Economics Letters 128 (2015) 83–86

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Optimal asymptotic least squares estimation in a singular set-up

Antonio Diez de los Rios [*]

Bank of Canada, Financial Markets Department, 234 Laurier Av. West, Ottawa (Ontario), K1A 0G9, Canada

HIGHLIGHTS

- I extend the optimal asymptotic least squares estimation framework.
- I focus on singularities in the asymptotic covariance of the distance function.
- The relationship with the maximum likelihood estimation framework is discussed.

ARTICLE INFO

Article history:
Received 18 July 2014
Received in revised form
8 December 2014
Accepted 8 January 2015
Available online 2 February 2015

JEL classification:
C12
C13

Keywords:
Asymptotic optimality
Auxiliary parameters
Minimum distance estimation
Generalized inverse
Singular covariance matrix

ABSTRACT

In this note, I extend the optimal asymptotic least squares estimation framework to deal with singularities in the asymptotic covariance of the distance function. Further, the relationship between the asymptotic least squares and maximum likelihood estimation frameworks in such a singular set-up is discussed.

1. Introduction

As noted by Gourieroux et al. (1982, 1985) (GMT hereafter), many empirical models can be formalized as a set of implicit equations $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ between the parameters of interest $\boldsymbol{\theta}$, and a set of auxiliary parameters $\boldsymbol{\pi}$, for which a consistent and asymptotically normal estimate $\widehat{\boldsymbol{\pi}}$ is available. In particular, these authors suggest estimating the structural parameters by trying to find the set of parameters $\boldsymbol{\theta}$ that makes $\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})$ as close as possible to zero, in the metric of a given weighting matrix $\mathbf{W}_T$. This approach, called asymptotic least-squares (ALS) delivers strongly consistent and asymptotically normal estimates, and it is especially appealing when the function $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})$ is linear in the set of parameters of interest because in that case, the ALS estimator is known in closed form.[1]

It is possible to select $\mathbf{W}_T$ such that the resulting ALS estimator of the parameters of interest is optimal in the sense that the difference between the asymptotic variance of the optimal ALS estimator and another ALS estimator based on any other quadratic form in the same distance function is negative semidefinite. In particular, GMT show that the optimal weighting matrix is equal to the inverse of the asymptotic covariance matrix of the set of implicit constraints between the reduced-form and the parameters of interest.

However, there are situations where such a covariance matrix has a reduced-rank structure and, therefore, the definition of an optimal ALS estimator provided by GMT breaks down. This can occur when the parameters of interest are subject to some equality constraints. For example, Diez de los Rios (forthcoming) shows that such a singularity arises naturally in the context of the estimation

* Tel.: +1 613 782 8500; fax: +1 613 782 7819.
E-mail address: diez@bankofcanada.ca.

[1] Applications of the ALS principle include, i.e., testing the rank of a matrix (Cragg and Donald, 1997), testing the specification of asset pricing models (Barone-Adesi et al., 2004), the estimation of dynamic stochastic general equilibrium models (Christiano et al., 2005), the estimation of dynamic games models (Pesendorfer and Schmidt-Dengler, 2008), and the estimation of Gaussian dynamic term structure models (Diez de los Rios, forthcoming).

http://dx.doi.org/10.1016/j.econlet.2015.01.006

Gaussian dynamic term structure models when the factors are chosen to be linear combinations of the yields as proposed by Joslin et al. (2011) given that the model has to price these portfolios of yields as well (which, in turn, restricts the parameters of interest).

In this note I borrow from Peñaranda and Sentana (2012), who study the problem of obtaining an optimal generalized method of moments (GMM) estimator when the asymptotic variance of the moment conditions is singular in the population, to extend the theory of optimal ALS estimation to cover the singular set-up (Section 2). In addition, I derive the conditions under which optimal ALS estimation in a singular set-up is asymptotically equivalent to maximum likelihood estimation (Section 3). An appendix with auxiliary results can be obtained upon request.

2. Asymptotic least squares estimation in a singular setup

Prior to defining an optimal ALS estimator in a singular set-up, I introduce some formal notation and briefly review the estimation framework proposed by GMT.

2.1. Asymptotic distribution of ALS estimators

Let $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^K$ be the vector of parameters of interest, and let $\boldsymbol{\pi} \in \boldsymbol{\Pi} \subset \mathbb{R}^H$ be the vector of auxiliary parameters. Both sets of parameters are related through a system of implicit equations of the form $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})$ is a $(G \times 1)$ twice continuously differentiable distance function with Jacobians given by

$$\mathbf{G}_\pi(\boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{\partial \mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})}{\partial \boldsymbol{\pi}'}, \qquad \mathbf{G}_\theta(\boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{\partial \mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}.$$

Let $\mathbf{p}(\boldsymbol{\theta})$ be a function satisfying $\mathbf{g}[\mathbf{p}(\boldsymbol{\theta}), \boldsymbol{\theta}] = \mathbf{0}$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, which implies that the system of implicit equations $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ has a unique solution for $\boldsymbol{\pi}$ given $\boldsymbol{\theta}$, and let this solution be given by $\boldsymbol{\pi} = \mathbf{p}(\boldsymbol{\theta})$. For example, the function $\mathbf{p}(\boldsymbol{\theta})$ can be thought of as the set of auxiliary parameters implied by the set of parameters $\boldsymbol{\theta}$.

Let $\widehat{\boldsymbol{\pi}}$ denote a strongly consistent and asymptotically normal estimator of the auxiliary parameters, such that as $T \to \infty$, $\widehat{\boldsymbol{\pi}} \to \boldsymbol{\pi}^0 = \mathbf{p}(\boldsymbol{\theta}^0)$, P_{θ^0} almost surely; and $\sqrt{T}(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0) \xrightarrow{d} N\left[\mathbf{0}, \mathbf{V}_\pi(\boldsymbol{\theta}^0)\right]$, where T denotes the number of observations in the sample and $\boldsymbol{\theta}^0$ denotes the true value of the parameters of interest; i.e., $\mathbf{g}(\boldsymbol{\pi}^0, \boldsymbol{\theta}^0) = \mathbf{g}\left[\mathbf{p}(\boldsymbol{\theta}^0), \boldsymbol{\theta}^0\right] = \mathbf{0}$.

GMT propose to minimize a quadratic form in the distance function evaluated at the estimates of the auxiliary parameters, $\widehat{\boldsymbol{\pi}}$:

$$\widehat{\boldsymbol{\theta}}_{ALS} = \arg\min_{\boldsymbol{\theta}} \mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta})' \mathbf{W}_T \mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}), \tag{1}$$

where $\mathbf{W}_T$ is a positive semi-definite weighting matrix that possibly depends on the observations. For this reason, the ALS estimation framework is also known as minimum distance estimation, and $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})$ is sometimes referred to as a distance function.[2]

Intuitively, the ALS estimation framework is especially appealing when (i) $\widehat{\boldsymbol{\pi}}$ can be obtained using linear regression methods and (ii) $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})$ is linear in the set of parameters of interest. In those cases, the solution to the optimization problem in (1) is known in closed form (see, for example, the empirical applications of Gourieroux et al., 1989, and Diez de los Rios, forthcoming).

Importantly, the ALS estimates are consistent and asymptotically normal. Specifically, let $\mathbf{W}_T$ converge P_{θ^0} almost surely to $\mathbf{W}$, a non-stochastic semi-definite weighting matrix of size G, and rank greater than or equal to K. If the true values of the parameters of interest and auxiliary parameters, $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$, both belong to the interior of $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$, respectively, and $\mathbf{G}_\theta' \mathbf{W} \mathbf{G}_\theta$ evaluated at $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$ is non-singular (which implies that the rank of $\mathbf{G}_\theta(\boldsymbol{\pi}^0, \boldsymbol{\theta}^0) = K$ and that $K \leq G$), then (see GMT for the proof) $\widehat{\boldsymbol{\theta}}_{ALS}$ is strongly consistent for every choice of $\mathbf{W}_T$, and its asymptotic distribution is given by

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{ALS} - \boldsymbol{\theta}^0) \xrightarrow{d} N\left[\mathbf{0}, (\mathbf{G}_\theta' \mathbf{W} \mathbf{G}_\theta)^{-1} \mathbf{G}_\theta' \mathbf{W} \mathbf{G}_\pi \mathbf{V}_\pi \mathbf{G}_\pi' \mathbf{W} \mathbf{G}_\theta (\mathbf{G}_\theta' \mathbf{W} \mathbf{G}_\theta)^{-1}\right], \tag{2}$$

where the various matrices in this equation are evaluated at $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$.

2.2. Optimal ALS estimation

As in the case of (overidentified) GMM estimation, it is possible to choose an "optimal" weighting matrix, in the sense that the difference between the asymptotic variance of the resulting ALS estimator and another ALS estimator based on any other quadratic form in the same distance function is positive definite. In particular, GMT show that when $\mathbf{G}_\pi \mathbf{V}_\pi \mathbf{G}_\pi'$ and $\mathbf{G}_\theta'(\mathbf{G}_\pi \mathbf{V}_\pi \mathbf{G}_\pi')^{-1}\mathbf{G}_\theta$ are non-singular when evaluated at $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}^0$ (which implies that the rank of $\mathbf{G}_\pi(\boldsymbol{\pi}^0, \boldsymbol{\theta}^0) = G$ and that $G \leq H$), then an optimal ALS exists and corresponds to the choice of a weighting matrix $\mathbf{W}_T$ that converges to $\mathbf{W} = (\mathbf{G}_\pi \mathbf{V}_\pi \mathbf{G}_\pi')^{-1}$. Note that, by the delta method, the optimal weighting matrix is simply the inverse of the asymptotic covariance of the distance function evaluated at the estimates of the auxiliary parameters $\mathbf{V}_g(\boldsymbol{\theta}^0) = avar\left[\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)\right] = \mathbf{G}_\pi(\boldsymbol{\pi}^0, \boldsymbol{\theta}^0)\mathbf{V}_\pi(\boldsymbol{\theta}^0)\mathbf{G}_\pi'(\boldsymbol{\pi}^0, \boldsymbol{\theta}^0)$. Specifically,

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{OALS} - \boldsymbol{\theta}^0) \xrightarrow{d} N\left[\mathbf{0}, (\mathbf{G}_\theta' \mathbf{V}_g^{-1} \mathbf{G}_\theta)^{-1}\right], \tag{3}$$

where $\widehat{\boldsymbol{\theta}}_{OALS}$ denotes the optimal ALS estimator.

Note, however, that using $\mathbf{W} = \mathbf{V}_g(\boldsymbol{\theta}^0)^{-1}$ is not feasible given that the weighting matrix needs to be evaluated at the true value of the parameters of interest, $\boldsymbol{\theta}^0$ (which is unknown). Instead, as in the case of optimal GMM estimation, it is possible to replace $\boldsymbol{\theta}^0$ in $\mathbf{V}_g(\boldsymbol{\theta}^0)$ with an initial consistent estimator of $\boldsymbol{\theta}^0$ (e.g., the ALS estimator that uses the identity matrix as the weighting matrix) to obtain an asymptotically equivalent but feasible two-step ALS estimator. Further, the minimized value of the ALS criterion function scaled by T has an asymptotic χ^2 distribution with degrees of freedom equal to $G - K$.

Finally, as noted by Kodde et al. (1990), if (i) the system of relationships $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ is complete,[3] and (ii) $\boldsymbol{\pi}$ is estimated by ML, or a method asymptotically equivalent to ML, then the optimal ALS estimator is asymptotically equivalent to the ML estimator of $\boldsymbol{\theta}$.

2.3. Optimal ALS estimation: the singular case

Unfortunately, GMT's definition of optimal ALS estimation breaks down when $\mathbf{V}_g(\boldsymbol{\theta}^0)$ has a reduced-rank structure, which occurs when (i) $G > H$, that is, the dimension of the distance function $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})$ is larger than the dimension of the vector of auxiliary parameters; and/or (ii) $\mathbf{V}_\pi(\boldsymbol{\theta}^0)$ has a reduced-rank structure.

An empirically relevant example (see, for example, Diez de los Rios, forthcoming) where $\mathbf{V}_g(\boldsymbol{\theta}^0)$ is singular is the case where the parameters of interests are constrained and the system $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ is complete (i.e., $G = H$ and $\mathbf{G}_\pi$ has full rank). Specifically, let $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ denote the set of S restrictions that $\boldsymbol{\theta}$ is subject to. Further, any restriction on $\boldsymbol{\theta}$ will constrain $\boldsymbol{\pi}$ as well given that, without

[2] Specifically, (classical) minimum distance estimation refers to the case where the distance function has the form $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \boldsymbol{\pi} - \mathbf{p}(\boldsymbol{\theta})$ (see, e.g., Chamberlain, 1982).

[3] The system $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ is complete if the dimension of the set of reduced-form parameters, $\boldsymbol{\pi}$, is equal to the dimension of $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta})$; and the Jacobian $\mathbf{G}_\pi(\boldsymbol{\pi}, \boldsymbol{\theta})$ has full rank (i.e., rank$[\mathbf{G}_\pi(\boldsymbol{\pi}_0, \boldsymbol{\theta}_0)] = G$) when evaluated at the true value. For example, the (classical) minimum distance function $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \boldsymbol{\pi} - \mathbf{p}(\boldsymbol{\theta})$ satisfies these conditions.