Regression result April 15

Xiaolan

4/19/2023

Preparation

The **expected profit** of crop i in province j and year t is defined as

$$E\pi_{ijt} = \tilde{y}_{ijt}(\tilde{p}_{it} - c_{ijt}) + \tilde{d}_{ijt}$$

where:

- $\tilde{p}_{it}$: future price (yuan per kilogram) averaged in March for Novermber delivery in year t
- c_{ijt}: the realized cost (yuan per kilogram) of crop i in province j and year t
- $\tilde{y}_{ijt}$: the yield (kilogram per mu) of crop i in province j averaged in the past three years, $\{t-3, t-2, t-1\}$.
- $\tilde{d}_{ijt}$: the expected subsidy (yuan per mu) of crop i in province j weighted for the past two years, with $\tilde{d}_{ijt} = 0.67 * d_{ij,t-1} + 0.33 * d_{ij,t-2}$

The **realized profit** of crop i in province j and year t is defined as

$$\pi_{ijt} = y_{ijt}(p_{ijt} - c_{ijt}) + d_{ijt}$$

where p_{ijt} is the realized price (yuan per kilogram) for crop i in province j and year t.

Thus, the expected revenue of crop i in province j and year t is

$$Er_{ijt} = \tilde{y}_{ijt}\tilde{p}_{it} + \tilde{d}_{ijt} = E\pi_{ijt} + \tilde{y}_{ijt}c_{ijt}$$

We define the relative revenue of crop i to a baseline crop wheat as:

$$rr_{ijt} = r_{ijt}/r_{3jt}$$

OLS regression

Suppose s_{ijt} is the share of cropland area for crop i in province j and year t. Let $s_{0jt} = 1 - \sum_{i=1}^{3} s_{ijt}$.

We define $z_{ijt} = \log(s_{ijt}/s_{0jt})$ as the dependent variable.

Model 1 : Use expected profit as explanatory variable

$$z_{ijt} = \beta_{i0} + \sum_{k=1}^{3} E\pi_{kjt}\beta_{ik} + u_j + \epsilon_{ijt}$$

where

- u_j: fixed effect for province j.
- ϵ_{ijt}: random error, assumed i.i.d. with normal distribution

Model 2 : Use expected revenue

$$z_{ijt} = \beta_{i0} + \sum_{k=1}^{3} Er_{kjt}\beta_{ik} + u_j + \epsilon_{ijt}$$

Model 3: Use expected relative revenue

$$z_{ijt} = \beta_{i0} + \sum_{k=1}^{2} Err_{kjt}\beta_{ik} + u_j + \epsilon_{ijt}$$

Some figures:

(1) Expected profit against acreage share for corn:

(2) Expected revenue against acreage share:

(3) Logriathm of expected revenue against acreage share:

```
lm_corn_pf <- lm(y2 ~ 0 + pfex_corn + pfex_rice + pfex_soy  + region,
              data = regdat %>% filter(crop == "corn"))

lm_rice_pf <- lm(y2 ~ 0 + pfex_corn + pfex_rice + pfex_soy  + region,
              data = regdat %>% filter(crop == "rice"))

lm_soy_pf <- lm(y2 ~ 0 + pfex_corn + pfex_rice + pfex_soy  + region,
              data = regdat %>% filter(crop == "soybean"))


stargazer(lm_corn_pf, lm_rice_pf, lm_soy_pf, title = "Model 1 Results",column.labels = c("Corn", "Rice"
```

Model 1 Regression

```
##
## Model 1 Results
## ================================================================
##                                          y2
##                                  Corn       Rice     Soybean
##                                   (1)       (2)        (3)
## ----------------------------------------------------------------
```

```
## pfex_corn                       -0.0005   -0.0003     -0.001
##                                (0.0003)   (0.0004)   (0.0004)
##
## pfex_rice                      -0.001*** -0.001***    0.0001
##                                (0.0001)   (0.0001)   (0.0001)
##
## pfex_soy                        0.0002     0.0001     0.001*
##                                (0.0004)   (0.0005)   (0.0005)
##
## regionNeimenggu                0.316***  -3.190***  -1.529***
##                                 (0.116)   (0.128)    (0.136)
##
## regionJilin                    1.854***    0.116    -1.161***
##                                 (0.095)   (0.105)    (0.111)
##
## regionLiaoning                 1.273***   -0.234**  -2.134***
##                                 (0.096)   (0.106)    (0.113)
##
## regionHeilongjiang             1.615***   0.959***   0.798***
##                                 (0.102)   (0.112)    (0.119)
##
## ---------------------------------------------------------------
## Observations                      80         80         80
## R2                               0.901     0.968      0.941
## Adjusted R2                      0.892     0.965      0.936
## Residual Std. Error (df = 73)    0.328     0.362      0.384
## F Statistic (df = 7; 73)      94.975*** 314.206*** 166.889***
## ===============================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01
##
## Model 1 Results
## =
## 3
## -
```

```
lm_corn_rev <- lm(y2 ~ 0 + revex_corn + revex_rice + revex_soy   + region,
              data = regdat %>% filter(crop == "corn"))

lm_rice_rev <- lm(y2 ~ 0 + revex_corn + revex_rice + revex_soy  + region,
              data = regdat %>% filter(crop == "rice"))

lm_soy_rev <- lm(y2 ~ 0 + revex_corn + revex_rice + revex_soy   + region,
              data = regdat %>% filter(crop == "soybean"))



stargazer(lm_corn_rev, lm_rice_rev, lm_soy_rev,  title = "Model 2 Results",column.labels = c("Corn", "R
```

Model 2 Regression

```
##
```

```
## Model 2 Results
## ===============================================================
##                                          y2
##                                  Corn       Rice     Soybean
##                                   (1)       (2)        (3)
## ---------------------------------------------------------------
## revex_corn                       0.001*     0.0005     -0.001
##                                 (0.0004)   (0.0004)   (0.0004)
##
## revex_rice                      0.001***   0.001**    0.001**
##                                 (0.0004)   (0.0004)   (0.0004)
##
## revex_soy                       0.002***   0.001**     0.001
##                                  (0.001)   (0.001)    (0.001)
##
## regionNeimenggu                -4.164*** -6.522***  -3.099***
##                                  (0.773)   (0.831)    (0.890)
##
## regionJilin                    -2.853*** -3.381***  -2.788***
##                                  (0.767)   (0.824)    (0.882)
##
## regionLiaoning                 -3.576*** -3.847***  -3.917***
##                                  (0.811)   (0.872)    (0.934)
##
## regionHeilongjiang             -2.734*** -2.275***    -0.779
##                                  (0.721)   (0.775)    (0.830)
##
## ---------------------------------------------------------------
## Observations                       80         80         80
## R2                                0.902      0.970      0.943
## Adjusted R2                       0.892      0.967      0.938
## Residual Std. Error (df = 73)     0.327      0.351      0.376
## F Statistic (df = 7; 73)       95.850*** 334.140*** 174.003***
## ===============================================================
## Note:                                *p<0.1; **p<0.05; ***p<0.01
##
## Model 2 Results
## =
## 3
## -
```

```
lm_corn_rr <- lm(y2 ~ 0 + rr_corn + rr_rice  + region, # rc_corn + rc_rice + rc_soy +
              data = regdat %>% filter(crop == "corn"))

lm_rice_rr <- lm(y2 ~ 0 + rr_corn + rr_rice  + region,
              data = regdat %>% filter(crop == "rice"))

lm_soy_rr <- lm(y2 ~ 0 +  rr_corn + rr_rice  +  region,
              data = regdat %>% filter(crop == "soybean"))

stargazer(lm_corn_rr, lm_rice_rr, lm_soy_rr, title = "Model 3 Results",column.labels = c("Corn", "Rice"
```

Model 3 Regression

```
##
## Model 3 Results
## ============================================================
##                                          y2
##                                  Corn       Rice     Soybean
##                                   (1)       (2)        (3)
## ------------------------------------------------------------
## rr_corn                        0.535***   0.484**     -0.166
##                                 (0.190)   (0.194)    (0.205)
##
## rr_rice                        -0.268*** -0.187***    0.022
##                                 (0.031)   (0.031)    (0.033)
##
## regionNeimenggu                 -0.268   -3.820***  -1.321***
##                                 (0.382)   (0.390)    (0.413)
##
## regionJilin                    1.283***    -0.425   -0.937***
##                                 (0.260)   (0.265)    (0.280)
##
## regionLiaoning                 0.802***  -0.690***  -1.914***
##                                 (0.239)   (0.244)    (0.258)
##
## regionHeilongjiang             1.131***    0.474*    0.972***
##                                 (0.267)   (0.272)    (0.288)
##
## ------------------------------------------------------------
## Observations                      80         80         80
## R2                               0.878     0.966      0.938
## Adjusted R2                      0.868     0.963      0.933
## Residual Std. Error (df = 74)    0.362     0.369      0.391
## F Statistic (df = 6; 74)      88.750*** 351.123*** 186.964***
## ============================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01
##
## Model 3 Results
## =
## 3
## -
```

Model 4

use the last year's true profit

```
lm_corn_pftr <- lm(y2 ~ 0 + pftr_corn + pftr_rice + pftr_soy + region,
              data = regdat %>% filter(crop == "corn"))

lm_rice_pftr <- lm(y2 ~ 0 + pftr_corn + pftr_rice + pftr_soy + region,
              data = regdat %>% filter(crop == "rice"))

lm_soy_pftr <- lm(y2 ~ 0 + pftr_corn + pftr_rice + pftr_soy +  region,
              data = regdat %>% filter(crop == "soybean"))
```

```
stargazer(lm_corn_pftr, lm_rice_pftr, lm_soy_pftr, title = "Model 4 Results",column.labels = c("Corn", "
```

```
##
## Model 4 Results
## ===============================================================
##                                          y2
##                                  Corn       Rice      Soybean
##                                   (1)       (2)        (3)
## ---------------------------------------------------------------
## pftr_corn                      -0.002**   -0.001**    -0.001
##                                 (0.001)   (0.001)    (0.0005)
##
## pftr_rice                       0.001**    0.0005    -0.001*
##                                (0.0004)   (0.0004)   (0.0003)
##
## pftr_soy                         0.001     0.0002     0.001
##                                 (0.001)   (0.001)    (0.001)
##
## regionNeimenggu                -0.448*** -3.622***  -1.330***
##                                 (0.162)   (0.144)    (0.124)
##
## regionJilin                    0.970***  -0.414***  -1.000***
##                                 (0.162)   (0.144)    (0.124)
##
## regionLiaoning                  0.335*   -0.783***  -1.925***
##                                 (0.179)   (0.160)    (0.137)
##
## regionHeilongjiang             0.851***   0.508***   0.956***
##                                 (0.144)   (0.128)    (0.110)
##
## ---------------------------------------------------------------
## Observations                      80         80         80
## R2                               0.780     0.953      0.944
## Adjusted R2                      0.759     0.949      0.939
## Residual Std. Error (df = 73)    0.489     0.436      0.374
## F Statistic (df = 7; 73)      36.949*** 213.016*** 175.621***
## ===============================================================
## Note:                               *p<0.1; **p<0.05; ***p<0.01
##
## Model 4 Results
## =
## 3
## -
```

Model 5

```
lm_corn_rr2 <- lm(y2 ~ 0 + rr_corn2 + rr_soy2  + region, # rc_corn + rc_rice + rc_soy +
              data = regdat %>% filter(crop == "corn"))

lm_rice_rr2 <- lm(y2 ~ 0 + rr_corn2 + rr_soy2  + region,
```

```
              data = regdat %>% filter(crop == "rice"))

lm_soy_rr2 <- lm(y2 ~ 0 +  rr_corn2 + rr_soy2  +  region,
              data = regdat %>% filter(crop == "soybean"))

stargazer(lm_corn_rr2, lm_rice_rr2, lm_soy_rr2, title = "Model 5 Results",column.labels = c("Corn", "Ri
```

```
##
## Model 5 Results
## ===============================================================
##                                          y2
##                                 Corn       Rice     Soybean
##                                  (1)       (2)        (3)
## ---------------------------------------------------------------
## rr_corn2                         0.312     0.076     -1.252*
##                                 (0.720)   (0.712)    (0.723)
##
## rr_soy2                         2.695**    2.110*     1.439
##                                 (1.116)   (1.104)    (1.122)
##
## regionNeimenggu                -1.226*** -4.237***  -1.289***
##                                 (0.172)   (0.170)    (0.173)
##
## regionJilin                      0.053   -1.137***  -0.975***
##                                 (0.172)   (0.170)    (0.173)
##
## regionLiaoning                 -0.446*** -1.436***  -1.999***
##                                 (0.167)   (0.165)    (0.168)
##
## regionHeilongjiang              -0.055     -0.200    0.934***
##                                 (0.157)   (0.155)    (0.157)
##
## ---------------------------------------------------------------
## Observations                      80         80         80
## R2                               0.865     0.965      0.941
## Adjusted R2                      0.854     0.962      0.936
## Residual Std. Error (df = 74)    0.381     0.377      0.383
## F Statistic (df = 6; 74)      78.798*** 336.412*** 195.100***
## ===============================================================
## Note:                               *p<0.1; **p<0.05; ***p<0.01
##
## Model 5 Results
## =
## 3
## -
```

Prediction Results True Result in 2021

```
share_true %>% data.frame() %>% "colnames<-"(unique(regdat$region)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	48.09	71.14	62.93	43.31
rice	1.77	13.53	12.03	25.67
soybean	10.22	4.08	2.4	25.81

Model 1 Prediction

```
share_est_pf %>% data.frame() %>% "colnames<-"(unique(regdat$region)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	43.21	68.54	62.06	42.75
rice	1.46	12.99	14.46	23.5
soybean	9.68	4.97	3.16	23.58

Model 2 Prediction

```
share_est_rev %>% data.frame() %>% "colnames<-"(unique(regdat$region)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	59.4	77.73	70.61	52.26
rice	1.67	12.13	14.3	24.36
soybean	6.15	2.72	2.21	16.69

Model 3 Prediction

```
share_est_rr %>% data.frame() %>% "colnames<-"(unique(regdat$region)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	49.81	68.25	58.61	42.53
rice	1.52	13.12	14.34	23.45
soybean	8.25	4.57	3.1	23.45

Model 4 Prediction

```
share_est_pftr %>% data.frame() %>% "colnames<-"(unique(regdat$region)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	26.95	59.42	56.45	31.44
rice	1.1	13.12	14.25	21.47

	Neimenggu	Jilin	Liaoning	Heilongjiang
soybean	12.04	6.33	3.66	33.75

Model 5 prediction

```
share_est_rr2 %>% data.frame() %>% "colnames<-"(unique(regdat$region)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	55.88	74.68	65.63	51.1
rice	1.63	12.71	14.9	24.87
soybean	6.27	3.09	2.54	16.8

```
MSE1 <- rowMeans((share_true - share_est_pf)^2)
MSE2 <- rowMeans((share_true - share_est_rev)^2)
MSE3 <- rowMeans((share_true - share_est_rr)^2)
MSE4 <- rowMeans((share_true - share_est_pftr)^2)
MSE5 <- rowMeans((share_true - share_est_rr2)^2)

cbind(MSE1, MSE2, MSE3, MSE4, MSE5) %>% "colnames<-"(paste0("Model", 1:5)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

MSE Results

	Model1	Model2	Model3	Model4	Model5
corn	7.911	77.61	7.645	191.8	35.3
rice	2.75	2.21	2.624	5.796	2.392
soybean	1.659	25.41	2.545	18.25	24.45

{Cross Validation}

We will use cross validation to compare the models.

```
# combine two datasets
alldat <- rbind(regdat, preddat) %>% arrange(region, crop, year)

alldat <- alldat %>%
    mutate(ex_revenue = yield_lag * ex_price + subsidy_lag,
         tcost = yield_lag * cost,
         rcratio = (yield_lag * ex_price + subsidy_lag)/(yield_lag * cost))

alldat <- alldat %>%
    mutate(share_wheat = alldat %>% filter(crop == "wheat") %>% "$"(share) %>%
           matrix(., nrow = 21) %>% # use 21 here since we have 21 years
           kronecker(rep(1, 4), .) %>% as.vector()) %>%
  mutate(y2 = log(share/(other + share_wheat)))   # get new y2 since now we combine other
```

```
alldat <- alldat %>%
  mutate(pfex_corn = alldat %>% filter(crop == "corn") %>% "$"(profit_ex) %>%
           matrix(., nrow = 21) %>% # "+"(400) %>% log() %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         pfex_rice = alldat %>% filter(crop == "rice") %>% "$"(profit_ex) %>%
           matrix(., nrow = 21) %>% # "+"(400) %>% log() %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         pfex_soy = alldat %>% filter(crop == "soybean") %>% "$"(profit_ex) %>%
           matrix(., nrow = 21) %>% # "+"(400) %>% log() %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         pfex_wheat = alldat %>% filter(crop == "wheat") %>% "$"(profit_ex) %>%
           matrix(., nrow = 21) %>% # "+"(400) %>% log() %>%
           kronecker(rep(1, 4), .) %>% as.vector(),

         pftr_corn = alldat %>% filter(crop == "corn") %>% "$"(profit_true_lag) %>%
           matrix(., nrow = 21) %>% # "+"(400) %>% log() %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         pftr_rice = alldat %>% filter(crop == "rice") %>% "$"(profit_true_lag) %>%
           matrix(., nrow = 21) %>% # "+"(400) %>% log() %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         pftr_soy = alldat %>% filter(crop == "soybean") %>% "$"(profit_true_lag) %>%
           matrix(., nrow = 21) %>% # "+"(400) %>% log() %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         pftr_wheat = alldat %>% filter(crop == "wheat") %>% "$"(profit_true_lag) %>%
           matrix(., nrow = 21) %>% # "+"(400) %>% log() %>%
           kronecker(rep(1, 4), .) %>% as.vector(),

         revex_corn = alldat %>% filter(crop == "corn") %>% "$"(ex_revenue) %>%
           matrix(., nrow = 21) %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         revex_rice = alldat %>% filter(crop == "rice") %>% "$"(ex_revenue) %>%
           matrix(., nrow = 21) %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         revex_soy = alldat %>% filter(crop == "soybean") %>% "$"(ex_revenue) %>%
           matrix(., nrow = 21) %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         revex_wheat = alldat %>% filter(crop == "wheat") %>% "$"(ex_revenue) %>%
           matrix(., nrow = 21) %>%
           kronecker(rep(1, 4), .) %>% as.vector(),

         tcost_corn = alldat %>% filter(crop == "corn") %>% "$"(tcost) %>%
           matrix(., nrow = 21) %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         tcost_rice = alldat %>% filter(crop == "rice") %>% "$"(tcost) %>%
           matrix(., nrow = 21) %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         tcost_soy = alldat %>% filter(crop == "soybean") %>% "$"(tcost) %>%
           matrix(., nrow = 21) %>%
           kronecker(rep(1, 4), .) %>% as.vector(),
         tcost_wheat = alldat %>% filter(crop == "wheat") %>% "$"(tcost) %>%
           matrix(., nrow = 21) %>%
           kronecker(rep(1, 4), .) %>% as.vector())%>%
  mutate(rr_corn = revex_corn/revex_soy,
```

```
        rr_rice = revex_rice/revex_soy,
        rc_corn = tcost_corn/tcost_soy,
        rc_rice = tcost_rice/tcost_soy)
```

Start the cross validation:

```
RMSE1_list <- NULL
RMSE2_list <- NULL
RMSE3_list <- NULL
RMSE4_list <- NULL
RMSE5_list <- NULL
for(i in 2001:2021)
{
  # Step 1: select regdat and preddat
  regdat <- alldat %>% filter(year != i)
  preddat <- alldat %>% filter(year == i)

  # Step 2: get the true share
  share_true <- preddat$share %>% matrix(., nrow = 4) %>% "["(-4,)

  # Step 3: run all the models

  # model 1
  lm_corn_pf <- lm(y2 ~ 0 + pfex_corn + pfex_rice + pfex_soy  + region,
              data = regdat %>% filter(crop == "corn"))

  lm_rice_pf <- lm(y2 ~ 0 + pfex_corn + pfex_rice + pfex_soy  + region,
                data = regdat %>% filter(crop == "rice"))

  lm_soy_pf <- lm(y2 ~ 0 + pfex_corn + pfex_rice + pfex_soy  + region,
                data = regdat %>% filter(crop == "soybean"))

  # model 2
  lm_corn_rev <- lm(y2 ~ 0 + revex_corn + revex_rice + revex_soy   + region,
              data = regdat %>% filter(crop == "corn"))

  lm_rice_rev <- lm(y2 ~ 0 + revex_corn + revex_rice + revex_soy  + region,
                data = regdat %>% filter(crop == "rice"))

  lm_soy_rev <- lm(y2 ~ 0 + revex_corn + revex_rice + revex_soy  + region,
                data = regdat %>% filter(crop == "soybean"))

  # model 3

  lm_corn_rr <- lm(y2 ~ 0 + rr_corn + rr_rice + region, # rc_corn + rc_rice + rc_soy +
              data = regdat %>% filter(crop == "corn"))

  lm_rice_rr <- lm(y2 ~ 0 + rr_corn + rr_rice + region,
                data = regdat %>% filter(crop == "rice"))

  lm_soy_rr <- lm(y2 ~ 0 +  rr_corn + rr_rice +  region,
                data = regdat %>% filter(crop == "soybean"))
```

```r
# model 4

lm_corn_pftr <- lm(y2 ~ 0 + pftr_corn + pftr_rice + pftr_soy + region,
            data = regdat %>% filter(crop == "corn"))

lm_rice_pftr <- lm(y2 ~ 0 + pftr_corn + pftr_rice + pftr_soy + region,
              data = regdat %>% filter(crop == "rice"))

lm_soy_pftr <- lm(y2 ~ 0 + pftr_corn + pftr_rice + pftr_soy +  region,
              data = regdat %>% filter(crop == "soybean"))

# model 5
lm_corn_rrrc <- lm(y2 ~ 0 + rr_corn + rr_rice + rc_corn + rc_rice + region, # rc_corn + rc_rice + rc_
            data = regdat %>% filter(crop == "corn"))

lm_rice_rrrc <- lm(y2 ~ 0 + rr_corn + rr_rice + rc_corn + rc_rice + region,
              data = regdat %>% filter(crop == "rice"))

lm_soy_rrrc <- lm(y2 ~ 0 +  rr_corn + rr_rice + rc_corn + rc_rice + region,
              data = regdat %>% filter(crop == "soybean"))

# Step 4: calculate the predict share:

# model 1
pd_corn_pf <- predict(lm_corn_pf, newdata = preddat %>% filter(crop == "corn"))
pd_rice_pf <- predict(lm_rice_pf, newdata = preddat %>% filter(crop == "rice"))
pd_soy_pf <- predict(lm_soy_pf, newdata = preddat %>% filter(crop == "soybean"))

y_pred_pf <- rbind(pd_corn_pf, pd_rice_pf, pd_soy_pf)

share_est_pf <- round(sweep(exp(y_pred_pf), MARGIN = 2, FUN = "/", STATS =
                              colSums(exp(y_pred_pf)) + 1) * 100, 2)

# model 2
pd_corn_rev <- predict(lm_corn_rev, newdata = preddat %>% filter(crop == "corn"))
pd_rice_rev <- predict(lm_rice_rev, newdata = preddat %>% filter(crop == "rice"))
pd_soy_rev <- predict(lm_soy_rev, newdata = preddat %>% filter(crop == "soybean"))

y_pred_rev <- rbind(pd_corn_rev, pd_rice_rev, pd_soy_rev)

share_est_rev <- round(sweep(exp(y_pred_rev), MARGIN = 2, FUN = "/", STATS =
                               colSums(exp(y_pred_rev)) + 1) * 100, 2)

# model 3:
pd_corn_rr<- predict(lm_corn_rr, newdata = preddat %>% filter(crop == "corn"))
pd_rice_rr <- predict(lm_rice_rr, newdata = preddat %>% filter(crop == "rice"))
pd_soy_rr <- predict(lm_soy_rr, newdata = preddat %>% filter(crop == "soybean"))

y_pred_rr <- rbind(pd_corn_rr, pd_rice_rr, pd_soy_rr)

share_est_rr <- round(sweep(exp(y_pred_rr), MARGIN = 2, FUN = "/", STATS =
                              colSums(exp(y_pred_rr)) + 1) * 100, 2)
```

```
  # model 4:
  pd_corn_pftr <- predict(lm_corn_pftr, newdata = preddat %>% filter(crop == "corn"))
  pd_rice_pftr <- predict(lm_rice_pftr, newdata = preddat %>% filter(crop == "rice"))
  pd_soy_pftr <- predict(lm_soy_pftr, newdata = preddat %>% filter(crop == "soybean"))

  y_pred_pftr <- rbind(pd_corn_pftr, pd_rice_pftr, pd_soy_pftr)

  share_est_pftr <- round(sweep(exp(y_pred_pftr), MARGIN = 2, FUN = "/", STATS =
                                colSums(exp(y_pred_pftr)) + 1) * 100, 2)

  # model 5:
  pd_corn_rrrc <- predict(lm_corn_rrrc, newdata = preddat %>% filter(crop == "corn"))
  pd_rice_rrrc <- predict(lm_rice_rrrc, newdata = preddat %>% filter(crop == "rice"))
  pd_soy_rrrc <- predict(lm_soy_rrrc, newdata = preddat %>% filter(crop == "soybean"))

  y_pred_rrrc <- rbind(pd_corn_rrrc, pd_rice_rrrc, pd_soy_rrrc)

  share_est_rrrc <- round(sweep(exp(y_pred_rrrc), MARGIN = 2, FUN = "/", STATS =
                                colSums(exp(y_pred_rrrc)) + 1) * 100, 2)

  # Step 5: calculate the RMSE

  RMSE1 <- rowMeans((share_true - share_est_pf)^2) %>% sqrt()
  RMSE2 <- rowMeans((share_true - share_est_rev)^2) %>% sqrt()
  RMSE3 <- rowMeans((share_true - share_est_rr)^2) %>% sqrt()
  RMSE4 <- rowMeans((share_true - share_est_pftr)^2) %>% sqrt()
  RMSE5 <- rowMeans((share_true - share_est_rrrc)^2) %>% sqrt()

  # Step 6: combine the results
  RMSE1_list <- rbind(RMSE1_list, RMSE1)
  RMSE2_list <- rbind(RMSE2_list, RMSE2)
  RMSE3_list <- rbind(RMSE3_list, RMSE3)
  RMSE4_list <- rbind(RMSE4_list, RMSE4)
  RMSE5_list <- rbind(RMSE5_list, RMSE5)
}

RMSE1_avg <- colMeans(RMSE1_list)
RMSE2_avg <- colMeans(RMSE2_list)
RMSE3_avg <- colMeans(RMSE3_list)
RMSE4_avg <- colMeans(RMSE4_list)
RMSE5_avg <- colMeans(RMSE5_list)
```

The RMSE table:

```
cbind(RMSE1_avg, RMSE2_avg, RMSE3_avg, RMSE4_avg, RMSE5_avg) %>%
  "colnames<-"(paste0("Model", 1:5)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

	Model1	Model2	Model3	Model4	Model5
corn	3.964	4.234	4.818	7.639	4.906
rice	1.747	1.857	1.761	2.503	1.668

	Model1	Model2	Model3	Model4	Model5
soybean	2.342	2.724	2.818	3.411	2.897

The standard deviation of RMSE

```
RMSE1_sd <- apply(RMSE1_list, MARGIN = 2, FUN = sd)
RMSE2_sd <- apply(RMSE2_list, MARGIN = 2, FUN = sd)
RMSE3_sd <- apply(RMSE3_list, MARGIN = 2, FUN = sd)
RMSE4_sd <- apply(RMSE4_list, MARGIN = 2, FUN = sd)
RMSE5_sd <- apply(RMSE5_list, MARGIN = 2, FUN = sd)

cbind(RMSE1_sd, RMSE2_sd, RMSE3_sd, RMSE4_sd, RMSE5_sd) %>%
  "colnames<-"(paste0("Model", 1:5)) %>%
  "rownames<-"(unique(regdat$crop)[-4]) %>% pander::pander()
```

	Model1	Model2	Model3	Model4	Model5
corn	0.9372	3.028	1.655	3.18	1.887
rice	0.6454	0.6957	0.7443	0.8889	0.7561
soybean	1.65	1.862	1.909	1.736	2.003