UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 00/5

A QUASI-DIFFERENCING APPROACH TO DYNAMIC MODELLING FROM A TIME SERIES OF INDEPENDENT CROSS SECTIONS

by Sourafel Girma

March 2000 DP 00/5

ISSN 1360-2438

UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 00/5

A QUASI-DIFFERENCING APPROACH TO DYNAMIC MODELLING FROM A TIME SERIES OF INDEPENDENT CROSS SECTIONS

by Sourafel Girma

Sourafel Girma is Research Associate in the School of Economics, University of Nottingham.

March 2000

Abstract

We motivate and describe a GMM method of estimating linear dynamic models from a time series of independent cross sections. This involves subjecting the model to a quasi-differencing transformation across pairs of individuals that belong to the same group. Desirable features of the model include the fact that: (i) no aggregation is involved, (ii) dynamic response parameters can vary across groups, (iii) the presence of unobserved individual specific heterogeneity is explicitly allowed for, and (iv) the GMM estimators are derived by realistically assuming group size asymptotics. The Monte Carlo experiments conducted to assess the finite sample performances of the proposed estimators provide us with encouragement.

1: Introduction

In this paper we propose a new method of using data from a time series of independent cross-sections (abbreviated to TISICS in what follows) in fitting linear dynamic models such as

$$y_{it} = \alpha \, y_{it-1} + \beta \, x_{it} + f_i + \varepsilon_{it} \;, \tag{1}$$

where i and t index individual units and time periods respectively, f is an individual effect and ε is a disturbance term. The main problem is the fact that such data sets do not track the same group of individuals over time, whereas the econometric models of interest require some information on intra-individual differences. Thus it is not possible to consistently estimate the parameters of interest unless additional information in the shape of instrumental variables is available or further identifying restrictions are imposed on the model.

Deaton (1985) proposes a pseudo-panel approach to circumvent the identification problem caused by the absence of repeated observations on individual units. This involves placing individuals into distinct groups or cohorts, say G_c, for c=1,…, C and working with the expectation of model (1) at each point in time conditional on cohort membership. Thus taking $E\{y_{it} | i(t) \in G_c\}$, we have

$$y^*_{ct} = a\ y^*_{ct-1} + b\ x^*_{ct\,it} + f^*_{ct} + e^*_{ct} \tag{2}$$

where starred superscripts denote the conditional expectations.

Equation (2) is a model of the average behaviour of a cohort, much in the spirit of the 'representative agent' construction. Since cohort averages are effectively the units of analysis and the number of cross sections is often small in practice, one might have to assume asymptotics on C to derive large sample properties of the ensuing econometric estimators. On the other hand as y^*_{ct-1} and f^*_{ct} are correlated by construction, it is desirable to difference the former out of the model. This requires that f^*_{ct} be invariant over time. One way to force this condition is to assume that all individuals within a cohort are homogeneous with respect to the unobserved heterogeneity factors. It will then be possible to apply standard dynamic panel data techniques by treating sample cohort averages as error-ridden observations on typical individuals (Collado, 1997).

Notwithstanding the fact that the pseudo-panel approach offers a framework for making joint use of independent cross sectional information, there are problems with some of its features. Firstly the assertion of intra-cohort homogeneity seems too strong. It is indeed difficult to assume unobserved individual heterogeneity away in micro data. Secondly the aggregation framework might not be desirable. An estimation method that is based on individual level data can make a more efficient use of the available information and potential problems of interpretation arising out of aggregation are likely to be mitigated. Thirdly, the practice of establishing the large sample properties of econometric estimators and test statistics by driving the number of cohorts to infinity is not satisfactory. There is often a physical limit beyond which one cannot increase the number of cohorts. The oft-cited example of date of birth cohorts is a case in point. Finally, and this pertains to dynamic modelling in particular, standard pooling estimators will lead to invalid inference if the response parameters are characterised by cohort-wise heterogeneity. This follows from Pesaran and Smith (1995) who showed that dynamic modelling from pooled heterogeneous panels is not trivial. Given that cohorts should be chosen so that they are as distinct from each other as possible (*cf.* Deaton, 1985), it is important to have a framework which allows for some sort of response heterogeneity.

The above discussion demonstrates that there is clearly scope for studying alternative estimators for dynamic models from TISICS. In a contribution to the literature Moffit (1993) suggests a two-stage least squares approach where the regressor y_{it-1} in (1) is instrumented by the predicted value $\hat{y}_{it-1}$ from a linear projection of the dependent variable on a vector of time-varying and time-invariant variables, say S_{it} and Z_i respectively. To make this operational, one needs knowledge of S_{it-1}. But it is not always clear where this information is going to come from as TISICS data sets do not typically contain the history of the variables for the cross sectional units. Moreover two-stage least squares is potentially problematic in that the fitted instruments (which need to be correlated with y_{it-1}) will also have a non-zero correlation with the individual effect f_i, since the two quantities are statistically related. The only obvious exception is if S_{it} is a function of time alone, hence uncorrelated with the time-invariant f_i. But this will result in instruments that do not have much variation across individuals in the likely case where the number of cross sections is much smaller than the cross sectionals units. As a result the ensuing instrumental variables estimators will be highly inaccurate. So, although Moffit's (1993) paper generalises Deaton (1985) in the sense that the possibility of instrumentation procedures other than grouping is discussed, it does not seem to be too promising due to its rather strong informational requirement (Verbeek, 1992). In this paper we explore the possibility of making inference on model (1) from individual level data, by performing pair-wise quasi-differences across different observed individuals within the same group. This produces the following regression function at each time period and for all i and j : $\mathrm{E}\left\{y_{it} | y_{jt-1}, x_{it}\right\}$. This is potentially estimable since $\mathrm{Cov}\left\{y_{it}, y_{jt-1}\right\}$ can be identified from the data as long there are group-wise cross sectional correlations.

The plan for the rest of the paper is as follows. In section 2 we formally describe the quasi-differenced model along with the maintained assumptions. We identify the maximum number of linearly independent quasi-differences that are available and those will be used as a basis for estimation. In section 3 we generate some population moment restrictions implied by the quasi-differenced model and suggest a generalised method of moments (GMM) estimation framework. The asymptotic properties of the GMM estimators are realistically based on having a large number of individuals per group-time cell. This compares favourably to the Deaton-type estimators in which the number of group/time periods is required to grow without limit. Another important advantage of the quasi-differencing approach is that slope parameter heterogeneity across groups can easily be allowed since the method can be applied on a single group. In section 4 we conduct some Monte Carlo experiments to assess the finite sample performances of the proposed estimators. Section 5 concludes.

2: Model description

2.1: Basic identifying assumptions

We focus on the estimation of the following class of models form TISICS:

$$y_{i(t)t} = \alpha_{i(t)} y^{*}_{i(t)t-1} + \beta^{'}_{i(t)} x_{i(t)t} + f_{i(t)} + \varepsilon_{i(t)t} \qquad (5)$$

t = 2,T ; i(t) = 1,..., N_t .

Here x is a k-dimensional vector of control variables, f represents individual-specific effects and the ε's are time-varying disturbances. The time index in parentheses is used to emphasise the non-panel nature of the data. However when no ambiguity is apparent, we will sometimes drop this index for the sake of notational elegance. It is assumed that we have cross sectional observations for t = 1,..T , with T$\geq$2. Notice that for $y^{*}_{i(t)t-1}$ the time index on the individual does not match to the (proper) time index. Such variables are not observed and they will often be denoted by a star superscript. We will next discuss some of the basic identifying assumptions of the model**.**

H.1: The population is partitioned into a finite set of mutually exclusive and exhaustive groups G_c, c = 1,…,C. Groups may be based on one or more underlying observed variables that are constant over time for all individuals. Without loss of generality, we assume that at each point in time *n* members are drawn at random from G_c, $\forall$ c.

H.2: $\alpha_i = \alpha_j = \alpha_c$ and $\beta_i = \beta_j = \beta_c$ for $\forall\ (i,\ j) \in G_c$.

This assumption restricts the individual level heterogeneity in response parameters by stating that those parameters are identical within a group, but we do the next best thing by allowing for the possibility of group-wise heterogeneity.

H.3: $|\alpha_c| < 1, \forall\ c$.

H.4: $y_{i(t)0} = k_1 f_c + k_2 \varepsilon_{i(t)0}$, where the k's are constants and f_c is a random group specific effect.

The above assumption implies that the initial (unobserved) conditions are random draws from possibly different populations, but share a common component in the specification of a conditional mean. Assumption **H.4** has an important implication: $\mathrm{Cov}\{y_{i(t)t}, y_{j(s)s}\} \neq 0$ for $\forall (i(t), j(s)) \in G_c$ and $\forall\ t, s$. That is, different individuals within the same group exhibit nonzero correlations. The information obtained from this source of correlations will prove useful in identifying the dynamic response parameters. While **H.4** may not be an easily testable restriction, it is not without its justifications. It has been noted in the econometric literature that group structures are a common feature of many cross sectional data sets (e.g., Frees, 1995). Researchers working with TISICS data often assume that each member of the group has some information about the average group value (Blundell *et al*, 1994).

H.5: $x_{i(t)t} = \upsilon_{i(t)} + \lambda' x_{i(t)t-1} + \theta_{ct} + w_{i(t)t}$,

where the eigen vectors of λ lie within the unit circle, $\upsilon_{i(t)}$ is a k-dimensional vector of individual specific drift and θ_{ct} is group-time specific error component which can come about due to some shared characteristics between individuals within a group-time cell. It can be seen that θ_{ct} is a further source of within group correlations. From the above two assumptions, it is clear that intra-group correlation is directly proportional to the variances of f_c and θ_{ct}.

Further assumptions are:

H.6: $\text{Cov}\{f_i, \varepsilon_{ti}\} = 0, \forall\ i \textit{ and } t$.

H.7: f_i, f_c, f_{ct}, θ_{ct} and ε_{it} are all mutually, independently distributed with mean zero and nonzero and finite variances.

The above assumption allows for purely individual-specific heterogeneity, even if individuals within a group share a common effect. This is in contrast to Deaton's (1985) approach where within-group heterogeneity is not explicitly allowed for.

H.8: $\text{Cov}\{x_{i(t)t}, f_{j(s)}\}=0$

This is an innocuous assumption since $f_{j(s)}$ is specific to individual *j(s).*

Making use of the above assumptions, we can show that $\text{Cov}\{y_{i(t)t}, y_{j(s)s}\} \neq 0$ and equal $\forall (i,\ j) \in G_c$. In other words, our assumptions imply an equicorrelation structure within a group-time cell. Intuitively, without equicorrelation there would not be enough independent information to estimate $O(n^2)$ covariances per cell from just n observations. We like to note that in one form or another, the assumption of equicorrelation amongst the units of analysis is prevalent in econometric modelling. For example, the two-way panel estimation relies on equicorrelation between individuals observed during the same time period (e.g., Hsiao, 1986). The covariance stationarity assumption employed in time series models is also a case in point.

2.2: A quasi-differencing approach

We begin by writing (3) in its reduced form representation as[1]

$$y_{i(t)t} = \pi_{t1} x^{*}_{i(t)1} + \ldots + \pi_{t(t-1)} x^{*}_{i(t)t-1} + \pi_{tt} x_{i(t)t} + \pi_{t0} y^{*}_{i(t)0} + v_{i(t)t} \qquad (4)$$

with

$$v_{i(t)t} = \sum_{s=0}^{t-1} \alpha^{s} \varepsilon_{i(t)(t-s)} + \frac{1-\alpha^{t}}{1-\alpha} f_{i(t)}$$

$$\pi_{ts} = \beta' \alpha^{t-s}, \text{ for } 1 \le s \le t,$$

and

$$\pi_{t0} = \alpha^{t}.$$

Quasi-differencing the model as $\{y_{i(t)t} - \alpha\, y_{j(t-1)t-1}; \forall [i(t), j(t-1)]\, and\, \forall t \geq 2\}$, we obtain the following potentially estimable model

$$y_{i(t)t} = \alpha\, y_{j(t-1)t-1} + \beta' x_{i(t)t} + \eta_{ijt} \qquad (6)$$

where

$$\eta_{ijt} = \alpha^{t}\left[y_{i(t)0} - y_{j(t-1)0}\right] + \Delta f_{ijt} + \Delta x_{ijt} + \Delta \varepsilon_{ijt}$$

with

$$\Delta f_{ijt} = \frac{1-\alpha^{t}}{1-\alpha} f_{i(t)} - \frac{\alpha - \alpha^{t}}{1-\alpha} f_{j(t-1)},$$

$$\Delta x_{ijt} = \beta' \sum_{r=0}^{t-2} \alpha^{r+1} \left\{ x^{*}_{i(t)(t-1-r)} - x^{*}_{j(t-1)(t-1-r)} \right\}$$ [2]

and

$$\Delta \varepsilon_{ijt} = \sum_{r=1}^{t-1} \alpha^{r} \left\{ \varepsilon_{i(t)t-r} - \varepsilon_{j(t-1)t-r} \right\} + \varepsilon_{i(t)t}.$$

---

[1] As the estimator proposed in this section deals with a single group , the index on α and β is dropped.

[2] Although $x_{j(t-1)t-1}$ is observed, it is left in the disturbances for reasons that will shortly become apparent.

Let $D_1 = \left[0_{(T-1)\times 1} : I_{T-1}\right]$; $D_2 = \left[I_{T-1} : 0_{(T-1)\times 1}\right]$ and $A_2 = D_2 \otimes l_n \otimes I_n$, where 0_a ; I_a *and* l_a are the null matrix, the identity matrix and the vector of ones of dimension *a* respectively. When all observations within a group are arranged first by individuals and then by time periods, and the transformation matrix $B(\alpha) = A_1 - \alpha A_2$.

is applied to the resulting $nT \times 1$ vector, it can be checked that the pairwise quasi-differenced model of equation (4) will be produced. Since the $n^2(T-1) \times nT$ matrix $B(\alpha)$ has rank equal to $nT-1$, the transformed error covariance matrix will be singular because $r = n^2(T-1) - (nT-1)$ equations in the quasi-differenced model are redundant. Consequently any estimation strategy has to make due allowance for the singularity of the disturbance covariance matrix. There are two main ways of dealing with this (e.g., Judge et al, 1985): (a) the use of appropriate generalised inverse estimators, and (b) the elimination of the redundant equations prior to estimation. Employing generalised inverses does not seem to be a terribly attractive alternative in our framework. This is because it involves a multiple of n^4 terms and with possibly large n, this will pose awkward computational problems. We thus attempt to identify *nT - 1* linearly independent pairwise quasi-differences within each group. To do so we first consider the following set of quasi-differences within each group between individuals with the same ordering in their respective time domains: $L_1 = \left\{ y_{i(t)t} - \alpha y_{i(t-1)t-1} ; \forall i(t), i(t-1)\, and\ t \geq 2 \right\}$.

Note that $i(t) = i(t-1)$ is to be understood that as the case where individual i at time t and individual i at time (t-1) correspond to the same ordering within their respective time domains. It can be verified that L_1 consists of n (T - 1) linearly independent quasi-differenced equations. By augmenting L_1 by the following *n -1* quasi-differences between the first observation at time t = 2 and all observations bar the first one at time t = 1, $L_2 = \{y_{1(2)2} - \alpha y_{j(1)1}; \forall j \geq 2 \in G_{c1}\}$, it can be checked that $L = L_1 \cup L_2$ will produce the desired full set of *nT - 1* linearly independent quasi-differences. As a result L will exhaust all information contained in the pairwise quasi-differenced model. Of course this is not the only set of linearly independent quasi-differences. However since any other such set is a nonsingular transformation of L, the same information will be preserved

Let $W_{ijt} = [y_{j(t-1)t-1}, x'_{i(t)t}]'$ and define $\gamma_0 = [\alpha \ , \beta\ ']'$ as the true parameter vector. When the above transformations are applied to equation (3), the following model will then be available for estimation for $\forall [i(t), j(t-1)] \in S \equiv S_1 \cup S_2$

$$y_{i(t)t} = \alpha\ y_{j(t-1)t-1} + \beta\ ' x_{i(t)t} + \eta_{ijt} \equiv \gamma_0' W_{ijt} + \eta_{ijt}\ , \tag{5}$$

where $S_1 = \{[i(t), i(t-1)]\, and\, t \geq 2\}$; $S_2 = \{\forall [1(2), j(1)],\ j(1) \geq 2)\}$.

3: Inference

3.1 Identification

OLS on model (5) would produce inconsistent estimators because η_{ijt} is correlated with both $y_{j(t-1)t-1}$ and $x_{i(t)t}$. This can easily be checked using the underlying assumptions of the model. Fortunately, these assumptions also imply that the composite error term η^t_{cij} satisfies the following two sets of linear moment conditions:

$$E\{y_{g(t-s(t-s}\eta_{ijt}\} = 0; t = 2,...T; s = 1,...,t-1;\ \ \forall\ g(t-1) \neq j(t-1)$$

and

$$E\{x_{g(t-s)t-s}\eta_{ijt}\} = 0; t = 2,...T; s = 0,...,t-1; \forall g(t-s) \mid g(t-s) \neq i(t)\ or\ g(t-1) = j(t-1).$$

That is past and present values of the dependent and explanatory variables within the same group can be used as instruments. Because of assumptions **H.4** and **H.5,** a nonzero IV-regressor correlation is ensured. Notice that since the correlations in the x's between the different individuals comes from the θ_{ct}'s, leaving the observable $x_{j(t-1)t-1}$ in the composite disturbance term ensures that the latter does not contain a θ term. Also notice that since η_{ijt} contains $y_{i(t)0} - y_{j(t-1)0}$, any $y_{g(t-s)(t-s)}$ will not be correlated with it because of the equicorrelation assumption imposed on the model.

A problem that we face is the possibility of an infinite number of instruments, given that the number of observations per group-time cell is allowed to grow. For example, $X_{i(t-2)t-2}; i = 1,...n$ can be used as instruments for the regressors in the quasi-differenced model. Since standard optimality results for GMM estimators (e.g., Hansen, 1982) are not trivially applicable when the vector of conditioning variables is infinite dimensional, it pays to look for some ways to economise on the number of instruments to be used. One apparent solution might be to average the instruments over cells. In this case there would not be enough internal variation in the instruments set, unless the number of groups grows with *n* at a suitable rate. This is bound to adversely affect the precision of the resulting estimators. In this paper we restrict ourselves to a set of finite orthogonality conditions by the simple, albeit ad hoc, device of using just one orthogonality condition within each subset of the moment restrictions. For example in the Monte Carlo experiments conducted in the next section, the following moment restrictions are chosen to be the basis for estimating γ:

$E\{Z_{ijt}\eta_{ijt}\}=0$ with $Z_{ijt}=\{y_{g(t-1)t-1}, x^{'}_{g(t-1)t-1}, x^{'}_{g(t)t}\}$, where g is made to correspond to j-1[3]. We have to emphasise, however, that this is entirely arbitrary. All that is needed is to assign distinct instruments to each observation.

3.2 GMM inference

GMM estimation proceeds by forming sample analogs of those population moment conditions. The sample analogs are usually sample averages of $Z_{iji}\eta_{ijt}$ (also known as the disturbances of GMM) and the latter are expected to satisfy appropriate laws of large numbers (LLNs) for all admissible points in the parameter space. As far as justifying the use of LLNs is concerned, the simplest case occurs when the random variables under consideration are independently and identically distributed. In our case, however, there is some dependence across observations due to the assumptions we employed and the quasi-differencing transformation of the data. In fact the disturbances of the model exhibit contemporaneous as well as temporal correlations within groups. They are also heterogeneous across time as can be verified by inspecting the structure of η^{t}_{cij}. Alternative approaches to deriving LLNs that are applicable to a variety of non i.i.d environments are proposed by several authors (e.g., White, 1984). These laws of large numbers typically place some restrictions on the dependence, heterogeneity and moments of the data process.

The main problem in the present context is the contemporaneous within group dependence introduced by the transformation function L_2 . One practical solution may be to average all quasi-differences produced by L_2 and assume that as n grows the dependence will become negligible (and accept the risk of any finite sample bias). In the sequel we work with the model produced by L_1 alone, because the large sample

[3] We adopt the convention of putting g = n for j = 1.

argument with a fixed number of groups and time periods is neater and the computations involved are simpler. Thus we have, for i = 1,…,n and t=2,...,T

$$y_{i(t)} = \alpha y_{i(t-1)(t-1)} + \beta' x_{i(t)t} + \eta_{iit} \equiv \gamma_0' W_{iit} + \eta_{iit}$$

and the resulting orthogonality conditions can be expressed as $\mathrm{E}\{Z_{iit}\eta_{iit}\} \equiv \mathrm{E}\{h_{it}\} = 0$, with $Z_{iit} = \{y_{i-1(t-1)t-1}, x'_{i-1(t-1)t-1}, x'_{i-1(t)t}\}$. In the Appendix A we prove that the following proposition holds:

Proposition 1: $\frac{1}{n(T-1)} \sum_{it-2}^{T} \sum_{i=1}^{n} h_{it} \xrightarrow{p} 0$ as $n \to \infty$ with T fixed.

Defining S_{nT} to be a sequence of weight matrices, the GMM estimator of γ_0 can thus be obtained upon solving

$$argmin_{\gamma} \left[\frac{1}{n(T-1)} \sum_{i,t} h_{it} \right]' S_{nT} \left[\frac{1}{n(T-1)} \sum_{i,t} h_{it} \right] .$$

The optimal GMM estimator of γ is obtained by setting the weight matrix in the GMM minimand function to the inverse of $\Omega \equiv Cov\left[\frac{1}{n(T-1)} \sum_{i,t} h_{it} \right]$ and it is given as

$\hat{\gamma} = \left[W' Z \Omega^{-1} Z' W\right]^{-1} \left[W' Z \Omega^{-1} Z' y\right]$, where the data and instruments matrices are constructed in an obvious way. In appendix B we show that the following proposition holds under the assumptions of the model:

Proposition 2: $\frac{1}{\sqrt{n(T-1)}} \sum_{t=2}^{T} \sum_{i}^{n} h_{it} \xrightarrow{d} N(0,\Omega)$ as $n \to \infty$ with T fixed.

Thus under some regularity conditions (*cf.* White 1984), $\hat{\gamma}_c$ will be asymptotically normally distributed as

$$\sqrt{L}\left(\hat{\gamma} - \gamma_0\right) \to N\left(0, V\right), \tag{6}$$

with $V = \left[\left(\frac{1}{n(T-1)} Z'W \right)' \Omega^{-1} \left(\frac{1}{n(T-1)} Z'W \right) \right]^{-1}$.

Obviously $\hat{\gamma}$ is not feasible since Ω is unknown, but the latter can be replaced by a consistent estimator without affecting the asymptotic properties of $\hat{\gamma}$ (*cf.* Hansen,1982). A Newey-West (1987) type estimator for Ω can be constructed nonparametrically as $\hat{\Omega} = \frac{1}{n(T-1)} \left[\sum_{i,t} h_{it} h'_{it} + \left(\sum_{i,t} h_{it} h'_{i(t-1)} + h_{i(t-1)} h'_{it} \right) \right]$.

An advantage of this approach is that it can be implemented with only two cross sections. And when we have more than one group and the parameters of the model are heterogeneous across these groups and the equation errors are not related, it can be implemented separately for each group without loss of efficiency. If the groups share a common parameter, however, a joint estimation can enhance efficiency. For example, with C $<\infty$ groups and only α varying across groups, the model can be estimated within a system of instrumental variables method by stacking the observation by groups.

<u>3.3: A simpler estimator</u>

For some M in each time period, randomly divide observations within a group into M groups. Let $\bar{y}_{mt}$ be the sample mean across all observations in subgroup m, m=1,M. Now consider the quasi-differenced equations: $\{y_{it} - \alpha \bar{y}_{mt-1}\}$ which gives rise to the following estimable model: $y_{it} = \alpha \, \bar{y}_{mt-1} + \beta' x_{it} + \bar{\eta}_{imt}$. One can then implement the GMM estimation strategy of this paper by using $\left[\bar{y}_{(m-1)t-1} \; \bar{x}_{(m-1)t} \; \bar{x}_{(m-1)t-1} \right]$ as instrumental variables[4]. The advantage of this approach is that for fixed M, the asymptotic of the resulting estimator is much simpler since for *sufficiently large* $\frac{n}{M}$ the disturbance of the quasi-differenced equation are asymptotically uncorrelated. As a result GMM can be routinely implemented. The disadvantage of this approach is that using instruments based on sub-group averages might throw away information contained in the separately lagged individual instruments. In Section 4 we assess the relative merits of these alternative estimators by conducting some simulation experiments.

3.4: Some extensions to the basic model

The study of the effects of observed time-invariant variables such as sex, region and other socio-economic background variables (assuming that they are not used to define groups) could be important in practice. To see how such variables fit into the quasi-differencing framework, define by W the vector of observed time-invariant regressors and consider the following model

$$y_{i(t)t} = \alpha y^{*}_{i(t)t-1} + \varphi W_{i(t)} + \varepsilon_{i(t)t} . \qquad (7)$$

Pairwise quasi-differencing equation (7) would yield

$$y_{i(t)t} = \alpha y_{j(t-1)t-1} + \rho_1 W_{i(t)} + \rho_2 W_{j(t-1)} + \theta^{t}_{ij} \qquad (8)$$

where $\theta^{t}_{ij} = \alpha^{t}\left[y_{0i} - y_{oj}\right] + \Delta f^{t}_{ij} + \Delta\varepsilon^{t}_{ij}$; $\rho_1 = \frac{1-\alpha^{t}}{1-\alpha}\varphi$ and $\rho_2 = -\frac{\alpha - \alpha^{t}}{1-\alpha}\varphi$.

---

[4] As an anonymous referee noted , the pairwise quasi-differencing estimators proposed in this paper can be seen as the limiting case of this simpler estimator as M approaches the sample size.

The identification of ϑ from equation (16) would be greatly simplified if $W_{i(t)} = W_{j(t-1)}$, that is if individuals *i(t) and j(t-1)* exhibit the same time-invariant characteristics. Equation (8) then simplifies to $y_{ti} = \alpha\, y_{j(t-1)t-1} + \rho\, W_{i(t)} + \theta^{t}_{cij}$

Thus a suitable choice of pairs for differencing would ease the estimation of ϑ, a parameter which is not identifiable in first-differenced dynamic (pseudo) panel models.

The model we have considered in this paper can also be easily extended to include lagged regressors of higher orders. For example, to estimate the pure AR(2) model

$y_{t\ t} = \ y_{t\ t-} + \alpha\ y_{t\ t-} + \varepsilon$

$$\{y_{i(t)t} - \alpha_1 y_{j(t-1)(t-1)} - \alpha_2 y_{k(t-2)(t-2)}; \forall i(t), j(t-1) \ \textit{and}\ k(t-2) \in G_c\}.$$

4: Monte Carlo Simulations

To get a feel for how the GMM estimators based on the pairwise quasi-differences (PQD) and the simpler average quasi-differences (AQD) perform in practice, we conducted some Monte Carlo simulations. We are mainly concerned with the consistency of the estimators and their respective standard errors. We examine two combinations of parameters, (α,β)={(0.8, -0.5), (0.3,0.8)}, to explore the possible sensitivity of the results to parameter values. Since an important factor affecting the behaviour of the estimators is the magnitude of IV-regressor correlation, we consider alternative data generating mechanism in which 25%, 50% and 75% of the total variability in the (scalar) exogenous variable and the initial values is attributed to a common group component. In each of these cases, the individual-specific variance is taken to be unity. Fixing the number of groups and cross sections to 1 and 5 respectively, we experimented with three different group-time cell sizes, 100, 200 and 400, which we think are not unrealistic. For each of the resulting 18 configurations, one hundred Monte Carlo experiments are conducted. For the AQD estimators the observations are randomly assigned to 10 subgroups.

Tables 1 and 2 summarise the results from the experiments. The means and standard deviations are computed over the 100 replications and the figures in parentheses represent the percentage biases from the true values. The sample means of the asymptotic standard errors calculated from the GMM formula are also reported, with the percentage deviation of these asymptotic values from the respective across replication standard deviations of estimated values given in parentheses.

A noteworthy result from the experiments is that the finite sample bias of the GMM estimator is more pronounced when the relative variance of the group effect (say ζ) is small. For example in Table 1, the bias in the estimator for α (β) decreases from 3.6% (3.8 %) to 1.5% (1.8 %) when n = 100 as ζ increases from 25% to 50%. But, it is encouraging to note that this bias seems to be tolerable for all sample sizes.

When it comes to the efficiency of the GMM estimators as measured by the across-simulation standard deviations, the role of ζ is more crucial. For given sample size, efficiency improves dramatically as ζ increases. For example, in Table 2 the standard deviations of the estimators is divided by a factor of 3 as ζ doubles to 50%, for n = 200. A marked gain in efficiency is also observed as ζ increases to 75%.

A worrying aspect of the results is the absolute magnitude of the standard deviations (and asymptotic standard errors) when ζ is small, irrespective of the sample sizes. Assuming that the estimators will be deemed precise enough if the margin of error at 5% level is within 20% for the true parameter value, the standard deviation should be less than a tenth of the corresponding parameter. Using this rough guide most estimates based on n = 100 and 200 are disappointing when $\zeta = 25\%$. In general good precision is obtained when ζ is set to 75%, although some reasonably precise estimates emerge with $\zeta = 50\%$.

Another important issue to be addressed is whether the asymptotic standard errors are consistently estimated. The only clue to this comes from a comparison with the simulated standard deviations of the estimators, which themselves are subject to error. Following the argument of Crepon and Mairesse, (1997) we conclude that the estimated asymptotic standard errors and the simulated standard deviations are not significantly different at the 5% level as they are mostly within 20% of each other. We thus find it encouraging that the nonparametric covariance matrix estimator performs quite well in small samples.

The above discussion has highlighted the desirability of having a reasonably high level of relative variation in the group-specific component (and hence a high degree of correlation between individuals within the same group) for the success of the quasi-differencing method proposed in this paper. This is not a surprising result in view of recent work which emphasises the pitfalls of IV-type estimators when the instrument-regressor correlation is low (Staiger and Stock, 1997). So an important implication for applied work is to routinely examine the correlations between the regressors and the instrumental variables candidates and exercise caution when these values are low.

The small sample performance of the simpler AQD proposed in Section 3.3 proves to be disappointing, except when the sample size is quite large (n=400) and most of the variation in the data comes from the group specific effect ($\zeta = 75\%$). A large number of observation per sub-group is needed to guarantee that the AQD disturbances are not correlated; while the consequences of throwing some individual specific information following the aggregation procedure is not serious if only a small part of the sample variation is due to individual-specific effects.

5: Concluding remarks

In this paper we motivated and described a new method of estimating linear dynamic models from a time series of independent cross sections. Some of the desirable features of the method include the fact that no aggregation is involved; the response parameters can freely vary across groups; the presence of unobserved individual specific heterogeneity is explicitly allowed for and the large sample results are derived by realistically assuming group size asymptotics. Moreover the Monte Carlo experiments conducted to assess the finite sample performances of the proposed estimators provide us with some encouragement. For these reasons, we think that it can be considered as a feasible alternative in applied work.

Appendix A: Proof of Consistency

Proposition 1: $\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}h_{it}\xrightarrow{p}E(h_{it})=0$ as $n\to\infty$ with T fixed.

For the sake of notational elegance we pretend that t runs from 1 to T and for ease of exposition we assume that *h* is scalar. We want to show that the sequence of random vector $\{h_{it}$; t = 1,…,T; n= 1,2,..} satisfies a Weak Law of Large Numbers (WLLN). Using the assumptions of the model and further assuming that the various fourth order moments exist and are finite, it can be established that the covariance of the GMM disturbances have the following general structure:

$$E[h_{it}h_{jt}]=\begin{cases} f_1(t)<\infty & if\ i=j\ and\ t=s \\ f_2\ (t,s)<\infty & if\ i=j\ and\,|t-s|=1 \\ f_3\ (t,s)<\infty & if\ i=j-1\ and\,|t-s|=1 \\ 0 & else. \end{cases}$$

Here $f_1(t)$ denotes the variance terms which is a function of time ; $f_2(t,s)$ represents covariance terms between two individuals which correspond to the same ordering within adjacent time periods, and $f_3(t,s)$ is the result of using the stochastic $y_{i-1(t-1)t-1}$ as an instrument. Thus we have a zero-mean random variable which is heterogeneous across time but whose covariance structure is independent of the individual index.

Now re-write $\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}h_{it}$ as $\frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}h_{i1}+\frac{1}{n}\sum_{i=1}^{n}h_{i2}+...+\frac{1}{n}\sum_{i=1}^{n}h_{iT}\right\}$ and consider the tth term $\frac{1}{n}\sum_{i=1}^{n}h_{it}$. Since $E\left(\frac{1}{n}\sum_{i=1}^{n}h_{it}\right)=0$ and by Chebyshev inequality, for ε >0

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}h_{it}\right|>\varepsilon\right)\leq\frac{Var\left[\frac{1}{n}\sum_{i=1}^{n}h_{it}\right]}{\varepsilon^2}\ .$$

Now, since $E(h_{it}h_{jt})=0$

$$Var\left[\frac{1}{n}\sum_{i=1}^{n} h_{it}\right] = \frac{1}{n^2}\left[\sum_{i=1}^{n} Var(h_{it})\right] = \frac{1}{n^2}\sum_{i=1}^{n} f_1(t) \to 0 \ as\, n \to \infty .$$

Hence $\frac{1}{n}\sum_{i=1}^{n} h_{it} \xrightarrow{p} 0 \ as \ n \to \infty$, for a given t. It then follows that the $T \times 1$

vector $H_n = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} h_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} h_{iT} \end{bmatrix} \xrightarrow{p} H = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ as $n \to \infty$.

Now let $g(H_n) = \frac{1}{T} H_n' i_T$, where i_T is the $T \times 1$ vector of ones. Since g(.) is a

continuous function , we use Proposition 6.1.4 in Brockwell and Davis (1991) to

establish that $g(H_n) = H_n' i_T = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} h_{it} \xrightarrow{p} g(H) = 0$ as $n \to \infty$ with T fixed.

QED.

Further assume that :

(a) the IV-regressor product matrix satisfies $\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} z_{it} w_{it} - Q_{nT} \xrightarrow{p} = 0$, where

$\{ Q_{nT} \}$ is O(1) and has uniformly full rank and (b) The GMM weighting matrix S_{nT}

converges in probability to S which is O(1) and uniformly positive definite.

Then it follows from the theorem in White (1984,p.25) that $\hat{g}$, which is the solution

to establish that $argmin_{g} \left[\frac{1}{n(T-1)}\sum_{i,t} h_{it}\right]' S_{nT} \left[\frac{1}{n(T-1)}\sum_{i,t} h_{it}\right]$ exists in probability

and $\hat{g} \xrightarrow{p} g_0$.

Appendix B: Proof of asymptotic normality

Proposition 2: $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i}^{n} h_{it} \xrightarrow{d} N(0,\Omega)$ as $n \to \infty$ with T fixed.

Proof:

Rewrite $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i}^{n} h_{it}$ as $\frac{1}{\sqrt{T}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_i$, where $H_i \equiv \sum_{t=1}^{T} h_{it} = [h_{i1} \cdots h_{iT}] i_T \equiv h_i' i_T$.

Consider $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_i$ with $E(H_i) = 0$, by the orthogonality conditions. The variance of H_i can be written as

$$Var(H_i) = i_T' Var(h_i) i_T$$

$$= i_T' \begin{bmatrix} E(h_{i1}h_{i1}) & E(h_{i1}h_{i2}) & \cdots & E(h_{i1}h_{iT}) \\ E(h_{i2}h_{i1}) & E(h_{i2}h_{i2}) & \cdots & E(h_{12}h_{iT}) \\ \vdots & \vdots & \ddots & \vdots \\ E(h_{iT}h_{i1}) & E(h_{iT}h_{i2}) & \cdots & E(h_{iT}h_{iT}) \end{bmatrix} i_T$$

$$= i_T' \begin{bmatrix} f_1(1) & f_2(1,2) & 0 & \cdots & \cdots & 0 \\ f_2(2,1) & f_1(2) & f_2(2,3) & 0 & \cdots & 0 \\ 0 & f_2(3,2) & f_1(3) & \ddots & 0 & \vdots \\ \vdots & 0 & & \ddots & \ddots & 0 \\ \vdots & \vdots & & & f_1(T-1) & f_2(T-1,T) \\ 0 & 0 & \cdots & 0 & f_2(T,T-1) & f_1(T) \end{bmatrix} i_T$$

$$= \sum_{t=1}^{T} f_1 + \sum_{t=1}^{T-1} \left[f_2(t,t+1) + f_2(t+1,t) \right] \equiv \Omega_1 .$$

Similarly we establish that $E(H_i H_{i-1}) = \sum_{t=1}^{T-1} \left[f_3(t,t+1) + f_3(t+1,t) \right] \equiv \Omega_2$. By assumption $E(H_i H_{i-j}) = 0 \;\; if \;\; j \geq 2$, implying that $\{H_i\}_{i=1}^{\infty}$ is "covariance stationary" . Thus we can use Wold's decomposition Theorem to write

$H_i = \mu + \lambda_i + \theta\lambda_{i-1}$,

where $\lambda_i \sim n(0, \sigma^2)$; $E(H_i) = \mu = 0$ and $\theta = \dfrac{\Omega_1 \pm \sqrt{\Omega_1^2 - 4\Omega_2^2}}{2\Omega_2}$.

Then by the theorem in Anderson (1971, p 429), $\sqrt{n}\left(\overline{H} - \mu\right) \overset{a}{\sim} N(0, \sum_{j=-\infty}^{j=\infty} \sigma_j)$ *as* $n \to \infty$,

where $\sigma_j = E\{(H_i - \mu)(H_{i-j} - \mu)\}$ and $\overline{H} = \dfrac{1}{n}\sum_{i=1}^{n} H_i$. Since $\sigma_j = 0$ *for* $|j| > 1$, we

have $\dfrac{1}{\sqrt{n}}\sum_{i=1}^{n} H_i \overset{a}{\sim} N(0, \Omega_1 + 2\Omega_2)$.

Hence $\dfrac{1}{\sqrt{nT}}\sum_{i=1}^{n} H_i \equiv \dfrac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n} h_{it} \overset{a}{\sim} N\left(0, \dfrac{1}{T}(\Omega_1 + 2\Omega_2)\right) \equiv N(0, \Omega)$ *as* $n \to \infty$ with T

fixed. *QED.*

If Ω is uniformly positive definite and O(1) and assumptions (a) and (b) of Appendix A hold, we can use the theorem in White (1984,p.69) to establish that

$$D^{-1/2}\sqrt{nT}(\hat{\gamma} - \gamma_0) \overset{a}{\sim} N(0, I),$$

where $D \equiv (Q'_{nT} S_{nT} Q_{nT})^{-1} Q'_{nT} S_{nT} \Omega S_{nT} Q_{nT} (Q'_{nT} S_{nT} Q_{nT})^{-1}$.

In practice, the asymptotic covariance matrix Ω is substituted by its positive semidefinite consistent estimator and Q_{nT} is replaced by $Z'W/nT$, where Z and W are the matrices of instruments and regressors respectively. Since $\{H_i\}_{i=1}^{\infty}$ is "covariance stationary ", the nonparametric estimators of asymptotic covariance matrices in the econometric literature , [e.g., Newey and West ,1987) should work under some regularity conditions regarding the underlying error terms.

References

Anderson,T.W ,1971, *The Statistical Analysis Time Series.* New York: Wiley.

Brockwell,P.J. and R.A. Davis ,1991 , *Time series: Theory and Methods*. New York: Springer-Verlag.

Blundell,R.; M. Browning and C. Meghir,1994, Consumer demand and the life-cycle allocation of household expenditure, *Review of Economic Studies* 61, 57-80.

Crepon, B. and J. Mairesse,1996, The Chamberlain approach, in L. Matyas and P. Sevestre, eds., *The Econometrics of Panel Data: Handbook of Theory and Applications* (Kluwer Academic Publishers, Dordrecht) 323-395.

Collado, M.D,1997, Estimating dynamic models from time series of independent cross sections, *Journal of Econometrics* 82, 37-62.

Deaton, A.,1985, Panel data from time series of cross sections, Journal of Econometrics 30, 109-126.

Frees, E.D.,1995, Assessing cross-sectional correlation in panel data, *Journal of Econometrics* 69, 393-414.

Hansen, L.P.,1982, Large sample properties of generalised method of moments estimators, *Econometrica* 50, 1029-1054.

Hsiao,C.,1986, *Analysis of Panel Data*,(Cambridge University Press,Cambridge).

Judge, G.; W.Griffiths; W. Hill; R.C. Lutkepohl and T.C. Lee, 1985, *The Theory and Practice of Econometrics* (Wiley, New York).

Moffitt, R.,1993, Identification and estimation of dynamic models with time series of repeated cross sections, *Journal of Econometrics* 59, 99-123.

Newey, W.K. and K.D. West, 1987, A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Pesaran, M.H., and R. Smith, 1995, Estimating long-run relationship from dynamic heterogeneous panels, *Journal of Econometrics* 68, 79-112.

Staiger,D., and J.H. Stock, 1997, Instrumental variables regression with weak instruments, *Econometrica* 65, 557-586.

Verbeek,M.,1992, Pseudo Panel Data, in L. Matyas and P. Sevestre, eds., *The Econometrics of Panel Data: Handbook of theory and applications* (Kluwer Academic Publishers, Dordrecht) 303-315.

White, H.,1984, *Asymptotic Theory for Econometricians* (Academic Press, New York).

Table 1

Mean (% bias), standard deviation and asymptotic standard error of the estimators for $\alpha = 0.8$; $\beta = -0.5$

Relative variance of group effect						
	25%		50%		75%	
n=100						
α	PQD	AQD	PQD	AQD	PQD	AQD
Mean	.771 (3.6%)	.358 (55.2%)	.812 (1.5%)	.539 (32.6%)	.810 (1.3%)	.669 (16.4%)
Std deviation	.294	.078	.209	.087	.071	.046
Asym s.error	.297 (1%)	.034 (56.4%)	.175 (16.3%)	.037 (47.5%)	.070 (1.4%)	.026 (43.5%)
β						
Mean	-.519 (3.8%)	-.796 (59.2%)	-.491 (1.8%)	-.665 (33%)	-.494 (1.2%)	-.589 (17.8%)
Std deviation	.211	.06	.135	.062	.057	.042
Asym s.error	.213 (1%)	.034 (43.3%)	.122 (9.6%)	.034 (45.1%)	.054 (5.3%)	.023 (45.2%)
n=200						
α	PQD	AQD	PQD	AQD	PQD	AQD
Mean	.818 (2.2%)	.462 (42.2%)	.805 (.6%)	.623 (22.2%)	.786 (1.8%)	.717 (10.4%)
Std deviation	.439	.090	.151	.073	.049	.041
Asym s.error	.421 (4.1%)	.042 (53.3%)	.131 (13.2%)	.029 (61.3%)	.050 (2%)	.019 (53.6%)
β						
Mean	-.463 (7.4%)	-.714 (42.8%)	-.496 (1%)	-.616 (23.2%)	-.506 (1%)	-.559 (11.8%)
Std deviation	.342	.085	.112	.057	.040	.029
Asym s.error	.299 (12.6%)	.039 (54.1%)	.094 (16.1%)	.026 (54.4%)	.038 (5%)	.017 (41.4%)
n=400						
α	PQD	AQD	PQD	AQD	PQD	AQD
Mean	.764 (4.5%)	.574 (28.2%)	.801 (0%)	.695 (13.1%)	.803 (0%)	.754 (5.7%)
Std deviation	.247	.084	.099	.043	.038	.028
Asym s.error	.238 (3.6%)	.032 (62%)	.091 (8.1%)	.022 (48.9%)	.037 (2.6%)	.014 (50%)
β						
Mean	-.546 (9.2%)	-.663 (32.6%)	-.504 (1%)	-.569 (13.8%)	-.498 (.4%)	-.532 (6.4%)
Std deviation	.196	.063	.059	.037	.026	.019
Asym s.error	.182 (7.2%)	.030 (52.4%)	.065 (10.2%)	.019 (48.6%)	.029 (11.5%)	.013 (31.6%)

Table 2

Mean (%bias) , standard deviation and asymptotic standard error of the estimators for $\alpha = 0.3$; $\beta = 0.8$

Relative variance of group effect						
	25%		50%		75%	
n=100						
α	PQD	AQD	PQD	AQD	PQD	AQD
Mean	.340 (13%)	.121 (60%)	.299 (0%)	.185 (38%)	.304 (1.3%)	.242 (19%)
Std deviation	.221	.029	.107	.028	.042	.024
Asym s.error	.251 (13.6%)	.023 (20%)	.105 (1.9%)	.024 (14.3%)	.041 (2.4%)	.016 (33.3%)
β						
Mean	.769 (3.9%)	.934 (16.8%)	.800 (0%)	.888 (11.1%)	.798 (0%)	.846 (5.75%)
Std deviation	.188	.027	.086	.026	.035	.022
Asym s.error	.211 (12.2%)	.022 (19.5%)	.089 (3.48%)	.022 (15.4%)	.034 (2.8%)	.014 (36.3%)
n=200						
α	PQD	AQD	PQD	AQD	PQD	AQD
Mean	.290 (3.3%)	.164 (45.3%)	.308 (2.6%)	.230 (23%)	.301 (0%)	.267 (11%)
Std deviation	.242	.046	.066	.026	.031	.019
Asym s.error	.206 (14.9%)	.027 (41.3%)	.071 (7.6%)	.018 (31.8%)	.030 (3.3)	.012 (36.8%)
β						
Mean	.815 (1.88%)	.902 (12.8%)	.791 (1.12%)	.854 (6.8%)	.799 (0%)	.826 (3.25%)
Std deviation	.177	.033	.063	.023	.025	.016
Asym s.error	.166 (6.2%)	.025 (24.2%)	.060 (4.8%)	.016 (23.1%)	.025 (0%)	.011 (31.3%)
n=400						
α	PQD	AQD	PQD	AQD	PQD	AQD
Mean	.324 (8%)	.208 (30.6%)	.294 (2%)	.258 (14%)	.297 (1%)	.279 (7%)
Std deviation	.164	.032	.049	.020	.029	.013
Asym s.error	.170 (3.7%)	.022 (31.2%)	.046 (6.1%)	.014 (30%)	.021 (27.6%)	.010 (27.6%)
β						
Mean	.789 (1.4%)	.871 (8.9%)	.805 (0.6%)	.832 (4%)	.803 (.4%)	.817 (2.1%)
Std deviation	.146	.029	.039	.017	.020	.010
Asym s.error	.139 (4.8%)	.020 (31%)	.040 (2.6%)	.013 (23.5%)	.017 (15%)	.007 (30%)