economics letters

Economics Letters 60 (1998) 11–18

Overparameterization in the seminonparametric density estimation

Ming Liu[a,*], Harold H. Zhang[b]

[a]Department of SE & EM, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
[b]Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 5 November 1997; accepted 3 March 1998

Abstract

The seminonparametric (SNP) density estimation proposed by Gallant and Nychka (1987) and Gallant and Tauchen (1989) has been applied in many econometric analyses. In this paper, we show that the information matrix may become singular when an SNP model is overparameterized. This singularity problem may occur even when a model selection criterion penalizes the size of a model, and thus cause problems in the sieves expansion model selection process. We propose to use the likelihood ratio test to safeguard against such overparameterization. © 1998 Elsevier Science S.A.

Keywords: Information matrix; Singularity; SNP density estimation

JEL classification: C14; C22; C52

1. Introduction

Since it was proposed in Gallant and Nychka (1987) and Gallant and Tauchen (1989), the Seminonparametric (SNP) density estimation has been widely used in the econometric analysis of time series data. For instance, Brunner (1992) uses the SNP to examine features of the real GNP; Gallant et al. (1991) apply the method to fit exchange rate series; Gallant et al. (1990) utilize the SNP to create the conditional version of the Hansen–Jagannathan bound for asset pricing models; Gallant et al., 1992, 1993 employ the method to investigate the relationship between prices and trading volume of the Standard & Poors 500 composite index; and Tauchen et al. (1996) further extend their dynamic analysis of the volume and return relationship to a panel of individual stocks. The SNP model consists of a polynomial expansion, with the leading term being a VAR Gaussian. Higher order terms of the polynomial expansion accommodate shape departures from the Gaussian density and the conditional dependence of the moments of the density on the history of the process. Additionally, the SNP model can fit not only the conditional heterogeneity associated with the first and second moments as parametric models like ARCH does but it can also accommodate more general forms of conditional heterogeneity associated with the higher order moments.

More recently, the SNP density estimate has been used as an auxiliary model in the method of moments estimation of structural models where the explicit likelihood function is not available.

*Corresponding author. Tel.: 0086 852 2609 8310; fax: 0086 852 2603 5505; e-mail: mliu@se.cuhk.edu.hk

PII: S0165-1765(98)00077-9

Because of the nonparametric flexibility of the SNP model, the estimation is asymptotically efficient and is thus called the efficient method of moments (EMM) (see Gallant and Tauchen, 1996, and Gallant and Long, 1996). The EMM was first proposed in Bansal et al., 1995; Tauchen et al. (1996) and Gallant and Tauchen (1996). It has since appeared in many applications, such as the estimation of stochastic volatility models (Anderson and Lund, 1996, 1997; Ghysels and Jasiak, 1994; Gallant et al., 1996 and Gallant and Tauchen, 1997, among others) and the equilibrium asset pricing models (Tallarini and Zhang, 1996).

Estimation of the SNP model is accomplished by using the standard maximum likelihood method. The preferred model is selected based on typical model selection criteria such as Bayesian information criterion, Akaike information criterion, or Hannan and Quinn information criterion, among others. In a more recent study, Liu and Zhang (1997) propose a chi-square test which utilizes the property that the scores of the SNP model should be Martingale when the model is correctly specified. An important condition in the process of searching for the most preferred model is that the information matrix for each model should be nonsingular. Violation of the nonsingularity condition will result in inefficient estimates and the inability to obtain the variance–covariance matrix in the SNP density estimation. Furthermore, a singular information matrix creates problems for the EMM estimation. This is because the weighting matrix in the objective function of the EMM estimation is the inverse of the information matrix of the SNP auxiliary model.

In this paper, we present examples in which the singularity of information matrix occurs as a result of overparameterizing the SNP model in certain dimensions. We propose that the likelihood ratio test be applied to detect the possible overparameterization in the SNP density estimation. In Section 2, we provide an overview of the SNP method. Section 3 presents two examples of a singular information matrix due to overparameterization in the SNP model expansion. In Section 4, we propose the likelihood ratio test to detect the overparameterization.

2. The SNP density estimator

The SNP density is a member of a class of parameterized conditional densities

$$\mathscr{H}_K = \{ f_K(y|x, \theta) : \theta = (\theta_1, \theta_2, \cdots, \theta_{l_K})\}, \tag{1}$$

which expands $\mathscr{H}_1 \subset \mathscr{H}_2 \subset \cdots$ as K increases. As K becomes large enough, the SNP density could accommodate various kinds of nonlinearities in time series data.

Specifically, the K^{th} model on the hierarchy is given by

$$f_K(y_t|x_{t-1}, \theta) = \frac{\{P_K[r^{-1}_{x_{t-1}}(y_t - \mu_{x_{t-1}}), x_{t-1}]\}^2 \phi[r^{-1}_{x_{t-1}}(y_t - \mu_{x_{t-1}})]}{r_{x_{t-1}} \int [P_K(u, x_{t-1})]^2 \phi(u)\, \mathrm{d}u}, \tag{2}$$

where $P_K(.,.)$ is a Hermite polynomial given as follows:

$$P_K(z, x) = \sum_{\alpha=0}^{K_z} \left(\sum_{|\beta|}^{K_x} a_{\alpha\beta} x^{\beta} \right) z^{\alpha}, \tag{3}$$

$z_t = r^{-1}_{x_{t-1}}(y_t - \mu_{x_{t-1}})$ where μ_x and r_x are the location and the scale function respectively, and $\phi(z) = \sqrt{2\pi} e^{-z'z/2}$. The location function μ_x is affine in x:

$$\mu_{x_{t-1}} = b_0 + b'x_{t-1}, \tag{4}$$

while the scale function r_x is affine in the absolute values of x:

$$r_{x_{t-1}} = \rho_0 + \rho'|x_{t-1}| \, .$$

Denote by $\mathbf{a}=[a_{\alpha\beta}]'$, $\alpha+\beta>0$ the coefficients of the Hermite polynomial, by $\mathbf{b}=(b_0, b')'$ the coefficients of the location function, and by $\mathbf{c}=(\rho_0, \rho')'$ the coefficients of the scale function. The vector of parameters of the SNP density can then be represented as $\theta=(\mathbf{a}', \mathbf{b}', \mathbf{c}')'$. To achieve identification, the coefficient $a_{0,0}$ is set to 1. The tuning parameters are the number of lags in the location function (L_μ), the scale function (L_r), the Hermite polynomial (L_p), and the degrees of the polynomial in z (K_z) and in x (K_x). In our examples discussed below, we assume that x is one dimensional. The nonsingular information matrix in the SNP density estimation thus means that the following matrix has full rank,

$$\mathscr{I}^0_K = \mathscr{E}\left[\frac{\partial}{\partial\theta} \log f_K(y_{t+\tau}|x_{t-1}, \theta^0)\right]\left[\frac{\partial}{\partial\theta} \log f_K(y_t|x_{t-1}, \theta^0)\right]'$$

where θ^0 is the vector of pseudo-true parameters. When the SNP model is properly parameterized, the above information matrix is nonsingular. However, when the model is overparameterized along certain dimensions, the information matrix may become singular.

Singularity of an information matrix does not necessarily imply nonidentifiability. Sargan (1983), for example, provides an example where a singular information matrix coexists with identifiability. Lee (1993) demonstrates a case of similar irregularity in the inference of the stochastic frontier function. Hall (1990) shows that singularity of the information matrix could occur in the context of Lagrange Multiplier tests with the seminonparametric alternatives.

3. Two examples of overparameterization

In this section, we present two examples of SNP models with a singular information matrix due to overparameterization in certain dimensions. The overparameterized model can be reduced to the true model when the additional parameters are set to 0. Denote by θ^a the vector of coefficients and by (L^a_μ, L^a_r, L^a_p, K^a_z, L^a_x) the tuning parameters of the overparameterized SNP model. For notational convenience, we denote

$$B(x_{t-1}) = r_{x_{t-1}} \int [P_K(u, x_{t-1})]^2 \phi(u) \, du,$$

$$f_{K,t} = f_K(y_t|x_{t-1}, \theta),$$

$$z_t = r^{-1}_{x_{t-1}}(y_t - \mu_{x_{t-1}}) \, .$$

The scores for concerned dimensions of the parameters are as follows:

$$s_{a_{\alpha\beta}} = \frac{2\phi(z_t)P_K(z_t, x_{t-1})}{B^2(x_{t-1})f_{K,t}},$$

$$\left(x_{t-1}^{\beta} z_t^{\alpha} B(x_{t-1}) - P_K(z_t, x_{t-1}) r_{x_{t-1}} \int P_K(u, x_{t-1}) x_{t-1}^{\beta} u^{\alpha} \phi(u)\, \mathrm{d}u\right), \tag{7}$$

$$s_{b_0} = \frac{1}{r_{x_{t-1}} P_K(z_t, x_{t-1})} \left(P_K(z_t, x_{t-1}) z - 2P_{Kz}(z_t, x_{t-1})\right). \tag{8}$$

3.1. Case 1

Suppose that the true data generating process can be represented by the SNP model with $K_x = 0$, $L_r = 0$, $L_p = 0$, and $L_\mu > 0$.[1] In this case, the Hermite polynomial $P_K(z, x)$ contains only terms in z,

$$P_K(z, x) = \sum_{\alpha=0}^{K_z} a_\alpha z^\alpha, \ a_0 = 1 .$$

We further assume that $P_K(z, x)$ contains only the even order terms, i.e., $a_\alpha = 0$ for $\alpha = 1, 3, \cdots, K_z - 1$ where K_z is an even integer. However, in the model selection process, we overparameterize the SNP model with $K_z^{\,a} = K_z + 1$ and without suppressing the odd order terms, but all the other tuning parameters are properly chosen. In this case the singularity problem emerges. First we note in this case, $r_{x_{t-1}}$ is now a constant, and

$$B(x_{t-1}) = Cr ,$$

where r denotes $r_{x_{t-1}}$ and C is a constant. Substituting the above into Eqs. (7) and (8), we have

$$s_{a_\alpha} = \frac{2\phi(z_t)P_K(z_t)}{C^2 r f_{K,t}} \left(C z_t^{\alpha} - P_K(z_t) \int P_K(u) u^{\alpha} \phi(u)\, \mathrm{d}u\right), \tag{10}$$

$$s_{b_0} = \frac{\phi(z_t)P_K(z_t)}{C r^2 f_{K,t}} \left(P_K(z_t) z_t - 2P_{Kz}(z_t)\right),$$

where P_{Kz} denotes the derivative of $P_K(z)$ with respect to z.

Because for the standard normal density $\phi(u)$ all the odd order moments equal to 0 and $P_K(u)$ contains even-order terms only, we have the scores for the odd order terms as follows:

$$s_{a_\alpha} = \frac{2\phi(z_t)P_K(z_t)}{C r f_{K,t}} z_t^{\alpha} . \tag{12}$$

The score for b_0, s_{b_0}, also consists of odd order terms only. It can thus be expressed as a linear combination of s_{a_α}, for $\alpha = 1, 2, \cdots, K_z^{\,a}$. We can summarize the result in the following lemma.

[1] Case 2.1 of Hall (1990) corresponds to this case with $K_z = 0$.

Lemma 3.1. *In the case described above, there exists a set of constants* c_i, $i=1, 2,\cdots, K_z^a$, *not all equal to zero, such that*

$$s_{b_0} = c_1 s_{a_1} + c_2 s_{a_2} + \cdot\cdot\cdot + c_{K_z^a} s_{a_{K_z^a}}.$$

Proof. Realize that s_{b_0} is a linear combination of $(\phi(z_t)P_K(z_t))/(f_{K,t})z_t^\alpha$ and s_{a_α} is proportional to $(\phi(z_t)P_K(z_t))/(f_{K,t})z_t^\alpha$ where α takes odd integers. Thus, choosing $c_1, c_3,\cdot\cdot\cdot, c_{K_z^a}$ appropriately and setting $c_2 = c_4 = \cdot\cdot\cdot = c_{K_z} = 0$, we obtain the above outcome.

The result indicates that the scores of the overparameterized SNP model are collinear. The information matrix generated by the scores is thus singular.

3.2. Case 2

We now present a more general case of the singular information matrix due to overparameterization in the SNP models. Suppose that an SNP model with tuning parameters L_μ, L_r, L_p, K_z, and K_x is exactly identified. However, one overfits the SNP model by setting $K_x{}^a = K_x + 1$. Since the overparameterized SNP model subsumes the exactly identified model, the parameter space of the latter is a subset of the parameter space of the former. Setting the additional parameters to 0, we can map the exactly identified family of models to the overparameterized family of models. Specifically, in this case, we can set the extra parameters $\theta^a_{\alpha K_x^a} = 0$. Identification of the SNP model implies that the scores of the overparameterized model $s^a_{a_{\alpha\beta}} = s_{a_{\alpha\beta}}$, for $\beta \le K_x$. The scores for θ^a_{α,K_x^a}, $s^a_{\alpha K_x^a}$ can be obtained using Eq. (7) by setting K_x at K_x^a. We examine the following summation:

$$S = \sum_{\alpha=0}^{K_z} \left(\sum_{\beta=1}^{K_x^a} a_{\alpha,\beta-1} \right) s^a_{\alpha\beta}. \tag{13}$$

Substituting scores given by Eq. (7) to the above yields:

$$S = \frac{2x_{t-1}\phi(z_t)P_K(z_t, x_{t-1})}{B^2(x_t - 1)f_{K,t}},$$

$$\left(B(x_{t-1}) \sum_{\alpha=0}^{K_z} \sum_{\beta=0}^{K_x} a_{\alpha\beta} x_{t-1}^\beta z_t^\alpha - P_K(z_t, x_{t-1}) r_{x_{t-1}} \int P_K(u, x_{t-1}) \sum_{\alpha=0}^{K_z} \sum_{\beta=0}^{K_x} a_{\alpha\beta} x_{t-1}^\beta u^\alpha \phi(u)\, \mathrm{d}u \right)$$

$$= \frac{2x_{t-1}\phi(z_t)P_K(z_t, x_{t-1})}{B^2(x_{t-1})f_{K,t}} (B(x_{t-1})P_K(z_t, x_{t-1}) - P_K(z_t, x_{t-1})B(x_{t-1}))$$

$$= 0$$

The above result shows that a linear combination of a subset of the scores is 0. We thus have the following lemma.

Lemma 3.2. *In the example given above, the scores for the overparameterized SNP model are collinear with each other*.

Proof. Setting the constants associated with scores $s_{\alpha\beta}$ to $a_{\alpha,\beta-1}$ and letting the constants for the rest of the scores be zero, we have a linear combination of all the scores equal to $\Sigma_{\alpha=0}^{K_z}\left(\Sigma_{\beta=1}^{K_x^a} a_{\alpha,\beta-1}\right)s_{\alpha\beta}^a$, which is zero as shown above. The scores are thus collinear with each other.

As a consequence of the above result, the information matrix of the SNP model will be singular.

4. Test for overparameterization

As the above examples show, when an SNP model is overparameterized, the information matrix of the model is singular. As a result, the parameter estimates will not be efficient and the covariance matrix of the estimates will be very imprecise. Any inferences drawn in this situation can be misleading. The consequence of a singular information matrix may be particularly severe in the efficient method of moments estimation. In the EMM, we use the inverse of the information matrix as the weighting matrix for different moments to form the GMM type objective function. The information matrix will not be invertible if it is singular. It is thus necessary to prevent overparameterization in the model selection process in the SNP estimation. While in general testing for overparameterization varies depending on specific situations, for the examples presented above, we can apply the regular likelihood ratio test. The basic idea is to use reparameterization to circumvent the singular information matrix problem. As shown in Lee (1993), in certain cases when the information matrix is singular, reparameterization can be used to make the information matrix nonsingular. By reparameterizing the parameter space, he demonstrates that the use of the likelihood ratio would still be valid in his case.

In case 1, while keeping all the other parameters the same, we reparameterize the set of parameters $\boldsymbol{d}=(b_0, a_1, a_3,\cdots,a_{K_z-1}, a_{K_z^a})$ as follows,

$$\boldsymbol{d}=(B_0+c_0\sqrt{A_{K_z^a}}, A_1+c_1\sqrt{A_{K_z^a}}, A_3+c_3\sqrt{A_{K_z^a}},\cdots,A_{K_z-1}+c_{K_z-1}\sqrt{A_{K_z^a}}, \sqrt{A_{K_z^a}})\,.$$

The new set of parameters would consist of the set $(B_0, A_1, A_3,\cdots,A_{K_z-1}, A_{K_z^a})$ in place of **d**.

In case 2, the following reparameterization yields a nonsingular information matrix. Denote by δ the new set of parameters, and by A a subset of parameters in δ that captures the high order dynamics in $P(z, x)$. Then $\delta=(A'\mathbf{b}'\mathbf{c}')$, where **b** and **c** are vectors of coefficients of the original parameterization. A is related to the original parameters **a** in the following manner,

$$\boldsymbol{a}=\left(A_{0,1}+a_{0,0}\sqrt{A_{K_z,K_x^a}}, A_{1,1}+a_{1,0}\sqrt{A_{K_z,K_x^a}},\cdots,A_{K_z,K_x}+a_{K_z,K_x-1}\sqrt{A_{K_z,K_x}}, \sqrt{A_{K_z,K_x^a}}\right)\,.$$

With the above transformation, in case 1, the scores with $A_{K_z^a}$ become the type of $\frac{0}{0}$ since at the point of identification $A_{K_z^a}=0$. In case 2, the scores with A_{K_z,K_x^a} become the type of $\frac{0}{0}$ since at the point of identification $A_{K_z,K_x^a}=0$. By using the L'Hospital rule, we have a nonsingular information matrix in both cases. The detailed derivation is available upon request.

Therefore, using reparameterization, we can obtain a nonsingular information matrix. This resolves the singularity problem in the SNP density estimation. The maximum likelihood estimation of the reparameterized SNP model becomes a regular MLE problem. A likelihood ratio test can then be applied to detect overparameterization. Specifically, we can test null hypotheses H_0: $A_{K_z^a}=0$ in case 1 and H_0: $A_{K_z,K_x^a}=0$ in case 2. The tests will be equivalent to the null hypotheses H_0: $a_{K_z^a}=0$ in case 1 and H_0: $a_{K_z,K_x^a}=0$ in case 2 in the original parameterization of the SNP models.

One may argue that the likelihood ratio test proposed is not necessary, given that in practice we select the SNP model based on information criteria penalizing the size of the model. However, since different information criteria penalize the size of a model in different ways, the model selected may not reflect our concern of overparameterization. The potential damage of the overparameterization thus justifies the use of the likelihood ratio test in this instance.

Acknowledgements

We thank George Tauchen for his helpful comments.

References

Anderson, T.G., Lund, J., 1997. Estimating continuous time stochastic volatility models of the short term interest rate. Journal of Econometrics 77 (2), 343–377.

Anderson, T.G., Lund, J., 1996. Stochastic volatility and mean drift in the short rate diffusion: Sources of steepness, level, and curvature of the yield curve. Working paper, Northwestern University, Evanston, IL.

Bansal, R., Gallant, A.R., Hussey, R., Tauchen, G., 1995. Nonparametric estimation of structural models for high-frequency currency market data. Journal of Econometrics 66, 251–287.

Brunner, A.D., 1992. Conditional asymmetries in real GNP: A seminonparametric approach. Journal of Business and Economic Statistics 10 (1), 65–72.

Gallant, A.R., Hansen, L.P., Tauchen, G., 1990. Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution. Journal of Econometrics 45, 141–180.

Gallant, A.R., Hsieh, D.A., Tauchen, G., 1991. On fitting a recalcitrant series: The pound/dollar exchange rate, 1974–83. In: Barnett, W.A., Powell, J., Tauchen, G. (Eds.) Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics, Cambridge University Press, Cambridge.

Gallant, A.R., Hsieh, D.A., Tauchen, G. 1996. Estimation of stochastic volatility models with diagnostics. Journal of Econometrics, forthcoming.

Gallant, A.R., Long, J.R. 1996. Estimating stochastic differential equations efficiently by minimum chi-square. Biometrika, forthcoming.

Gallant, A.R., Nychka, D.W., 1987. Seminonparametric maximum likelihood estimation. Econometrica 55, 363–390.

Gallant, A.R., Rossi, P.E., Tauchen, G., 1992. Stock prices and volume. Review of Financial Studies 5, 199–242.

Gallant, A.R., Rossi, P.E., Tauchen, G., 1993. Nonlinear dynamic structures. Econometrica 61, 871–907.

Gallant, A.R., Tauchen, G., 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. Econometrica 57, 1091–1120.

Gallant, A.R., Tauchen, G., 1996. Which moment to match?. Econometric Theory 12, 657–681.

Gallant, A.R., Tauchen, G., 1997. Estimation of continuous time models for stock returns and interest rates. Macroeconomic Dynamics 1, 135–168.

Ghysels, E., Jasiak, J. 1994. Stochastic volatility and time deformation: An application to trading volume and leverage effects. Working paper, Universite de Montreal.

Hall, A., 1990. Lagrange multiplier tests for normality against semi-nonparametric alternatives. Journal of Business & Economic Statistics 8, 417–426.

Lee, L., 1993. Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix. Econometric Theory 9, 413–430.

Liu, M., Zhang, H.H. 1997. Specification tests in the efficient method of moments framework with applications to the stochastic volatility models. Manuscript, Chinese University of Hong Kong and Carnegie Mellon University.

Sargan, J.D., 1983. Identification and lack of identification. Econometrica 51, 1605–1633.

Tallarini, T.D., Zhang, H.H. 1996. External habit and the cyclicality of the expected stock returns. Manuscript, Carnegie Mellon University.

Tauchen, G., Zhang, H., Liu, M., 1996. Volume, volatility, and leverage: A dynamic analysis. Journal of Econometrics 74, 177–208.