Heteroscedasticity

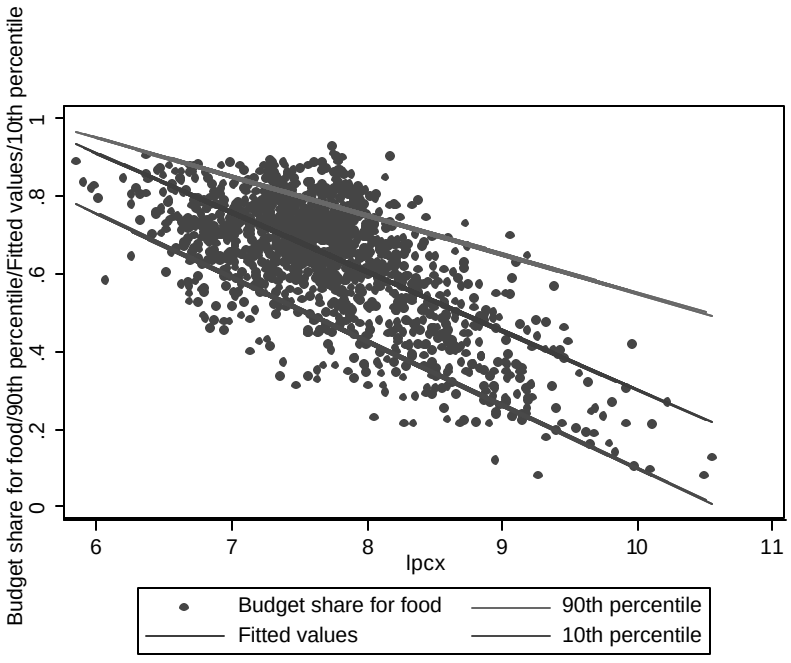
Chapter 8

Heteroscedasticity: Definition

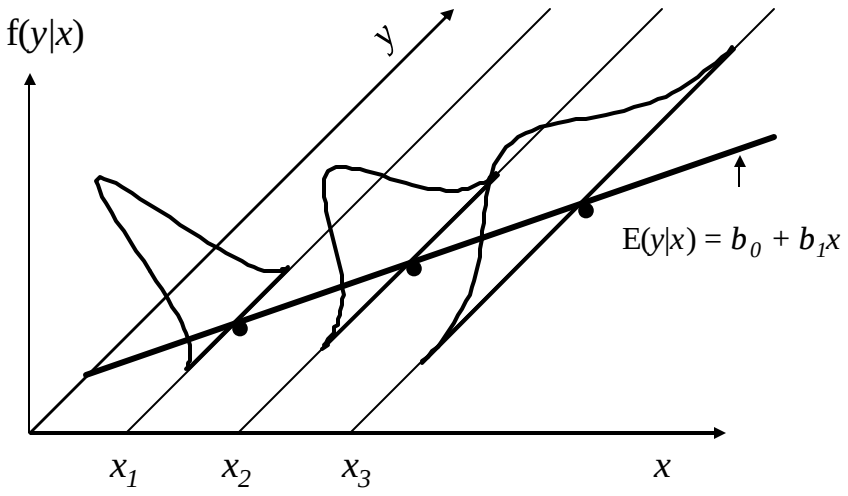
- Heteroskedasticity occurs when the variance of the error terms differ across observations
- The model now becomes
  - $Y_i = b_0 + b_1 X_i + \varepsilon_i$,
    - Variance of the errors is no longer assumed to be a constant
      - $\text{Var}(\varepsilon_i) = \sigma^2_i$
- s_i^2 may vary with x_i, with y_i or with something left out of the model (e.g. a quadratic term)
  - s_i^2 may vary for each observation, or for groups of observations, especially in clustered samples

Causes

- Why might we observe heteroskedasticity?
  - Suppose 100 students enroll in a typing class—some of which have typing experience and some of which do not
    - After the first class there would be a great deal of dispersion in the number of typing mistakes
    - After the final class the dispersion would be smaller
      - The error variance is nonconstant—it falls as time increases
  - If we gathered data on the income and food expenditures of a large number of families, those with high levels of income may have a greater dispersion in food expenditures than those at lower income levels
    - With high incomes, can afford to eat whatever individual tastes dictate
    - With low incomes, everyone forced to eat the cheapest foods
- Heteroskedasticity arises most often with cross-sectional data
  - But finance data often has time-varying volatility (ARCH)



Example of Heteroskedasticity



Why Worry About Heteroskedasticity?

- OLS is still unbiased and consistent, even if we do not assume homoskedasticity
- The standard errors of the estimates <u>are</u> biased if we have heteroskedasticity
  - If the standard errors are biased, we can not use the usual *t* statistics or *F* statistics or *LM* statistics for drawing inferences
  - OLS, even if standard errors could be correctly measured, is no longer efficient

Robust Standard Errors

- The most common response to the (potential) presence of heteroscedasticity of an unknown form is to use a heteroscedastically-robust estimator for the covariance matrix of the regression parameters
- Hence, any subsequent inferences and hypothesis tests should also be robust to heteroscedasticity (including Wald tests, which can be used in this context because the robust estimator is a less restricted model than the OLS estimator of the covariance)
- These estimators go by various names (Huber, Eicker, "sandwich" – because of the way the formula looks) but were introduced to econometrics by White (1980)

Robust Standard Errors

- The covariance matrix with heteroscedasticity is: $V(\hat{b}) = (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'(s^2\Omega)\mathbf{X}](\mathbf{X}'\mathbf{X})^{-1}$
- so it looks like we need an estimate of $s^2\Omega$ but we don't know what Ω is. But a consistent estimator of $\frac{1}{n}s^2\mathbf{X}'\Omega\mathbf{X}$ is given by $\frac{1}{n}s^2\sum_{i=1}^{n} e_i^2 x_i x_i'$ where e_i is the *ith* least squares residual

$$V(\hat{b}) = n\left(X'X\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i'\right)\left(X'X\right)^{-1}$$

$$\textit{compared to } s^2\left(X'X\right)^{-1} \text{ with homoscedasticity}$$

Robust Standard Errors

- Important to remember that these robust standard errors only have asymptotic justification – with small sample sizes *t* statistics formed with robust standard errors will not have a distribution close to the *t*, and inferences will not be correct
- In Stata, robust standard errors are easily obtained using the robust option e.g *reg y x, robust*
  - Stata automatically gives robust standard errors if using the survey estimators (or using *pweights* – which are sample survey weights)

An Aside: Serial Correlation-Robust Standard Errors

- In time series, with autocorrelated residuals, if you don't want to use approaches like Cochrane-Orcutt transformation that rely on estimating ρ and quasi-differencing
  - can calculate serial correlation-robust standard errors, along the same lines as heteroskedasticity robust standard errors
  - One difference, you have to choose the autocorrelation lag length, whereas don't have to choose anything for the heteroskedastic robust standard error
- "Newey-West std errors" in Stata

Testing for Heteroskedasticity

- Essentially want to test H_0: Var($e|x_1, x_2,..., x_k$) = s^2, which is equivalent to H_0: E($e^2|x_1, x_2,..., x_k$) = E(e^2) = s^2
- If assume the relationship between e^2 and x_j will be linear, can test as a linear restriction
- So, for $e^2 = d_0 + d_1x_1 +...+ d_k x_k + v$) this means testing H_0: $d_1 = d_2 = ... = d_k = 0$

The Breusch-Pagan Test

- Don't observe the error, but can estimate it with the residuals from the OLS regression
- After regressing the residuals squared on all of the *x*'s, can use the R^2 to form a *Lagrange Multiplier* (LM) test
- The *LM* statistic is $LM = nR^2$, which is distributed χ^2_k

Example of Breusch-Pagan Test

```
. qui reg  ltobacco lpcx lmales lfemales urban
. predict uhat, resid
. gen uhat2=uhat^2
. reg uhat2  lpcx urban lmales lfemales
      Source |       SS       df       MS              Number of obs =     351
-------------+------------------------------           F(  4,   346) =    0.41
       Model |  2.89034004     4   .72258501           Prob > F      =  0.8045
    Residual |  616.134701   346  1.78073613           R-squared     =  0.0047
-------------+------------------------------           Adj R-squared = -0.0068
       Total |  619.025041   350  1.76864297           Root MSE      =  1.3344
------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lpcx |  -.0583671   .1015975    -0.57   0.566    -.2581935    .1414593
       urban |  -.1430389   .1600626    -0.89   0.372    -.4578571    .1717792
      lmales |  -.0074283   .1341613    -0.06   0.956    -.2713026     .256446
    lfemales |    .042792   .1259268     0.34   0.734    -.2048863    .2904704
       _cons |   1.44119   .8151355     1.77   0.078    -.1620543    3.044434
------------------------------------------------------------------------------
. scalar bptest=e(N)*e(r2)
. scalar list bptest
    bptest =  1.6388826
```

Compare this value to the critical value for chi-sq with 4 degrees of freedom

Weighted Least Squares

- If OLS estimators are not BLUE in the presence of heteroskedasticity
  - What are the best estimators?
- Can weight the observations so that more weight is put on observations associated with levels of X having a smaller error variance
- Transform the model so that the errors no longer exhibit heteroskedasticity
- The basic model with heteroskedasticity is
  - $Y_i = b_0 + b_1 X_i + \varepsilon_i$
  - $\text{Var}(\varepsilon_i) = \sigma_i^2$

Weighted Least Squares (WLS)

- Dividing each observation by the associated standard deviation of the error transforms the model

$$\frac{Y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{X_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}$$

  - OLS estimators for this model are BLUE in the presence of heteroskedasticity
  - WLS is an example of a Generalized Least Squares (GLS) estimator – more efficient than OLS on the unweighted data

Weighted Least Squares

- The error term in the transformed model is no longer heteroskedastic

$$\text{var}\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{\text{var}(\varepsilon_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

- The transformed model is called a weighted least squares
  - Each observation is now weighted by the inverse of the standard deviation of the error
- Major difficulty in estimating weighted least squares
  - Don't observe σ_i so can't get the exact weights needed
  - A regression using estimated weights, is an example of "Feasible GLS"

Weighted Least Squares

- Methods of estimating σ_i include
  - Assume σ_i^2 is a function of an explanatory variable
    - Divide each observation by the value of X_{ji}
      - Must be careful when interpreting the resulting coefficients
        - E.g. the former coefficient on the X_{ji} term appears as the constant

$$if \quad \mathrm{var}(e_i) = g^2 X_{1i}^2 \quad WLS\ is:$$

$$\frac{Y_i}{X_{1i}} = \frac{b_0}{X_{1i}} + b_1 + b_2 \frac{X_{2i}}{X_{1i}} + \frac{e_i}{X_{1i}}$$

  - Grouped data (e.g. State-level averages)
    - From CLT, know that the variance of means of more populated States will be smaller than variance for less populated States
    - So WLS, with square root of State population as weight

Survey data

- Samples are already weighted, so WLS used to give estimates for the population, regardless of concerns about heteroscedasticity
- Variance is likely to be non-constant across clusters and there also may be lack of independence within clusters (spatial autocorrelation)
- The variance estimators used by Stata's svy commands adjust for both the heteroscedasticity and within-cluster autocorrelation
  - Special case of the Huber-White 'robust' variance estimation