Regression with time series

Class Notes

Manuel Arellano

February 22, 2018

1 Classical regression model with time series

Model and assumptions The basic assumption is

$$E(y_t \mid x_1, ..., x_T) = E(y_t \mid x_t) = x_t'\beta.$$

The first equality is always satisfied with *iid* observations whereas the second imposes linearity in the relationship. With dependent observations the first equality imposes restrictions on the stochastic process of (y_t, x_t). In later sections we will study the nature of these restrictions and consider generalizations that are suitable for time series.

The previous assumption can be written in the form of equation as follows

$$y_t = x_t'\beta + u_t \qquad E(u_t \mid X) = 0$$

where $X = (x_1, ..., x_T)'$ and $y = (y_1, ..., y_T)'$. This is appropriate for a linear relationship between x and y and unobservables that are mean independent of past and future values of x.

The second assumption in the classical regression model is $Var(y \mid X) = \sigma^2 I_T$, which amounts to

$$\begin{aligned} E\left(u_t^2 \mid X\right) &= \sigma^2 \quad \text{for all } t \\ E(u_t u_{t-j} \mid X) &= 0 \quad \text{for all } t, j. \end{aligned}$$

The case where $E(u_t u_{t-j} \mid X) \neq 0$ is called autocorrelation. An example is a Cobb-Douglas production function in which u represents multiplicative measurement error in the output level, independent of inputs at all periods. The error u is autocorrelated possibly as a result of temporal aggregation in the data.

In the context of the classical regression model with time series it is convenient to distinguish between conditional heteroskedasticity and unconditional heteroskedasticity. If $E\left(u_t^2 \mid X\right)$ does not depend on X there is no conditional heteroskedasticity and

$$E\left(u_t^2 \mid X\right) = E\left(u_t^2\right).$$

However, if u_t is not stationary in variance then

$$E\left(u_t^2\right) = \sigma_t^2$$

where σ_t^2 is a function of t. In this situation we speak of unconditional heteroskedasticity.

On the other hand if u_t is stationary then $E\left(u_t^2\right) = \sigma^2$, which is compatible with the possibility of conditional heteroskedasticity. For example,

$$\begin{aligned} E\left(u_t^2 \mid X\right) &= \delta_0 + \delta_1 x_t^2 \\ E\left(u_t^2\right) &= \delta_0 + \delta_1 E\left(x_t^2\right) = \sigma^2. \end{aligned}$$

In this case there will be unconditional homoskedasticity if $E\left(x_t^2\right)$ is constant for all t.

The same reasoning can be applied to conditional and unconditional autocovariances. If u_t is stationary then

$$E\left(u_t u_{t-j}\right) = \gamma_j \quad \text{for all } t.$$

However, it is possible that conditional autocovariances

$$E\left(u_t u_{t-j} \mid X\right) = \gamma_j\left(X\right)$$

depend on X, in which case we would have both conditional heteroskedasticity and autocorrelation.

OLS with dependent observations: robust inference Consider the usual representation for the scaled estimation error:

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t.$$

Letting $w_t = x_t u_t = x_t\left(y_t - x_t'\beta\right)$, we have that $E\left(w_t\right) = 0$. Moreover, if $\left(y_t, x_t\right)$ is stationary, so is w_t. Under some conditions, $\overline{w}_T \xrightarrow{p} 0$, $T^{-1}\sum_{t=1}^{T} x_t x_t' \xrightarrow{p} E\left(x_t x_t'\right) > 0$, and $\sqrt{T}\overline{w}_T \xrightarrow{d} \mathcal{N}\left(0, V\right)$ with $V = \sum_{j=-\infty}^{\infty} \Gamma_j$ and $\Gamma_j = E\left(w_t w_{t-j}'\right) = E\left(u_t u_{t-j} x_t x_{t-j}'\right)$. It then follows that $\widehat{\beta}$ is consistent and asymptotically normal:

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, W\right)$$

where

$$W = \left[E\left(x_t x_t'\right)\right]^{-1} V \left[E\left(x_t x_t'\right)\right]^{-1}.$$

Recall that if observations are *iid* $V = E\left(u_t^2 x_t x_t'\right)$. If in addition there is absence of conditional heteroskedasticity $V = \sigma^2 E\left(x_t x_t'\right)$. On the other hand if $\gamma_j\left(X\right) = \gamma_j$ for all j:

$$V = \textstyle\sum_{j=-\infty}^{\infty} E\left(u_t u_{t-j}\right) E\left(x_t x_{t-j}'\right).$$

The Newey-West estimate of V is

$$\widehat{V} = \widehat{\Gamma}_0 + \sum_{j=1}^{m}\left(1 - \frac{j}{m+1}\right)\left(\widehat{\Gamma}_j + \widehat{\Gamma}_j'\right)$$

where

$$\widehat{\Gamma}_j = \frac{1}{T-j}\sum_{t=j+1}^{T}\widehat{u}_t\widehat{u}_{t-j}x_t x_{t-j}'$$

and $\widehat{u}_t$ are OLS residuals. The corresponding estimate of W is

$$\widehat{W} = \left(\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\right)^{-1}\widehat{V}\left(\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\right)^{-1},$$

which boils down to the heteroskedasticity-consistent White standard error formula if $m = 0$ so that $\widehat{V} = \widehat{\Gamma}_0$.

If there is autocorrelation but no heteroskedasticity an alternative consistent estimator of V is

$$\widetilde{V} = \widetilde{\Gamma}_0 + \sum_{j=1}^{m}\left(\widetilde{\Gamma}_j + \widetilde{\Gamma}_j'\right)$$

where

$$\widetilde{\Gamma}_j = \widetilde{\gamma}_j\left(\frac{1}{T-j}\sum_{t=j+1}^{T}x_t x_{t-j}'\right)$$

and $\widetilde{\gamma}_j = (T-j)^{-1}\sum_{t=j+1}^{T}\widehat{u}_t\widehat{u}_{t-j}$.

The previous limiting results do not depend on the validity of the assumptions of the classical linear regression model. They are valid for inference about regression coefficients that are regarded as estimates of linear projections from stationary and ergodic stochastic processes (or some alternative mixing conditions under which the results hold).

For example, the results can be used if x_t is y_{t-1} or contains lags of y among other variables, regardless of whether the model is autoregressive or not.

This is not to say that OLS will necessarily be a consistent estimator of a dynamic model with autocorrelation, in fact in general it will not be. For example, take an ARMA$(1,1)$ model:

$$\begin{aligned} y_t &= \alpha + \rho y_{t-1} + u_t \\ u_t &= v_t - \theta v_{t-1}. \end{aligned}$$

The linear projection of y_t on y_{t-1} will not coincide with $\alpha + \rho y_{t-1}$ because in the ARMA$(1,1)$ model $Cov\left(y_{t-1}, u_t\right) \neq 0$. The previous results for regressions would allow us to make inference about the linear projection coefficients, not about (α, ρ) in the ARMA(1,1) model.

Generalized least squares Under the assumption $E\left(u_t \mid X\right) = 0$, OLS is not only consistent but also unbiased. Letting $u = (u_1, ..., u_T)'$, its variance given X in finite samples is given by

$$Var\left(\widehat{\beta} \mid X\right) = \left(X'X\right)^{-1}X'E\left(uu' \mid X\right)X\left(X'X\right)^{-1}.$$

If there is autocorrelation there is a changing dependence among data points. This suggests considering weighted least squares estimates in which different pairs of observations receive different weights. That is, estimators of the form

$$\widetilde{\beta} = \left(X'H'HX\right)^{-1} X'H'Hy$$

or equivalently OLS in the transformed regression

$$Hy = HX\beta + Hu$$

where H is a $T \times T$ matrix of weights.

We have previously dealt with the case in which H is diagonal. Here since the estimator only depends on $H'H$ we can limit ourselves to consider weight matrices that are lower triangular:

$$H = \begin{pmatrix} h_{11} & 0 & \dots & 0 \\ h_{21} & h_{22} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ h_{T1} & h_{T2} & & h_{TT} \end{pmatrix},$$

so that

$$\widetilde{\beta} - \beta = \left(\sum_{t=1}^{T} x_t^* x_t^{*\prime}\right)^{-1} \sum_{t=1}^{T} x_t^* u_t^*$$

where $u_t^* = h_{t1}u_1 + h_{t2}u_2 + ... + h_{tt}u_t$ and similarly for x_t^*. The elements of H may be constant or functions of X: $H = H(X)$.

Under the strict exogeneity assumption $E(u \mid X) = 0$, $\widetilde{\beta}$ is unbiased (and consistent) for any H. In general, $\widetilde{\beta}$ will not be consistent for β in a best linear predictor or in a conditional expectation model in the absence of the strict exogeneity assumption. We can obtain an asymptotic normality result for $\widetilde{\beta}$ similar to the one for OLS simply replacing (y_t, x_t) with (y_t^*, x_t^*).

Under some regularity conditions, as long as $Var(u \mid X) = \Omega(X) = (H'H)^{-1} = H^{-1}H'^{-1}$, we get the optimal GLS estimator, which satisfies:

$$\sqrt{T}\left(\widetilde{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \left[\operatorname*{plim}_{T\to\infty} \frac{1}{T} X'\Omega(X)^{-1} X\right]^{-1}\right).$$

Thus, a triangular factorization of the inverse covariance matrix of y given X is an efficient choice of weight matrix. Intuitively, this transformation produces errors such that their autocovariance matrix is an identity. Once again we obtain a generalized least squares statistic similar to the one we encountered before:

$$\widetilde{\beta} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y.$$

Feasible GLS with AR(1) errors A popular GLS parametric transformation is when u_t is a first-order autoregressive process. The details are as follows. Assuming that

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim iid\left(0, \sigma^2\right),$$

we have $\Omega = \sigma^2 V$ with

$$V = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \vdots & & & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{pmatrix}.$$

It can be shown by direct multiplication that $V^{-1} = H'H$ with

$$H = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}.$$

Thus, in this case

$$u^* = Hu = \begin{pmatrix} \sqrt{1-\rho^2}u_1 \\ u_2 - \rho u_1 \\ \vdots \\ u_T - \rho u_{T-1} \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}u_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}.$$

If the first observation is omitted, GLS is equivalent to OLS in the transformed equation:

$$y_t - \rho y_{t-1} = \left(x_t - \rho x_{t-1}\right)' \beta + \varepsilon_t,$$

which is equivalent to MLE given the first observation.

Letting $\widehat{\rho}$ be the autoregressive coefficient estimated from OLS residuals,[1] the Cochrane-Orcutt procedure consists in constructing the pseudo differences $y_t - \widehat{\rho} y_{t-1}$ and $x_t - \widehat{\rho} x_{t-1}$ and estimating the transformed model by OLS.

Alternatively, the full log-likelihood can be maximized with respect to all parameters:

$$L\left(\beta, \sigma^2, \rho\right) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{1}{2}(y - X\beta)' \Omega^{-1} (y - X\beta).$$

[1] Specifically, $\widehat{\rho} = \sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1} / \sum_{t=2}^{T} \widehat{u}_{t-1}^2$ where $\widehat{u}_t = y_t - x_t' \widehat{\beta}$ and $\widehat{\beta} = \left(X'X\right)^{-1} X'y$.

2 Distributed lags and partial adjustment

2.1 Distributed lags

A generalization of the classical regression model that maintains the strict exogeneity assumption but allows for dynamic responses of y to changes in x is:

$$E\left(y_t \mid x_1, ..., x_T\right) = E\left(y_t \mid x_t, x_{t-1}, ..., x_{t-p}\right) = \delta + \beta_0 x_t + \beta_1 x_{t-1} + ... + \beta_p x_{t-p}.$$

This is called a distributed lag model. We use a scalar regressor for simplicity. Formally, it is the same as a model with $p+2$ exogenous regressors $(1, x_t, ..., x_{t-p})$.

The coefficient β_0 is the short run multiplier whereas the long run multiplier is given by

$$\gamma = \beta_0 + \beta_1 + ... + \beta_p.$$

Letting $x_1 = 1$ and $x_j = 0$ for $j \neq 1$, the distribution of lag responses is obtained from:

$$\begin{aligned} E_x\left(y_0\right) &= \delta \\ E_x\left(y_1\right) &= \delta + \beta_0 \\ E_x\left(y_2\right) &= \delta + \beta_1 \\ &\vdots \\ E_x\left(y_{p+1}\right) &= \delta + \beta_p \\ E_x\left(y_{p+2}\right) &= \delta. \end{aligned}$$

The mean lag is given by $\left(\sum_{j=0}^{p} j\beta_j\right) / \sum_{j=0}^{p} \beta_j$ whereas the median lag is the lag at which 50 percent of the total effect has taken place.

If one is interested in long run effects it may be convenient to reformulate the equation. For example, if $p = 2$ we can go from

$$y_t = \delta + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

to the following reparameterization:

$$y_t = \delta + \gamma x_t + \beta_1\left(x_{t-1} - x_t\right) + \beta_2\left(x_{t-2} - x_t\right) + u_t$$

where $\gamma = \beta_0 + \beta_1 + \beta_2$ or also

$$y_t = \delta + \gamma z_t + \left(\beta_1 - \beta_0\right)\left(x_{t-1} - z_t\right) + \left(\beta_2 - \beta_0\right)\left(x_{t-2} - z_t\right) + u_t$$

where $z_t = \left(x_t + x_{t-1} + x_{t-2}\right)/3$.

Sometimes p is estimated as part of a model selection procedure. It is also common to model the lag structure, that is, to specify $\left(\beta_0, \beta_1, ..., \beta_p\right)$ as functions of a smaller set of parameters (e.g. the polynomial model introduced by Shirley Almon in 1965), especially when the x_{t-j} are highly colinear and the information about individual β_j is small.

Koyck distributed lags A long-standing popular model specifies a geometric lag structure:

$$\beta_j = \beta_0 \alpha^j \qquad (j = 1, ..., p).$$

If we let $p \to \infty$ we have

$$E\left(y_t \mid \{x_s\}_{s=-\infty}^{T}\right) = \delta + \beta_0 x_t + \alpha\beta_0 x_{t-1} + \alpha^2 \beta_0 x_{t-2} + ... \tag{1}$$

or in equation form

$$y_t = \delta + \beta_0 \textstyle\sum_{j=0}^{\infty} \alpha^j x_{t-j} + u_t \qquad E\left(u_t \mid ...\ x_{-1}, x_0, x_1, ..., x_T\right) = 0.$$

The long run effect in this case is $\gamma = \beta_0 / (1 - \alpha)$ if $|\alpha| < 1$.

Subtracting (1) from the lagged equation multiplied by α:

$$E\left(y_t - \alpha y_{t-1} \mid \{x_s\}_{s=-\infty}^{T}\right) = \delta^* + \beta_0 x_t$$

where $\delta^* = (1 - \alpha)\,\delta$. Similarly,

$$y_t = \delta^* + \alpha y_{t-1} + \beta_0 x_t + \varepsilon_t \tag{2}$$

where $\varepsilon_t = u_t - \alpha u_{t-1}$ so that $E\left(\varepsilon_t \mid ...\ x_{-1}, x_0, x_1, ..., x_T\right) = 0$.

In general, $(\delta^*, \alpha, \beta_0)$ are not the coefficients of a population linear regression of y_t on $(1, y_{t-1}, x_t)$ because in equation (2) y_{t-1} is correlated with ε_t through u_{t-1}. A similar situation arose in the case of ARMA$(1,1)$ models.

For a given value of p, nonlinear least squares estimates of $(\delta, \alpha, \beta_0)$ solve:

$$\min \sum_{t=p+1}^{T} \left(y_t - \delta - \beta_0 \textstyle\sum_{j=0}^{p} \alpha^j x_{t-j}\right)^2.$$

2.2 Partial adjustment

An equation that includes lags of y_t together with other explanatory variables and an error term (with properties to be discussed below) is often called a partial adjustment model. A simple version is:

$$y_t = \delta + \alpha y_{t-1} + \beta_0 x_t + u_t. \tag{3}$$

The name comes from a hypothesis of gradual adjustment to an optimal target value y_t^* when adjustment is costly:

$$y_t - y_{t-1} = \gamma\left(y_t^* - y_{t-1}\right)$$

or

$$y_t = (1 - \gamma)\, y_{t-1} + \gamma y_t^*. \tag{4}$$

Equation (4) gives rise to (3) if

$$y_t^* = \delta^* + \beta^* x_t + u_t^*$$

with $\delta = \gamma\delta^*$, $\beta_0 = \gamma\beta^*$, $u_t = \gamma u_t^*$, and $\alpha = (1-\gamma)$. The last coefficient captures the speed of adjustment, which is typically a quantity of interest (e.g. in models of investment, labor demand or consumption with habits).

The empirical content of a relationship like (3) depends on its statistical interpretation. One possibility is to regard $(\delta + \alpha y_{t-1} + \beta_0 x_t)$ as the expectation of y_t given $(y_1, ..., y_{t-1}, x_1, ..., x_T)$ so that

$$E(u_t \mid y_1, ..., y_{t-1}, x_1, ..., x_T) = 0.$$

This implies that $E(u_t \mid u_1, ..., u_{t-1}) = 0$ and therefore lack of serial correlation in u_t: $E(u_t u_{t-j}) = 0$ for $j > 0$. This interpretation is incompatible with the geometric distributed lag model and more generally with any model in which it is intended to allow for both dynamics and serial correlation.

Partial adjustment vs serial correlation In a *static model* with serial correlation:

$$y_t = \delta + \beta x_t + u_t$$

the response of y_t to a change in x_t is static. There is just persistence in the error term u_t (in the same way that there may be persistence in x_t). Examples include production functions and wage equations.

In a *partial adjustment* model

$$y_t = \delta + \alpha y_{t-1} + \beta_0 x_t + u_t$$

the effect of x on y is dynamic (as seen in the discussion of geometric distributed lags).

In a static model with strictly exogenous x, serial correlation in u does not alter the fact that

$$\beta = \frac{Cov(x_t, y_t)}{Var(x_t)}.$$

In contrast, under the assumptions of the static model with autocorrelation, the linear projection of y_t on $z_t = (y_{t-1}, x_t)'$ does not provide consistent estimates of the static model parameters:

$$\begin{aligned}
\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} &= [Var(z_t)]^{-1} Cov(z_t, y_t) = \begin{pmatrix} Var(y_{t-1}) & Cov(y_{t-1}, x_t) \\ Cov(y_{t-1}, x_t) & Var(x_t) \end{pmatrix}^{-1} \begin{pmatrix} Cov(y_{t-1}, y_t) \\ Cov(x_t, y_t) \end{pmatrix} \\
&= \begin{pmatrix} \beta^2 Var(x_t) + \sigma_u^2 & \beta Cov(x_{t-1}, x_t) \\ \beta Cov(x_{t-1}, x_t) & Var(x_t) \end{pmatrix}^{-1} \begin{pmatrix} \beta^2 Cov(x_t, x_{t-1}) + Cov(u_t, u_{t-1}) \\ \beta Var(x_t) \end{pmatrix}.
\end{aligned}$$

In the special case in which $Cov(x_{t-1}, x_t) = 0$ we have

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} \frac{Cov(u_t, u_{t-1})}{\beta^2 Var(x_t) + \sigma_u^2} \\ \beta \end{pmatrix}.$$

Thus, $\psi_2 = \beta$ but $\psi_1 \neq 0$ unless $Cov(u_t, u_{t-1}) = 0$.

Common factor restrictions Is it possible to distinguish empirically between partial adjustment and serial correlation? A static model with AR(1) errors:

$$
\begin{aligned}
y_t &= \delta + \beta x_t + u_t \\
u_t &= \rho u_{t-1} + \varepsilon_t
\end{aligned}
$$

upon substitution can be written in the form

$$(y_t - \delta - \beta x_t) = \rho\,(y_{t-1} - \delta - \beta x_{t-1}) + \varepsilon_t$$

or

$$y_t = (1-\rho)\,\delta + \beta x_t - \rho\beta x_{t-1} + \rho y_{t-1} + \varepsilon_t.$$

This equation can be regarded as a special case of a partial adjustment model without serial correlation:

$$y_t = \pi_0 + \psi_0 x_t + \psi_1 x_{t-1} + \alpha y_{t-1} + \varepsilon_t$$

subject to the restriction

$$\psi_1 = -\alpha\psi_0. \tag{5}$$

This type of constraint and its generalizations are Sargan's Comfac or common factor restrictions.[2] They can be easily tested using a Wald statistic because the estimation under the alternative hypothesis can be done by OLS.

In fact, a Comfac restriction is always satisfied under the following null hypothesis:

$$H_0: \quad y_t = \delta + \beta x_t + u_t \quad E\,(u_t \mid x_1, ..., x_T) = 0.$$

To see this, let the linear projection of y_t on $(1, x_t, x_{t-1}, y_{t-1})$ be

$$y_t = \pi_0 + \psi_0 x_t + \psi_1 x_{t-1} + \alpha y_{t-1} + \varepsilon_t \tag{6}$$

where $E^*\,(\varepsilon_t \mid x_t, x_{t-1}, y_{t-1}) = 0$. Due to the law of iterated projections

$$E^*\,(y_t \mid x_t, x_{t-1}) = \pi_0 + \psi_0 x_t + \psi_1 x_{t-1} + \alpha E^*\,(y_{t-1} \mid x_t, x_{t-1})$$

or

$$(\delta + \beta x_t) = \pi_0 + \psi_0 x_t + \psi_1 x_{t-1} + \alpha\,(\delta + \beta x_{t-1})\,.$$

Matching coefficients:

$$
\begin{aligned}
\pi_0 + \alpha\delta &= \delta \implies \pi_0 = (1-\alpha)\,\delta \\
\psi_0 &= \beta \\
\psi_1 + \alpha\beta &= 0 \implies \text{Comfac restriction}
\end{aligned}
$$

[2] Sargan, J. D. (1980): "Some Tests of Dynamic Specification for a Single Equation," *Econometrica*, 48, 879–897.

Thus, a test of the hypothesis H_0: $\psi_1 + \alpha\beta = 0$ in the linear projection (6) has power against a static model with serial correlation and exogenous regressors even if the form of the autocorrelation is not AR(1).

Partial adjustment with autocorrelation A Comfac test has power to reject a static model with autocorrelation. However, identification of a dynamic model may be problematic if the model combines both partial adjustment and serial correlation. For example, OLS is not consistent in the partial adjustment model

$$y_t = \delta + \alpha y_{t-1} + \beta_0 x_t + u_t$$

if u_t is serially correlated. Nevertheless, if $u_t \sim$AR(1): $u_t = \rho u_{t-1} + \varepsilon_t$ then

$$(y_t - \delta - \alpha y_{t-1} - \beta_0 x_t) = \rho\left(y_{t-1} - \delta - \alpha y_{t-2} - \beta_0 x_{t-1}\right) + \varepsilon_t$$

or

$$y_t = (1-\rho)\,\delta + (\alpha+\rho)\, y_{t-1} - \alpha\rho y_{t-2} + \beta_0 x_t - \rho\beta_0 x_{t-1} + \varepsilon_t$$

giving rise to a new level of Comfac restrictions, which can be tested and enforced in estimation.

This type of model is an example of a stochastic relationship between variables in which the regressors are not independent of the errors. The estimation problem for models of this type will be considered in greater generality in the context of instrumental variable estimation.

3 Predetermined variables

In the classical regression model

$$y_t = x_t'\beta + u_t$$

the variable x_t is strictly exogenous in the sense that

$$E\left(x_t u_s\right) = 0 \text{ for all } t, s.$$

We say that a variable is predetermined if

$$E\left(x_t u_t\right) = 0, \quad E\left(x_t u_{t+1}\right) = 0, \quad E\left(x_t u_{t+2}\right) = 0, ...$$

but we do not exclude the possibility that

$$E\left(x_t u_{t-1}\right) \neq 0, \quad E\left(x_t u_{t-2}\right) \neq 0, ...$$

An example of predetermined variable is y_{t-1} in the AR(1) model. However, y_{t-1} is not predetermined in the geometric distributed lag model. Another example of predetermined variable arises in a relation between female labor force participation and children. Yet another example is in testing for market efficiency in foreign exchange markets (Hansen and Hodrick, 1980).[3]

An alternative weaker condition for predetermined variables is to simply require $E(u_t \mid x_t) = 0$ or

$$E(x_t u_t) = 0,$$

which is the assumption that implies consistency of OLS under standard regularity conditions.

Dynamic regression with sequential conditioning Here we consider partial adjustment models for conditional means of the form

$$E(y_t \mid y_1, ..., y_{t-1}, x_1, ..., x_t).$$

In a linear specification with first order lags we have

$$y_t = \delta + \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t$$

$$E(u_t \mid y_1, ..., y_{t-1}, x_1, ..., x_t) = 0.$$

In this type of model u_t is serially uncorrelated by construction. The regressor x_t is predetermined in the sense of being correlated with past values of u or y (feedback).

If x is strictly exogenous then

$$E(u_t \mid y_1, ..., y_{t-1}, x_1, ..., x_t) = E(u_t \mid y_1, ..., y_{t-1}, x_1, ..., x_T),$$

so that a test of strict exogeneity in this context is a test of significance of future values of x in the regression (a Sims' type test).[4]

Conditional means of the form $E(y_t \mid y_1, ..., y_{t-1}, x_1, ..., x_t)$ are natural predictors in the time series context, but they may or may not correspond to quantities of economic interest in an application.

Estimation In regression models with sequential conditioning OLS is consistent but not unbiased. In small samples the bias can be a problem. In an AR(1) model with a positive autoregressive parameter the bias is negative.[5]

When x is a predetermined variable, regressions in linear transformations of the data such as GLS are not justified in general and may lead to inconsistent estimates even if OLS is consistent.

[3] Hansen, L. P. and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829–853.

[4] Sims, C. A. (1972): "Money, Income, and Causality," *American Economic Review*, 62, 540–552.

[5] Hurwicz, L. (1950): "Least Squares Bias in Time Series," in Koopmans, T. C. (ed.), *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph No. 10, John Wiley, New York.

4 Granger causality

Let the observed time series be $y^T = (y_1, ..., y_T)$ and $x^T = (x_1, ..., x_T)$. The joint distribution of the data admits the following factorizations:[6]

$$f\left(y^T, x^T\right) = f\left(y^T \mid x^T\right) f\left(x^T\right) = \prod_{t=1}^T f\left(y_t \mid y^{t-1}, x^T\right) \prod_{t=1}^T f\left(x_t \mid x^{t-1}\right) \tag{7}$$

Also,[7]

$$f\left(y^T, x^T\right) = \prod_{t=1}^T f\left(y_t, x_t \mid y^{t-1}, x^{t-1}\right) = \prod_{t=1}^T f\left(y_t \mid y^{t-1}, x^t\right) \prod_{t=1}^T f\left(x_t \mid y^{t-1}, x^{t-1}\right) \tag{8}$$

- In an autoregressive univariate time series analysis one models the mean or other characteristics of the distribution of $f\left(x_t \mid x^{t-1}\right)$.
- In a VAR multivariate time series analysis one models the means of the joint distribution $f\left(y_t, x_t \mid y^{t-1}, x^{t-1}\right)$.
- In a dynamic regression with sequential conditioning one models the mean of $f\left(y_t \mid y^{t-1}, x^t\right)$.
- In a classical regression model one models the means of $f\left(y^T \mid x^T\right)$ (in a static model assuming that $E\left(y_t \mid x^T\right) = E\left(y_t \mid x_t\right)$).

All these are different aspects of the joint distribution of the data we may be interested to study.

Granger non-causality We say that y does not Granger cause x if[8]

$$E^*\left(x_t \mid x^{t-1}, y^{t-1}\right) = E^*\left(x_t \mid x^{t-1}\right), \tag{9}$$

or using a definition based on distributions if

$$f\left(x_t \mid x^{t-1}, y^{t-1}\right) = f\left(x_t \mid x^{t-1}\right). \tag{10}$$

It can be shown that (9) is equivalent to the Sims' strict exogeneity condition:

$$E^*\left(y_t \mid x^T\right) = E^*\left(y_t \mid x^t\right) \tag{11}$$

and also that (10) is equivalent to the Chamberlain-Sims distributional strict exogeneity condition:[9]

$$f\left(y_t \mid y^{t-1}, x^T\right) = f\left(y_t \mid y^{t-1}, x^t\right). \tag{12}$$

[6] Arellano, M. (1992): "On Exogeneity and Identifiability," *Investigaciones Económicas*, 16, 401–409.

[7] With some abuse of notation $f\left(x_1 \mid x^0\right)$ denotes $f\left(x_1\right)$ and $f\left(y_1 \mid y^0, x^T\right)$ denotes $f\left(y_1 \mid x^T\right)$. Similarly, $f\left(y_1, x_1 \mid y^0, x^0\right)$ denotes $f\left(y_1, x_1\right)$, $f\left(y_1 \mid y^0, x^1\right)$ denotes $f\left(y_1 \mid x_1\right)$ and $f\left(x_1 \mid y^0, x^0\right)$ denotes $f\left(x_1\right)$.

[8] Granger, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–438.

[9] Chamberlain, G. (1982): "The General Equivalence of Granger and Sims Causality," *Econometrica*, 50, 569–581.

Note that if (10) holds, the second components in the factorizations (7) and (8) of the distribution of the data will coincide, so that the first components must also coincide.

These are notions of causality based on the idea of temporal ordering of predictors: the effect cannot happen before the cause.

Finding Granger causality is not in itself evidence of causality. Due to the operation of unobservables and omitted variables, Granger causality does not imply nor is it implied by causality.

5 Cointegration

Error correction mechanism representation (ECM) Consider a dynamic regression model

$$y_t = \delta + \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t.$$

If we subtract y_{t-1} from both sides of the equation and add $-\beta_0 x_{t-1} + \beta_0 x_{t-1}$ to the l.h.s. we get:

$$y_t - y_{t-1} = \delta - (1-\alpha) y_{t-1} + \beta_0 (x_t - x_{t-1}) + (\beta_0 + \beta_1) x_{t-1} + u_t$$

and also

$$\Delta y_t = \delta + \beta_0 \Delta x_t - (1-\alpha)(y_{t-1} - \gamma x_{t-1}) + u_t \tag{13}$$

where γ is the long run effect

$$\gamma = \frac{\beta_0 + \beta_1}{1-\alpha}.$$

Thus, $(y_{t-1} - \gamma x_{t-1})$ can be seen as the error in the long run relationship between y and x. According to (13) a large deviation in the long run error will have a negative impact on the change Δy given Δx, hence the term "error correction mechanism" representation applied to (13).

Equation (13) is convenient for enforcing long-run restrictions in the estimation of partial adjustment models. For example, Davidson et al. (1978) imposed a long-run income elasticity of unity in the estimation of a consumption function using an equation like (13) subject to $\gamma = 1$.[10]

Cointegration The ECM representation is specially useful in the case in which $y_t \sim I(1)$ and $x_t \sim I(1)$ but $y_t - \gamma x_t \sim I(0)$. In this situation one says that (y_t, x_t) are cointegrated.[11]

More generally, we say that the variables in an $m \times 1$ time series vector w_t are cointegrated if all the variables are $I(1)$ but there is a linear combination that is $I(0)$:

$$a' w_t \sim I(0)$$

[10] Davidson, J. E. H., D. F Hendry, F. Srba, and S. Yeo (1978): "Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom," *The Economic Journal*, 88, 661–692.

[11] Engle, R. F. and C. W. Granger (1987): "Co-integration and error correction: Representation, estimation and testing," *Econometrica*, 55, 251–276.

for some vector a different from zero, which is called the cointegrating vector.

As an example let us consider the following model for $w_t = (y_t, x_t)'$:

$$\begin{aligned} y_t &= \beta x_t + u_t \\ x_t &= x_{t-1} + \varepsilon_t \end{aligned}$$

where (u_t, ε_t) are white noise.

In this example x_t is a random walk and therefore $I(1)$. The variable y_t is also $I(1)$ but $y_t - \beta x_t$ is $I(0)$. The cointegration vector is $(1, -\beta)$. The idea is that while there may be permanent changes over time in the individual time series, there is a long run relationship that keeps together the individual components, which is represented by $a' w_t$.

An early example of error-correction model is Sargan's 1964 study of wages and prices in the UK.[12] In Sargan's model Δy_t and Δx_t are wage and price inflation respectively, whereas the error correction term is a deviation of real wages from a productivity trend. This equilibrium term captures the role of real-wage resistance in wage bargains as a mechanism for regaining losses from unanticipated inflation.

Literature development Some important developments in the cointegration literature are:

- Representations for VAR process of cointegrated multivariate time series. In general, it can be shown that if $w_t \sim I(1)$ is a cointegrated vector of time series then an ECM representation exits (Granger representation theorem).[13]

- Estimation of the cointegrating vector. Given that the time series are $I(1)$, estimators of the cointegrating vector are "superconsistent".

- Cointegration tests (with and without knowledge of the cointegrating vector).

[12] Sargan, J. D. (1964): "Wages and prices in the United Kingdom: A Study in Econometric Methodology." In P.E. Hart, G. Mills and J. Whitaker (eds), *Econometric Analysis for National Economic Planning*, Colston Papers 16, London.

[13] Granger, C. W. J. (1981): "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121–130.