In finance, you may sometimes hear about factor models, where a factor is something observable (inflation, growth,...) that affects prices of stocks (because they cannot be diversified away). A special case is the CAPM, that I usually teach in macro II, where there is one "market factor." In economics, we usually have models with unobservable factors in mind. I will use finance notation here for a start.

Unobserved Factors (Static):
In the case of unobserved factors, we have the model

$$R_{it} \quad = \quad \alpha_i + b_{i1}F_t^1 + b_{i2}F_t^2 + u_{it}\,,$$

where we assume F^1 and F^2 are mean zero, variance one, independent (maybe normal) random variables and the u terms are white noise, with variance σ_{ui}^2, independent of each other. (The first two of these assumptions are mere normalizations.) We refer to the b terms as factor loadings. If we define $R_t = \{R_{1t}, ..., R_{Nt}\}$ for a sample of N stock- or portfolio-returns, and

$$B \;=\; \begin{pmatrix} b_{11} & b_{12} \\ . & . \\ . & . \\ b_{N1} & b_{N2} \end{pmatrix} .$$

We have

$$var(R_t) = BB' + \sigma_u I_N \ .$$

Ignoring the mean for now, we can estimate the parameters of B and σ_u by (non-linear) ML estimation (or least-squares). The variance matrix has $N * (N+1)/2$ separate values and B has $2\,N$ parameters, so for $N = 4$, we have 4*5/2=10 different elements in the variance matrix and 2*4+4 parameters so the model is not identified and you might assume only one factor or that the different returns have similar error variances. We would typically use the model for larger values of N, but the point is that you should always consider if you have an identified model.

For a statistician, this is all the content of a basic factor model: as set of restrictions on the variance covariance matrix.

If there are more than one factor, Newton algorithms may struggle estimating B because it is not unique (actually, even for one factor, the sign is not identified, because the loading "are like standard deviations," but you can easilty force a positive estimate. Some softward comes with algorithms to break non-uniquness, for example, by chosing to maximize the loadings on the first factor.

Factor analysis and Principal Components Factors look a lot like principal components and, indeed, we will sometimes use principal components as estimates of factors. But in principle, there are many difference

1. Principal components are direct functions of the data
2. "Factors" don't necessary"exist, but are constraints on the covariance matrix.
3. There are N principal components, but often one or two factors are assumed
4. If factor model is literally true with, say, 2 factors, there should be two principal components with many loadings absorbing those two and the other principal components should be (asymptotically) orthogonal.
5. Factors are not uniquely identified, only the $B'B$ matrix, but the solution for B is not unique.
6. In various applications, people use Principal Components to "extract" the factors.
7. You can, and often should, test the constraints on the number of factors.

Application in Finance The "APT" (Asset Pricing Theory) application posits, reasonbly, that if stocks follow a factor model there should only be risk premiums associated with what cannot be diversified away; i.e., the factors. If we have N stocks, the APT model assumes that each factor i comes with a risk premium λ so that the mean returns to stocks (after subtracting the save return), are $a = \{\alpha_1, ..., \alpha_N\})$ and $\Lambda = \{\lambda_1, \lambda_2\ \}$'):

$$a = B\Lambda\ ;$$

i.e.: $\alpha_i = b_{i1}\lambda_1 + b_{i2}\lambda_2$, which gives us more degrees of freedom for $N > 2$ as we add two parameters to fit N mean terms. You could now write down the likelihood function in terms a sum of T terms involving the $N-$ dimensional multivariate normal distribution.

The mean is

$$\alpha_i = r^f + b_{i1}\lambda^1 + b_{i2}\lambda^2\ ,$$

where λ^k, $k = 1, 2$ are risk premiums associated with the factors. In general, mean-variance restrictions will help identify parameters.

Dynamic Factor Models A dynamic factor model with one factor will look like this:

$$x_{it} = \alpha_i + \beta F_t + b(L)u_{it}\ ,$$

where

$$\gamma(L)F_t = e_t\ ,$$

for variables $i = 1, ..., N$. Sometimes we will say a "latent" factor model, because F is not observed (if it were, we would just do regression). Depending on the economic model that you have in mind, the $b(L)$ lag-polynomium may just be the identity (i.e., all dynamics is captured by the factor) or often the inverse of an AR-model. Writing it in terms of an MA model is just for convenience. All error terms such as e or u are, sometimes implicitly, i.i.d. and independent of each other—the whole point of the factor model is to parameterize dependencies.
Very often, we have just one lag in the AR polynomium, so

$$F_t = \gamma F_{t-1} + e_t \,,$$

which captures most time variation, but of course there are applications where you want more complicated lag structures.

A natural generalization is

$$x_{it} = \alpha_i + \Sigma_{k=1}^{K} \beta_k F_{kt} + b(L) u_{it} \,,$$

where

$$F_{kt} = a_k(L) e_{kt} \,,$$

where there are K factors and each has it own dynamics.

Factor analysis and co-integration Cointegration is also capturing latent non-stationary variables. Factor analysis most commonly assume stationarity (often implicitly), but does not have to. But if you have a large number of variables, you will run out of degrees of freedom in a Johansen co-integration analysis (the paper of mine in *REStat* that I have mentioned, shows that the critical values have to be adjusted in small samples even for 5–10 variables. Most commonly, if people want to study a large number of variables, they use stationary variables, such a the growth rate of GDP rather than the level.

Estimation of Dynamic Factor models I Linear. Unobserved variables. That spells Kalman filter. Tthe Kalman smoother to extract factors if you need to, although I suspect you may have to think about how to deal with non-uniqueness (haven't tried that).

Estimation of Dynamic Factor models II or Principal Components Redux In forecasting, people sometimes use a very large number of variables, so Kalman filter estimation becomes cumbersome (slow and/or convergence problems). If you are mainly interested in forecasting, or not particular focused on testing for number of factors, you can do the shortcut and, if N is large, you extract the principal components P_{jt}, $j = 1, ...N$ from the vector X_t of the N variables observed at period t. Now, you "factors" are observable and can be used as a regular variables in further analysis. Often only the largest PC or at most a few of the largest. (I am not sure as I write this

if the weights used to construct the PC typically are assumed constant; i.e. $P_{jt} = w_{jt}X_t$ versus $P_{jt} = w_jX_t$. Which is correct? You need a model to answer that and in many application there are only a foggy implicit model, but you should try and be explicit or you can end up in big trouble. My prior would be that there is some constant structure in the economy driving variables so that the weights should be constant. How to do that? One way would be to stack all the time periods and extract the loadings and principal components, which you then split up by t.

FAVAR models If you extract the factors using principal components, they become observables. You can then include the in a VAR (or SVAR) together with other variables. The gold-standard is always to do everything simultaneously using ML, but maybe you want to factor(s) to capture some latent pattern affecting many variables. But a VAR cannot have to many variables, because the degrees of freedom used increased quadratically in number of obs., so you do something a bit more ad hoc. One can, and ought to, evaluate if ones methods is valid by Monte Carlo.

Other estimation One may use simulation methods, which very roughly implies drawing a large number of errors, simulationg the model and picking what fits the data. Gibbs sampling which is often used in Bayesian estimation is a simulation approach which is often surprisingly efficient. I do not have first-hand experience, but of course software is widely available.

The Handbook article by Stock and Watson posted on the class-page gives more details and, in particular, references that you should consult to see the current standards for empirical work in this area (assuming you want to use such methods).