Modelling Optimal Instrumental Variables for Dynamic Panel Data Models

Manuel Arellano[*]
CEMFI, Madrid

Abstract

Two-step instrumental variable estimators for panel data models with general predetermined variables are considered that are asymptotically efficient under some auxiliary assumptions, but remain consistent when the assumptions are violated. Asymptotic efficiency is defined in relation to the information bound for the conditional mean specification of the model. Unlike in standard panel GMM, optimal instruments are parameterized using a fixed number of coefficients for any value of T. Thus, the properties of the resulting estimators are not fundamentally affected by the relative dimensions of T and N. Empirical illustrations are reported using firm- and country-level panel data.

JEL classification: C23.

Keywords: Panel data; optimal instruments; general predetermined variables; double asymptotics; VAR with random effects; country growth equations.

Address for correspondence: CEMFI, Casado del Alisal 5, 28014 Madrid, Spain. Tel. +34 91 429 0551, Fax +34 91 429 1056, E-mail: arellano@cemfi.es

1. Introduction

In this paper we discuss a new method for estimating panel data models with general predetermined or endogenous explanatory variables. In doing so we propose a framework for modelling optimal instruments in panel data analysis.

A popular method in dynamic panel data estimation is GMM, which is consistent in short panels, robust, has general applicability, and provides a well defined notion of optimality (Holtz-Eakin, Newey and Rosen, 1988; Arellano and Bond, 1991). However, in practice the application of GMM often entails too many instruments for acceptable sampling properties in either finite or large samples when the time series dimension is not fixed (Alvarez and Arellano, 2003).

For autoregressive models there are also available likelihood-based methods which exhibit better finite sample properties than GMM but can be seriously biased if certain auxiliary assumptions are violated. Moreover, these methods cannot be readily extended to cover models with endogenous or general predetermined variables. There is therefore an acute robustness-efficiency trade-off in the choice among existing techniques. In addition, some methods are designed for short panels of large cross-sections, while others target long panels of small cross-sections, but there is a vacuum in between. The literature does not seem to have much to offer to researchers interested in "small N, small T" panels or other panels that are not easily classifiable.

GMM estimators can be regarded as providing implicit models for the optimal instruments. The problem is that in the panel context these models are often overparameterized, leading to poor properties in finite samples and double asymptotics. This paper provides a framework for parsimonious modelling of optimal instruments in panel data, which is, first, coherent with the fixed T, large N perspective; second, has good theoretical properties in a double asymptotic setup, and thirdly provides estimators with the same robustness features as popular GMM methods. More research is needed on panel data methodology from a time series perspective, and this paper is intended as a contribution towards a marriage of the cross-sectional (fixed T) and time-series (long T) perspectives.

Recent results by Newey and Smith (2004) on the higher order properties of empirical likelihood (EL) and GMM indicate that panel EL estimators may exhibit better finite sample properties than their GMM counterparts. However, while the double asymptotic properties of panel EL estimation remain to be explored, the

fact that a panel EL estimator will be inconsistent for fixed N, large T (as long as it is based on an increasing number of moment conditions) suggests that EL estimation will be less robust to double asymptotic plans than the instrumental variable methods considered in this paper.

The paper is organized as follows. Section 2 presents the model, links GMM with the parametric optimal-instrument perspective, and introduces projection-restricted simple IV estimation (SIV). Section 3 discusses the asymptotic biases of one-step GMM when both T and N tend to infinity. It is shown that the order of magnitude of the bias depends on whether the explanatory variables are pre-determined or endogenous, and that in the latter situation GMM is inconsistent. In Section 4 we present an auxiliary random effects VAR model for the vector of instruments, and obtain sequential linear projections of the effects. Section 5 describes the form of optimal instruments and provides several examples. Section 6 considers pseudo maximum likelihood estimation (PML) of the auxiliary VAR model. Section 7 discusses the properties of feasible projection-restricted IV estimators with and without completely endogenous variables, and the calculation of asymptotic variance matrix estimates. Section 8 contains empirical illustrations and Monte Carlo simulations. Subsection 8.1 reports estimates of autoregressive employment and wage equations from firm panel data; Subsection 8.2 presents the results of a simulation exercise calibrated to the previous firm panel, and Subsection 8.3 reports estimates of country growth convergence rates using a panel of 92 countries observed at five-year intervals. Finally, Section 9 ends with some concluding remarks and plans for future work. All proofs and technical details are contained in Appendices.

2. Model and Optimal Instruments: Overview

2.1. A Sequential Conditional Mean Model

Let us consider a fixed effects panel data model of the form

$$y_{it} = x'_{it}\beta + \eta_i + \varepsilon_{it} \quad (t = 1, ..., T; i = 1, ..., N), \tag{2.1}$$

together with the conditional mean assumption

$$E(\varepsilon_{it}|z_i^t) \equiv E_t(\varepsilon_{it}) = 0 \tag{2.2}$$

where $z_i^t = (z'_{i1}, ..., z'_{it})'$ and $\left(y_i^T, x_i^T, z_i^T, \eta_i\right)$ are *iid* random variables; η_i represents an unobservable individual effect and z_{it} is a vector of instrumental variables.

The following remarks about the nature of the explanatory variables and the instruments are relevant. First, if an explanatory variable x_{kit} is predetermined for $\varepsilon_{i(t+j)}$, then $x_{ki(t-j)}$ is a component of z_{it}. Second, an x_{kit} may also be completely endogenous in the sense of not being predetermined for any lead of ε. Finally, z_{it} may contain external instruments that are not part of x_{it} or its lags.

An example of this type of model is an equation from a VAR with individual effects. Other examples are partial adjustment equations with predetermined regressors, or a structural relationship between endogenous variables. As illustrations of the latter we consider below cross-country growth and household consumption Euler equations.

2.2. Information Bound and Optimal Instruments

Since the distribution of $\eta_i \mid z_i^T$ is unrestricted, all information about β is in the conditional moments for the errors in differences or forward orthogonal deviations

$$E_t\left(y_{it}^* - x_{it}^{*'}\beta\right) = E_t\left(\varepsilon_{it}^*\right) = 0 \quad (t = 1, ..., T - 1) \tag{2.3}$$

where starred variables denote orthogonal deviations:

$$\varepsilon_{it}^* = \left(\frac{T - t}{T - t + 1}\right)^{1/2}\left[\varepsilon_{it} - \frac{1}{(T - t)}(\varepsilon_{i(t+1)} + ... + \varepsilon_{iT})\right]. \tag{2.4}$$

The advantage of orthogonal deviations is that if ε_{it} is homoskedastic and serially uncorrelated so is ε_{it}^* (Arellano and Bover, 1995).

Two-wave Panel   If $T = 2$ there is just one equation in deviations (which coincides with first-differences):

$$E\left(y_{i1}^* - x_{i1}^{*\prime}\beta \mid z_i^1\right) = E_1\left(\varepsilon_{i1}^*\right) = 0 \tag{2.5}$$

and the optimal instrument is $h_{i1} = E_1\left(x_{i1}^*\right)/E_1\left(\varepsilon_{i1}^{*2}\right)$, in the sense that the un-feasible IV estimator

$$\widetilde{\beta}_1 = \left(\sum_{i=1}^N h_{i1} x_{i1}^{*\prime}\right)^{-1} \sum_{i=1}^N h_{i1} y_{i1}^* \tag{2.6}$$

attains the variance bound for this problem.

A parametric approach to feasible estimation is to specify functional forms for $E_1\left(x_{i1}^*\right)$ and $E_1\left(\varepsilon_{i1}^{*2}\right)$ and substitute suitable estimates in the IV formula (2.6). 2SLS is an example of this approach that uses the sample linear projection of x_{i1}^* on z_i^1 as an estimate of the optimal instrument. That is, 2SLS attains the variance bound when $E_1\left(x_{i1}^*\right)$ is linear and $E_1\left(\varepsilon_{i1}^{*2}\right)$ is constant.[1]

Thus, in a parametric approach to feasible IV estimation there are two levels of assumptions: the substantive conditional moment restrictions used in estimation, and the auxiliary assumptions used in estimating the optimal instruments. The former are related to consistency and the latter to asymptotic efficiency. This distinction between substantive and auxiliary assumptions is central to the perspective adopted in this paper.

Multi-wave Panel   If $T > 2$ the form of the optimal instrument is compli-cated by the fact that neither conditional heteroskedasticity or autocorrelation in ε_{it}^* are ruled out. The unfeasible optimal IV estimator in the general case solves

$$\sum_{i=1}^N \sum_{t=1}^{T-1} h_{it}\left(\widetilde{y}_{it} - \widetilde{x}_{it}'\widetilde{\beta}\right) = 0 \tag{2.7}$$

where $h_{it} = E_t\left(\widetilde{x}_{it}\right)/E_t\left(\widetilde{\varepsilon}_{it}^2\right)$ and

$$\widetilde{\varepsilon}_{i(T-1)} = \varepsilon_{i(T-1)}^* \tag{2.8}$$

---

[1]Incorporating these assumptions in estimation may reduce the variance bound for β. A trade-off between robustness and efficiency arises, since estimates of β that exploit the extra restrictions may be inconsistent if they are false.

$$\widetilde{\varepsilon}_{it} = \varepsilon_{it}^* - \tau_{t1}\widetilde{\varepsilon}_{i(t+1)} - ... - \tau_{t(T-t-1)}\widetilde{\varepsilon}_{i(T-1)} \ \ (t = T - 2, ..., 1) \tag{2.9}$$

with $\tau_{tj} = E_{t+j}\left(\varepsilon_{it}^*\widetilde{\varepsilon}_{i(t+j)}\right)/E_{t+j}\left(\widetilde{\varepsilon}_{i(t+j)}^2\right)$.

The $\widetilde{\varepsilon}_{it}$ are forward filtered errors such that $E_t\left(\widetilde{\varepsilon}_{it}\right) = 0$ and $E_{t+j}\left(\widetilde{\varepsilon}_{it}\widetilde{\varepsilon}_{i(t+j)}\right) = 0$ so that the bound for $T - 1$ periods is the sum of the bounds for each period (Chamberlain, 1992):

$$J_{T-1} = \sum_{t=1}^{T-1} E\left(\frac{E_t\left(\widetilde{x}_{it}\right)E_t\left(\widetilde{x}_{it}\right)'}{E_t\left(\widetilde{\varepsilon}_{it}^2\right)}\right). \tag{2.10}$$

If the ε_{it}^*'s are conditionally homoskedastic and serially uncorrelated $\widetilde{\varepsilon}_{it} = \varepsilon_{it}^*$. The optimal instrument is $h_{it} \equiv h_t\left(z_i^t\right) = E_t(x_{it}^*)$, and the unfeasible IVE

$$\widetilde{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T-1} h_{it}x_{it}^{*\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T-1} h_{it}y_{it}^*\right) \tag{2.11}$$

attains the fixed-T, large-N variance bound.

Generalized Method of Moments   A One-step GMM estimator (GMM1) is an example of the parametric approach to feasible IVE that uses the cross-sectional sample linear projections of x_{it}^* on z_i^t as an estimate of the optimal instruments. GMM1 is based on the specification

$$h_t\left(z_i^t\right) = z_{i1}'\pi_{t1} + ... + z_{it}'\pi_{tt} \equiv z_i^{t\prime}\pi_t, \tag{2.12}$$

together with assumptions of homoskedasticity and lack of serial correlation.

In a time series context (large T, fixed N), these projections cannot be consistently estimated without further restrictions. But as $N \to \infty$ for fixed T, the IVE that uses the sample projection $z_i^{t\prime}\widehat{\pi}_t$ has the same asymptotic distribution as the unfeasible IVE based on the population projection $z_i^{t\prime}\pi_t$.

However, when both T and N tend to infinity at the same rate, the feasible and unfeasible estimators differ as shown for autoregressive models in Alvarez and Arellano (2003), who found that GMM1 had an asymptotic bias of order $1/N$. In the next Section we obtain the asymptotic bias of GMM1 for model (2.1)-(2.2) when both T and N tend to infinity. We show that the order of magnitude of the bias depends on whether the explanatory variables are endogenous or predetermined. In the predetermined case the bias is of order $1/N$—as in Alvarez

and Arellano (2003)—but in the endogenous case the bias is of order T/N, and hence a potentially more serious problem. These results provide theoretical support for the approach developed in this paper, because the proposed estimators are immune to asymptotic bias in a double asymptotics.

Under heteroskedasticity or autocorrelation, it is possible to obtain a GMM estimator based on the moments $E\left(z_i^t v_{it}^*\right) = 0$ that is more efficient than $\widehat{\beta}_{GMM1}$ by using as weight matrix the inverse of a robust estimate of the variance matrix of the orthogonality conditions. These are the standard two-step robust GMM2 estimators (Arellano and Bond, 1991). A GMM2 estimator, however, will not attain the efficiency bound in general, although it may attain it under more general conditions than GMM1. Specifically, GMM2 will attain the bound under serial correlation and time series heteroskedasticity of the form $E(v_{it}^2 | z_i^t) = E(v_{it}^2) = \sigma_t^2$ and $E(v_{it} v_{i(t+j)} | z_i^{t+j}) = E(v_{it} v_{i(t+j)}) = \sigma_{t(t+j)}$ for $j > 0$.[2]

Semiparametric Asymptotically Efficient Estimation for Fixed T It is also possible to devise estimators that attain the fixed-T efficiency bound under more general assumptions than GMM1 or GMM2. This could be achieved by considering a semiparametric estimator that replaces the unknown functions in (2.11) or (2.7) by nonparametric estimates. The former would be asymptotically efficient when the conditional expectations $E(x_{it}^* | z_i^t)$ are nonlinear but the errors are classical. The latter might be efficient in the general case. Estimators of this type have been considered by Hahn (1997). Nevertheless, since these estimators use even more flexible estimates of $E(x_{it}^* | z_i^t)$ than GMM1, they would not be expected to perform better than GMM1 unless T/N is sufficiently close to zero.

2.3. The Crude Time-Series Parametric Approach

The conventional time-series parametric approach specifies

$$h_t\left(z_i^t\right) = \sum\nolimits_{j=0}^{q} z_{i(t-j)}' \gamma_j \tag{2.13}$$

---

[2] A more restrictive two-step GMM estimator based on a weight matrix that only depends on the data second-order moments (Arellano and Bond, 1991, footnote 2), will attain the bound under the same conditions as the standard GMM2.

for some q.[3] Equation (2.13) has constant coefficients γ_j that can be consistently estimated time series-wise. It is based on assumptions about the stability and degree of dependence in the z_{it} process.

Panel data examples of the time-series approach are the stacked IV estimators for autoregressive models due to Anderson and Hsiao (1982). From the point of view of fixed-T optimal IV estimation, however, the crude time-series approach has two undesirable features. First, (2.13) cannot be calculated for periods without sufficient observations of the initial lags, hence requiring trimming of the time series and loss of some cross-sections.

A second less obvious problem is that if the z_{it} process contains heterogeneous intercepts—as may be expected given model (2.1)— $h_t(z_i^t)$ will depend on all lags, even if at the individual level z is characterized by stable low-order dependence of the kind that supports assumption (2.13) in a time series context. The reason is that all lags are predictors of the effects (as described in Section 4). Since the predictions are updated each period, the coefficients of the predictor change over time. This is precisely the motivation behind (2.12).

2.4. Projection-Restricted IV Estimation

The problem with $z_i^{t\prime}\pi_t$ is that it often contains too many coefficients for good finite sample inference. If we have an individual-effects, stable process for z characterized by a parameter vector γ, the π_t are functions of γ. Therefore, instead of an unrestricted linear projection we may consider a restricted one:

$$h_t\left(z_i^t\right) = z_i^{t\prime}\pi_t\left(\gamma\right). \tag{2.14}$$

These restrictions are similar to the low order, stable dependence assumptions implicit in the time-series approach. The difference is that we are using them in a fashion consistent with the fixed-T panel perspective.

For a long-T panel the outcome is essentially the same as in the crude time-series approach. For a very short-T panel the outcome is essentially the same as in the standard GMM approach. But for other panels the number of first-stage coefficients is kept constant, while using all waves and the predictive ability due to unobserved heterogeneity.

---

[3] In this case it is better to think of $h_t(z_i^t)$ as $E_t\left(x_{it} - x_{i(t+1)}\right)$ because x_{it}^* involves a different number of terms for each t.

The suggested strategy is to specify a VAR for z_{it} with individual effects and unrestricted initial conditions. The status of this assumption is similar to that of (2.13) in the time-series approach. Then use as instruments the estimated restricted projections $z_i^{t\prime}\pi_t(\gamma)$, whose evaluation involves, as we shall see, a straight-forward recursive Kalman-filter calculation.

Unrestricted π_t in GMM are estimated by cross-sectional OLS. Restricted $\pi_t(\gamma)$ will be estimated using a modified WG estimator, which is a multivariate generalization of the random effects PML in Alvarez and Arellano (2003).

We then obtain

$$\widehat{\beta} = \left(\sum_i \sum_t \widehat{h}_{it} x_{it}^{*\prime}\right)^{-1} \sum_i \sum_t \widehat{h}_{it} y_{it}^*. \tag{2.15}$$

GMM sets $\widehat{h}_{it} = z_i^{t\prime}\widehat{\pi}_t$ where $\widehat{\pi}_t$ is an OLS estimate, and the projection-restricted IVE sets $\widehat{h}_{it} = z_i^{t\prime}\pi_t(\widehat{\gamma})$ where $\widehat{\gamma}$ is a PML estimate.

For fixed T and large N both estimators have similar robustness properties in the sense of being consistent under the same assumptions, but when T is not fixed the latter is immune to asymptotic biases because the number of first-stage coefficients does not increase with T. If the auxiliary assumptions $\pi_t = \pi_t(\gamma)$ are violated, projection-restricted IV remains a consistent estimate of β.

The goal is, therefore, to use the time-series parametric approach in such a way that the resulting estimator can still achieve the fixed T efficiency bound under certain assumptions, in the same way as standard GMM only achieves the fixed T bound under certain auxiliary assumptions.

3. Double-Asymptotic Biases

In this Section we obtain the asymptotic bias of the one-step GMM estimator when both T and N tend to infinity. We show that the order of magnitude of the bias depends on whether the explanatory variables are endogenous or predetermined.

It is well known that if the number of first-stage coefficients is large relative to the sample size, standard asymptotic approximations may be a poor guide to the finite sample properties of IV estimates, specially when the instruments are weak. In the panel context, double-asymptotic results provide formal approximations to the impact of nonnegligible T (and hence of many instruments) relative to N

on the properties of GMM estimates. They also provide a theoretical motivation for the strategy for reducing the number of first-stage coefficients pursued in this paper.

The form of the GMM estimator is

$$\widehat{\beta} = \left(\sum_{t=1}^{T-1} X_t^{*\prime} M_t X_t^* \right)^{-1} \sum_{t=1}^{T-1} X_t^{*\prime} M_t y_t^*, \tag{3.1}$$

where $M_t = Z_t \left(Z_t^\prime Z_t \right)^{-1} Z_t^\prime$ is $N \times N$, $X_t^* = (x_{1t}^*, ..., x_{Nt}^*)'$ is $N \times k$, $Z_t = (z_1^t, ..., z_N^t)'$ is $N \times mt$, and $y_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$ is $N \times 1$, or in a more compact notation

$$\widehat{\beta} = (X^{*\prime} M X^*)^{-1} X^{*\prime} M y^*. \tag{3.2}$$

3.1. Assumptions

Let us define the vector of variables $w_{it} = (\varepsilon_{it}, x_{it}', z_{it}')'$. Depending on the model, x and z may contain elements in common. Also, some of the xs may be lags of y, and some of the zs may be lags of y and/or x. We make the following assumption:

Assumption 1: w_{it} can be represented as a vector MA(∞) of the form

$$w_{it} = \mu_i + \zeta_{it} + \Psi_1 \zeta_{i(t-1)} + \Psi_2 \zeta_{i(t-2)} + ... \tag{3.3}$$

where $\{\Psi_j\}_{j=0}^{\infty}$ is absolutely summable, and $\{\zeta_{it}\}_{t=-\infty}^{\infty}$ is an i.i.d. sequence independent of the individual-specific mean vector μ_i and with finite fourth-order moments. Moreover, the first element of μ_i is zero, and

$$\sum_{s=-\infty}^{\infty} |s| \gamma_s = B_\infty < \infty \tag{3.4}$$

where $\gamma_s = E \left(x_{it} \varepsilon_{i(t-s)} \right)$.

GMM estimators are typically motivated under less restrictive conditions than Assumption 1, which is only made for simplicity, since our purpose is to exhibit some consequences of double-asymptotics in a leading situation, rather than providing an alternative basis for inference with GMM estimates. The results that follow are expected to hold under more general conditions.

Predeterminedness and Endogeneity   If x_{it} is predetermined for ε_{it}, then $x_{it} \vdash \left(\varepsilon_{it}, \varepsilon_{i(t+1)}, \varepsilon_{i(t+2)}, ...\right)$ but $x_{it} \nvdash \left(\varepsilon_{i(t-1)}, \varepsilon_{i(t-2)}, ...\right)$, so that $\gamma_0 = \gamma_{-1} = ... = 0$ but $\gamma_s \neq 0$ for $s > 0$. If on the other hand x_{it} is completely endogenous

$$x_{it} \nvdash \left(..., \varepsilon_{i(t-2)}, \varepsilon_{i(t-1)}, \varepsilon_{it}, \varepsilon_{i(t+1)}, \varepsilon_{i(t+2)}, ...\right),$$

in which case $\gamma_s \neq 0$ for all s. Finally, if $x_{it} \vdash \left(\varepsilon_{i(t+j)}, \varepsilon_{i(t+j+1)}, ...\right)$ we say that x_{it} is endogenous but predetermined for $\varepsilon_{i(t+j)}$, and the corresponding γs vanish.

3.2. The Order of the Estimation Error

Let us write the standardized estimation error as

$$\widehat{\beta} - \beta = \left(\frac{1}{NT} \sum_{t=1}^{T-1} X_t^{*\prime} M_t X_t^*\right)^{-1} \frac{1}{NT} \sum_{t=1}^{T-1} X_t^{*\prime} M_t \varepsilon_t^* \equiv A_{NT}^{-1} b_{NT}. \qquad (3.5)$$

A key result is given in the following theorem.

Theorem 1. *Letting $b_{NT} = (NT)^{-1} \sum_{t=1}^{T-1} X_t^{*\prime} M_t \varepsilon_t^*$, under Assumption 1 if x_{it} is predetermined, in the sense that $E\left(x_{it} \varepsilon_{i(t+j)}\right) = 0$ for $j \geq 0$, then*

$$E\left(b_{NT}\right) = O\left(\frac{m}{N}\right). \qquad (3.6)$$

If x_{it} is endogenous, in the sense that $E\left(x_{it} \varepsilon_{it}\right) \neq 0$, then

$$E\left(b_{NT}\right) = O\left(\frac{mT}{N}\right). \qquad (3.7)$$

Intuitively, in the predetermined case the "endogeneity bias" vanishes as $T \to \infty$ because the within-group OLS (WG) estimator is large-T consistent. However, WG is not T-consistent in the endogenous explanatory variable case. These results also highlight the impact of m—the dimension of the instrument vector z_{it}—on the order of magnitude of the bias.

3.3. The asymptotic bias of GMM

The previous setup can be used to extend the results in Alvarez and Arellano (2003) to analyzing the asymptotic properties of GMM estimators with general predetermined or endogenous explanatory variables. Here we provide an informal

discussion of consistency by drawing parallels with the corresponding Alvarez–Arellano results.

Let us first consider the probability limit of $(NT)^{-1} (X^{*\prime}MX^{*})$. It is useful to introduce at this point the time series individual-specific linear projection of x_{it} on $\{z_{it}, z_{i(t-1)}, z_{i(t-2)}, ...\}$:

$$p_{it} \equiv E^{*}\left(x_{it} \mid z_{it}, z_{i(t-1)}, ...\right) = \phi_{i} + \sum_{j=0}^{\infty} \Phi_{j} z_{i(t-j)}. \tag{3.8}$$

Letting ξ_{it} be the corresponding projection error, we have

$$x_{it} = p_{it} + \xi_{it}. \tag{3.9}$$

If x_{it} is predetermined then $x_{it} = p_{it}$ and the errors ξ_{it} are identically zero for all i and t. In such a case, using similar arguments as in Alvarez and Arellano (2003), when $T/N \to c < \infty$ we can obtain

$$\operatorname*{plim}_{T\to\infty, N\to\infty} \frac{1}{NT} X^{*\prime}MX^{*} = E\left[(x_{it} - \mu_{xi})(x_{it} - \mu_{xi})'\right] \tag{3.10}$$

where μ_{xi} corresponds to the partition $\mu_{i} = (0, \mu_{xi}', \mu_{zi}')'$.

However, when x_{it} is endogenous we have

$$\frac{1}{NT} X^{*\prime}MX^{*} = \frac{1}{NT} P^{*\prime}MP^{*} + \frac{1}{NT} \Xi^{*\prime}MP^{*} + \frac{1}{NT} P^{*\prime}M\Xi^{*} + \frac{1}{NT} \Xi^{*\prime}M\Xi^{*} \tag{3.11}$$

where $X^{*} = P^{*} + \Xi^{*}$ denotes the matrix and orthogonal deviation counterpart to $x_{it} = p_{it} + \xi_{it}$. By analogy with the Alvarez–Arellano results, when $T/N \to c < \infty$ we have

$$\frac{1}{NT} P^{*\prime}MP^{*} \xrightarrow{p} E\left[(p_{it} - \phi_{i})(p_{it} - \phi_{i})'\right] \tag{3.12}$$

$$\frac{1}{NT} P^{*\prime}M\Xi^{*} = O\left(\frac{m}{N}\right) \tag{3.13}$$

$$\frac{1}{NT} \Xi^{*\prime}M\Xi^{*} = O\left(\frac{mT}{N}\right). \tag{3.14}$$

Decomposition (3.9) can also be used to examine the probability limit of $(NT)^{-1} (X^{*\prime}M\varepsilon^{*})$. We have

$$\operatorname*{plim}_{T\to\infty, N\to\infty} \frac{1}{NT} X^{*\prime}M\varepsilon^{*} = \operatorname*{plim}_{T\to\infty, N\to\infty} \frac{1}{NT} P^{*\prime}M\varepsilon^{*} + \operatorname*{plim}_{T\to\infty, N\to\infty} \frac{1}{NT} \Xi^{*\prime}M\varepsilon^{*}. \tag{3.15}$$

11

Extending the Alvarez–Arellano results (Lemma 2), it turns out that the first term on the right hand side of (3.15) vanishes, but $(NT)^{-1}\left(\Xi^{*\prime}M\varepsilon^{*}\right)$ is of order $O\left(mT/N\right)$.

Thus, when x_{it} is endogenous and $T/N \to c < \infty$, the form of the asymptotic bias is given by

$$
\begin{aligned}
\plim_{T\to\infty,N\to\infty}\left(\widehat{\beta}-\beta\right) &= \left\{E\left[\left(p_{it}-\phi_{i}\right)\left(p_{it}-\phi_{i}\right)'\right] + \plim_{T\to\infty,N\to\infty}\frac{1}{NT}\Xi^{*\prime}M\Xi^{*}\right\}^{-1}\\
&\times \plim_{T\to\infty,N\to\infty}\frac{1}{NT}\Xi^{*\prime}M\varepsilon^{*}
\end{aligned} \tag{3.16}
$$

or

$$
\plim_{T\to\infty,N\to\infty}\left(\widehat{\beta}-\beta\right) = \left\{E\left[\left(p_{it}-\phi_{i}\right)\left(p_{it}-\phi_{i}\right)'\right] + E\left(\xi_{it}\xi_{it}'\right)\frac{mc}{2}\right\}^{-1}\gamma_{0}\frac{mc}{2}. \tag{3.17}
$$

Note that when x is predetermined $\gamma_{0} = 0$ and $\xi_{it} \equiv 0$, so that the bias vanishes as long as $c < \infty$. The conclusion is that GMM is inconsistent in the endogenous explanatory variable situation. In the predetermined case it is consistent, but it can be expected to exhibit a bias in the asymptotic distribution of order $O\left(m/N\right)$, similar to those reported in Alvarez and Arellano (2003) for autoregressive models.

The intuition for these results is that in the predetermined case the endogeneity bias tends to disappear as T increases, whereas in the endogenous case it does not. So having an increasing number of moment conditions results in a larger bias when the explanatory variables are endogenous.

4. Auxiliary VAR with Individual Effects

Let us consider a stable VAR(1) process for an $m \times 1$ vector z_{it} :

$$
z_{it} = Az_{i(t-1)} + (I-A)\mu_{i} + v_{it} \ (t = 1,...,T) \tag{4.1}
$$

$$
E\left(v_{it} \mid z_{i}^{t-1}, \mu_{i}\right) = 0 \tag{4.2}
$$

where $z_{i}^{t-1} = \left(z_{i0}', ..., z_{i(t-1)}'\right)'$, and $Var\left(v_{it}\right) = \Omega_{t}$. For notational convenience we assume here that z_{i0} is also observed.

It is appropriate to conduct our discussion at the multivariate level because typically the conditioning variable will be a vector rather than a scalar, even if we are dealing with a single-equation model. However, we restrict attention to

a first-order process for simplicity. Generalization to higher-order processes is cumbersome but straightforward.

If the process for individual i started in the infinite past:

$$z_{i0} = \mu_i + \sum_{j=0}^{\infty} A^j v_{i(-j)}.$$ (4.3)

Since we wish to allow for the possibility that the process started in any given period, and that different individuals started at different times, we treat (z_{i0}, μ_i) as realizations of some arbitrary cross-sectional joint distribution.

Let $E(\mu_i) = \mu$, $Var(\mu_i) = \Omega_\mu$, and let the linear projection of z_{i0} on μ_i be

$$z_{i0} = \tau_0 + \Upsilon_1 \mu_i + v_{i0}^\dagger$$ (4.4)

where $E\left(v_{i0}^\dagger\right) = 0$ and $Var\left(v_{i0}^\dagger\right) = \Gamma_0$. In general, τ_0, Υ_1, and Γ_0 are free parameters, but if the process is mean stationary $\tau_0 = 0$, and $\Upsilon_1 = I_m$. If the process is also stationary in variance, $\Omega_t = \Omega$ for all t, and $\Gamma_0 = \sum_{j=0}^{\infty} A^j \Omega A^{j\prime}$. In any event, μ_i denotes the mean of the steady state distribution of the process for individual i, and μ is the cross-sectional mean of μ_i.

VAR Forecasts   For $s > 0$ we have

$$z_{i(t+s)} = \mu_i + A^s \left(z_{it} - \mu_i\right) + v_{i(t+s)} + A v_{i(t+s-1)} + ... + A^{s-1} v_{i(t+1)}.$$ (4.5)

Therefore,

$$E_t \left(z_{i(t+s)}\right) = E_t(\mu_i) + A^s \left[z_{it} - E_t(\mu_i)\right]$$ (4.6)

and

$$E_t(z_{it}^*) = c_t \left[z_{it} - \frac{1}{T-t} \sum_{s=1}^{T-t} E_t\left(z_{i(t+s)}\right)\right]$$ (4.7)

so that

$$E_t(z_{it}^*) = c_t \left(I - \frac{1}{T-t} \sum_{s=1}^{T-t} A^s\right) \left[z_{it} - E_t(\mu_i)\right].$$ (4.8)

Notice that $E_t(z_{it}^*)$ only depends on z_{it} given $E_t(\mu_i)$. This is so because we are dealing with a first-order VAR. However, since $E_t(\mu_i)$ depends on all lags, so does $E_t(z_{it}^*)$.

Similarly,

$$E_t \left(z_{i(t-j)}^* \right) = c_t \left\{ z_{i(t-j)} - \frac{1}{T-t} \left[z_{i(t-j+1)} + ... + z_{it} + E_t \left(z_{i(t+1)} + ... + z_{i(T-j)} \right) \right] \right\} \tag{4.9}$$

where

$$E_t \left(z_{i(t+1)} + ... + z_{i(T-j)} \right) = (T-t-j) E_t \left(\mu_i \right) + \left(\sum_{s=1}^{T-t-j} A^s \right) \left[z_{it} - E_t \left(\mu_i \right) \right].$$

The same is true if conditional expectations $E_t \left(. \right)$ are replaced by linear projections $E_t^* \left(. \right)$. The form of $E_t^* \left(\mu_i \right)$ is given in the following Theorem.

Theorem 2. *(Sequential linear projections of the effects) For the VAR(1) model presented above, the linear projection of the vector of individual effects μ_i on z_i^t is given by the following recursive updating formula for $t \geq 0$:*

$$E_t^* \left(\mu_i \right) = \mathcal{H}_t^{-1} \delta_{it} \tag{4.10}$$

where

$$\mathcal{H}_0 = I + \Omega_\mu \Upsilon_1' \Gamma_0^{-1} \Upsilon_1 \tag{4.11}$$

$$\delta_{i0} = \mu + \Omega_\mu \Upsilon_1' \Gamma_0^{-1} \left(z_{i0} - \tau_0 \right) \tag{4.12}$$

and for $t \geq 1$:

$$\mathcal{H}_t = \mathcal{H}_{t-1} + \Omega_\mu \left(I - A \right)' \Omega_t^{-1} \left(I - A \right) \tag{4.13}$$

$$\delta_{it} = \delta_{i(t-1)} + \Omega_\mu \left(I - A \right)' \Omega_t^{-1} \left(z_{it} - A z_{i(t-1)} \right). \tag{4.14}$$

For the stationary case, by inserting $\tau_0 = 0$, $\Upsilon_1 = I_m$, and $\Omega_t = \Omega$ for all t in the result given in the Theorem, we get:

$$E_t^* \left(\mu_i \right) = \left[I + \Lambda_0 + t\Lambda_1 (I - A) \right]^{-1} \left[\mu + \Lambda_0 z_{i0} + \Lambda_1 \sum_{s=1}^{t} \left(z_{is} - A z_{i(s-1)} \right) \right]. \tag{4.15}$$

where $\Lambda_0 = \Omega_\mu \Gamma_0^{-1}$ and $\Lambda_1 = \Omega_\mu (I - A)' \Omega^{-1}$ (further details on the stationary case are in the Appendix).

Note that the linear projection of the vector of individual effects in a VAR(1) model with mean stationarity and constant Ω but no covariance stationarity is of

14

the same form as the result in (4.15), but treating Γ_0 as an unrestricted covariance matrix.

Another intermediate possibility is one in which $\Omega_t = \Omega$ for all t, but initial conditions are left fully unrestricted, so that τ_0, Υ_1, and Γ_0 are treated as free parameters. This is a case of special interest because $E_t^* (\mu_i)$ depends on a fixed number of parameters which does not increase with t.

5. The Form of the Optimal Instruments

An expression for the optimal instruments is given by[4]

$$
\begin{aligned}
E_t \left(x_{it}^* \right) &= B_0 E_t \left(z_{it}^* \right) + B_1 E_t \left(z_{i(t-1)}^* \right) + ... + B_q E_t \left(z_{i(t-q)}^* \right) \quad (5.1) \\
&= B^\dagger E_t \left(z_{it(q)}^* \right)
\end{aligned}
$$

where $B^\dagger = (B_0, ..., B_q)$ and $z_{it(q)} = \left(z_{it}', ..., z_{i(t-q)}' \right)'$.

If all x's are predetermined for some lead of ε, the form of $E_t \left(x_{it}^* \right)$ and $B^\dagger$ is known given knowledge of the parameters of the VAR for z_{it}, so that estimation of β can be based on the sample moments

$$
\sum_i \sum_t \widehat{E}_t \left(x_{it}^* \right) \left(y_{it}^* - x_{it}^{*\prime} \beta \right) \quad (5.2)
$$

where $\widehat{E}_t \left(x_{it}^* \right)$ is an estimate of $E_t \left(x_{it}^* \right)$. Since the optimal instrument has the same dimension as β no further weighting of the moments is required.

If on the other hand part of the x's are completely endogenous, then q and some of the elements of $B^\dagger$ are unknown. Thus, we consider GMM estimates of β based on the moments

$$
\sum_i \sum_t \widehat{E}_t \left(z_{it(q)}^* \right) \left(y_{it}^* - x_{it}^{*\prime} \beta \right) . \quad (5.3)
$$

In this case the instruments may have a larger dimension than β so that weighting of the moments may be required. Indeed, the choice of weight matrix

---

[4]Expression (5.1) is saying that

$$
E \left(x_{it}^* - B_0 z_{it}^* - B_1 z_{i(t-1)}^* - ... - B_q z_{i(t-q)}^* \mid z_i^t \right) = 0
$$

or that the time-series individual-specific projection of x_{it} on z_i^t only depends on the first q lags of z_{it} and satisfies $E \left(x_{it} - B_0 z_{it} - B_1 z_{i(t-1)} - ... - B_q z_{i(t-q)} - \xi_i \mid z_i^t \right) = 0$.

here is equivalent to choosing an estimation method for $B^\dagger$. A suggested weighting is discussed below.[5] Next, we consider some examples.

Example 1: An Equation from a VAR   We have

$$w_{1it} = a_1' w_{i(t-1)} + \eta_i + \varepsilon_{it} \tag{5.4}$$

$$E\left(\varepsilon_{it} \mid w_i^{t-1}\right) = 0. \tag{5.5}$$

In this case $z_{it} = w_{i(t-1)}$ and $x_{it} = z_{it}$, so that $B_0 = I$, and $B_j = 0$ for $j > 0$.

When the model of interest is a VAR, the auxiliary and substantive models coincide, except for the fact that the auxiliary model is based on stronger assumptions than the substantive model. In contrast, in conditional models the auxiliary assumptions give a parametric form to the feedback processes, which in the substantive model will typically remain unspecified.[6]

Example 2: Partial Adjustment Regression   We have

$$y_{it} = \alpha y_{i(t-1)} + \beta_0 w_{it} + \beta_1 w_{i(t-1)} + \eta_i + \varepsilon_{it} \tag{5.6}$$

$$E\left(\varepsilon_{it} \mid y_i^{t-1}, w_i^t\right) = 0. \tag{5.7}$$

In this case $z_{it} = \left(y_{i(t-1)}, w_{it}\right)'$, $x_{it} = \left(y_{i(t-1)}, w_{it}, w_{i(t-1)}\right)'$,

$$x_{it} = \begin{pmatrix} y_{i(t-1)} \\ w_{it} \\ w_{i(t-1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} z_{it} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} z_{i(t-1)} \equiv B_0 z_{it} + B_1 z_{i(t-1)}. \tag{5.8}$$

Example 3: Cross-Country Growth   The next example is an augmented Solow model of the determinants of growth as in Caselli, Esquivel, and Lefort (1996). The equation is

$$y_{it} = \alpha y_{i(t-1)} + s_{i(t-1)}\gamma + f_{i(t-1)}'\delta + \eta_i + \varepsilon_{it} \tag{5.9}$$

$$E\left(\varepsilon_{it} \mid y_i^{t-1}, s_i^{t-1}, f_i^{t-2}\right) = 0. \tag{5.10}$$

---

[5] Note that estimation of an extended VAR for (x_{it}', z_{it}') subject to exclusion restrictions (of lagged x's) is not warranted. The reason is that the $B^\dagger$ are meant to be the coefficients of a linear projection of x_{it} on z_i^t and individual effects, and there is no reason why the errors in this projection should be serially uncorrelated.

[6] Optimal instruments for scalar autoregressive models are discussed in Appendix C.

The time interval is 5 years, y_{it} is log per-capita GDP, $f_{i(t-1)}$ is a vector of flow variables containing the rates of investment and population growth, and $s_{i(t-1)}$ is a stock variable measuring the secondary school enrollment rate.

In this case $z_{it} = \left(y_{i(t-1)}, s_{i(t-1)}, f'_{i(t-2)} \right)'$, and

$$
\begin{pmatrix} y_{i(t-1)} \\ s_{i(t-1)} \\ f_{i(t-1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} y_{i(t-1)} \\ s_{i(t-1)} \\ f_{i(t-2)} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ u_{fi(t-1)} \end{pmatrix} \tag{5.11}
$$

where the last equation coincides with the equation for $f_{i(t-2)}$ in the VAR model of z_{it}. Thus,

$$
E_t \left(x_{it}^* \right) = B_0 E_t \left(z_{it}^* \right). \tag{5.12}
$$

Since B_0 is in general a squared, nonsingular matrix, it is irrelevant in the construction of optimal instruments, and the optimal instrument in the growth example can be simply taken to be $E \left(z_{it}^* \mid z_i^t \right)$. So in both the VAR and growth examples we have $B_0 = I$ and $B_j = 0$ for $j > 0$.

Example 4: Euler Equation for Household Consumption The last example is an Euler equation of the type considered by Zeldes (1989) and others:

$$
\ln \left(1 + r_{it} \right) = \alpha \Delta \ln c_{it} + \beta' w_{it} + \eta_i + \varepsilon_{it} \tag{5.13}
$$

where w_{it} is a vector of changes in family size variables, r_{it} is the rate of return on a riskless asset, and c_{it} is consumption; η_i captures heterogeneity in discount rates, and ε_{it} unobservable changes in tastes and expectation errors.

The vector of instruments z_{it} contains w_{it}, lagged income, and marginal taxes, so that both returns and consumption growth are treated as completely endogenous variables.

In this case, letting $z_{it} = (w'_{it}, z'_{2it})'$, we have

$$
\Delta \ln c_{it} = \sum_{j=0}^{q} \pi'_j z_{i(t-j)} + \zeta_i + \xi_{it} \tag{5.14}
$$

$$
w_{it} = (I, 0) z_{it} \tag{5.15}
$$

or

$$
E_t \begin{pmatrix} \left(\Delta \ln c_{it} \right)^* \\ w_{it}^* \end{pmatrix} = \begin{pmatrix} \pi'_0 \\ I, 0 \end{pmatrix} E_t \left(z_{it}^* \right) + .. + \begin{pmatrix} \pi'_q \\ 0 \end{pmatrix} E_t \left(z_{i(t-J)}^* \right). \tag{5.16}
$$

Thus, there is a nontrivial specification of the $(B_0 ... B_q)$ due to the presence of completely endogenous variables and external instruments.

17

6. PML estimation of the VAR model

To obtain feasible estimators, the coefficients parameterizing the optimal instruments must be replaced by sample estimates. The suggested estimates are Gaussian pseudo-maximum likelihood statistics of the VAR auxiliary model. The estimates are obtained under the assumptions that the data are normally distributed and the error variance matrices are homoskedastic, but the multivariate linear projection of the initial observations on the effects is left unrestricted. Estimates of this type were considered by Blundell and Smith (1991), and Alvarez and Arellano (2003) for a scalar autoregressive model.[7] Alvarez and Arellano obtained a useful concentrated likelihood that only depended on the autoregressive parameter, and found that the maximizer of this criterion behaved very well in simulations.[8] This section reports a multivariate generalization of the PML results of Alvarez and Arellano, which will be required in practice for obtaining feasible optimal instruments.[9]

As shown in the Appendix, the log-likelihood given z_{i0} (under homoskedasticity and a joint normal distribution for μ_i and z_{i0} with unrestricted mean and covariance matrix) can be written as:

$$\ln f\left(z_{i1}, ..., z_{iT} \mid z_{i0}\right) = -\frac{(T-1)}{2} \ln \det \Omega - \frac{1}{2} \sum_{t=1}^{T-1} u_{it}^{*\prime} \Omega^{-1} u_{it}^{*}$$

$$-\frac{1}{2} \ln \det \Theta_0 - \frac{1}{2} \left(\overline{u}_i - \phi_0 - \Phi_1 z_{i0}\right)' \Theta_0^{-1} \left(\overline{u}_i - \phi_0 - \Phi_1 z_{i0}\right). \tag{6.1}$$

where $u_{it}^{*} = z_{it}^{*} - A z_{i(t-1)}^{*}$, $\overline{u}_i = \overline{z}_i - A\overline{z}_{i(-1)}$ and

$$\overline{u}_i \mid z_{i0} \sim \mathcal{N}\left(\phi_0 + \Phi_1 z_{i0}, \Theta_0\right) \tag{6.2}$$

Concentrating ϕ_0, Φ_1, Θ_0 and Ω, the PMLE of A solves

$$\widetilde{A} = \arg\min\{\ln \det \left(Z^{*\prime} - A Z_{-1}^{*\prime}\right)\left(Z^{*} - Z_{-1}^{*} A'\right)$$

---

[7] Blundell and Smith (1991) considered a generalized least squares estimator of the same model, which has been further discussed by Blundell and Bond (1998).

[8] This is the PML estimate that does not restrict the individual effect variance to be non-negative. Alvarez and Arellano also discussed an alternative PML that enforced non-negativity, but in such a case a boundary solution may occur.

[9] One problem with this method as an estimator of substantive VAR parameters is that it is not robust to lack of time series homoskedasticity.

$$+\frac{1}{T-1}\ln\det\left(\overline{Z}'-A\overline{Z}'_{-1}\right)S_0\left(\overline{Z}-\overline{Z}_{-1}A'\right)\}\tag{6.3}$$

where $S_0 = I - F\left(F'F\right)^{-1}F'$ and F is the $N \times (m+1)$ matrix of constants and initial observations $f_{i0} = (1, z'_{i0})'$.

PML estimates of the remaining parameters Given $\widetilde{A}$, they are given by

$$\widetilde{\Omega} = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T-1}(z^*_{it} - \widetilde{A}z^*_{i(t-1)})(z^*_{it} - \widetilde{A}z^*_{i(t-1)})'\tag{6.4}$$

$$\widetilde{\Phi} = \left(\widetilde{\phi}_0, \widetilde{\Phi}_1\right) = \left(\overline{Z}' - \widetilde{A}\overline{Z}'_{-1}\right)F\left(F'F\right)^{-1}\tag{6.5}$$

and letting $\widetilde{\overline{u}}_i = \overline{z}_i - \widetilde{A}\overline{z}_{i(-1)}$:

$$\widetilde{\Theta}_0 = \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\overline{u}}_i - \widetilde{\phi}_0 - \widetilde{\Phi}_1 z_{i0}\right)\left(\widetilde{\overline{u}}_i - \widetilde{\phi}_0 - \widetilde{\Phi}_1 z_{i0}\right)'.\tag{6.6}$$

The estimated mean and variance matrix of $\eta_i = (I - A)\mu_i$ can be obtained as:

$$\widetilde{\Omega}_\eta = \widetilde{\Theta}_0 + \widetilde{\Phi}_1\widetilde{\Sigma}_0\widetilde{\Phi}'_1 - \widetilde{\Omega}/T\tag{6.7}$$

$$\widetilde{\eta} = \widetilde{\phi}_0 + \widetilde{\Phi}_1\overline{z}_0\tag{6.8}$$

where $\overline{w}_0$ and $\widetilde{\Sigma}_0$ are the sample mean and variance of w_{i0}.

Finally, estimates of Υ_1 and τ_0 are given by

$$\widetilde{\Upsilon}_1 = \widetilde{\Sigma}_0\widetilde{\Phi}'_1\widetilde{\Omega}^{-1}_\eta\left(I - \widetilde{A}\right)\tag{6.9}$$

$$\widetilde{\tau}_0 = \overline{z}_0 - \widetilde{\Sigma}_0\widetilde{\Phi}'_1\widetilde{\Omega}^{-1}_\eta\widetilde{\eta}.\tag{6.10}$$

7. Inference with Feasible SIV Estimators

An unfeasible estimator for a parameterization of the instruments takes the form

$$\widehat{\beta}_{UF} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T-1}h_t(z^t_i, \gamma)x^{*\prime}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T-1}h_t(z^t_i, \gamma)y^*_{it}\right),\tag{7.1}$$

where γ is a pseudo true value that in practice will be defined by the probability limit of some sample statistic.

Under standard regularity conditions, for fixed T and large N,

$$\sqrt{N(T-1)}(\widehat{\beta}_{UF} - \beta) \xrightarrow{d} \mathcal{N}(0, V_{UF}) \tag{7.2}$$

with asymptotic variance matrix given by

$$V_{UF} = (T-1)\left[E\left(H_i'X_i^*\right)\right]^{-1} E\left(H_i'\varepsilon_i^*\varepsilon_i^{*\prime}H_i\right)\left[E\left(X_i^{*\prime}H_i\right)\right]^{-1} \tag{7.3}$$

where $H_i = \left(h_{i1}, ..., h_{i(T-1)}\right)'$, $h_{it} = h_t(z_i^t, \gamma)$, $X_i^* = \left(x_{i1}^*, ..., x_{i(T-1)}^*\right)'$, and $\varepsilon_i^* = \left(\varepsilon_{i1}^*, ..., \varepsilon_{i(T-1)}^*\right)'$.

Provided $E_t(\varepsilon_{it}^2) = \sigma^2$ and $E_{t+j}(\varepsilon_{it}\varepsilon_{i(t+j)}) = 0$ for $j > 0$ it will be the case that

$$V_{UF} = \sigma^2(T-1)\left[E\left(H_i'X_i^*\right)\right]^{-1} E\left(H_i'H_i\right)\left[E\left(X_i^{*\prime}H_i\right)\right]^{-1}. \tag{7.4}$$

If in addition $E_t(x_{it}^*) = h_t(z_i^t, \gamma)$ then V_{UF} coincides with the variance bound.

A feasible estimator $\widehat{\beta}_F$ is of the same form as $\widehat{\beta}_{UF}$ but γ is replaced by an estimate $\widehat{\gamma}$. Provided, $p\lim_{N\to\infty}\widehat{\gamma} = \gamma$ and $\sqrt{N}(\widehat{\gamma} - \gamma)$ is bounded in probability, the feasible and unfeasible estimators are asymptotically equivalent.

To see this first note that

$$\sqrt{N}(\widehat{\beta}_F - \beta) = \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-1} h_t(z_i^t, \gamma)x_{it}^{*\prime}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T-1} h_t(z_i^t, \widehat{\gamma})v_{it}^* + o_p(1).$$

Moreover, using the mean value theorem

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T-1} h_t(z_i^t, \widehat{\gamma})v_{it}^* = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T-1} h_t(z_i^t, \gamma)v_{it}^*$$
$$+ \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\frac{\partial h_t(z_i^t, \gamma)}{\partial\gamma'}v_{it}^*\right)\sqrt{N}(\widehat{\gamma} - \gamma) + o_p(1).$$

Since $E\left(v_{it}^* \mid z_i^t\right) = 0$, we have that $p\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-1} v_{it}^*\partial h_t(z_i^t, \gamma)/\partial\gamma' = 0$. Therefore, the second term on the *rhs* of the previous expression is $o_p(1)$. From this it follows that $\sqrt{N}(\widehat{\beta}_F - \beta) \xrightarrow{d} \mathcal{N}(0, V_{UF})$. Similar arguments can be used for the cases where T or both N and T tend to infinity.

Consistent Estimation of the Asymptotic Variance Matrix A natural estimate of the "Hessian" component of V_{UF} is

$$\widehat{\Psi}_{NT} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \widehat{h}_{it} x_{it}^{*\prime}, \tag{7.5}$$

where $\widehat{h}_{it} = h_t(z_i^t, \widehat{\gamma})$. $\widehat{\Psi}_{NT}$ is an N-consistent estimator of $(T-1)^{-1} E(H_i' X_i^*)$, but also an N and T consistent estimator of $p \lim_{T \to \infty, N \to \infty} \widehat{\Psi}_{NT}$.

The conventional robust fixed-T estimate of the "outer-product" component of V_{UF} is

$$\widetilde{\Upsilon}_{NT} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \widehat{H}_i' \widehat{\varepsilon}_i^* \widehat{\varepsilon}_i^{*\prime} \widehat{H}_i \tag{7.6}$$

where $\widehat{\varepsilon}_{it}^* = y_{it}^* - x_{it}^{*\prime} \widehat{\beta}_F$ and $\widehat{H}_i$ contains the estimated instruments $\widehat{h}_{it}$.

It can be shown that

$$\widetilde{\Upsilon}_{NT} = \widehat{\Omega}_0 + \sum_{\ell=1}^{T-2} \left(1 - \frac{\ell}{T-1}\right) \left(\widehat{\Omega}_\ell + \widehat{\Omega}_\ell'\right) \tag{7.7}$$

where

$$\widehat{\Omega}_\ell = \frac{1}{N(T-1-\ell)} \sum_{i=1}^{N} \sum_{t=\ell+1}^{T-1} \widehat{\varepsilon}_{it}^* \widehat{\varepsilon}_{i(t-\ell)}^* \widehat{h}_{it} \widehat{h}_{i(t-\ell)}' \tag{7.8}$$

for $\ell = 0, 1, ..., T-2$.[10]

From a non-fixed T perspective, we consider the following estimator of the outer-product component:

$$\widehat{\Upsilon}_{NT} = \widehat{\Omega}_0 + \sum_{\ell=1}^{r} \left(1 - \frac{\ell}{r+1}\right) \left(\widehat{\Omega}_\ell + \widehat{\Omega}_\ell'\right). \tag{7.9}$$

where $r \leq T-2$.

For $r = T-2$, $\widehat{\Upsilon}_{NT}$ particularizes to the standard robust fixed-T formula. In a fixed T context, it is natural to use $\widetilde{\Upsilon}_{NT}$, since the number of $\Omega_{i\ell}$ terms is fixed,

---

[10]Note that

$$\widehat{H}_i' \widehat{\varepsilon}_i^* \widehat{\varepsilon}_i^{*\prime} \widehat{H}_i = \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \widehat{\varepsilon}_{it}^* \widehat{\varepsilon}_{is}^{*\prime} \widehat{h}_{it} \widehat{h}_{is}' = (T-1)\Omega_{i0} + \sum_{\ell=1}^{T-2} (T-1-\ell)(\Omega_{i\ell} + \Omega_{i\ell}')$$

where $\Omega_{i\ell} = (T-1-\ell)^{-1} \sum_{t=\ell+1}^{T-1} \widehat{\varepsilon}_{it}^* \widehat{\varepsilon}_{i(t-\ell)}^* \widehat{h}_{it} \widehat{h}_{i(t-\ell)}'$ and $\widehat{\Omega}_\ell = N^{-1} \sum_{i=1}^{N} \Omega_{i\ell}$ for $\ell = 0, 1, ..., T-2$.

and large N ensures consistent estimation of all of them, including $\Omega_{i(T-2)}$, which is a time series average with a single observation. However, from a non-fixed T perspective, the bound r on the number of autocovariances used to form $\widehat{\Upsilon}_{NT}$ should be chosen as a suitable function of T to ensure consistent estimation for a given asymptotic arrangement in N and T.

Formula (7.9) is a panel data version of the estimator proposed in Newey and West (1987) for time series data. It has the attractive feature of providing a positive semidefinite estimator for any value of r.[11]

In the time series context, the term $[1 - \ell/(r + 1)]$ appearing in (7.9) is motivated as a damping factor for reducing the sampling error induced by higher-order sample covariances in the truncated estimate (called modified Bartlett weights in Anderson,1971, 511-513). It is interesting that these weights appear naturally as a feature of the sample covariance formula (7.7) from a cross-sectional perspective.

In the fixed-T panel data context, formula (7.6) is often used with instrument matrices whose number of columns is of order T or T^2, so that the dimension of $\widetilde{\Upsilon}_{NT}$ itself increases with T. Here, however, we are considering optimal instruments h_{it} with a fixed dimension for any value of T.

The proposal is, therefore, to base inference on the following estimate of the asymptotic covariance matrix of $\sqrt{N(T-1)}(\widehat{\beta}_F - \beta)$ for a chosen value of r:

$$\widehat{V}_{UF}^{\dagger} = \widehat{\Psi}_{NT}^{-1} \left[\widehat{\Omega}_0 + \sum_{\ell=1}^{r} \left(1 - \frac{\ell}{r+1} \right) \left(\widehat{\Omega}_\ell + \widehat{\Omega}_\ell' \right) \right] \widehat{\Psi}_{NT}'^{-1}. \tag{7.10}$$

Optimal Instruments with Completely Endogenous Explanatory Variables: A Two-Step GMM Method Now we can pursue the discussion on estimation when part of the x's are completely endogenous and the instruments have a larger dimension than β. We suggest using the inverse of $\widehat{\Upsilon}_{NT}$ as the weight matrix. Thus, we consider estimators of the form

$$\widetilde{\beta} = \left[\left(\sum_i \sum_t x_{it}^* \widehat{h}_{it}' \right) \widehat{\Upsilon}_{NT}^{-1} \left(\sum_i \sum_t \widehat{h}_{it} x_{it}^{*\prime} \right) \right]^{-1}$$
$$\left(\sum_i \sum_t x_{it}^* \widehat{h}_{it}' \right) \widehat{\Upsilon}_{NT}^{-1} \left(\sum_i \sum_t \widehat{h}_{it} y_{it}^* \right) \tag{7.11}$$

---

[11] The estimator in (7.9) relies on cross-sectional independence. A related formula for large T, fixed N within-group standard errors that allow for arbitrary cross-sectional dependence is proposed in Arellano (2003).

where $\widehat{h}_{it}$ is an estimate of $E_t\left(z^*_{it(q)}\right)$ as introduced in (5.3).

Note that $\widetilde{\beta}$ can be regarded as the SIV estimator that uses the following moments that have the same dimension of β:

$$\sum_i \sum_t \widehat{x}^{\dagger}_{it} \left(y^*_{it} - x^{*\prime}_{it}\widetilde{\beta}\right) = 0 \tag{7.12}$$

where $\widehat{x}^{\dagger}_{it} = \left(\sum_i \sum_t x^*_{it}\widehat{h}'_{it}\right) \widehat{\Upsilon}^{-1}_{NT}\widehat{h}_{it}$.

8. Empirical Illustrations

8.1. VAR for Firm Panel Data

We estimate autoregressive employment and wage equations from firm panel data. We first consider the dataset used by Alonso-Borrego and Arellano (1999). This is a balanced panel of 738 Spanish manufacturing companies, for which there are available annual observations for 1983-1990. Secondly we consider a longer panel of 385 firms from the same source (Bank of Spain, Central de Balances) on which 14 years of data are available, also starting in 1983. The average size of the firms in the second panel is more than twice as large as that of those in the first one.

We estimate a first-order VAR model for the logs of employment and wages, denoted n_{it} and w_{it} respectively. Individual and time effects are included in both equations. Time effects are removed prior to estimation by taking data in deviations from period-specific cross-sectional means.

Within-groups (WG), GMM, PML, and projection-restricted simple IV (SIV) estimates are reported. In this illustration PML appears in two different roles. First, it is an input in calculating feasible optimal instruments for SIV estimation. Second, since the auxiliary and substantive models coincide in this illustration, PML can be also regarded as an estimator of the parameters of interest that imposes time-series homoskedasticity.[12]

Table 1 shows the results. Focusing on the leading coefficient in the employment equation, first notice the larger size of the within-group estimates from the longer panel, which is to be expected since WG is known to have a small-T

---

[12]The PML estimates that we report use unrestricted initial conditions. According to estimates of Υ_1 for the two panels (reported in Table 1) there is some evidence of nonstationary initial conditions.

downward bias (Nickell, 1981). Next, the discrepancy between GMM and PML estimates is as noticeable as that between WG and GMM. Finally, SIV estimates are in between GMM and PML. The difference between GMM and PML may be due to finite-sample bias in GMM, heteroskedasticity bias in PML, or a combination of both. SIV estimates are as robust as GMM under a fixed-T asymptotics, but less prone to bias in a double asymptotics. In particular, SIV is robust to time-series heteroskedasticity in short panels whereas PML is not. Reported SIV estimates are half way between GMM and PML.

8.2. Monte Carlo Simulations

Next we performed a simulation exercise loosely calibrated to the previous firm panel dataset. The design was chosen from the partial adjustment representation as follows:

$$
\begin{align}
y_{it} &= 1 + 0.8y_{i(t-1)} - 0.5x_{it} + 0.3x_{i(t-1)} + \eta_i + v_{it} \tag{8.1} \\
x_{it} &= 0.5 + 0.3x_{i(t-1)} + \xi_i + \varepsilon_{it}, \tag{8.2}
\end{align}
$$

where all unobservables are *iid* normally distributed (over both i and t) with zero mean and $\sigma_v^2 = \sigma_\varepsilon^2 = 0.01$, $Corr(v_{it}, \varepsilon_{it}) = 0$, $\sigma_\eta^2 = \sigma_\xi^2 = 0.09$, and $Corr(\eta_i, \xi_i) = 0.6$. Therefore, the variance of the fixed effects is 9 times that of the random errors. There is no feedback from lagged y into x, and the long run effect of x on y is unity. Initial observations are generated from the stationary distribution of the process. The corresponding VAR is

$$
\begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.8 & 0.15 \\ 0 & 0.3 \end{pmatrix} \begin{pmatrix} y_{i(t-1)} \\ x_{i(t-1)} \end{pmatrix} + \begin{pmatrix} c_i + e_{it} \\ \xi_i + \varepsilon_{it} \end{pmatrix} \tag{8.3}
$$

where $e_{it} = v_{it} - 0.5\varepsilon_{it}$, $c_i = \eta_i - 0.5\xi_i$,

$$
Var\begin{pmatrix} e_{it} \\ \varepsilon_{it} \end{pmatrix} = \Omega = \begin{pmatrix} 0.0125 & -0.005 \\ -0.005 & 0.01 \end{pmatrix} \tag{8.4}
$$

and

$$
Var\begin{pmatrix} c_i \\ \xi_i \end{pmatrix} = \Omega_\eta = \begin{pmatrix} 0.0585 & 0.009 \\ 0.009 & 0.09 \end{pmatrix}. \tag{8.5}
$$

The implied correlations are $Corr(e_{it}, \varepsilon_{it}) = -0.447$ and $Corr(c_i, \xi_i) = 0.124$.

In Table 2 we report medians and median absolute errors of the WG, GMM, PML, and SIV estimators for $\{N = 738, T = 8\}$ and $\{N = 385, T = 14\}$.

In another experiment with $N = 738$, $T = 8$ we generated data with trending variances. The specification in this case was

$$\sigma_{vt}^2 = \sigma_{\varepsilon t}^2 = 0.01 + 0.001t, \tag{8.6}$$

with $t = 0$ selected in such a way that the resulting sequence of variances ranged from .005 to .012. This is also in Table 2. Finally, in Table 3 we report further simulations for $\{N = 200, T = 8\}, \{N = 100, T = 6\}$, and $\{N = 50, T = 15\}$. For all cases we conducted 1000 replications.

Focusing again on the leading coefficient in the first equation (a_{11} with true value of 0.8), Table 2 shows that GMM is downward biased but PML and SIV are virtually median unbiased. The robustness of SIV comes at the cost of a larger median absolute error than PML. The third panel of Table 2 for the experiment with trending variance shows that PML is upward biased, which goes in the direction of the empirical findings. Table 3 reports results for other sample sizes. Those in the first two experiments are similar to the Arellano–Bond firm-level data, and the cross-country panels used in growth studies, respectively. The last one illustrates the situation in a smaller panel that would be difficult to classify as either long T or large N, small T. In all cases SIV is unbiased and has reasonable median absolute errors, which is in contrast to some spectacular GMM biases.

8.3. Estimating Country Growth Convergence Rates

Using panel GMM, Caselli, Esquivel, and Lefort (1996) found a surprisingly large estimate of the convergence rate of about 10 percent. This was in sharp contrast with earlier cross-sectional estimates of Barro and Sala-i-Martin, who found convergence rates of 2-3 percent.[13] Caselli et al. claimed that earlier estimates were biased due to lack of proper control of country effects and predeterminedness.

The worry is that their estimates have finite sample downward biases in the GDP autoregressive coefficient that translate into upward biases in estimated convergence rates, as noted by Bond, Hoeffler, and Temple (2001).

We obtained the Caselli et al. data and re-estimated an augmented Solow model of the form given in (5.9) with the new projection-restricted IV estimator. We have found a smaller convergence rate of about 4 percent, but imprecisely

---

[13]See Barro and Sala-i-Martin (1995) for details and further references.

estimated (Table 4).[14] From a substantive point of view, however, explaining the uncovered heterogeneity in steady state income levels seems at least as crucial as finding good estimates of the convergence coefficient.

The standard errors reported in Table 4 were calculated from formula (7.10) with $r = T - 2$. Using smaller values of r made very little difference. Further exploration of the effects on the estimates and standard errors of using optimal instruments based on higher-order VAR models would be of some interest.

9. Concluding Remarks

We developed a new methodology for instrumental variable estimation of panel data models with general predetermined or endogenous explanatory variables. The suggested instruments are linear forecasts of the explanatory variables constructed under the assumption that the vector of conditioning variables follows a panel VAR process. These are linear combinations of all available lags as in ordinary GMM, but with the crucial difference that the number of first-stage coefficients is kept constant regardless of the value of T. We show analytically and through Monte Carlo simulations that this fundamentally alters the properties of the estimators in double asymptotics and finite samples. The new estimators eliminate double-asymptotic biases while retaining similar robustness and optimality properties as GMM in fixed T environments.

Specific discussions on unbalanced panels and higher-order auxiliary models are clearly of practical importance, but they are left for future work. GMM is not well suited to cope with the complications derived from typical patterns of unbalancedness or rotation in firm and household panels, which magnify the number of first stage coefficients. In the GMM context this has been avoided at the expense of introducing ad-hoc restrictions in the form of the cross-sectional projections across sub-panels—see Arellano and Bond (1991) for discussion. In contrast, projection-restricted IV estimators offer the possibility of a coherent specification of optimal instruments across sub-panels based on a small number of common coefficients.

---

[14]We use similar data as Caselli et al. and Bond et al. Our sample is the same as the one in Table 2 of Bond et al., except for the exclusion of 9 countries and 17 observations in order to have a balanced panel.

Appendix

A. Double-Asymptotic Results

Proof of Theorem 1

We shall use the following implication of Assumption 1. For all $r, s \geq 0$ we have

$$E_t \left(x_{i(t+r)} \varepsilon_{i(t+s)} \right) = E \left(x_{i(t+r)} \varepsilon_{i(t+s)} \right). \tag{A.1}$$

To see this note that

$$
\begin{aligned}
E \left(w_{i(t+r)} w'_{i(t+s)} \mid w_i^t, \mu_i \right) &= \mu_i \mu'_i + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Psi_j E \left(\zeta_{i(t+r-j)} \zeta'_{i(t+s-k)} \mid \zeta_i^t, \mu_i \right) \Psi'_k \\
&= \mu_i \mu'_i + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Psi_j E \left(\zeta_{i(t+r-j)} \zeta'_{i(t+s-k)} \right) \Psi'_k. \tag{A.2}
\end{aligned}
$$

By the law of iterated expectations

$$
\begin{aligned}
E \left(w_{i(t+r)} w'_{i(t+s)} \mid z_i^t \right) &= E \left[E \left(w_{i(t+r)} w'_{i(t+s)} \mid w_i^t, \mu_i \right) \mid z_i^t \right] \\
&= E \left(w_{i(t+r)} w'_{i(t+s)} \right) + E_t \left(\mu_i \mu'_i \right) - E \left(\mu_i \mu'_i \right). \tag{A.3}
\end{aligned}
$$

Equation (A.1) holds because ε_{it} does not have an individual effect (i.e the first component of μ_i is zero).

Let us examine the form of $E \left(X^{*\prime} M \varepsilon^* \right)$. We have

$$E \left(X^{*\prime} M \varepsilon^* \right) = \sum_{t=1}^{T-1} E \left(X_t^{*\prime} M_t \varepsilon_t^* \right) = m \sum_{t=1}^{T-1} t E \left(x_{it}^* \varepsilon_{it}^* \right). \tag{A.4}$$

To see this note that

$$
\begin{aligned}
E \left(X_t^{*\prime} M_t \varepsilon_t^* \right) &= \left\{ E \left(x_{\ell t}^{*\prime} M_t \varepsilon_t^* \right) \right\} = \left\{ E \operatorname{tr} \left[M_t E_t \left(\varepsilon_t^* x_{\ell t}^{*\prime} \right) \right] \right\} \\
&= m t E \left(\varepsilon_{it}^* x_{it}^* \right). \tag{A.5}
\end{aligned}
$$

The last equality comes from (A.1), $\operatorname{tr}(M_t) = mt$, and the fact that due to cross-sectional independence:

$$
\begin{aligned}
E_t \left(\varepsilon_t^* x_{\ell t}^{*\prime} \right) &= \begin{pmatrix} E_t \left(\varepsilon_{1t}^* x_{\ell 1 t}^* \right) & \cdots & E_t \left(\varepsilon_{1t}^* x_{\ell N t}^* \right) \\ \vdots & \ddots & \vdots \\ E_t \left(\varepsilon_{Nt}^* x_{\ell 1 t}^* \right) & \cdots & E_t \left(\varepsilon_{Nt}^* x_{\ell N t}^* \right) \end{pmatrix} \tag{A.6} \\
&= \begin{pmatrix} E_t \left(\varepsilon_{1t}^* x_{\ell 1 t}^* \right) & \cdots & E_t \left(\varepsilon_{1t}^* \right) E_t \left(x_{\ell N t}^* \right) \\ \vdots & \ddots & \vdots \\ E_t \left(\varepsilon_{Nt}^* \right) E_t \left(x_{\ell 1 t}^* \right) & \cdots & E_t \left(\varepsilon_{Nt}^* x_{\ell N t}^* \right) \end{pmatrix} = E_t \left(\varepsilon_{it}^* x_{\ell i t}^* \right) I_N.
\end{aligned}
$$

27

Next, consider

$$E\left(x_{it}^{*}\varepsilon_{it}^{*}\right) = \left(\frac{T-t}{T-t+1}\right)\left\{E\left(\varepsilon_{it}x_{it}\right) - \frac{1}{(T-t)}E\left[\varepsilon_{it}\left(x_{i(t+1)} + ... + x_{iT}\right)\right]\right.$$
$$-\frac{1}{(T-t)}E\left[\left(\varepsilon_{i(t+1)} + ... + \varepsilon_{iT}\right)x_{it}\right]$$
$$\left.+\frac{1}{(T-t)^2}E\left[\left(\varepsilon_{i(t+1)} + ... + \varepsilon_{iT}\right)\left(x_{i(t+1)} + ... + x_{iT}\right)\right]\right\}. \quad (A.7)$$

Collecting terms we obtain

$$E\left(x_{it}^{*}\varepsilon_{it}^{*}\right) = \gamma_0 - \frac{1}{(T-t)(T-t+1)}$$
$$\times\left\{(T-t)\gamma_{-(T-t)} + ... + \gamma_{-1} + \gamma_1 + ... + (T-t)\gamma_{(T-t)}\right\}. (A.8)$$

Thus

$$E\left(X^{*\prime}M\varepsilon^{*}\right) = m\sum_{t=1}^{T-1}tE\left(x_{it}^{*}\varepsilon_{it}^{*}\right)$$
$$= m\left(\gamma_0\sum_{t=1}^{T-1}t - \sum_{t=1}^{T-1}\frac{t}{(T-t)(T-t+1)}B_{(T-t)}\right) \quad (A.9)$$

where

$$B_{(T-t)} = (T-t)\gamma_{-(T-t)} + ... + \gamma_{-1} + \gamma_1 + ... + (T-t)\gamma_{(T-t)}. \quad (A.10)$$

We now use the following facts:

$$\sum_{t=1}^{T-1}t = \frac{1}{2}(T-1)T \quad (A.11)$$

$$\sum_{t=1}^{T-1}\frac{t}{(T-t)(T-t+1)} = T - \sum_{t=1}^{T}\frac{1}{t} \quad (A.12)$$

$$\left|B_{(T-t)}\right| < |B_{\infty}| \quad (A.13)$$

Hence, when x_{it} is *endogenous* $(\gamma_0 \neq 0)$ we have $E\left(X^{*\prime}M\varepsilon^{*}\right) = O\left(mT^2\right)$ or

$$E\left(\frac{X^{*\prime}M\varepsilon^{*}}{NT}\right) = O\left(\frac{mT}{N}\right). \quad (A.14)$$

In contrast, when x_{it} is *predetermined* we have $E\left(X^{*\prime}M\varepsilon^{*}\right) = O\left(mT\right)$ or

$$E\left(\frac{X^{*\prime}M\varepsilon^{*}}{NT}\right) = O\left(\frac{m}{N}\right). \quad (A.15)$$

B. Auxiliary VAR with Individual Effects

B.1. Sequential linear projections of the effects

Proof of Theorem 2

I show that for the VAR(1) model presented in the main text, the linear projection of the vector of individual effects μ_i on $z_i^t = (z_{i0}', ..., z_{it}')'$ is given by

$$
E^* \left(\mu_i \mid z_i^t \right) = \left[\Omega_\mu^{-1} + \Upsilon_1' \Gamma_0^{-1} \Upsilon_1 + (I - A)' \left(\sum\nolimits_{s=1}^t \Omega_s^{-1} \right) (I - A) \right]^{-1}
$$

$$
\times \left[(\Omega_\mu^{-1} \mu - \Upsilon_1' \Gamma_0^{-1} \tau_0) + (I - A)' \sum\nolimits_{s=1}^t \Omega_s^{-1} \left(z_{is} - A z_{i(s-1)} \right) + \Upsilon_1' \Gamma_0^{-1} z_{i0} \right]
$$

$$
= \left[I + \Lambda_0 \Upsilon_1 + \sum_{s=1}^t \Lambda_{1s} (I - A) \right]^{-1} \left[\mu + \Lambda_0 \left(z_{i0} - \tau_0 \right) + \sum_{s=1}^t \Lambda_{1s} \left(z_{is} - A z_{i(s-1)} \right) \right]
$$

$$(\text{B.1})$$

for $t \geq 1$, and for $t = 0$:

$$
\begin{aligned}
E^* \left(\mu_i \mid z_i^t \right) &= \left[\Omega_\mu^{-1} + \Upsilon_1' \Gamma_0^{-1} \Upsilon_1 \right]^{-1} \left[\Omega_\mu^{-1} \mu + \Upsilon_1' \Gamma_0^{-1} \left(z_{i0} - \tau_0 \right) \right] \\
&= \left[I + \Lambda_0 \Upsilon_1 \right]^{-1} \left[\mu + \Lambda_0 \left(z_{i0} - \tau_0 \right) \right]
\end{aligned}
\qquad (\text{B.2})
$$

where $\Lambda_0 = \Omega_\mu \Upsilon_1' \Gamma_0^{-1}$ and $\Lambda_{1s} = \Omega_\mu (I - A)' \Omega_s^{-1}$.

The VAR model implies that

$$
z_{it} = (I - A^t)\mu_i + \left(v_{it} + A v_{i(t-1)} + ... + A^{t-1} v_{i1} \right) + A^t z_{i0}. \qquad (\text{B.3})
$$

Substituting (4.4) in (B.3) we obtain

$$
z_{it} = \left[I + A^t \left(\Upsilon_1 - I \right) \right] \mu_i + A^t \tau_0 + z_{it}^* \qquad (\text{B.4})
$$

where

$$
z_{it}^* = v_{it} + A v_{i(t-1)} + ... + A^{t-1} v_{i1} + A^t v_{i0}^\dagger. \qquad (\text{B.5})
$$

Let G denote the $(t+1) m \times m$ matrix $G = (I_m, A', A^{2\prime}, ..., A^{t\prime})'$. Then

$$
z_i^t = F \mu_i + G \tau_0 + z_i^{*t} \qquad (\text{B.6})
$$

where

$$
F = (\iota \otimes I) + G \left(\Upsilon_1 - I \right), \qquad (\text{B.7})
$$

ι is a $(t+1) \times 1$ vector of ones, and $z_i^{*t} = (z_{i0}^{*\prime}, ..., z_{it}^{*\prime})'$. Therefore,

$$
E \left(z_i^t \right) = F \mu + G \tau_0 \qquad (\text{B.8})
$$

29

$$Var\left(z_i^t\right) = F\Omega_\mu F' + V \tag{B.9}$$

where $V = Var\left(z_i^{*t}\right)$. Moreover, note that

$$Cov\left(z_i^t, \mu_i\right) = F\Omega_\mu. \tag{B.10}$$

The linear projection is given by

$$E^*\left(\mu_i \mid z_i^t\right) = \psi_t + \Pi_t' z_i^t \tag{B.11}$$

where

$$\psi_t = \mu - \Pi_t' E\left(z_i^t\right) \tag{B.12}$$

and

$$\Pi_t = \left[Var\left(z_i^t\right)\right]^{-1} Cov\left(z_i^t, \mu_i\right). \tag{B.13}$$

Letting $\Omega_\mu = PP'$ and $M = FP$, and using the matrix inversion lemma

$$\left[Var\left(z_i^t\right)\right]^{-1} = (MM' + V)^{-1} = V^{-1} - V^{-1}M\left(I + M'V^{-1}M\right)^{-1}M'V^{-1}.$$

Moreover, letting $V^{-1} = B'B$ where B is a block-lower triangular matrix

$$
\begin{aligned}
\Pi_t' z_i^t &= \Omega_\mu F'\left[V^{-1} - V^{-1}M\left(I + M'V^{-1}M\right)^{-1}M'V^{-1}\right]z_i^t \\
&= PM'\left[B'B - B'BM\left(I + M'B'BM\right)^{-1}M'B'B\right]z_i^t \\
&= P\left[I - \left(M'B'\right)\left(BM\right)\left(I + M'B'BM\right)^{-1}\right]\left(M'B'\right)\left(Bz_i^t\right) \\
&= P\left[I + \left(M'B'\right)\left(BM\right)\right]^{-1}\left(M'B'\right)\left(Bz_i^t\right). \tag{B.14}
\end{aligned}
$$

Since $V = Var\left(z_i^{*t}\right)$, the matrix B has to be such that $Var\left(Bz_i^{*t}\right) = I$. Therefore, since B must satisfy

$$Bz_i^{*t} = \begin{pmatrix} \Gamma_0^{-1/2}v_{i0}^\dagger \\ \Omega_1^{-1/2}v_{i1} \\ \vdots \\ \Omega_t^{-1/2}v_{it} \end{pmatrix}, \tag{B.15}$$

it is given by

$$B = \begin{pmatrix} \Gamma_0^{-1/2} & 0 & 0 & \dots & 0 & 0 \\ -\Omega_1^{-1/2}A & \Omega_1^{-1/2} & 0 & & 0 & 0 \\ 0 & -\Omega_2^{-1/2}A & \Omega_2^{-1/2} & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & & \Omega_{t-1}^{-1/2} & 0 \\ 0 & 0 & 0 & \dots & -\Omega_t^{-1/2}A & \Omega_t^{-1/2} \end{pmatrix}. \tag{B.16}$$

Direct multiplication gives

$$BM = B\left[(\iota \otimes I) + G\left(\Upsilon_1 - I\right)\right]P = \begin{pmatrix} \Gamma_0^{-1/2}\Upsilon_1 \\ \Omega_1^{-1/2}\left(I - A\right) \\ \vdots \\ \Omega_t^{-1/2}\left(I - A\right) \end{pmatrix} P \qquad (\text{B.17})$$

and

$$Bz_i^t = \begin{pmatrix} \Gamma_0^{-1/2}z_{i0} \\ \Omega_1^{-1/2}\left(z_{i1} - Az_{i0}\right) \\ \vdots \\ \Omega_t^{-1/2}\left(z_{it} - Az_{i(t-1)}\right) \end{pmatrix}. \qquad (\text{B.18})$$

Hence

$$\left(M'B'\right)\left(BM\right) = P'\left[\Upsilon_1'\Gamma_0^{-1}\Upsilon_1 + \left(I - A\right)'\left(\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\right)\left(I - A\right)\right]P \qquad (\text{B.19})$$

and

$$\left(M'B'\right)\left(Bz_i^t\right) = P'\left[\Upsilon_1'\Gamma_0^{-1}z_{i0} + \left(I - A\right)'\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\left(z_{is} - Az_{i(s-1)}\right)\right]. \qquad (\text{B.20})$$

Inserting these terms in the expression given above

$$\begin{aligned} \Pi_t'z_i^t &= \left[P^{-1'}P^{-1} + \Upsilon_1'\Gamma_0^{-1}\Upsilon_1 + \left(I - A\right)'\left(\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\right)\left(I - A\right)\right]^{-1} \\ &\quad \left[\Upsilon_1'\Gamma_0^{-1}z_{i0} + \left(I - A\right)'\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\left(z_{is} - Az_{i(s-1)}\right)\right]. \end{aligned} \qquad (\text{B.21})$$

Finally, to obtain the vector of intercepts, note that since $E\left(z_{i0}\right) = \tau_0 + \Upsilon_1\mu$ and

$$E\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\left(z_{is} - Az_{i(s-1)}\right) = \left(\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\right)\left(I - A\right)\mu, \qquad (\text{B.22})$$

we have

$$\begin{aligned} \Pi_t'E\left(z_i^t\right) &= \left[\Omega_\mu^{-1} + \Upsilon_1'\Gamma_0^{-1}\Upsilon_1 + \left(I - A\right)'\left(\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\right)\left(I - A\right)\right]^{-1} \\ &\quad \left[\Upsilon_1'\Gamma_0^{-1}\tau_0 + \Upsilon_1'\Gamma_0^{-1}\Upsilon_1\mu + \left(I - A\right)'\left(\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\right)\left(I - A\right)\mu\right]. \end{aligned} \qquad (\text{B.23})$$

Collecting terms we obtain

$$\begin{aligned} \psi_t &= \mu - \Pi_t'E\left(z_i^t\right) \\ &= \left[\Omega_\mu^{-1} + \Upsilon_1'\Gamma_0^{-1}\Upsilon_1 + \left(I - A\right)'\left(\sum\nolimits_{s=1}^{t}\Omega_s^{-1}\right)\left(I - A\right)\right]^{-1} \\ &\quad \left[\Omega_\mu^{-1}\mu - \Upsilon_1'\Gamma_0^{-1}\tau_0\right] \end{aligned} \qquad (\text{B.24})$$

31

from which the result follows.

A Stationary VAR(1) Process

The stationary case is obtained as a specialization of the initial VAR model to $\tau_0 = 0$, $\Upsilon_1 = I_m$, $\Omega_t = \Omega$ for all t, and $\Gamma_0 = \sum_{j=0}^{\infty} A^j \Omega A^{j\prime}$, so that Γ_0 satisfies $\Gamma_0 = A\Gamma_0 A' + \Omega$. In such case we have

$$E\left(w_i^t\right) = \iota \otimes \mu \tag{B.25}$$

$$Var\left(w_i^t\right) = \left(\iota\iota' \otimes \Omega_\mu\right) + V \tag{B.26}$$

$$Cov\left(w_i^t, \mu_i\right) = \iota \otimes \Omega_\mu. \tag{B.27}$$

Moreover, V is given by

$$V = \begin{pmatrix} \Gamma_0 & \Gamma_1' & \cdots & \Gamma_t' \\ \Gamma_1 & \Gamma_0 & & \Gamma_{t-1}' \\ \vdots & & \ddots & \\ \Gamma_t & \Gamma_{t-1} & \cdots & \Gamma_0 \end{pmatrix}. \tag{B.28}$$

where $\Gamma_j = A^j \Gamma_0$. Nevertheless, the matrix B in the decomposition of the inverse of V is of the same form as in the general case but with constant Ω.

There are two special cases of equation (4.15) that are of some interest:

1. *Uncorrelated Multivariate Error Components ($A = 0$).* In this case $\Gamma_0 = \Omega$ and $\Lambda_1 = \Lambda_0$, so that

$$E_t\left(z_{it}^*\right) = c_t\left[z_{it} - E_t\left(\mu_i\right)\right] \tag{B.29}$$

    and

$$E_t\left(\mu_i\right) = \left[I + (1+t)\Lambda_0\right]^{-1}\left[\mu + \Lambda_0 \sum_{s=0}^{t} z_{is}\right]. \tag{B.30}$$

2. *Homogeneous VAR(1) process ($\Omega_\mu = 0$):* In this case $\delta_{it} = \delta_{i0} = \mu$ and $\mathcal{H}_t = \mathcal{H}_0 = I$, so that $E_t\left(\mu_i\right) = \mu$ and

$$E_t\left(z_{it}^*\right) = c_t\left(I - \frac{1}{T-t}\sum_{s=1}^{t} A^s\right)(z_{it} - \mu). \tag{B.31}$$

B.2. VAR Log Likelihood Given Initial Observations

The VAR model can be written as

$$
\begin{pmatrix}
I & 0 & \dots & 0 & 0 \\
-A & I & & 0 & 0 \\
\vdots & & \ddots & & \vdots \\
0 & 0 & & I & 0 \\
0 & 0 & \dots & -A & I
\end{pmatrix}
\begin{pmatrix}
z_{i1} \\
z_{i2} \\
\vdots \\
z_{i(T-1)} \\
z_{iT}
\end{pmatrix}
=
\begin{pmatrix}
I_m \\
0 \\
\vdots \\
0 \\
0
\end{pmatrix}
A z_{i0} +
\begin{pmatrix}
u_{i1} \\
u_{i2} \\
\vdots \\
u_{i(T-1)} \\
u_{iT}
\end{pmatrix},
\quad \text{(B.32)}
$$

or using a compact notation

$$
B z_i = D A z_{i0} + u_i \tag{B.33}
$$

$$
u_i = (\iota \otimes \eta_i) + v_i \tag{B.34}
$$

where $u_{it} = \eta_i + v_{it}$, $\eta_i = (I - A)\mu_i$, ι is a vector of ones of order T and $v_i = (v_{i1}', ..., v_{iT}')'$. Thus, under time series homoskedasticity

$$
E(u_i) = \iota \otimes \eta \tag{B.35}
$$

and

$$
Var(u_i) = (\iota\iota' \otimes \Omega_\eta) + (I_T \otimes \Omega) \tag{B.36}
$$

where $\eta = (I - A)\mu$ and $\Omega_\eta = (I - A)\Omega_\mu(I - A)'$.

The conditional density of z_i given z_{i0} is related to that of u_i by

$$
f(z_i \mid z_{i0}) = f(u_i \mid z_{i0}) \det(B) \tag{B.37}
$$

but $\det(B) = 1$ because B is a triangular matrix. Moreover

$$
f(u_i \mid z_{i0}) = f(\overline{u}_i, u_i^* \mid z_{i0}) \left|\det(H \otimes I_m)\right| \tag{B.38}
$$

where $H = (\iota/T, \mathcal{A}')'$ is a $T \times T$ transformation matrix, and $\mathcal{A}$ is the $(T-1) \times T$ forward orthogonal deviations operator, that produces $(H \otimes I_m) u_i = (\overline{u}_i', u_i^{*\prime})'$. The determinant of the transformation satisfies $\left|\det(H \otimes I_m)\right| = T^{-m/2}$, which is an irrelevant constant.

Note that

$$
E\left[(H \otimes I_m) u_i\right] = H\iota \otimes \eta = \begin{pmatrix} \eta \\ 0 \end{pmatrix} \tag{B.39}
$$

$$
\begin{aligned}
Var\left[(H \otimes I_m) u_i\right] &= (H \otimes I_m)\left[(\iota\iota' \otimes \Omega_\eta) + (I_T \otimes \Omega)\right](H' \otimes I_m) \\
&= (H\iota\iota'H' \otimes \Omega_\eta) + (H'H \otimes \Omega) \\
&= \begin{pmatrix} \Omega_\eta + T^{-1}\Omega & 0 \\ 0 & I_{T-1} \otimes \Omega \end{pmatrix}.
\end{aligned} \tag{B.40}
$$

33

In order to obtain the mean and variance matrix of $\{(H \otimes I_m) u_i \mid z_{i0}\}$ it is convenient to introduce some additional notation. Let $\Sigma_0 = Var(z_{i0}) = \Upsilon_1 \Omega_\mu \Upsilon_1' + \Gamma_0$ and let the linear projection of η_i on z_{i0} be

$$E^* (\eta_i \mid z_{i0}) = \phi_0 + \Phi_1 z_{i0} = (\phi_0, \Phi_1) \begin{pmatrix} 1 \\ z_{i0} \end{pmatrix} \equiv \Phi f_{i0} \qquad (B.41)$$

so that $\Phi_1 = (I - A)\Omega_\mu \Upsilon_1' \Sigma_0^{-1}$.[15] Thus under joint normality of $\eta_i, z_{i0}, z_{i1}, ..., z_{iT}$:

$$E[(H \otimes I_m) u_i \mid z_{i0}] = \begin{pmatrix} \phi_0 + \Phi_1 z_{i0} \\ 0 \end{pmatrix} \qquad (B.42)$$

and

$$\begin{aligned} Var[(H \otimes I_m) u_i \mid z_{i0}] &= Var[(H \otimes I_m) u_i] - Var \begin{pmatrix} \phi_0 + \Phi_1 z_{i0} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \Theta_0 & 0 \\ 0 & I_{T-1} \otimes \Omega \end{pmatrix} \end{aligned} \qquad (B.43)$$

where $\Theta_0 = (\Omega_\eta - \Phi_1 \Sigma_0 \Phi_1') + T^{-1}\Omega$.[16]

Then, under normality

$$\begin{aligned} \ln f(z_{i1}, ..., z_{iT} \mid z_{i0}) &= -\frac{(T-1)}{2} \ln \det \Omega - \frac{1}{2} u_i^{*\prime} (I_{T-1} \otimes \Omega^{-1}) u_i^* \qquad (B.44) \\ &\quad -\frac{1}{2} \ln \det \Theta_0 - \frac{1}{2} (\overline{u}_i - \phi_0 - \Phi_1 z_{i0})' \Theta_0^{-1} (\overline{u}_i - \phi_0 - \Phi_1 z_{i0}). \end{aligned}$$

C. Estimating Scalar Autoregressive Models

Let us consider first a scalar AR(1) model with individual effects of the form

$$y_{it} = \alpha y_{i(t-1)} + (1 - \alpha) \mu_i + v_{it} \quad \mid \alpha \mid < 1 \qquad (C.1)$$

$$E(v_{it} \mid y_{i0}, y_{i1}, ..., y_{i(t-1)}) = 0 \qquad (C.2)$$

---

[15]Note that

$$\begin{aligned} \Phi_1 &= Cov(\eta_i, w_{i0})\Sigma_0^{-1} = (I - A)Cov(\mu_i, w_{i0})\Sigma_0^{-1} \\ &= (I - A)\Omega_\mu \Omega_\mu^{-1} Cov(\mu_i, w_{i0})\Sigma_0^{-1} = (I - A)\Omega_\mu \Upsilon_1' \Sigma_0^{-1}. \end{aligned}$$

[16]Note that $Var(\eta_i \mid w_{i0}) = (\Omega_\eta - \Phi_1 \Sigma_0 \Phi_1')$.

so that in the notation of (2.1) and (2.2) x_{it} and z_{it} are both scalar variables and $x_{it} = z_{it} = y_{i(t-1)}$. For convenience we assume that y_{i0} is observed. In this case

$$\frac{1}{c_t} E\left(x_{it}^* \mid y_i^{t-1}\right) = y_{i(t-1)} - \frac{1}{T-t}\left[E\left(y_{it} \mid y_i^{t-1}\right) + ... + E\left(y_{i(T-1)} \mid y_i^{t-1}\right)\right]. \quad \text{(C.3)}$$

Note that since

$$
\begin{aligned}
E\left(y_{i(t+s)} \mid y_i^{t-1}\right) &= \alpha^{s+1} y_{i(t-1)} + (1 + \alpha + ... + \alpha^s)E\left((1-\alpha)\mu_i \mid y_i^{t-1}\right) \\
&= E\left(\mu_i \mid y_i^{t-1}\right) + \alpha^{s+1}\left[y_{i(t-1)} - E\left(\mu_i \mid y_i^{t-1}\right)\right] \quad (s = 0, ..., T-1-t),
\end{aligned}
$$

the instrument $E\left(x_{it}^* \mid y_i^{t-1}\right)$ is just a function of $y_{i(t-1)}$ and $E\left(\mu_i \mid y_i^{t-1}\right)$ given by

$$E\left(x_{it}^* \mid y_i^{t-1}\right) = c_t\left(1 - \frac{1}{T-t}\sum_{s=1}^{T-t}\alpha^s\right)\left[y_{i(t-1)} - E\left(\mu_i \mid y_i^{t-1}\right)\right]. \quad \text{(C.4)}$$

In general $E\left(\mu_i \mid y_i^{T-1}\right)$ can be a nonlinear function of y_i^{T-1}, but in the auxiliary models we shall assume it coincides with the linear projection

$$E^*\left(\mu_i \mid y_i^{T-1}\right) = \psi_{T-1} + \pi_{T-1}' y_i^{t-1}. \quad \text{(C.5)}$$

In view of the results of the previous section, it turns out that the coefficients of this projection are unrestricted when the unconditional variances $E\left(v_{it}\right) = \sigma_t^2$ are allowed to change with t in an unspecified way. But even in this case note that due to the law of iterated projections

$$E^*\left(\mu_i \mid y_i^{t-1}\right) = E^*\left[E^*\left(\mu_i \mid y_i^t\right) \mid y_i^{t-1}\right],$$

all the instruments $E\left(x_{it}^* \mid y_i^{t-1}\right)$ can be written as functions of α, ψ_{T-1}, and γ_{T-1}. This substantially reduces the number of parameters relative to the unrestricted linear projections for $E\left(x_{it}^* \mid y_i^{t-1}\right)$ used implicitly by standard GMM, despite being based on the same auxiliary model. Nevertheless, the number of coefficients still increases with T.

A Strictly Stationary Auxiliary Model   As an auxiliary model we may consider a strictly stationary AR(1) process, in which case, using the results from Appendix B, we obtain

$$E^*\left(\mu_i \mid y_i^{t-1}\right) = m_{t-1}\left(y_i^{t-1}, \theta\right) \equiv \frac{\mu + \phi\left[(1-\alpha)\sum_{s=1}^{t-1} u_{is} + \left(1-\alpha^2\right)y_{i0}\right]}{1 + \phi\left[(t-1)(1-\alpha)^2 + 1 - \alpha^2\right]} \quad \text{(C.6)}$$

where $\phi = \sigma_\mu^2/\sigma^2$, $\theta = (\alpha, \phi, \mu)'$, and $u_{is} = y_{is} - \alpha y_{i(s-1)}$.

Letting $\lambda = \sigma_\eta^2/\sigma^2 = (1-\alpha)^2\phi$, an equivalent expression is given by

$$
E^* \left(\mu_i \mid y_i^{t-1} \right) = \frac{\mu}{1 + \lambda \left(t - 1 + \frac{1+\alpha}{(1-\alpha)} \right)}
$$

$$
+ \frac{\lambda}{1 + \lambda \left(t - 1 + \frac{1+\alpha}{(1-\alpha)} \right)} \left[\sum_{s=0}^{t-1} y_{is} + \frac{\alpha}{1-\alpha} \left(y_{i0} + y_{i(t-1)} \right) \right]. \quad \text{(C.7)}
$$

Note that as $t \to \infty$ $E^* \left(\mu_i \mid y_i^{t-1} \right)$ converges to the limit of $t^{-1} \sum_{s=0}^{t-1} y_{is}$ (which is given by μ_i), but for small values of t the approximation of $E^* \left(\mu_i \mid y_i^{t-1} \right)$ by $E^* \left(\mu_i \mid \sum_{s=0}^{t-1} y_{is} \right)$ may be poor, even if y_{it} is a strictly stationary process (it would only be appropriate if $\alpha = 0$).[17]

Using as the auxiliary model the assumptions of strict stationarity together with the linearity of $E \left(\mu_i \mid y_i^{t-1} \right)$, the instrument $E \left(x_{it}^* \mid y_i^{t-1} \right)$ becomes a linear function of $\sum_{s=0}^{t-1} u_{is}$ and y_{i0} with coefficients that depend exclusively on α, ϕ and μ. Since these assumptions need not be true, the previous coefficients should be understood as pseudo true values for which we use the notation $c = (a, f, m)'$. Thus our parameterization of the instrument is given by

$$
h_t \left(y_i^{t-1}, c \right) = c_t \left[1 - \frac{a}{1-a} \left(\frac{1 - a^{T-t}}{T-t} \right) \right] \left[y_{i(t-1)} - m_{t-1} \left(y_i^{t-1}, c \right) \right]. \quad \text{(C.8)}
$$

An Auxiliary Model with Unrestricted Initial Conditions   An alternative, more general auxiliary model that retains a fixed number of parameters is one in which the assumption of stationarity of initial observations is removed. This adds three extra parameters to the instrument function.

The previous auxiliary model assumed that the linear projection of y_{i0} on μ_i

$$
y_{i0} = \tau_0 + \tau_1 \mu_i + v_{i0} \quad \text{(C.9)}
$$

was such that $\tau_0 = 0$, $\tau_1 = 1$ and $Var(v_{i0}) = \gamma_0^2 = \sigma^2/(1-\alpha^2)$. In the alternative model τ_0, τ_1 and γ_0^2 are free parameters. Using again the results from Appendix B we obtain

$$
E^* \left(\mu_i \mid y_i^{t-1} \right) = m_{t-1}^\dagger \left(y_i^{t-1}, \theta \right) \equiv \frac{(\mu - r_0 \tau_1 \tau_0) + \phi (1-\alpha) \sum_{s=1}^{t-1} u_{is} + r_0 \tau_1 y_{i0}}{1 + \phi(t-1)(1-\alpha)^2 + r_0 \tau_1^2}
$$

$$
\text{(C.10)}
$$

---

[17]With $T = 2$, we only need to consider $t = 1$, and there is just one instrument given by

$$
E(x_{i1}^* \mid y_{i0}) = 2^{-1/2} \frac{(1-\alpha)^2}{(1-\alpha) + \lambda(1+\alpha)} (y_{i0} - \mu).
$$

where $r_0 = \sigma_\mu^2/\gamma_0^2$ and $\theta = (\alpha, \phi, \mu, \tau_0, \tau_1, r_0)'$. Note that under stationarity $r_0 = \phi(1 - \alpha^2)$.

For large t the difference between (C.6) and (C.10) will be small regardless of the values of τ_0, τ_1 and γ_0^2, but in short panels the difference may be important if the steady state distribution of the process is not a good approximation to the distribution of initial observations. In such a case a better choice of instrument will be

$$h_t^\dagger\left(y_i^{t-1}, c\right) = c_t \left[1 - \frac{a}{1-a}\left(\frac{1-a^{T-t}}{T-t}\right)\right]\left[y_{i(t-1)} - m_{t-1}^\dagger\left(y_i^{t-1}, c\right)\right]. \qquad \text{(C.11)}$$

AR(p) Processes The previous discussion can be extended to a stable autoregressive process of order p:

$$y_{it} = \alpha_1 y_{i(t-1)} + ... + \alpha_p y_{i(t-p)} + \eta_i + v_{it} \qquad \text{(C.12)}$$

$$E(v_{it} \mid y_{i0}, y_{i1}, ..., y_{i(t-1)}) = 0, \qquad \text{(C.13)}$$

so that in terms of the notation of model (2.1) and (2.2), we have $x_{it} = (y_{i(t-1)}, ..., y_{i(t-p)})'$ and $z_{it} = y_{i(t-1)}$. For convenience we assume that $y_{i0}, y_{i(-1)}, ..., y_{i(-p+1)}$ are observed (i.e. the time series dimension of the panel is $T + p$), and write the model in companion form

$$x_{i(t+1)} = \Pi x_{it} + d_1\left(\eta_i + v_{it}\right) \qquad \text{(C.14)}$$

where $d_1 = (1, 0, ..., 0)'$ of order p, and Π is the $p \times p$ matrix

$$\Pi = \begin{pmatrix} \alpha_1 & \cdots & \alpha_p \\ I_{p-1} & \vdots & 0 \end{pmatrix}. \qquad \text{(C.15)}$$

Therefore,

$$\frac{1}{c_t}E\left(x_{it}^* \mid y_i^{t-1}\right) = x_{it} - \frac{1}{T-t}\left[E\left(x_{i(t+1)} \mid y_i^{t-1}\right) + ... + E\left(x_{iT} \mid y_i^{t-1}\right)\right]. \qquad \text{(C.16)}$$

Moreover, since

$$E\left(x_{i(t+s)} \mid y_i^{t-1}\right) = \Pi^s x_{it} + \left(I_p + \Pi + ... + \Pi^{s-1}\right)d_1 E\left(\eta_i \mid y_i^{t-1}\right) \quad (s = 1, ..., T-t), \qquad \text{(C.17)}$$

the vector of instruments $E\left(x_{it}^* \mid y_i^{t-1}\right)$ is a function of x_{it} and $E\left(\eta_i \mid y_i^{t-1}\right)$ given by

$$E\left(x_{it}^* \mid y_i^{t-1}\right) = c_t \left[I - \frac{1}{T-t}\Pi\left(I - \Pi^{T-t}\right)\left(I - \Pi\right)^{-1}\right]\left[x_{it} - \iota_p E\left(\mu_i \mid y_i^{t-1}\right)\right] \qquad \text{(C.18)}$$

where $\eta_i = (1 - \alpha_1 - ... - \alpha_p)\mu_i$.[18]

Under the strictly stationary auxiliary model, the linear projection $E^* \left(\mu_i \mid y_i^{t-1} \right)$ is still of the form

$$
\begin{aligned}
E^* \left(\mu_i \mid y_i^{t-1} \right) &= \mu + \phi\iota' \left(\phi\iota\iota' + V \right)^{-1} \left(y_i^{t-1} - \mu\iota \right) \\
&= \frac{\mu}{1 + \phi \left(\iota'V^{-1}\iota \right)} + \frac{\phi}{1 + \phi \left(\iota'V^{-1}\iota \right)} \iota'V^{-1}y_i^{t-1}, \qquad \text{(C.19)}
\end{aligned}
$$

but now $V = V(\alpha_1, ..., \alpha_p)$ corresponds to the autoregressive covariance matrix of order p.

---

[18]Note that
$$
(I - \Pi)^{-1} d_1 E \left(\eta_i \mid y_i^{t-1} \right) = \iota_p E \left(\mu_i \mid y_i^{t-1} \right)
$$
since $(1 - \alpha_1 - ... - \alpha_p)d_1 = (I - \Pi)\iota_p$.

References

Alonso-Borrego, C. and M. Arellano (1999): "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data", *Journal of Business & Economic Statistics*, 17, 36-49.

Alvarez, J. and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators", *Econometrica*, 71, 1121-1159.

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, John Wiley & Sons, New York.

Anderson, T.W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, 18, 47-82.

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

Arellano, M. and S.R. 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.

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

Barro, R. J. and X. Sala-i-Martin (1995), *Economic Growth*, McGraw Hill, New York.

Blundell, R. and R. Smith (1991): "Initial Conditions and Efficient Estimation in Dynamic Panel Data Models", *Annales d'Economie et de Statistique*, 20/21, 109-123.

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

Bond, S. R., A. Hoeffler, and J. Temple (2001): "GMM Estimation of Empirical Growth Models", CEPR Discussion Paper 3048, London.

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.

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

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

Hahn, J. (1997): "Efficient Estimation of Panel Data Models with Sequential Moment Restrictions", *Journal of Econometrics*, 79, 1-21.

Holtz-Eakin, D., W. Newey and H. Rosen (1988): "Estimating Vector Autoregressions with Panel Data, *Econometrica*, 56, 1371-1395.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-708.

Newey, W. K. and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *Econometrica*, 72, 219-255.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1417-1426.

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

Table 1
Employment and Wage VAR Model
Panel Data of Spanish Firms

	WG	GMM	PML	SIV
$N = 738, T = 8$				
Employment equation				
$n_{i(t-1)}$	0.71	0.86	1.00	0.93
	(0.03)	(0.06)		(0.07)
$w_{i(t-1)}$	0.08	0.12	0.08	0.14
	(0.03)	(0.07)		(0.08)
Wage equation				
$n_{i(t-1)}$	0.06	−0.03	0.01	−0.02
	(0.02)	(0.08)		(0.08)
$w_{i(t-1)}$	0.44	0.29	0.68	0.32
	(0.03)	(0.10)		(0.10)
$N = 385, T = 14$				
Employment equation				
$n_{i(t-1)}$	0.86	0.83	0.995	0.90
	(0.05)	(0.05)		(0.05)
$w_{i(t-1)}$	0.26	0.29	0.28	0.30
	(0.09)	(0.13)		(0.15)
Wage equation				
$n_{i(t-1)}$	0.01	−0.09	0.03	−0.15
	(0.02)	(0.04)		(0.05)
$w_{i(t-1)}$	0.45	0.34	0.62	0.32
	(0.07)	(0.10)		(0.12)

All data in deviations from period-specific
cross-sectional means.

Estimates of Υ_1 using SIV estimates of A:
$(N = 738, T = 8)$ sample:

$$\widehat{\Upsilon}_{1(738)} = \begin{pmatrix} 0.863 & 0.155 \\ 0.008 & 0.849 \end{pmatrix}$$

$(N = 385, T = 14)$ sample:

$$\widehat{\Upsilon}_{1(385)} = \begin{pmatrix} 0.983 & 0.441 \\ -0.004 & 0.905 \end{pmatrix}.$$

Table 2

Monte Carlo Simulations for the VAR Model

	WG		GMM		PML		SIV	
	median	mae	median	mae	median	mae	median	mae
$N = 738, T = 8$								
a_{11}	0.48	0.32	0.72	0.08	0.80	0.02	0.80	0.05
a_{12}	0.09	0.06	0.13	0.03	0.15	0.02	0.15	0.03
a_{21}	0.03	0.03	0.01	0.04	0.00	0.01	-0.01	0.05
a_{22}	0.12	0.18	0.29	0.03	0.30	0.01	0.30	0.03
$N = 385, T = 14$								
a_{11}	0.63	0.17	0.74	0.06	0.80	0.01	0.80	0.03
a_{12}	0.12	0.03	0.13	0.02	0.15	0.01	0.15	0.02
a_{21}	0.02	0.02	0.01	0.02	-0.00	0.01	-0.00	0.02
a_{22}	0.21	0.09	0.29	0.02	0.30	0.01	0.30	0.02
$N = 738, T = 8$ Data with trend in variance								
a_{11}	0.50	0.30	0.70	0.11	0.86	0.06	0.81	0.07
a_{12}	0.11	0.04	0.12	0.04	0.17	0.02	0.16	0.04
a_{21}	0.03	0.03	0.01	0.05	-0.01	0.01	-0.01	0.06
a_{22}	0.11	0.19	0.29	0.03	0.31	0.02	0.30	0.03

$a_{11} = 0.8, a_{12} = 0.15, a_{21} = 0, a_{22} = 0.3$

1000 replications. *mae* is median absolute error.

Table 3

Monte Carlo Simulations for the VAR Model

	WG		GMM		PML		SIV	
	median	mae	median	mae	median	mae	median	mae
				$N = 200, T = 8$				
a_{11}	0.48	0.32	0.59	0.21	0.80	0.05	0.80	0.09
a_{12}	0.09	0.06	0.08	0.07	0.15	0.03	0.16	0.06
a_{21}	0.03	0.03	0.03	0.07	-0.00	0.02	0.00	0.09
a_{22}	0.12	0.18	0.27	0.05	0.30	0.02	0.30	0.05
				$N = 100, T = 6$				
a_{11}	0.35	0.45	0.40	0.40	0.80	0.10	0.80	0.22
a_{12}	0.08	0.07	0.04	0.13	0.15	0.06	0.15	0.12
a_{21}	0.04	0.05	0.04	0.14	-0.00	0.05	-0.01	0.19
a_{22}	0.04	0.26	0.22	0.10	0.30	0.05	0.29	0.11
				$N = 50, T = 15$				
a_{11}	0.64	0.16	0.62	0.18	0.80	0.04	0.80	0.06
a_{12}	0.12	0.04	0.10	0.06	0.15	0.04	0.15	0.05
a_{21}	0.02	0.03	0.02	0.04	0.00	0.02	0.00	0.06
a_{22}	0.22	0.09	0.24	0.06	0.30	0.03	0.30	0.04

$a_{11} = 0.8, a_{12} = 0.15, a_{21} = 0, a_{22} = 0.3$

1000 replications. *mae* is median absolute error.

Table 4
Augmented Solow Model
$N = 92, T = 5$

	OLS	WG	GMM	SIV
$(1+\beta)$	0.947	0.680	0.698	0.808
(s.e.)	(0.017)	(0.057)	(0.107)	(0.248)
$\ln(enr_t)$	0.035	-0.049	-0.140	-0.114
(s.e.)	(0.014)	(0.029)	(0.066)	(0.229)
$\ln(s_t)$	0.081	0.138	0.144	0.090
(s.e.)	(0.017)	(0.039)	(0.055)	(0.075)
$\ln(n_t + g + d)$	-0.094	-0.033	0.230	0.227
(s.e.)	(0.053)	(0.152)	(0.339)	(1.384)
Implied λ	0.011	0.077	0.072	0.043
(s.e.)	(0.004)	(0.017)	(0.031)	(0.062)

All data in deviations from period-specific cross-sectional means.
Data for 5-year intervals 1960-1985.
s.e. robust to heterosk. & autocorrelation, $m = T - 2$.
n = population growth rate; g = rate of technical change;
d = rate of depreciation of physical capital ($g + d = 0.05$);
s = saving rate; enr = secondary-school enrollment rate.