Bootstrap confidence intervals in functional regression under dependence

Paula Raña*, Germán Aneiros, Juan Vilar and Philippe Vieu
Departamento de Matemáticas, Universidade da Coruña
Institut de Mathématiques, Université Paul Sabatier

Abstract

This study proposes naive and wild bootstrap procedures to construct pointwise confidence intervals for two functional regression models: the functional nonparametric regression model, considering scalar response and functional predictor, and the semi-functional partial linear regression model, in which we add linear effect of scalar covariates. By means of these two bootstrap procedures we can approximate the asymptotic distribution of the estimators in both regression models. The validity of these two methods has been proved theoretically in the setting of dependent data, assuming α-mixing conditions on the sample, and they were used to construct pointwise confidence intervals for each component of the functional regression models. A simulation study was carried out to show the performance of the proposed procedures in the functional nonparametric model, in addition to an application to electricity demand and price from the Spanish Electricity Market which illustrates its usefulness in practice for both regression models.

1 Introduction

The aim of our paper is to investigate the question of practical use of functional time series predictions by providing a bootstrapping procedure for overcoming the difficulty related to the estimation of the constants in the limit distribution. We are focus on two functional regression models: Functional Nonparametric (FNP) and Semi-Functional Partial Linear (SFPL).

We consider first the FNP model:

$$Y_i = m(\boldsymbol{\chi}_i) + \varepsilon_i, \tag{1}$$

*Corresponding author. E-mail: paula.rana@udc.es

where the process $\{(\boldsymbol{\chi}_i, Y_i)\}$ is α-mixing and identically distributed as $(\boldsymbol{\chi}, Y)$. The response, Y, is scalar while the covariate, $\boldsymbol{\chi}$, is valued in some infinite-dimensional space $\mathcal{H}$, which is endowed with a semi-metric $d(\cdot,\cdot)$. Finally, $m(\cdot)$ is an unknown smooth real-valued operator and the corresponding random errors $\{\varepsilon_i\}$ are i.i.d. as ε, and we assume that $\mathbb{E}(\varepsilon|\boldsymbol{\chi}) = 0$ and $\mathbb{E}(\varepsilon^2|\boldsymbol{\chi}) = \sigma_\varepsilon^2(\boldsymbol{\chi}) < \infty$. Given a fixed element χ of the space $\mathcal{H}$, the first part of this study focuses on inference on $m(\chi)$ in model (1); specifically, the aim is to construct confidence intervals for $m(\chi)$. We consider the the functional kernel estimator $\widehat{m}_h(\cdot)$ of the regression function $m(\cdot) = \mathbb{E}(Y \mid \boldsymbol{\chi} = \cdot)$ as Ferraty et al. (2010) and Delsol (2009).

The bootstrap algorithms are based on resampling the residuals of the model and build a bootstrap version for the regression function estimator. Two different procedures are proposed, one using naive bootstrap for homoscedastic models and also wild bootstrap when it is heteroscedastic. We must take into account the use of two bandwidths when dealing with bootstrap concerning regression.

When dealing with Semi-Functional Partial Linear Regression model (SFPL model), we will consider the next model:

$$Y_i = \boldsymbol{X}_i^T\boldsymbol{\beta} + m(\boldsymbol{\chi}_i) + \varepsilon_i,\ i = 1, \ldots, n, \tag{2}$$

where the sequence $\{(\boldsymbol{X}_i, \boldsymbol{\chi}_i, Y_i)\}$ is α-mixing. Using the notation referred in the extended version of this study, we consider the estimators $\widehat{\boldsymbol{\beta}}_h$ and $\widehat{m}_h(\cdot)$ of the vector parameter $\boldsymbol{\beta}$ and the function $m(\cdot)$ in (2) as $\widehat{\boldsymbol{\beta}}_h = (\widetilde{\mathbf{X}}_h^T\widetilde{\mathbf{X}}_h)^{-1}\widetilde{\mathbf{X}}_h^T\widetilde{\mathbf{Y}}_h$ and $\widehat{m}_h(\chi) = \sum_{i=1}^n w_h(\boldsymbol{\chi}_i, \chi)(Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_h)$, respectively, as in Aneiros and Vieu (2008).

We develop also two bootstrap procedures, for homoscedastic (naive bootstrap) and heteroscedastic models (wild bootstrap). In both cases we follow the same idea as in the FNP model, resampling the residuals of the SFPL model (2) and building the correspondent bootstrap version of the estimators for each component of the SFPL model.

2 Contributions

The present study establishes and proves theoretically the validity of the proposed bootstrap procedures and apply them to build confidence intervals. Main reference when dealing with the validity of the bootstrap in the FNP model concerns the asymptotic distribution of $\widehat{m}_h(\chi)$. We use the result by Delsol (2009), which gives this asymptotic distribution under α-mixing conditions (see its Theorem 2.7). Then, we consider the same assumptions related to its Theorem 2.7, together with Ferraty et al. (2010) to attain the validity of the bootstrap in the independent case. Using the remaining considerations and notation indicated in the extended version of this study, we present our first theorem.

Theorem 1 *Under assumptions indicated above, for the wild bootstrap procedure, we have that*

$$\sup_{y\in\mathbb{R}}\left|P^{\mathcal{S}}\left(\sqrt{nF_{\chi}(h)}(\widehat{m}^*_{hb}(\chi)-\widehat{m}_b(\chi))\leq y\right)-P\left(\sqrt{nF_{\chi}(h)}(\widehat{m}_h(\chi)-m(\chi))\leq y\right)\right|\to 0 \ a.s.$$

In addition, if the model is homoscedastic (i.e. $\sigma^2_{\varepsilon}(\cdot)=\sigma^2_{\varepsilon}$), then the same result holds for the naive bootstrap.

In the case of SFPL mode, we use assumptions established in Theorem 1 in Aneiros and Vieu (2008), to prove the asymptotic normality of $\widehat{\boldsymbol{\beta}}$ also under dependence. We may consider two different theorems, one for each part of the SFPL model.

Theorem 2 *Under assumptions mentioned above, if the model is homoscedastic, for the naive bootstrap and in other case if, in addition $|\varepsilon_i|<\infty, i=1,\ldots,n$, $F(h)^{-1}n^{-1/4+1/r}logn(loglogn)^{1/4}\to 0$, $\mathbb{E}|\eta\eta^T|<\infty$ and $\mathbb{E}|\eta|^3<\infty$, for the wild bootstrap procedure we have that*

$$\sup_{y\in\mathbb{R}}\left|P^{\mathcal{S}}\left(\sqrt{n}\mathbf{a}^T(\widehat{\boldsymbol{\beta}}^*_b-\widehat{\boldsymbol{\beta}}_b)\leq y\right)-P\left(\sqrt{n}\mathbf{a}^T(\widehat{\boldsymbol{\beta}}_b-\boldsymbol{\beta})\leq y\right)\right|\to_P 0$$

Theorem 3 *Under Assumptions mentioned above if, in addition $||\boldsymbol{X}_i||_{\infty}\leq C<\infty$, if the model is homoscedastic, for the naive bootstrap procedure, and in any case for the wild bootstrap procedure we have:*

$$\sup_{y\in\mathbb{R}}\left|P^{\mathcal{S}}\left(\sqrt{nF(h)}(\widehat{m}^*_{hb}(\chi)-\widehat{m}_b(\chi))\leq y\right)-P\left(\sqrt{nF(h)}(\widehat{m}_h(\chi)-m(\chi))\leq y\right)\right|\to_P 0$$

Those theorems had been proved theoretically and they were applied to build bootstrap confidence intervals within the two functional regression models. Simulations shown the accuracy of the proposed procedures for FNP model, meanwhile its usefulness is shown in practice through an application to electricity demand ad price in the Spanish Electricity Market using in this case both functional regression models.

3 State of the art

Considering the FNP model in the setting of independent data, Ferraty et al. (2007) obtained the asymptotic normality of a properly standardized estimator, $\widehat{m}_h(\chi)$; then, by estimating the constants involved in the standardized estimator one can construct the corresponding confidence intervals. The main drawback of this procedure is that such constants could be difficult to estimate (for some simple examples, see Proposition 1 in Ferraty et al., 2007). This drawback was overcome in Ferraty et al. (2010) by means of bootstrapping techniques, by approximating directly the distribution of the estimation error without having to estimate the constants involved in the standardized estimator. On the other hand, some studies exist in the case of dependent data $\{(\boldsymbol{\chi}_i, Y_i)\}$. For instance, Masry (2005) and Delsol (2009) obtained the asymptotic normality of a properly standardized estimator, $\widehat{m}_h(\chi)$, under α-mixing conditions. The main advantage of the results in Delsol (2009) against the ones in Masry (2005) is the fact that Delsol obtained explicit constants, which is not the case of Masry (2005). As in the setting of independent data recently referred, there exist situations where the constants given in Delsol (2009) are difficult to estimate, and this drawback could be overcome, again, through implementation of bootstrap techniques. Thus, the present study represents the extension, to the case of dependent data, of Ferraty et al. (2010), and can be found more in detail in Raña et al (to appear).

Nevertheless, considering now the SFPL model, as fas as we know, there is no preceding study in the literature regarding validity of the bootstrap in this model (even for the independent case). Moreover, it is even difficult to find applications of this kind of bootstrap procedures applied to classical partial linear regression. One can find in Liang et al. (2000) and You and Chen (2006) proposals for bootstrap approximation in partial linear regression but in the case of fixed design, independent data and regarding the linear component of the model. For all these reasons, this study is the first approach to the validity of the bootstrap procedures developed in the context of SFPL model with dependent data (and, as a particular case, to independent data), considering both linear and nonparametric parts of the model.

References

[1] Aneiros-Pérez, G. and Vieu, P. (2008). Nonparametric time series prediction: A semi-functional partial linear modeling. *J Multivariate Anal*, 99:834–857.

[2] Delsol, L. (2009). Advances on asymptotic normality in non-parametric functional time series analysis. *Statistics*, 43:13–33.

[3] Ferraty, F., Mas, A., and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical apects. *Aust. N. Z. J. Stat.*, 49:267–286.

[4] Ferraty, F., Van Keilegom, I., and Vieu, P. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scand. J. Statist.*, 37:286–306.

[5] Liang, H., Hardle, W. and Sommerfeld, V. (2000) Bootstrap approximation in a partially linear regression model, *Journal of Statistical Planning and Inference*, 91: 413–426.

[6] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.*, 115:155–177.

[7] Raña, P., Aneiros, G., Vilar, J. and Vieu, P. (2016) Bootstrap confidence intervals in functional nonparametric regression under dependence. *(To appear)*.

[8] You, J. and Chen, G. (2006) Wild bootstrap estimation in partially linear models with heteroscedasticity. *Statistics & Probability Letters*, 76: 340–348.