Regression result April 15

Xiaolan

4/15/2023

Preparation

The true share for provinces in 2021 are

```
share_true %>% data.frame() %>% "colnames<-"(unique(regdat$region)) %>%
  "rownames<-"(unique(regdat$crop)) %>% 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
wheat	5.06	0.07	0.06	0.45

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:

$$rr_{ijt} = r_{ijt}/r_{4jt}$$

OLS regression

Suppose s_{ijt} is the share of cropland area for crop i in province j and year t. Let $s_{0jt} = 1 - \sum_{i=1}^{4} s_{ijt}$. We define $z_{ijt} = \log(s_{ijt}/s_{0jt})$ as the dependent variable.

Model 1 : Use expected profit as explanatory variable

$$z_{ijt} = \beta_{i0} + \sum_{k=1}^{4} E\pi_{kjt}\beta_{ik} + u_j + \epsilon_{ijt}$$

where

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

Model 2 : Use expected revenue

$$z_{ijt} = \beta_{i0} + \sum_{k=1}^{4} Er_{kjt}\beta_{ik} + u_j + \epsilon_{ijt}$$

Model 3: Use expected relative revenue

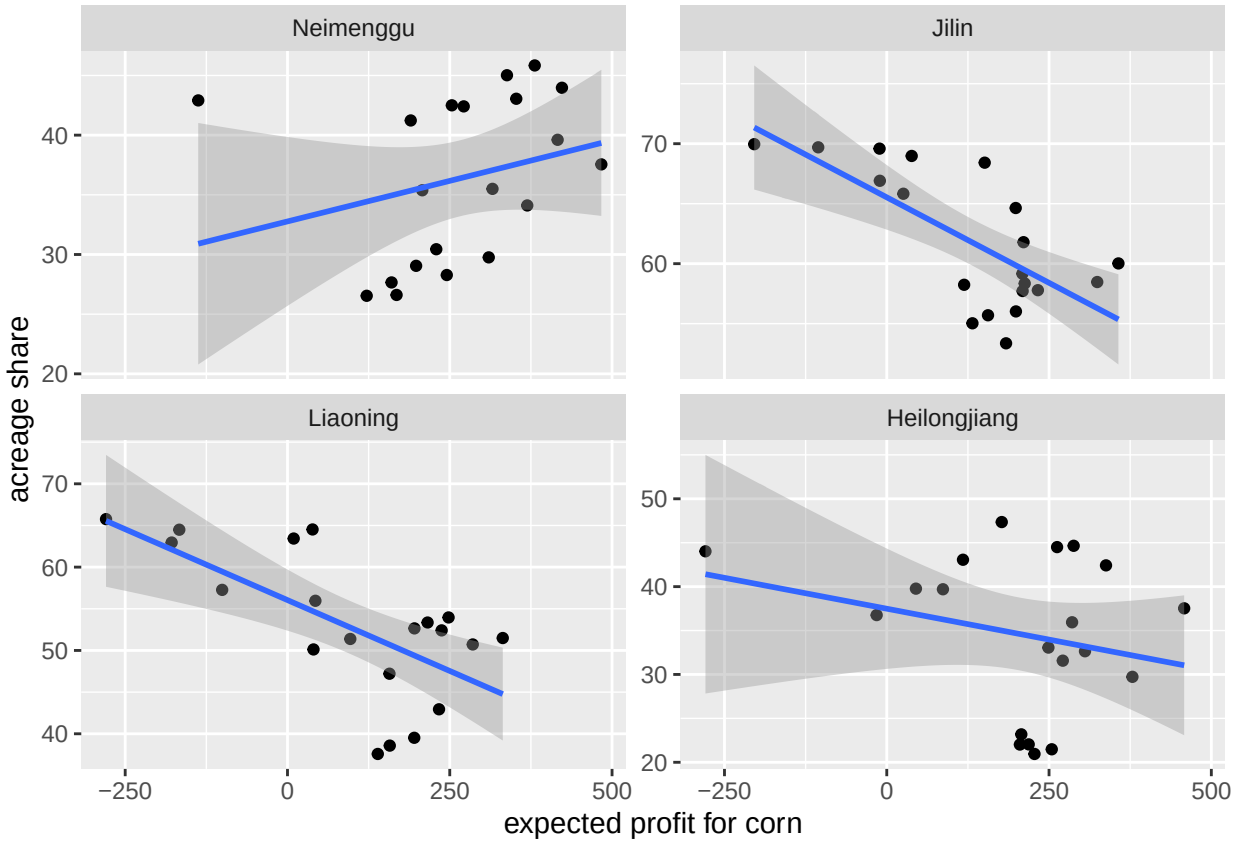
$$z_{ijt} = \beta_{i0} + \sum_{k=1}^{3} Err_{kjt}\beta_{ik} + u_j + \epsilon_{ijt}$$

Some figures:

(1) Expected profit against acreage share for corn:

```
ggplot(aes(x = pfex_corn, y = share), data = regdat %>% filter(crop == "corn")) + geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~region, nrow = 2, scales = "free_y")+
  labs(x = "expected profit for corn", y = "acreage share")
```

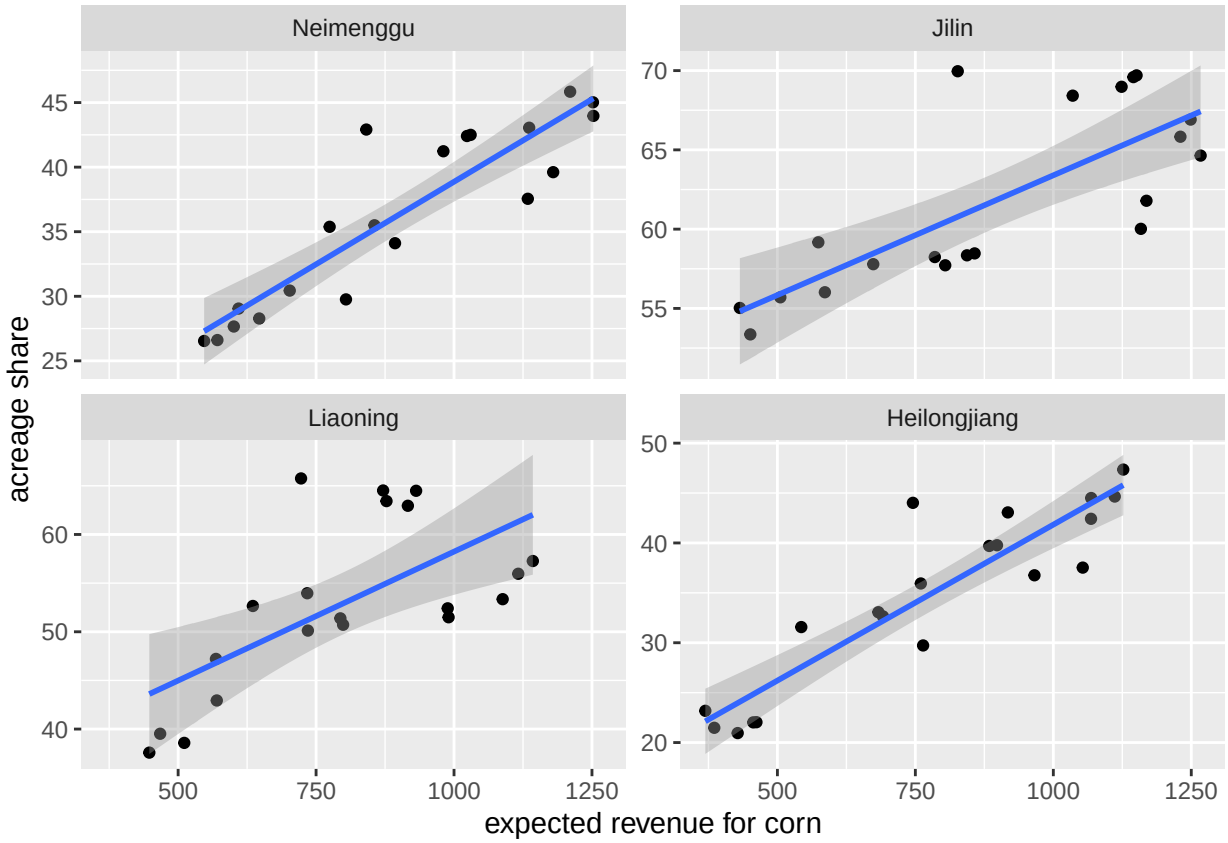
```
## `geom_smooth()` using formula 'y ~ x'
```



(2) Expected revenue against acreage share:

```
ggplot(aes(x = revex_corn, y = share), data = regdat %>% filter(crop == "corn")) + geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~region, nrow = 2, scales = "free_y")+
  labs(x = "expected revenue for corn", y = "acreage share")
```

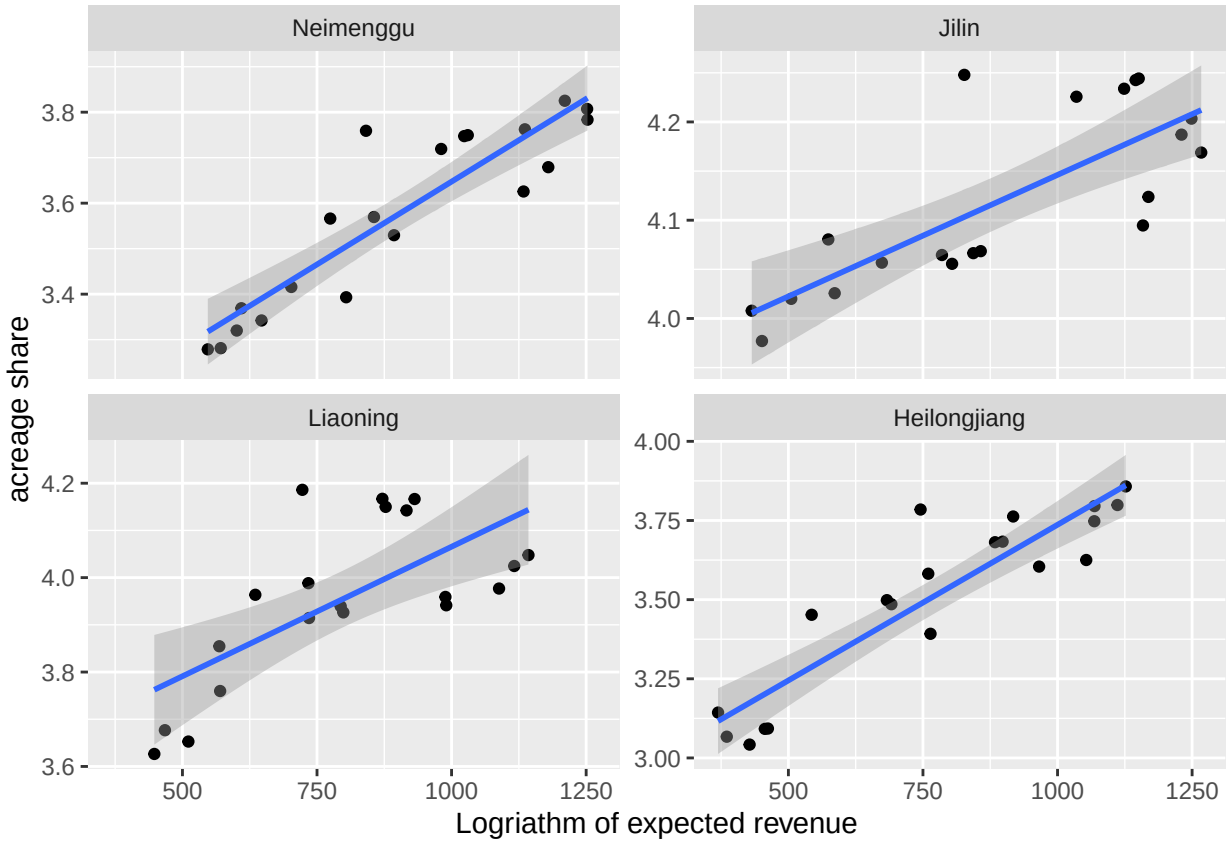
```
## `geom_smooth()` using formula 'y ~ x'
```



(3) Logriathm of expected revenue against acreage share:

```
ggplot(aes(x = revex_corn, y = log(share)), data = regdat %>% filter(crop == "corn")) + geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~region, nrow = 2, scales = "free_y")+
  labs(x = "Logriathm of expected revenue", y = "acreage share")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




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

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

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

lm_wheat_pf <- lm(y ~ 0 + pfex_corn + pfex_rice + pfex_soy + pfex_wheat + region,
              data = regdat %>% filter(crop == "wheat"))
#summary(lm_corn_pf)
#summary(lm_rice_pf)
#summary(lm_soy_pf)
#summary(lm_wheat_pf)
stargazer(lm_corn_pf, lm_rice_pf, lm_soy_pf, lm_wheat_pf, title = "Model 1 Results",column.labels = c("
```

Model 1 Regression

```
##
## Model 1 Results
## ================================================================================
##                                                 y
```

```
##                                 Corn       Rice     Soybean     Wheat
##                                  (1)       (2)        (3)       (4)
## ------------------------------------------------------------------------
## pfex_corn                      -0.0005    -0.001     -0.001     0.001
##                               (0.0004)   (0.0004)   (0.0004)   (0.001)
##
## pfex_rice                     -0.001*** -0.001***   0.00002    0.001***
##                               (0.0001)   (0.0001)   (0.0001)   (0.0002)
##
## pfex_soy                       0.0003     0.0003    0.001**     0.001
##                               (0.0004)   (0.0005)   (0.001)    (0.001)
##
## pfex_wheat                     0.0004     0.001*     0.001     -0.0002
##                               (0.0005)   (0.001)    (0.001)    (0.001)
##
## regionNeimenggu               0.493***  -2.937***  -1.312***  -2.418***
##                                (0.134)    (0.142)    (0.154)    (0.244)
##
## regionJilin                   1.816***    0.019    -1.230***  -5.911***
##                                (0.108)    (0.115)    (0.124)    (0.197)
##
## regionLiaoning                1.306***    -0.157   -2.078***  -5.432***
##                                (0.105)    (0.112)    (0.121)    (0.192)
##
## regionHeilongjiang            1.699***    0.996***   0.857***  -2.624***
##                                (0.111)    (0.118)    (0.128)    (0.202)
##
## ------------------------------------------------------------------------
## Observations                     80         80         80         80
## R2                              0.901      0.966      0.938      0.978
## Adjusted R2                     0.890      0.963      0.931      0.975
## Residual Std. Error (df = 72)   0.340      0.361      0.390      0.619
## F Statistic (df = 8; 72)      82.015*** 258.523*** 136.528*** 399.086***
## ========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
##
## Model 1 Results
## =
## 3
## -
```

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

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

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

lm_wheat_rev <- lm(y ~ 0 + revex_corn + revex_rice + revex_soy + revex_wheat  + region,
```

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

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

Model 2 Regression

```
##
## Model 2 Results
## ================================================================================
##                                                 y
##                                    Corn        Rice       Soybean      Wheat
##                                     (1)         (2)         (3)         (4)
## --------------------------------------------------------------------------------
## revex_corn                          0.001      0.0004     -0.0003     -0.0004
##                                    (0.0004)   (0.0004)   (0.0005)    (0.001)
##
## revex_rice                        0.001***    0.001**     0.001**     -0.0003
##                                    (0.0004)   (0.0004)   (0.0005)    (0.001)
##
## revex_soy                           0.001      0.001       0.001       -0.002
##                                    (0.001)    (0.001)     (0.001)     (0.001)
##
## revex_wheat                         0.001**     0.001      -0.0003      0.0002
##                                    (0.0005)   (0.001)     (0.001)     (0.001)
##
## regionNeimenggu                   -4.362*** -6.576***   -2.896***     -0.373
##                                     (0.775)    (0.850)     (0.929)     (1.543)
##
## regionJilin                       -2.789*** -3.347***   -2.807***    -3.663**
##                                     (0.759)    (0.832)     (0.910)     (1.511)
##
## regionLiaoning                    -3.774*** -3.948***   -3.844***    -3.197*
##                                     (0.807)    (0.885)     (0.967)     (1.606)
##
## regionHeilongjiang                -2.668*** -2.184***     -0.644      -0.567
##                                     (0.714)    (0.783)     (0.856)     (1.420)
##
## --------------------------------------------------------------------------------
## Observations                          80          80          80          80
## R2                                   0.910       0.967       0.939       0.976
## Adjusted R2                          0.901       0.964       0.932       0.974
## Residual Std. Error (df = 72)        0.323       0.355       0.388       0.644
## F Statistic (df = 8; 72)          91.522*** 267.781*** 138.425*** 368.499***
## ================================================================================
## Note:                                            *p<0.1; **p<0.05; ***p<0.01
##
## Model 2 Results
## =
## 3
## -
```

```
#summary(lm_corn_rev)
#summary(lm_rice_rev)
```

```r
#summary(lm_soy_rev)
#summary(lm_wheat_rev)
```

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

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

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

lm_wheat_rr <- lm(y ~ 0 +  rr_corn + rr_rice + rr_soy +  region,
              data = regdat %>% filter(crop == "wheat"))

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

Model 3 Regression

```
##
## Model 3 Results
## ================================================================================
##                                                 y
##                                     Corn       Rice     Soybean     Wheat
##                                      (1)       (2)        (3)        (4)
## --------------------------------------------------------------------------------
## rr_corn                            0.368     0.324      -0.311     -0.829
##                                   (0.296)   (0.294)    (0.321)    (0.513)
##
## rr_rice                          -0.350*** -0.254***    0.017     0.269***
##                                   (0.040)   (0.040)    (0.044)    (0.070)
##
## rr_soy                             0.384     0.138      0.581      -0.096
##                                   (0.439)   (0.437)    (0.476)    (0.762)
##
## regionNeimenggu                   -0.012   -3.341***  -1.422***  -1.443***
##                                   (0.271)   (0.270)    (0.294)    (0.471)
##
## regionJilin                      1.404***    -0.112   -1.306***  -4.934***
##                                   (0.392)   (0.390)    (0.425)    (0.680)
##
## regionLiaoning                   0.830***   -0.468*   -2.227***  -4.660***
##                                   (0.281)   (0.279)    (0.305)    (0.487)
##
## regionHeilongjiang               1.371***   0.901***   0.767**   -1.761***
##                                   (0.335)   (0.333)    (0.364)    (0.581)
##
## --------------------------------------------------------------------------------
## Observations                        80         80         80         80
## R2                                 0.883     0.965      0.935      0.976
```

```
## Adjusted R2                       0.872     0.962      0.928      0.974
## Residual Std. Error (df = 73)     0.367     0.365      0.398      0.636
## F Statistic (df = 7; 73)      78.822*** 288.269*** 149.405*** 430.786***
## ============================================================================
## Note:                                          *p<0.1; **p<0.05; ***p<0.01
##
## Model 3 Results
## =
## 3
## -
```

```
#summary(lm_corn_rr)
#summary(lm_rice_rr)
#summary(lm_soy_rr)
#summary(lm_wheat_rr)
```

Prediction Results True Result in 2021

```
share_true %>% data.frame() %>% "colnames<-"(unique(regdat$region)) %>%
  "rownames<-"(unique(regdat$crop)) %>% 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
wheat	5.06	0.07	0.06	0.45

Model 1 Prediction

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

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	43.26	68.81	62.04	43.66
rice	1.45	12.37	15.22	22.58
soybean	9.66	4.85	3.25	23.33
wheat	7.64	0.06	0.15	1

Model 2 Prediction

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

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	56.49	73.76	67.52	46.33
rice	1.72	13.1	14.74	24.68

	Neimenggu	Jilin	Liaoning	Heilongjiang
soybean	7.43	3.7	2.61	21.48
wheat	2.49	0.02	0.05	0.37

Model 3 Prediction

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

	Neimenggu	Jilin	Liaoning	Heilongjiang
corn	48.71	71.74	60.2	45.64
rice	1.57	12.84	14.4	23.63
soybean	9.09	3.98	3.21	22.23
wheat	3.73	0.03	0.13	0.56

```
MSE1 <- rowMeans((share_true - share_est_pf)^2)
MSE2 <- rowMeans((share_true - share_est_rev)^2)
MSE3 <- rowMeans((share_true - share_est_rr)^2)

cbind(MSE1, MSE2, MSE3) %>% "colnames<-"(paste0("Model", 1:3)) %>%
  "rownames<-"(unique(regdat$crop)) %>% pander::pander()
```

MSE Results

	Model1	Model2	Model3
corn	7.418	26.9	3.407
rice	5.293	2.128	2.574
soybean	1.945	6.68	3.69
wheat	1.742	1.653	0.4469