# Test Exercise 3: Answers to the Questions

(a) Use *general-to-specific* to come to a model. Start by regressing the *federal funds rate* on the other 7 variables and eliminate 1 variable at a time.

- Starting with the all the 7 explanatory variables and dropping the variables 1 at a time (w.r.t. the lowest **p-value**) we get the following:

| | (Intercept) | INFL | PROD | UNEMPL | COMMPRI | PCE | PERSINC | HOUST | |
|---|---|---|---|---|---|---|---|---|---|
| *coeff* | -0.2147 | 0.6968 | -0.0576 | 0.1041 | -0.0055 | 0.3426 | 0.2464 | -0.0194 | |
| *p.val* | 0.3808 | 0 | 0.1489 | $\boxed{0.2816}$ | 0.0636 | 0 | 1e-04 | 0 | **Drop UNEMPL** |
| *R^2* | 0.639 | | | | | | | | |

| | (Intercept) | INFL | PROD | COMMPRI | PCE | PERSINC | HOUST | |
|---|---|---|---|---|---|---|---|---|
| *coeff* | -0.2856 | 0.6942 | -0.0248 | -0.0065 | 0.3671 | 0.2512 | -0.021 | |
| *p.val* | 0.2263 | 0 | $\boxed{0.3357}$ | 0.0209 | 0 | 0 | 0 | **Drop PROD** |
| *R^2* | 0.638 | | | | | | | |

| | (Intercept) | INFL | COMMPRI | PCE | PERSINC | HOUST | |
|---|---|---|---|---|---|---|---|
| *coeff* | -0.2363 | 0.7177 | -0.0075 | 0.3398 | 0.2402 | -0.0205 | |
| *p.val* | 0.3051 | 0 | 0.0046 | 0 | 1e-04 | 0 | **All variables are significant at 5% level** |
| *R^2* | 0.637 | | | | | | |

```
##
## Call:
## lm(formula = INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST,
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1918 -1.5298 -0.0974  1.3829  7.6603
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.236287   0.230212  -1.026  0.30509
## INFL         0.717720   0.056972  12.598  < 2e-16 ***
## COMMPRI     -0.007499   0.002639  -2.842  0.00463 **
## PCE          0.339822   0.058989   5.761 1.29e-08 ***
## PERSINC      0.240157   0.059265   4.052 5.68e-05 ***
## HOUST       -0.020519   0.004382  -4.682 3.45e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.188 on 654 degrees of freedom
## Multiple R-squared:  0.6374, Adjusted R-squared:  0.6347
## F-statistic:   230 on 5 and 654 DF,  p-value: < 2.2e-16
```

**Final Model**

**Stepwise Backward Regression using AIC**

```
## Start:  AIC=1041.17
## INTRATE ~ INFL + PROD + UNEMPL + COMMPRI + PCE + PERSINC + HOUST
##
##           Df Sum of Sq    RSS    AIC
## - UNEMPL   1      5.56 3125.4 1040.3
## <none>                 3119.8 1041.2
## - PROD     1      9.99 3129.8 1041.3
## - COMMPRI  1     16.52 3136.3 1042.7
## - PERSINC  1     79.31 3199.1 1055.7
## - HOUST    1     82.70 3202.5 1056.4
## - PCE      1    117.21 3237.0 1063.5
## - INFL     1    603.72 3723.5 1155.9
##
## Step:  AIC=1040.34
## INTRATE ~ INFL + PROD + COMMPRI + PCE + PERSINC + HOUST
##
##           Df Sum of Sq    RSS    AIC
## - PROD     1      4.44 3129.8 1039.3
## <none>                 3125.4 1040.3
## - COMMPRI  1     25.67 3151.0 1043.7
## - PERSINC  1     82.88 3208.2 1055.6
## - HOUST    1    108.45 3233.8 1060.8
## - PCE      1    150.68 3276.0 1069.4
## - INFL     1    600.15 3725.5 1154.3
##
## Step:  AIC=1039.28
## INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST
##
##           Df Sum of Sq    RSS    AIC
## <none>                 3129.8 1039.3
## - COMMPRI  1     38.65 3168.4 1045.4
## - PERSINC  1     78.58 3208.4 1053.7
## - HOUST    1    104.91 3234.7 1059.0
## - PCE      1    158.82 3288.6 1070.0
## - INFL     1    759.48 3889.3 1180.7
```

**Final Model (using AIC)**

```
##
## Call:
## lm(formula = INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST,
##     data = df)
##
## Coefficients:
## (Intercept)         INFL      COMMPRI          PCE      PERSINC
##   -0.236287     0.717720    -0.007499     0.339822     0.240157
##       HOUST
##   -0.020519
```

**Stepwise Backward Regression using BIC**

```
## Start:  BIC=1077.1
## INTRATE ~ INFL + PROD + UNEMPL + COMMPRI + PCE + PERSINC + HOUST
##
##           Df Sum of Sq    RSS    BIC
## - UNEMPL   1      5.56 3125.4 1071.8
## - PROD     1      9.99 3129.8 1072.7
## - COMMPRI  1     16.52 3136.3 1074.1
## <none>                 3119.8 1077.1
## - PERSINC  1     79.31 3199.1 1087.2
## - HOUST    1     82.70 3202.5 1087.9
## - PCE      1    117.21 3237.0 1095.0
## - INFL     1    603.72 3723.5 1187.4
##
## Step:  AIC=1071.79
## INTRATE ~ INFL + PROD + COMMPRI + PCE + PERSINC + HOUST
##
##           Df Sum of Sq    RSS    BIC
## - PROD     1      4.44 3129.8 1066.2
## - COMMPRI  1     25.67 3151.0 1070.7
## <none>                 3125.4 1071.8
## - PERSINC  1     82.88 3208.2 1082.6
## - HOUST    1    108.45 3233.8 1087.8
## - PCE      1    150.68 3276.0 1096.4
## - INFL     1    600.15 3725.5 1181.2
##
## Step:  AIC=1066.23
## INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST
##
##           Df Sum of Sq    RSS    BIC
## <none>                 3129.8 1066.2
## - COMMPRI  1     38.65 3168.4 1067.8
## - PERSINC  1     78.58 3208.4 1076.1
## - HOUST    1    104.91 3234.7 1081.5
## - PCE      1    158.82 3288.6 1092.4
## - INFL     1    759.48 3889.3 1203.1
```

**Final Model (using BIC)**

```
##
## Call:
## lm(formula = INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST,
##     data = df)
##
## Coefficients:
## (Intercept)         INFL      COMMPRI          PCE      PERSINC
##   -0.236287     0.717720    -0.007499     0.339822     0.240157
##       HOUST
##   -0.020519
```

(b) Use *specific-to-general* to come to a model. Start by regressing the *federal funds rate* on only a constant and add 1 variable at a time. Is the model the same as in $(a)$?

- Starting regression only with constant and adding 1 variable at a time (the most significant one, i.e., with the least p-value) we obtain the following:

| | INFL | PROD | UNEMPL | COMMPRI | PCE | PERSINC | HOUST |
|---|---|---|---|---|---|---|---|
| intercept | 1.644129e+00 | 5.394704e+00 | 4.542960e+00 | 5.401737e+00 | -3.331470e-01 | 5.126281e+00 | 5.403626e+00 |
| coef | 9.446992e-01 | -1.610555e-02 | 4.534938e-01 | -1.152313e-02 | 8.289447e-01 | 1.034739e-01 | -3.095185e-02 |
| intercept.pval | 1.932206e-23 | 1.591793e-139 | 4.255914e-95 | 6.372498e-169 | 2.371635e-01 | 3.017588e-95 | 5.054543e-173 |
| coef.pval | 2.291631e-119 | 5.870581e-01 | 1.978701e-10 | 6.133917e-03 | 7.367723e-80 | 1.502290e-01 | 4.313714e-07 |
| R^2 | 5.598580e-01 | | | | | | |

```
## Start:  AIC=1698.89
## INTRATE ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + INFL     1    4833.0 3799.5 1159.3
## + PCE      1    3625.0 5007.5 1341.5
## + UNEMPL   1     515.5 8117.0 1660.2
## + HOUST    1     329.0 8303.5 1675.2
## + COMMPRI  1      98.0 8534.5 1693.3
## + PERSINC  1      27.1 8605.4 1698.8
## <none>                 8632.5 1698.9
## + PROD     1       3.9 8628.6 1700.6
##
## Step:  AIC=1159.26
## INTRATE ~ INFL
##
##           Df Sum of Sq    RSS    AIC
## + PERSINC  1    456.15 3343.4 1076.8
## + PCE      1    383.68 3415.9 1091.0
## + UNEMPL   1    281.56 3518.0 1110.4
## + PROD     1    129.22 3670.3 1138.4
## + COMMPRI  1     16.08 3783.5 1158.5
## <none>                 3799.5 1159.3
## + HOUST    1      4.95 3794.6 1160.4
##
## Step:  AIC=1076.85
## INTRATE ~ INFL + PERSINC
##
##           Df Sum of Sq    RSS    AIC
## + PCE      1    58.683 3284.7 1067.2
## + HOUST    1    38.112 3305.3 1071.3
## + UNEMPL   1    35.684 3307.7 1071.8
## + COMMPRI  1    25.996 3317.4 1073.7
## <none>                 3343.4 1076.8
## + PROD     1     0.494 3342.9 1078.8
##
## Step:  AIC=1067.16
## INTRATE ~ INFL + PERSINC + PCE
##
##           Df Sum of Sq    RSS    AIC
## + HOUST    1   116.254 3168.4 1045.4
## + COMMPRI  1    49.991 3234.7 1059.0
## + PROD     1    11.147 3273.5 1066.9
## <none>                 3284.7 1067.2
## + UNEMPL   1     7.229 3277.5 1067.7
##
## Step:  AIC=1045.38
## INTRATE ~ INFL + PERSINC + PCE + HOUST
##
##           Df Sum of Sq    RSS    AIC
## + COMMPRI  1    38.647 3129.8 1039.3
## + PROD     1    17.416 3151.0 1043.7
## <none>                 3168.4 1045.4
## + UNEMPL   1     0.210 3168.2 1047.3
##
## Step:  AIC=1039.28
## INTRATE ~ INFL + PERSINC + PCE + HOUST + COMMPRI
##
##          Df Sum of Sq    RSS    AIC
## <none>                3129.8 1039.3
## + PROD    1    4.4425 3125.3 1040.3
## + UNEMPL  1    0.0077 3129.8 1041.3
```

**Stepwise Forward Regression using AIC**

**As we can see, in both the cases (a) and (b) we obtain the same final model.**

```
##
## Call:
## lm(formula = INTRATE ~ INFL + PERSINC + PCE + HOUST + COMMPRI,
##     data = df)
##
## Coefficients:
## (Intercept)         INFL      PERSINC          PCE        HOUST
##   -0.236287     0.717720     0.240157     0.339822    -0.020519
##     COMMPRI
##   -0.007499
```

**Final Model (using AIC)**

As can be seen from above, the **unrestricted model** is $INTRATE = X_1\beta_1 + X_2\beta_2 + \epsilon$ and

the **restricted model** is $INTRATE = X_1\beta_1 + \epsilon$,

where $X_1 = X(INFL, COMMPRI, PCE, PERSINC, HOUST)$ and $X_2 = X(PROD, UNEMPL)$.

*Sum of square residual* for the unrestricted model = $3119.796$, whereas the *Sum of square residual* for the restricted model = $3129.795$.

Now the $F - statistic$ for our hypothesis test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$ is $F = \frac{(R_1^2 - R_0^2)/g}{(1-R_1^2)/(n-k)} = \frac{(0.6386 - 0.63744)/2}{(1-0.6386)/(660-8)} = 1.046375$, and the *p-value* is $0.3517967$ with degrees of freedoms $2, 653$. Similar results we obtain from **wald test** and **anova**, so we can't reject our **null hypothesis** at 5% significance level, it confirms that we shall be better off without the extra variables (with the **restricted model**).

(c) Compare your model from $(a)$ and the *Taylor rule* of equation $i_t = \beta_1 + \beta_2 \pi_t + \beta_3 y_t + \epsilon_t$. Consider $R^2$, *AIC* and *BIC*. Which of the models do you prefer?

- We compare both the unrestricted (with all 7 explanatory variables) and the restricted (final) model (with 5 explanatory variables) obtained from part (a) with Taylor model (with 2 explanatory variables). The results are as shown below.

| | Taylors (restricted) | unrestricted | restricted |
|---|---|---|---|
| *R^2* | 0.57483 | 0.6386 | 0.63744 |
| *AIC* | 3013.42062 | 2916.16577 | 2914.2777 |
| *BIC* | 3031.38958 | 2956.59593 | 2945.72338 |

**Taylor's (restricted) model**

```
INTRATE ~ INFL + PROD

Residuals:
    Min      1Q  Median      3Q     Max
-5.1842 -1.6619  0.0085  1.3745  7.9237

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.25075    0.17609   7.103 3.19e-12 ***
INFL         0.97441    0.03271  29.792  < 2e-16 ***
PROD         0.09475    0.01970   4.810 1.88e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.364 on 657 degrees of freedom
Multiple R-squared:  0.5748, Adjusted R-squared:  0.5735
F-statistic: 444.1 on 2 and 657 DF,  p-value: < 2.2e-16
```

**restricted model from (a)**

```
INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST

Residuals:
    Min      1Q  Median      3Q     Max
-7.1918 -1.5298 -0.0974  1.3829  7.6603

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.236287   0.230212  -1.026  0.30509
INFL         0.717720   0.056972  12.598  < 2e-16 ***
COMMPRI     -0.007499   0.002639  -2.842  0.00463 **
PCE          0.339822   0.058989   5.761 1.29e-08 ***
PERSINC      0.240157   0.059265   4.052 5.68e-05 ***
HOUST       -0.020519   0.004382  -4.682 3.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.188 on 654 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.6374, Adjusted R-squared:  0.6347
F-statistic:   230 on 5 and 654 DF,  p-value: < 2.2e-16
```

**unrestricted model from (a)**

```
INTRATE ~ INFL + PROD + UNEMPL + COMMPRI + PCE +PERSINC + HOUST

Residuals:
    Min      1Q  Median      3Q     Max
-7.4286 -1.4409 -0.1142  1.3547  7.7212

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.214701   0.244829  -0.877   0.3808
INFL         0.696766   0.062031  11.233  < 2e-16 ***
PROD        -0.057581   0.039848  -1.445   0.1489
UNEMPL       0.104141   0.096640   1.078   0.2816
COMMPRI     -0.005525   0.002973  -1.858   0.0636 .
PCE          0.342641   0.069230   4.949 9.49e-07 ***
PERSINC      0.246418   0.060525   4.071 5.25e-05 ***
HOUST       -0.019371   0.004660  -4.157 3.65e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.187 on 652 degrees of freedom
Multiple R-squared:  0.6386, Adjusted R-squared:  0.6347
F-statistic: 164.6 on 7 and 652 DF,  p-value: < 2.2e-16
```

Now the $F-statistic$ for our hypothesis test to test the Taylor's (restricted) model against the unrestricted model, we found $F = 23.01$, and the *p-value* is $< 0.05$. Similar results we obtain from **wald test** and **anova**, so we can reject our **null hypothesis** at 5% significance level, it confirms that Taylor's model is not better than the **unrestricted model**.

```
## Wald test
##
## Model 1: INTRATE ~ INFL + PROD + UNEMPL + COMMPRI + PCE + PERSINC + HOUST
## Model 2: INTRATE ~ INFL + PROD
##   Res.Df Df     F    Pr(>F)
## 1    652
## 2    657 -5 23.01 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Variance Table
##
## Model 1: INTRATE ~ INFL + PROD + UNEMPL + COMMPRI + PCE + PERSINC + HOUST
## Model 2: INTRATE ~ INFL + PROD
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1    652 3119.8
## 2    657 3670.3 -5   -550.51 23.01 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Similarly with **anova** test we can see that the *p-value* is small, so we can reject the null hypothesis again and conclude that Taylor's model is not better than the **restricted model** (we can't do **walds test** because these models are not nested).

```
## Analysis of Variance Table
##
## Model 1: INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST
## Model 2: INTRATE ~ INFL + PROD
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    654 3129.8
## 2    657 3670.3 -3   -540.51 37.648 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(d) Test the *Taylor rule* of equation $i_t = \beta_1 + \beta_2 \pi_t + \beta_3 y_t + \epsilon$ using the *RESET test*, *Chow break* and *forecast test* (with in both tests as break date January $1985$) and a *Jarque-Bera test*. What do you conclude?

- *Sum of square residual* for the unrestricted model = $S1 = 3119.796$, whereas the *Sum of square residual* for the restricted Taylor's model = $S0 = 3670.306$ and the *F-statistic* for the **RESET Test** $F = \frac{(S_0 - S_1)/g}{S_1/(n-k)} = \frac{(3119.796 - 3670.306)/1}{3119.796/(660-4)} = 115.7558$, with *p-value=0*, we can reject the *null hypothesis* at 5% significance level, concluding that *we reject that the model is a linear regression model.*

$S_0$

- The total *Sum of square residual* for the Taylor's model for *1960:1 - 2014:12* is $= 3670.306$, whereas the *Sum of square residual* before 1985 January (*1960:1 - 1984:12*) $= S_1 = 1757.802$, whereas the *Sum of square residual* for on or after 1985 January ((*1985:1 - 2014:12*)) $= S_2 = 1813.22$. Hence, the *F-statistic* for the **Chow break Test** $F = \frac{(S_0 - (S_1 + S_2))/k}{(S_1 + S_2)/(n - 2*k)} = 4.578194$, with *p-value=0.003568257*, we can reject the *null hypothesis* at 5% significance level that *the model parameters do not suffer from the break*

- The *F-statistic* for the **Chow forecast Test** $F = \frac{(S_0 - S_1)/n_2}{S_1/(n_1 - k)} = 0.8976073$, with *p-value* as $0.4422323$, where we have $n_1 = 300,\ n_2 = 360$. The *p-value* is not *significant* at 5% level, hence we can't reject the *null hypothesis* that *there is no structural change in the prediction period.*

- The *Jarque-Bera test statistic* is $JB = \sqrt{(\frac{n}{6} S)^2} + \sqrt{(\frac{n}{24}(K-3))^2} = 9.171319$, where *skewness* of residuals $S = 0.3213945$ and *kurtosis* $K = 3.164455$, with *p-value* $0.01019702$ so that we can reject the *null hypothesis* $H_0$ that residuals are *normally distributed*, at 5% level of significance.

**Conclusion:** we can conclude that the model requires non-linear terms of the regressors and linear regression is not appropriate here (also rejecting the hypothesis that the error is normally distributed). The chow break test indicates that the model parameters sufffer from the break, although surprisingly enough, chow forecast test indicates that there is no structural change during the prediction period.