USPS-LR-2

BEFORE THE
POSTAL RATE COMMISSION
WASHINGTON, DC 20260-0001

EXPERIMENTAL PRIORITY MAIL FLAT-RATE BOX, 2004	Docket No. MC2004-2

Library Reference USPS-LR-2/MC2004-2:

PRIORITY MAIL PARCEL SIZE DISTRIBUTION AND DENSITY STUDY,

SUPPORTING MATERIALS OF WITNESS LOETSCHER (USPS-T-3)

TABLE OF CONTENTS

1. SUMMARY 1

2. METHODOLOGY 1

   A. SAMPLE DESIGN 1
   B. DATA COLLECTION 2

3. DATA ANALYSIS 3
   A. ESTIMATION 3
   B. MODEL 3
   C. CALCULATION OF CUBIC FEET DISTRIBUTIONS 5
   D. DENSITY CALCULATIONS 6

APPENDIX A – LISTING OF ELECTRONIC FILES 7
APPENDIX B – EXCELL MACRO “DISTMAC” 8
APPENDIX C – STATA COMMANDS 12
APPENDIX D – STATA OUTPUT 14

Priority Mail Parcel Size Distribution and Density Study

1. Summary

This report provides national estimates of the size distributions and density for Priority Mail parcels. These size distributions consist of estimates of the distribution of piece sizes by zone and by pound step for Priority Mail parcels. The data needed to produce these estimates were obtained from a study conducted at ten sample sites in October 2002–January 2003, where piece characteristics were collected from a sample of destinating Priority Mail parcels. The size distributions are presented in Tables 1-6 of USPS-T-3.

The study plan and data analysis are documented below. The sample design, data collection, and data processing are described in Section 2. Data analysis is discussed in Section 3. Associated documentation is provided in the Appendices.

2. Methodology

A. Sample design

Destinating Priority Mail parcels were sampled at the plant prior to final distribution to delivery unit. In some areas this is a Priority Mail Processing Center (PMPC), in others an Area Distribution Center (ADC) or Sectional Center Facility (SCF). Sampling destinating mail at the SCF captures all Priority Mail flows except any pieces that originate and destinate in the same five-digit areas, and in doing so would bypass the SCF. This mail volume is negligible. Assuming the mail piece characteristics for these parcels are on average similar to the sampled mail, its exclusion should not bias the results. Since no *a priori* basis has been identified that would suggest the characteristics of excluded pieces are different, and given its negligible volume, these national estimates are statistically robust.

The universe for this survey is defined as the 464 SCFs (some of which are also ADCs) for which destinating Priority Mail volumes were recorded in ODIS for FY2002. Ten sample sites were randomly selected with probability proportional to total destinating Priority Mail volume in FY2002. Since sites were selected with probability proportional to destinating SCF ODIS volume and ADCs tend to be larger SCFs, the draw produced a sample of nine ADCs and one SCF. It was determined during initial contact with the sample sites that Priority Mail for some of the sites is processed at a nearby PMPC and transported directly to the delivery unit from the PMPC, or only crossdocked at the plant. In these cases sampling was conducted at the PMPC, as noted.

The population was not stratified before selecting sample sites, as typical stratification variables (e.g., geographic region, volume, revenue) were not *a priori* thought to be correlated with the variable under study (density characteristics). The sample selected does have some geographic dispersion. The sample includes the two highest destinating volume sites. Although the lowest ranked sample site is only the 181st site (out of 464), the top 181 sites represent over 80 percent of all Priority Mail volume.

Priority Mail Density Survey – Sample Sites		
	FY02 Destinating Priority Mail	
Sample Site	Volume	Rank
Phoenix, AZ 852	14,8511,957	(1)
New York, NY 100	14,646,514	(2)
North Metro, GA 300	9,765,566	(20)
Detroit, MI 481	8,507,889	(32)
Jacksonville, FL 320	6,108,588	(59)
Kansas City, MO 640	4,232,420	(93)
Hartford, CT 060	4,002,593	(100)
Everett, WA 982	3,534,212	(108)
Syracuse, NY 130	2,599,272	(147)
Wichita, KS 670	1,887,280	(181)

B. Data Collection

Only Priority Mail parcels were sampled for this study.

Sampling was done at ADCs or PMPCs where Priority Mail for multiple SCFs was typically being processed in one (SPBS or manual sortation). In order to sample only those parcels for the selected SCFs, sample pieces were selected either before or after processing to 5-digit ZIP Code, depending on the point in operation the mail for the sample SCF could be identified and sampled separately. Sample selection was determined by how best to sample for the selected SCF while not interfering with operations and retaining sufficient time to select and record sample pieces for the sample site.

Priority Mail parcels usually arrive at a plant loose in rolling stock. The rolling stock generally consists of two streams, machinable parcels and non-machinable outsides (NMO's). Because of this, a two-stage sampling process was used for each stream. Separate skip factors were used to select sample rolling stock and sample pieces within each selected rolling stock. Sampling rates at each sample site across rolling stock and across pieces were determined by expected volume.

Heavy parcels (pieces over 12 pounds) were over-sampled to permit more precise estimates for the higher weight ranges. For density calculations the observations are weighted by national RPW estimates of piece volume by pound increment.

The piece characteristics recorded for each sample piece were:

- Weight
- Dimensions (length, width, and height; or length and girth)
- Indicia
- Postage (where available)
- Ancillary services
- Use of Postal Service packaging
- Origin ZIP Code
- Destination ZIP Code
- Transportation mode (if known)[1]

After data processing and cleaning, 5,368 sample pieces were used in this analysis.

3. Data Analysis

A. Estimation

The distribution of cubic feet within a pound increment is estimated by first using ordinary least squares to estimate the mean cubic feet/weight relationship and the variance of cubic feet around the mean cubic feet value. The residual terms of this regression are tested for normality. Finally distributions of cubic feet are calculated using the estimated parameters from the regression.

B. Model

The equation used to estimate the relationship of cubic feet and weight is the translog function recommended by the Commission in Docket No. R94-1[2] for Parcel Post and used by witness Eggleston in LR-J-66 in Docket No. R2001-1, with additional terms included to measure the impact of postal zone on the weight/cube relationship. The estimation equation used is:

$$\ln(CUFT_i) = \beta_0 + \beta_1 \ln(WT_i) + \beta_2 \ln(WTSQ_i) + \beta_j ZONE_i + \varepsilon_i$$

[1] Transportation mode was determined using information from the ACT tag or from information received from transportation and logistics offices at the plant.
[2] PRC Op., Docket No. R94-1, page V-116.

The variable definitions are as follows:

$CUFT_i$ – is the cubic feet of the ith piece
WT_i - is the weight of the ith piece in pounds
$WTSQ_i$ – is the weight of the ith piece squared
$ZONE_i$ – is the postal zone (1-3, 4, 5, 6, 7, or 8) of the ith piece
j = 1-3,4,5,6,7,8.
ε_i - is a residual term representing the distribution of cubic feet

In the lower pound increments pieces exceeding 15 inches in length and thickness of less than 0.25 inches were observed (X-ray films, posters etc.). These pieces are parcels by DMM definitions but have very low cube. It was hypothesized because of these pieces the cube/weight relationship for pieces under 2 pounds would be different than for pound increments over 2 pounds. For pieces over 2 pounds the length + girth restrictions were more likely to be binding for thin pieces. The two populations (pieces under 2 pounds and pieces over 2 pounds) are compared and an F test rejects pooling the two populations. For this reason separate regressions are used for pieces under 2 pounds and for pieces over 2 pounds.[3]

The estimated regression statistics and parameters (with t-statistics in parentheses) are presented below:

	Regression Statistics and Parameter Estimates	
	Pieces Under 2 Pounds	**Pieces Over 2 Pounds**
Observations	2,880	2,488
Standard Error (σ)	0.6443	0.7006
R-Squared	0.2369	0.4585
Adjusted R-Squared	0.2351	0.4569
Intercept – β_0	-1.9906 (-84.42)	-2.6101 (-31.50)
WT – β_1	0.7361	1.5377

[3] The F statistic comparing separate regressions for the two groups relative to a restricted (i.e. pooled) regression is 16.75. With 8 restrictions and 5,352 degrees of freedom, the F test rejects pooling the two groups with greater than 99 percent confidence. In addition, tests of the error terms fail to reject the assumption that they are normally distributed at the 5 percent significance level, though for pieces over 2 pounds normality is rejected at the 10 percent significance level. The test for normality is based on D'Agostino, Balanger, and D'Agostino Jr., *The American Statistician* 44(4) (1990) as implemented in the Stata statistical software version 6.0. Stata commands are provided in Appendix C and electronically as the Stata file "errortest.do". Stata output of these tests are presented in Appendix D.

	(25.60)	(16.65)
WTSQ – β_2	0.1017 (3.06)	-0.1782 (-7.74)
Zone 4 – β_4	0.1035 (2.72)	0.2150 (4.85)
Zone 5 – β_5	0.0646 (1.86)	0.1947 (4.71)
Zone 6 – β_6	0.0500 (1.13)	0.2114 (4.50)
Zone 7 – β_7	0.0820 (1.62)	0.2198 (3.83)
Zone 8 – β_8	0.0223 (0.56)	0.2675 (6.06)

C. Calculation of Cubic Feet Distributions

The distributions are calculated by pound increment and zone. The distributions represent the probability that a piece lies between two cubic feet thresholds and are calculated by evaluating the integral:

$$pr(t_l < \ln(cuft_{wj}) < t_h) = \int_{w_l}^{w_h} \int_{t_l - \beta_0 - \beta_1 \ln(w) - \beta_2 \ln(w^2) - \beta_{zj}}^{t_h - \beta_0 - \beta_1 \ln(w) - \beta_2 \ln(w^2) - \beta_{zj}} (f(\varepsilon)\partial e) g(w) \partial w$$

Where:

w - is the piece weight

$f(\varepsilon)$ - is the probability density function of the residual term, ε, assumed to be normally distributed with mean 0 and variance given by the regression standard error squared (σ^2).

$g(w)$ - is the probability density function of weight within a pound increment, assumed to be uniformly distributed across the pound increment

w_h - is the upper limit of the pound increment

w_l - is the lower limit of the pound increment

t_h - is the natural log of the upper limit of cubic feet threshold

t_l - is the natural log of the lower limit of cubic feet threshold

This integral is calculated using the Excel macro “distmac” in the Excel workbook “Cubic Feet Distribution LR” provided in Appendix B.

D. Density Calculations

At the request of witness Scherer density is calculated for 0.34 cubic foot pieces. These pieces are estimated to have an average density of 6.70 pounds per cubic foot.[4] This density (pounds per cubic foot) for Priority Mail parcel-shaped pieces is calculated as total Priority Mail parcel weight divided by total Priority Mail parcel cubic feet for parcels between 0.335 and 0.345 cubic feet. Total Priority Mail parcel weight is calculated as the sum across pound increments of average pound increment weight times the number of pieces between 0.335 and 0.345 cubic feet.

Total Priority Mail Parcel Weight = $\sum_{j=1}^{70} \overline{w}_j \times pcs_j$

Where

$\overline{w}_j$ - is the average weight of parcels in pound increment j

pcs_j - is the total number of 0.34 cubic foot parcels in pound increment j

Total Priority Mail parcel cubic feet is calculated as the sum across pound increments of estimated pound increment average cubic feet times the number of pieces between 0.335 and 0.345 cubic feet.

Total Priority Parcel Cubic Feet = $\sum_{j=1}^{70} \overline{cu}_j \times pcs_j$

Where

$\overline{cu}_j$ - is the average cubic feet of parcels in pound increment j assumed to be 0.34 cubic feet.

pcs_j - is the total number of 0.34 cubic foot Priority Mail parcels in pound increment j

[4] The density calculation, 6.70 pounds per cubic foot, relies on distributional Priority Mail parcels volume data that are not disclosed in the workbook. I am informed by the Postal Service that these data are commercially sensitive because they are similar to the GFY 2003 Priority Mail Billing Determinants which will not be reported to the Postal Rate Commission until 2005. Moreover, the data are more revelatory than the Billing Determinants because they disaggregate Priority Mail volume into parcels and pieces of other shapes (mainly flats).

Appendix A - Listing of Electronic Files

A. Excel workbook "Cubic Feet Distribution LR.xls"
Worksheets:

1. "Sample Data" – Listing of raw sample data
2. "2 > Regression" – Parameter estimates, regression statistics and residuals of regression for Priority Mail parcels greater than 2 pounds.
3. "1&2 Pound" - Parameter estimates, regression statistics and residuals of regression for Priority Mail parcels less than 2 pounds.
4. "Distribution" – Output values of macro "distmac". Displays the distributional estimates of Priority Mail parcels by zone, pound increment, and size.

Macro - "distmac" – Macro used in the development of distributional estimates.

B. Stata "do" file "errortest.do" – Listing of Stata commands used in data analysis. Stata version 6.0.

C. Stata input file "cubic_ft_dist" stata input file. Stata version 6.0.

Appendix B – Excel Macro "distmac"

```
Sub distmac()
'
' distmac Macro
' Macro recorded 2/12/2003 by plloetscher
'
' Keyboard Shortcut: Ctrl+d
'


    Dim wt As Double
    Dim beta0 As Double
    Dim beta1 As Double
    Dim beta2 As Double
    Dim beta10 As Double
    Dim beta11 As Double
    Dim beta12 As Double
    Dim mu As Double
    Dim mu2 As Double
    Dim lim As Double
    Dim limu As Double
    Dim prob As Double
    Dim aprob As Double
    Dim rect As Double
    Dim sig As Double
    Dim sig1 As Double
    Dim value As Double
    Dim mean As Double
    Dim p108 As Double
    Dim mu108 As Double
    Dim zone1(6) As Double
    Dim zone3(6) As Double
    Dim iz As Integer


    Dim i As Integer
    Dim iwt As Integer
    Dim idelta As Integer
    Dim imu As Integer

    Dim ilim As Integer
    Dim wwt As Double
    Dim upper As Integer
```

```
Application.Calculation = xlCalculationManual
Application.ScreenUpdating = False



beta11 = Worksheets("1&2 Pound").Cells(18, 2).value
beta12 = Worksheets("1&2 Pound").Cells(19, 2).value
sig1 = Worksheets("1&2 Pound").Cells(7, 2).value
zone1(1) = 0
For iz = 2 To 6
   zone1(iz) = Worksheets("1&2 Pound").Cells(18 + iz, 2).value
Next

For iz = 1 To 6
    beta10 = Worksheets("1&2 Pound").Cells(17, 2).value
    beta10 = beta10 + zone1(iz)
   For iwt = 1 To 2
      If (iwt = 1) Then
         wt = Worksheets("Distribution").Cells(12 + iwt, 3).value - 0.7
         rect = 1 / 7
         upper = 7

      Else
         wt = Worksheets("Distribution").Cells(12 + iwt, 3).value - 1
         rect = 0.1
         upper = 10
      End If

      For ilim = 1 To 2
         lim = Worksheets("Distribution").Cells(12, 4 + ilim).value
         limu = Worksheets("Distribution").Cells(12, 4 + ilim + 1).value
         aprob = 0

         wwt = wt + 0.05

         For i = 1 To upper
            If ilim < 2 Then
               mu = lim - beta10 - beta11 * Log(wwt) - beta12 * (Log(wwt) ^ 2)
               mu2 = limu - beta10 - beta11 * Log(wwt) - beta12 * (Log(wwt) ^ 2)
               mu108 = Log(8.5944) - beta10 - beta11 * Log(wwt) - beta12 *
(Log(wwt) ^ 2)
               prob = Excel.WorksheetFunction.NormDist(mu2, 0, sig1, True) -
Excel.WorksheetFunction.NormDist(mu, 0, sig1, True)
```

```
                    prob = prob / Excel.WorksheetFunction.NormDist(mu108, 0, sig1,
True)
                    aprob = aprob + rect * prob
                    wwt = wwt + 0.1
                 Else
                    mu = lim - beta10 - beta11 * Log(wwt) - beta12 * (Log(wwt) ^ 2)
                    mu108 = Log(8.5944) - beta10 - beta11 * Log(wwt) - beta12 *
(Log(wwt) ^ 2)
                    prob = Excel.WorksheetFunction.NormDist(mu108, 0, sig1, True)
- Excel.WorksheetFunction.NormDist(mu, 0, sig1, True)
                    prob = prob / Excel.WorksheetFunction.NormDist(mu108, 0, sig1,
True)
                    aprob = aprob + rect * prob
                    wwt = wwt + 0.1
                 End If
              Next
              Worksheets("Distribution").Cells(12 + iwt, 4 + (iz - 1) * 3 + ilim).value =
aprob
           Next
        Next
    Next

    beta0 = Worksheets("2 > Regression").Cells(17, 2).value
    beta1 = Worksheets("2 > Regression").Cells(18, 2).value
    beta2 = Worksheets("2 > Regression").Cells(19, 2).value
    sig = Worksheets("2 > Regression").Cells(7, 2).value
    For iz = 2 To 6
    zone3(iz) = Worksheets("2 > Regression").Cells(18 + iz, 2).value
    Next
    zone1(1) = 0

    For iz = 1 To 6
        beta0 = Worksheets("2 > Regression").Cells(17, 2).value
        beta0 = beta0 + zone3(iz)


        For iwt = 1 To 68
        wt = Worksheets("Distribution").Cells(14 + iwt, 3).value - 1
           For ilim = 1 To 2
              lim = Worksheets("Distribution").Cells(12, 4 + ilim).value
              limu = Worksheets("Distribution").Cells(12, 4 + ilim + 1).value
              aprob = 0

               wwt = wt + 0.05

              For i = 1 To upper
```

```
                If ilim < 2 Then
                   mu = lim - beta0 - beta1 * Log(wwt) - beta2 * (Log(wwt) ^ 2)
                   mu2 = limu - beta0 - beta1 * Log(wwt) - beta2 * (Log(wwt) ^ 2)
                   mu108 = Log(8.5944) - beta0 - beta1 * Log(wwt) - beta2 *
(Log(wwt) ^ 2)
                   prob = Excel.WorksheetFunction.NormDist(mu2, 0, sig, True) -
Excel.WorksheetFunction.NormDist(mu, 0, sig, True)
                   prob = prob / Excel.WorksheetFunction.NormDist(mu108, 0, sig,
True)
                   aprob = aprob + rect * prob
                   wwt = wwt + 0.1
                Else
                   mu = lim - beta0 - beta1 * Log(wwt) - beta2 * (Log(wwt) ^ 2)
                   mu108 = Log(8.5944) - beta0 - beta1 * Log(wwt) - beta2 *
(Log(wwt) ^ 2)
                   prob = Excel.WorksheetFunction.NormDist(mu108, 0, sig, True) -
Excel.WorksheetFunction.NormDist(mu, 0, sig, True)
                   prob = prob / Excel.WorksheetFunction.NormDist(mu108, 0, sig,
True)
                   aprob = aprob + rect * prob
                   wwt = wwt + 0.1
                End If
             Next
             Worksheets("Distribution").Cells(14 + iwt, 4 + ilim + (iz - 1) * 3).value =
aprob
          Next
       Next
    Next


      Application.Calculation = xlCalculationAutomatic
      Application.ScreenUpdating = True



End Sub
```

Appendix C - Stata Commands

```
#delimit ; set more 1 ;
log using errortest, replace ;

insheet using cubic_ft_dist.txt, names ;
gen z2 = izone == 2 ;
gen z3 = izone == 3 ;
d ;
sum volume weight if over_2 == 0, detail ;
sum volume weight if over_2 == 1, detail ;

*** regressions: small flats ;
regress lnvol lnwt lnwt2 z4 z5 z6 z7 z8 if over_2 == 0 ;
scalar ess_s = _result(4) ;

* recover residuals ;
predict e_2under, resid, if e(sample) ;
sum e_2under, detail ;

* test residuals for pieces under two pounds ;
** sktest presents a test for normality based on skewness and another based on
kurtosis and then combines the two tests into an overall test statistic. sktest
implements the test described by D'Agostino, Balanger, and D'Agostino (The
American Statistician 44(4)) with the correction developed by Royston. ;
sktest e_2under ;

*** regressions: large flats ;
regress lnvol lnwt lnwt2 z4 z5 z6 z7 z8 if over_2 == 1 ;
scalar ess_l = _result(4) ;

* recover residuals ;
predict e_over_2, resid, if e(sample) ;
sum e_over_2, detail ;

* test residuals for pieces over two pounds ;
sktest e_over_2 ;

**** Ftest - 2 equivalent ways. 1 significance of difference in coeffs * 2 restricted
and unrestricted sum of squares ;

xi: regress lnvol i.over_2|lnwt i.over_2|lnwt2 i.over_2|z4 i.over_2|z5 i.over_2|z6
i.over_2|z7 i.over_2|z8 over_2 if over_2 == 0 | over_2 == 1 ;

test IoXlnw_1 IoXlnwa1 IoXz4_1 IoXz5_1 IoXz6_1 IoXz7_1 IoXz8_1 over_2 ;
```

```
regress lnvol lnwt lnwt2 z4 z5 z6 z7 z8 if over_2 == 0 | over_2 == 1 ;

scalar ess_r = _result(4) ;
scalar k = _result(3) + 1 ;
scalar N = _result(1) ;
scalar N_2k = (N - 2*k) ;
scalar fstat = ((ess_r - (ess_s + ess_l))/k)/((ess_s + ess_l)/(N_2k)) ;

disp k, N_2k, fstat ;
disp fprob(k,N_2k,fstat) ;

log close ;
```

Appendix D – Stata Output

```
. insheet using cubic_ft_dist_w_flats.txt, names ;
(23 vars, 5368 obs)

. gen z2 = izone == 2 ;

. gen z3 = izone == 3 ;

. d ;

Contains data
  obs:         5,368
 vars:            25
 size:       450,912 (55.9% of memory free)
-------------------------------------------------------------------------------
   1. obsid     int    %8.0g
   2. site      str12  %12s
   3. indicia   str2   %9s
   4. weight    float  %9.0g
   5. length    float  %9.0g
   6. width     float  %9.0g
   7. height    float  %9.0g
   8. girth     float  %9.0g
   9. volume    float  %9.0g
  10. volume2   float  %9.0g
  11. izone     byte   %8.0g
  12. wti       byte   %8.0g
  13. len_grth  float  %9.0g
  14. density   float  %9.0g
  15. lnvol     float  %9.0g
  16. lnwt      float  %9.0g
  17. lnwt2     float  %9.0g
  18. z4        byte   %8.0g
  19. z5        byte   %8.0g
  20. z6        byte   %8.0g
  21. z7        byte   %8.0g
  22. z8        byte   %8.0g
  23. over_2    byte   %8.0g
  24. z2        float  %9.0g
  25. z3        float  %9.0g
-------------------------------------------------------------------------------
Sorted by:
     Note:  dataset has changed since last saved
```

```
. sum volume weight if over_2 == 0, detail ;

                          volume
-------------------------------------------------------------
      Percentiles      Smallest
 1%        .031           .014
 5%        .047           .016
10%        .053           .019       Obs                2880
25%        .083            .02       Sum of Wgt.        2880

50%        .156                      Mean           .1926167
                        Largest       Std. Dev.      .1701152
75%        .247           1.638
90%        .344           1.989       Variance       .0289392
95%        .465           2.144       Skewness       3.934503
99%        .757           2.155       Kurtosis       31.75183

                          weight
-------------------------------------------------------------
      Percentiles      Smallest
 1%         .21             .1
 5%        .355             .1
10%         .45            .12       Obs                2880
25%         .75            .14       Sum of Wgt.        2880

50%         1.1                      Mean           1.119177
                        Largest       Std. Dev.      .4748549
75%         1.5              2
90%        1.78              2       Variance       .2254872
95%         1.9              2       Skewness        .02008
99%           2              2       Kurtosis       2.012347

. sum volume weight if over_2 == 1, detail ;

                          volume
-------------------------------------------------------------
      Percentiles      Smallest
 1%        .069           .015
 5%        .116           .035
10%        .173           .037       Obs                2488
25%       .3075           .039       Sum of Wgt.        2488

50%       .6535                      Mean           .9182886
                        Largest       Std. Dev.      .9144502
75%       1.1965          5.919
90%        1.997           5.99       Variance       .8362192
```

```
95%       2.781        6.426      Skewness      2.337783
99%       4.615        8.398      Kurtosis      10.64185

                      weight
-------------------------------------------------------------
     Percentiles      Smallest
 1%        2.05          2.01
 5%        2.15          2.01
10%         2.3          2.02      Obs               2488
25%         2.8          2.02      Sum of Wgt.       2488

50%         4.2                    Mean          7.116884
                     Largest       Std. Dev.     7.371555
75%        8.15           58
90%     15.3125          60.5      Variance      54.33983
95%       23.02          63.4      Skewness         2.985
99%        37.9          67.8      Kurtosis      14.88144

. *** regressions: small flats ;
. regress lnvol lnwt lnwt2 z4 z5 z6 z7 z8 if over_2 == 0 ;

  Source |       SS       df       MS              Number of obs =    2880
---------+------------------------------           F(  7,  2872) =  127.37
   Model |  370.209537     7  52.8870767           Prob > F      =  0.0000
Residual |  1192.48535  2872  .415210777           R-squared     =  0.2369
---------+------------------------------           Adj R-squared =  0.2350
   Total |  1562.69489  2879  .542790861           Root MSE      =  .64437

------------------------------------------------------------------------------
   lnvol |      Coef.   Std. Err.      t    P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lnwt |   .7359721   .0287456    25.603   0.000        .679608    .7923363
   lnwt2 |   .1016238   .0332043     3.061   0.002        .036517    .1667305
      z4 |   .1034841   .0380711     2.718   0.007       .0288346    .1781337
      z5 |   .0645834   .0346439     1.864   0.062       -.003346    .1325127
      z6 |   .0499192   .0378161     1.320   0.187      -.0242302    .1240686
      z7 |   .0819458   .0505758     1.620   0.105      -.0172228    .1811144
      z8 |   .0223322   .0400009     0.558   0.577      -.0561012    .1007655
   _cons |  -1.990562    .023579   -84.421   0.000      -2.036796   -1.944329
------------------------------------------------------------------------------

. scalar ess_s = _result(4) ;

. * recover residuals ;
. predict e_2under, resid, if e(sample) ;
(2488 missing values generated)
```

```
. sum e_2under, detail ;

                      Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -1.528214      -2.295791
 5%    -1.067262      -2.075853
10%     -.811833      -1.846391       Obs               2880
25%    -.4561105      -1.811292       Sum of Wgt.       2880

50%     .0105006                      Mean          3.13e-10
                      Largest       Std. Dev.      .6435847
75%     .4480011       2.319657
90%     .7960996       2.330959       Variance       .4142012
95%     1.036086       2.449442       Skewness       .0134676
99%     1.448895       2.522949       Kurtosis        3.10552

. * test residuals for pieces under two pounds ;
. ** sktest presents a test for normality based on skewness and another based o
> n kurtosis and then combines the two tests into an overall test statistic. sk
> test implements the test described by D'Agostino, Balanger, and D'Agostino (T
> he American Statistician 44(4)) with the correction developed by Royston. ;
. sktest e_2under ;

               Skewness/Kurtosis tests for Normality
                                       ------- joint -------
 Variable |  Pr(Skewness)   Pr(Kurtosis)  adj chi-sq(2)  Pr(chi-sq)
----------+-------------------------------------------------------
 e_2under |      0.767         0.242           1.46       0.4822

. *** regressions: large flats ;
. regress lnvol lnwt lnwt2 z4 z5 z6 z7 z8 if over_2 == 1 ;

  Source |       SS       df       MS              Number of obs =    2488
---------+------------------------------           F(  7,  2480) =  299.93
   Model |  1030.54435     7  147.220622           Prob > F      =  0.0000
Residual |  1217.30911  2480  .490850446           R-squared     =  0.4585
---------+------------------------------           Adj R-squared =  0.4569
   Total |  2247.85346  2487  .903841359           Root MSE      =  .70061

------------------------------------------------------------------------------
   lnvol |      Coef.   Std. Err.      t    P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lnwt |   1.537416   .0923222    16.653   0.000       1.356379    1.718452
   lnwt2 |  -.1781381   .0230298    -7.735   0.000      -.2232976   -.1329785
```

```
      z4 |   .2150284   .0443729      4.846   0.000       .1280167    .3020402
      z5 |   .1947027     .04131      4.713   0.000       .113697    .2757083
      z6 |   .2115059   .0470042      4.500   0.000       .1193343    .3036776
      z7 |   .2197955    .057442      3.826   0.000       .1071562    .3324348
      z8 |   .2675006   .0441529      6.059   0.000       .1809203    .3540809
   _cons |  -2.609846   .0828441    -31.503   0.000      -2.772296   -2.447395
------------------------------------------------------------------------------

. scalar ess_l = _result(4) ;

. * recover residuals ;
. predict e_over_2, resid, if e(sample) ;
(2880 missing values generated)

. sum e_over_2, detail ;

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -1.633945      -2.924945
 5%    -1.208992      -2.286752
10%    -.9440451      -2.242628       Obs               2488
25%    -.4692211      -2.097646       Sum of Wgt.       2488

50%     .0416828                      Mean          -2.04e-10
                        Largest       Std. Dev.      .6996205
75%     .4782287       2.010518
90%     .8772044       2.049768       Variance       .4894689
95%     1.105505       2.270331       Skewness      -.1107105
99%     1.611614       2.455381       Kurtosis       2.970103

. * test residuals for pieces over two pounds ;
. sktest e_over_2 ;

                Skewness/Kurtosis tests for Normality
                                          ------- joint -------
 Variable |  Pr(Skewness)   Pr(Kurtosis)  adj chi-sq(2)  Pr(chi-sq)
----------+-------------------------------------------------------
 e_over_2 |      0.024         0.813           5.12       0.0774

. **** Ftest - 2 equivalent ways. 1 significance of difference in coeffs * 2 re
> stricted and unrestricted sum of squares ;
. xi: regress lnvol i.over_2|lnwt i.over_2|lnwt2 i.over_2|z4 i.over_2|z5 i.over
> _2|z6 i.over_2|z7 i.over_2|z8 over_2 if over_2 == 0 | over_2 == 1 ;
i.over_2             Iover__0-1   (naturally coded; Iover__0 omitted)
i.over_2|lnwt        IoXlnw_#     (coded as above)
```

```
i.over_2|lnwt2        IoXlnwa#     (coded as above)
i.over_2|z4           IoXz4_#      (coded as above)
i.over_2|z5           IoXz5_#      (coded as above)
i.over_2|z6           IoXz6_#      (coded as above)
i.over_2|z7           IoXz7_#      (coded as above)
i.over_2|z8           IoXz8_#      (coded as above)

  Source |       SS       df       MS              Number of obs =    5368
---------+------------------------------           F( 15,  5352) =  602.17
   Model |  4066.97751    15  271.131834           Prob > F      =  0.0000
Residual |  2409.79446  5352  .450260549           R-squared     =  0.6279
---------+------------------------------           Adj R-squared =  0.6269
   Total |  6476.77197  5367  1.20677696           Root MSE      =  .67101

------------------------------------------------------------------------------
   lnvol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lnwt |   .7359721   .0299343    24.586   0.000     .6772887    .7946556
IoXlnw_1 |   .8014434   .0933522     8.585   0.000     .6184351    .9844517
   lnwt2 |   .1016238   .0345774     2.939   0.003     .0338379    .1694096
IoXlnwa1 |  -.2797618   .0410135    -6.821   0.000     -.360165   -.1993586
      z4 |   .1034841   .0396455     2.610   0.009     .0257629    .1812054
 IoXz4_1 |   .1115443   .0581197     1.919   0.055     -.002394    .2254826
      z5 |   .0645834   .0360765     1.790   0.073    -.0061412     .135308
 IoXz5_1 |   .1301193   .0535436     2.430   0.015     .0251521    .2350865
      z6 |   .0499192   .0393799     1.268   0.205    -.0272813    .1271198
 IoXz6_1 |   .1615867    .059812     2.702   0.007     .0443309    .2788425
      z7 |   .0819458   .0526673     1.556   0.120    -.0213035     .185195
 IoXz7_1 |   .1378498   .0761615     1.810   0.070    -.0114578    .2871573
      z8 |   .0223322    .041655     0.536   0.592    -.0593286    .1039929
 IoXz8_1 |   .2451684   .0593583     4.130   0.000     .128802    .3615349
  over_2 |  -.6192833   .0830573    -7.456   0.000    -.7821095   -.4564571
   _cons |  -1.990562   .0245541   -81.068   0.000    -2.038698   -1.942426
------------------------------------------------------------------------------

. test IoXlnw_1 IoXlnwa1 IoXz4_1 IoXz5_1 IoXz6_1 IoXz7_1 IoXz8_1 over_2 ;

 ( 1)  IoXlnw_1 = 0.0
 ( 2)  IoXlnwa1 = 0.0
 ( 3)  IoXz4_1 = 0.0
 ( 4)  IoXz5_1 = 0.0
 ( 5)  IoXz6_1 = 0.0
 ( 6)  IoXz7_1 = 0.0
 ( 7)  IoXz8_1 = 0.0
 ( 8)  over_2 = 0.0
```

```
      F(  8,  5352) =   16.75
          Prob > F =    0.0000


. regress lnvol lnwt lnwt2 z4 z5 z6 z7 z8 if over_2 == 0 | over_2 == 1 ;

  Source |       SS       df       MS              Number of obs =    5368
---------+------------------------------           F(  7,  5360) = 1242.03
   Model |  4006.65603     7  572.379432           Prob > F      =  0.0000
Residual |  2470.11594  5360  .460842527           R-squared     =  0.6186
---------+------------------------------           Adj R-squared =  0.6181
   Total |  6476.77197  5367  1.20677696           Root MSE      =  .67885

------------------------------------------------------------------------------
   lnvol |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lnwt |   .8024534   .0155126     51.729   0.000       .7720424    .8328643
   lnwt2 |   .0146059   .0063838      2.288   0.022       .0020911    .0271208
      z4 |   .1583457   .0292939      5.405   0.000       .1009178    .2157737
      z5 |   .1290061   .0269401      4.789   0.000       .0761925    .1818196
      z6 |   .1254682   .0299518      4.189   0.000       .0667505    .1841858
      z7 |   .1562119   .0384382      4.064   0.000       .0808574    .2315665
      z8 |   .1525932   .0299306      5.098   0.000       .0939169    .2112694
   _cons |  -1.997857   .0180387   -110.754   0.000      -2.03322   -1.962493
------------------------------------------------------------------------------

. scalar ess_r = _result(4) ;

. scalar k = _result(3) + 1 ;

. scalar N = _result(1) ;

. scalar N_2k = (N - 2*k) ;

. scalar fstat = ((ess_r - (ess_s + ess_l))/k)/((ess_s + ess_l)/(N_2k)) ;

. disp k, N_2k, fstat ;
8 5352 16.746272

. disp fprob(k,N_2k,fstat) ;
9.016e-25

. log close ;
```