asymptotic (large sample) χ^2 null distributions, with the number of restrictions imposed by the null hypothesis as degrees of freedom. The likelihood-ratio and Wald tests (and the less commonly used score or Lagrange multiplier test) are asymptotically equivalent to each other but may produce different conclusions in small samples.

Hypothesis tests for individual regression coefficients

The most commonly used hypothesis test concerns an individual regression parameter, say β_2, with null hypothesis

$$H_0\text{: } \beta_2 = 0$$

versus the two-sided alternative $H_a\text{: } \beta_2 \neq 0$.

The Wald statistic for testing the null hypothesis is

$$w = \left(\frac{\widehat{\beta}_2}{\widehat{\mathrm{SE}}(\widehat{\beta}_2)} \right)^2$$

with a χ^2 null distribution with 1 degree of freedom since the null hypothesis imposes one restriction. In practice, the test statistic

$$z = \frac{\widehat{\beta}_2}{\widehat{\mathrm{SE}}(\widehat{\beta}_2)}$$

is usually used, which has a standard normal null distribution (because its square has a χ^2 distribution with 1 degree of freedom).

The z statistic is reported as z in the Stata output. For instance, in the output from **xtmixed** on page 100, the z statistic for the regression parameter of smoking is -12.02, which gives a two-sided p-value of less than 0.001.

The likelihood-ratio test statistic is less commonly used for testing individual regression parameters but takes the same form as explained below for several coefficients.

Joint hypothesis tests for several regression coefficients

Consider now the null hypothesis that the regression coefficients of two covariates x_{2ij} and x_{3ij} are both zero,

$$H_0\text{: } \beta_2 = \beta_3 = 0$$

versus the alternative hypothesis that at least one of the parameters is nonzero. For example, for the smoking and birthweight application, we may want to test the null hypothesis that the quality of prenatal care (as measured by the Kessner index) makes no difference to birthweight (controlling for the other covariates), where the Kessner index is represented by two dummy variables, kessner2 and kessner3.

Let $\widehat{\beta}_2$ and $\widehat{\beta}_3$ be maximum likelihood estimates from the model including the covariates. The Wald and likelihood-ratio tests have asymptotic χ^2 null distributions with 2 degrees of freedom since the null imposes two restrictions.

The Wald statistic can be expressed as

$$
\begin{aligned}
w &= (\widehat{\beta}_2, \widehat{\beta}_3) \left\{ \begin{array}{cc} \widehat{\mathrm{SE}}(\widehat{\beta}_2)^2 & \widehat{\mathrm{Cor}}(\widehat{\beta}_2, \widehat{\beta}_3)\, \widehat{\mathrm{SE}}(\widehat{\beta}_2)\, \widehat{\mathrm{SE}}(\widehat{\beta}_3) \\ \widehat{\mathrm{Cor}}(\widehat{\beta}_2, \widehat{\beta}_3)\, \widehat{\mathrm{SE}}(\widehat{\beta}_2)\, \widehat{\mathrm{SE}}(\widehat{\beta}_3) & \widehat{\mathrm{SE}}(\widehat{\beta}_3)^2 \end{array} \right\}^{-1}_{(\widehat{\beta}_2, \widehat{\beta}_3)'} \\
&= \frac{1}{1 - \widehat{\mathrm{Cor}}(\widehat{\beta}_2, \widehat{\beta}_3)^2} \left\{ \left(\frac{\widehat{\beta}_2}{\widehat{\mathrm{SE}}(\widehat{\beta}_2)} \right)^2 + \left(\frac{\widehat{\beta}_3}{\widehat{\mathrm{SE}}(\widehat{\beta}_3)} \right)^2 - 2\, \widehat{\mathrm{Cor}}(\widehat{\beta}_2, \widehat{\beta}_3)\, \frac{\widehat{\beta}_2}{\widehat{\mathrm{SE}}(\widehat{\beta}_2)}\, \frac{\widehat{\beta}_3}{\widehat{\mathrm{SE}}(\widehat{\beta}_3)} \right\}
\end{aligned}
$$

The Wald test for the null hypothesis that the Kessner index does not have an effect, i.e., that the coefficients of `kessner2` and `kessner3` are both zero, $\beta_{10} = \beta_{11} = 0$, can be performed by using the `testparm` command:

```
. quietly xtreg birwt smoke male mage hsgrad somecoll collgrad married
> black kessner2 kessner3 novisit pretri2 pretri3, i(momid) mle
. testparm kessner2 kessner3
( 1)  [birwt]kessner2 = 0
( 2)  [birwt]kessner3 = 0
           chi2(  2) =   26.94
         Prob > chi2 =    0.0000
```

We can reject the null hypothesis at the 5% level with $w = 26.94$, df $= 2$, $p < 0.001$. A more robust version of the test is obtained by specifying robust standard errors in the estimation command using the `vce(robust)` option for `xtreg` (only with the `re` option) or the `robust` option for `gllamm` followed by the `testparm` command.

We can also test the simultaneous hypothesis that three or more regression coefficients are all zero, but the expression for the Wald statistic becomes convoluted unless matrix expressions are used.

The analogous likelihood-ratio test statistic is

$$L = 2(l_1 - l_0)$$

where l_1 and l_0 are now the maximized log likelihoods for the models including and excluding both `kessner2` and `kessner3`, respectively. A likelihood-ratio test for the null hypothesis that the Kessner index does not have an effect (given the other covariates) can be performed by using the `lrtest` command:

```
. estimates store full
. quietly xtreg birwt smoke male mage hsgrad somecoll collgrad married black
> novisit pretri2 pretri3, i(momid) mle
. lrtest full .
Likelihood-ratio test
(Assumption: . nested in full)                 LR chi2(2)  =     26.90
                                               Prob > chi2 =    0.0000
```

Note that likelihood-ratio tests for regression coefficients cannot be based on "log likelihoods" from restricted maximum likelihood (REML) estimation. When using `xtmixed`, the `mle` option must therefore be specified.

Sometimes it is required to test hypotheses regarding linear combinations of coefficients as demonstrated in section 1.8. In section 3.7.4, we will encounter a special case of this when testing the null hypothesis that two regression coefficients are equal, or in other words that the difference between the coefficients is 0, a simple example of a *contrast*. Wald tests of such hypotheses can be performed in Stata using the `lincom` command.

3.6.2 Predicted means and confidence intervals

We can use the `adjust` command to obtain predicted means for mothers and pregnancies with particular covariate values. This is useful for interpreting the results from regression modeling.

For example, if we want to interpret the effects of smoking and education on birthweight, we can set all other covariates to some meaningful values and produce a table of predicted means by smoking and education. It is useful to first define a categorical variable for level of education and give the categories value labels

```
. generate education = hsgrad*1 + somecoll*2 + collgrad*3
. label define ed 0 "no HS grad" 1 "HS grad" 2 "some Coll" 3 "Coll grad"
. label values education ed
```

After retrieving the estimates of the full model, we then use the `adjust` command, setting all dummy variables either to 0 or 1 and all continuous values to their mean (the latter is accomplished by simply listing the variables). We do not specify values for `smoking` and the education dummy variables, but specify `smoke` and `education` in the `by()` option instead:

(*Continued on next page*)

Cut

The estimated coefficient of smoke is close to that in the previous model where all cluster means were included.

A great advantage of clustered data is that we can investigate and address the endogeneity problem at the cluster level for unit-level covariates (correlation between ζ_j and x_{ij}). However, we cannot handle cluster-level endogeneity for cluster-level covariates (correlation between ζ_j and x_j), which is tackled by an approach suggested by Hausman and Taylor (1981) and implemented in the `xthtaylor` command. However, it is not straightforward to check for endogeneity at the unit level, i.e., to check if ϵ_{ij} is correlated with either cluster-level or unit-level covariates. To correct for this kind of endogeneity, external instrumental variables are usually required.

3.7.5 Hausman endogeneity test

The Hausman test (Hausman 1978), more aptly called the Durbin–Wu–Hausman test, can be used to compare two alternative estimators of $\boldsymbol{\beta}$, both of which are consistent (the estimates approach the true parameter values as the sample size tends to infinity) if the model is true. In its simplest form, one of the estimators is efficient (is the most precise estimator as the sample size tends to infinity) if the model is true, but inconsistent when the model is misspecified. The other estimator is consistent also under misspecification but not efficient when the model is true.

For instance, both the fixed-effects estimator $\widehat{\boldsymbol{\beta}}^W$ and the generalized least-squares estimator $\widehat{\boldsymbol{\beta}}^R$ are consistent, whereas only $\widehat{\boldsymbol{\beta}}^R$ is efficient, if the random-intercept model is correctly specified. However, for certain model violations, such as correlations between the random intercept and any of the covariates, $\widehat{\boldsymbol{\beta}}^R$ becomes inconsistent, whereas $\widehat{\boldsymbol{\beta}}^W$ remains consistent.

Consider first the simple case of a model with a single covariate x_{ij} that varies both between and within clusters. The Hausman test statistic for endogeneity then takes the form

$$h = \frac{(\widehat{\beta}^W - \widehat{\beta}^R)^2}{\widehat{\mathrm{SE}}(\widehat{\beta}^W)^2 - \widehat{\mathrm{SE}}(\widehat{\beta}^R)^2} \tag{3.13}$$

which has a χ^2 null distribution with 1 degree of freedom. The denominator of the test statistic would usually take the form $\widehat{\mathrm{SE}}(\widehat{\beta}^W)^2 + \widehat{\mathrm{SE}}(\widehat{\beta}^R)^2 - 2\,\widehat{\mathrm{Cov}}(\widehat{\beta}^W, \widehat{\beta}^R)$, where the covariance between the within and between estimators would be hard to obtain. However, it can be shown that the denominator simplifies to the one in (3.13) because the random-effects estimator is efficient when the random-intercept model is true.

Consider now the case where there are several covariates that all vary both between and within clusters. Let the fixed effects and GLS estimates be denoted $\widehat{\boldsymbol{\beta}}^W$ and $\widehat{\boldsymbol{\beta}}^R$, respectively, and let the corresponding estimated covariance matrices be denoted $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}^W)$ and $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}^R)$. The Hausman test statistic then takes the form

$$h = (\widehat{\beta}^W - \widehat{\beta}^R)\left\{\widehat{\mathrm{Cov}}(\widehat{\beta}^W) - \widehat{\mathrm{Cov}}(\widehat{\beta}^R)\right\}^{-1}(\widehat{\beta}^W - \widehat{\beta}^R)'$$

The h statistic has a χ^2 null distribution with degrees of freedom given as the number of overlapping estimated regression coefficients from the two approaches, that is the number of covariates with both between- and within-cluster variation.

We can use the `hausman` command to perform the Hausman test in Stata, following estimation of $\widehat{\beta}^W$ using `xtreg` with the `fe` option and $\widehat{\beta}^R$ using `xtreg` with the `re` option:

```
. quietly xtreg birwt smoke male mage hsgrad somecoll collgrad married
> black kessner2 kessner3 novisit pretri2 pretri3, i(momid) fe
. estimates store fixed
. quietly xtreg birwt smoke male mage hsgrad somecoll collgrad married
> black kessner2 kessner3 novisit pretri2 pretri3, i(momid) re
. estimates store random
. hausman fixed random
```

	Coefficients (b) fixed	(B) random	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
smoke	-104.5494	-217.7488	113.1995	22.71343
male	125.6355	120.9874	4.648084	5.297981
mage	23.15832	8.137158	15.02116	2.687211
kessner2	-91.49483	-92.89604	1.401212	12.44845
kessner3	-128.091	-150.6366	22.54563	24.87574
novisit	-4.805898	-29.9223	25.11641	41.66561
pretri2	81.29039	92.73087	-11.44048	13.94097
pretri3	153.059	178.4334	-25.37443	30.76114

```
                b = consistent under Ho and Ha; obtained from xtreg
    B = inconsistent under Ha, efficient under Ho; obtained from xtreg
Test:  Ho:  difference in coefficients not systematic
              chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =       60.07
            Prob>chi2 =      0.0000
```

There is strong evidence for model misspecification since the Hausman test statistic is 60.07 with df = 8. The Hausman statistic is almost identical to the Wald statistic for the joint null hypothesis that all regression coefficients of the cluster means are 0 shown in the previous section. Indeed, these tests become equivalent in large samples.

A significant Hausman test is often taken to mean that the random-intercept model should be abandoned in favor of a fixed-effects model that only utilizes within information. However, this would preclude estimation of the coefficients of covariates that vary only between clusters (admittedly, these estimates must be interpreted with caution since they may be inconsistent unless the covariates are exogenous). Moreover, if there are covariates having the same within- and between-effects, we obtain more precise estimates of these coefficients by exploiting both within- and between-cluster information. We therefore prefer using a random-intercept model where cluster means are included for some covariates, as demonstrated for smoking and mother's age in the previous sec-

tion. Indeed, when mn_smok and mn_mage are included in the random-intercept model, the Hausman test is no longer significant at the 5% level.

3.8 Fixed versus random effects revisited

In section 2.4, we discussed whether the effects of clusters should be treated as random or fixed. We argued that this depends on whether inferences are for the population of clusters or only for the clusters included in the sample. In table 3.3, we consider these as the main questions and then ask questions related to each main question. The answers to these questions delineate the main differences between fixed-effects and random-effects approaches.

Table 3.3: Overview of distinguishing features of fixed- and random-effects approaches

Questions:	Answers:	
	Fixed effects	**Random effects**
Inference for population of clusters?	No	Yes
Minimum number of clusters required?	Any number	For estimating ψ, at least 10 or 20
What assumptions are required?	None for distribution of fixed intercepts α_j	Random intercepts ζ_j normal, constant variance ψ, exogenous covariates, etc.
Can estimate effects of cluster-level covariates?	No	Yes
Inference for clusters in particular sample?	Yes	No, not for βs, Yes, for ζ_j by using EB
Minimum cluster size required?	Any sizes if many ≥ 2, but large for est. α_j	Any sizes if many ≥ 2
Is the model parsimonious?	No, J parameters α_j but can eliminate α_j	Yes, one variance parameter ψ for all J clusters
Can estimate within-cluster effects of covariates?	Yes	Yes, by including cluster means

Unlike the fixed-effects model, the random-effects model can be used to make inferences regarding the population of clusters, but at the cost of requiring many clusters if inference regarding ψ is desired and making additional assumptions regarding the random-intercept distribution. The additional assumptions include the specification of a normal distribution for ζ_j, a constant variance ψ for the ζ_j, and exogeneity of the

observed covariates $\mathbf{x}_{ij}$ with respect to ζ_j. An advantage of the random-effects model is that it can be used to estimate the effects of between-cluster covariates, in contrast to the fixed-effects model.

While the fixed-effects model is designed for making inferences regarding the clusters in the sample, the random-effects model can also to some extent be used for this purpose by predicting the ζ_j using empirical Bayes (EB). However, inferences regarding the regression coefficients from random-effects models, such as estimated standard errors, are for the population of clusters as discussed in section 2.7.2. The fixed-effects approach requires large cluster sizes if we want to estimate the intercepts α_j and is much less parsimonious than the random-intercept model because it includes one parameter α_j for each cluster, whereas the random-intercept model has only one parameter ψ for the variance of the random intercepts ζ_j. Eliminating the α_j by mean centering as shown in section 3.7.2 simplifies the estimation problem but does not make the estimates of the remaining parameters any more efficient. Unlike the random-effects approach, the fixed-effects approach controls for clusters, providing estimates of within-cluster effects of covariates. The random-effects model can provide estimates of within-cluster effects only with extra effort, namely, by including cluster means of those covariates for which the between effect differs from the within effect.

3.9 Residual diagnostics

We now consider residual diagnostics for assessing the normality assumptions for ζ_j and ϵ_{ij}.

In section 2.9.2, we discussed empirical Bayes (EB) prediction of the random intercepts for different clusters j. Such predictions $\widetilde{\zeta}_j$ can be interpreted as estimated level-2 residuals for the mothers. We can obtain corresponding estimated level-1 residuals for birth i of mother j as

$$\widetilde{\epsilon}_{ij} = \widehat{\xi}_{ij} - \widetilde{\zeta}_j$$

where

$$\widehat{\xi}_{ij} = y_{ij} - (\widehat{\beta}_1 + \widehat{\beta}_2 x_{2ij} + \cdots + \widehat{\beta}_p x_{pij})$$

In linear mixed models (but not in the generalized linear mixed models discussed later), these estimated level-2 and level-1 residuals have normal sampling distributions if the model is true. We can therefore use histograms or normal quantile–quantile plots of the estimated residuals to assess the assumptions that the true residuals ζ_j and ϵ_{ij} are normally distributed.

We can also try to find outliers by using standardized residuals and looking for values that are unlikely under the standard normal distribution, e.g., values outside the range ± 2. In section 2.9.3, we discussed the sampling standard deviation, or diagnostic standard error, of the empirical Bayes predictions. A standardized level-2 residual can therefore be obtained as