**Chapter 4 Summary**

In this chapter we examined the problem of multicollinearity, a problem commonly encountered in empirical work, especially if there are several correlated explanatory variables in the model. As long as collinearity is not perfect, we can work within the framework of the classical linear regression model, provided the other assumptions of the CLRM are satisfied.

If collinearity is not perfect, but high, several consequences ensue. The OLS estimators are still BLUE, but one or more regression coefficients have large standard errors relative to the values of the coefficients, thereby making the *t* ratios small. Therefore one may conclude (misleadingly) that the true values of these coefficients are not different from zero. Also, the regression coefficients may be very sensitive to small changes in the data, especially if the sample is relatively small (see Exercise 4.6).

There are several diagnostic tests to detect collinearity, but there is no guarantee that they will yield satisfactory results. It is basically a trial and error process.

The best practical advice is to do nothing if you encounter collinearity, for very often we have no control over the data. However, it is very important that the variables included in the model are chosen carefully. As our illustrative example shows, redefining a model by excluding variables that may not belong in the model may attenuate the collinearity problem, provided we do not omit variables that are relevant in a given situation. Otherwise, in reducing collinearity we will be committing model specification errors, which are discussed in Chapter 7. So, think about the

model carefully before you estimate the regression model.

There is one caveat. If there is multicollinearity in a model and if your objective is forecasting, multicollinearity may not be bad, provided the collinear relationship observed in the sample continues to hold in the forecast period.

Finally, there is a statistical technique, called **principal components analysis**, which will “resolve” the problem of near-collinearity. In PCA we construct artificial variables in such a way that they are orthogonal to each other. These artificial variables, called principal components (PC), are extracted from the original *X* regressors.

We can then regress the original regressand on the principal components. We showed how the PCs are computed and interpreted, using our illustrative example.

One advantage of this method is that the PCs are usually smaller in number than the original number of regressors. But one practical disadvantage of the PCA is that the PCs very often do not have viable economic meaning, as they are (weighted) combinations of the original variables which may be measured in different units of measurement. Therefore, it may be hard to interpret the PCs. That is why they are not much used in economic research, although they are used extensively in psychological and education research.