Winter is coming: What do the futures hold?

ECON207 G1

Bharat Gangwani

19 November 2022

Introduction

Natural gas supports household/commercial heating, electricity generation and applications in industrial production. Consequently, securing stable supply for natural gas is of significant interest to many industry and government stakeholders. Its importance has been underscored this year as Russia largely shut its natural gas supply to European countries in retaliation against the sanctions imposed for its invasion of Ukraine. Natural gas consumption in Europe has declined by more than 10% so far this year compared to last year while prices in the United States have climbed to their highest summertime peak since 2008 (IEA 2022). While Europe has been able to secure its natural gas reserves for the coming winter, significant challenges may arise next year when these reserves are depleted and require replenishment (England and Wilson 2022). It's this necessity to safeguard and protect future supply which motivates this paper's central question: **Does a decrease in forecasted temperatures in the United States increase the price of natural gas futures?**

Natural gas futures are contracts which obligate the buyer to purchase a given quantity of natural gas at a predetermined price sometime in the future. Both consumers and producers of natural gas can use these contracts to hedge their risk by safeguarding themselves from significant price movements. As a result, obtaining a reliable estimate for the responsiveness of the price of natural gas futures to changes in temperature and temperature forecasts can allow companies to be able to better hedge their risk within volatile global commodity markets.

In this project, I specifically utilize the Henry Hub Natural Gas futures due to their high liquidity and standardized terms of contract (CME Group 2022). Additional controls incorporated will include natural gas imports and amount held in storage. One would expect the amount of contemporary natural gas imports and storage in the country to be positively and negatively correlated with the price of natural gas futures respectively. Nick and Thoenes (2014) make a more rigorous case for the inclusion of natural gas imports and amount held in storage within their model. Their paper informs my variable selection as well. Historical price, storage and imports data is obtained from EIA (2022). The average monthly temperature data is obtained from the GSOM, University of Minnesota St. Paul Station dataset (Lawrimore et al. 2016). Within my analysis, I utilize leading temperature data as a proxy for temperature forecasts making the assumption that temperature forecasts would be close to the actual temperatures. This is due to the unavailability of historical data for temperature forecasts.

Since the futures price data is time series, I expected and experienced a significant amount of noise in the dataset along with serial correlation between the error terms. There is a genuine concern of inconsistent coefficients in cases where I include AR terms and yet experience serial correlation. Consequently, I utilize ARDL and difference models to both exploit serial correlation to increase the model's predictive power and minimize the likelihood of inconsistent coefficient estimates. In the end, I create a dynamically complete model incorporating a linear deterministic trend, seasonal variation and appropriate autocorrelation terms to explain the changes in the price of natural gas futures. The final model validates part of my initial hypothesis. While a contemporaneous decrease in average monthly temperature increases prices of natural gas futures, temperature forecasts do not seem to have a significant effect on them.

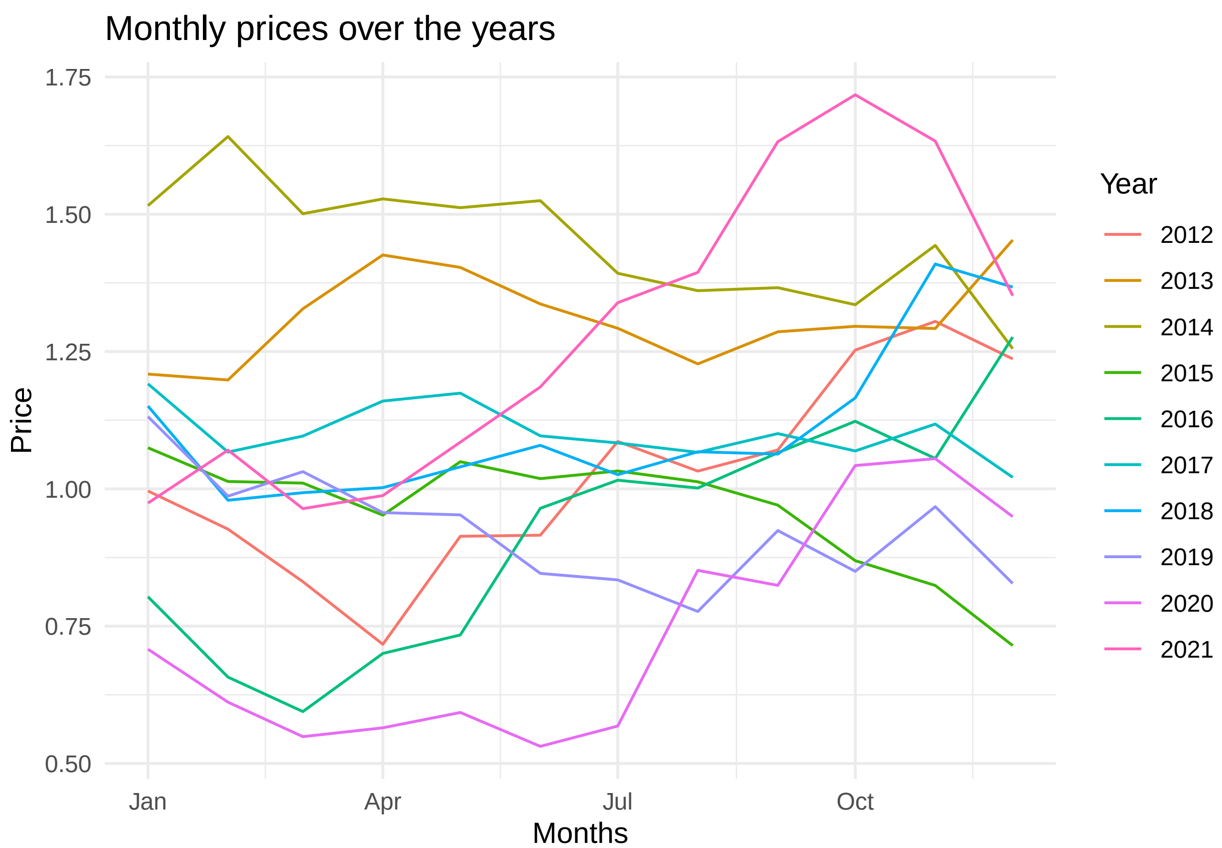
Data and Methodology

The dataset I use throughout the analysis contains 5 predictors and 1 predicted variable. The dataset has 388 observations at the country level and the unit of time is a month. In the case of Price, daily data was grouped by and averaged on the year-month level. Storage was also available at a weekly frequency; it was upscaled to the month level as well. Other series were directly available at the monthly frequency. A short description of the variable and its source is provided in the table below:

Variable	Description	Min	Median	Max	Source
Price	Log price of the natural gas future at month t	0.15	1.09	2.6	EIA
t	A sequence of integers [1, 388]	1	194.5	388	na
Month	A set of dummy variables for the 12 months	na	na	na	na
TAVG	The average monthly temperature recorded at the University of Minnesota	-15.40	9.00	25.62	NCEI
Imports	Volume of natural gas imported by the United States	117,009	258,011	267,573	EIA
Storage	Volume of natural gas held in storage in the United States	5,041,971	6,776,630	6,737,241	EIA

The price time series seems to possess weak seasonality and an additive trend, as suggested by the changes in its average price level over the years in Figure 1. This justifies the addition of a trend regressor and a vector of seasonality (month) dummies to our regression specification.

```
ggplot(dfm_plot, aes(x = xaxis, y = Price)) +
  geom_line(aes(color = Year )) +
  xlab("Months") +
  ylab("Price") +
  scale_x_date(labels = date_format("%b")) +
  ggtitle("Monthly prices over the years")
```



Consequently, the first model I propose is the following:

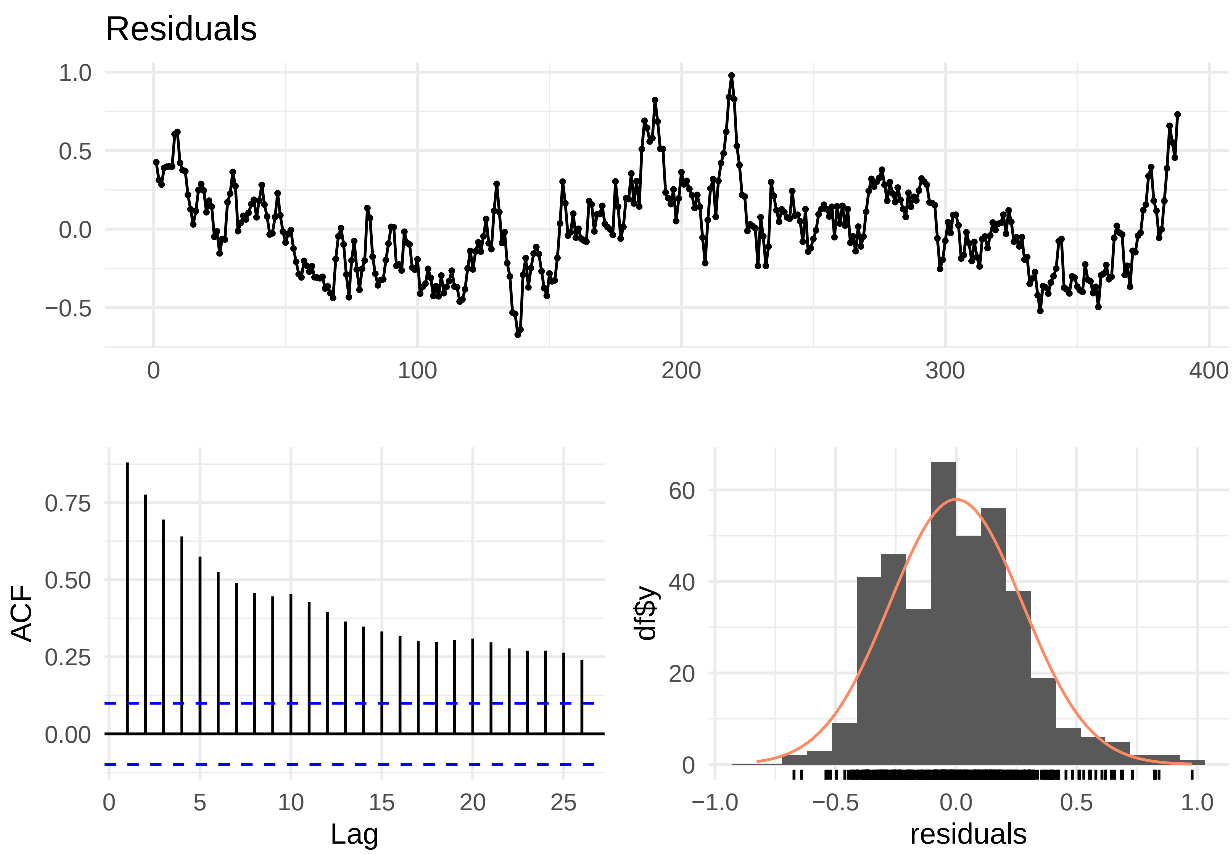
$$Price_t = \beta_0 + \beta_1 TAVG_t + \beta_2 Imports_t + \beta_3 Storage_t + \alpha_0 t + \sum_{i=2}^{12} \alpha_i Month_i + \epsilon_t$$

where $\alpha_i Month_i$ are the set of 11 dummy vectors for the months excluding January.

```
# Linear Regression
lr2 = lm(Price ~ . - Date,df)
summary(lr2, vcov = NeweyWest)$r.squared
```

```
## [1] 0.7135817
```

```
# Correlogram for residuals
checkresiduals(lr2, lag = 12)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 12
##
## data:  Residuals
## LM test = 315.26, df = 12, p-value < 2.2e-16
```

However, I observe that the while the model has a $R^2 = 71.36\%$, it also has significant serial correlation among its noise terms. This is confirmed by the Breusch-Godfrey test for lag 12 as well. The test statistic is 46.785 and is hence quite significant.

Consequently, I would need lagged terms in the specification to address the serial correlation. After several iterations trying to minimize serial correlation without introducing significant multicollinearity and increasing the variance of the estimators, I find AR(1) with 6 leading $TAVG$ terms to serve as proxies for temperature forecasts as the most appropriate model. Its population regression function is as follows:

$$Price_t = \beta_0 + \sum_{i=0}^{6} \beta_{1+i} TAVG_{t+i} + \beta_2 Imports_t + \beta_3 Storage_t + \beta_4 Price_{t-1} + \alpha_0 t + \sum_{i=2}^{12} \alpha_i Month_i + \epsilon_t$$

The significance of the coefficients of the leading terms would determine the validity of my initial hypothesis. Additionally, the joint F-test restricting the coefficients of the $TAVG_t$ terms to be 0 will determine whether their addition to the model adds significant explanatory power. These are discussed further in the next section.

Results

The final output of the model specified above is estimated below. We can observe that all the seasonality dummies are significant at least at the 10% level. All other variables are statistically significant at the 5% level as well except some of the $TAVG$ leading terms. The seasonality dummies suggest that futures prices tend higher relative to January throughout the rest of the year and peak around June. This may be a result of producers and consumers locking in trades at the given price levels for the coming winter. As expected, the price of natural gas futures is affected positively by the quantity of imports and negatively by the amount held in storage. Prices are also significantly persistent, as suggested by the the coefficient of LPrice1 taking the highest numerical value in the estimation. Finally, the model output suggests that a contemporaneous drop in temperature increases price of natural gas futures today by 0.935%. However, a unit increase in the forecasted temperature for 4 months later appears to increase the price of gas futures while a unit increase in the forecasted temperature for 3 months later appears to conversely decrease it. Since the significance and magnitude of these coefficients is almost similar, they appear to cancel out. These effects of changes in temperature are independent of seasonality.

```r
dfardl = df

createLag = function(dfc, n){
  c(rep(NA, n), dfc[1:(length(dfc)-n)])
}

createLead = function(dfc, n){
  c(dfc[(1+n):length(dfc)], rep(NA, n))
}
# ARDL(p, q)
p = 1
q = 6

for (i in 1:p){
  dfardl[, sprintf("LPrice%s", i)] = createLag(df$Price, i)
}

for (i in 1:q){
  dfardl[, sprintf("LTAVG%s", i)] = createLead(df$TAVG, i)
}

ardl1 = lm(Price ~ . - Date, dfardl)
sumreg = summary(ardl1, vcov = NeweyWest)
coef(sumreg)
```

```
##                     Estimate   Std. Error    t value      Pr(>|t|)
## (Intercept)    -5.240501e-02 1.590935e-01 -0.3293975  7.420479e-01
## MonthApril      3.806328e-01 8.836433e-02  4.3075385  2.135015e-05
## MonthMay        4.804872e-01 1.086189e-01  4.4236057  1.289673e-05
## MonthJune       5.631179e-01 1.245182e-01  4.5223738  8.326540e-06
## MonthJuly       5.274980e-01 1.316271e-01  4.0075188  7.467498e-05
## MonthAugust     5.351861e-01 1.277306e-01  4.1899600  3.518220e-05
## MonthSeptember  5.038227e-01 1.126674e-01  4.4717689  1.042910e-05
## MonthOctober    4.425178e-01 9.064305e-02  4.8819827  1.585086e-06
## MonthNovember   2.737414e-01 6.593857e-02  4.1514606  4.133186e-05
## MonthDecember   1.040396e-01 4.099814e-02  2.5376671  1.158287e-02
## MonthFebruary   8.234036e-02 4.117820e-02  1.9996104  4.629761e-02
## MonthMarch      2.089439e-01 6.488904e-02  3.2200186  1.399103e-03
## TAVG           -9.347420e-03 2.522421e-03 -3.7057337  2.441093e-04
## Imports         8.032209e-07 1.515807e-07  5.2989658  2.040060e-07
## Storage        -4.323788e-08 2.171145e-08 -1.9914785  4.718780e-02
## t               1.767025e-04 9.052551e-05  1.9519637  5.172181e-02
## LPrice1         8.871399e-01 1.972396e-02 44.9777855 2.317154e-149
## LTAVG1         -1.613120e-03 2.458839e-03 -0.6560493  5.122138e-01
```

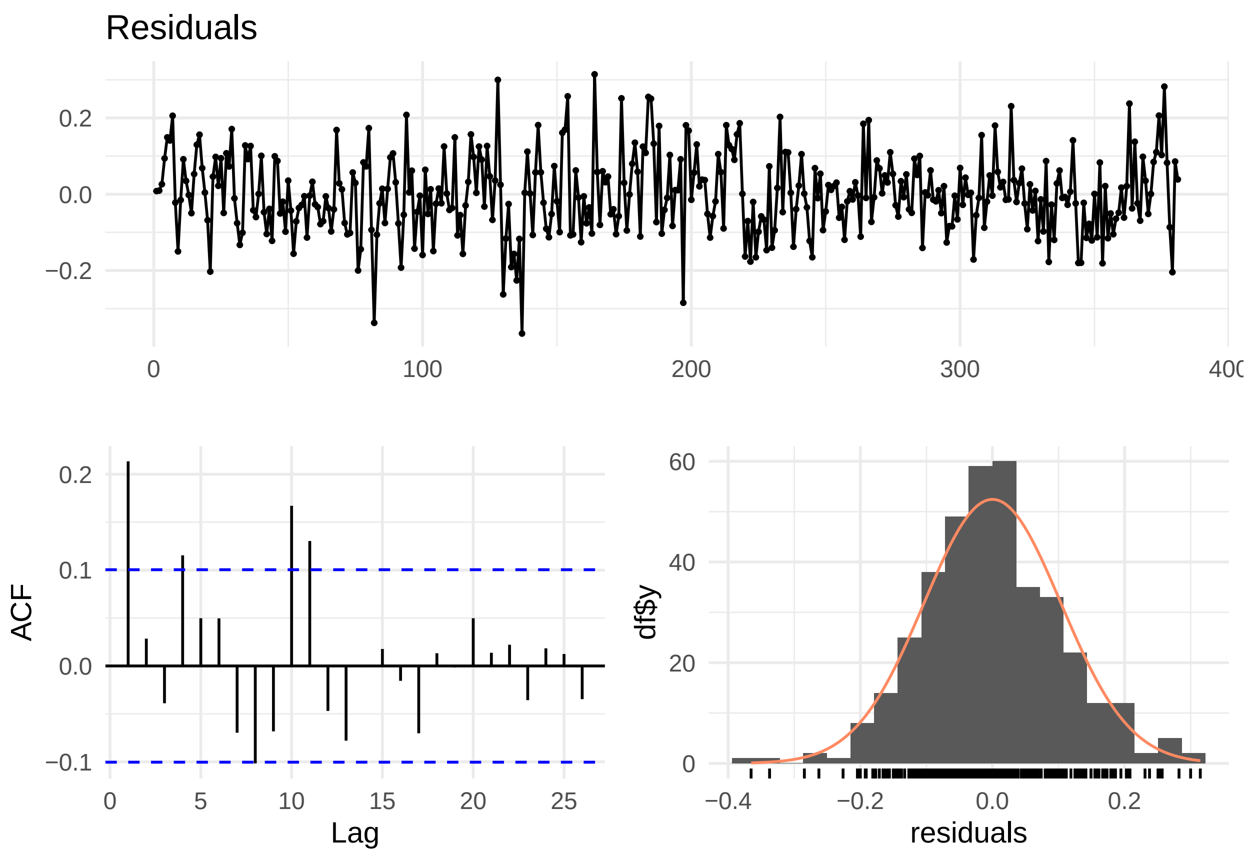
```
## LTAVG2          -3.079655e-03 2.462730e-03 -1.2505042  2.119321e-01
## LTAVG3          -4.733272e-03 2.477545e-03 -1.9104685  5.687156e-02
## LTAVG4           5.241082e-03 2.470822e-03  2.1211897  3.459262e-02
## LTAVG5           1.256890e-03 2.402926e-03  0.5230666  6.012515e-01
## LTAVG6          -2.057626e-03 2.345069e-03 -0.8774266  3.808434e-01
```

```
sumreg$r.squared
```

```
## [1] 0.9584459
```

Due to the high persistence in the data, the $R^2 = 95.84\%$ is also quite high.

```
# Correlogram for residuals
checkresiduals(ardl1, lag = 12)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 12
##
## data:  Residuals
## LM test = 56.566, df = 12, p-value = 9.467e-08
```

Finally, the inclusion of the AR(1) term appears to have alleviated the serial correlation slightly. The residuals appear more normally distributed around 0 while their correlations are more erratic and hence harder to formalize into a model. Consequently, following the principle of parsimony and due to a lack of further evident patterns to exploit in the serial correlations, I conclude with this model despite Breusch-Godfrey test still suggesting significant serial correlation.

Utilizing a joint F-test, we test the restrictive hypothesis $(TAVG_{t+i} = 0) \forall i \in [0, 6]$ The result of the joint F-test suggests that there is evidence at the 1% level to reject the hypothesis that the coefficients of these lagged terms are restricted to 0. Hence, these coefficients retain significance within the model and improve its explanatory power.

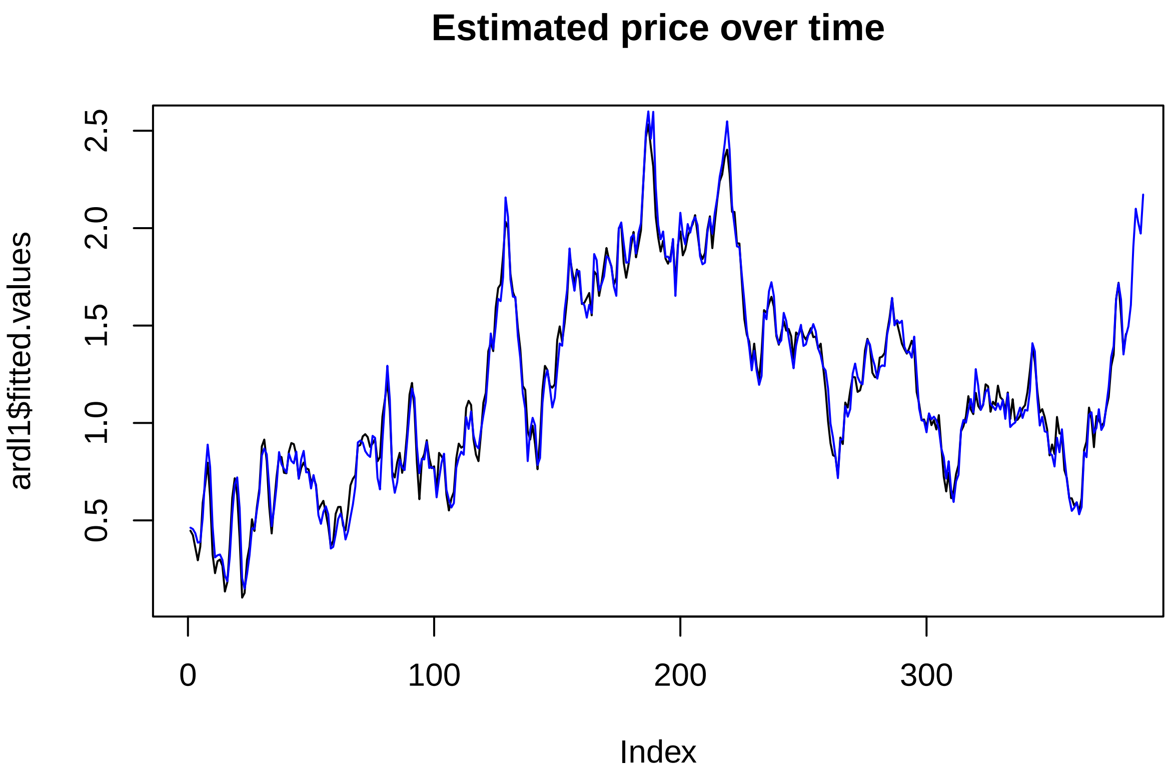
```r
restrictions = c("TAVG = 0", "LTAVG1 = 0", "LTAVG2 = 0", "LTAVG3 = 0",
                 "LTAVG4 = 0", "LTAVG5 = 0", "LTAVG6 = 0")
linearHypothesis(ardl1, restrictions)
```

```
## Linear hypothesis test
##
## Hypothesis:
## TAVG = 0
## LTAVG1 = 0
## LTAVG2 = 0
## LTAVG3 = 0
## LTAVG4 = 0
## LTAVG5 = 0
## LTAVG6 = 0
##
## Model 1: restricted model
## Model 2: Price ~ (Date + Month + TAVG + Imports + Storage + t + LPrice1 +
##     LTAVG1 + LTAVG2 + LTAVG3 + LTAVG4 + LTAVG5 + LTAVG6) - Date
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    365 4.4249
## 2    358 4.0885  7   0.33636 4.2075 0.0001784 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion

In this project, I develop a dynamically complete model to predict the price of natural gas futures. Using the model, I identify the impact that changes in the average temperature and average temperature forecasts in the United States have on the price of natural gas futures. I propose AR(1) with 6 leading $TAVG$ terms as the final model due to its significant R^2, inclusion of sufficient leading temperature terms as proxies for temperature forecasts, a high number of significant coefficients and minimized serial correlation among the error terms. The model tracks the original Price series well. However, that is primarily a consequence of strong persistence within the series as suggested by the high coefficient estimate of the AR(1) term.

```r
# Fitted values plot
plot(ardl1$fitted.values, type = "l", main = "Estimated price over time")
lines(dfardl$Price, type = "l", col = "blue")
```



Despite the persistence, however, 3 out of the 6 $TAVG$ terms are statistically significant at the 10% level. The joint F-test also delivers sufficient evidence to suggest that the coefficients of all 6 terms are not equal to 0 and hence they improve the explanatory power of the model. As a result, I confidently conclude in response to the original question that while an increase in contemporaneous temperature decreases the prices of natural gas futures, temperature forecasts appear to have little effect on them.

References

CME Group. 2022. "Henry Hub Natural Gas Overview - CME Group." CME Group. November 18, 2022. https://www.cmegroup.com/markets/energy/natural-gas/natural-gas.html.

EIA. 2022. "Natural Gas Data - U.S. Energy Information Administration (EIA)." U.S. Energy Information Administration. October 31, 2022. https://www.eia.gov/naturalgas/data.php.

England, Andrew, and Tom Wilson. 2022. "Europe at Risk of 'Much Worse' Energy Crisis Next Year, Warns Qatar." *Financial Times*, October 18, 2022. https://www.ft.com/content/96611190-3c09-48b7-8ded-7b2651590f94.

IEA. 2022. "Natural Gas Markets Expected to Remain Tight into 2023 as Russia Further Reduces Supplies to Europe - News." International Energy Agency. October 3, 2022. https://www.iea.org/news/natural-gas-markets-expected-to-remain-tight-into-2023-as-russia-further-reduces-supplies-to-europe.

Lawrimore, Jay, Ron Ray, Scott Applequist, Bryant Korzeniewski, and Matthew J. Menne. 2016. "Global Summary of the Month (GSOM), Version 1." NOAA National Centers for Environmental Information. https://doi.org/10.7289/V5QV3JJ5.

Nick, Sebastian, and Stefan Thoenes. 2014. "What Drives Natural Gas Prices? — A Structural VAR Approach." *Energy Economics* 45 (September): 517–27. https://doi.org/10.1016/j.eneco.2014.08.010.