The other full-systems estimation method that is widely used is **nonlinear FIML**. For this, it is convenient to write the equation system to be estimated not as (18.79) but rather as

$$\boldsymbol{h}_t(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{\theta}) = \boldsymbol{U}_t, \quad \boldsymbol{U}_t \sim \mathrm{NID}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{18.84}$$

where $\boldsymbol{\theta}$ is still a p–vector of parameters, $\boldsymbol{h}_t$ is a $1 \times g$ vector of nonlinear functions, and $\boldsymbol{U}_t$ is a $1 \times g$ vector of error terms. There need be no conflict between (18.79) and (18.84) if we think of the i^{th} element of $\boldsymbol{h}_t(\cdot)$ as being the same as the t^{th} element of $\boldsymbol{f}_i(\cdot)$.

The density of the vector $\boldsymbol{U}_t$ is

$$(2\pi)^{-g/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\Big(-\tfrac{1}{2}\boldsymbol{U}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_t^{\top}\Big).$$

To obtain the density of $\boldsymbol{Y}_t$, we must replace $\boldsymbol{U}_t$ by $\boldsymbol{h}_t(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{\theta})$ and multiply by the Jacobian factor $|\det \boldsymbol{J}_t|$, where $\boldsymbol{J}_t \equiv \partial\boldsymbol{h}_t(\boldsymbol{\theta})/\partial\boldsymbol{Y}_t$, that is, the $g \times g$ matrix of derivatives of $\boldsymbol{h}_t$ with respect to the elements of $\boldsymbol{Y}_t$. The result is

$$(2\pi)^{-g/2}|\det \boldsymbol{J}_t||\boldsymbol{\Sigma}|^{-1/2} \exp\Big(-\tfrac{1}{2}\boldsymbol{h}_t(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}\boldsymbol{h}_t^{\top}(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{\theta})\Big).$$

It follows immediately that the loglikelihood function is

$$\begin{aligned}\ell(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = &-\frac{ng}{2}\log(2\pi) + \sum_{t=1}^{n}\log|\det \boldsymbol{J}_t| - \frac{n}{2}\log|\boldsymbol{\Sigma}| \\ &-\frac{1}{2}\sum_{t=1}^{n}\boldsymbol{h}_t(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}\boldsymbol{h}_t^{\top}(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{\theta}).\end{aligned} \tag{18.85}$$

This may then be maximized with respect to $\boldsymbol{\Sigma}$ and the result substituted back in to yield the concentrated loglikelihood function

$$\begin{aligned}\ell^c(\boldsymbol{\theta}) = &-\frac{ng}{2}\big(\log(2\pi) + 1\big) + \sum_{t=1}^{n}\log|\det \boldsymbol{J}_t| \\ &-\frac{n}{2}\log\Big|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{h}_t^{\top}(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{\theta})\boldsymbol{h}_t(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{\theta})\Big|.\end{aligned} \tag{18.86}$$

Inevitably, there is a strong resemblance between (18.85) and (18.86) and their counterparts (18.28) and (18.30) for the linear case. The major difference is that the Jacobian term in (18.85) and (18.86) is the sum of the logs of n different determinants. Thus every time one evaluates one of these loglikelihood functions, one has to calculate n different determinants. This can be very expensive if g or n is large. Of course, the problem goes away if the model is linear in the endogenous variables, since $\boldsymbol{J}_t$ will then be the same for all t.