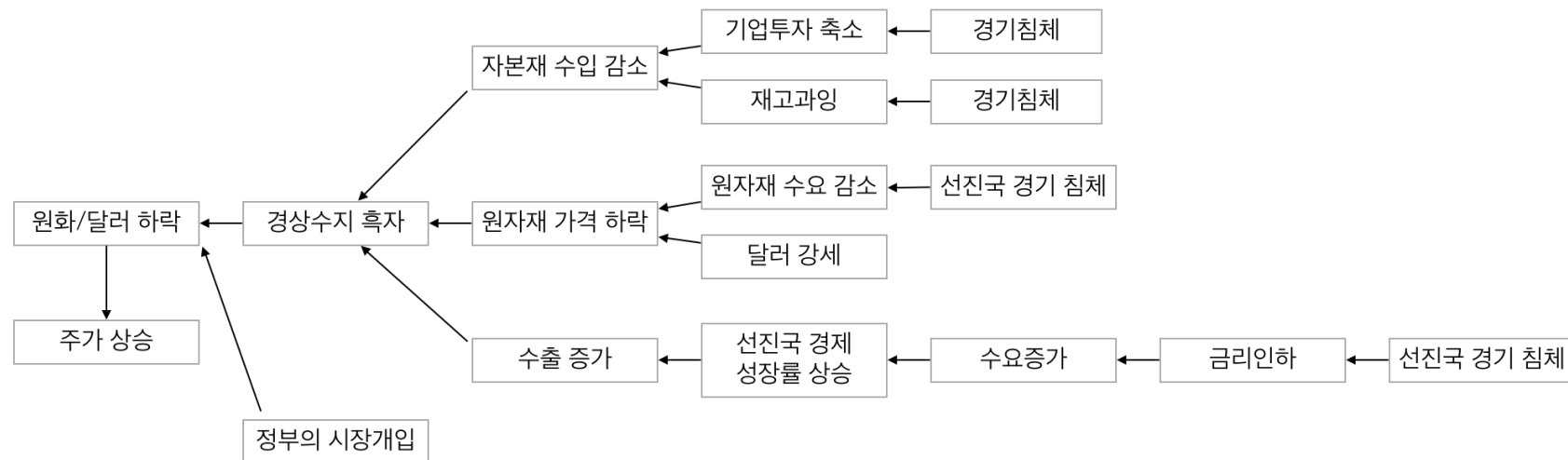
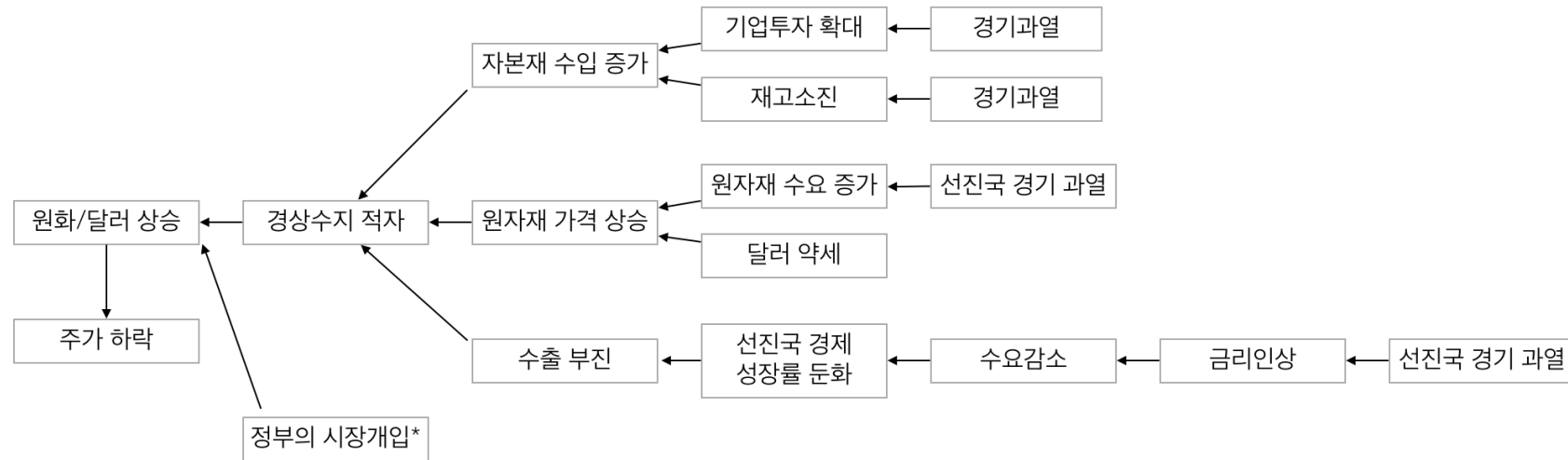


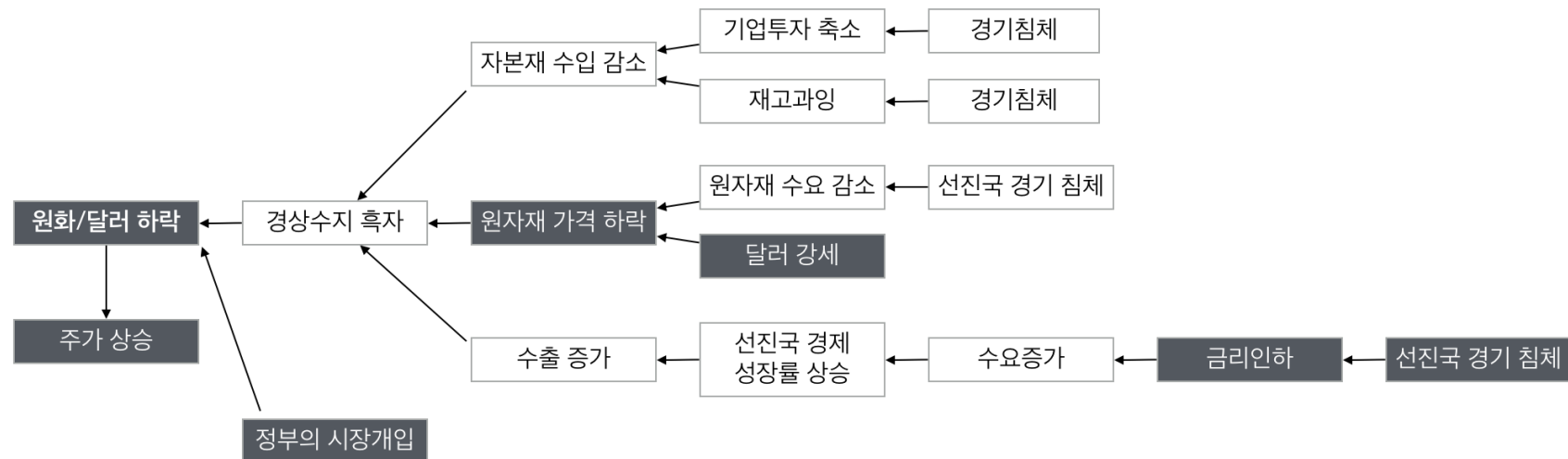
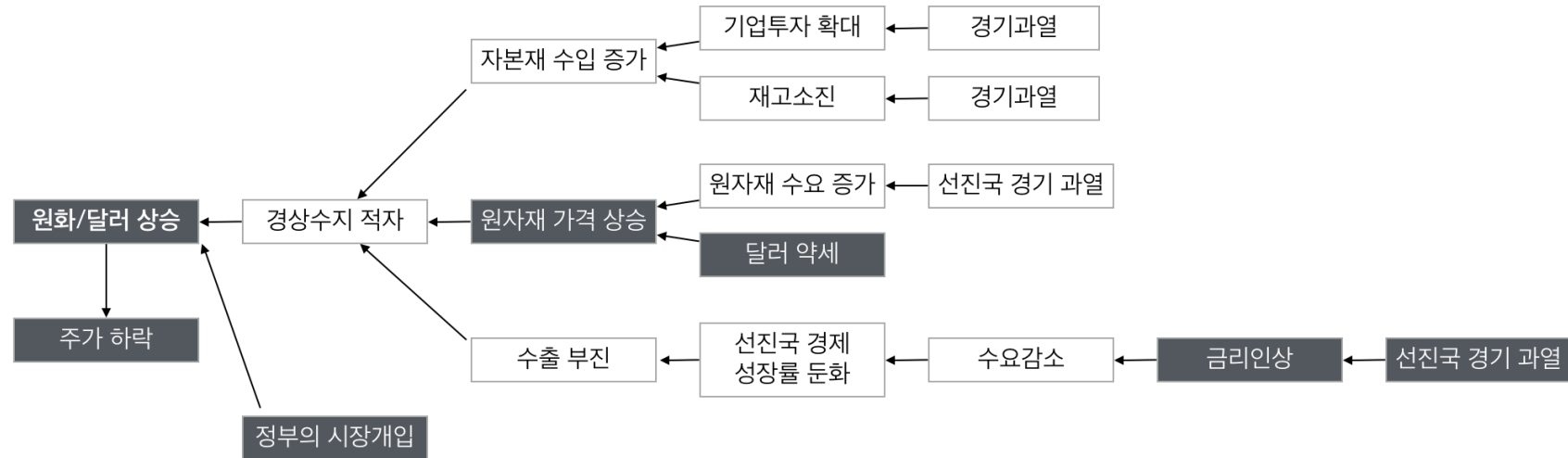
환율 예측 모델링

2018.05

1. 환율 변동 메커니즘



2. 변수 발굴



3. Data 수집 (Qaundl)

No	국내외 구분	변수그룹	변수명	설명	데이터 원천
1	국내	목표변수	fx_krw_usd	미국 달러화 환율	Quandl("QUANDL/USDKRW")
2		금리	yield_krx_b	KRX 채권 지수	KRX채권지수 - 유형별/기간별지수 (시장가격지수 사용)
3			yield_krx_g	KRX 국고채프라임지수	국고채프라임지수 - 만기물별/일별지수
4			yield_krx_t	KRX-Korea Treasury Bond Index	KRX-Korea Treasury Bond Index - 일자별지수
5		KRX 섹터 지수	stck_kospi	코스피 지수	Quandl("YAHOO/INDEX_KS11")
6			stck_krx_100	KRX_100	KRX_100
7			stck_krx_autos	KRX Autos	KRX Autos
8			stck_krx_energy	KRX Energy & Chemicals	KRX Energy & Chemicals
9			stck_krx_it	KRX IT	KRX IT
10			stck_krx_semicon	KRX Semicon	KRX Semicon
11			stck_krx_ship_buid	KRX Shipbuilding	KRX Shipbuilding
12			stck_krx_steels	KRX Steels	KRX Steels
13			stck_krx_trans	KRX Transportation	KRX Transportation
14	주요국 환율		fx_krw_aud	호주 달러화 환율	Quandl("QUANDL/AUDKRW")
15			fx_krw_cny	중국 위안화 환율	Quandl("QUANDL/CNYKRW")
16			fx_krw_gbp	영국 파운드화 환율	Quandl("QUANDL/GBPKRW")
17			fx_krw_eur	유로화 환율	Quandl("QUANDL/EURKRW")
18	해외	USD 지수	usd_index	US Dollar Index	Quandl("FRED/DTWEXB")
19		원자재 가격	cmd_copper	동 가격	Quandl("WSJ/COPPER")
20			cmd_corn	옥수수 가격	Quandl("WSJ/CORN_FEED")
21			cmd_gold	금 가격	Quandl("LBMA/GOLD")
22			cmd_oil_brent	브렌트유 가격	Quandl("FRED/DCOILBRENTU")
23			cmd_oil_wti	서부텍사스유 가격	Quandl("FRED/DCOILWTICO")
24			cmd_gas	천연가스 가격	Quandl("YAHOO/INDEX_XNG")
25			cmd_silver	은 가격	Quandl("LBMA/SILVER")

No	국내외 구분	변수그룹	변수명	설명	데이터 원천
26	주요국 주가지수		stck_cac40	프랑스 주가지수	Quandl("YAHOO/INDEX_FCHI")
27			stck_dax	독일 주가지수	Quandl("YAHOO/INDEX_GDAXI")
28			stck_nasdaq	미국 나스닥지수	Quandl("NASDAQOMX/NDX")
29			stck_nikkei	일본 주가지수	Quandl("YAHOO/INDEX_N225")
30			stck_nyse	미국 나이스지수	Quandl("YAHOO/INDEX_NYA")
31			stck_snp500	미국 S&P 500 지수	Quandl("YAHOO/INDEX_GSPC")
32			stck_ssec	중국 주가지수	Quandl("YAHOO/INDEX_SSEC")
33		금리	yield_ca	캐나다 금리	Quandl("YIELDCURVE/CAN")
34			yield_jp	일본 금리	Quandl("YIELDCURVE/JPN")
35			yield_fr	프랑스 금리	Quandl("YIELDCURVE/FRA")
36	주요국 대달러 환율		yield_nz	뉴질랜드 금리	Quandl("YIELDCURVE/NZL")
37			yield_uk	영국 금리	Quandl("YIELDCURVE/GBR")
38			yield_us	미국 금리	Quandl("YIELDCURVE/USA")
39			fx_aud_usd	대 미국 달러 호주 달러화 환율	Quandl("QUANDL/USDAUD")
40			fx_cad_usd	대 미국 달러 캐나다 달러화 환율	Quandl("QUANDL/USDCAD")
41			fx_cny_usd	대 미국 달러 중국 위안화 환율	Quandl("QUANDL/USDCNY")
42			fx_eur_usd	대 미국 달러 유로화 환율	Quandl("BNP/USDEUR")
43			fx_gbp_usd	대 미국 달러 영국 파운드화 환율	Quandl("QUANDL/USDGBP")
44			fx_jpy_usd	대 미국 달러 일본 엔화 환율	Quandl("QUANDL/USDJPY")
45			fx_nzd_usd	대 미국 달러 뉴질랜드 달러화 환율	Quandl("QUANDL/USDNZD")

5. Data 수집 (KRX)

KRX Indices

[KRX지수](#) |
 [산업분류](#) |
 [정보센터](#) |
 [고객서비스](#) |
 [About Us](#)

KRX지수

[전체지수 시세](#)
[주가지수](#) -
KRX시리즈
 • KOSPI시리즈
 • KOSDAQ시리즈
 • 테마지수

[파생상품지수](#) +
[전략형지수](#) +
[글로벌지수](#)
[채권지수](#) +

> KRX지수 > 주가지수 > KRX시리즈

KRX시리즈

2005년 한국거래소의 통합출범을 계기로 산출하는 지수 시리즈로 유가증권시장과 코스닥시장의 보통주를 대상으로 하여 그중 대표종목을 선별하여 지수를 산출하고 있습니다.
KRX 시리즈 지수로 KRX 100과 KRX 섹터지수 17종 등 총 18종이 산출되고 있습니다.

(2014/09/05 오후 03:34:15)

구분	지수명	기준시점	발표시점	기준지수	현재지수	전일비	등락률	시가총액
대표지수	KRX 100	2001. 1. 2	2005. 6. 1	1,000	4,332.16	▼ 22.89	-0.53	9,377,068,069
섹터지수	KRX Autos	2006. 1. 2	2006. 1.23	1,000	2,201.74	▼ 18.91	-0.85	1,284,599,409
	KRX Semicon	2006. 1. 2	2006. 1.23	1,000	1,752.13	▲ 18.78	1.08	2,165,857,489
	KRX Health Care	2006. 1. 2	2006. 1.23	1,000	1,413.27	▲ 12.31	0.88	120,263,923
	KRX Banks	2006. 1. 2	2006. 1.23	1,000	898.88	▼ 7.36	-0.81	804,520,005
	KRX IT	2006. 1. 2	2006. 1.23	1,000	1,475.32	▼ 3.17	-0.21	3,051,896,322

6. Mart 만들기

(1) Merge Data(merge, cbind, sqldf 활용)

① sqldf 사용

sql의 left outer join을 이용하여 Date를 기준으로 자료를 나란히 붙인다. 해당 날짜에 없는 data는 NA로 세팅된다.

② cbind 사용

cbind는 순서를 그대로 유지하고 연결할 때 사용한다. 수집한 data 중 KRX의 금리 지수가 모든 일자를 포함하고 있어 cbind를 사용할 수 있다. 사용 전에 두 data frame의 Date가 순서가 맞는지를 table 함수를 이용해 확인하였다.

③ merge 사용

merge는 공통 변수가 있는 경우 이를 기준으로 데이터를 나란히 연결한다. 여기서는 Date를 이용하여 연결하였다.

(2) Missing Data Handling과 Date Shift

Missing Value는 전일자의 수치를 이용했다. 이 때 주의할 점은 목표변수도 전일자로 채워서는 안된다.

나머지 변수만 채우고, 목표변수의 NA는 na.omit으로 제거해야 한다. 이 때도 순서를 지켜야 한다. 채우고 제거해야지, 제거하고 채우면 다른 값들이 들어오게 된다.

7. Data 분리

모델링을 위해서는 data를 다음과 같이 셋으로 나누어야 한다.

- train data set: 모델링을 위해 투입하는 data set으로 해당기간 내 data를 test set과 7:3 정도의 비율로 나눈다.
- test data set: train data set과 동일 기간의 모델 검증용 data set이다.

```
ind<-sample(2,nrow(model_data),prob=c(0.7,0.3),replace=T)
train_data<-model_data[ind==1,]
train_diff<-model_diff[ind==1]
test_data<-model_data[ind==2,]
test_diff<-model_diff[ind==2]
```

8. Algorithm 적용

```
lm_fx<-lm(f_krw_usd~.,train_data[,-1])
```

```
summary(lm_fx)
```

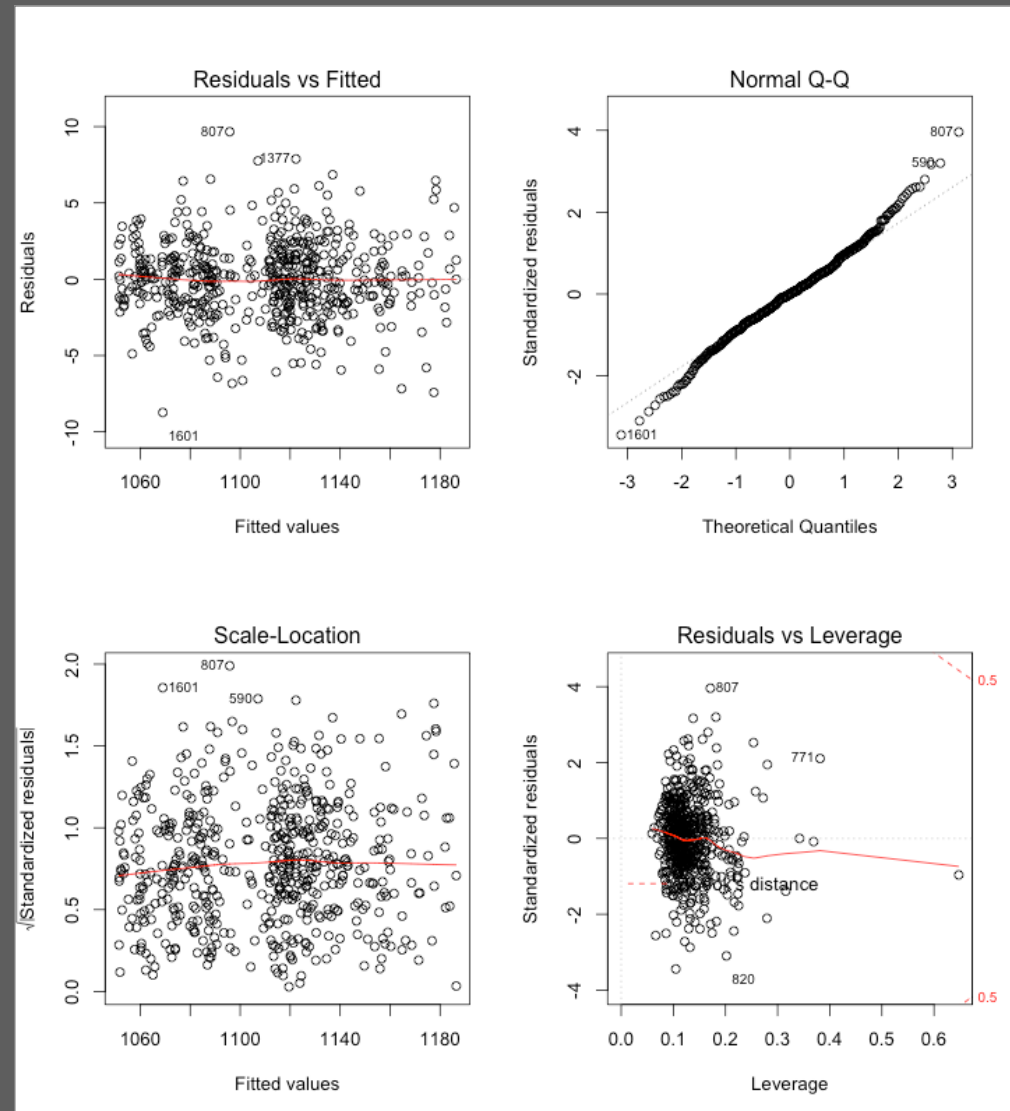
```
## Call:
## lm(formula = f_krw_usd ~ ., data = train_data[, -1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.734  -1.532  -0.002   1.427   9.663
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.77e+02  1.25e+02  -4.60  5.4e-06 ***
## c_copper      -1.88e+00  1.88e+00  -1.00  0.31766
## c_oil_wti      2.29e-01  8.11e-02   2.83  0.00487 **
## c_gas         -1.85e-02  1.46e-02  -1.27  0.20522
## f_aud_usd     -1.43e+02  1.05e+02  -1.36  0.17380
## f_cad_usd     -1.63e+02  3.35e+01  -4.86  1.6e-06 ***
## f_cny_usd      3.50e+01  4.24e+01   0.83  0.40956
## f_eur_usd      1.13e+02  4.24e+01   2.66  0.00819 **
## f_gbp_usd      1.14e+03  3.82e+02   2.99  0.00292 **

<< 종략>>

## s_krx_autos    -1.08e-02  3.08e-03  -3.50  0.00051 ***
## s_krx_energy    4.62e-03  2.73e-03   1.69  0.09166 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.68 on 478 degrees of freedom
## Multiple R-squared:  0.994, Adjusted R-squared:  0.993
## F-statistic: 1.15e+03 on 71 and 478 DF, p-value: <2e-16
```


9. Model 검증

```
par(mfrow=c(2,2))  
plot(lm_fx)
```



10. Step

(1)전진 선택법 (Forward selection)

```
lm_fx_fwd<-step(l_lm,direction="forward")
```

(2)후진 제거법 (Backward elimination)

```
lm_fx_bwd<-step(lm_fx,direction="backward")
```

(3)단계적 방법 (Stepwise method)

```
lm_fx_bth<-step(lm_fx,direction="both")
```

EOD