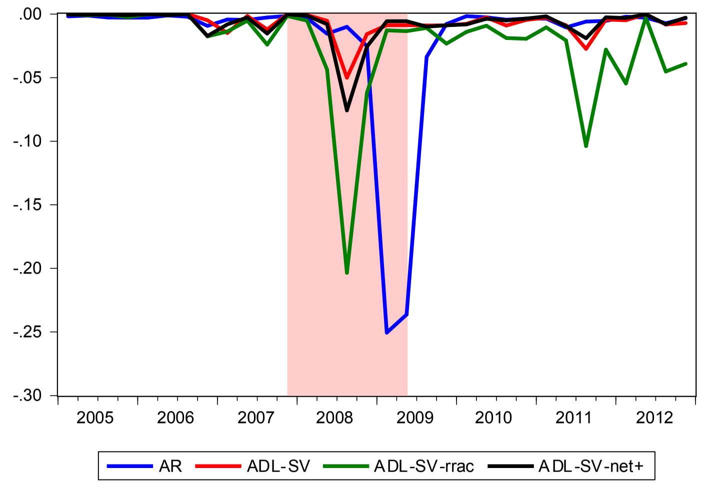
(c) Ex-Post Revised: Forecast horizon $h = 5$



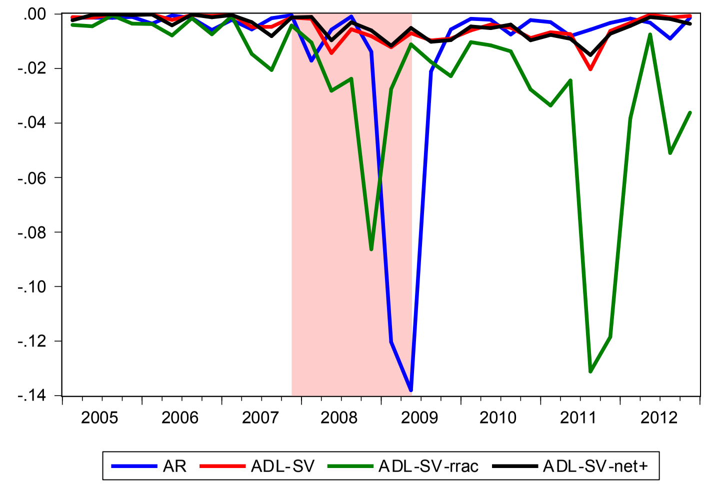
(d) Real-Time: Forecast horizon $h = 5$

Notes: $cusum_t^{crps} = \sum_{N+1}^{t} \text{CRPS}_t^b - \text{CRPS}_t^a, t = N+1, ..., T,\ N = 1989\text{Q4},\ T = 2012\text{Q4},\ a =$ alternative model, $b =$ AR benchmark model.

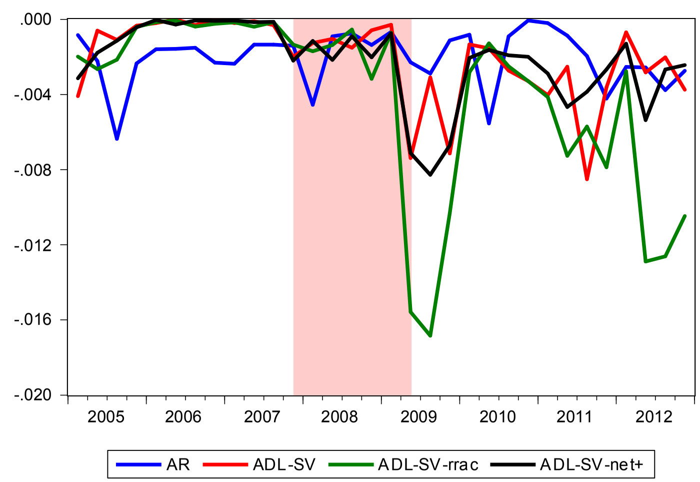
Figure 3: Risk of Recession



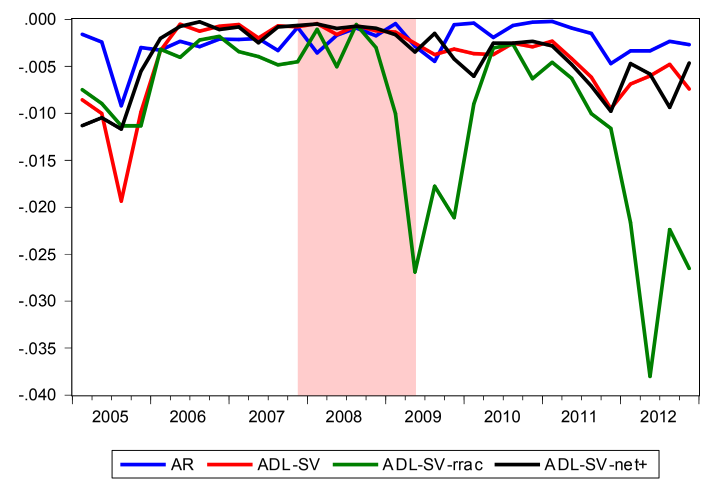
(a) Ex-Post Revised: $h = 1$



(b) Real-Time: $h = 1$



(c) Ex-Post Revised: $h = 5$



(d) Real-Time: $h = 5$

Notes: The risk of recession from 7 is computed around the Great Recession using the AR(4), AR-SV, ADL-SV^{rrac}, and ADL-SV^{net+} forecast densities.