Dynamic Panel Data Models II: Lags and Predetermined Variables[1]

Class Notes

Manuel Arellano

Revised: October 18, 2009

1 Models with Strictly Exogenous and Lagged Dependent Variables

1.1 The Nature of the Model

In this section we extend the fixed effect model with strictly exogenous regressors to include lags of the dependent variable, allowing for error serial correlation of unknown form. The prototypical equation takes the form

$$y_{it} = \alpha y_{i(t-1)} + x'_{it}\beta + \eta_i + v_{it}, \tag{1}$$

together with the assumption

$$E\left(v_{it} \mid x_{i1}, ..., x_{iT}, \eta_i\right) = 0 \;\; (t = 1, ..., T). \tag{2}$$

An equation of this type might also contain lags of x and/or additional lags of y, but (1) captures the essential feature of the model that we wish to discuss. Namely, a dynamic effect of x on y for which the speed of adjustment is governed by the coefficient of lagged y.

Assumption (2) implies that x is uncorrelated to past, present and future values of v, and hence it is a strictly exogenous variable. It does not rule out correlation between x and the individual effect η. Lagged y will be correlated by construction with η and with lagged v, but it may also be correlated with contemporaneous v if v is serially correlated, which is not ruled out by (2). Thus, lagged y is effectively an endogenous explanatory variable in equation (1) with respect to both η and v.

Examples include partial adjustment models of firm investment or labour demand, and household consumption or labour supply models with habits. In these applications the coefficient α captures the magnitude of adjustment costs or the extent of habits. It therefore has a structural significance. Moreover, there are often reasons to expect serial correlation in the transitory errors v of the economic model. In those cases lagged y must be treated as an endogenous explanatory variable.

Assumption (2) implies that for all t and s

$$E\left[x_{is}\left(\Delta y_{it} - \alpha \Delta y_{i(t-1)} - \Delta x'_{it}\beta\right)\right] = 0. \tag{3}$$

Thus, the model generates internal moment conditions that, subject to a rank condition, will ensure identification in spite of serial correlation of unspecified form and the endogeneity of lagged y. Essen-

[1] This is an abridged version of Part III in Arellano (2003).

tially, we are exploiting the strict exogeneity of x in order to use lags and leads of x that do not have a direct effect on Δy_{it} as instruments for $\Delta y_{i(t-1)}$.

For example, if the model contains the contemporaneous and first lag of a scalar variable x and $T = 3$, we have three instruments x_1, x_2 and x_3 for the single equation in first differences

$$y_3 - y_2 = \alpha (y_2 - y_1) + \beta_0 (x_3 - x_2) + \beta_1 (x_2 - x_1) + (v_3 - v_2), \tag{4}$$

so that the coefficients α, β_0, β_1 are potentially just-identifiable from the moment conditions $E(x_{is} \Delta v_{i3}) = 0$, $(s = 1, 2, 3)$.

The models in this section should not be regarded as an extension of the pure autoregressive model. The purpose of AR models is to capture time series dependence, so that it is natural to start with serially uncorrelated errors. Here, however, lagged y appears in a structural role, and we consider models where its effect is identified regardless of serial correlation.

1.2 An Example: Cigarette Addiction

As an illustration, we consider Becker, Grossman, and Murphy (1994)'s analysis of cigarette consumption for US state panel data. The empirical model is

$$c_{it} = \theta c_{i(t-1)} + \beta\theta c_{i(t+1)} + \gamma p_{it} + \eta_i + \delta_t + v_{it} \tag{5}$$

where:

c_{it} = Annual per capita cigarette consumption in packs by state.

p_{it} = Average cigarette price per pack.

θ = A measure of the extent of addiction (for $\theta > 0$).

β = Discount factor.

Becker et al. are interested in testing whether smoking is addictive by considering the response of cigarette consumption to tax-induced exogenous changes in cigarette prices.

Equation (5) can be obtained as an approximation to the first-order conditions of utility maximization in a life-cycle model with certainty and habits, in which utility in period t depends on cigarette consumption in t and $t-1$. The degree of addiction is measured by θ, which will be positive if smoking is addictive. Furthermore, the price coefficient γ should be negative due to concavity of the utility. With certainty, the marginal utility of wealth is constant over time but not cross-sectionally. The state specific intercept η_i is intended to capture such variation, although according to the theory γ would also be a function of the marginal utility of wealth. Finally, the δ_t's represent aggregate shocks, possibly correlated with prices, which are treated as period specific parameters.

Equation (5) captures the fact that addictive behaviour implies that past consumption increases current consumption, holding the current price and the marginal utility of wealth fixed. Moreover, a rational addict will decrease current consumption in response to an anticipated decrease in future consumption.

The errors v_{it} represent unobserved life-cycle utility shifters, which are likely to be autocorrelated. Therefore, even in the absence of addiction ($\theta = 0$) and serial correlation in prices, we would expect to find dependence over time in c_{it}. As spelled out below, current consumption depends on prices in all periods through the effects of past and future consumption, but it is independent of past and future prices when $c_{i(t-1)}$ and $c_{i(t+1)}$ are held fixed. Exploiting this fact, the strategy of Becker et al. is to identify θ, β, and γ from the assumption that prices are strictly exogenous relative to the unobserved utility shift variables, which enables them to use lagged and future prices as instrumental variables. The required exogenous variation in prices comes from changes in cigarette tax rates. A crucial ingredient of this identification arrangement is the assumption that agents are able to anticipate future prices without error.

Becker, Grossman, and Murphy use annual US state data over the period 1955-1985 ($N = 50$, $T = 31$). Price variation arising from differences in excise taxes on cigarettes across states and time is an essential source of exogenous variation in this exercise. In addition, thanks to the panel nature of the data, the aggregate component of the errors can be held fixed through the use of time dummies. For these reasons a similar exercise with aggregate time series data, although technically possible, would lack the empirical justification for using prices as instruments that the state-level analysis has. On the other hand, individual-level panel data, while potentially useful in characterizing heterogeneity in the degree of addiction, would not add identifying content to the model if the only source of exogenous price variation remained state-level differences in fiscal policies.

Relation to the Joint Process of Consumption and Prices Finally, it is instructive to consider the statistical nature of model (5) and its relation to the bivariate autoregressive representation of the joint process of c_{it} and p_{it}. Letting u_{it} be the composite error term in (5), and L and L^{-1} denote the lag and forward operators, respectively, the equation can be written as

$$\left(1 - \theta L - \beta\theta L^{-1}\right) c_{it} = \gamma p_{it} + u_{it} \tag{6}$$

or

$$\frac{\theta}{\lambda}\left(1 - \lambda L\right)\left(1 - \beta\lambda L^{-1}\right) c_{it} = \gamma p_{it} + u_{it} \tag{7}$$

where λ is the stable root of the equation $\beta\theta\lambda^2 - \lambda + \theta = 0$. Thus, we can express current consumption as a function of past and future prices of the form

$$c_{it} = \gamma \sum_{j=-\infty}^{\infty} \psi_j p_{i(t+j)} + \sum_{j=-\infty}^{\infty} \psi_j u_{i(t+j)} \tag{8}$$

where the coefficients ψ_j are functions of θ and β.

Equation (8) is a regression of c_{it} on past and future prices. Becker et al.'s model is effectively placing a set of restrictions on the coefficients of this regression. Aside from stationarity, the error process is left unrestricted, and so is the price process. In conjunction with univariate processes for u and p one can obtain the autoregressive or moving average representations of the joint process of c_{it} and p_{it}.

It is interesting to note that while p is a strictly exogenous variable relative to v in equation (5), it is nevertheless Granger-caused by c. What is meant by this is that, regardless of the form of the univariate process of p, as long as $\psi_j \neq 0$ for some $j > 0$ the projection of p on lagged p and lagged c will have nonzero coefficients on some lagged c. Therefore, p would not be described as "strictly exogenous" in the sense of Sims (1972), even if it is strictly exogenous relative to u in model (5). Granger non-causality and Sims strict exogeneity will only occur if $\beta = 0$, in which case $\psi_j = 0$ for all $j > 0$, which corresponds to the model with "myopic habits" also considered by Becker et al.

1.3 GMM Estimation

1.3.1 2SLS Estimation from a Large T Perspective

Becker et al. treated the individual effects in (5) as parameters to be jointly estimated with the remaining coefficients. They employed 2SLS estimators using $p_{i(t-1)}$, p_{it}, $p_{i(t+1)}$, and individual and time dummies as instrumental variables. This is a natural perspective given the sample size of the state panel they used where $N = 50$ and $T = 31$. It is nevertheless useful to relate this type of estimators to estimators in deviations in order to exhibit some equivalences and the connection with the perspective adopted in small T large N environments.

Let the stacked form of a generic model that includes (1) and (5) as special cases be

$$y = W\delta + C\eta + v \tag{9}$$

where η is an $N \times 1$ vector of individual effects and $C = I_N \otimes \iota_T$ is an $NT \times N$ matrix of individual dummies. The NT-row matrix of explanatory variables W will contain observations of $y_{i(t-1)}$ and x_{it} in model (1), and of $c_{i(t-1)}$, $c_{i(t+1)}$, p_{it} and time dummies in model (5). Moreover, let Z be an NT-row matrix of instruments with at least as many columns as W. In a version of Becker's et al. example Z contains $p_{i(t-1)}$, p_{it}, $p_{i(t+1)}$ and time dummies, whereas in model (1) it may contain observations of x_{it}, $x_{i(t-1)}$, ..., $x_{i(t-j)}$ for some given j, and the actual value of T will be adjusted accordingly.

A 2SLS estimator of (δ, η) in (9) using Z and the individual dummies as instruments, $Z^{\dagger} = (Z, C)$ say, is given by

$$\begin{aligned} \begin{pmatrix} \widetilde{\delta} \\ \widetilde{\eta} \end{pmatrix} &= \left[W^{\dagger\prime} Z^{\dagger} \left(Z^{\dagger\prime} Z^{\dagger} \right)^{-1} Z^{\dagger\prime} W^{\dagger} \right]^{-1} W^{\dagger\prime} Z^{\dagger} \left(Z^{\dagger\prime} Z^{\dagger} \right)^{-1} Z^{\dagger\prime} y \\ &\equiv \left(\widehat{W}^{\dagger\prime} \widehat{W}^{\dagger} \right)^{-1} \widehat{W}^{\dagger\prime} y \end{aligned} \tag{10}$$

where $W^{\dagger} = (W, C)$ and $\widehat{W}^{\dagger} = \left(\widehat{W}, C \right)$ denotes the fitted value of $W^{\dagger}$ in a regression on (Z, C).

The estimator $\widetilde{\delta}$ is numerically the same as the *within-group 2SLS estimator* based on all variables y, W and Z in deviations from time means:

$$\widetilde{\delta}=\left[W'\overline{Q}Z\left(Z'\overline{Q}Z\right)^{-1}Z'\overline{Q}W\right]^{-1}W'\overline{Q}Z\left(Z'\overline{Q}Z\right)^{-1}Z'\overline{Q}y \tag{11}$$

where $\overline{Q}=I_N\otimes Q$ and Q is the within-group operator.

We can see this by taking into account that $\overline{Q}=I_{NT}-C\left(C'C\right)^{-1}C$ and using the result from partitioned regression:

$$\widetilde{\delta}=\left(\widehat{W}'\overline{Q}\widehat{W}\right)^{-1}\widehat{W}'\overline{Q}y. \tag{12}$$

Since $\widehat{W}$ is the fitted value in a regression of W on Z and C, it turns out that $\overline{Q}\widehat{W}$ is the fitted value in a regression of $\overline{Q}W$ on $\overline{Q}Z$, which is given by $\overline{Q}\widehat{W}=\overline{Q}Z\left(Z'\overline{Q}Z\right)^{-1}Z'\overline{Q}W$. Substituting this expression in (12), the equivalence with (11) follows.

The estimator $\widetilde{\delta}$ will be consistent as $T\rightarrow\infty$ as long as $E\left(z_{it}v_{it}\right)=0$, so that it will retain time-series consistency even if z_{it} is only predetermined. Consistency as $N\rightarrow\infty$ for fixed T, however, requires that

$$E\left[\left(z_{it}-\overline{z}_i\right)\left(v_{it}-\overline{v}_i\right)\right]=0. \tag{13}$$

Such condition will be satisfied if z_{it} is strictly exogenous for v_{it} as it is the case in both models (1) and (5). In what follows we consider the estimation of model (1) from a small T, large N perspective. This will let us provide further discussion of the link with the within-group 2SLS estimator (11).

1.3.2 Optimal IV Estimation in a Small T, Large N Context

Let us rewrite model (1)-(2) with an explicit intercept as

$$y_{it} = \gamma_0+w'_{it}\gamma_1+u_{it}=\left(1,w'_{it}\right)\gamma+u_{it} \tag{14}$$

$$u_{it} = \eta_i+v_{it} \tag{15}$$

$$E\left(v_{it}\mid x_i,\eta_i\right)=0\ \ \left(t=1,...,T\right), \tag{16}$$

where $w_{it}=\left(y_{i(t-1)},x'_{it}\right)'$, $\gamma_1=\left(\alpha,\beta'\right)'$, and γ_0 denotes a constant term, so that the individual effect η_i has been redefined to have zero mean. Also $\gamma=\left(\gamma_0,\gamma'_1\right)'$ and $x_i=\left(x'_{i0},...,x'_{iT}\right)'$. Moreover, for notational convenience we assume that y_{i0} and x_{i0} are observed. Thus, we start with a panel with $T+1$ time series observations, but we only have T observations of the vector w_{it} for individual i.

Letting $v_i=\left(v_{i1},...,v_{iT}\right)'$, $y_i=\left(y_{i1},...,y_{iT}\right)'$, and $W_i=\left(w_{i1},...,w_{iT}\right)'$, the unfeasible optimal instrumental-variable estimator of γ_1 based on the conditional moment restriction for the errors in orthogonal deviations $E\left(v_i^*\mid x_i\right)=0$ is

$$\widehat{\gamma}_{1UIV}=\left[\sum_{i=1}^{N}E\left(W_i^{*\prime}\mid x_i\right)\Omega^{-1}\left(x_i\right)W_i^*\right]^{-1}\sum_{i=1}^{N}E\left(W_i^{*\prime}\mid x_i\right)\Omega^{-1}\left(x_i\right)y_i^* \tag{17}$$

where $\Omega(x_i) = Var(v_i^* \mid x_i)$, $v_i^* = Av_i$, $y_i^* = Ay_i$, $W_i^* = AW_i$ and A denotes the $(T-1)\times T$ orthogonal deviations transformation matrix. A feasible counterpart requires estimates of $E(W_i^* \mid x_i)$ and $\Omega(x_i)$.

The within-group 2SLS estimator corresponding to (11) can be written as

$$\widetilde{\gamma}_1 = \left(\sum_{i=1}^{N} \widetilde{W}_i^{*\prime} W_i^*\right)^{-1} \sum_{i=1}^{N} \widetilde{W}_i^{*\prime} y_i^* \tag{18}$$

where

$$\widetilde{W}_i^* = Z_i^* \widetilde{\Pi} \equiv Z_i^* \left(\sum_{i=1}^{N} Z_i^{*\prime} Z_i^*\right)^{-1} \left(\sum_{i=1}^{N} Z_i^{*\prime} W_i^*\right) \tag{19}$$

and Z_i^* is a matrix of instruments in orthogonal deviations. Thus, $\widetilde{\gamma}_1$ is an estimator of the form of (17) with $\widetilde{W}_i^*$ and an identity matrix in place of $E(W_i^{*\prime} \mid x_i)$ and $\Omega(x_i)$, respectively.

For example, if Z_i^* contains x_{it}^* and $x_{i(t-1)}^*$, the instruments used by $\widetilde{\gamma}_1$ consist of the sample counterparts of the linear projections

$$E^*\left(w_{it}^* \mid x_{it}^*, x_{i(t-1)}^*\right) = \pi_0' x_{it}^* + \pi_1' x_{i(t-1)}^* \tag{20}$$

with the same coefficients for all t. However, in a panel with large N and small T we may consider estimators based on less restrictive projections of the form

$$E^*\left(w_{it}^* \mid x_{i0}, ..., x_{iT}\right) = \pi_{t0}' x_{i0} + ... + \pi_{tT}' x_{iT} \tag{21}$$

with unrestricted coefficients for each t. In contrast with (20), the projection (21) not only depends on all lags and leads of x, but also the coefficients are period specific. Naturally, in a large T, fixed N environment it would not be possible to obtain consistent estimates of the coefficients of (21) without restrictions. Next we turn to consider estimators of this kind.

1.3.3 GMM with the Number of Moments Increasing with T

Let us consider GMM estimators based on the moment conditions $E(v_i^* \otimes x_i) = 0$. Letting $Z_i = I_{(T-1)} \otimes x_i'$, these estimators take the form

$$\widehat{\gamma}_{1GMM} = \left[\left(\sum_i W_i^{*\prime} Z_i\right) A_N \left(\sum_i Z_i' W_i^*\right)\right]^{-1} \left(\sum_i W_i^{*\prime} Z_i\right) A_N \left(\sum_i Z_i' y_i^*\right), \tag{22}$$

where A_N is a weight matrix that needs to be chosen.

The model can be regarded as an incomplete system of $(T-1)$ simultaneous equations with T endogenous variables $y_{i0}^*, ..., y_{i(T-1)}^*$, a vector of instruments given by x_i, and cross-equation restrictions (since the same coefficients occur in the equations for different periods). From this perspective, a 2SLS estimator of the system uses

$$A_N = \left(\sum_i Z_i' Z_i\right)^{-1} = I_T \otimes \left(\sum_i x_i x_i'\right)^{-1}. \tag{23}$$

The difference between this 2SLS estimator and (18) is that the latter uses (20) to form predictions of w_{it}^*, whereas the former uses period-specific projections on all lags and leads as in (21).

Similarly, a three-stage least squares estimator (3SLS) is based on

$$A_N = \left(\sum_i Z_i' \widetilde{\Omega} Z_i \right)^{-1} = \widetilde{\Omega}^{-1} \otimes \left(\sum_i x_i x_i' \right)^{-1} \tag{24}$$

where $\widetilde{\Omega} = N^{-1} \sum_i \widetilde{v}_i^* \widetilde{v}_i^{*\prime}$ and the $\widetilde{v}_i^*$ are 2SLS residuals. Finally, a weight matrix that is robust to both heteroskedasticity and serial correlation is

$$A_N = \left(\sum_i Z_i' \widetilde{v}_i^* \widetilde{v}_i^{*\prime} Z_i \right)^{-1}. \tag{25}$$

The latter gives rise to Chamberlain's (1984) robust generalization of 3SLS for model (1).

Estimating the intercept Having estimated γ_1, a consistent estimate of γ_0 can be obtained as

$$\widehat{\gamma}_0 = \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - w_{it}' \widehat{\gamma}_{1GMM} \right). \tag{26}$$

Alternatively, we may consider the joint estimation of γ_0 and γ_1 by GMM from the moment conditions

$$E \left(y_i - \iota_T \gamma_0 - W_i \gamma_1 \right) = 0 \tag{27}$$

$$E \left[Z_i' \left(y_i^* - W_i^* \gamma_1 \right) \right] = 0, \tag{28}$$

or equivalently

$$E \left[\begin{pmatrix} I_T & 0 \\ 0 & Z_i \end{pmatrix}' \begin{pmatrix} u_i \\ u_i^* \end{pmatrix} \right] = E \left(Z_i^{\dagger\prime} u_i^\dagger \right) = 0, \tag{29}$$

where $u_i^\dagger = H^\dagger u_i$ and $H^\dagger = (I_T, A')'$. This leads to estimators of the form

$$\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{pmatrix} = \left[\left(\sum_i W_i^{\dagger\prime} Z_i^\dagger \right) A_N^\dagger \left(\sum_i Z_i^{\dagger\prime} W_i^\dagger \right) \right]^{-1} \left(\sum_i W_i^{\dagger\prime} Z_i^\dagger \right) A_N^\dagger \left(\sum_i Z_i^{\dagger\prime} y_i^\dagger \right) \tag{30}$$

where $y_i^\dagger = H^\dagger y_i$ and $W_i^\dagger = H^\dagger \left(\iota_T, W_i \right)$.

Expression (30) is a "levels & differences" or "system" estimator in the sense of combining moment conditions for errors in levels and deviations. The same argument can be used to estimate coefficients on time-invariant explanatory variables that are assumed to be uncorrelated with the effects.

1.3.4 Explanatory Variables Uncorrelated with the Effects

The previous setting also suggests a generalization to a case where a subset of the x's are uncorrelated with the effects. Suppose that a subset $x_{1i} = (x'_{1i1}, ..., x'_{1iT})'$ of x_i are uncorrelated with the effects whereas the remaining x_{2i} are correlated, as in static Hausman-Taylor models. In such case we obtain a GMM estimator of the same form as (30), but using an augmented matrix of instruments given by

$$Z_i^\dagger = \begin{pmatrix} I_T \otimes (1, x'_{1i}) & 0 \\ 0 & I_{(T-1)} \otimes x'_{2i} \end{pmatrix}. \tag{31}$$

If all x's are uncorrelated with the effects, the second block of moments in orthogonal deviations drops out and we are just left with the moments in levels

$$E\left(u_i \otimes x_{\ell i}\right) = E\left(Z'_{\ell i} u_i\right) = E\left[Z'_{\ell i}\left(y_i - W_{\ell i}\gamma\right)\right] = 0 \tag{32}$$

where $Z_{\ell i} = (I_T \otimes x'_{\ell i})$, $x_{\ell i} = (1, x'_i)'$ and $W_{\ell i} = (\iota_T, W_i)$.

1.3.5 Enforcing Restrictions in the Covariance Matrix

Let the marginal covariance matrix of u_i be $E\left(u_i u'_i\right) = \Omega$. The general form of Ω is

$$\Omega = \sigma_\eta^2 \iota_T \iota'_T + E\left(v_i v'_i\right), \tag{33}$$

which is not restrictive if $E\left(v_i v'_i\right)$ is unrestricted. However, we may consider restricting Ω by restricting the time series properties of v_i. For example, by considering the standard error components structure

$$\Omega = \sigma_\eta^2 \iota_T \iota'_T + \sigma^2 I_T, \tag{34}$$

or some other ARMA process for v_i, so that Ω can be expressed as a function of a smaller set of coefficients $\Omega\left(\theta\right)$.

We may consider estimating γ imposing the constraints in Ω. Even if Ω does not depend directly on γ, enforcing the covariance restrictions in Ω will in general lead to more efficient but less robust estimates of γ. This is so because the model is a simultaneous system as opposed to a multivariate regression. It may also help identification in a way that we shall pursue below.

The set of moments is now

$$E\left(Z_i^{\dagger\prime} u_i^\dagger\right) = 0 \tag{35}$$

$$E\left\{vech\left[u_i u'_i - \Omega\left(\theta\right)\right]\right\} = 0 \tag{36}$$

and this can be used as the basis for covariance restricted GMM estimators for models with all, part, or none of the x's correlated with the effects, something that will be reflected in the choice of $Z_i^\dagger$.

It is possible that some of the covariance restrictions may be expressed as simple instrumental variable restrictions. For example, if v_{it} is serially uncorrelated then the error in orthogonal deviations in period t v^*_{it} is not only orthogonal to x_i but also to $y_{i0}, ..., y_{i(t-1)}$, which suggests the use of GMM estimators with an increasing number of instruments.

2 Predetermined Variables

Regressors in static fixed-effect models may be correlated to a time-invariant error component, but are strictly exogenous in the sense of being uncorrelated to past, present and future time-varying errors.

These models can be extended to structural equations with strictly exogenous instruments and endogenous regressors that can be correlated with time varying errors at all lags and leads. The previous section was devoted to a prominent special model of this kind. Namely, a dynamic model containing lags of the dependent variable and strictly exogenous explanatory variables. The lagged dependent variable was treated as an endogenous variable since we allowed for unrestricted error serial correlation, and the strict exogeneity assumption was key to identification.

These "all-or-nothing" settings may be too restrictive in a time series environment, since it is possible to imagine situations in which explanatory variables may be correlated with errors at certain periods but not others, and such patterns may provide essential information for identification.

A familiar example is the white-noise measurement error model in which mismeasured regressors are correlated with contemporaneous errors but not with lagged or future errors. Another example is an autoregressive model in which the lagged dependent variable is a regressor correlated to past errors but not to current or future errors. In the two examples, regressors are correlated with the effects.

In this section we consider models with time varying errors that are uncorrelated to current and lagged values of certain conditioning variables but not to their future values, so that they are *predetermined* with respect to time varying errors. Some of these variables may be explanatory variables or lags of them, but may also be external predetermined instruments. Moreover, the equation may contain explanatory endogenous variables whose lags may or may not be part of the conditioning set.

Autoregressive models and dynamic regressions with feedback are specific examples. The emphasis in this section, however, is in an incomplete model that specifies orthogonality conditions between a structural error and predetermined instruments. In our context, predeterminedness is defined relative to a structural error. By "predetermined variables" we just refer to variables that are potentially correlated to lagged values of the structural error but are uncorrelated to present and future values.

An alternative approach to models with predetermined and/or endogenous regressors would be to consider complete systems. For example, VAR models or structural transformations of them. In this section, however, we have in mind situations in which a researcher is interested in modelling certain equations but not others. Thus, we focus in incomplete models with unspecified feedback processes, and restrictions specifying that errors are mean independent to certain variables in a sequential way.

2.1 Introduction and Examples

The previous section considered econometric models with lagged dependent variables whose errors were mean independent of past and future values of certain variables z_{it}

$$E\left(v_{it} \mid z_{i1}, ..., z_{iT}\right) = 0, \tag{37}$$

and referred to these variables as strictly exogenous with respect to v_{it}. We now consider models whose errors satisfy *sequential moment conditions* of the form

$$E\left(v_{it} \mid z_{i1}, ..., z_{it}\right) = 0. \tag{38}$$

Autoregressive Processes An example of (38) is an AR model with individual effects:

$$v_{it} = y_{it} - \alpha y_{i(t-1)} - \eta_i$$

and $E\left(v_{it} \mid y_{i0}, ..., y_{i(t-1)}, \eta_i\right) = 0$, so that (38) is satisfied with $z_{it} = y_{i(t-1)}$.

In what follows we present other instances to illustrate the scope of sequential moment assumptions.

2.1.1 Partial Adjustment with Feedback

Another example is a sequential version of the partial adjustment model. The equation is as (1):

$$y_{it} = \alpha y_{i(t-1)} + x_{it}'\beta + \eta_i + v_{it} \tag{39}$$

but in this instance the errors are assumed to satisfy

$$E\left(v_{it} \mid y_i^{t-1}, x_i^t, \eta_i\right) = 0, \tag{40}$$

so that (39) is regarded as a parameterization of the regression function $E\left(y_{it} \mid y_i^{t-1}, x_i^t, \eta_i\right)$.

There are two main differences between the properties of v's and x's in this model and model (1)-(2). Firstly, (40) implies lack of autocorrelation in v_{it} since lagged v's are linear combinations of the variables in the conditioning set, whereas (2) by itself does not restrict the serial dependence of errors. Secondly, in contrast with (40), assumption (2) implies that y does not Granger-cause x, in the sense that forecasts of x_{it} given x_i^{t-1}, y_i^{t-1} and additive effects are not affected by y_i^{t-1}. Thus, (2) rules out the possibility of feedback from lagged y to current x's, whereas feedback is allowed in (40).

Without further restrictions the two models are not nested. This situation can be illustrated using the example with $T = 3$ given in (4):

$$\Delta y_3 = \alpha \Delta y_2 + \beta_0 \Delta x_3 + \beta_1 \Delta x_2 + \Delta v_3.$$

Under assumption (2) the coefficients α, β_0, β_1 are potentially just-identifiable from the moments

$$E\left[\begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{pmatrix} \Delta v_{i3}\right] = 0. \tag{41}$$

Under assumption (40) the three parameters are also potentially just-identifiable from

$$E\left[\begin{pmatrix} y_{i1} \\ x_{i1} \\ x_{i2} \end{pmatrix} \Delta v_{i3}\right] = 0, \tag{42}$$

but the two models only have two moment restrictions in common, which in this example are not enough to identify the three parameters.

2.1.2 Euler Equation for Household Consumption

The next example is an intertemporal consumption model with uncertainty as in Zeldes (1989), who used individual after-tax returns and food consumption from the PSID. Suppose each period t a family i chooses consumption c_{it} and portfolio shares to maximize the expected value of a time-separable life-cycle utility function. In the absence of liquidity constraints, optimal consumption must satisfy the following Euler equation

$$E_{t-1}\left[\left(\frac{1+r_{it}}{1+\delta_i}\right) U'_{it}\left(c_{it}\right)\right] = U'_{i(t-1)}\left(c_{i(t-1)}\right) \tag{43}$$

where $U'_{it}(.)$ denotes the marginal utility of consumption, $E_{t-1}(.)$ is a conditional expectation given information available at time $t-1$, δ_i is a household-specific rate of time preference, and r_{it} is the rate of return of a riskless asset. Equivalently, we can write

$$\frac{(1+r_{it})\, U'_{it}\left(c_{it}\right)}{(1+\delta_i)\, U'_{i(t-1)}\left(c_{i(t-1)}\right)} = 1 + \varepsilon_{it} \tag{44}$$

where ε_{it} is an expectational error that satisfies $E_{t-1}(\varepsilon_{it}) = 0$ and is therefore uncorrelated with information known to the consumer at time $t-1$.

Suppose the utility of consumption has constant relative risk aversion coefficient α

$$U_{it}\left(c_{it}\right) = \frac{c_{it}^{1-\alpha}}{1-\alpha} e^{\theta_{it}} \tag{45}$$

and θ_{it} captures differences in preferences across families and time, which are specified as

$$\theta_{it} = \beta' x_{it} + \phi_t + \zeta_i + \xi_{it}. \tag{46}$$

In Zeldes' model, the vector x_{it} contains age and family size variables, and ζ_i, ϕ_t and ε_{it} are, respectively, family, time, and residual effects. With this specification the log of (44) is given by

$$\ln(1+r_{it}) - \alpha \Delta \ln c_{it} - \beta' \Delta x_{it} - \delta_t - \eta_i = v_{it} + \ln(1+\varepsilon_{it}) \tag{47}$$

where $\eta_i = \ln(1+\delta_i)$, $\delta_t = \Delta\phi_t$ and $v_{it} = \Delta\xi_{it}$. Moreover, using a second-order Taylor approximation $\ln(1+\varepsilon_{it}) \simeq \varepsilon_{it} - \varepsilon_{it}^2/2$, we can write

$$\ln(1+\varepsilon_{it}) = E_{t-1}\ln(1+\varepsilon_{it}) + e_{it} \simeq -\frac{1}{2} Var_{t-1}(\varepsilon_{it}) + e_{it} \tag{48}$$

where by construction $E_{t-1}(e_{it}) = 0$. The conditional variance $Var_{t-1}(\varepsilon_{it})$ may contain additive individual and time effects that would be subsumed into δ_t and η_i, but otherwise it is assumed not to change with time $t-1$ variables. Hence, the basic empirical equation becomes

$$\ln(1 + r_{it}) = \alpha \Delta \ln c_{it} + \beta' \Delta x_{it} + \delta_t + \eta_i + u_{it} \tag{49}$$

$$u_{it} = v_{it} + e_{it}. \tag{50}$$

Thus, the equation's error term u_{it}, net of individual and time effects, is made of two components: the unobservable change in tastes v_{it} and the expectational error e_{it}.[2]

In this model both returns and consumption growth at time t can be correlated with e_{it}, and are therefore treated as endogenous variables. Zeldes' identifying assumption is of the form of (38) with the instrument vector z_{it} containing Δx_{it}, lagged income, and marginal tax rates.

Lagged e's should be uncorrelated to current e's as long as they are in the agents' information sets, but v's may be serially correlated (unless ξ is a random walk). So serial correlation in u cannot be ruled out. Moreover, lagged consumption may be correlated to v even if it is not correlated to e. Thus, the presence of the taste shifter v rules out consumption lags as instruments. Lack of orthogonality between lagged c and u in this model does not necessarily imply a violation of the Euler condition.

Finally, note that the model's martingale property $E_{t-1}e_{it} = 0$ implies that for a variable z_{it} in the information set a time average of the form

$$\frac{1}{T} \sum_{t=1}^{T} z_{it} e_{it} \to 0$$

as $T \to \infty$. However, the cross-sectional average $N^{-1} \sum_{i=1}^{N} z_{it} e_{it}$ need not converge to zero as $N \to \infty$ if e_{it} contains aggregate shocks (cf. Chamberlain, 1984). The cross-sectional limit will only vanish if the e's are independent idiosyncratic shocks, what in model (49) requires us to assume that aggregate shocks affect all households in the same way and can be captured by time dummies.

2.1.3 Cross-Country Growth and Convergence

Our last example is a cross-country panel equation of growth determinants as in Caselli, Esquivel, and Lefort (1996). They estimated five 5-year period growth equations for 97 countries of the form

$$y_{it} - y_{i(t-5)} = \beta y_{i(t-5)} + s'_{i(t-5)} \gamma + \overline{f}'_{i(t-5)} \delta + \xi_t + \eta_i + v_{it} \tag{51}$$

$$(t = 1965, 1970, 1975, 1980, 1985)$$

where y_{it} denotes log per-capita GDP in country i in year t, $s_{i(t-5)}$ is a vector of stock variables measured in year $t-5$ (including human capital indicators such as the log- secondary school enrollment rate), $\overline{f}_{i(t-5)}$ is a vector of flow variables measured as averages from $t-5$ to $t-1$: $\overline{f}_{i(t-5)} =$

[2] Without the approximation, the model is nonlinear in variables and it has multiplicative individual effects.

$\left(f_{i(t-5)} + ... + f_{i(t-1)}\right)/5$ (such as rates of investment, population growth, and government expenditure), ξ_t captures global shocks, and η_i is a country effect that may represent differences in technology.

Caselli, Esquivel and Lefort completed the specification with the following identifying assumptions:

$$E\left(v_{i1965} \mid y_{i1960}, s_{i1960}\right) = 0 \tag{52}$$

$$E\left(v_{it} \mid y_{i1960}, s_{i1960}, .., y_{i(t-5)}, s_{i(t-5)}, \overline{f}_{i1960}, .., \overline{f}_{i(t-10)}\right) = 0 \tag{53}$$

$$(t = 1970, ..., 1985).$$

That is, the country-specific time-varying shock in year t, v_{it}, is uncorrelated to stock variables dated $t-5$ or earlier, including the output variable $y_{i(t-5)}$, and to the average flow variables dated $t-10$ or earlier. Thus, stock variables dated $t-5$ are treated as predetermined for v_{it}, whereas flow variables dated $t-5$ are predetermined for $v_{i(t+5)}$, and the assumptions restrict the dependence over time of v_{it}. Moreover, all explanatory variables are allowed to be correlated with the country effects η_i.

Islam (1995) considered an equation similar to (51) and allowed for correlation between country effects and the determinants of growth in s and f, but treated them as strictly exogenous for v. Islam's identifying assumptions did not restrict the form of serial correlation in v. However, as argued by Caselli et al., except for indicators of a country's geography and climate, strict exogeneity assumptions do not seem very useful in the growth context. Variables like the investment rate or the population growth rate are potentially both effects and causes of economic growth.

The Caselli et al. (1996) and Islam (1995) models illustrate our discussion of the contrast between partial adjustment equations with and without strictly exogenous variables in (1)-(2) and (39)-(40).

If β is negative, equation (51) describes the convergence of output from an initial level to a steady-state level, and it is broadly consistent with a variety of neoclassical growth models. The variables s, f and η are therefore interpreted as determinants of a country's steady state level of income. The smaller the value of β the faster the (conditional) convergence to the steady state. If on the contrary $\beta = 0$ then there is no convergence, and s, f and η measure differences in steady-state growth rates.

Panel data analyses were preceded by cross-sectional analyses in which growth over the 25 year interval 1960-1985 was related to 1960 income and other determinants. These studies typically found a negative but small effect of initial income, that implied a convergence rate in the range of 2-3 percent.

Caselli et al. and others studied panels over 5 or 10 year subperiods as a way of controlling for unobserved constant differences in steady state growth, and lack of exogeneity of observed determinants. Growth over long time spans of at least 5 years is chosen to abstract from business cycle effects.

The GMM panel estimates of β reported by Caselli, Esquivel and Lefort are also negative but of a larger magnitude than the cross-sectional estimates, implying a 10 percent convergence rate.[3] An open question is the extent to which these estimates are affected by GMM finite sample biases.

[3] The coefficient β can be regarded as approximating $-\left(1 - e^{-\lambda\tau}\right)$, where λ is the convergence rate and $\tau = 5$, so that $\lambda \simeq -(1/\tau)\ln(1+\beta)$.

2.2 Large T Within-Group Estimation

Let us consider a generic linear model that includes the previous examples as special cases:

$$y_{it} = w_{it}'\delta + \eta_i + v_{it}, \tag{54}$$

so that the errors satisfy condition (38), which for convenience we rewrite as

$$E\left(v_{it} \mid z_i^t\right) = 0 \tag{55}$$

for some vector z_{it} and $z_i^t = (z_{i1}, ..., z_{it})$.

If T is large, the sample realizations of the effects η_i may be treated as parameters that are jointly estimated with the common parameter vector δ. Thus, if all the variables in w_{it} are predetermined (i.e. if the w_{it} are functions of the variables in the conditioning set z_i^t), the within-group OLS estimator of δ is consistent as $T \to \infty$:

$$\widehat{\delta}_{WG} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T-1} w_{it}^{*} w_{it}^{*\prime}\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T-1} w_{it}^{*} y_{it}^{*} \tag{56}$$

where as usual starred variables denote forward orthogonal deviations. An example in which all w_{it} are predetermined is the partial adjustment regression model (39)-(40).

Large T consistency of $\widehat{\delta}_{WG}$ hinges on the condition $E\left(w_{it}v_{it}\right) = 0$. This estimator, however, is not unbiased for fixed T since $E\left(w_{it}^{*}v_{it}^{*}\right) \neq 0$. Moreover, the bias does not tend to zero as N increases. Intuitively, as $N \to \infty$ the cross-sectional sample average

$$\frac{1}{N}\sum_{i=1}^{N} w_{it}^{*} v_{it}^{*}$$

approaches $E\left(w_{it}^{*}v_{it}^{*}\right)$, but it does not get closer to zero for a given value of T. Therefore, $\widehat{\delta}_{WG}$ is inconsistent in a fixed T, large N setting (as in a pure autoregressive model).

If some of the w_{it} are endogenous explanatory variables (that is, if they are not functions of z_i^t), subject to identification, a within-group 2SLS estimator of the type described in (11) or (18) is consistent for large T:

$$\widehat{\delta}_{WG,2SLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T-1} \widehat{w}_{it}^{*} w_{it}^{*\prime}\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T-1} \widehat{w}_{it}^{*} y_{it}^{*}. \tag{57}$$

The vector $\widehat{w}_{it}^{*}$ denotes the estimated linear projection of w_{it}^{*} on $z_{it}^{*}, ..., z_{i(t-J)}^{*}$ for some pre-specified maximum lag value J:

$$\widehat{w}_{it}^{*} = \widehat{\pi}_0' z_{it}^{*} + ... + \widehat{\pi}_j' z_{i(t-J)}^{*} \;\; (t = 1, ..., T-1). \tag{58}$$

Examples with both endogenous and predetermined explanatory variables are the consumption Euler equation and cross-country growth models.

In parallel with the previous case, large T consistency of $\widehat{\delta}_{WG,2SLS}$ requires that $E\left(z_{i(t-j)}v_{it}\right)=0$ for $j=1,..,J$, a condition that is satisfied under the model's assumptions. But fixed T large N consistency would require $E\left(z^{*}_{i(t-j)}v^{*}_{it}\right)=0$, which does not hold because z_{it} is only a predetermined variable relative to v_{it}.

2.3 Small T GMM Estimation

Moments and Weight Matrices GMM estimators of δ in (54) based on the moments for the errors in first differences

$$E\left[z_i^t\left(v_{i(t+1)}-v_{it}\right)\right]=0 \;\; (t=1,...,T-1) \tag{59}$$

or in orthogonal deviations

$$E\left(z_i^t v_{it}^{*}\right)=0 \;\; (t=1,...,T-1) \tag{60}$$

are consistent for large N and fixed T. The orthogonality between v_{it}^{*} and z_i^t is due to the fact that v_{it}^{*} is a forward deviation that only depends on current and future values of the errors.

A generic compact expression that encompasses (59) and (60) is

$$E\left(Z_i'Ku_i\right)\equiv E\left[Z_i'K\left(y_i-W_i\delta\right)\right]=0 \tag{61}$$

where Z_i is a block-diagonal matrix whose t-th block is given by $z_i^{t\prime}$, $y_i=(y_{i1},...,y_{iT})'$, $W_i=(w_{i1}',...,w_{iT}')'$, $u_i=(u_{i1},...,u_{iT})'$, and $u_{it}=\eta_i+v_{it}$. Moreover, K represents any $(T-1)\times T$ upper-triangular transformation matrix of rank $(T-1)$, such that $K\iota=0$, where ι is a $T\times 1$ vector of ones.

Orthogonality between K and ι ensures that the transformation eliminates any fixed effects, whereas by being upper triangular the transformed vector of errors Ku_i may depend on present and future errors but not on lagged ones. Note that both the first-difference matrix operator and the forward orthogonal deviations operator satisfy these requirements.

The form of a GMM estimator of δ based on (61) is

$$\widehat{\delta}=\left[\left(\sum_{i=1}^{N}W_i'K'Z_i\right)A_N\left(\sum_{i=1}^{N}Z_i'KW_i\right)\right]^{-1}\left(\sum_{i=1}^{N}W_i'K'Z_i\right)A_N\left(\sum_{i=1}^{N}Z_i'Ky_i\right). \tag{62}$$

Given identification, $\widehat{\delta}$ is consistent and asymptotically normal as $N\rightarrow\infty$ for fixed T. An optimal choice of the weight matrix A_N is a consistent estimate of the inverse of $E\left(Z_i'Ku_iu_i'K'Z_i\right)$ up to scale.

We consider three alternative estimators depending on the choice of A_N. Firstly, a one-step weight matrix given by

$$A_N=\left(\sum_{i=1}^{N}Z_i'KK'Z_i\right)^{-1}. \tag{63}$$

This choice is optimal under conditional homoskedasticity $E(v_{it}^2|z_i^t) = \sigma^2$ and lack of autocorrelation $E(v_{it}v_{i(t+j)}|z_i^{t+j}) = 0$ for $j > 0$, since in this case we have

$$E\left(Z_i' K v_i v_i' K' Z_i\right) = \sigma^2 E\left(Z_i' K K' Z_i\right). \tag{64}$$

A second, more general, two-step choice of A_N is given by

$$A_N = \left(\sum_{i=1}^{N} Z_i' K \widetilde{\Omega} K' Z_i\right)^{-1}, \tag{65}$$

where $\widetilde{\Omega} = N^{-1}\sum_{i=1}^{N} \widetilde{u}_i \widetilde{u}_i'$, and $\widetilde{u}_i$ is a vector of one-step residuals.[4] This choice is optimal if conditional variances and autocovariances are constant:

$$\begin{aligned} E(v_{it}^2|z_i^t) &= E(v_{it}^2) && (66)\\ E\left(v_{it}v_{i(t+j)}|z_i^{t+j}\right) &= E\left(v_{it}v_{i(t+j)}\right). && (67) \end{aligned}$$

Thirdly, the standard two-step robust choice is

$$A_N = \left(\sum_{i=1}^{N} Z_i' K \widetilde{u}_i \widetilde{u}_i' K' Z_i\right)^{-1}, \tag{68}$$

which is an optimal weight matrix even if the conditional variances and autocovariances of the errors are not constant. In contrast, the two-step estimator based on (65) does not depend on the data fourth-order moments but is asymptotically less efficient than the estimator that uses (68) unless (66) and (67) are satisfied, in which case they are asymptotically equivalent (Arellano and Bond, 1991). GMM estimators of autoregressive models can be regarded as a special case.

It is also useful to compare the estimator in (22) for dynamic models with strictly exogenous variables with that in (62). In both cases we have a system of transformed equations for different periods, but while in (22) the same instruments are valid for all equations (as in a standard simultaneous equations model), in (62) different instruments are valid for different equations, and we have an increasing set of instruments available as time progresses. Clearly, it is possible to combine features of the two settings (e.g. a model containing both predetermined and strictly exogenous variables).

The Irrelevance of Filtering A GMM estimator of the form given in (62) is invariant to the choice of K, provided K satisfies the required conditions, and A_N depends on K as in (63), (65) or (68) (Arellano and Bover, 1995). A requirement of the invariance result is that all the available instruments are used each period, so that Z_i is a block-diagonal matrix with an increasing number of instruments per block as indicated above.

[4] Note that $\widetilde{\Omega}$ is a consistent estimate of $\Omega = E\left(u_i u_i'\right)$, but not of $E\left(v_i v_i'\right)$. However, $K\widetilde{\Omega}K'$ is a consistent estimate of $KE\left(v_i v_i'\right)K'$.

Despite the irrelevance of filtering, specific choices of K may be computationally advantageous. Recall that we obtained the forward orthogonal deviations operator as

$$A = (DD')^{-1/2} D$$

where D is the first-difference operator, so that $AA' = I_{(T-1)}$. Therefore, for $K = A$, the one-step weight matrix (63) simply becomes $\left(\sum_{i=1}^{N} Z_i' Z_i\right)^{-1}$. Hence, the one-step GMM estimator can be obtained as a matrix-weighted average of cross-sectional IV estimators:

$$\widehat{\delta} = \left(\sum_{t=1}^{T-1} W_t^{*\prime} Z_t \left(Z_t' Z_t\right)^{-1} Z_t' W_t^*\right)^{-1} \sum_{t=1}^{T-1} W_t^{*\prime} Z_t \left(Z_t' Z_t\right)^{-1} Z_t' y_t^*, \tag{69}$$

where $W_t^* = (w_{1t}^{*\prime}, ..., w_{Nt}^{*\prime})'$, $y_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$, and $Z_t = (z_1^{t\prime}, ..., z_N^{t\prime})'$. This is a useful computational feature of orthogonal deviations when T is not a very small number.

Similarly, for K given by

$$K = \left(D\widetilde{\Omega}D'\right)^{-1/2} D \tag{70}$$

where $\left(D\widetilde{\Omega}D'\right)^{-1/2}$ denotes the upper-triangular Cholesky decomposition of $\left(D\widetilde{\Omega}D'\right)^{-1}$, the two-step weight matrix (65) also becomes $\left(\sum_{i=1}^{N} Z_i' Z_i\right)^{-1}$. Thus, the corresponding two-step GMM estimator can also be written in the form of (69) after replacing orthogonal deviations by observations transformed according to (70) (Keane and Runkle, 1992).

2.4 Optimal Instruments

Let us consider the form of the information bound and the optimal instruments for model (54)-(55) in a small T context. Since $E(\eta_i | z_i^T)$ is unrestricted, all the information about δ is contained in $E(v_{it} - v_{i(t+1)} | z_i^t) = 0$ for $t = 1, ..., T-1$.

For a single period the information bound is

$$J_{0t} = E\left(\frac{d_{it} d_{it}'}{\omega_{it}}\right)$$

where $d_{it} = E(w_{it} - w_{i(t+1)} | z_i^t)$ and $\omega_{it} = E[(v_{it} - v_{i(t+1)})^2 | z_i^t]$ (cf. Chamberlain, 1987). Thus, for a single period the optimal instrument is $m_{it} = d_{it}/\omega_{it}$, in the sense that under suitable regularity conditions the unfeasible IV estimator

$$\widetilde{\delta}_{(t)} = \left(\sum_{i=1}^{N} m_{it} \Delta w_{i(t+1)}'\right)^{-1} \left(\sum_{i=1}^{N} m_{it} \Delta y_{i(t+1)}\right) \tag{71}$$

satisfies $\sqrt{N}\left(\widetilde{\delta}_{(t)} - \delta\right) \xrightarrow{d} \mathcal{N}\left(0, J_{0t}^{-1}\right)$.

If the transformed errors were conditionally serially uncorrelated, the total information would be the sum of the information bounds for each period. Forward orthogonal deviations was obtained as a filter applied to the differenced data that removed the moving average serial correlation induced by differencing when v_{it} is *iid*. Generalizing this idea we can obtain a forward filter that removes conditional serial correlation to arbitrary first-differenced errors that satisfy sequential moment restrictions. This is achieved by the following recursive transformation proposed by Chamberlain (1992):

$$\widetilde{v}_{i(T-1)} = v_{i(T-1)} - v_{iT}$$

$$\widetilde{v}_{it} = (v_{it} - v_{i(t+1)}) - \tau_{t1}\left(z_i^{t+1}\right)\widetilde{v}_{i(t+1)} - \tau_{t2}\left(z_i^{t+2}\right)\widetilde{v}_{i(t+2)} - \ldots - \tau_{t(T-t-1)}\left(z_i^{T-1}\right)\widetilde{v}_{i(T-1)} \tag{72}$$

for $t = T-2, \ldots, 1$, where

$$\tau_{tj}\left(z_i^{t+j}\right) = \frac{E[(v_{it} - v_{i(t+1)})\widetilde{v}_{i(t+j)}|z_i^{t+j}]}{E(\widetilde{v}^2_{i(t+j)}|z_i^{t+j})}. \tag{73}$$

The interest in this transformation is that it satisfies the same conditional moment restrictions as the original errors in first-differences, namely

$$E(\widetilde{v}_{it}|z_i^t) = 0, \tag{74}$$

but additionally it satisfies by construction the lack of dependence requirement:[5]

$$E(\widetilde{v}_{it}\widetilde{v}_{i(t+j)}|z_i^{t+j}) = 0 \text{ for } j = 1, \ldots, T-t-1. \tag{75}$$

Therefore, in terms of the transformed errors the information bound can be written as

$$J_0 = \sum_{t=1}^{T-1} E\left(\frac{\widetilde{d}_{it}\widetilde{d}'_{it}}{\widetilde{\omega}_{it}}\right) \tag{76}$$

where $\widetilde{d}_{it} = E\left(\widetilde{w}_{it}|z_i^t\right)$ and $\widetilde{\omega}_{it} = E\left(\widetilde{v}^2_{it}|z_i^t\right)$. The variables $\widetilde{w}_{it}$ and $\widetilde{y}_{it}$ denote the corresponding transformations to the first-differences of w_{it} and y_{it} such that $\widetilde{v}_{it} = \widetilde{y}_{it} - \widetilde{w}'_{it}\delta$.

Thus, the optimal instruments for all periods are $\widetilde{m}_{it} = \widetilde{d}_{it}/\widetilde{\omega}_{it}$, in the sense that under suitable regularity conditions the unfeasible IV estimator

$$\widetilde{\delta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T-1}\widetilde{m}_{it}\widetilde{w}'_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T-1}\widetilde{m}_{it}\widetilde{y}_{it}\right) \tag{77}$$

[5] To see that this is the kind of conditional lack of serial correlation that is required, notice that $E\left(\widetilde{v}_{it} \mid z_i^t\right) = 0$ and $E\left(\widetilde{v}_{i(t+j)} \mid z_i^{t+j}\right) = 0$ imply, respectively, that $E\left[h_t\left(z_i^t\right)\widetilde{v}_{it}\right] = 0$ and $E\left[h_{t+j}\left(z_i^{t+j}\right)\widetilde{v}_{i(t+j)}\right] = 0$ for any functions $h_t(.)$ and $h_{t+j}(.)$. If $E\left(\widetilde{v}_{it}\widetilde{v}_{i(t+j)} \mid z_i^{t+j}\right) = 0$ then

$$E\left[\widetilde{v}_{it}\widetilde{v}_{i(t+j)}h_t\left(z_i^t\right)h_{t+j}\left(z_i^{t+j}\right)\right] = 0,$$

so that the information bound for any $h_t(.)$ and $h_{t+j}(.)$ is the sum of the information bounds for each period.

satisfies $\sqrt{N}\left(\widetilde{\delta}-\delta\right) \xrightarrow{d} \mathcal{N}\left(0, J_0^{-1}\right)$.

The optimal IV estimator (77) is unfeasible on two accounts. Firstly, it uses a data dependent filter whose weights are unknown functions of the conditioning variables. Secondly, the optimal instruments $\widetilde{m}_{it}$ depend on unknown conditional expectations of the filtered data. A feasible estimator that achieves the bound could be potentially constructed by replacing the unknown functions with nonparametric regression estimators. Alternatively, one could use a GMM estimator based on an expanding set of instruments as N tends to infinity.

Lack of Serial Correlation Suppose that the original errors are conditionally serially uncorrelated so that

$$E\left(v_{it}v_{i(t+j)} \mid z_i^{t+j}\right) = 0.$$

This is, for example, a property of the partial adjustment regression model (39)-(40), but not necessarily of Zeldes' Euler equation (49) since unobserved taste changes may be serially correlated.

In this case the weights (73) are equal to zero for $j > 1$ and the optimal filter becomes

$$\widetilde{v}_{i(T-1)} = v_{i(T-1)} - v_{iT}$$

$$\widetilde{v}_{it} = \left(v_{it} - v_{i(t+1)}\right) + \frac{\sigma^2_{i(t+1)}}{\widetilde{\sigma}^2_{i(t+1)}}\widetilde{v}_{i(t+1)} \quad (t = T-2, ..., 1) \tag{78}$$

where $\sigma^2_{it} = E\left(v^2_{it} \mid z^t_i\right)$ and $\widetilde{\sigma}^2_{it} = E\left(\widetilde{v}^2_{it} \mid z^t_i\right)$.

If the conditional variances are constant but there is unconditional time series heteroskedasticity, so that $E(v^2_{it}|z^t_i) = \sigma^2_t$ and $E\left(\widetilde{v}^2_{it} \mid z^t_i\right) = \widetilde{\sigma}^2_t$, we have

$$\widetilde{v}_{it} = v_{it} - \frac{1}{\left(\sigma^{-2}_{(t+1)} + ... + \sigma^{-2}_T\right)}\left(\sigma^{-2}_{(t+1)}v_{i(t+1)} + ... + \sigma^{-2}_T v_{iT}\right) \quad (t = T-1, ..., 1) \tag{79}$$

and

$$\widetilde{\sigma}^2_t = \sigma^2_t + \frac{1}{\left(\sigma^{-2}_{(t+1)} + ... + \sigma^{-2}_T\right)}. \tag{80}$$

Other Special Cases If the v_{it}'s are conditionally homoskedastic and serially uncorrelated, so that $E(v^2_{it}|z^t_i) = \sigma^2$ and $E(v_{it}v_{i(t+j)}|z^{t+j}_i) = 0$ for $j > 0$, the $\widetilde{v}_{it}$'s blow down to ordinary forward orthogonal deviations:

$$\widetilde{v}_{it} = v_{it} - \frac{1}{(T-t)}(v_{i(t+1)} + ... + v_{iT}) \equiv \frac{1}{c_t}v^*_{it} \text{ for } t = T-1, ..., 1.$$

where $c_t^2 = (T-t)/(T-t+1)$. In such case the optimal instrument is $\widetilde{m}_{it} = c_t \sigma^{-2} E(w_{it}^*|z_i^t)$ so that

$$\widetilde{\delta} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T-1} E\left(w_{it}^*|z_i^t\right) w_{it}^{*\prime}\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T-1} E\left(w_{it}^*|z_i^t\right) y_{it}^* \tag{81}$$

$$= \left[\sum_{i=1}^{N} \mathcal{E}_{\mathcal{T}}\left(W_i^{*\prime}\right) W_i^*\right]^{-1} \sum_{i=1}^{N} \mathcal{E}_{\mathcal{T}}\left(W_i^{*\prime}\right) y_i^* \tag{82}$$

and

$$J_0 = \frac{1}{\sigma^2}\sum_{t=1}^{T-1} E[E(w_{it}^*|z_i^t)E(w_{it}^{*\prime}|z_i^t)] = \frac{1}{\sigma^2} E\left[\mathcal{E}_{\mathcal{T}}\left(W_i^{*\prime}\right)\mathcal{E}_{\mathcal{T}}\left(W_i^*\right)\right], \tag{83}$$

where we have introduced the notation

$$\mathcal{E}_{\mathcal{T}}\left(W_i^{*\prime}\right) = \left[E\left(w_{i1}^*|z_i^1\right), ..., E\left(w_{i(T-1)}^*|z_i^{T-1}\right)\right]. \tag{84}$$

If we further assume that the conditional expectations $E(w_{it}^*|z_i^t)$ are linear:

$$E(w_{it}^*|z_i^t) = \Pi_t z_i^t \tag{85}$$

with

$$\Pi_t = E(w_{it}^* z_i^{t\prime})[E(z_i^t z_i^{t\prime})]^{-1}, \tag{86}$$

then

$$J_0 = \frac{1}{\sigma^2}\sum_{t=1}^{T-1} E(w_{it}^* z_i^{t\prime})[E(z_i^t z_i^{t\prime})]^{-1} E(z_i^t w_{it}^{*\prime}) \tag{87}$$

which coincides with the inverse of the asymptotic covariance matrix of the standard GMM estimator (62) under the stated assumptions.

Note that the one-step GMM estimator is of the form of (81) with the unrestricted sample linear projection $\widehat{\Pi}_t z_i^t$ in place of $E(w_{it}^*|z_i^t)$. On the other hand, the within-group OLS and 2SLS estimators (56) and (57) are also of the form of (81), but in place of $E(w_{it}^*|z_i^t)$ they use w_{it}^* or $\widehat{w}_{it}^*$, neither of which are in the admissible set of instruments for fixed T, large N consistency.

If we assume that the conditional variances and autocovariances are constant, but we allow for constant autocorrelation and unconditional time series heteroskedasticity, as in (66) and (67), the $\widetilde{v}_{it}$ are equivalent to the population counterpart of the Keane and Runkle filter (70). Note that this filter can also be expressed as a GLS transformation of the errors in orthogonal deviations (instead of first-differences):

$$K = \left(A\Omega A'\right)^{-1/2} A. \tag{88}$$

Thus, in this case

$$\widetilde{\delta} = \left[\sum_{i=1}^{N} \mathcal{E}_{\mathcal{T}} \left(W_i^{*\prime} \right) \Omega^{*-1} W_i^* \right]^{-1} \sum_{i=1}^{N} \mathcal{E}_{\mathcal{T}} \left(W_i^{*\prime} \right) \Omega^{*-1} y_i^* \tag{89}$$

and

$$J_0 = E \left[\mathcal{E}_{\mathcal{T}} \left(W_i^{*\prime} \right) \Omega^{*-1} \mathcal{E}_{\mathcal{T}} \left(W_i^* \right) \right] \tag{90}$$

where $\Omega^* = E \left(u_i^* u_i^{*\prime} \right) = A \Omega A'$.

Note that we do not consider the question of the impact on the bound for δ of assuming conditional homoskedasticity, lack of serial correlation, and linearity of $E(w_{it}^* | z_i^t)$. Here, we have merely particularized the bound for δ based on $E(v_{it} | z_i^t) = 0$ to cases where the additional restrictions happen to occur in the population but are not used in the calculation of the bound.

2.5 Instruments Uncorrelated with the Effects

2.5.1 System Estimators

We have already discussed system estimators that combined moment restrictions in levels and deviations in various contexts. We considered models that contained a subset of strictly exogenous variables that were uncorrelated with the effects in both static and dynamic settings. Moreover, we also considered mean-stationary AR models in which lagged dependent variables in first differences were valid instruments for the equation errors in levels. The use of moment restrictions in levels may afford large information gains, but also biases if the restrictions are violated.

We now consider instrumental variables that are uncorrelated with the effects in the context of sequential moment conditions. Let us suppose as before that

$$y_{it} = \delta' w_{it} + \eta_i + v_{it}$$

$$E(v_{it} | z_i^t) = 0,$$

but in addition a subset z_{1it} of the predetermined instruments in z_{it} are assumed to be uncorrelated with the effects:

$$E \left[z_{1it} \left(\eta_i - \eta \right) \right] = 0 \tag{91}$$

where $\eta = E \left(\eta_i \right)$. The implication is that for $s \leq t$:

$$E \left[z_{1is} \left(\eta_i - \eta + v_{it} \right) \right] = 0. \tag{92}$$

Given the basic moment restrictions for the errors in first differences considered earlier

$$E \left[z_i^{t-1} \left(\Delta y_{it} - \delta' \Delta w_{it} \right) \right] = 0 \tag{93}$$

and the moments involving the levels' intercept η

$$E\left(y_{it}-\eta-\delta' w_{it}\right)=0, \tag{94}$$

assumption (91) adds the following restrictions in levels[6]

$$E\left[z_{1it}\left(y_{it}-\eta-\delta' w_{it}\right)\right]=0. \tag{95}$$

These two sets of moments can be combined using the Arellano & Bover (1995) GMM estimator. A generic compact expression for the full set of moments is

$$E\left[\left(\begin{array}{cc} Z_{\ell i} & 0 \\ 0 & Z_i \end{array}\right)'\left(\begin{array}{c} u_i \\ Ku_i \end{array}\right)\right]=E\left(Z_i^{\dagger\prime}u_i^{\dagger}\right)=E\left[Z_i^{\dagger\prime}\left(y_i^{\dagger}-W_i^{\dagger}\delta^{\dagger}\right)\right]=0 \tag{96}$$

where u_i, K, and Z_i are as in (61), $Z_{\ell i}$ is the following block-diagonal matrix

$$Z_{\ell i}=\left(\begin{array}{ccc} (1,z_{1i1}') & \dots & 0 \\ & \ddots & \\ 0 & & (1,z_{1iT}') \end{array}\right), \tag{97}$$

and $\delta^{\dagger}=\left(\delta',\eta\right)'$. With these changes, the form of the GMM estimator is similar to that of (62):

$$\widehat{\delta}^{\dagger}=\left[\left(\sum_{i=1}^{N}W_i^{\dagger\prime}Z_i^{\dagger}\right)A_N\left(\sum_{i=1}^{N}Z_i^{\dagger\prime}W_i^{\dagger}\right)\right]^{-1}\left(\sum_{i=1}^{N}W_i^{\dagger\prime}Z_i^{\dagger}\right)A_N\left(\sum_{i=1}^{N}Z_i^{\dagger\prime}y_i^{\dagger}\right). \tag{98}$$

The standard robust choice of A_N is the inverse of an unrestricted estimate of the variance matrix of the moments $N^{-1}\sum_{i=1}^{N}Z_i^{\dagger\prime}\widetilde{u}_i^{\dagger}\widetilde{u}_i^{\dagger\prime}Z_i^{\dagger}$ for some preliminary consistent residuals $\widetilde{u}_i^{\dagger}$.The difference in this case is that, contrary to GMM estimators based exclusively in moments for the errors in differences, an efficient one-step estimator under restrictive assumptions does not exist. Since $E\left(Z_i^{\dagger\prime}\widetilde{u}_i^{\dagger}\widetilde{u}_i^{\dagger\prime}Z_i^{\dagger}\right)$ depends on the errors in levels, at the very least it will be a function of the ratio of variances of η_i and v_{it}. Moreover, since some of the instruments for the equations in levels are not valid for those in differences, and conversely, not all the covariance terms between the two sets of moments will be zero.

2.5.2 Stationarity Restrictions

A leading case of uncorrelated predetermined instruments are first differences of predetermined instruments that exhibit constant correlation with the effects (Arellano and Bover, 1995). Suppose that we

[6] Note, for example, that $E\left[z_{1i(t-1)}\left(y_{it}-\eta-\delta' w_{it}\right)\right]=0$ is not an extra restriction since it follows from two existing restrictions:

$$E\left[z_{1i(t-1)}\left(y_{it}-\eta-\delta' w_{it}\right)\right]=E\left[z_{1i(t-1)}\left(\Delta y_{it}-\delta'\Delta w_{it}\right)\right]+E\left[z_{1i(t-1)}\left(y_{i(t-1)}-\eta-\delta' w_{i(t-1)}\right)\right].$$

can partition

$$z_{it} = \begin{pmatrix} z_{Ait} \\ z_{Bit} \end{pmatrix}$$

such that the variables in z_{Ait} have constant correlation with the effects over time, in which case

$$E\left(\eta_i \Delta z_{Ait}\right) = 0. \tag{99}$$

Constant correlation over time will be expected when z_{Ait} is conditionally stationary in mean (Blundell and Bond, 1998). That is, when

$$E\left(\Delta z_{Ait} \mid \eta_i\right) = 0. \tag{100}$$

This situation leads to a list of orthogonality conditions of the form (96) with $z_{1it} = \Delta z_{Ait}$, $z_{2i1} = z_{i1}$, and $z_{2it} = z_{Bit}$ for $t > 1$.

Blundell and Bond (2000) employed moment restrictions of this type in their empirical analysis of Cobb-Douglas production functions using firm panel data. They found that moment conditions for the production function in first differences of the form (93) were not very informative, due to the fact that firm output, capital and employment were highly persistent. In contrast, the first-difference instruments for production function errors in levels turned out to be informative and empirically valid.

2.5.3 Time-Invariant Explanatory Variables

Sometimes the effect of a vector of time-invariant explanatory variables is of interest, a parameter vector γ, say, in a model of the form

$$y_{it} = \delta' w_{it} + \gamma' \zeta_i + \eta_i + v_{it} \tag{101}$$

$$E\left(v_{it} \mid z_i^t\right) = 0$$

where for convenience we assume that ζ_i contains an intercept and therefore $E\left(\eta_i\right) = 0$ without lack of generality. If ζ_i is orthogonal to $(\eta_i + v_{it})$ then δ and γ can be jointly estimated using a system GMM estimator that combines the moments (93) with

$$E\left[\zeta_i \left(y_{it} - \delta' w_{it} - \gamma' \zeta_i\right)\right] = 0. \tag{102}$$

However, if the components of ζ_i (other than the intercept) are not orthogonal to $(\eta_i + v_{it})$, the corresponding elements of γ cannot be identified from the basic moments because the time-invariant explanatory variables are absorbed into the individual effect.

In principle, the availability of predetermined instruments that are uncorrelated with the effects might help to identify such parameters. Indeed, this was Hausman and Taylor (1981)'s motivation for considering uncorrelated strictly exogenous variables. However, moments in levels that are derived from stationarity restrictions, as in (99), are unlikely to serve that purpose. The problem is that if $E(\eta_i \Delta z_{Ait}) = 0$ holds we may expect that $E(\zeta_i \Delta z_{Ait}) = 0$ will also hold, in which case changes in z_{Ait} would not help the identification of γ.

2.5.4 Levels Moments Implied by Lack of Serial Correlation

The basic model with sequential moment conditions that we consider in this section (54)-(55) may or may not restrict the serial correlation properties of the errors depending on the nature of the variables in the conditioning set. Moreover, restriction (55) does not rule out the correlation between the errors v_{it} and the effects η_i. Sometimes, a natural starting point for an empirical analysis will be the stronger assumption

$$E\left(v_{it} \mid z_i^t, \eta_i\right) = 0 \tag{103}$$

which implies (55) and also lack of correlation between v_{it} and η_i. If z_{it} includes $y_{i(t-1)}$ and $w_{i(t-1)}$ then (103) implies that the v_{it} are not autocorrelated. Examples of this situation are the autoregressive and partial adjustment models (39)-(40).

The point we wish to make here is that if the v_{it} are serially uncorrelated and have constant (or lack of) correlation with the effects, then $\Delta v_{i(t-1)}$ can be used as an instrument for the errors in levels:

$$E\left(\Delta v_{i(t-1)} u_{it}\right) = 0. \tag{104}$$

In effect, we have

$$E\left(\Delta v_{i(t-1)} u_{it}\right) = E\left(\Delta v_{i(t-1)} \eta_i\right) + E\left(\Delta v_{i(t-1)} v_{it}\right) = 0$$

as long as the v's are not autocorrelated and $E\left(v_{it}\eta_i\right)$ does not vary with t.[7]

Condition (104) is the type of quadratic moment restriction considered by Ahn and Schmidt (1995) for autoregressive models. In principle, their use makes the estimation problem nonlinear, although an alternative is to evaluate $\Delta v_{i(t-1)}$ using a preliminary consistent estimator of δ, and use the resulting residual as an instrument for the levels equation as part of a system estimator.

Note that if $y_{i(t-1)}$ and $w_{i(t-1)}$ belong to z_{it} and they are mean stationary in such a way that

$$E\left[\begin{pmatrix} \Delta y_{i(t-1)} \\ \Delta w_{i(t-1)} \end{pmatrix} u_{it}\right] = 0, \tag{105}$$

then the moment condition (104) becomes redundant given (105). The reason is that since $\Delta v_{i(t-1)} = \Delta y_{i(t-1)} - \delta' \Delta w_{i(t-1)}$, (104) is a linear combination of (105).

References

[1] Ahn, S. and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data", *Journal of Econometrics*, 68, 5-27.

[7] If the v's are uncorrelated with the effects it is also true that $E\left(v_{i(t-1)} u_{it}\right) = 0$, but $v_{i(t-1)}$ cannot be used as an instrument since it depends on the unobservable individual effects.

[2] Arellano, M. (2003): *Panel Data Econometrics*, Oxford University Press, Oxford.

[3] Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, 277-297.

[4] Arellano, M. and O. Bover (1995): "Another Look at the Instrumental-Variable Estimation of Error-Components Models", *Journal of Econometrics*, 68, 29-51.

[5] Becker, G., M. Grossman, and K. Murphy (1994): "An Empirical Analysis of Cigarette Addiction", *American Economic Review*, 84, 396-418.

[6] Blundell, R. and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, 87, 115-143.

[7] Blundell, R. and S. Bond (2000): "GMM Estimation with Persistent Panel Data: An Application to Production Functions", *Econometric Reviews*, 19, 321-340.

[8] Caselli, F., G. Esquivel, and F. Lefort (1996): "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics", *Journal of Economic Growth*, 1, 363-389.

[9] Chamberlain, G. (1984): "Panel Data", in Griliches, Z. and M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, Elsevier Science, Amsterdam.

[10] Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, 305-334.

[11] Chamberlain (1992): Comment: Sequential Moment Restrictions in Panel Data", *Journal of Business & Economic Statistics*, 10, 20-26.

[12] Hausman, J. A. and W. E. Taylor (1981): "Panel Data an Unobservable Individual Effects", *Econometrica*, 49, 1377-1398.

[13] Islam, N. (1995): "Growth Empirics: A Panel Data Approach", *Quarterly Journal of Economics*, 110, 1127-1170.

[14] Keane, M. and D. Runkle (1992): "On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous", *Journal of Business & Economic Statistics*, 10, 1-9.

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

[16] Zeldes, S. P. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation", *Journal of Political Economy*, 97, 305-346.