

Biørn, Erik

**Working Paper**

## Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences

Memorandum, No. 19/1998

**Provided in Cooperation with:**

Department of Economics, University of Oslo

*Suggested Citation:* Biørn, Erik (1998) : Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences, Memorandum, No. 19/1998, University of Oslo, Department of Economics, Oslo

This Version is available at:

<https://hdl.handle.net/10419/90806>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**PANEL DATA WITH MEASUREMENT ERRORS:  
INSTRUMENTAL VARIABLES AND GMM PROCEDURES  
COMBINING LEVELS AND DIFFERENCES \*)**

by

**ERIK BIØRN**

ABSTRACT

The estimation of linear, static regression equations from panel data with measurement errors in the regressors is considered. If the latent regressor is autocorrelated or non-stationary, several consistent instrumental variables (IV) and generalized method of moments (GMM) estimators usually exist, provided some structure is imposed on the disturbances and measurement errors. Applications of GMM procedures (i) on equations in differences with IV's in levels and (ii) on equations in levels with IV's in differences – the difference transformations eliminating unobserved individual heterogeneity – are considered. The IV's may be constructed from values of the regressors or the regressand. Some of the orthogonality conditions delimiting the valid IV's are redundant. Illustrations based on data on inputs and outputs from an eight year panel of manufacturing firms are given.

**Key words:** Panel Data. Errors-in-Variables. Repeated Measurement.  
Moment Conditions. GMM Estimation. Returns to scale

**JEL classification:** C13, C23, C33.

---

\* Invited paper presented at the Eighth International Conference on Panel Data, Göteborg, June 1998. I am grateful to Jørgen Aasness, Tor Jakob Klette, Terje Skjerpen, and Tom Wansbeek for valuable comments.

# 1 Introduction

A familiar and notable property of the Ordinary Least Squares (OLS) when there are random measurement errors (errors-in-variables, EIV) in the regressors is that the slope coefficient estimator is inconsistent. In the one regressor case (or the multiple regressor case with uncorrelated regressors), under standard assumptions, the estimator is biased towards zero, often denoted as *attenuation*. More seriously, unless some ‘extraneous’ information is available, *e.g.* the existence of valid parameter restrictions or valid instruments for the error-ridden regressors, slope coefficients cannot (in general) be identified from standard data [see Fuller (1987, section 1.1.3)].<sup>1</sup> This conclusion of *lack of identification* in EIV models, however, relates to *uni-dimensional* data, *i.e.*, pure (single or repeated) cross-sections or pure time-series. If the variables are observed as panel data, exhibiting *two-dimensional* variation, it may be possible to handle the EIV identification problem and estimate slope coefficients consistently without extraneous information, provided that the distribution of the latent regressors and the measurement errors satisfy certain weak conditions.

Briefly and intuitively, the reason why the existence of variables observed along two dimensions makes the EIV identification problem more manageable, is partly (i) the *repeated measurement* property of panel data – each individual and each period is ‘replicated’ – so that the measurement error problem can be reduced by taking averages, which, in turn, may show sufficient variation to permit consistent estimation<sup>2</sup>, and partly (ii) the larger set of *other linear data transformations* available for estimation. Such transformations, involving several individuals or several periods, may be needed to take account of uni-dimensional ‘nuisance variables’ like unobserved individual or period specific *heterogeneity*, which are potentially correlated with the regressor.

The focus of this paper is on the estimation of linear, static regression equations from balanced panel data with additive, random measurement errors in the regressors by means of methods utilizing instrumental variables (IV’s). The panel data available to an econometrician are frequently from individuals, firms, or other kinds

---

<sup>1</sup>Note, however, (i) that identification under non-normality of the true regressor is possible, by utilizing moments of the distribution of the observable variables of order higher than the second, [see Reiersøl (1950)] and (ii) that even under non-identification, bounds on the parameters can be established from the distribution of the observable variables [see Fuller (1987, p. 11)]. These bounds may be wide or narrow, depending on the covariance structure of the variables; see Klepper and Leamer (1984), Bekker *et al.* (1987), Erickson (1993), and Willassen (1998).

<sup>2</sup>The repeated measurement argument in relation to latent variables models is elaborated in Aigner *et al.* (1984, section 3.10). See also Willassen (1979, section 3)

of micro units, where not only observation errors in the narrow sense, but also departures between theoretical variable definitions and their observable counterparts in a wider sense may be present. The models we will consider superimpose the EIV problem on the heterogeneity problem.

From the panel data literature disregarding the EIV problem we know that the effect of, for example, additive (fixed or random) individual heterogeneity within a linear model can be eliminated by deducting individual means, taking differences over periods, etc. [see Hsiao (1986, Section 1.1) and Baltagi (1995, Chapter 2)]. Such transformations, however, may magnify the variation in the measurement error component of the observations relative to the variation in the true structural component, *i.e.*, they may increase the ‘noise/signal ratio’. Data transformations intended to ‘solve’ the latent heterogeneity problem may then aggravate the EIV problem. Several familiar estimators for panel data models, including the fixed effects within-group and between-group estimators, and the random effects Generalized Least Squares (GLS) estimators will then be inconsistent, although to a degree depending, *inter alia*, on the way in which the number of individuals and and/or periods tend to infinity and on the heterogeneity of the measurement error process; see Griliches and Hausman (1986) and Biørn (1992, 1996) for a substantial number of examples for one regressor models. Such inconsistency problems will not be dealt with in the present paper.

If the distribution of the latent regressor vector is *not* time invariant and the second order moments of the measurement errors and disturbances are structured to some extent, a large number of consistent IV estimators of the coefficient of the latent regressor vector exist. The consistency of these estimators is robust to potential correlation between the individual heterogeneity and the latent regressor. A notable point is that serial correlation or non-stationarity of the latent regressor is favourable from the point of view of identification and estimability of the coefficient vector,<sup>3</sup> but restrictions has to be imposed on the distribution of the measurement errors and disturbances, although they do not need to be independently, identically distributed (IID). Briefly, there should not be ‘too much structure’ on the second order moments of the latent exogenous regressors across the panel, and not ‘too little structure’ on the second order moments of the errors and disturbances.

The procedures we will consider for static panel data EIV models in this paper resemble, to some extent, procedures for autoregressive (AR) models for panel data without measurement errors (mostly AR(1) equations with individual heterogeneity

---

<sup>3</sup>This point, in general terms, is made by Griliches (1986, pp. 1482 – 1483).

and often with exogenous regressors added) discussed, *inter alia*, by Anderson and Hsiao (1981, 1982), Holtz-Eakin *et al.* (1988), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), Blundell and Bond (1995), and Sevestre and Trognon (1996). Notable elements in this literature are (a) linear transformations of the AR equation, especially differencing, intended to remove individual heterogeneity, (b) lagged level values of the predetermined variables used as instruments for the lagged difference to ‘control for’ its correlation with the differenced disturbance, and (c) substantial attention devoted to the orthogonality conditions (OC’s) derived from the model’s structure (including the initial condition of the AR process) in the context of IV or Generalized Method of Moments (GMM) procedures.

The estimation procedures we will consider in the paper have two basic characteristics: First, *a mixture of level and difference variables* are involved. Second, the OC’s derived from the EIV structure – involving levels and differences over one or more than one periods – are not all essential, some are redundant. Our estimation procedures are of two kinds: (A) The equation is transformed to differences to remove individual heterogeneity (the within individual estimator being inconsistent in the presence of EIV), and is estimated by IV or GMM. As IV’s we use level values of the regressors and/or regressands for other periods. (B) The equation is kept in level form, and is estimated by IV or GMM. As IV’s we use differenced values of the regressors and/or regressands for other periods. The conditions needed for consistency of the (B) procedures under individual heterogeneity are stronger than for the (A) procedures. Our (A) procedures extend and modify those proposed in Griliches and Hausman (1986) (which may be considered a seminal contribution to this literature), Wansbeek and Koning (1991), Arellano and Bover (1995), Biørn (1992, 1996), and Biørn and Klette (1997a,b). Since levels are used as IV’s for differences or differences are used as IV’s for levels, the correlation may be low, and the ‘weak IV problem’ to which increasing attention has been paid in recent literature on IV and GMM estimation [see, *e.g.*, Nelson and Startz (1990), Davidson and MacKinnon (1993, pp. 217 – 224), and Staiger and Stock (1996, 1997)], also is a potential problem in our setting.

The rest of the paper is organized as follows. In Section 2, we present our model, with the joint occurrence of EIV and *individual heterogeneity*, not only the standard case with EIV as white noise, but also *autocorrelation*, represented by a finite moving average (MA) process or an *error components structure*. The implied OC’s and their potential redundancy are discussed. In Section 3, we consider estimators constructed from period means and their consistency when the number of individuals

goes to infinity. In Section 4, several IV and GMM procedures are considered, with regard to efficiency and robustness to heteroskedasticity and autocorrelation of the measurement errors and the disturbances. In Section 5, we present empirical illustrations, based on panel data on inputs and outputs from Norwegian manufacturing firms. Section 6 concludes.

## 2 Model and orthogonality conditions

### 2.a Model, notation, and basic assumptions

Consider a set of panel data with  $N$  ( $\geq 2$ ) individuals observed in  $T$  ( $\geq 2$ ) periods and a relationship between  $y$  (observable scalar) and a  $(1 \times K)$  vector  $\boldsymbol{\xi}$  (latent),

$$(2.1) \quad y_{it} = c + \boldsymbol{\xi}_{it}\boldsymbol{\beta} + \alpha_i + u_{it}, \quad i = 1, \dots, N; t = 1, \dots, T,$$

where  $(y_{it}, \boldsymbol{\xi}_{it})$  is the value of  $(y, \boldsymbol{\xi})$  for individual  $i$  in period  $t$ ,  $c$  is a scalar constant,  $\boldsymbol{\beta}$  is a  $(K \times 1)$  vector and  $\alpha_i$  is a zero (marginal) mean individual effect, which we consider as random and potentially correlated with  $\boldsymbol{\xi}_{it}$ , and  $u_{it}$  is a zero mean disturbance, which may also contain a measurement error in  $y_{it}$ . We observe

$$(2.2) \quad \mathbf{x}_{it} = \boldsymbol{\xi}_{it} + \mathbf{v}_{it}, \quad i = 1, \dots, N; t = 1, \dots, T,$$

where  $\mathbf{v}_{it}$  is a zero mean vector of measurement errors. Hence,

$$(2.3) \quad y_{it} = c + \mathbf{x}_{it}\boldsymbol{\beta} + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + u_{it} - \mathbf{v}_{it}\boldsymbol{\beta}, \quad i = 1, \dots, N; t = 1, \dots, T,$$

or in vector form, arranged by individuals and by periods, as

$$(2.4) \quad \mathbf{y}_{i\cdot} = \mathbf{e}_T c + \mathbf{X}_{i\cdot}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{i\cdot}, \quad \boldsymbol{\epsilon}_{i\cdot} = \mathbf{e}_T \alpha_i + \mathbf{u}_{i\cdot} - \mathbf{V}_{i\cdot}\boldsymbol{\beta}, \quad i = 1, \dots, N,$$

$$(2.5) \quad \mathbf{y}_{\cdot t} = \mathbf{e}_N c + \mathbf{X}_{\cdot t}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{\cdot t}, \quad \boldsymbol{\epsilon}_{\cdot t} = \boldsymbol{\alpha} + \mathbf{u}_{\cdot t} - \mathbf{V}_{\cdot t}\boldsymbol{\beta}, \quad t = 1, \dots, T,$$

where  $\mathbf{y}_{i\cdot} = (y_{i1}, \dots, y_{iT})'$ ,  $\mathbf{X}_{i\cdot} = (\mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT})'$ ,  $\mathbf{y}_{\cdot t} = (y_{1t}, \dots, y_{Nt})'$ ,  $\mathbf{X}_{\cdot t} = (\mathbf{x}'_{1t}, \dots, \mathbf{x}'_{Nt})'$ , etc.,  $\mathbf{e}_m$  denoting the  $(m \times 1)$  vector of ones and  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N)'$ . In the following, we refer to the  $\epsilon_{it}$ 's as composite errors/disturbances.

Let

$$\mathbf{d}_{t\theta} = \begin{bmatrix} (1 \times T) \text{ vector with} \\ \text{element } t = 1, \text{ element } \theta = -1, \\ \text{and zero otherwise} \end{bmatrix}, \quad \begin{matrix} t = 2, \dots, T, \\ \theta = 1, \dots, t-1, \end{matrix}$$

so that we can write arbitrary differences as  $\Delta y_{it\theta} = y_{it} - y_{i\theta} = \mathbf{d}_{t\theta}\mathbf{y}_{i\cdot}$ ,  $\Delta \mathbf{x}_{it\theta} = \mathbf{x}_{it} - \mathbf{x}_{i\theta} = \mathbf{d}_{t\theta}\mathbf{X}_{i\cdot}$ , etc., and let  $\Delta \mathbf{y}_{t\theta} = [\Delta y_{1t\theta} \cdots \Delta y_{Nt\theta}]' = \mathbf{y}_{\cdot t} - \mathbf{y}_{\cdot \theta}$ ,  $\Delta \mathbf{X}_{t\theta} =$

$[\Delta \mathbf{x}'_{1t\theta} \cdots \Delta \mathbf{x}'_{Nt\theta}]' = \mathbf{X}_{\cdot t} - \mathbf{X}_{\cdot \theta}$ , etc. Premultiplying (2.4) by  $\mathbf{d}_{t\theta}$ , we get<sup>4</sup>

$$(2.6) \quad \Delta y_{it\theta} = \Delta \mathbf{x}_{it\theta} \boldsymbol{\beta} + \Delta \epsilon_{it\theta},$$

$$(2.7) \quad \Delta \mathbf{y}_{t\theta} = \Delta \mathbf{X}_{t\theta} \boldsymbol{\beta} + \Delta \boldsymbol{\epsilon}_{t\theta}, \quad i = 1, \dots, N; \quad t, \theta = 1, \dots, T,$$

We will use both (2.3) and (2.6) for estimation of the coefficient vector  $\boldsymbol{\beta}$ , exploiting appropriate moment conditions for the  $\epsilon_{it}$ 's,  $y_{it}$ 's, and  $\mathbf{x}_{it}$ 's to be described below.

We assume that  $(\boldsymbol{\xi}_{it}, u_{it}, \mathbf{v}_{it}, \alpha_i)$  are independent across individuals (which excludes random period specific components) and make the following *basic orthogonality assumptions*:

$$\text{ASSUMPTION (A):} \quad \begin{cases} \mathbf{E}(\mathbf{v}'_{it} u_{i\theta}) = \mathbf{0}_{K1}, & \mathbf{E}(\boldsymbol{\xi}'_{it} u_{i\theta}) = \mathbf{0}_{K1}, & i = 1, \dots, N, \\ \mathbf{E}(\boldsymbol{\xi}'_{i\theta} \otimes \mathbf{v}_{it}) = \mathbf{0}_{KK}, & & t, \theta = 1, \dots, T, \\ \mathbf{E}(\alpha_i \mathbf{v}_{it}) = \mathbf{0}_{1K}, & \mathbf{E}(\alpha_i u_{it}) = 0, & \end{cases}$$

where  $\mathbf{0}_{mn}$  denotes the  $(m \times n)$  zero matrix and  $\otimes$  is the Kronecker product operator.<sup>5</sup> We also need some *additional assumptions* about the measurement errors and disturbances. The first and strongest are:

$$\text{ASSUMPTION (B1):} \quad \mathbf{E}(\mathbf{v}'_{it} \mathbf{v}_{i\theta}) = \mathbf{0}_{KK}, \quad t \neq \theta,$$

$$\text{ASSUMPTION (C1):} \quad \mathbf{E}(u_{it} u_{i\theta}) = 0, \quad t \neq \theta,$$

which impose non-autocorrelation on both of them. Weaker assumptions are:

$$\text{ASSUMPTION (B2):} \quad \mathbf{E}(\mathbf{v}'_{it} \mathbf{v}_{i\theta}) = \mathbf{0}_{KK}, \quad |t - \theta| > \tau,$$

$$\text{ASSUMPTION (C2):} \quad \mathbf{E}(u_{it} u_{i\theta}) = 0, \quad |t - \theta| > \tau,$$

$$\text{ASSUMPTION (B3):} \quad \mathbf{E}(\mathbf{v}'_{it} \mathbf{v}_{i\theta}) \text{ is invariant to } t, \theta, \quad t \neq \theta,$$

$$\text{ASSUMPTION (C3):} \quad \mathbf{E}(u_{it} u_{i\theta}) \text{ is invariant to } t, \theta, \quad t \neq \theta,$$

of which (B2) and (C2) allow for a (vector) moving average (MA) structure up to order  $\tau$  ( $\geq 1$ ), and (B3) and (C3) allow for time invariance of the autocorrelation. The latter will, for instance, be satisfied if the measurement errors and the disturbances have individual components, say  $\mathbf{v}_{it} = \mathbf{v}_{1i} + \mathbf{v}_{2it}$ ,  $u_{it} = u_{1i} + u_{2it}$ , where  $\mathbf{v}_{1i}$ ,  $\mathbf{v}_{2it}$ ,  $u_{1i}$ , and  $u_{2it}$  are independent IID processes. Homoskedasticity of  $\mathbf{v}_{it}$  and/or  $u_{it}$  across  $i$  and  $t$  is not assumed, but may be imposed.

---

<sup>4</sup>Premultiplication by  $\mathbf{d}_{t\theta}$  is, of course, only one way of eliminating  $\alpha_i$  from (2.4). Any  $(1 \times T)$  vector  $\mathbf{c}_{t\theta}$  such that  $\mathbf{c}_{t\theta} \mathbf{e}_T = 0$ , has this property, for example the rows of the within individual transformation matrix  $\mathbf{I}_T - \mathbf{e}_T \mathbf{e}'_T / T$ , where  $\mathbf{I}_T$  is the  $T$  dimensional identity matrix.

<sup>5</sup>The last two assumptions are stronger than strictly needed; time invariance of  $\mathbf{E}(\alpha_i \mathbf{v}_{it})$  and  $\mathbf{E}(\alpha_i u_{it})$  is sufficient. For practical applications, this modification seems to be of little importance.

Our final assumptions relate to the *distribution of the latent regressor vector*  $\boldsymbol{\xi}_{it}$ . We here consider:

ASSUMPTION (D1):	$\mathbf{E}(\boldsymbol{\xi}_{it})$	is invariant to $t$ ,
ASSUMPTION (D2):	$\mathbf{E}(\alpha_i \boldsymbol{\xi}_{it})$	is invariant to $t$ ,
ASSUMPTION (E1):	$\mathbf{E}(\boldsymbol{\xi}_{it}' \boldsymbol{\xi}_{i\theta})$	has rank $K$ for some $t \neq \theta$ and is not invariant to $t, \theta$ ,
ASSUMPTION (E2):	$\text{rank}(\mathbf{E}[\boldsymbol{\xi}_{ip}' (\Delta \boldsymbol{\xi}_{it\theta})]) = K$	for some $p \neq t \neq \theta$ ,
ASSUMPTION (E3):	$\text{rank}(\mathbf{E}[(\Delta \boldsymbol{\xi}_{ipq})' \boldsymbol{\xi}_{it}]) = K$	for some $p \neq q \neq t$ .

Assumptions (D1) and (D2) hold when  $\boldsymbol{\xi}_{it}$  is stationary for all  $i$  [(D1) alone imposing mean stationarity], but  $\boldsymbol{\xi}_{it}$  and  $\alpha_i$  do not need to be uncorrelated. Assumption (E1) and the stronger versions (E2) and (E3) *impose non-IID and some form of autocorrelation or non-stationarity on*  $\boldsymbol{\xi}_{it}$ . Assumption (E1) – (E3) exclude, for instance, the case where  $\boldsymbol{\xi}_{it}$  has an individual component, so that  $\boldsymbol{\xi}_{it} = \boldsymbol{\xi}_{1i} + \boldsymbol{\xi}_{2it}$ , where  $\boldsymbol{\xi}_{1i}$  and  $\boldsymbol{\xi}_{2it}$  are independent (vector) IID processes.<sup>6</sup>

## 2.b Identification and the structure of the second order moments

The nature of the conditions that the distribution of  $(\boldsymbol{\xi}_{it}, u_{it}, \mathbf{v}_{it}, \alpha_i)$  must satisfy to make identification of  $\boldsymbol{\beta}$  possible can be illustrated as follows. Assume, for simplicity, that this distribution is the same for all individuals and that (A) holds, and let

$$\begin{aligned} \mathbf{C}(\boldsymbol{\xi}_{it}, \boldsymbol{\xi}_{i\theta}) &= \boldsymbol{\Sigma}_{t\theta}^{\xi\xi}, & \mathbf{E}(\boldsymbol{\xi}_{it} \alpha_i) &= \boldsymbol{\Sigma}_t^{\xi\alpha}, & \mathbf{E}(\alpha_i^2) &= \sigma^{\alpha\alpha}, & i &= 1, \dots, N, \\ \mathbf{C}(\mathbf{v}_{it}' \mathbf{v}_{i\theta}) &= \boldsymbol{\Sigma}_{t\theta}^{vv}, & \mathbf{E}(u_{it} u_{i\theta}) &= \sigma_{t\theta}^{uu}, & & & t, \theta &= 1, \dots, T, \end{aligned}$$

where  $\mathbf{C}$  denotes the covariance matrix operator. It then follows from (2.1) and (2.2) that the second order moments of the observable variables can be expressed as

$$(2.8) \quad \begin{cases} \mathbf{C}(\mathbf{x}_{it}, \mathbf{x}_{i\theta}) = \boldsymbol{\Sigma}_{t\theta}^{\xi\xi} + \boldsymbol{\Sigma}_{t\theta}^{vv}, \\ \mathbf{C}(\mathbf{x}_{it}, y_{i\theta}) = \boldsymbol{\Sigma}_{t\theta}^{\xi\xi} \boldsymbol{\beta} + \boldsymbol{\Sigma}_t^{\xi\alpha}, \\ \mathbf{C}(y_{it}, y_{i\theta}) = \boldsymbol{\beta}' \boldsymbol{\Sigma}_{t\theta}^{\xi\xi} \boldsymbol{\beta} + \boldsymbol{\Sigma}_t^{\xi\alpha} \boldsymbol{\beta} + \boldsymbol{\beta}' (\boldsymbol{\Sigma}_{t\theta}^{\xi\alpha})' + \sigma^{\alpha\alpha} + \sigma_{t\theta}^{uu}, \end{cases} \quad i = 1, \dots, N; \quad t, \theta = 1, \dots, T.$$

---

<sup>6</sup>It might be possible, depending on the particular problem, to impose more structure on the first and second order moments of the  $u_{it}$ 's,  $\mathbf{v}_{it}$ 's,  $\boldsymbol{\xi}_{it}$ 's and  $\alpha_i$ 's – confer the ‘structural approach’ to EIV modelling. In this way we might obtain more efficient (but potentially less robust) estimators by operating on the full covariance matrix of the  $y_{it}$ 's and the  $\mathbf{x}_{it}$ 's rather than eliminating the  $\alpha_i$ 's by differencing, as elaborated in the following sections. This alternative approach, which is more in the spirit of the LISREL or LIML approach to simultaneous equation systems modelling and estimation, will not be considered.



The identifiability of  $\beta$  from second order moments in general depends on whether or not knowledge of  $\mathbb{C}(\mathbf{x}_{it}, \mathbf{x}_{i\theta})$ ,  $\mathbb{C}(\mathbf{x}_{it}, y_{i\theta})$ , and  $\mathbb{C}(y_{it}, y_{i\theta})$  is sufficient for obtaining a unique solution for  $\beta$  from (2.8), given the restrictions imposed on the  $\Sigma_{i\theta}^{\xi\xi}$ 's,  $\Sigma_t^{\xi\alpha}$ 's,  $\sigma_{i\theta}^{uu}$ 's, and  $\sigma^{\alpha\alpha}$ . The answer in general depends on  $T$  and  $K$ . With no further information, the number of elements in  $\mathbb{C}(\mathbf{x}_{it}, \mathbf{x}_{i\theta})$  and  $\mathbb{C}(y_{it}, y_{i\theta})$  (all of which can be consistently estimated from corresponding sample moments under weak conditions) equal the number of unknown elements in  $\Sigma_{i\theta}^{vv}$  and  $\sigma_{i\theta}^{uu}$ , which is  $KT(KT + 1)/2$  and  $T(T + 1)/2$ , respectively. Then  $\sigma^{\alpha\alpha}$  cannot be identified and  $\mathbb{C}(\mathbf{x}_{it}, y_{i\theta})$  contains the only ‘additional information’ available for identification of  $\beta$ ,  $\Sigma_{i\theta}^{\xi\xi}$ , and  $\Sigma_t^{\xi\alpha}$ .

Consider two (extreme) examples. If  $T = 1$ , *i.e.*, cross-section data, and no additional restrictions are imposed, we have an identification problem for any  $K$ . On the other hand, if  $T$  is arbitrarily large and all  $\xi_{it}$ ,  $u_{it}$ ,  $\mathbf{v}_{it}$ ,  $\alpha_i$  are IID,

$$\xi_{it} \sim \text{IID}(\mu^\xi, \Sigma^{\xi\xi}), \quad \mathbf{v}_{it} \sim \text{IID}(\mathbf{0}_{T,1}, \Sigma^{vv}), \quad u_{it} \sim \text{IID}(0, \sigma^{uu}), \quad \alpha_i \sim \text{IID}(0, \sigma^{\alpha\alpha}),$$

we also have lack of identification in general.<sup>7</sup> From (2.8) we then get

$$(2.9) \quad \begin{cases} \mathbb{C}(\mathbf{x}_{it}, \mathbf{x}_{it}) &= \Sigma^{\xi\xi} + \Sigma^{vv}, & i = 1, \dots, N, \\ \mathbb{C}(\mathbf{x}_{it}, y_{it}) &= \Sigma^{\xi\xi}\beta, & t = 1, \dots, T, \\ \mathbb{C}(y_{it}, y_{it}) &= \beta'\Sigma^{\xi\xi}\beta + \sigma^{\alpha\alpha} + \sigma^{uu}, \end{cases}$$

and

$$(2.10) \quad \begin{cases} \mathbb{C}(\mathbf{x}_{it}, \mathbf{x}_{i\theta}) &= \mathbf{0}_{K,K}, & i = 1, \dots, N, \\ \mathbb{C}(\mathbf{x}_{it}, y_{i\theta}) &= \mathbf{0}_{K,1}, & t, \theta = 1, \dots, T; \theta \neq t, \\ \mathbb{C}(y_{it}, y_{i\theta}) &= \sigma^{\alpha\alpha}, \end{cases}$$

and are basically in the same situation with regard to identifiability of  $\beta$  when  $T > 1$  as when  $T = 1$ . The ‘cross-period’ equations (2.10) serve no other purpose than identification of  $\sigma^{\alpha\alpha}$ , and whether  $T = 1$  or  $T > 1$  realizations of  $\mathbb{C}(\mathbf{x}_{it}, \mathbf{x}_{it})$ ,  $\mathbb{C}(\mathbf{x}_{it}, y_{it})$ , and  $\mathbb{C}(y_{it}, y_{it})$  are available in (2.9) is immaterial to the identifiability of  $\beta$ ,  $\Sigma^{\xi\xi}$ ,  $\Sigma^{vv}$ , and  $\sigma^{uu}$ . In intermediate situations, identification may be ensured when  $T > 1$ .

Briefly, these examples illustrate that in order to ensure identification of the slope coefficient vector when panel data are available, there should not be ‘too much structure’ on the second order moments of the latent exogenous regressors across the panel, and not ‘too little structure’ on the second order moments of the errors and disturbances.

---

<sup>7</sup>We get an essentially similar conclusion when  $\xi_{it}$  is time invariant and IID across  $i$ .

## 2.c Moment equations and orthogonality conditions

A substantial number of (linear and non-linear) moment conditions involving  $y_{it}$ ,  $\mathbf{x}_{it}$ , and  $\epsilon_{it}$  can be derived from the (A) – (E) assumptions made above.

Since (2.1) – (2.3) and (A) imply

$$\begin{aligned} E(\mathbf{x}'_{it}\mathbf{x}_{i\theta}) &= E(\boldsymbol{\xi}'_{it}\boldsymbol{\xi}_{i\theta}) + E(\mathbf{v}'_{it}\mathbf{v}_{i\theta}), \\ E(\mathbf{x}'_{it}y_{i\theta}) &= E(\boldsymbol{\xi}'_{it}\boldsymbol{\xi}_{i\theta})\boldsymbol{\beta} + E[\boldsymbol{\xi}'_{it}(\alpha_i + c)], \\ E(y_{it}y_{i\theta}) &= c^2 + E(\alpha_i^2) + \boldsymbol{\beta}'E(\boldsymbol{\xi}'_{it}\boldsymbol{\xi}_{i\theta})\boldsymbol{\beta} \\ &\quad + \boldsymbol{\beta}'E[\boldsymbol{\xi}'_{it}(\alpha_i + c)] + E[(\alpha_i + c)\boldsymbol{\xi}_{i\theta}]\boldsymbol{\beta} + E(u_{it}u_{i\theta}), \\ E(\mathbf{x}'_{it}\epsilon_{i\theta}) &= E(\boldsymbol{\xi}'_{it}\alpha_i) - E(\mathbf{v}'_{it}\mathbf{v}_{i\theta})\boldsymbol{\beta}, \\ E(y_{it}\epsilon_{i\theta}) &= \boldsymbol{\beta}'E(\boldsymbol{\xi}'_{it}\alpha_i) + E(\alpha_i^2) + E(u_{it}u_{i\theta}), \end{aligned}$$

we can derive the following moments equations involving observable variables in levels and differences:<sup>8</sup>

$$(2.11) \quad E[\mathbf{x}'_{ip}(\Delta\mathbf{x}_{it\theta})] = E[\boldsymbol{\xi}'_{ip}(\Delta\boldsymbol{\xi}_{it\theta})] + E[\mathbf{v}'_{ip}(\Delta\mathbf{v}_{it\theta})],$$

$$(2.12) \quad E[\mathbf{x}'_{ip}(\Delta y_{it\theta})] = E[\boldsymbol{\xi}'_{ip}(\Delta\boldsymbol{\xi}_{it\theta})]\boldsymbol{\beta},$$

$$(2.13) \quad E[(\Delta\mathbf{x}_{ipq})'y_{it}] = E[(\Delta\boldsymbol{\xi}_{ipq})'\boldsymbol{\xi}_{it}]\boldsymbol{\beta} + E[(\Delta\boldsymbol{\xi}_{ipq})'(\alpha_i + c)],$$

and moment equations involving observable variables and errors/disturbances:

$$(2.14) \quad E[\mathbf{x}'_{ip}(\Delta\epsilon_{it\theta})] = -E[\mathbf{v}'_{ip}(\Delta\mathbf{v}_{it\theta})]\boldsymbol{\beta},$$

$$(2.15) \quad E[y_{ip}(\Delta\epsilon_{it\theta})] = E[u_{ip}(\Delta u_{it\theta})],$$

$$(2.16) \quad E[(\Delta\mathbf{x}_{ipq})'\epsilon_{it}] = E[(\Delta\boldsymbol{\xi}_{ipq})'\alpha_i] - E[(\Delta\mathbf{v}_{ipq})'\mathbf{v}_{it}]\boldsymbol{\beta},$$

$$(2.17) \quad E[(\Delta y_{ipq})\epsilon_{it}] = \boldsymbol{\beta}'E[(\Delta\boldsymbol{\xi}_{ipq})'\alpha_i] + E[(\Delta u_{ipq})u_{it}],$$

$$t, \theta, p, q = 1, \dots, T.$$

Not all of the equations in (2.11) – (2.17), whose number is substantial even for small or moderate  $T$ , are, of course, independent. Depending on which (B), (C), and (D) assumptions are valid, some of the terms on the right hand side of (2.14) – (2.17), or all, will vanish. Note, with respect to (2.13), that (D1) implies  $E[(\Delta\boldsymbol{\xi}_{ipq})'c] = \mathbf{0}_{K1}$  for any  $c$ , and that (D2) implies  $E[(\Delta\boldsymbol{\xi}_{ipq})'\alpha_i] = \mathbf{0}_{K1}$ , so that if both are satisfied,  $E[(\Delta\mathbf{x}_{ipq})'y_{it}] = E[(\Delta\boldsymbol{\xi}_{ipq})'\boldsymbol{\xi}_{it}]\boldsymbol{\beta}$ .

---

<sup>8</sup>Wansbeek (1996) has focused on the importance of including not only the moment conditions involving the  $x$ 's in panel data analyses of EIV models [as in Griliches and Hausman (1986)], but also those involving the  $y$ 's.

Hence, for  $T > 2$ , (2.3), (2.6), and (2.14) – (2.17) imply the following moment conditions on the observable variables and the composite errors/disturbances<sup>9</sup>

$$(2.18) \quad \begin{cases} \text{When (B1), or (B3), holds and } t \neq \theta \neq p, \\ \text{or when (B2) holds and } |t - p|, |\theta - p| > \tau, \text{ then} \\ \mathbf{E}[\mathbf{x}'_{ip}(\Delta\epsilon_{it\theta})] = \mathbf{E}[\mathbf{x}'_{ip}(\Delta y_{it\theta})] - \mathbf{E}[\mathbf{x}'_{ip}(\Delta\mathbf{x}_{it\theta})]\boldsymbol{\beta} = \mathbf{0}_{K1}. \end{cases}$$

$$(2.19) \quad \begin{cases} \text{When (C1), or (C3), holds and } t \neq \theta \neq p, \\ \text{or when (C2) holds and } |t - p|, |\theta - p| > \tau, \text{ then} \\ \mathbf{E}[y_{ip}(\Delta\epsilon_{it\theta})] = \mathbf{E}[y_{ip}(\Delta y_{it\theta})] - \mathbf{E}[y_{ip}(\Delta\mathbf{x}_{it\theta})]\boldsymbol{\beta} = 0. \end{cases}$$

$$(2.20) \quad \begin{cases} \text{When (B1), or (B3), (D1), and (D2) hold, and } t \neq p \neq q, \\ \text{or when (B2), (D1), and (D2) hold, and } |t - p|, |t - q| > \tau, \text{ then} \\ \mathbf{E}[(\Delta\mathbf{x}_{ipq})'\epsilon_{it}] = \mathbf{E}[(\Delta\mathbf{x}_{ipq})'y_{it}] - \mathbf{E}[(\Delta\mathbf{x}_{ipq})'\mathbf{x}_{it}]\boldsymbol{\beta} = \mathbf{0}_{K1}. \end{cases}$$

$$(2.21) \quad \begin{cases} \text{When (C1), or (C3), (D1), and (D2) hold, and } t \neq p \neq q, \\ \text{or when (C2), (D1), and (D2) hold, and } |t - p|, |t - q| > \tau, \text{ then} \\ \mathbf{E}[(\Delta y_{ipq})\epsilon_{it}] = \mathbf{E}[(\Delta y_{ipq})y_{it}] - \mathbf{E}[(\Delta y_{ipq})\mathbf{x}_{it}]\boldsymbol{\beta} = 0. \end{cases}$$

Furthermore, from (2.11) – (2.13) it follows for  $T > 2$  that

$$(2.22) \quad \text{When (E2) holds, } \text{rank}(\mathbf{E}[\mathbf{x}'_{ip}(\Delta\mathbf{x}_{it\theta})]) = K, \text{rank}(\mathbf{E}[y_{ip}(\Delta\mathbf{x}_{it\theta})]) = 1.$$

$$(2.23) \quad \text{When (D1), (D2), and (E3) hold,}$$

$$\text{rank}(\mathbf{E}[(\Delta\mathbf{x}_{ipq})'\mathbf{x}_{it}]) = K, \text{rank}(\mathbf{E}[(\Delta y_{ipq})\mathbf{x}_{it}]) = 1.$$

## 2.d Essential and redundant orthogonality conditions

Before we proceed to IV or GMM estimation of  $\boldsymbol{\beta}$  using OC's from (2.18) – (2.21) an examination of the relationships between the  $\frac{1}{2}KT(T-1)(T-2)$  OC's in (2.18) and between the  $\frac{1}{2}KT(T-1)(T-2)$  OC's in (2.20) is needed. Throughout, we assume that  $T > 2$ . Some of these conditions are redundant, since they can be derived as linear combinations of other conditions.<sup>10</sup> Confining attention to the OC's constructed from  $\mathbf{x}$ 's, we have<sup>11</sup>

<sup>9</sup>Modifications of (2.20) and (2.21) when the mean stationarity assumption (D1) is relaxed, so that less structure is imposed on the distribution of  $\boldsymbol{\xi}_{it}$ , will be discussed in section 4.d.

<sup>10</sup>Essential and redundant moment conditions in the context of AR models for panel data are discussed in Ahn and Schmidt (1995), Arellano and Bover (1995), and Blundell and Bond (1995). A general treatment of redundancy of moment conditions in GMM estimation, although with no reference to panel data models, can be found in Breusch *et al.* (1997).

<sup>11</sup>The OC's involving  $y$ 's can be treated similarly.

- ( $\alpha$ ) When (B1), or (B3), and (C1), or (C3), are satisfied, all OC's (2.18) can be constructed from *all admissible OC's relating to equations differenced over one period and a subset of OC's relating to differences over two periods.*
- ( $\beta$ ) When (B1), or (B3), and (C1), or (C3), are satisfied, all OC's (2.20) can be constructed from *all admissible OC's relating to IV's differenced over one period and a subset of IV's differenced over two periods.*
- ( $\gamma$ ) When (B2) and (C2) are satisfied, all OC's (2.18) can be constructed from *all admissible OC's relating to equations differenced over one period and a subset of OC's relating to differences over  $2(\tau+1)$  periods.*
- ( $\delta$ ) When (B2) and (C2) are satisfied, all OC's (2.20) can be constructed from *all admissible OC's relating to IV's differenced over one period and a subset of IV's differenced over  $2(\tau+1)$  periods.*

We denote the non-redundant conditions defined by ( $\alpha$ ) – ( $\delta$ ) as *essential* OC's for the specific cases. Since (2.18) and (2.20) are symmetric, we prove only ( $\alpha$ ) [and ( $\gamma$ )] and derive ( $\beta$ ) [and ( $\delta$ )] by way of analogy.

Since  $\mathbf{x}'_{ip} \Delta \epsilon_{it\theta} = \mathbf{x}'_{ip} (\sum_{j=\theta+1}^t \Delta \epsilon_{ij,j-1})$ , we see that if (hypothetically) all  $p = 1, \dots, T$  combined with all  $t > \theta$  would give admissible OC's, (2.18) for differences over  $2, 3, \dots, T-1$  periods could have been constructed from the conditions relating to one-period differences only. However, since  $(t, \theta) = (p, p-1), (p+1, p)$  are inadmissible, and (when (B1), or (B3), holds)  $(t, \theta) = (p+1, p-1)$  is admissible, we have to distinguish between the cases where  $p$  is (I) strictly outside and (II) strictly inside the interval  $(\theta, t)$ .<sup>12</sup> From the identities

$$\begin{aligned} (I) \quad \mathbf{x}'_{ip} \Delta \epsilon_{it\theta} &= \mathbf{x}'_{ip} (\sum_{j=\theta+1}^t \Delta \epsilon_{ij,j-1}) \quad \text{for } p = 1, \dots, \theta-1, t+1, \dots, T, \\ (II) \quad \mathbf{x}'_{ip} \Delta \epsilon_{it\theta} &= \mathbf{x}'_{ip} (\sum_{j=\theta+1}^{p-1} \Delta \epsilon_{ij,j-1} + \Delta \epsilon_{i,p+1,p-1} + \sum_{j=p+2}^t \Delta \epsilon_{ij,j-1}) \\ &\quad \text{for } p = \theta+1, \dots, t-1, \end{aligned}$$

taking expectations, we then obtain

**Proposition 1:** *When (B1) or (B3), and (C1) or (C3) are satisfied,*

- (a)  $\mathbf{E}[\mathbf{x}'_{ip} (\Delta \epsilon_{it,t-1})] = \mathbf{0}_{K,1}$  for  $p = 1, \dots, t-2, t+1, \dots, T$ ;  $t = 2, \dots, T$  are  $K(T-1)(T-2)$  essential OC's for equations differenced over one period.
- (b)  $\mathbf{E}[\mathbf{x}'_{it} (\Delta \epsilon_{it+1,t-1})] = \mathbf{0}_{K,1}$  for  $t = 2, \dots, T-1$  are  $K(T-2)$  essential OC's for equations differenced over two periods.
- (c) *The other OC's are redundant: among the  $\frac{1}{2}KT(T-1)(T-2)$  conditions in (2.18) when  $T > 2$ , only a fraction of  $2/(T-1)$ , are essential.*

---

<sup>12</sup>See also Crépon and Mairesse (1996, pp. 357 – 358).

Symmetrically, from (2.20) we have

**Proposition 2:** *When (B1), or (B3), and (C1), or (C3), are satisfied,*

- (a)  $E[(\Delta \mathbf{x}_{ip,p-1})' \epsilon_{it}] = \mathbf{0}_{K,1}$  for  $t = 1, \dots, p-2, p+1, \dots, T$ ;  $p = 2, \dots, T$  are  $K(T-1)(T-2)$  essential OC's for equations in levels, with IV's differenced over one period.
- (b)  $E[(\Delta \mathbf{x}_{it+1,t-1})' \epsilon_{it}] = \mathbf{0}_{K,1}$  for  $t = 2, \dots, T-1$  are  $K(T-2)$  essential OC's for equations in levels, with IV's differenced over two periods.
- (c) *The other OC's are redundant: among the  $\frac{1}{2}KT(T-1)(T-2)$  conditions in (2.19) when  $T > 2$ , only a fraction of  $2/(T-1)$ , are essential.*

These propositions can be generalized to the case where  $\epsilon_{it}$  is a  $MA(\tau)$  process. For any finite  $\tau = 0, 1, 2, \dots$ , we have

**Proposition 1\*:** *When (B2) and (C2) are satisfied,*

- (a)  $E[\mathbf{x}'_{ip}(\Delta \epsilon_{it,t-1})] = \mathbf{0}_{K,1}$  for  $p = 1, \dots, t-\tau-2, t+\tau+1, \dots, T$ ;  $t = 2, \dots, T$  are essential OC's for equations differenced over one period.
- (b)  $E[\mathbf{x}'_{it}(\Delta \epsilon_{it+\tau+1,t-\tau-1})] = \mathbf{0}_{K,1}$  for  $t = \tau+2, \dots, T-\tau-1$  are essential OC's for equations differenced over  $2(\tau+1)$  periods.
- (c) *The other OC's in (2.18) are redundant.*

**Proposition 2\*:** *When (B2) and (C2) are satisfied,*

- (a)  $E[(\Delta \mathbf{x}_{ip,p-1})' \epsilon_{it}] = \mathbf{0}_{K,1}$  for  $t = 1, \dots, p-\tau-2, p+\tau+1, \dots, T$ ;  $p = 2, \dots, T$  are essential OC's for equations in levels, with IV's differenced over one period.
- (b)  $E[(\Delta \mathbf{x}_{it+\tau+1,t-\tau-1})' \epsilon_{it}] = \mathbf{0}_{K,1}$  for  $t = \tau+2, \dots, T-\tau-1$  are essential OC's for equations in levels, with IV's differenced over  $2(\tau+1)$  periods.
- (c) *The other OC's in (2.20) are redundant.*

These propositions can be (trivially) modified to include also the essential and redundant OC's in the  $y$ 's or the  $\Delta y$ 's, given in (2.19) and (2.21).

### 3 Estimation by OLS on period means

In this section, we consider estimators of  $\beta$  constructed from period means, to illustrate applications of the repeated measurement property of panel data.

Premultiplying (2.7) by  $\mathbf{e}'_N/N$  and substituting  $\theta = t - s$ , we get the following equation in differenced period means:

$$(3.1) \quad \Delta_s \bar{y}_{\cdot t} = \Delta_s \bar{\mathbf{x}}_{\cdot t} \beta + \Delta_s \bar{\epsilon}_{\cdot t}, \quad s = 1, \dots, T-1; t = s+1, \dots, T,$$

where  $\bar{y}_{\cdot t} = \sum_i y_{it}/N$ ,  $\bar{\mathbf{x}}_{\cdot t} = \sum_i \mathbf{x}_{it}/N$ , etc., and  $\Delta_s$  is the operator differencing over  $s$  periods. Taking differences between period means and global means, we obtain from (2.5)

$$(3.2) \quad (\bar{y}_{\cdot t} - \bar{y}) = (\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}})\boldsymbol{\beta} + (\bar{\epsilon}_{\cdot t} - \bar{\epsilon}), \quad t = 1, \dots, T,$$

where  $\bar{y} = \sum_i \sum_t y_{it}/(NT)$ ,  $\bar{\mathbf{x}} = \sum_i \sum_t \mathbf{x}_{it}/(NT)$ , etc.

Now, the (weak) law of the large numbers, when (A) is satisfied, implies under weak conditions, not including (B2), (B3), (C2), (C3), (D1), and (D2) [cf. McCabe and Tremayne (1993, section 3.5)]<sup>13</sup>,

$$\text{plim}(\bar{\epsilon}_{\cdot t}) = 0, \quad \text{plim}(\bar{\mathbf{x}}_{\cdot t} - \bar{\boldsymbol{\xi}}_{\cdot t}) = \mathbf{0}_{1K}, \quad \text{plim}[\bar{\mathbf{x}}'_{\cdot t} \bar{\epsilon}_{\cdot t}] = \mathbf{0}_{K1},$$

even if  $\text{plim}[(1/N) \sum_{i=1}^N \mathbf{x}'_{it} \epsilon_{it}] \neq \mathbf{0}_{K1}$ . Therefore, (3.1) and (3.2) imply

$$(3.3) \quad \text{plim}[(\Delta_s \bar{\mathbf{x}}_{\cdot t})' (\Delta_s \bar{y}_{\cdot t})] = \text{plim}[(\Delta_s \bar{\mathbf{x}}_{\cdot t})' (\Delta_s \bar{\mathbf{x}}_{\cdot t})] \boldsymbol{\beta},$$

$$(3.4) \quad \text{plim}[(\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}})' (\bar{y}_{\cdot t} - \bar{y})] = \text{plim}[(\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}})' (\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}})] \boldsymbol{\beta}.$$

Hence, if  $E[(\Delta_s \bar{\boldsymbol{\xi}}_{\cdot t})' (\Delta_s \bar{\boldsymbol{\xi}}_{\cdot t})]$  and  $E[(\bar{\boldsymbol{\xi}}_{\cdot t} - \bar{\boldsymbol{\xi}})' (\bar{\boldsymbol{\xi}}_{\cdot t} - \bar{\boldsymbol{\xi}})]$  have rank  $K$ , which are variants of assumptions (E1) – (E3), consistent estimators of  $\boldsymbol{\beta}$  can be obtained by OLS on (3.1) or on (3.2). This gives

$$(3.5) \quad \hat{\boldsymbol{\beta}}_{\Delta_s} = \left[ \sum_{t=s+1}^T (\Delta_s \bar{\mathbf{x}}_{\cdot t})' (\Delta_s \bar{\mathbf{x}}_{\cdot t}) \right]^{-1} \left[ \sum_{t=s+1}^T (\Delta_s \bar{\mathbf{x}}_{\cdot t})' (\Delta_s \bar{y}_{\cdot t}) \right], \quad s = 1, \dots, T-1,$$

$$(3.6) \quad \hat{\boldsymbol{\beta}}_{BP} = \left[ \sum_{t=1}^T (\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}})' (\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}}) \right]^{-1} \left[ \sum_{t=1}^T (\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}})' (\bar{y}_{\cdot t} - \bar{y}) \right],$$

provided that a sufficient number of degrees of freedom remains. The latter is the ‘between period’ (BP) estimator.

These estimators simply exploit the fact that averages of a large number of repeated measurements of an error-ridden variable give under weak conditions an error-free measure of the true average at the limit, *provided that this true average shows variation along the remaining dimension, i.e., across periods*. This intuition explains why the consistency of  $\hat{\boldsymbol{\beta}}_{\Delta_s}$  and  $\hat{\boldsymbol{\beta}}_{BP}$  are at the same time robust to individual heterogeneity and utilize the *average-of-repeated-measurements* possibility allowed by panel data. However, since they do not exploit any inter-individual variation in the data, their efficiency may be low, because this kind of variation often tends to dominate.

---

<sup>13</sup>Here and in the following plim always denotes probability limits when  $N$  goes to infinity and  $T$  is finite.

Basic to the above conclusions is the assumption that *the measurement error has no period specific component*. This assumption is important, since if such a component were present, it would not vanish when taking plims of period means, *i.e.*,  $\text{plim}(\bar{\mathbf{v}}_t)$  would no longer be zero, (3.3) and (3.4) would no longer hold, and so  $\hat{\boldsymbol{\beta}}_{\Delta_s}$  and  $\hat{\boldsymbol{\beta}}_{BP}$  would not be consistent. ‘Within period’ transformations, *e.g.*, differences taken across individuals, would, however, eliminate this kind of measurement error heterogeneity.<sup>14</sup>

## 4 Estimation by IV and GMM – differences vs. levels

In this section, we construct estimators of  $\boldsymbol{\beta}$  by replacing the expectations in (2.18) – (2.21) by sample means taken over  $i$  and minimizing their distances from the zero vector. There are several ways in which this idea can be operationalized; we can

- (i) *Estimate equations in differences, with instruments in levels, using (2.18) and/or (2.19) for (a) one  $(t, \theta)$  and one  $p$ , (b) one  $(t, \theta)$  and several  $p$ , or (c) several  $(t, \theta)$  and several  $p$  jointly.*
- (ii) *Estimate equations in levels, with instruments in differences, using (2.20) and/or (2.21) for (a) one  $t$  and one  $(p, q)$ , (b) one  $t$  and several  $(p, q)$ , or (c) several  $t$  and several  $(p, q)$  jointly.*

In cases (i.a) and (ii.a), we obtain an empirical distance equal to the zero vector, so no minimization is needed. This corresponds, formally, to the situation with ‘exact identification’ (exactly as many OC’s as needed) in classical IV estimation. In cases (i.b), (i.c), (ii.b), and (ii.c), we have, in a formal sense, ‘overidentification’ (more than the necessary number of OC’s), and therefore construct ‘compromise estimators’ by minimizing appropriate quadratic forms in the corresponding empirical distances.

### 4.a Simple IV estimators, one level or one difference

As a starting point, we consider the simplest cases, (i.a) and (ii.a).

**Equation in differences, IV’s in levels.** The sample mean counterpart to (2.18) for one  $(t, \theta, p)$  gives the estimator<sup>15</sup>

$$(4.1) \quad \hat{\boldsymbol{\beta}}_{xp(t\theta)} = [\sum_{i=1}^N \mathbf{x}'_{ip}(\Delta \mathbf{x}_{it\theta})]^{-1} [\sum_{i=1}^N \mathbf{x}'_{ip}(\Delta y_{it\theta})].$$

<sup>14</sup>See Biørn (1996, sections 10.3 and 10.5) for an elaboration of this.

<sup>15</sup>This kind of estimator is considered in Griliches and Hausman (1986, section 3) and Biørn (1996, section 10.2.3). A closely related estimator for an AR(1) equation is proposed in Anderson and Hsiao (1982, p. 59).

If  $K = 1$ , the sample mean counterparts to (2.19) for one  $(t, \theta, p)$  gives

$$(4.2) \quad \widehat{\beta}_{yp(t\theta)} = [\sum_{i=1}^N y_{ip}(\Delta \mathbf{x}_{i\theta})]^{-1} [\sum_{i=1}^N y_{ip}(\Delta y_{i\theta})].$$

If  $K > 1$ , the latter estimator is infeasible, but we may modify (4.1) by replacing one element in  $\mathbf{x}_{ip}$  by  $y_{ip}$ .

**Equation in levels, IV's in differences.** The sample mean counterpart to (2.20) for one  $(t, p, q)$  gives the estimator

$$(4.3) \quad \widehat{\beta}_{x(pq)t} = [\sum_{i=1}^N (\Delta \mathbf{x}_{ipq})' \mathbf{x}_{it}]^{-1} [\sum_{i=1}^N (\Delta \mathbf{x}_{ipq})' y_{it}].$$

If  $K = 1$ , the sample mean counterparts to (2.21) for one  $(t, p, q)$  gives

$$(4.4) \quad \widehat{\beta}_{y(pq)t} = [\sum_{i=1}^N (\Delta y_{ipq}) \mathbf{x}_{it}]^{-1} [\sum_{i=1}^N (\Delta y_{ipq}) y_{it}].$$

If  $K > 1$ , the latter estimator is infeasible, but we may modify (4.3) by replacing one element in  $\Delta \mathbf{x}_{ipq}$  by  $\Delta y_{ipq}$ .

From (2.18) – (2.21) we note that

- Assumption (B1) or (B3) – non-autocorrelation or time invariant autocorrelation of the (level) measurement errors (in the  $x$ 's) – is necessary for consistency of  $\widehat{\beta}_{xp(t\theta)}$  for all  $p \neq \theta, t$  and consistency of  $\widehat{\beta}_{x(pq)t}$  for all  $t \neq p, q$ .
- Assumption (C1) or (C3) – non-autocorrelation or time invariant correlation of the (level) disturbances (including the measurement errors in the  $y$ 's) – is necessary for consistency of  $\widehat{\beta}_{yp(t\theta)}$  for all  $p \neq \theta, t$  and consistency of  $\widehat{\beta}_{y(pq)t}$  for all  $t \neq q, p$ .
- Assumptions (D1) and (D2) – time invariance of  $E(\boldsymbol{\xi}_{it})$  and  $E(\alpha_i \boldsymbol{\xi}_{it})$  – are necessary for consistency of  $\widehat{\beta}_{x(pq)t}$  and  $\widehat{\beta}_{y(pq)t}$ , but they are not necessary for consistency of  $\widehat{\beta}_{xp(t\theta)}$  and  $\widehat{\beta}_{yp(t\theta)}$ .<sup>16</sup>

If we replace (B1) and (C1) by the weaker assumptions (B2) and (C2), we must ensure that the IV set has a lag or lead of at least  $\tau + 1$  periods to the regressor to ‘get clear of’ the  $\tau$  period memory of the MA process:

- If  $|t - p|, |\theta - p| > \tau$ , (B2) and (E2) ensure consistency of  $\widehat{\beta}_{xp(t\theta)}$ , and (C2) and (E2) ensure consistency of  $\widehat{\beta}_{yp(t\theta)}$ .
- If  $|p - t|, |q - t| > \tau$ , (B2), (D1), (D2), and (E3) ensure consistency of  $\widehat{\beta}_{x(pq)t}$ , and (C2), (D1), (D2), and (E3) ensure consistency of  $\widehat{\beta}_{y(pq)t}$ .

---

<sup>16</sup>Modifications of  $\widehat{\beta}_{x(pq)t}$  and  $\widehat{\beta}_{y(pq)t}$  which retain consistency when (D1) is relaxed, will be discussed in Section 4.d.



Since the correlation between the regressors and the instruments, say between  $\mathbf{x}_{ip}$  and  $\Delta\mathbf{x}_{it\theta}$ , may be low, (4.1) – (4.4) may suffer from the ‘weak instrument problem’, discussed in Nelson and Startz (1990) and Staiger and Stock (1997). Composite IV or GMM estimates may be an answer to this problem.

#### 4.b Composite IV estimators, one level or one difference

We next consider estimation of  $\beta$  in (2.6) for one pair of periods  $(t, \theta)$ , utilizing as IV’s for  $\Delta\mathbf{x}_{it\theta}$  all admissible  $\mathbf{x}_{ip}$ ’s, and estimation of  $\beta$  in (2.3), *i.e.*, for one period  $(t)$ , utilizing as IV’s for  $\mathbf{x}_{it}$  all admissible  $\Delta\mathbf{x}_{ipq}$ ’s.<sup>17</sup>

To formalize this, we define the selection matrices

$$\mathbf{P}_{t\theta} = \left[ \begin{array}{l} ((T-2) \times T) \text{ matrix} \\ \text{obtained by deleting} \\ \text{from } \mathbf{I}_T \text{ rows } t \text{ and } \theta \end{array} \right], \quad t, \theta = 1, \dots, T,$$

and the stacked  $[(T-2) \times T]$  differencing matrices

$$\mathbf{D}_t = \left[ \begin{array}{c} \mathbf{d}_{21} \\ \vdots \\ \mathbf{d}_{t-1, t-2} \\ \mathbf{d}_{t+1, t-1} \\ \mathbf{d}_{t+2, t+1} \\ \vdots \\ \mathbf{d}_{T, T-1} \end{array} \right], \quad t = 1, \dots, T.$$

The latter are one-period differencing matrices, except that  $\mathbf{d}_{t, t-1}$  and  $\mathbf{d}_{t+1, t}$  are replaced by their sum,  $\mathbf{d}_{t+1, t-1}$ .<sup>18</sup> We use the notation

$$\begin{aligned} \mathbf{y}_{i(t\theta)} &= \mathbf{P}_{t\theta} \mathbf{y}_{i\cdot}, & \mathbf{X}_{i(t\theta)} &= \mathbf{P}_{t\theta} \mathbf{X}_{i\cdot}, & \mathbf{x}_{i(t\theta)} &= \text{vec}(\mathbf{X}_{i(t\theta)})', \\ \Delta\mathbf{y}_{i(t)} &= \mathbf{D}_t \mathbf{y}_{i\cdot}, & \Delta\mathbf{X}_{i(t)} &= \mathbf{D}_t \mathbf{X}_{i\cdot}, & \Delta\mathbf{x}_{i(t)} &= \text{vec}(\Delta\mathbf{X}_{i(t)})', \end{aligned}$$

etc. Here  $\mathbf{X}_{i(t\theta)}$  denotes the  $[(T-2) \times K]$  matrix of  $\mathbf{x}$  levels obtained by *deleting* rows  $t$  and  $\theta$  from  $\mathbf{X}_{i\cdot}$ , and  $\Delta\mathbf{X}_{i(t)}$  denotes the  $[(T-2) \times K]$  matrix of  $\mathbf{x}$  differences obtained by stacking all one-period differences between rows of  $\mathbf{X}_{i\cdot}$  *not including* period  $t$  and the single two-period difference between the columns for periods  $t+1$  and  $t-1$ . The vectors  $\mathbf{y}_{i(t\theta)}$  and  $\Delta\mathbf{y}_{i(t)}$  are constructed from  $\mathbf{y}_{i\cdot}$  in a similar way. In general, we let subscripts  $(t\theta)$  and  $(t)$  on a matrix or vector denote deletion of  $(t\theta)$  differences and  $t$  levels, respectively. Stacking  $\mathbf{y}'_{i(t\theta)}$ ,  $\Delta\mathbf{y}'_{i(t)}$ ,  $\mathbf{x}_{i(t\theta)}$ , and  $\Delta\mathbf{x}_{i(t)}$ ,

<sup>17</sup>The former, for  $K = 1$ , is considered by Griliches and Hausman (1986). See also Biørn and Klette (1997a, b).

<sup>18</sup>The two-period difference is effective only for  $t = 2, \dots, T-1$ .

by individuals, we get

$$\mathbf{Y}_{(t\theta)} = \begin{bmatrix} \mathbf{y}'_{1(t\theta)} \\ \vdots \\ \mathbf{y}'_{N(t\theta)} \end{bmatrix}, \Delta \mathbf{Y}_{(t)} = \begin{bmatrix} \Delta \mathbf{y}'_{1(t)} \\ \vdots \\ \Delta \mathbf{y}'_{N(t)} \end{bmatrix}, \mathbf{X}_{(t\theta)} = \begin{bmatrix} \mathbf{x}_{1(t\theta)} \\ \vdots \\ \mathbf{x}_{N(t\theta)} \end{bmatrix}, \Delta \mathbf{X}_{(t)} = \begin{bmatrix} \Delta \mathbf{x}_{1(t)} \\ \vdots \\ \Delta \mathbf{x}_{N(t)} \end{bmatrix},$$

which have dimensions  $(N \times (T-2))$ ,  $(N \times (T-2))$ ,  $(N \times (T-2)K)$ , and  $(N \times (T-2)K)$ , respectively. These four matrices contain the IV's to be considered below.

**Equation in differences, IV's in levels.** Using  $\mathbf{X}_{(t\theta)}$  as IV matrix for  $\Delta \mathbf{X}_{t\theta}$  in (2.7), we obtain the following estimator of  $\beta$ , *specific to period  $(t, \theta)$  differences and utilizing all admissible  $\mathbf{x}$  level IV's*,

$$(4.5) \quad \begin{aligned} \hat{\beta}_{x(t\theta)} &= \left[ (\Delta \mathbf{X}_{t\theta})' \mathbf{X}_{(t\theta)} \left( \mathbf{X}'_{(t\theta)} \mathbf{X}_{(t\theta)} \right)^{-1} \mathbf{X}'_{(t\theta)} (\Delta \mathbf{X}_{t\theta}) \right]^{-1} \\ &\quad \times \left[ (\Delta \mathbf{X}_{t\theta})' \mathbf{X}_{(t\theta)} \left( \mathbf{X}'_{(t\theta)} \mathbf{X}_{(t\theta)} \right)^{-1} \mathbf{X}'_{(t\theta)} (\Delta \mathbf{y}_{t\theta}) \right] \\ &= \left[ \left[ \sum_i (\Delta \mathbf{x}_{it\theta})' \mathbf{x}_{i(t\theta)} \right] \left[ \sum_i \mathbf{x}'_{i(t\theta)} \mathbf{x}_{i(t\theta)} \right]^{-1} \left[ \sum_i \mathbf{x}'_{i(t\theta)} (\Delta \mathbf{x}_{it\theta}) \right] \right]^{-1} \\ &\quad \times \left[ \left[ \sum_i (\Delta \mathbf{x}_{it\theta})' \mathbf{x}_{i(t\theta)} \right] \left[ \sum_i \mathbf{x}'_{i(t\theta)} \mathbf{x}_{i(t\theta)} \right]^{-1} \left[ \sum_i \mathbf{x}'_{i(t\theta)} (\Delta y_{it\theta}) \right] \right]. \end{aligned}$$

It exists if  $\mathbf{X}'_{(t\theta)} \mathbf{X}_{(t\theta)}$  has rank  $(T-2)K$ , which requires  $N \geq (T-2)K$ . This GMM estimator utilizes the OC  $\mathbf{E}[\mathbf{x}'_{i(t\theta)} (\Delta \epsilon_{it\theta})] = \mathbf{0}_{(T-2)K,1}$ , which follows from (2.18), and minimizes the quadratic form:

$$\hat{\beta}_{x(t\theta)} = \operatorname{argmin} \left( \frac{1}{N} \mathbf{X}'_{(t\theta)} \Delta \epsilon_{t\theta} \right)' \left( \frac{1}{N^2} \mathbf{X}'_{(t\theta)} \mathbf{X}_{(t\theta)} \right)^{-1} \left( \frac{1}{N} \mathbf{X}'_{(t\theta)} \Delta \epsilon_{t\theta} \right).$$

The weight matrix  $(N^{-2} \mathbf{X}'_{(t\theta)} \mathbf{X}_{(t\theta)})^{-1}$  is proportional to the inverse of the (asymptotic) covariance matrix of  $N^{-1} \mathbf{X}'_{(t\theta)} \Delta \epsilon_{t\theta}$  when  $\Delta \epsilon_{it\theta}$  is IID across  $i$ , possibly with a variance depending on  $(t, \theta)$ . The consistency of  $\hat{\beta}_{x(t\theta)}$  relies on (B1), or (B3), and the validity of (E2) for all  $p$ .

Two modifications of  $\hat{\beta}_{x(t\theta)}$  exist:

- (1) If  $\operatorname{var}(\Delta \epsilon_{it\theta}) = \omega_{it\theta}$  varies with  $i$  and is known, we can increase the efficiency of (4.5) by replacing  $\mathbf{x}'_{i(t\theta)} \mathbf{x}_{i(t\theta)}$  by  $\mathbf{x}'_{i(t\theta)} \omega_{it\theta} \mathbf{x}_{i(t\theta)}$ , which gives an asymptotically optimal GMM estimator.<sup>19</sup> Estimation of  $\sum_i \mathbf{x}'_{i(t\theta)} \omega_{it\theta} \mathbf{x}_{i(t\theta)}$  for unknown  $\omega_{it\theta}$  is elaborated in Appendix A.
- (2) Instead of using  $\mathbf{X}_{(t\theta)}$  as IV matrix for  $\Delta \mathbf{X}_{t\theta}$  in (2.7), we may either, if  $K = 1$ , use  $\mathbf{Y}_{(t\theta)}$ , or, for arbitrary  $K$ ,  $\mathbf{Z}_{(t\theta)} = (\mathbf{X}_{(t\theta)} : \mathbf{Y}_{(t\theta)})$ .

---

<sup>19</sup>For a more general treatment of asymptotic efficiency in estimation with moment conditions, see Chamberlain (1987) and Newey and McFadden (1994).

**Equation in levels, IV's in differences.** Using  $\Delta\mathbf{X}_{(t)}$  as IV matrix for  $\mathbf{X}_t$  in (2.5) (omitting, for notational simplicity, the 'dot' subscript on  $\mathbf{X}_{\cdot,t}$  and  $\mathbf{y}_{\cdot,t}$ ) we get the following estimator of  $\beta$ , *specific to period  $t$  levels, utilizing all admissible  $x$  difference IV's*,

$$(4.6) \quad \begin{aligned} \widehat{\beta}_{x(t)} &= \left[ \mathbf{X}'_t(\Delta\mathbf{X}_{(t)}) \left( (\Delta\mathbf{X}_{(t)})'(\Delta\mathbf{X}_{(t)}) \right)^{-1} (\Delta\mathbf{X}_{(t)})' \mathbf{X}_t \right]^{-1} \\ &\quad \times \left[ \mathbf{X}'_t(\Delta\mathbf{X}_{(t)}) \left( (\Delta\mathbf{X}_{(t)})'(\Delta\mathbf{X}_{(t)}) \right)^{-1} (\Delta\mathbf{X}_{(t)})' \mathbf{y}_t \right] \\ &= \left[ \left[ \sum_i \mathbf{x}'_{it}(\Delta\mathbf{x}_{i(t)}) \right] \left[ \sum_i (\Delta\mathbf{x}_{i(t)})'(\Delta\mathbf{x}_{i(t)}) \right]^{-1} \left[ \sum_i (\Delta\mathbf{x}_{i(t)})' \mathbf{x}_{it} \right] \right]^{-1} \\ &\quad \times \left[ \left[ \sum_i \mathbf{x}'_{it}(\Delta\mathbf{x}_{i(t)}) \right] \left[ \sum_i (\Delta\mathbf{x}_{i(t)})'(\Delta\mathbf{x}_{i(t)}) \right]^{-1} \left[ \sum_i (\Delta\mathbf{x}_{i(t)})' \mathbf{y}_{it} \right] \right]. \end{aligned}$$

It exists if  $(\Delta\mathbf{X}_{(t)})'(\Delta\mathbf{X}_{(t)})$  has rank  $(T-2)K$ , which again requires  $N \geq (T-2)K$ . This GMM estimator utilizes the OC  $\mathbf{E}[(\Delta\mathbf{x}_{i(t)})'\epsilon_{it}] = \mathbf{0}_{(T-2)K,1}$ , which follows from (2.20), and minimizes the quadratic form:

$$\widehat{\beta}_{x(t)} = \operatorname{argmin} \left( \frac{1}{N}(\Delta\mathbf{X}_{(t)})'\epsilon_t \right)' \left[ \frac{1}{N^2}(\Delta\mathbf{X}_{(t)})'(\Delta\mathbf{X}_{(t)}) \right]^{-1} \left( \frac{1}{N}(\Delta\mathbf{X}_{(t)})'\epsilon_t \right).$$

The weight matrix  $[N^{-2}(\Delta\mathbf{X}_{(t)})'(\Delta\mathbf{X}_{(t)})]^{-1}$  is proportional to the inverse of the (asymptotic) covariance matrix of  $N^{-1}(\Delta\mathbf{X}_{(t)})'\epsilon_t$  when  $\epsilon_{it}$  is IID across  $i$ , possibly with a variance depending on  $t$ . The consistency of  $\widehat{\beta}_{x(t)}$  relies on (B1), or (B3), (D1), (D2), and the validity of (E3) for all  $(p, q)$ .

Two modifications of  $\widehat{\beta}_{x(t)}$  exist:

- (1) If  $\operatorname{var}(\epsilon_{it}) = \omega_{it}$  varies with on  $i$  and is known, we can increase the efficiency of (4.6) by replacing  $(\Delta\mathbf{x}_{i(t)})'(\Delta\mathbf{x}_{i(t)})$  by  $(\Delta\mathbf{x}_{i(t)})'\omega_{it}(\Delta\mathbf{x}_{i(t)})$ , which gives an asymptotically optimal GMM estimator. Estimation of  $\sum_i (\Delta\mathbf{x}_{i(t)})'\omega_{it}(\Delta\mathbf{x}_{i(t)})$  for unknown  $\omega_{it}$  is elaborated in Appendix A.
- (2) Instead of using  $\Delta\mathbf{X}_{(t)}$  as IV matrix for  $\mathbf{X}_t$  in (2.5), we may either, if  $K = 1$ , use  $\Delta\mathbf{Y}_{(t)}$ , or, for arbitrary  $K$ ,  $\Delta\mathbf{Z}_{(t)} = (\Delta\mathbf{X}_{(t)} \dot{\vdots} \Delta\mathbf{Y}_{(t)})$ .

If we find (B1) or (C1) too restrictive and replace them by (B2) or (C2), we must ensure that the variables in the IV matrix have a lead or lag of at least  $\tau + 1$  periods to the regressor, to 'get clear of' the  $\tau$  period memory of the  $\operatorname{MA}(\tau)$  process. Formally, we then replace  $\mathbf{P}_{t\theta}$  and  $\mathbf{D}_t$  by<sup>20</sup>

$$\mathbf{P}_{t\theta(\tau)} = \left[ \begin{array}{c} \text{matrix obtained by} \\ \text{deleting from } \mathbf{I}_T \\ \text{rows } \theta - \tau, \dots, \theta + \tau \\ \text{and } t - \tau, \dots, t + \tau \end{array} \right], \quad t, \theta = 1, \dots, T,$$

<sup>20</sup>The number of rows in these matrices depends, in general, on  $\tau$ .

and

$$\mathbf{D}_{t(\tau)} = \begin{bmatrix} \mathbf{d}_{21} \\ \vdots \\ \mathbf{d}_{t-\tau-1,t-\tau-2} \\ \mathbf{d}_{t+\tau+1,t-\tau-1} \\ \mathbf{d}_{t+\tau+2,t+\tau+1} \\ \vdots \\ \mathbf{d}_{T,T-1} \end{bmatrix}, \quad t = 1, \dots, T.$$

and otherwise proceed as above. The matrix  $\mathbf{D}_{t(\tau)}$  is a one-period differencing matrix, except that  $\mathbf{d}_{t-\tau,t-\tau-1}, \dots, \mathbf{d}_{t+\tau+1,t+\tau}$  are replaced by their sum,  $\mathbf{d}_{t+\tau+1,t-\tau-1}$ .

#### 4.c GMM estimators, all levels or differences

We finally consider GMM estimation of  $\beta$  when we combine all essential OC's as given in Proposition 1 and in Proposition 2.<sup>21</sup>

**Equation in differences, IV's in levels.** Consider (2.6) for all  $\theta = t - 1$  and all  $\theta = t - 2$ . These  $(T-1) + (T-2)$  equations stacked for individual  $i$  read

$$(4.7) \quad \begin{bmatrix} \Delta y_{i21} \\ \Delta y_{i32} \\ \vdots \\ \Delta y_{i,T,T-1} \\ \Delta y_{i31} \\ \Delta y_{i42} \\ \vdots \\ \Delta y_{i,T,T-2} \end{bmatrix} = \begin{bmatrix} \Delta \mathbf{x}_{i21} \\ \Delta \mathbf{x}_{i32} \\ \vdots \\ \Delta \mathbf{x}_{i,T,T-1} \\ \Delta \mathbf{x}_{i31} \\ \Delta \mathbf{x}_{i42} \\ \vdots \\ \Delta \mathbf{x}_{i,T,T-2} \end{bmatrix} \beta + \begin{bmatrix} \Delta \epsilon_{i21} \\ \Delta \epsilon_{i32} \\ \vdots \\ \Delta \epsilon_{i,T,T-1} \\ \Delta \epsilon_{i31} \\ \Delta \epsilon_{i42} \\ \vdots \\ \Delta \epsilon_{i,T,T-2} \end{bmatrix},$$

or, compactly,

$$\Delta \mathbf{y}_i = (\Delta \mathbf{X}_i) \beta + \Delta \epsilon_i.$$

---

<sup>21</sup>The procedures to be described below assume that (B1) and (C1) are satisfied. If we find these conditions too restrictive and replace them by (B2) or (C2), we must ensure once again that the variables in the IV matrix have a lead or lag of at least  $\tau + 1$  periods to the regressor, to 'get clear of' the  $\tau$  period memory of the MA( $\tau$ ) process, cf. Propositions 1\* and 2\*. Otherwise, we can proceed as described below.

The IV matrix, according to Proposition 1,<sup>22</sup> is the  $((2T-3) \times KT(T-2))$  matrix<sup>23</sup>

$$(4.8) \quad \mathbf{Z}_i = \begin{bmatrix} \mathbf{x}_{i(21)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i(32)} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{i(T,T-1)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{x}_{i2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{x}_{i3} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{i,T-1} \end{bmatrix}.$$

Let

$$\begin{aligned} \Delta \mathbf{y} &= [(\Delta \mathbf{y}_1)', \dots, (\Delta \mathbf{y}_N)']', & \Delta \boldsymbol{\epsilon} &= [(\Delta \boldsymbol{\epsilon}_1)', \dots, (\Delta \boldsymbol{\epsilon}_N)']', \\ \Delta \mathbf{X} &= [(\Delta \mathbf{X}_1)', \dots, (\Delta \mathbf{X}_N)']', & \mathbf{Z} &= [\mathbf{Z}'_1, \dots, \mathbf{Z}'_N]'. \end{aligned}$$

The GMM estimator corresponding to (2.18), which we now write as  $\mathbb{E}[\mathbf{Z}'_i(\Delta \boldsymbol{\epsilon}_i)] = \mathbf{0}_{T(T-2)K,1}$ , minimizing  $[N^{-1}(\Delta \boldsymbol{\epsilon})' \mathbf{Z}][N^{-2} \mathbf{V}]^{-1}[N^{-1} \mathbf{Z}'(\Delta \boldsymbol{\epsilon})]$  for  $\mathbf{V} = \mathbf{Z}' \mathbf{Z}$ , can be written as

$$(4.9) \quad \begin{aligned} \hat{\boldsymbol{\beta}}_{Dx} &= [(\Delta \mathbf{X})' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' (\Delta \mathbf{X})]^{-1} [(\Delta \mathbf{X})' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' (\Delta \mathbf{y})] \\ &= \left[ \sum_i (\Delta \mathbf{X}_i)' \mathbf{Z}_i \right] \left[ \sum_i \mathbf{Z}_i' \mathbf{Z}_i \right]^{-1} \left[ \sum_i \mathbf{Z}_i' (\Delta \mathbf{X}_i) \right]^{-1} \\ &\quad \times \left[ \sum_i (\Delta \mathbf{X}_i)' \mathbf{Z}_i \right] \left[ \sum_i \mathbf{Z}_i' \mathbf{Z}_i \right]^{-1} \left[ \sum_i \mathbf{Z}_i' (\Delta \mathbf{y}_i) \right]. \end{aligned}$$

If  $\Delta \boldsymbol{\epsilon}$  has a non-scalar covariance matrix, a more efficient GMM estimator is obtained for  $\mathbf{V} = \mathbf{V}_{Z(\Delta \boldsymbol{\epsilon})} = \mathbb{E}[\mathbf{Z}'(\Delta \boldsymbol{\epsilon})(\Delta \boldsymbol{\epsilon})' \mathbf{Z}]$ , which gives

$$(4.10) \quad \tilde{\boldsymbol{\beta}}_{Dx} = [(\Delta \mathbf{X})' \mathbf{Z} \mathbf{V}_{Z(\Delta \boldsymbol{\epsilon})}^{-1} \mathbf{Z}' (\Delta \mathbf{X})]^{-1} [(\Delta \mathbf{X})' \mathbf{Z} \mathbf{V}_{Z(\Delta \boldsymbol{\epsilon})}^{-1} \mathbf{Z}' (\Delta \mathbf{y})].$$

We can estimate  $\mathbf{V}_{Z(\Delta \boldsymbol{\epsilon})}/N$  consistently from the residuals obtained from (4.9),  $\widehat{\Delta \boldsymbol{\epsilon}}_i = \Delta \mathbf{y}_i - (\Delta \mathbf{X}_i) \tilde{\boldsymbol{\beta}}_{Dx}$ , by means of [see White (1984, sections IV.3 and VI.2) and (1986, section 3)]

$$(4.11) \quad \frac{\widehat{\mathbf{V}}_{Z(\Delta \boldsymbol{\epsilon})}}{N} = \frac{1}{N} \sum_{i=1}^N \mathbf{Z}'_i (\widehat{\Delta \boldsymbol{\epsilon}}_i) (\widehat{\Delta \boldsymbol{\epsilon}}_i)' \mathbf{Z}_i.$$

<sup>22</sup>It is possible to include not only the essential OC's in this kind of GMM estimation, but also the redundant OC's. The singularity problems involved may be treated by replacing standard inverses in the estimation formulae by Moore-Penrose inverses, see Biørn and Klette (1997a,b).

<sup>23</sup>Formally, we here use different IV's for the  $(T-1) + (T-2)$  different equations in (4.7), with  $\boldsymbol{\beta}$  as a common slope coefficient. This kind of flexibility is one of the attractions of GMM as compared with standard two stage and three stage least squares; see Schmidt (1990).

Inserting (4.11) in (4.10), we get the asymptotically optimal (feasible) GMM estimator

$$(4.12) \quad \tilde{\beta}_{Dx} = \left[ \sum_i (\Delta \mathbf{X}_i)' \mathbf{Z}_i \left[ \sum_i \mathbf{Z}_i' \widehat{\Delta \epsilon}_i \widehat{\Delta \epsilon}_i' \mathbf{Z}_i \right]^{-1} \sum_i \mathbf{Z}_i' (\Delta \mathbf{X}_i) \right]^{-1} \\ \times \left[ \sum_i (\Delta \mathbf{X}_i)' \mathbf{Z}_i \left[ \sum_i \mathbf{Z}_i' \widehat{\Delta \epsilon}_i \widehat{\Delta \epsilon}_i' \mathbf{Z}_i \right]^{-1} \sum_i \mathbf{Z}_i' (\Delta \mathbf{y}_i) \right].$$

These estimators can be modified by extending all  $\mathbf{x}_{i(t,t-1)}$  to  $(\mathbf{x}_{i(t,t-1)}; \mathbf{y}'_{i(t,t-1)})$  and all  $\mathbf{x}_{it}$  to  $(\mathbf{x}_{it}; \mathbf{y}_{it})$  in (4.8), which also exploit the OC's in the  $\mathbf{y}$ 's.

**Equation in levels, IV's in differences.** We next consider the procedures for the equation in levels with the IV's in differences. The  $T$  stacked level equations for individual  $i$  are [cf. (2.3)]

$$(4.13) \quad \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} c \\ \vdots \\ c \end{bmatrix} + \begin{bmatrix} \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{iT} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{iT} \end{bmatrix},$$

or, compactly [cf. (2.5), omitting the 'dot' subscript],

$$\mathbf{y}_i = \mathbf{e}_T c + \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i.$$

The IV matrix, according to Proposition 2, is the  $(T \times T(T-2)K)$  matrix<sup>24</sup>

$$(4.14) \quad \Delta \mathbf{Z}_i = \begin{bmatrix} \Delta \mathbf{x}_{i(1)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Delta \mathbf{x}_{i(T)} \end{bmatrix}.$$

Let

$$\mathbf{y} = [\mathbf{y}'_1, \dots, \mathbf{y}'_N]', \quad \boldsymbol{\epsilon} = [\boldsymbol{\epsilon}'_1, \dots, \boldsymbol{\epsilon}'_N]', \\ \mathbf{X} = [\mathbf{X}'_1, \dots, \mathbf{X}'_N]', \quad \Delta \mathbf{Z} = [(\Delta \mathbf{Z}'_1)', \dots, (\Delta \mathbf{Z}'_N)']'.$$

The GMM estimator corresponding to (2.20), which we now write as  $E[(\Delta \mathbf{Z}_i)' \boldsymbol{\epsilon}_i] = \mathbf{0}_{T(T-2)K,1}$ , minimizing  $[N^{-1} \boldsymbol{\epsilon}' (\Delta \mathbf{Z})] (N^{-2} \mathbf{V}_\Delta)^{-1} [N^{-1} (\Delta \mathbf{Z})' \boldsymbol{\epsilon}]$  for  $\mathbf{V}_\Delta = (\Delta \mathbf{Z})' (\Delta \mathbf{Z})$ , can be written as

$$(4.15) \quad \hat{\beta}_{Lx} = \left[ \mathbf{X}' (\Delta \mathbf{Z}) [(\Delta \mathbf{Z})' (\Delta \mathbf{Z})]^{-1} (\Delta \mathbf{Z})' \mathbf{X} \right]^{-1} \\ \times \left[ \mathbf{X}' (\Delta \mathbf{Z}) [(\Delta \mathbf{Z})' (\Delta \mathbf{Z})]^{-1} (\Delta \mathbf{Z})' \mathbf{y} \right] \\ = \left[ \sum_i \mathbf{X}_i' (\Delta \mathbf{Z}_i) \left[ \sum_i (\Delta \mathbf{Z}_i)' (\Delta \mathbf{Z}_i) \right]^{-1} \sum_i (\Delta \mathbf{Z}_i)' \mathbf{X}_i \right]^{-1} \\ \times \left[ \sum_i \mathbf{X}_i' (\Delta \mathbf{Z}_i) \left[ \sum_i (\Delta \mathbf{Z}_i)' (\Delta \mathbf{Z}_i) \right]^{-1} \sum_i (\Delta \mathbf{Z}_i)' \mathbf{y}_i \right].$$

---

<sup>24</sup>Again, we formally use different IV's for different equations, considering (4.13) as  $T$  different equations with  $\boldsymbol{\beta}$  as a common slope coefficient.

If  $\epsilon$  has a non-scalar covariance matrix, a more efficient GMM estimator is obtained for  $\mathbf{V}_\Delta = \mathbf{V}_{(\Delta Z)\epsilon} = \mathbf{E}[(\Delta \mathbf{Z})' \epsilon \epsilon' (\Delta \mathbf{Z})]$ , which gives

$$(4.16) \quad \tilde{\boldsymbol{\beta}}_{Lx} = \left[ \mathbf{X}'(\Delta \mathbf{Z}) \mathbf{V}_{(\Delta Z)\epsilon}^{-1} (\Delta \mathbf{Z})' \mathbf{X} \right]^{-1} \left[ \mathbf{X}'(\Delta \mathbf{Z}) \mathbf{V}_{(\Delta Z)\epsilon}^{-1} (\Delta \mathbf{Z})' \mathbf{y} \right].$$

We can estimate  $\mathbf{V}_{(\Delta Z)\epsilon}/N$  consistently from the residuals obtained from (4.15),<sup>25</sup>  $\hat{\boldsymbol{\epsilon}}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{Lx}$ , by

$$(4.17) \quad \widehat{\mathbf{V}}_{(\Delta Z)\epsilon} = \frac{1}{N} \sum_{i=1}^N (\Delta \mathbf{Z}_i)' \hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i' (\Delta \mathbf{Z}_i).$$

Inserting (4.17) in (4.16), we get the asymptotically optimal (feasible) GMM estimator

$$(4.18) \quad \tilde{\boldsymbol{\beta}}_{Lx} = \left[ \sum_i \mathbf{X}_i' (\Delta \mathbf{Z}_i) \left[ \sum_i (\Delta \mathbf{Z}_i)' \hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i' (\Delta \mathbf{Z}_i) \right]^{-1} \left[ \sum_i (\Delta \mathbf{Z}_i)' \mathbf{X}_i \right] \right]^{-1} \\ \times \left[ \sum_i \mathbf{X}_i' (\Delta \mathbf{Z}_i) \left[ \sum_i (\Delta \mathbf{Z}_i)' \hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i' (\Delta \mathbf{Z}_i) \right]^{-1} \left[ \sum_i (\Delta \mathbf{Z}_i)' \mathbf{y}_i \right] \right].$$

These estimators can be modified by extending all  $\Delta \mathbf{x}_{i(t)}$  to  $(\Delta \mathbf{x}_{i(t)}; \Delta \mathbf{y}_{i(t)})$  in (4.14), which also exploit the OC's in the  $\Delta \mathbf{y}$ 's. Other moment estimators, which we do not consider in our EIV context, are discussed for situations with predetermined IV's in Ziliak (1997), with the purpose of reducing the potential finite sample bias of the asymptotically optimal GMM estimators.

#### 4.d Non-stationarity and the intercept term in the level equation

The treatment of the intercept term  $c$  in estimating the level equation when using the OC's (2.20) and (2.21) needs a comment. A basic point is that when (D1) holds, using IV's in differences annihilates  $c$  in the moment equations, since then  $\mathbf{E}(\Delta \mathbf{x}_{ipq}) = \mathbf{0}_{1K}$  and  $\mathbf{E}(\Delta y_{ipq}) = 0$ .

If, however, we relax the mean stationarity assumption (D1), which cannot be assumed to hold in many practical situations, we get

$$\mathbf{E}[(\Delta \mathbf{x}_{ipq})' \epsilon_{it}] = \mathbf{E}[(\Delta \mathbf{x}_{ipq})' y_{it}] - \mathbf{E}[(\Delta \mathbf{x}_{ipq})' c] - \mathbf{E}[(\Delta \mathbf{x}_{ipq})' \mathbf{x}_{it}] \boldsymbol{\beta} = \mathbf{0}_{K1}, \\ \mathbf{E}[(\Delta y_{ipq}) \epsilon_{it}] = \mathbf{E}[(\Delta y_{ipq}) y_{it}] - \mathbf{E}[(\Delta y_{ipq}) c] - \mathbf{E}[(\Delta y_{ipq}) \mathbf{x}_{it}] \boldsymbol{\beta} = 0.$$

Using  $\mathbf{E}(\epsilon_{it}) = \mathbf{E}(y_{it}) - c - \mathbf{E}(\mathbf{x}_{it}) \boldsymbol{\beta} = 0$  to eliminate  $c$  leads to the following modifi-

---

<sup>25</sup>We can here omit the intercept  $c$ ; see Section 4.d.

cations of (2.20) and (2.21):

$$(4.19) \quad \left\{ \begin{array}{l} \text{When (B1), or (B3), and (D2) hold, and } t \neq p \neq q, \\ \text{or when (B2) and (D2) hold, and } |t-p|, |t-q| > \tau, \text{ then} \\ \mathbf{E}[(\Delta \mathbf{x}_{ipq})' \epsilon_{it}] = \mathbf{E}[(\Delta \mathbf{x}_{ipq})'(y_{it} - \mathbf{E}(y_{it}))] \\ \quad - \mathbf{E}[(\Delta \mathbf{x}_{ipq})'(\mathbf{x}_{it} - \mathbf{E}(\mathbf{x}_{it}))] \boldsymbol{\beta} = \mathbf{0}_{K1}. \end{array} \right.$$

$$(4.20) \quad \left\{ \begin{array}{l} \text{When (C1), or (C3), and (D2) hold, and } t \neq p \neq q, \\ \text{or when (C2), and (D2) hold, and } |t-p|, |t-q| > \tau, \text{ then} \\ \mathbf{E}[(\Delta y_{ipq}) \epsilon_{it}] = \mathbf{E}[(\Delta y_{ipq})(y_{it} - \mathbf{E}(y_{it}))] \\ \quad - \mathbf{E}[(\Delta y_{ipq})(\mathbf{x}_{it} - \mathbf{E}(\mathbf{x}_{it}))] \boldsymbol{\beta} = \mathbf{0}. \end{array} \right.$$

To implement these modified OC's in the GMM procedures described in Sections 4.a – 4.c for the level equation, we could replace  $\mathbf{E}(y_{it})$  and  $\mathbf{E}(\mathbf{x}_{it})$  by corresponding global or period specific sample means and otherwise proceed as above. An illustration will be given in Section 5.

## 5 Illustration: Input elasticities in manufacturing

In this section, we present an empirical application of some of the estimators in Section 4 for a very simple relationship with a single regressor ( $K = 1$ ). The data are from eight successive annual Norwegian manufacturing censuses, collected by Statistics Norway, for the sector *Manufacture of textiles* (ISIC 32), with  $N = 215$  firms observed in the  $T = 8$  years 1983 – 1990. Two normalizations of the relationship are considered. In version A, the  $y_{it}$  and the  $x_{it}$  variables are, respectively, the log of the material input and the log of gross production, both measured as values at constant prices. In version B, their interpretation is reversed. This means that  $\boldsymbol{\beta}$  can be interpreted as the *marginal input elasticity of materials* in version A and as its inverse in version B, both assumed to be one in simple (Leontief) input-output analysis. The input and the output measure, not least the latter, may be thought to contain potential measurement errors both in the strict and wide sense. With this interpretation, the heterogeneity variable,  $\alpha_i$ , may pick up, *inter alia*, firm specific differences in technology.

The results are given in Tables 1 and 2 for the equations in differences and in Tables 3 and 4 for the equations in levels. Our notation convention is that ‘hat’ estimates are based on standard unweighted GMM formula, ‘tilded’ estimates are the asymptotically efficient GMM estimates based on the estimated optimal weighting matrices, as described in Sections 4.b and 4.c. The subscripts  $x$ ,  $y$ , and  $xy$  refer to



the instruments. The procedures used for calculating asymptotic standard deviation estimates are described in Appendix A.<sup>26</sup>

*Table 1* gives period specific GMM estimates of  $\beta$  obtained from the *differenced equation* for 5 selected differences (among the  $\frac{1}{2}T(T-1) = 28$ ), the IV's being level  $x$ 's (columns 1 and 4), level  $y$ 's (columns 2 and 5), or both of them (columns 3 and 6). The 'hat' estimates in columns 1 – 3 are based on (4.5), the 'tilde' estimates in columns 4 – 6 are the corresponding asymptotically efficient GMM estimates based on the estimated optimal weighting matrix, as described in Section 4.b. The estimates in panel A are uniformly higher than those in panel B, which agrees with the fact that they can be interpreted as obtained by running two stage least squares on the 'original' and on the 'reverse regression', respectively. Under both normalizations, the estimates utilizing the  $y$  instruments tend to exceed those based on the  $x$  instruments. The estimates combining the  $x$  and  $y$  instruments are not invariably between those based on either of them, and the former are not invariably more precise. Using the optimal weighting, however, we find that the estimates combining the  $x$  and  $y$  instruments are more precise, according to the asymptotic standard deviation estimates, than those based on either the  $x$  or the  $y$  instruments.

*Table 2* contains the overall GMM estimates obtained from the *differenced equation*, utilizing, respectively, all essential OC's (row 1) and only the OC's constructed from one-period differences (row 2). We find that adding the  $T-2 = 6$  essential two-period OC's to the  $(T-1)(T-2) = 42$  one-period OC's, may significantly affect the result (cf. Proposition 1). This holds both when the IV's are level  $x$ 's (columns 1 and 3) and level  $y$ 's (columns 2 and 4). Sargan-Hansen orthogonality test statistics, which are asymptotically distributed as  $\chi^2$  with a number of degrees of freedom equal to the number of OC's imposed (cf. Hansen (1982), Newey (1985), Arellano and Bond (1991), and Appendix B), corresponding to the asymptotically efficient estimates in columns 3 and 4, are also reported. These statistics indicate non-rejection of the full set of OC's when using the  $x$ 's as IV's in panel A [cf. (2.18)] and the  $y$ 's as IV's in panel B [cf. (2.18)] – *i.e.*, the output variable in both cases – with  $P$  value equal to 0.295 and 0.111, respectively. The full set of OC's when using the  $y$ 's as IV's in panel A and the  $x$ 's as IV's in panel B – *i.e.*, the material input variable in both cases – is however rejected. All the results in Table 2 uniformly indicate a marginal input elasticity of materials larger than one. The estimates in panel A are, however, lower than the (inconsistent) estimate obtained by running OLS regression

---

<sup>26</sup>All numerical calculations reported below are performed by means of procedures constructed by the author in the GAUSS software.

on differences (without intercept), which is 1.1608, and the estimates in panel B are higher than the (inconsistent) estimate obtained by running reverse OLS regression on differences (without intercept), which is as low as 0.5894.

*Table 3* contains period specific GMM estimates obtained from the *level equation* for 3 selected periods among the  $T = 8$ , using as IV's, respectively,  $x$  differences (columns 1 and 3) and  $y$  differences (columns 2 and 4). The results in panels A1 and B1 are based on the OC's corresponding to (2.20) and (2.21), with untransformed levels, for the 'original' and the 'reverse regression', respectively. All these estimates are very precise, the former clustering around 0.92 – 0.94, the latter around 1.06 – 1.08, regardless of which IV set is used. Using OC's corresponding to (4.19) and (4.20), *i.e.*, deducting year means from the level variables, we get the estimates given in panels A2 and B2. They are more sensitive to the choice of estimator and IV set and have standard deviation estimates which are substantially larger.

Finally, *Table 4* contains the overall GMM estimates obtained from the *level equation*, using all essential OC's in combination with the untransformed observations (row 1) or with the observations measured from their year means (row 2). The four sets of estimates differ substantially. The orthogonality test statistics (columns 5 and 6) give conclusions similar to those for the differenced equation in Table 2: Non-rejection of the full set of OC's when using the  $x$ 's as IV's in panel A [cf. (2.20)] and the  $y$ 's as IV's in panel B [cf. (2.21)] – *i.e.*, the output variable in both cases – and rejection when using the  $y$ 's as IV's in panel A and the  $x$ 's as IV's in panel B – *i.e.*, the material input variable in both cases. Note that the set of moment conditions we test here is larger than in Table 2, since it also includes assumption (D2), time invariance of the covariance between the firm specific effect  $\alpha_i$  and the latent regressor  $\xi_{it}$ . These estimates, unlike those for the differenced equation in Table 2, however, do not uniformly give marginal input elasticity estimates of materials exceeding one. Using level observations measured from year means, and hence relaxing mean stationarity of the latent regressor, we get estimates exceeding one, while using untransformed observations, and hence imposing mean stationarity, we get estimates smaller than one. For comparison, we refer the (inconsistent) OLS estimate obtained from the complete data set when deducting global mean and period means from the observations, which are 1.1450 and 1.1476, respectively, for the 'original regression' and 0.7888 and 0.7864, respectively, for the 'reverse regression'. The corresponding (consistent) 'between period' estimates, operating on the  $T = 8$  year means only [cf. (3.6)] are, however, very close to one, 1.0027 and 0.9859, respectively.

Tentatively, we then conclude from these examples that GMM estimates of the input elasticity of materials tend to be larger than one, regardless of the normalization of the equation, if we use either the equation in differences with IV's in levels or the equation in levels, with observations measured from their year means, and with IV's in differences. If we use the strict equation in levels and with IV's in differences, the GMM estimates tend to be smaller than one.

## 6 Concluding remarks

In this paper, we have demonstrated that several, rather simple, GMM estimators which may “solve” jointly the heterogeneity problem and the measurement error problem in panel data, exist. These problems may be “unsolvable” when only pure (single or repeated) cross section data or pure time series data are available. The estimators considered use either (A) equations in differences with level values as instruments, or (B) equations in levels with differenced values as instruments. In both cases, the differences may be taken over one period or more. Estimators combining approaches (A) and (B) are not considered. Ideas similar to the latter – although exploiting one-period differences only – have been followed recently in GMM estimation of AR(1) equations for panel data without measurement errors, by Ahn and Schmidt (1995), Arellano and Bover (1995), and Blundell and Bond (1995), *inter alia* for increasing estimation efficiency by utilizing the initial conditions on the process.

Even if the equation is static, instruments constructed from the regressors ( $x$ 's) as well as from the regressands ( $y$ 's) may be of interest. Our empirical illustrations, using input and output data for firms in a single regressor case, indicate that for both normalizations of the equation, GMM estimates using  $y$  instruments tend to exceed those using  $x$  instruments. GMM estimates combining both instrument sets in an ‘optimal’ way, are more precise than those using either one of them. Although a substantial number of orthogonality conditions constructed from differences taken over two periods or more are redundant, our numerical examples suggest that adding two-period difference orthogonality conditions to the one-period conditions in the GMM algorithm may significantly affect the result. From this we conclude that utilizing the information contained in two-period differences may increase estimation efficiency. Using levels as instruments for differences or *vice versa* as a general estimation strategy within a GMM framework, however, may raise problems related to ‘weak instruments’. Finding operational ways of identifying such instruments and

reducing their potential damage in terms of inefficiency, or even inconsistency, are challenges for future research.

Finally, we find that GMM estimates based on the level equation are more precise (in terms of asymptotic standard deviations) than those based on differenced equations. Deducting period means from levels, to compensate for non-stationarity of the latent regressor, give estimates for the level equation which are less precise and more sensitive to the choice of instrument set than those operating on untransformed levels.

Table 1:

PERIOD SPECIFIC GMM ESTIMATES OF  $\beta$ .  $N = 215$ ,  $T = 8$   
DIFFERENCED EQUATION  $(t, \theta)$ . INSTRUMENTS IN LEVELS.  
Standard deviation estimates in parenthesis.

*A.  $y = \ln(\text{material input})$ ,  $x = \ln(\text{output})$ .*

$t$	$\theta$	$\hat{\beta}_{x(t\theta)}$	$\hat{\beta}_{y(t\theta)}$	$\hat{\beta}_{xy(t\theta)}$	$\tilde{\beta}_{x(t\theta)}$	$\tilde{\beta}_{y(t\theta)}$	$\tilde{\beta}_{xy(t\theta)}$
2	1	1.1382	1.1555	1.1401	1.1693	1.1436	1.1577
		(0.1007)	0.1060	0.1082	0.0907	0.0872	0.0682)
5	4	1.0447	1.1555	1.0449	1.0006	1.1127	0.9993
		(0.0798)	0.0982	0.0812	0.0747	0.0941	0.0717)
8	7	0.9875	0.9352	1.0845	0.9610	0.9867	1.1035
		(0.1754)	0.1488	0.1372	0.1666	0.1348	0.1116)
5	1	1.0845	1.1563	1.0754	1.0782	1.0639	1.0579
		(0.0348)	0.0499	0.0327	0.0316	0.0293	0.0234)
8	4	1.0910	1.3334	1.0981	1.0987	1.3564	1.0563
		(0.0698)	0.0823	0.0673	0.0647	0.0766	0.0566)
8	1	1.0728	1.2175	1.0812	1.0566	1.0851	1.0332
		(0.0440)	0.0676	0.0446	0.0379	0.0380	0.0277)

*B.  $y = \ln(\text{output})$ ,  $x = \ln(\text{material input})$ .*

$t$	$\theta$	$\hat{\beta}_{x(t\theta)}$	$\hat{\beta}_{y(t\theta)}$	$\hat{\beta}_{xy(t\theta)}$	$\tilde{\beta}_{x(t\theta)}$	$\tilde{\beta}_{y(t\theta)}$	$\tilde{\beta}_{xy(t\theta)}$
2	1	0.8087	0.8256	0.7665	0.8485	0.8357	0.8354
		(0.0652)	0.0678	0.0639	0.0619	0.0642	0.0496)
5	4	0.8164	0.9165	0.8338	0.8301	0.9546	0.8782
		(0.0713)	0.0732	0.0743	0.0691	0.0696	0.0632)
8	7	0.8966	0.8796	0.7230	0.9308	0.8330	0.7694
		(0.1187)	0.1420	0.0981	0.1114	0.1343	0.0769)
5	1	0.8274	0.9134	0.8378	0.9297	0.9216	0.9312
		(0.0452)	0.0306	0.0419	0.0238	0.0266	0.0201)
8	4	0.7438	0.9021	0.7756	0.7265	0.8663	0.7826
		(0.0457)	0.0565	0.0448	0.0418	0.0517	0.0407)
8	1	0.7897	0.9215	0.7942	0.9074	0.9399	0.9210
		(0.0514)	0.0388	0.0492	0.0291	0.0330	0.0258)

Table 2:

GMM ESTIMATES OF  $\beta$  BASED ON DIFFERENCED EQUATION.  $N = 215$ ,  $T = 8$   
 INSTRUMENTS IN LEVELS. ALL ESSENTIAL ORTHOGONALITY CONDITIONS

Standard deviation estimates in parenthesis.

A.  $y = \ln(\text{material input})$ ,  $x = \ln(\text{output})$ .

	$\hat{\beta}_{Dx}$	$\hat{\beta}_{Dy}$	$\tilde{\beta}_{Dx}$	$\tilde{\beta}_{Dy}$
<i>Estimates</i>				
All OC's	1.0821 (0.0331)	1.1275 (0.0346)	1.0546 (0.0173)	1.0825 (0.0169)
One period OC's only	1.1215 (0.0366)	1.2398 (0.0496)	1.0690 (0.0184)	1.1164 (0.0219)
<i>Orthogonality tests</i>				
All OC's			51.7111	70.3871
<i>P</i> value			0.2950	0.0152
One period OC's only			56.4225	91.7643
<i>P</i> value			0.0550	0.0000

B.  $y = \ln(\text{output})$ ,  $x = \ln(\text{material input})$ .

	$\hat{\beta}_{Dx}$	$\hat{\beta}_{Dy}$	$\tilde{\beta}_{Dx}$	$\tilde{\beta}_{Dy}$
<i>Estimates</i>				
All OC's	0.8404 (0.0283)	0.8931 (0.0283)	0.8917 (0.0143)	0.9244 (0.0148)
One period OC's only	0.7451 (0.0361)	0.8475 (0.0305)	0.8531 (0.0175)	0.9042 (0.0155)
<i>Orthogonality tests</i>				
All OC's			86.5531	59.0766
<i>P</i> value			0.0004	0.1112
One period OC's only			112.4654	71.3243
<i>P</i> value			0.0000	0.0023

Table 3:

PERIOD SPECIFIC GMM ESTIMATES OF  $\beta$ .  $N = 215$ ,  $T = 8$   
 LEVEL EQUATION ( $t$ ). INSTRUMENTS IN DIFFERENCES.  
 Standard deviation estimates in parenthesis.

A1.  $y = \ln(\text{material input})$ ,  $x = \ln(\text{output})$ . No transformation

$t$	$\hat{\beta}_{x(t)}$	$\hat{\beta}_{y(t)}$	$\tilde{\beta}_{x(t)}$	$\tilde{\beta}_{y(t)}$
2	0.9262 (0.0038)	0.9203 (0.0079)	0.9274 (0.0034)	0.9215 (0.0061)
5	0.9346 (0.0032)	0.9426 (0.0064)	0.9349 (0.0030)	0.9403 (0.0057)
8	0.9306 (0.0034)	0.9329 (0.0051)	0.9325 (0.0032)	0.9378 (0.0047)

A2.  $y = \ln(\text{material input})$ ,  $x = \ln(\text{output})$ . Year mean deduction

$t$	$\hat{\beta}_{x(t)}$	$\hat{\beta}_{y(t)}$	$\tilde{\beta}_{x(t)}$	$\tilde{\beta}_{y(t)}$
2	1.0129 (0.0854)	1.2082 (0.1301)	0.9936 (0.0698)	1.0964 (0.1073)
5	0.9797 (0.1283)	1.1843 (0.1963)	0.9294 (0.1213)	1.1128 (0.1481)
8	1.0193 (0.1799)	1.3767 (0.1654)	1.0471 (0.1661)	1.1849 (0.1452)

B1.  $y = \ln(\text{output})$ ,  $x = \ln(\text{material input})$ . No transformation

$t$	$\hat{\beta}_{x(t)}$	$\hat{\beta}_{y(t)}$	$\tilde{\beta}_{x(t)}$	$\tilde{\beta}_{y(t)}$
2	1.0862 (0.0093)	1.0796 (0.0044)	1.0849 (0.0072)	1.0782 (0.0039)
5	1.0605 (0.0072)	1.0698 (0.0036)	1.0632 (0.0064)	1.0696 (0.0034)
8	1.0712 (0.0058)	1.0745 (0.0039)	1.0660 (0.0054)	1.0723 (0.0037)

B2.  $y = \ln(\text{output})$ ,  $x = \ln(\text{material input})$ . Year mean deduction

$t$	$\hat{\beta}_{x(t)}$	$\hat{\beta}_{y(t)}$	$\tilde{\beta}_{x(t)}$	$\tilde{\beta}_{y(t)}$
2	0.7815 (0.0852)	0.9600 (0.0740)	0.8236 (0.0761)	0.9740 (0.0652)
5	0.7211 (0.1106)	0.9263 (0.1144)	0.7668 (0.0939)	0.9765 (0.1051)
8	0.6109 (0.1060)	0.8401 (0.1314)	0.6941 (0.1010)	0.8634 (0.1294)

Table 4:

GMM ESTIMATES OF  $\beta$  BASED ON LEVEL EQUATION.  $N = 215$ ,  $T = 8$   
 INSTRUMENTS IN DIFFERENCES. ALL ESSENTIAL ORTHOGONALITY CONDITIONS

Col. 1 – 4: Standard deviation estimates in parenthesis.

Col. 5 – 6:  $\chi^2$  orthogonality test statistics,  $P$  values in parenthesis.

A.  $y = \ln(\text{material input})$ ,  $x = \ln(\text{output})$ .

	$\hat{\beta}_{Lx}$	$\hat{\beta}_{Ly}$	$\tilde{\beta}_{Lx}$	$\tilde{\beta}_{Ly}$	$\chi^2(\tilde{\beta}_{Lx})$	$\chi^2(\tilde{\beta}_{Ly})$
No transformation	0.9308	0.9325	0.9351	0.9404	56.7641	81.4882
	(0.0031)	(0.0052)	(0.0024)	(0.0022)	(0.1557)	(0.0013)
Year mean deduction	1.0219	1.2148	1.0739	1.1749	54.6578	73.5559
	(0.0644)	(0.1202)	(0.0289)	(0.0316)	(0.2065)	(0.0079)

B.  $y = \ln(\text{output})$ ,  $x = \ln(\text{material input})$ .

	$\hat{\beta}_{Lx}$	$\hat{\beta}_{Ly}$	$\tilde{\beta}_{Lx}$	$\tilde{\beta}_{Ly}$	$\chi^2(\tilde{\beta}_{Lx})$	$\chi^2(\tilde{\beta}_{Ly})$
No transformation	1.0718	1.0743	1.0628	1.0690	80.6409	56.6942
	(0.0060)	(0.0035)	(0.0025)	(0.0028)	(0.0016)	(0.1572)
Year mean deduction	0.7345	0.9392	0.7428	0.8834	64.4824	52.4157
	(0.0730)	(0.0559)	(0.0225)	(0.0242)	(0.0460)	(0.2720)



# Appendix

In this appendix, we elaborate (A) the procedures for estimating asymptotic covariance matrices of the GMM estimators and (B) the orthogonality tests.

## A. Estimation of covariance matrices

All models in the main text, with suitable interpretations of  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\boldsymbol{\epsilon}$ , and  $\boldsymbol{\Omega}$ , have the form:

$$(A.1) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \mathbb{E}(\mathbf{Z}'\boldsymbol{\epsilon}) = \mathbf{0}, \quad \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{\Omega},$$

where  $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_N)'$ ,  $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_N)'$ ,  $\mathbf{Z} = (\mathbf{Z}'_1, \dots, \mathbf{Z}'_N)'$ , and  $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}'_1, \dots, \boldsymbol{\epsilon}'_N)'$ ,  $\mathbf{Z}_i$  being the IV matrix of  $\mathbf{X}_i$ . The two generic GMM estimators considered are

$$(A.2) \quad \hat{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{P}_Z\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{P}_Z\mathbf{y}], \quad \mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}',$$

$$(A.3) \quad \tilde{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{P}_Z(\boldsymbol{\Omega})\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{P}_Z(\boldsymbol{\Omega})\mathbf{y}], \quad \mathbf{P}_Z(\boldsymbol{\Omega}) = \mathbf{Z}(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z})^{-1}\mathbf{Z}'.$$

Let the residual vector calculated from  $\hat{\boldsymbol{\beta}}$  be  $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ , and use the notation

$$\begin{aligned} \mathbf{S}_{XZ} &= \frac{\mathbf{X}'\mathbf{Z}}{N}, \quad \mathbf{S}_{ZX} = \frac{\mathbf{Z}'\mathbf{X}}{N}, \quad \mathbf{S}_{ZZ} = \frac{\mathbf{Z}'\mathbf{Z}}{N}, \quad \mathbf{S}_{\boldsymbol{\epsilon}Z} = \frac{\boldsymbol{\epsilon}'\mathbf{Z}}{N}, \quad \mathbf{S}_{Z\boldsymbol{\epsilon}} = \frac{\mathbf{Z}'\boldsymbol{\epsilon}}{N}, \\ \mathbf{S}_{Z\boldsymbol{\Omega}Z} &= \frac{\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}}{N}, \quad \mathbf{S}_{Z\boldsymbol{\epsilon}\boldsymbol{\epsilon}Z} = \frac{\mathbf{Z}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{Z}}{N}, \quad \mathbf{S}_{Z\hat{\boldsymbol{\epsilon}}Z} = \frac{\mathbf{Z}'\hat{\boldsymbol{\epsilon}}\hat{\boldsymbol{\epsilon}}'\mathbf{Z}}{N}. \end{aligned}$$

Inserting for  $\mathbf{y}$  from (A.1) in (A.2) and (A.3), we get

$$\begin{aligned} \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{N}[\mathbf{X}'\mathbf{P}_Z\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{P}_Z\boldsymbol{\epsilon}] \\ &= [\mathbf{S}_{XZ}\mathbf{S}_{ZZ}^{-1}\mathbf{S}_{ZX}]^{-1}[\mathbf{S}_{XZ}\mathbf{S}_{ZZ}^{-1}\frac{\mathbf{Z}'\boldsymbol{\epsilon}}{\sqrt{N}}], \\ \sqrt{N}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{N}[\mathbf{X}'\mathbf{P}_Z(\boldsymbol{\Omega})\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{P}_Z(\boldsymbol{\Omega})\boldsymbol{\epsilon}] \\ &= [\mathbf{S}_{XZ}\mathbf{S}_{Z\boldsymbol{\Omega}Z}^{-1}\mathbf{S}_{ZX}]^{-1}[\mathbf{S}_{XZ}\mathbf{S}_{Z\boldsymbol{\Omega}Z}^{-1}\frac{\mathbf{Z}'\boldsymbol{\epsilon}}{\sqrt{N}}], \end{aligned}$$

and hence,

$$\begin{aligned} N(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' &= [\mathbf{S}_{XZ}\mathbf{S}_{ZZ}^{-1}\mathbf{S}_{ZX}]^{-1}[\mathbf{S}_{XZ}\mathbf{S}_{ZZ}^{-1}\mathbf{S}_{Z\boldsymbol{\epsilon}\boldsymbol{\epsilon}Z}\mathbf{S}_{ZZ}^{-1}\mathbf{S}_{ZX}][\mathbf{S}_{XZ}\mathbf{S}_{ZZ}^{-1}\mathbf{S}_{ZX}]^{-1}, \\ N(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})' &= [\mathbf{S}_{XZ}\mathbf{S}_{Z\boldsymbol{\Omega}Z}^{-1}\mathbf{S}_{ZX}]^{-1}[\mathbf{S}_{XZ}\mathbf{S}_{Z\boldsymbol{\epsilon}\boldsymbol{\epsilon}Z}\mathbf{S}_{Z\boldsymbol{\Omega}Z}^{-1}\mathbf{S}_{ZX}][\mathbf{S}_{XZ}\mathbf{S}_{Z\boldsymbol{\Omega}Z}^{-1}\mathbf{S}_{ZX}]^{-1}. \end{aligned}$$

The asymptotic covariance matrices of  $\sqrt{N}\hat{\boldsymbol{\beta}}$  and  $\sqrt{N}\tilde{\boldsymbol{\beta}}$  can then, under suitable regularity conditions, be written as [see Bowden and Turkington (1984, pp. 26, 69)]

$$\begin{aligned} a\mathbf{V}(\sqrt{N}\hat{\boldsymbol{\beta}}) &= \lim \mathbb{E}[N(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] = \text{plim}[N(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'], \\ a\mathbf{V}(\sqrt{N}\tilde{\boldsymbol{\beta}}) &= \lim \mathbb{E}[N(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'] = \text{plim}[N(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})']. \end{aligned}$$

Since  $\mathbf{S}_{Z\boldsymbol{\epsilon}\boldsymbol{\epsilon}Z}$  and  $\mathbf{S}_{Z\boldsymbol{\Omega}Z}$  coincide asymptotically, we get, using bars to denote plims,

$$(A.4) \quad a\mathbf{V}(\sqrt{N}\hat{\boldsymbol{\beta}}) = [\bar{\mathbf{S}}_{XZ}\bar{\mathbf{S}}_{ZZ}^{-1}\bar{\mathbf{S}}_{ZX}]^{-1}[\bar{\mathbf{S}}_{XZ}\bar{\mathbf{S}}_{ZZ}^{-1}\bar{\mathbf{S}}_{Z\boldsymbol{\Omega}Z}\bar{\mathbf{S}}_{ZZ}^{-1}\bar{\mathbf{S}}_{ZX}][\bar{\mathbf{S}}_{XZ}\bar{\mathbf{S}}_{ZZ}^{-1}\bar{\mathbf{S}}_{ZX}]^{-1},$$

$$(A.5) \quad a\mathbf{V}(\sqrt{N}\tilde{\boldsymbol{\beta}}) = [\bar{\mathbf{S}}_{XZ}\bar{\mathbf{S}}_{Z\boldsymbol{\Omega}Z}^{-1}\bar{\mathbf{S}}_{ZX}]^{-1}.$$

Replacing the plims  $\bar{\mathbf{S}}_{XZ}$ ,  $\bar{\mathbf{S}}_{ZX}$ ,  $\bar{\mathbf{S}}_{ZZ}$  and  $\bar{\mathbf{S}}_{Z\Omega Z}$  by their sample counterparts,  $\mathbf{S}_{XZ}$ ,  $\mathbf{S}_{ZX}$ ,  $\mathbf{S}_{ZZ}$  and  $\mathbf{S}_{Z\hat{\epsilon}\hat{\epsilon}Z}$  and dividing by  $N$ , we get from (A.4) and (A.5) the following estimators of the asymptotic covariance matrices of  $\hat{\boldsymbol{\beta}}$  and  $\tilde{\boldsymbol{\beta}}$ :

$$(A.6) \quad \mathbf{V}(\hat{\boldsymbol{\beta}}) = \frac{1}{N} [\mathbf{S}_{XZ} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{ZX}]^{-1} [\mathbf{S}_{XZ} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{Z\hat{\epsilon}\hat{\epsilon}Z} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{ZX}] [\mathbf{S}_{XZ} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{ZX}]^{-1} \\ = [\mathbf{X}' \mathbf{P}_Z \mathbf{X}]^{-1} [\mathbf{X}' \mathbf{P}_Z \hat{\boldsymbol{\epsilon}} \hat{\boldsymbol{\epsilon}}' \mathbf{P}_Z \mathbf{X}] [\mathbf{X}' \mathbf{P}_Z \mathbf{X}]^{-1},$$

$$(A.7) \quad \mathbf{V}(\tilde{\boldsymbol{\beta}}) = \frac{1}{N} [\mathbf{S}_{XZ} \mathbf{S}_{Z\hat{\epsilon}\hat{\epsilon}Z}^{-1} \mathbf{S}_{ZX}]^{-1} \\ = [\mathbf{X}' \mathbf{Z} (\mathbf{Z}' \hat{\boldsymbol{\epsilon}} \hat{\boldsymbol{\epsilon}}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{X}]^{-1} = [\mathbf{X}' \mathbf{P}_Z (\hat{\boldsymbol{\epsilon}} \hat{\boldsymbol{\epsilon}}') \mathbf{X}]^{-1}.$$

These are the generic expressions which we use for estimating variances and covariances of the GMM estimators considered.

When calculating  $\tilde{\boldsymbol{\beta}}$  from (A.3) in practice, we replace  $\mathbf{P}_Z(\boldsymbol{\Omega})$  by  $\mathbf{P}_Z(\hat{\boldsymbol{\epsilon}} \hat{\boldsymbol{\epsilon}}') = \mathbf{Z} (\mathbf{Z}' \hat{\boldsymbol{\epsilon}} \hat{\boldsymbol{\epsilon}}' \mathbf{Z})^{-1} \mathbf{Z}'$  [see White (1982, 1984)].

## B. Orthogonality testing

For testing the (vector) orthogonality condition  $\mathbf{E}(\mathbf{Z}'_i \boldsymbol{\epsilon}_i) = \mathbf{0}$ , we use the Sargan-Hansen statistic [cf. Hansen (1982), Newey (1985), and Arellano and Bond (1991)]

$$(A.8) \quad H = N [\mathbf{S}_{\hat{\epsilon}Z} \mathbf{S}_{Z\hat{\epsilon}\hat{\epsilon}Z}^{-1} \mathbf{S}_{Z\hat{\epsilon}}]^{-1} = [\hat{\boldsymbol{\epsilon}}' \mathbf{Z} (\mathbf{Z}' \hat{\boldsymbol{\epsilon}} \hat{\boldsymbol{\epsilon}}' \mathbf{Z})^{-1} \mathbf{Z}' \hat{\boldsymbol{\epsilon}}]^{-1}.$$

Under the null, it is asymptotically distributed as  $\chi^2$  with a number of degrees of freedom equal to the number of orthogonality conditions, *i.e.*, the dimension of the vector  $\mathbf{Z}'_i \boldsymbol{\epsilon}_i$ .

## References

- Ahn, S.C. and Schmidt, P. (1995): Efficient Estimation of Models for Dynamic Panel Data. *Journal of Econometrics*, 68, 5 – 27.
- Aigner, D.J., Hsiao, C., Kapteyn, A., and Wansbeek, T. (1984): Latent Variable Models in Econometrics. Chapter 23 in *Handbook of Econometrics, Vol. II*, ed. by Z. Griliches and M.D. Intriligator. Amsterdam: North-Holland.
- Anderson, T.W. and Hsiao, C. (1981): Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association*, 76, 598 – 606.
- Anderson, T.W. and Hsiao, C. (1982): Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics*, 18, 47 – 82.
- Arellano, M. and Bond, S. (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 Bover, O. (1995): Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics*, 68, 29 – 51.
- Baltagi, B.H. (1995): *Econometric Analysis of Panel Data*. Chichester: Wiley.
- Bekker, P., Kapteyn, A., and Wansbeek, T. (1987): Consistent Sets of Estimates for Regressions with Correlated or Uncorrelated Measurement Errors in Arbitrary Subsets of All Variables. *Econometrica*, 55, 1223 – 1230.
- Biørn, E. (1992): The Bias of Some Estimators for Panel Data Models with Measurement Errors. *Empirical Economics*, 17, 51 – 66.
- Biørn, E. (1996): Panel Data with Measurement Errors. Chapter 10 in *The Econometrics of Panel Data. Handbook of the Theory with Applications*, ed. by L. Mátyás and P. Sevestre. Dordrecht: Kluwer.
- Biørn, E. and Klette, T.J. (1997a): Variable Differencing and GMM Estimation with Panel Data with Errors-in-Variables. Department of Economics, University of Oslo, Memorandum No. 16/1997.
- Biørn, E. and Klette, T.J. (1997b): Panel Data with Errors-in-Variables: A Note on Essential and Redundant Orthogonality Conditions in GMM-Estimation. Statistics Norway, Discussion Paper No. 190. Forthcoming in *Economics Letters*.
- Blundell, R. and Bond, S. (1995): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. Mimeo, The Institute for Fiscal Studies.
- Bowden, R.J. and Turkington, D.A. (1984): *Instrumental Variables*. Cambridge: Cambridge University Press.
- Breusch, T., Qian, H., Schmidt, P., and Wyhowski, D.J. (1997): Redundancy of Moment Conditions. Forthcoming in *Journal of Econometrics*.
- Chamberlain, G. (1987): Asymptotic Efficiency in Estimation With Conditional Moment Restrictions. *Journal of Econometrics*, 34, 305 – 334.
- Crépon, B. and Mairesse, J. (1996): The Chamberlain Approach. Chapter 14 in *The Econometrics of Panel Data. Handbook of the Theory with Applications*, ed. by L. Mátyás and P. Sevestre. Dordrecht: Kluwer.

- Davidson, R. and MacKinnon, J.G. (1993): *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.
- Erickson, T. (1993): Restricting Regression Slopes in the Errors-in-Variables Model by Bounding the Error Correlation. *Econometrica*, 91, 959 – 969.
- Fuller, W.A. (1987): *Measurement Error Models*. New York: Wiley.
- Griliches, Z. (1986): Economic Data Issues. Chapter 25 in *Handbook of Econometrics, Vol. III*, ed. by Z. Griliches and M.D. Intriligator. Amsterdam: North-Holland.
- Griliches, Z. and Hausman, J.A. (1986): Errors in Variables in Panel Data. *Journal of Econometrics*, 31, 93 – 118.
- Hansen, L.P. (1982): Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50, 1029 – 1054.
- Holtz-Eakin, D., Newey, W., and Rosen, H.S. (1988): Estimating Vector Autoregressions with Panel Data. *Econometrica*, 56, 1371 – 1395.
- Hsiao, C. (1986): *Analysis of Panel Data*. Cambridge: Cambridge University Press.
- Klepper, S. and Leamer, E. (1984): Consistent Sets of Estimates for Regressions with Errors in All Variables. *Econometrica*, 52, 163 – 183.
- McCabe, B. and Tremayne, A. (1993): *Elements of Modern Asymptotic Theory with Statistical Applications*. Manchester: Manchester University Press.
- Nelson, C.R. and Startz, R. (1990): Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, 58, 967 – 976.
- Newey, W.K. (1985): Generalized Method of Moments Specification Testing. *Journal of Econometrics*, 29, 229 – 256.
- Newey, W.K. and D. McFadden (1994): Large Sample Estimation and Hypothesis Testing. Chapter 36 in *Handbook of Econometrics, Vol. IV*, ed. by R.F. Engle and D.L. McFadden. Amsterdam: North-Holland.
- Reiersøl, O. (1950): Identifiability of a Linear Relation Between Variables which are Subject to Error. *Econometrica*, 18, 375 – 389.
- Schmidt, P. (1990): Three-Stage Least Squares with Different Instruments for Different Equations. *Journal of Econometrics*, 43, 389 – 394.
- Sevestre, P. and Trognon, A. (1996): Dynamic Linear Models. Chapter 7 in *The Econometrics of Panel Data. Handbook of the Theory with Applications*, ed. by L. Mátyás and P. Sevestre. Dordrecht: Kluwer.
- Staiger, D. and Stock, J.H. (1996): Asymptotics for GMM Estimators with Weak Instruments. National Bureau of Economic Research, Technical Working Paper No. 198.
- Staiger, D. and Stock, J.H. (1997): Instrumental Variables Regression With Weak Instruments. *Econometrica*, 65, 557 – 586.
- Wansbeek, T.J. (1996): Additional Moment Restrictions in Panel Data Models with Measurement Error. Paper presented at the Econometric Society European Meeting 1996, Istanbul.

- Wansbeek, T.J. and Koning, R.H. (1991): Measurement Error and Panel Data. *Statistica Neerlandica*, 45, 85 – 92.
- White, H. (1982): Instrumental Variables Regression with Independent Observations. *Econometrica*, 50, 483 – 499.
- White, H. (1984): *Asymptotic Theory for Econometricians*. Orlando: Academic Press.
- White, H. (1986): Instrumental Variables Analogs of Generalized Least Squares Estimators. In *Advances in Statistical Analysis and Statistical Computing. Theory and Applications, vol. 1*, ed. by R.S. Mariano, JAI Press, pp. 173 – 227.
- Willassen, Y. (1979): Two Clarifications on the Likelihood Surface in Functional Models. *Journal of Multivariate Analysis*, 9, 138 – 149.
- Willassen, Y. (1998): Deriving Bounds on the Structural Vector when the Measurement Errors are Correlated: An Elaboration of the Frisch/Reiersøl Approach. Department of Economics, University of Oslo, Memorandum No. 6/1998.
- Ziliak, J.P. (1997): Efficient Estimation With Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators. *Journal of Business and Economic Statistics*, 15, 419 – 431.