On the estimation of exponential regression models: an integrated GMM approach

Esmeralda A. Ramalho and Joaquim J.S. Ramalho

Department of Economics and *CEFAGE, Universidade de Évora*

Abstract

Exponential regression models are widely used in economic applications. Several estimators have been proposed for these models, e.g. Gamma and Poisson-based quasi-maximum likelihood estimators for cases where regressors are exogenous and generalized method of moments (GMM) estimators based on a residual function from a transformed version of the original model for cases of endogeneity. Building on the latter approach, we propose an unifying new class of GMM estimators that nests the previous estimators and, additionally, incorporates more flexible estimators that use spline functions of the regressors or the instrumental variables to approximate the optimal instruments.

Keywords: exponential regression model, GMM, Poisson, Gamma, splines.

1 Introduction

Exponential regression models are widely used in economic applications where the variable of interest can only take nonnegative values. Recently, Santos Silva and Tenreyo (2006) have demonstrated that the standard practice of using log-transformed models often leads to inconsistent estimators for the parameters of interest. They advocate the use of quasi-maximum likelihood (QML) estimators for direct estimation of exponential models, which possess optimality properties when the skedastic function of the data generating process (DGP) is correctly specified up to a scala parameter.

In the presence of endogenous regressors, direct estimation of exponential models often follows the traditional approach in nonlinear models of employing two stage estimators that rely on the specification of an auxiliary reduced model relating endogenous and instrumental variables (IV). The alternative generalized method of moments (GMM) estimator of Mullahy (1997) circumvents this assumption, being based on a transformed model that generates a set of orthogonality conditions between the (transformed) error term and the instrumental variables. An efficient version of this estimator that uses spline functions of IV to approximate optimal instruments was recently proposed by Boes (2010).

In this paper we suggest a class of GMM estimators for exponential models, extension of that of Mullahy's (1997), that unifies the diversity of approaches for exogenous and endogenous regressors and incorporates optimal estimators for both cases where the skedastic function of the DGP is known and unknown.

2 Background

Consider a nonnegative variable of interest (y) described by

$$y = \exp(x\beta)\, u, \tag{1}$$

where x is a set of explanatory variables, β denotes the parameters of interest, and u is a disturbance term. Under assumption $E(u|\, x) = 1$, there are several methods that may be used to estimate β consistently, such as Poisson- and Gamma-based QML and nonlinear least squares (which is equivalent to normal QML). The degree of efficiency of each estimator depends on the true specification of the conditional variance of u given x.

Consider a power proportional variance function

$$V(u|x) = \delta \exp(x\beta)^{\lambda-2}, \tag{2}$$

where $\delta, \lambda > 0$, which implies:

$$V(y|x) = \delta \exp(x\beta)^{\lambda}. \tag{3}$$

Normal (NQML), Poisson (PQML), and Gamma (GQML) QML estimators assume $\lambda = 0$, $\lambda = 1$, and $\lambda = 2$, respectively, and satisfy a sample version of the conditional moment restriction $E[\rho(y, x, \beta)\, |x] = 0$, where

$$\rho(y, x, \beta) = y - \exp(x\beta) \tag{4}$$

is the residual function of (1). In particular, those estimators are defined by the first-order conditions $g(y, x, z, \beta)^{QML} = \rho(y, x, \beta) \exp(x\beta)^{1-\lambda} x$.

In cases of endogenous regressors, in the presence of a suitable z vector of IV, such that $E(u|\, z) = 1$ and $E(y|\, x, u, z) = E(y|\, x, u)$, direct IV estimation of exponential models cannot be based on $E[\rho(y, x, \beta)\, |z]$, as this expectation is only 0 for $z = x$; see Mullahy (1997). Two different approaches have been proposed in the literature. The first relies on the specification of a reduced form model, $x = f(z\pi) + e$, where π is a parameter vector associated to z, to provide a residual function $\hat{e}$ to be included as additional regressor in (1), which becomes $y = \exp(x^{*}\gamma)\, u^{*}$, where $x^{*} = (x, \hat{e})$. This additional assumption of the reduced model is avoided by the IV GMM estimator proposed by Mullahy (1997), which is based on the transformed model

$$y \exp(-x\beta) - 1 = u - 1 \tag{5}$$

that gives rise to a residual function

$$\rho(y, x, \beta)^{M} = y \exp(-x\beta) - 1 = \rho(y, x, \beta) \exp(-x\beta), \tag{6}$$

which is orthogonal to functions of z, since $E\left[\rho(y, x, \beta)^{M} |z\right] = 0$.

Actually, as it is clear from both (4) and (6), the residual function (6) may be employed to construct conditional moment models that cope with both exogenous and endogenous regressors and encompass all the estimators discussed previously, as discussed next

3 General GMM estimators for exponential models

Consider the conditional moment restrictions defined by $E\left[\rho(y,x,\beta)^M|z\right]=0$. The resulting unconditional moment indicators

$$g(y,x,z,\beta)=\rho(y,x,\beta)^M q(z), \tag{7}$$

where $q(z)$ are functions of the IV, yield GMM estimators consistent for β with a degree of efficiency that depends on $q(z)$, for which the optimal choice can be written as

$$q^*(z)=E\left[\frac{\partial\rho(y,x,\beta)^M}{\partial\beta}|z\right]\Omega^{-1}, \tag{8}$$

where $\Omega=E\left(g(y,x,z,\beta)\,g(y,x,z,\beta)'|z\right)$.

According to the information employed to estimate $q^*(z)$, different are the estimators obtained. In cases where $V(u|x)$ is specified and regressors are exogenous, $\Omega=diag\left(\delta\exp(x\beta)^{\lambda-2}\right)$, $\lambda>0$, and $E\left[\frac{\partial\rho^M(y,x,z,\beta)}{\partial\beta}|x\right]=\exp(-x\beta)E(y|x)x=x$, yielding $q^*(x)=\exp(x\beta)^{2-\lambda}x$ and the optimal moment indicators

$$g(y,x,z,\beta)^*=\rho(y,x,\beta)^M\exp(x\beta)^{2-\lambda}x. \tag{9}$$

It is clear that (9): (i) reduces to the orthogonality conditions of QML, $g(y,x,z,\beta)^{QML}$, for $\lambda=\{0,1,2\}$ (the use of optimal instruments adjusts for the differences of the two alternative conditional moment restrictions $E[\rho(y,x,\beta)|x]=E\left[\rho(y,x,\beta)^M|x\right]=0$); (ii) the original estimator of Mullahy (1997) employed under exogeneity, arises for $\lambda=2$, being thus a GQML estimator; and (iii) because 2SIV estimators are QML estimators that use the extended set of regressors x^*, an optimal version of these estimators can be obtained from (9), by replacing x by x^*, stacked with the estimating function of the first step $g(x,z,\pi)=[x-f(z\pi)]z$ (the standard 2SIV estimator, obtained for $\lambda=1$, is optimal only when the error variance possess a Poisson structure).

In cases of endogenous regressors and/or the skedastic function remains unspecified, k low order approximating functions $q^k(z)$ for $q^*(z)$ may be used to produce an efficient estimator for model (7). A possible choice for $q^k(z)$ are spline functions $q^k(z)=(1,z,...,z^s,1[(z-t_1)>0]z,...,1[(z-t_{k-s-1})>0]z)$, with t_1, ..., t_{k-s-1} the knots, $1(A)$ denoting an indicator function for event A, and s being the order of the spline; see Donald, Imbens and Newey (2009). When applied to endogenous regressors, the resulting GMM estimator coincides with the estimator proposed by Boes (2010). On the other hand, the use of approximating functions with exogenous regressors has never been exploited, but in fact solves the problem of choosing one of the alternative QML estimators when the skedastic function is unknown. In fact, the use of $q^k(x)$ promotes the inclusion of information in the estimation procedure, producing gains of precision relative to just-identified estimators that assume an incorrect form for $V(u|x)$.

4 Simulation study

This section presents an illustration of the finite sample properties of the estimators discussed previously. Our Monte Carlo experiments are based on the sample design of Santos Silva and Tenreyo (2006). The dependent variable is generated according to (1), with u following a log-normal distribution with mean 1 and variance given by (2), with λ ranging from 0 to 3. We consider $E\left(y|x\right) = \exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2\right)$, where $\beta_0 = 0$ and x_2 is a dummy variable that equals 1 with probability 0.4. In cases of exogenous regressors x_1 is a standard normal and $\beta_1 = \beta_2 = 1$. Otherwise, $x_1 = z\pi + \alpha e$, where z and e follow, respectively, a $v-$multivariate normal and a standard normal distribution, $\pi \in (0.1, 0.9)$ and $\alpha \in (0, 1)$ (note that π and α control, respectively, the quality of the IV and the degree of endogeneity of x_1) and $\beta_1 = \beta_2 = 0.5$. For exogenous regressors, we consider PQML, GQML and two estimators based on splines, that approximate, respectively, x_1 (SGMM-X) and both x_1 and x_2 (SGMM-XX). For endogenous regressors, the estimators considered are two 2SIV estimators, which use the linear reduced form of the DGP and instrument functions that assume a Poisson (P2SIV) or Gamma (G2SIV) error variance, the estimator of Mullahy (1997), and a spline GMM estimator that corresponds to the best performance in the exogenous case (SGMM-ZZ).

Figure 1 presents the bias and root mean squared error (RMSE) for estimators for exogenous regressors for several values of λ, in steps of 0.2. Although the performance in terms of bias is satisfactory and similar for all estimators, the differences in terms of RMSE become important among PQML and GQML estimators as the differences relative to the optimal choice of $\lambda = 1$ and $\lambda = 2$, respectively, increase. On the other hand, the performance of SGMM-X and SGMM-XX is similar for β_1 but not for β_2, where SGMM-XX is clearly superior, because SGMM-X approximates only for x_1 while SGMM-XX accounts also for x_2. The RMSE of the last estimator is close to that of the PQML and GQML estimators for $\lambda = 1$ and $\lambda = 2$, when they are optimal, and is the smallest in cases where the wrong error variance is assumed.

Figures 2, 3, and 4 display results for cases of endogenous regressors for $\lambda = \{0, 1, 2, 3\}$ and changing α, π, and k, respectively. While all estimators appear to be insensitive to the number of IV employed and perform similarly in terms of bias for β_2, they behave differently in which concerns the bias of β_1, presenting a decay when either α grows or π decreases, especially for SGMM estimators. The biases of SGMM estimators are compensated by an increase in precision, in such a way that their RMSE is in general close to that of the 2SIV estimator that uses the correct error variance.

5 References

Boes, S. (2010), "Count data models with correlated unobserved heterogeneity", Scandinavian Journal of Statistics, 37, 382-402.

Donald, S.G., Imbens, G.W and Newey, W.K. (2009), "Choosing instruments with conditional moment restrictions models", Journal of Econometrics, 152, 28-36.

Mullahy, J. (1997), "Instrumental-variable estimation of count data models: applications to cigarette smoking behavior", Review of Economics and Statistics, 79(4), 586-593.

Santos Silva, J.M.C. and Tenreyro, S. (2006), "The Log of Gravity", Review of Economics and Statistics, 88(4), 641-658.