Semiparametric estimation of nonlinear panel data quarterly models with annual sample selection*

Montezuma B.G. Dumangane[1,2], Delger Enkhbayar[3], Nicoletta Rosati[2,4]

[1] Portuguese Competition Authority, [2] CEMAPRE,
[3] London School of Economics, [4] ISEG - Technical University of Lisbon

Abstract

PRELIMINARY AND INCOMPLETE This work aims at developing a semiparametric estimating procedure to correct for sample selection of quarterly data, by using complementary annual information. Estimation of annual models first allow to build appropriate weights, which are later combined into the quarterly models, yielding a consistent and asymptotically normal estimator, with sample selection correction, for the quarterly parameters.

1 Introduction

The estimation of quarterly models, which has gained increasing interest in many areas of economics, has often to deal with deficiencies of the data, many times due to non-random selection of the samples. In fact, because of the higher frequency of the data, in order to reduce the sample size, the units under observation are usually selected according to specific mechanisms, which often lead to nonrandom samples. This paper aims at giving some insight into this type of sample selection models, where the selection mechanism is known, but outlines a procedure which can easily be adapted to other cases.

The PTE survey (Quarterly Firms' Panel survey) is the only available source of quarterly information on Portuguese firms, but is contaminated by a selection problem. The IES (Simplified Firms' Information) survey and other annual sources, which include almost all firms in the population, can be used jointly with the quarterly data, available only for selected firms, to correct the quarterly models. Estimation of such models have to take into account the nonrandom nature of the sampling procedure of the quarterly dataset. In addition to non-random sampling, non-linear modelling has to be addressed, since the variables being modelled are non-negative by nature, and the log transformation is known to produce inconsistent estimates of at least the constant term (see Santos Silva e Tenreyro 2006).

*This paper benefited from valuable contributions from Idílio Freire, Paula Gonçalves, César Leitão, Margarida Martins, Carla Grosa. The hospitality of INE (Portuguese National Bureau of Statistics), where this work was initially developed, is gratefully acknowledged. Corresponding author: Nicoletta Rosati; email: nicoletta@iseg.utl.pt

A semiparametric procedure is developed to produce consistent estimates of the quarterly models. The functional form issue induced by the non-negative nature of the variables is address by specifying an exponential conditional expectation, whose parameters are estimated by GMM with almost efficient instruments. The first step of the procedure consists of specifying a model for the annual Sales, estimated with a panel of companies from annual surveys like the Simplified Firms' Information (IES). The estimated errors from this model are then used to produce weights to correct sample selection in the quarterly data, and yield consistent estimates of quarterly models of interest in a semiparametric way.

The note is organized as follows. In Section 2 the quarterly dataset, the PTE - Painel Trimestral de Empresas - and its non-random nature is briefly described. Section 3 develops the estimating procedure for the quarterly models under non-random sampling. This requires estimation of annual models, whose results will be used to obtain consistent estimation of the quarterly models by a weighted GMM procedure. Section 4 presents asymptotic results of consistency and normality of the semiparametric estimator. Section 5 presents the empirical results, and Section 6 concludes.

2 The Quarterly Firms' Panel survey

The Quarterly Firms' Panel survey (PTE) is a statistical survey of accounting data on a panel of non-financial companies. Its aim is to produce quarterly accounting information for the non-financial sector of the National Quarterly Accounts. It is composed by a panel of constant companies throughout the calendar year, assuring the maximum sample overlap in consecutive years.

From the statistical point of view the PTE is a non-random and stratified survey. Non-randomness comes from the fact that only companies with Sales above 600 000€ (this threshold might vary in certain activity sectors) in the 4th quarter of year t may be included in the sample of year $t+1$. The sample is then stratified according to the number of workers, being some strata exhaustive. Details on the survey methodology of the PTE can be found on the "Documento metodológico do INE - Departamento de Estatísiticas Económicas EP0033".

As such when using the PTE in an estimation procedure we face a typical *sample selection* problem. Dealing with this issue in the context of nonlinear models will be the main objective of this note.

3 Estimation of quarterly models

In this section the estimation of the quarterly models taking into account the sample selection problem is described. The methodology is applied to the estimation of reduced forms of costs and sales functions, but can be extended to the estimation of any quarterly regression function whose parameters might be inconsistently estimated if the sample selection is not taken into account.

3.1 The quarterly models

Interest lies in the estimation of Sales and Costs up to a given quarter, since the aim is to use these models in the national quarterly accounts and this is how the data from the PTE is registered (cumulatively). As such, observations from different quarters are not directly comparable. As a consequence, a model for each quarter is estimated, and the observations for each quarter are a panel of the companies in the PTE with an annual frequency. The models have the same regressors but allow for quarter specific parameters, which may be mutually restricted.

Let $t = 2002, ..., 2007$ be the year index and $q = 1, 2, 3, 4$ the quarters. For each company, indexed by $i = 1, ..., N$, let S_{it}^q be the sales of the company in the quarter q of year t , and define costs C_{it}^q similarly. The regression models for sales and costs can be expressed as

$$E(S_{it}^q|x_{s,it}^q, \alpha_{s,i}^q) = \mu_{s,it}^q \ \alpha_{s,i}^q \qquad \mu_{s,it}^q = \exp(x_{s,it}^q \beta_s^q) \tag{1}$$

$$E(C_{it}^q|x_{c,it}^q, \alpha_{c,i}^q) = \mu_{c,it}^q \ \alpha_{c,i}^q \qquad \mu_{c,it}^q = \exp(x_{c,it}^q \beta_c^q)$$

where $x_{s,it}^q$ and $x_{c,it}^q$ are respectively the observed regressors for the sales and cost function for quarter t, and $\alpha_{s,i}^q$ and $\alpha_{c,i}^q$ are unobserved and time constant random individual specific effects satisfying $E(\alpha_{s,i}^q|x_{s,it}^q) = E(\alpha_{c,i}^q|x_{c,it}^q) = 1$. The set of regressors considered for the sales function is: L_{it}^q the log of the number of workers, K_{it}^q the log of capital (imobilizado corpóreo), RKL_{it}^q the ratio of the standardized number of workers to capital, KI_{it}^q the log of imobilizado incorpóreo, $ICVN_{it-1}$ an index of concentration of sales of the A60 class of the firm defined in (?), $ICSE_t^q$ a quarterly index of economic performance and $DGDP_t^q$ the GDP growth rate. The cost function in addition considers a LS_{t-1}^q the log of sales of the previous year.

These models were estimated for the following institutional economic sectors (CAE = economic activity code)

Group	**Sector**
1	Electricity, gas and water supply (CAE 40 - 41)
2	Manufacturing (CAE 10 - 37)
3	Building (CAE 45)
4	Retail (CAE 50 - 52)
5	Restaurants and Hotels (CAE 55)
6	Transports and Telecommunications (CAE 60 - 64)
7	Services (CAE 70 - 74)
8	Other activities (other CAE)

3.2 The sample selection problem

The sample selection problem arises because only firms with annual sales above a certain threshold are selected into the following year PTE. As such two types of firms are selected.

Those with large sales induced by large observed covariates, and those with large sales because of large systematic errors. If these annual errors are correlated with the errors of the quarterly models of sales and costs, then not taking it into account will lead to inconsistent estimation of the quarterly parameters.

The first step is then to specify a model for the selection equation which is a regression model for the annual sales. Let S_{it} be the annual sales, and assume that

$$E(S_{it}|x_{s,it},\alpha_{s,i}) = \mu_{s,it}\alpha_{s,i}, \qquad t = 2001, ..., 2007 \tag{2}$$

where as before $\alpha_{s,i}$ is an individual specific effect. Also define the annual multiplicative error as $\varepsilon_{s,it} = S_{it}(\mu_{s,it}\alpha_{s,i})^{-1}$ which by construction has conditional mean $E(\varepsilon_{s,it}|x_{s,it},\alpha_{s,i}) = 1$.

Note that, since annual data is available on almost all firms in the population in the IES, the errors, or some functions of interest of them, can be observed . As such this is not a typical selection problem (quote something...) where the selection equation is a binary model and the equations of interest have to be conditioned on the binary outcome of the selection variable.

3.2.1 The traditional approach

Under sample selection the conditional expectation for the quarterly model of interest is not (1) but needs to allow for correlation between the errors of the selection equation and the quarterly model. From now on the exposition will focus on estimation of the quarterly sales equation but it is the same for all others.

Traditionally the parameters of exponential conditional expectations have been estimated by taking logs and applying linear regression techniques, ignoring the Jensen inequality. In fact estimating the regression parameters this way always leads to biased constant terms, and only in a special unlikely case can the regressor coefficients be estimated consistently (see for example Santos Silva e Tenreyro). In this linear context a possible parametric solution would be to condition on the selection equation error

$$E(\ln S^q_{it}|x^q_{s,it},\alpha^q_{s,i},\varepsilon_{s,it}) = \ln\mu^q_{s,it} + \ln\alpha^q_{s,i} + E(\ln\varepsilon^q_{s,it}|\varepsilon_{s,it})$$

where $\varepsilon^q_{s,it}$ is the quarterly error, and assume for example

$$E(\ln\varepsilon^q_{s,it}|\varepsilon_{s,it}) = \omega_i + \theta^q \ln\varepsilon_{s,it}$$

allowing for a quarter specific parameter and an individual effect ω_i. Upon substitution the parameters could be estimated with standard linear panel data techniques. If instead the exponential regression were to be preserved and the quarterly error was multiplicative the selection problem could be written as

$$E(S^q_{it}|x^q_{s,it},\alpha^q_{s,i},\varepsilon_{s,it}) = \mu^q_{s,it}\alpha^q_{s,i}E(\varepsilon^q_{s,it}|\varepsilon_{s,it}), \tag{3}$$

with for for example $E(\varepsilon^q_{s,it}|\varepsilon_{s,it}) = \exp(\omega_i + \theta^q \ln\varepsilon_{s,it})$. If $\varepsilon_{s,it}$ could be estimated consistently this procedure would be consistent if the conditional expectation on the right hand side of (3) was correctly specified.

3.2.2 A semiparametric approach

Assuming a parametric specification for the conditional expectation of the quarterly error is a nonrobust procedure to deal with the selection problem. We now propose a semiparametric method that overcomes this issue, and the need to make assumptions on ω_i. The procedure is inspired by the work presented in Powell (1987) and Kyriazidou (1997), and deals with selection effect as a fixed effect that can be differenced out exploiting the panel structure of the data.

Define the multiplicative conditional error of the annual equation as

$$\varepsilon_{s,it} = S_{it}/E(S_{it}|x_{s,it}, \alpha_{s,i}),$$

satisfying $E(\varepsilon_{s,it}|x_{s,it}, \alpha_{s,i}) = 1$, and the additive conditional errors of the quarterly equations

$$\varepsilon^q_{s,it} = S^q_{it} - E(S^q_{it}|x^q_{s,it}, \alpha^q_{s,i})$$

with $E(\varepsilon^q_{s,it}|x^q_{s,it}, \alpha^q_{s,i}) = 0$.

The conditional on selection quarterly equations now become

$$E(S^q_{it}|x^q_{s,it}, \alpha^q_{s,i}, \varepsilon_{s,it}) = \mu^q_{s,it}\alpha^q_{s,i} + \lambda_q(\varepsilon_{s,it})$$

where $\lambda_q(\varepsilon_{s,it}) = E(\varepsilon^q_{s,it}|\varepsilon_{s,it})$ is the selection effect which is left unspecified. The procedure proposed in the next subsection deals with this function as it it were an individual effect that can be differenced out.

Differencing out the selection effect Let us go back to the error representation of the quarterly model and take annual first differences

$$\triangle S^q_{it} = \alpha^q_{s,i} \triangle \mu^q_{s,it} + \triangle \varepsilon^q_{s,it}$$

Let $\varepsilon_{s,i} = (\varepsilon_{s,it}, \varepsilon_{s,it-1})$ and $\mathbf{x}^q_{s,it} = (x^q_{s,it}, x^q_{s,it-1})$. Taking expectations conditional on these two vectors

$$E[(\triangle S^q_{it} - \alpha^q_{s,i} \triangle \mu^q_{s,it})|\mathbf{x}^q_{s,it}, \alpha^q_{s,i}, \varepsilon_{s,i}] = E(\triangle \varepsilon^q_{s,it}|\mathbf{x}^q_{s,it}, \alpha^q_{s,i}\varepsilon_{s,i})$$

At this stage we require the additional assumptions that $\alpha^q_{s,i}$ is a random effect and $\varepsilon^q_{s,it}$ is exogenous. By iterated expectations, and remembering that we have set previously $E(\alpha^q_{s,i}|x^q_{s,it}) = 1$, we obtain:

$$E[(\triangle S^q_{it} - \triangle \mu^q_{s,it})|\mathbf{x}^q_{s,it}, \varepsilon_{s,i}] = E(\triangle \varepsilon^q_{s,it}|\varepsilon_{s,i}) \quad (4)$$

Let the selection effects now be defined as $\lambda^q_{it} \equiv E(\varepsilon^q_{s,it}|\varepsilon_{s,i})$ and $\lambda^q_{it-1} \equiv E(\varepsilon^q_{s,it-1}|\varepsilon_{s,i})$. Then we can rewrite (4) as

$$E[(\triangle S^q_{it} - \triangle \mu^q_{s,it})|\mathbf{x}^q_{s,it}, \varepsilon_{s,i}] = \lambda^q_{it} - \lambda^q_{it-1}$$

Unless there is no selection effect, i.e., the selection and quarterly errors are uncorrelated, $\lambda^q_{it}, \lambda^q_{it-1} \neq 0$. As such any estimation procedure based on first differencing will produce inconsistent estimates.

Semiparametric estimation Consistent estimation under sample selection requires specifying conditions under which the right hand side of (4) equals zero, so that it provides a moment condition suitable for consistent estimation. In order to proceed in that direction an additional set of assumptions has to be imposed.

1. λ_{it}^q does not depend on t, i.e. the functional form of the selection effect in the same quarter of different years is the same.

2. A symmetry assumption: $E(\varepsilon_{s,ip}^q|\varepsilon_{s,ip},\varepsilon_{s,p-1}) = E(\varepsilon_{s,ip}^q|\varepsilon_{s,ip},\varepsilon_{s,p+1})$

A stronger but more intuitive version of assumption 2 postulates that

2a. $E(\varepsilon_{s,it}^q|\varepsilon_{s,i}) = E(\varepsilon_{s,it}^q|\varepsilon_{s,it})$ and $E(\varepsilon_{s,it-1}^q|\varepsilon_{s,i}) = E(\varepsilon_{s,it-1}^q|\varepsilon_{s,it-1})$

Under these assumptions it is easy to see that if $\varepsilon_{s,it} = \varepsilon_{s,it-1}$ then $\lambda_{it}^q(\varepsilon_{s,it}) - \lambda_{it-1}^q(\varepsilon_{s,it-1}) = 0$, i.e., the selection effects behave like an individual effect that can be differenced out As a consequence, the following moment conditions are satisfied:

$$E[(\triangle S_{it}^q - \triangle\mu_{s,it}^q)|x_{s,it}^q, \varepsilon_{s,it} = \varepsilon_{s,it-1}] = 0 \tag{5}$$

and for some set of instruments $z_{s,it}^q$ a GMM estimator based on

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi_{it}(\triangle S_{it}^q - \triangle\mu_{s,it}^q)z_{s,it}^q \approx 0$$

where $\psi_{it} = 1(\varepsilon_{s,it} = \varepsilon_{s,it-1})$. This operator selects the observations for which the selection effects vanish with the annual differencing. In practice ψ_{it} is replaced by the kernel

$$\hat{\psi}_{it} = \frac{1}{h_n}K\left(\frac{\triangle\varepsilon_{s,it}}{h_n}\right), \qquad \lim_{n\to\infty} h_n = 0.$$

which has the role of weighting the observations according to the proximity of $\triangle\varepsilon_{s,it}$ to zero.

Since the annual error is specified as being multiplicative, estimation of the observed component of $\mu_{s,it}$ in (2) does not allow us to separate it from the individual effect. Instead, what can be consistently estimated is quantity $e_{s,it} = a_{s,i}\varepsilon_{s,it}$, where $a_{s,i}$ includes the individual effect and all other fixed regressors if a fixed effect estimation was performed. Consider the following transformation

$$\frac{e_{s,it}}{e_{s,it} + e_{s,it-1}} = \frac{\varepsilon_{s,it}}{\varepsilon_{s,it} + \varepsilon_{s,it-1}}$$

This is independent of $a_{s,i}$, and quantity $\triangle\varepsilon_{s,it}$ is close to zero if and only if quantity $\tau(\triangle e_{s,it}) \equiv \frac{\triangle e_{s,it}}{e_{s,it}+e_{s,it-1}}$ is close to zero, so that the kernel can be based on

$$\hat{\psi}_{it} = \frac{1}{h_n}K\left(\frac{\tau(\triangle\hat{e}_{s,it})}{h_n}\right) \tag{6}$$

which can be estimated consistently[1].

[1] Instead of the ratio we could have defined the transformation $e_{v,it} = \ln(a_{s,i}\varepsilon_{s,it})$ with similar results.

Additional moment conditions Since asymptotically this procedure uses only observations for which $\varepsilon_{s,it} = \varepsilon_{s,it-1}$, the sample size can be drastically reduced, with implications in the efficiency of the estimator. To circumvent this problem a possible solution is to extend the number of moment conditions to be used. This can be done by noting that under the assumptions (5) above can be generalised for any other lag, providing a family of moment conditions as

$$E[\psi^p_{it}(\triangle^p S^q_{it} - \triangle^p \mu^q_{s,it})|x^q_{s,it}] = 0, \qquad p \geq 1 \tag{7}$$

Accordingly this requires defining the appropriate set of weights

$$\hat{\psi}^p_{it} = h_n^{-1} K\left(\frac{\tau(\triangle^p \hat{e}_{s,it})}{h_n}\right)$$

to control for the sample selection effect. Estimation can then proceed in GMM framework for NSURE - Non Linear Seemingly Unrelated Regression - as now the estimator is defined by the p set of moment conditions.

3.3 Estimation of the annual error

In this section the estimation procedure of the selection equation is described in more detail. Note that this equation is the basis to produce the weights that will correct any estimation procedure that uses the PTE dataset and that might be affected by the selection bias. The aim of this estimation is solely to produce the errors, or a function of them that allows us to determine when the selection effect in the quarterly equations will be differenced out.

To estimate the annual data, a panel of 51346 firms observed from 2001 to 2007 from IES, IEH and IRC, complementing the information with the BDCI (ask Paula for more details on the sampling procedure).

Since the only purpose of the estimation of the selection equation is computing the error $e_{s,it}$ that might include any time constant effect, a fixed effects strategy is adopted.

Let $\mu_{s,it} \equiv c_{s,i} m_{s,it}$, where $m_{s,it}$ is the time varying component of regression function and $c_{s,i}$ includes all time constant observed regressors. Consider the Wooldridge (fixed effect) transformation for multiplicative error models; notice that $S_{it}/m_{s,it} = e_{s,it}$, so that:

$$e_{s,it} - e_{s,it-1} = a_{s,i}(\varepsilon_{s,it} - \varepsilon_{s,it-1}) \tag{8}$$

where $a_{s,i} \equiv c_{s,i}\alpha_{s,i}$. Under very mild assumptions for the errors and regressors, quantities in (8) satisfies $E(e_{s,it} - e_{s,it-1}|\mathbf{x}_{s,it}) = 0$ where $\mathbf{x}_{s,it} = (x_{s,it}, x_{s,it-1})$ are the observed time varying regressors. Note that, because of the presence of $\alpha_{s,i}$ in the right hand side of (8), the list of regressors may not include lag dependent variables and requires that the regressors are exogenous. Of course this requirements can be relaxed if suitable instruments are found that make the expectation equal to zero.

Let β_m be the parameter vector associated to the time varying regressors. To find the instruments to perform GMM estimation note that the optimal instruments for the conditional moment based on $g_{it} \equiv e_{s,it} - e_{s,it-1}$ give

$$E\left(\frac{\partial}{\partial \beta_m} g_{it} | \mathbf{x}_{s,it}, c_{s,i}, \alpha_{s,i}\right) = - \triangle\, x_{s,it} c_{s,i} \alpha_{s,i}$$

which suggests the use of the $(T \times KT)$ matrix $\triangle \mathbf{X}_{s,i}$ the block diagonal matrix with $\triangle x_{s,it}$ in the tth block as instruments

$$\triangle \mathbf{X}_{s,i} = \begin{bmatrix} \triangle x_{s,i1} & 0 & 0 & \cdots & 0 \\ 0 & \triangle x_{s,i2} & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & \triangle x_{s,iT} \end{bmatrix} \tag{9}$$

leading to the unconditional moment

$$E[\triangle \mathbf{X}'_{s,i} \mathbf{g}_i(\beta_m)] = 0$$

where $\mathbf{g}_i$ is $(T \times 1)$ vector. Given this moment condition, for this set of instruments the efficient estimator was obtained by solving

$$\min_{\beta_m} \sum_{i=1}^{N} [\triangle \mathbf{X}'_{s,i} \mathbf{g}_i(\beta_m)]' \hat{\Lambda}^{-1} \sum_{i=1}^{N} [\triangle \mathbf{X}'_{s,i} \mathbf{g}_i(\beta_m)]$$

where $\hat{\Lambda} \equiv N^{-1} \sum_{i=1}^{N} \triangle \mathbf{X}'_{s,i} \mathbf{g}_i(\hat{\beta}_m) \mathbf{g}_i(\hat{\beta}_m)' \triangle \mathbf{X}_{s,i}$ for a preliminary estimate of β_m. Under standard regularity conditions this estimator is consistent and asymptotically normal with estimated asymptotic variance

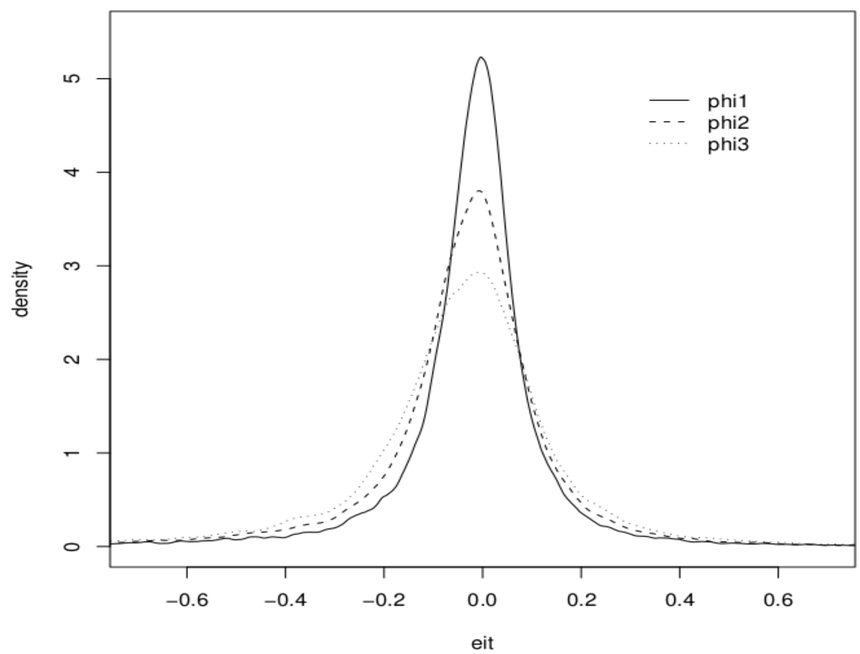
$$\widehat{\text{Avar}}(\hat{\beta}_m) = \hat{G}' \hat{\Lambda} \hat{G} / N, \qquad \hat{G} = N^{-1} \sum_{i=1}^{N} \triangle \mathbf{X}'_{s,i} \nabla_{\beta_m} \mathbf{g}_i(\beta_m)$$

3.3.1 Estimation of the error and the weights

Upon estimation of the parameters, the errors $e_{s,it}$ can be consistently estimated by $\hat{e}_{s,it} = S_{it} m^{-1}_{s,it}(\hat{\beta}_m)$. The estimates of the weights are the values of the non-parametric estimator of the probability density function of $\triangle^p \hat{e}_{s,it}$. Here we consider a kernel density estimate with a normal kernel and a bandwidth $0.9 \min(\hat{\sigma}, R/1.34) n^{-1/5}$ which is the default in the statistical package R. The choice of bandwidth is a compromise between smoothing enough to remove insignificant bumps and not smoothing too much to smear out real peaks.

As shown in Figure 1, the density becomes more disperse as p grows.

Figure 1: Density estimation



3.4 Random effects and efficient estimation of quarterly models

The focal point of this procedure is the estimation of the quarterly models from the PTE sample selected sample. Unfortunately the data quality of this survey does not allow the use of the information of all sampled firms. In fact many inconsistencies were found in some observations, namely the fact that the implicit sales and costs were found to be negative for some quarters in several companies. Moreover, it seems like the accounting informations provided by firms in the 4th quarter has a slightly different nature . An exercise of eliminating the main inconsistent observations was made, whose outcome is here assumed to have a completely random nature. After this cleansing, the dataset used for estimation included 3222 firms observed for 2002 to 2007.

3.4.1 The moment conditions

For this particular exercise the aim was to estimate jointly the quarterly parameters of both sales and costs equations. For each equation, the parameters were estimated using two moment conditions by setting $p = 2$ in (7). This increased the complexity of the estimation procedure but was found to be essential for the efficiency and identifiability of the parameters[2]. So let

$$g^{q,p}_{s,it}(S^q_{it}, \mu^q_{s,it}) \equiv \triangle^p S^q_{it} - \triangle^p \mu^q_{s,it}$$

be the function with weighted conditional expectation zero, for $p = 1, 2$, $q = 1, 2, 3, 4$. A similar expression can be derived for the cost function.

Because of nonlinear functional form of the conditional expectation of the sales and cost functions, the differencing operation still allows identification of all parameters, in particular the time constant parameters[3]. The quarterly parameters are restricted by specifying the regression function as

$$\mu^q_{s,it} = \exp(x^q_{s,it}\beta^{(q)}_s), \qquad \beta^{(q)}_s = \beta^{(4)}_s + \delta^{(q)}_s I(q < 4)$$

The $\delta^{(q)}_s$ are the differences of the parameter vector with respect to the 4^{th} quarter.

Finding the instruments The next step is to find a suitable set of instruments for GMM estimation. The choice of instruments has implications in the efficiency of the estimation procedure, but since the sample sizes used in some cases are not very large, they are important for the algorithm to be able to converge. This is because the way instruments must reflect the marginal effect of the regressors on the conditional expectation and also the conditional variance of the moment conditions.

Having defined a conditional expectation under some assumptions, it would be possible to find the optimal set of instruments. That would require computing a very complex covariance matrix of the moment conditions with non zero off diagonal elements due to the presence of lagged errors. Instead, a simpler form of computing the instruments was adopted, which ignored this serial correlation but still retained a "flavour" of efficiency. The instruments were defined as

$$z^{q,p}_{s,it} = E\left[\frac{\partial}{\beta^{(q)}_s} g^{q,p}_{s,it}(S^q_{it}, \mu^q_{s,it})|\mathbf{x}^q_{s,it}\right] / Var[g^{q,p}_{s,it}(S^q_{it}, \mu^q_{s,it})|\mathbf{x}^q_{s,it}] \tag{10}$$

Assuming that $Var[S^q_{it}|x^q_{s,it}, \alpha^q_i] = E[S^q_{it}|x^q_{s,it}, \alpha^q_i]$ and that $Var[\alpha^q_i|x^q_{s,it}] = \eta^2$, by iterated expectations it is easy to see that

$$E\left[\frac{\partial}{\beta^{(q)}_s} g^{q,p}_{s,it}(S^q_{it}, \mu^q_{s,it})|\mathbf{x}^q_{s,it}\right] = -\mu^q_{s,it}x^q_{s,it} + \mu^q_{s,it-1}x^q_{s,it-1}$$

$$Var[g^{q,p}_{s,it}(S^q_{it}, \mu^q_{s,it})|\mathbf{x}^q_{s,it}] = \mu^q_{s,it} + \mu^q_{s,it-1} + \eta^2(\mu^q_{s,it} - \mu^q_{s,it-1})^2 + 2\mu^q_{s,it}\mu^q_{s,it-1}$$

The conditional variance expression relies on the equivariance assumption which may not be valid . In any case this misspecification does not have consistency implications for the estimator, since it only affects the instruments. The instruments chosen this

[2]The importance of this is stressed by the fact that, when $p = 1$, in most activity sectors the estimation algorithm did not converge.

[3]This would not be the case if the model were to be linear.

way allow estimation to be more efficient and "smoother" than if we considered $x^q_{s,it}$ and $x^q_{s,it-1}$ or even $x^q_{s,it} - x^q_{s,it-1}$.

The unconditional $T \times 1$ moment condition associated to $\mathbf{g}^{q,p}_{s,i}$ can now be defined as $E[\mathbf{Z}^{q,p\prime}_{s,i}\mathbf{g}^{q,p}_{s,i}(\beta^{(q)}_s)] = 0$, where the matrices $\mathbf{Z}^{q,p}_{s,i}$ are defined analogously to (9).

Joint estimation of equations Let $\mathbf{r}^{q,p}_{s,i}(\beta^{(q)}_s) \equiv \mathbf{Z}^{q,p\prime}_{s,i}\mathbf{g}^{q,p}_{s,i}(\beta^{(q)}_s)$ be the $KT \times 1$ vector with zero unconditional moment. Estimation is joint in two different ways.

Firstly because for each equation (sales and costs) all quarterly parameters are estimated simultaneously using for each quarter the two moment conditions, i.e., $p = 1, 2$. This requires defining vector $\mathbf{r}^q_{s,i}(\beta^{(q)}_s)' = [\mathbf{r}^{q,1}_{s,i}(\beta^{(q)}_s)' \vdots \mathbf{r}^{q,2}_{s,i}(\beta^{(q)}_s)']$ and vector $\mathbf{r}_{s,i}(\theta_s)' = [\mathbf{r}^1_{s,i}(\beta^{(1)}_s)' \vdots \cdots \vdots \mathbf{r}^4_{s,i}(\beta^{(4)}_s)']$ the $8KT \times 1$ vector, and $\theta'_s = [\beta^{(1)\prime}_s \vdots \cdots \vdots \beta^{(4)\prime}_s]$ the complete set of parameters for the sales equation.

Secondly, joint estimation exploits possible correlations between the sales and the costs equations which have the same structure. So let us define vector $\mathbf{r}_{c,i}(\theta_c)' = [\mathbf{r}^1_{c,i}(\beta^{(1)}_c)' \vdots \cdots \vdots \mathbf{r}^4_{c,i}(\beta^{(4)}_c)']$ and vector $\theta'_c = [\beta^{(1)\prime}_c \vdots \cdots \vdots \beta^{(4)\prime}_c]$ such that

$$E \begin{bmatrix} \mathbf{r}_{s,i}(\theta_s) \\ \mathbf{r}_{c,i}(\theta_c) \end{bmatrix} = 0$$

The N3SLS estimator of $\theta' = (\theta'_s, \theta'_c)$, the $k = k_s + k_c$ parameter vector of interest, can now be defined as

$$\min_\theta \sum_{i=1}^N \mathbf{r}_i(\theta)'\hat{\Phi}^{-1}\sum_{i=1}^N \mathbf{r}_i(\theta)$$

where $\hat{\Phi} \equiv NT^{-1}\sum_{i=1}^N \mathbf{r}_i(\theta)\mathbf{r}_i(\theta)'$ is an estimate of the covariance matrix of the moment conditions and allows for possible correlation between the sales and the costs equation. Again the estimator is asymptotically normal.

4 Asymptotics

4.1 Assumptions

Assumption A1 (Regularity conditions): The data $\{S_{it}, x_{it}, S^q_{it}, x^q_{it}, z^q_{it}\}$ is a random sample. The set B is compact and β_0 lies in its interior. The function $\eta_i(\beta)$ is twice continuously-differentiable on B for almost all i. For $t = 2, ..., T$, $E(||z_{it}||^4)$ and $E(\varepsilon^4_{it})$ are finite. $E[Z_i\Omega_i\frac{\partial}{\partial\beta_0}\eta_i(\beta_0)]$ has full column rank around $\triangle\varepsilon$ in a neighbourhood of 0. The moment condition $E(Z_i\Omega_i\eta_i(\beta_0)) = 0$ holds.

Assumption A2 (Smoothness): For $t = 2, ..., T$, $f_t(\triangle\varepsilon_{it})$ is bounded from above for all $\triangle\varepsilon_{it}$ and is strictly positive for $\triangle\varepsilon_{it}$ in a neighbourhood of 0. In addition $f_t(\triangle\varepsilon_{it})$ and $E(\omega_r(\triangle\varepsilon_{it})z_{it}\eta_{it}(\beta))$ are $\kappa + 1$-times continuously differentiable in $\triangle\varepsilon_{it}$ for all $\triangle\varepsilon_{it}$ in a neighbourhood of 0, where $\kappa \geq 1$ is an integer.

Assumption A3 (Kernel weights): The kernel, k , is bounded on its support and of order κ. Moreover, $\int k(v)dv = 1$, $\int v^j k(v)dv = 0$, for $|j| = 1, ..., \kappa-1$, $\int |v^j||k(v)|dv < \infty$ for $|j| \in \{0, \kappa\}$. The bandwidth h_n is non-negative and $o(1)$ as $n \to \infty$, while $\sqrt{nh_n} \to \infty$, $\sqrt{nh_n}h_n^{\kappa} \to \tilde{h}$, where $0 \leq \tilde{h} < \infty$.

Assumption A4 (Selection effects): λ_{it}^q does not depend on t, i.e. the functional form of the selection effect in the same quarter of different years is the same. Moreover, assume symmetry: $E(\varepsilon_{it}^q|\varepsilon_{it},\varepsilon_{i,t-1}) = E(\varepsilon_{it}^q|\varepsilon_{it}, \varepsilon_{i,t+1})$

Assumption A5 (Higher order moments):

4.2 Proof of consistency and asymptotic normality

Subscripts s and c, indicating sales and costs, respectively, are dropped for notational ease.

The annual model (the selection equation) is given by:

$$E[S_{it}|X_{it}, \alpha_i] = \alpha_i e^{x_{it}\beta_0}$$

where the multiplicative error term satisfies

$$E(\varepsilon_{it}|X_{it}, \alpha_i) = 1$$

The quarterly model can be written in this form, if we condition on the selection equation error (allowing for correlation between the errors of the annual model and the errors of the quarterly model):

$$E[S_{it}^q|X_{it}^q, \alpha_i^q, \varepsilon_{it}] = \alpha_i^q e^{x_{it}^q\beta_0} + E(\varepsilon_{it}^q|\varepsilon_{it})$$

where we assume strict exogeneity of ε_{it}^q, $E(\varepsilon_{it}^q|X_{it}^q, \alpha_i^q) = 0$.

Define $\underline{X}_{it}^q = (X_{it}^q, X_{i,t-1}^q)$, $\underline{\varepsilon}_{it} = (\varepsilon_{it},\varepsilon_{i,t-1})$ and let $\lambda_{it} \equiv E(\varepsilon_{it}^q|\varepsilon_{it})$ be the selection effect. Conditional first-differencing yields:

$$\begin{aligned} E[(\triangle S_{it}^q - \alpha_i^q \triangle e^{x_{it}^q\beta_0})|\underline{X}_{it}^q, \alpha_i^q, \underline{\varepsilon}_{it}] &= \lambda_{it}^q - \lambda_{i,t-1}^q \\ E[(\triangle S_{it}^q - \triangle e^{x_{it}^q\beta_0})|\underline{X}_{it}^q, \underline{\varepsilon}_{it}] &= \lambda_{it}^q - \lambda_{i,t-1}^q \end{aligned}$$

assuming $E(\alpha_i^q|\underline{X}_{it}^q) = 1$. Following Kyriazidou (1997), a semiparametric approach is used to eliminate the selection effect, by exploiting the quarterly and the annual nature of the data. The general idea underlying the approach is that, if $\triangle\varepsilon_{it} \equiv \varepsilon_{it} - \varepsilon_{i,t-1} \approx 0$, then $\lambda_{it}^q - \lambda_{i,t-1}^q$ will tend to be small (given assumptions on the selection equation) and in the limit, they can be treated as being equal. Given our assumptions on the selection equation we obtain the following moment conditions:

$$E[(\triangle S_{it}^q - \triangle e^{x_{it}^q\beta_0})|\underline{X}_{it}^q, \varepsilon_{it} = \varepsilon_{i,t-1}] = 0 \quad (11)$$

which are free of the sample selection, hence free of incidental functions. Following Holtz-Eakin, Newey and Rosen (1988), the idea is to replace the first differences with local first differences in order to get a local GMM estimator based on :

$$E[z_{it}^q(\triangle S_{it}^q - \triangle e^{x_{it}^q \beta_0})|\varepsilon_{it} = \varepsilon_{i,t-1}] = 0$$

where z_{it}^q is a vector of instrumental variables, including $\underline{x}_{it}^q$ as well as external instrumental variables.

Let $\psi_{it} \equiv \omega_0(\triangle\varepsilon_{it})$, where $\omega_r(v) \equiv \delta(v - r)$ for Dirac's delta $\delta(\cdot)$, and $\eta_{it}(\beta) \equiv \triangle S_{it}^q - \triangle e^{x_{it}^q \beta}$, then we have:

$$E[\psi_{it} z_{it}^q \eta_{it}(\beta_0)] = 0$$

Given our assumptions, the above set of moment conditions can be generalised to $p \geq 1$ lags:

$$E[\psi_{it}^p z_{it} \eta_{it}^p(\beta_0)] = 0$$

where $\psi_{it}^p = \omega_0(\varepsilon_{it} - \varepsilon_{i,t-p})$ and $\eta_{it}^p(\beta) = \triangle^p S_{it}^q - \triangle^p e^{x_{it}^q \beta_0} = S_{it}^q - S_{i,t-p}^q - (e^{x_{it}^q \beta} - e^{x_{i,t-p}^q \beta})$, with z_{it} accordingly being a vector of instrumental variables. The estimation can then proceed in GMM framework for NSURE (Nonlinear Seemingly Unrelated Regression Estimation). However, for simplicity of exposition, I will focus on $p = 1$ from now on. One can extend the proofs to the above estimation procedure.

The dimension of x_{it}^q is indicated by k. Typically z_{it}^q will include $\underline{x}_{it}^q$, but it also may include more lags and leads as well as external instruments, so to allow for the instrument set to expand over time, the length of z_{it}^q is written as m_t, with $m = \sum_{t=2}^{T} m_t$. Let Z_i be the $m \times (T-1)$ matrix of instrument vectors z_{it}^q. Then the set of moment conditions may be written as:

$$E[Z_i \Psi_i \eta_i(\beta_0)] = 0 \tag{12}$$

where $\Psi_i = diag(\psi_{i2}, ..., \psi_{iT})$.

A natural way to proceed is to construct a GMM estimator based on a sample analogue of (12), replacing Ψ_i by a smoother such as a kernel weight.

Our estimator of β_0 then is given by:

$$\hat{\beta}_n = \arg\min q_n(\beta)' V_n q_n(\beta) \tag{13}$$

where $q_n(\beta) \equiv \frac{1}{n} \sum_{i=1}^{n} Z_i W_i \eta_i(\beta)$ for a parameter space B and a positive-definite weight matrix $V_n \xrightarrow{p} V_0$. Also,

$$\begin{aligned} W_i &\equiv diag(w_{i2}, ..., w_{iT}) \\ w_{it} &\equiv \frac{1}{h_n} k(\frac{\triangle\varepsilon_{it}}{h_n}) \end{aligned}$$

where $k(\cdot) : \mathbb{R} \to \mathbb{R}$ is a kernel function with a scalar bandwidth $h_n = h(n)$.

Asymptotically, only observations for which $\triangle\varepsilon_{it}$ lies in a shrinking neighbourhood of zero will contribute to the objective function (2). Asymptotic properties are derived for when we know $\{\varepsilon_{it}\}_{t=1}^{T}$. Since in practice we do not observe ε_{it}, then we may replace ε_{it} by GMM estimator, $\hat{\varepsilon}_{it}$, which is described in more detail in section 4.3 of the paper. Under conditions laid out below, the introduction of an estimator for ε_{it} will not change the asymptotic properties of $\hat{\beta}_n$. The proof is not included here, but it follows Theorem 1 of Kyriazidou (1997).

Let us now consider the asymptotic behaviour of $\hat{\beta}_n$. The assumptions given above are used throughout with explanations when necessary. Firstly, I will prove three lemmas that are used in the proof of Theorem 1:

Theorem 1 *Let assumptions B and D hold. Let V_0 be a positive definite non-stochastic matrix such that $V_n \xrightarrow{p} V_0$. Define $Q_0 \equiv Q_0(\beta_0)$. Then as $n \to \infty$, $\hat{\beta}_n \xrightarrow{p} \beta_0$ and $\sqrt{nh_n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, where $\Sigma = (Q_0' V_0 Q_0)^{-1} Q_0' V_0 \Lambda V_0 Q_0 (Q_0' V_0 Q_0)^{-1}$.*

Lemma 1 *Let ι_k be a $k \times 1$ column of 1's and $\Xi = (\iota_{m_2}' f_2(0) ... \iota_{m_T}' f_T(0))'$, where $f_t(.)$ denotes the marginal density of $\triangle\varepsilon_{it}$. Then $q_n(\beta) \xrightarrow{p} q_0(\beta) \equiv E[Z_i \Psi_i \eta_i(\beta)] \odot \Xi$ uniformly on B.*

Lemma 2 *Let $Q_n(\beta) = \frac{\partial q_n(\beta)}{\partial \beta'} = \frac{1}{n} \sum_{i=1}^{n} Z_i W_i \frac{\partial \eta_i(\beta)}{\partial \beta'}$. Then $Q_n(\beta) \xrightarrow{p} Q_0(\beta) \equiv E[Z_i \Psi_i \frac{\partial \eta_i(\beta)}{\partial \beta'}] \odot \Xi \iota_k'$ uniformly on B.*

Lemma 3 *Let G be $m \times m$ matrix of $(T-1)^2$ blocks where the $(t-1, s-1)$th block is given by $\iota_{m_t} \iota_{m_s}' f_{ts}(0,0)$, with $f_{ts}(.)$ denoting the joint density of $\triangle\varepsilon_{it}$ and $\triangle\varepsilon_{is}$. Let $\Lambda \equiv E[Z_i \Psi_i \eta_i(\beta_0) \eta_i(\beta_0)' \Psi_i Z_i'] \odot G \int k(v)^2 dv$. Then $\sqrt{nh_n}(q_n(\beta_0) - Eq_n(\beta_0)) \xrightarrow{d} \mathcal{N}(0, \Lambda)$*

Corollary 1 *(Variance estimator) Let $\Lambda_n \equiv \frac{h_n}{n} \sum_{i=1}^{n} Z_i W_i \eta_i(\hat{\beta}_n) \eta_i(\hat{\beta}_n)' W_i Z_i'$. $\Lambda_n \xrightarrow{p} \Lambda$ given assumptions.*

Proof. The same arguments as in Lemma 1 and 2 can be used to prove this corollary. ■

Corollary 2 *(Optimally-weighted estimator) The efficient weighting scheme sets $V_n = \Lambda_n^{-1}$. Then following,*

$$\sqrt{nh_n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(0, (Q_0' \Lambda^{-1} Q_0)^{-1})$$

Proof. The steps to prove this claim follow directly from Theorem 1 and Corollary 1. It can be shown that $(Q_0'\Lambda^{-1}Q_0)^{-1}$ is the lower bound on the asymptotic variance of $\hat{\beta}_n$.

■

Since ε_{it} is unobservable we replace it with $\hat{\varepsilon}_{it}$, an estimator which we will describe below. It is shown to be consistent and root-n asymptotically normal, given S_{it} and $x_{it.}$We propose an estimator for β_0 which replaces $q_n(\beta)$ by $\tilde{q}_n(\beta)$ in the moment conditions (1):

$$\tilde{\beta}_n = \arg\min \tilde{q}_n(\beta)' V_n \tilde{q}_n(\beta)$$

where $\tilde{q}_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} Z_i \tilde{W}_i \eta_i(\beta)$ with $\tilde{W}_i = diag(\frac{1}{h_n}k(\frac{\triangle\hat{\varepsilon}_{i2}}{h_n}), ..., \frac{1}{h_n}k(\frac{\triangle\hat{\varepsilon}_{iT}}{h_n}))$.

Theorem 2 *Let Assumptions A1-A5 hold. If* $\sqrt{nh_n}h_n^{\kappa} \to \tilde{h}$, *with* $0 \leq \tilde{h} < \infty$, *then* $\sqrt{nh_n}(\tilde{\beta}_n - \hat{\beta}_n) = o_p(1)$.

Since $\left(\tilde{\beta}_n - \beta\right) = (\tilde{\beta}_n - \hat{\beta}_n) + (\hat{\beta}_n - \beta_0)$,

$$\sqrt{nh_n}(\tilde{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

Its asymptotic properties remain unchanged by the fact that we are using an estimator of W_i, $\tilde{W}_i$. If the estimator of ε_{it}, $\hat{\varepsilon}_{it}$, is a consistent estimator that satisfies $\hat{\varepsilon}_{it} - \varepsilon_{it} = O_p(n^{-p})$, where $\frac{2}{5} < p \leq \frac{1}{2}$ and $h_n = h \cdot n^{-\mu}$, where $0 < h < \infty$, and $1 - 2p < \mu < \frac{p}{2}$, this claim follows Theorem 1 of Kyriazidou (1997).

It can be shown that the rate of convergence of $\hat{\beta}_n$ is lower if we choose h_n such that $\sqrt{nh_n}h_n^{\kappa} \to 0$, rather than $\sqrt{nh_n}h_n^{\kappa} \to \tilde{h} > 0$. The rate of convergence in distribution of $\hat{\beta}_n$ is maximised by setting $\mu = \frac{1}{2\kappa+1}$, in which case it becomes $n^{-\frac{\kappa}{2\kappa+1}}$. If we use a kernel symmetric around 0 ($\kappa = 2$), $\mu = \frac{1}{5}$ and the rate of convergence of $\hat{\beta}_n$ is $n^{-\frac{2}{5}}$. Thus, for κ large enough, the estimator converges at a rate that can be arbitrarily close to $n^{-\frac{1}{2}}$, provided that $\hat{\varepsilon}_{it}$ is estimated fast enough with $\frac{1}{2} = p > \frac{\kappa}{2\kappa+1}$, which is satisfied. Similarly, following Corollary of the same paper, it can be shown that we can eliminate the asymptotic bias while maintaining the maximal rate of convergence in the manner suggested by Bierens (1987).

Here I examine the asymptotic distribution of $\hat{\varepsilon}_{it}$, which is a function of $\hat{\beta}_m$, given in section 3.3 of the paper. We get the following moment conditions:

$$E[\triangle X_i' g_i(\beta_m)] = 0$$

where β_m is the parameter vector associated with the time varying regressors.

The GMM estimator of β_m is given by:

$$\hat{\beta}_m = \arg\min \left[\frac{1}{n}\sum_{i=1}^{n} \triangle X_i' g_i(\beta_m)\right]' W_n \left[\frac{1}{n}\sum_{i=1}^{n} \triangle X_i' g_i(\beta_m)\right]$$

where $W_n \xrightarrow{p} W_0$ with a positive definite matrix W_0. The efficient weighting scheme assigns $W_0 = E\left[\triangle X_i' g_i(\beta_m) g_i(\beta_m)' \triangle X_i\right]$.

For 2-step GMM estimation, we can use $W_n = \frac{1}{n}\sum_{i=1}^{n} \triangle X_i' g_i(\tilde{\beta}_m) g_i(\tilde{\beta}_m)' \triangle X_i$ where $\tilde{\beta}_m = \arg\min \left[\frac{1}{n}\sum_{i=1}^{n} \triangle X_i' g_i(\beta_m)\right]' \left[\frac{1}{n}\sum_{i=1}^{n} \triangle X_i' g_i(\beta_m)\right]$.

Under standard regularity conditions this estimator (with the efficient weighting scheme) is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\beta}_m - \beta_m) \xrightarrow{d} \mathcal{N}(0, (G'M^{-1}G)^{-1})$$

where $G = E\left[\frac{\partial}{\partial \beta_m} \triangle X_i' g_i(\beta_m)\right]$ and $M = E\left[\triangle X_i' g_i(\beta_m) g_i(\beta_m)' \triangle X_i\right]$

By delta-method, $\hat{\varepsilon}_{it}$ can be shown to be root-n consistent and asymptotically normal. The proofs follow the same arguments as the proof of Theorem 1.

5 Results for the quarterly models

Table 1 shows the estimation results for the quartely models, where both the uncorrected and the selection-bias adjusted estimates are presented. The regressors are as follows: LL = log(labour), LK = log(capital), ICSE = index of economic performance, DGDP = GDP growth rate.

Notice that the uncorrected procedure yields, in general, underestimation of significant parameters for the first three quartes, and overestimation for the last quarter. This reflects the fact that the quartely parameters for the first three quarters are estimated as diferences with the fourth quarter's effects, which is taken as the reference quarter, since the annual selection information is received in the fourth quarter.

6 Conclusions

A semiparametric estimating procedure was developed to correct for sample selection of quarterly data. The selection mechanism was known, but its frequency was different from the frequency of the data, since firms to be included in the sample were selected annually, and then would stay in the sample during all the four quarters after selection.

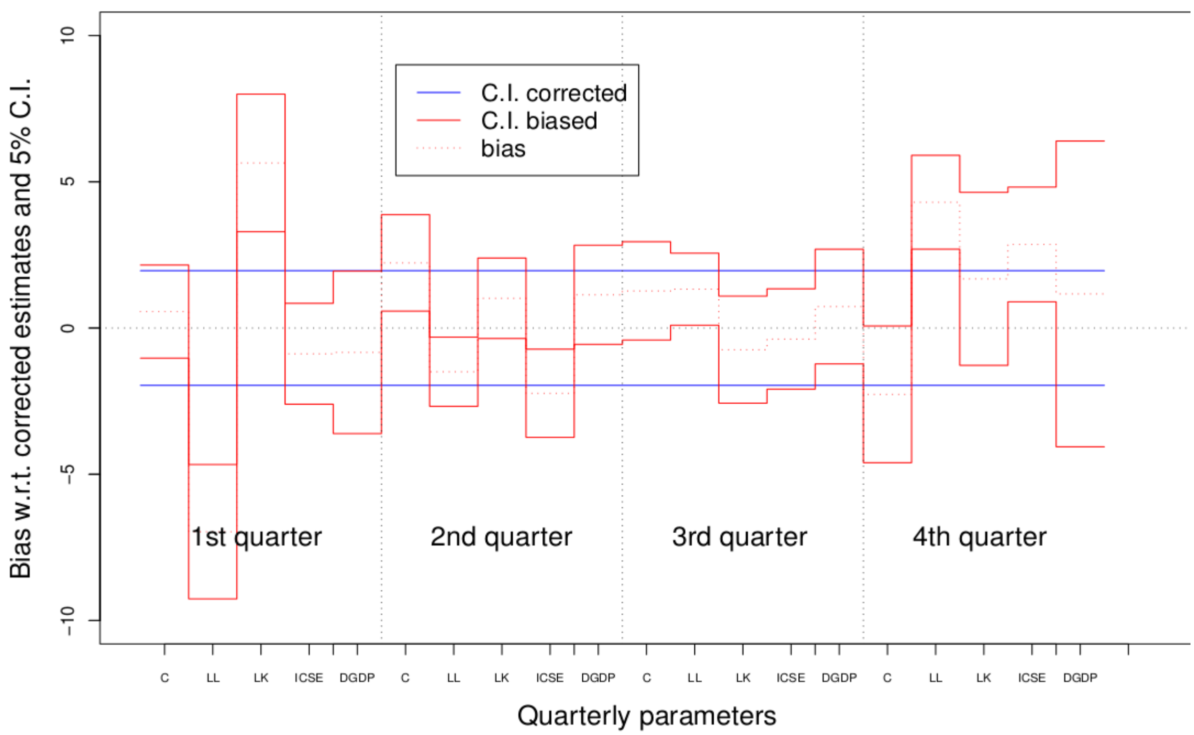
Estimation of annual models first allowed to build appropriate kernel weights, which were later combined into the quarterly model, yielding a consistent and asymptotically normal estimator, with sample selection correction, for the quarterly parameters.

The semiparametric procedure is similar to Kyriazidou (1997), but considers nonlinear specification for the regression model, and including the information about the selection mechanism, which for this specific data-set is completely known.

Table 1: Estimation results. LL = log(labour), LK = log(capital), ICSE = index of economic performance, DGDP = GDP growth rate.

	Uncorr (U)			GMM (C)			
Variable	Estimate	SE	pvalue	Estimate	SE	pvalue	(C-U)/U
Q1							
C	-3.7976	0.8018	0	-4.3514	0.9867	0	0.1458
LL	0.8787	0.0382	0	1.1057	0.0326	0	0.2583
LK	0.1347	0.0313	0	-0.0126	0.0261	0.63	-1.0935
ICSE	0.0092	0.0022	0	0.0114	0.0025	0	0.2391
DGDP	0.0086	0.0034	0.011	0.0106	0.0024	0	0.2326
Q2							
C	-1.2622	0.2177	0	-1.838	0.2587	0	0.4562
LL	0.6816	0.0649	0	0.8425	0.1076	0	0.2361
LK	0.4245	0.0331	0	0.3765	0.0472	0	-0.1131
ICSE	0.0037	0.001	0	0.0066	0.0013	0	0.7838
DGDP	0.005	0.0032	0.12	0.0008	0.0037	0.82	-0.8400
Q3							
C	-1.0503	0.2494	0	-1.4196	0.2908	0	0.3516
LL	0.6221	0.0421	0	0.5332	0.067	0	-0.1429
LK	0.3183	0.0312	0	0.343	0.0334	0	0.0776
ICSE	0.0056	0.0007	0	0.0059	0.0008	0	0.0536
DGDP	0.0024	0.0015	0.105	0.0013	0.0015	0.373	-0.4583
Q4							
C	-0.7441	0.2987	0.013	-0.176	0.2505	0.482	-0.7635
LL	0.5295	0.0442	0	0.2973	0.054	0	-0.4385
LK	0.3414	0.0439	0	0.2925	0.0291	0	-0.1432
ICSE	0.006	0.0007	0	0.004	0.0007	0	-0.3333
DGDP	-0.001	0.0016	0.543	-0.0017	0.0006	0.006	0.7000

Figure 2: Bias corrected estimates



7 Appendix: Proofs of Lemmas and Theorems

Proof of Lemma 1.

By standard ULLN, $\sup_{\beta \in B} ||q_n(\beta) - E[q_n(\beta]|| = \sup \left|\left| \frac{1}{n} \sum Z_i W_i \eta_i(\beta) - E(Z_i W_i \eta_i(\beta)\right|\right| = O(\frac{1}{\sqrt{n}})$ Checking the conditions for ULLN:

(i) $q_n(\beta)$ is continuous in β

(ii) The data are i.i.d.

(iii) The parameter space B is compact

(iv) $E \sup_{\beta \in B} ||Z_i \eta_i(\beta)|| < \infty$ (if we assume $E||z_{it}^4|| < \infty, E(\varepsilon_{it}^q)^4 < \infty$).

This guarantees that

$$E||Z_i W_i \eta_i(\beta)|| \leq \frac{1}{h_n} \sup_{v \in \mathbb{R}} k(v) E \sup_{\beta \in B} ||Z_i \eta_i(\beta)|| < \infty$$

Assuming we use a symmetric kernel:

$$\begin{aligned} \sup ||E[q_n(\beta)] - q_0(\beta)|| &\leq T \max_{t \in [1,\ldots,T]} \sup_{\beta \in B} ||E[z_{it} w_{it} \eta_{it}(\beta)] - E[z_{it} \omega_0(\triangle \varepsilon_{it}) \eta_{it}(\beta)]|| \\ &= O(h_n^2) \end{aligned}$$

Since $\lim_{n \to \infty} h_n = 0$, $\sup_{\beta \in B} ||q_n(\beta) - q_0(\beta)|| = o_p(1)$ ■

Proof of Lemma 2. Using the same arguments, we can show that $\sup_{\beta \in B} ||Q_n(\beta) - Q_0(\beta)|| = o_p(1)$. The key additional assumption is given by: $E \sup ||Z_i \frac{\partial \eta_i(\beta)}{\partial \beta'}|| < \infty$, which is satisfied given our conditions on the boundedness of moments. Similarly we can easily show that $Q_n(\beta)$ is abs. continuous at each $\beta \in B$. ■

Proof of Lemma 3. We show first that $\sqrt{nh_n} E[Z_i W_i \eta_i(\beta_0)] = o(1)$. Assuming we use a bias-reducing kernel of order κ, by a κth-order expansion of $E[z_{it} w_0(\triangle \varepsilon_{it}) \eta_{it}(\beta_0)] f_t(r)$ around $r = 0$, we get:

$$E[z_{it} w_{it} \eta_{it}(\beta_0)] = E[z_{it} w_0(\triangle \varepsilon_{it}) \eta_{it}(\beta_0)] + O(h_n^\kappa)$$

Assuming $\sqrt{nh_n} h_n^\kappa \to 0$, the bias vanishes.

We can prove the lemma by following Honoré and Kyriazidou (2000) in checking that the regularity conditions of Lyapunov's CLT for double arrays hold. Let $K_i \equiv h_n W_i = diag(k(\frac{\triangle \varepsilon_{i2}}{h_n}), \ldots k(\frac{\triangle \varepsilon_{iT}}{h_n}))$ and define

$$s_i \equiv \tau' \frac{Z_i K_i \eta_i(\beta_0) - E[Z_i K_i \eta_i(\beta_0)]}{\sqrt{h_n}}$$

for any vector of constants τ such that $\tau'\tau = 1$.

Then $\sqrt{nh_n}\tau'(q_n(\beta_0) - E\left(q_n(\beta_0)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i$.

Clearly, $E(s_i) = 0$ and $E(s_i s_i') = h_n[\tau' E(Z_i W_i \eta_i(\beta_0)\eta_i(\beta_0)' W_i Z_i')\tau] -$
$- h_n \tau' E[Z_i W_i \eta_i(\beta_0)] E[Z_i W_i \eta_i(\beta_0)]'\tau = \tau'\Lambda\tau + O\left(h_n\right) = \tau'\Lambda\tau + o(1) = O(1)$

Moreover, for any $\gamma \in (0,1)$:

$$\sum_{i=1}^{n} E\left[||\frac{s_i}{\sqrt{n}}||^{2+\gamma}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{||s_i||^{2+\gamma}}{\sqrt{n}^{\gamma}}\right] = O\left(\frac{1}{\sqrt{nh_n}}\right)^{\gamma} = o(1)$$

Lyapunov's theorem together with Cramér-Wold device implies that $\sqrt{nh_n}(q_n(\beta_0) - E\left(q_n(\beta_0)\right) \xrightarrow{d} \mathcal{N}(0,\Lambda)$ ■

Now we can use the Lemmas 1-3 to prove Theorem 1.

Proof of Theorem 1 (Consistency and asymptotic normality). Because β_0 is in the interior of the compact set B, $q_0(\beta_0)'V_0 q_0(\beta_0) < q_0(\beta)'V_0 q_0(\beta)$ for all $\beta \in B$, $\sup ||q_n(\beta) - q_0(\beta)|| = o_p(1)$ and $q_0(\beta)$ is continuous at each $\beta \in B$, $\hat{\beta}_n$ is consistent.

Moving on to asymptotic normality we want to get the asymptotic distribution of:

$$\hat{\beta}_n = \arg\min q_n(\beta)' V_n q_n(\beta)$$

FOC gives:

$$Q_n(\hat{\beta}_n)' V_n q_n(\hat{\beta}_n) = 0$$

By Mean Value Theorem we get:

$$Q_n(\hat{\beta}_n)' V_n [q_n(\beta_0) + Q_n(\beta^*)(\hat{\beta}_n - \beta_0)] = 0$$

for some β^* between $\hat{\beta}_n$ and β_0. Hence,

$$\sqrt{nh_n}(\hat{\beta}_n - \beta_0) = -[Q_n(\hat{\beta}_n)' V_n Q_n(\beta^*)]^{-1}[Q_n(\hat{\beta}_n)' V_n \sqrt{nh_n} q_n(\beta_0)]$$

$Q_n(\hat{\beta}_n) \xrightarrow{p} Q_0$ by lemma 2 (ULLN) and consistency of $\hat{\beta}_n$. By the same arguments $Q_n(\beta^*) \xrightarrow{p} Q_0$. Using the fact that $\sqrt{nh_n} E[q_n(\beta_0)] = o(1)$ and Lemma 3, we get:

$$\sqrt{nh_n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(0, (Q_0' V_0 Q_0)^{-1} Q_0' V_0 \Lambda V_0 Q_0 (Q_0' V_0 Q_0)^{-1}) = \mathcal{N}(0,\Sigma)$$

■

8 References

BIERENS H. J. (1987). 'Kernel estimators of regression functions'. In *Advances in Econometrics: Fifth world Congress*, Vol. 1, ed. by T. F. Bewley. Cambridge: Cambridge University Press.

HOLTZ-EAKIN D., W. NEWEY AND H. S. ROSEN (1988). 'Estimating vector autoregressions with panel data'. *Econometrica*, **56**, 1371-1395.

HONORÉ B. E. AND E. KYRIAZIDOU (2000). 'Estimation of Tobit-type models with individual-specific effects'. *Econometric Reviews*, **19**, 341-366.

KYRIAZIDOU E. (1997). 'Estimation of panel data sample selection model'. *Econometrica*, **65**, 1335-1364.

POWELL J. L. (1987). 'Semiparametric estimation of bivariate latent variable models'. Working Paper no. 8704, Social Systems Rsearch Institute, University of Wisconsin-Madison.

SANTOS SILVA J. M. C AND S. TENREYRO (2006). 'The log of gravity'. *The Review of Economics and Statistics*, **88**, 641-658.