2014/56

Instrumental variable estimation
in functional linear models

Jean-Pierre Florens and Sébastien Van Bellegem

CORE

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium

<http://www.uclouvain.be/core>

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve, Belgium.

Tel (32 10) 47 43 04

Fax (32 10) 47 43 01

E-mail: corestat-library@uclouvain.be

<http://www.uclouvain.be/en-44508.html>

CORE DISCUSSION PAPER
2014/56

**Instrumental variable estimation
in functional linear models**

Jean-Pierre FLORENS¹ and Sébastien VAN BELLEGEM²

September 2014

Abstract

In an increasing number of empirical studies, the dimensionality measured e.g. as the size of the parameter space of interest, can be very large. Two instances of large dimensional models are the linear regression with a large number of covariates and the estimation of a regression function with many instrumental variables. An appropriate setting to analyze high dimensional problems is provided by a functional linear model, in which the covariates belong to Hilbert spaces. This paper considers the case where covariates are endogenous and assumes the existence of instrumental variables (that are functional as well). The paper shows that estimating the regression function is a linear ill-posed inverse problem, with a known but data-dependent operator. The first contribution is to analyze the rate of convergence of the penalized least squares estimator. Based on the result, we discuss the notion of “instrument strength” in the high dimensional setting. We also discuss a generalized version of the estimator, when the problem is premultiplied by an instrument-dependent operator. This extends the technology of Generalized Method of Moments to high dimensional, functional data. A central limit theorem is also established on the inner product of the estimator. The studied estimators are easy and fast to implement, and the finite-sample performance is discussed through simulations and an application to the impact of age-specific fertility rate curves on yearly economic growth in the United Kingdom.

Keywords: high dimensional model, penalized least squares, instrumental variable, functional data, fertility rate, growth.

JEL Classification: Primary C26; secondary C14

¹ Toulouse School of Economics

² Center for Operations Research and Econometrics

We thank Rytis Bagdziunas, Eric Gauthier, Frédérique Fève, Maik Schwarz and Anna Simoni for helpful comments on a preliminary version of this work. Various parts of this work have been presented during the workshop in nonparametric econometrics in Banff (2009), the CIREQ workshop on high dimensional econometrics in Montréal (2012), the Joint Statistical Meeting in San Diego (2012), the LATAM workshop in Sao Paulo (2013) and during seminars at IHS (Wien), ECORE (Brussels) and the University of Luxembourg. Comments from the participants of those meetings were greatly appreciated. Address correspondence to Sébastien Van Bellegem, Voie du Roman Pays, 34, L1.03.01, B-1348 Louvain-la-Neuve, sebastien.vanbellegem@uclouvain.be.

1 Introduction

In an increasing number of economic studies, the dimensionality of the statistical model, measured as the size of the parameter space of interest, can be very large. One canonical example is provided by the multiple linear regression model

$$y = z'\varphi + \varepsilon \tag{1.1}$$

where y is a real random variable, z is a vector of p covariates and ε is an exogenous error term. To reflect that the number of covariates p is large, asymptotic inference is usually derived under the condition that p increases with the sample size (e.g. Fan, Fan, & Lv, 2008; Belloni & Chernozhukov, 2009; Belloni, Chernozhukov, & Hansen, 2011). Standard estimators such as Generalized Least Squares (GLS) thus need to be modified because of the unboundedness of the parameter space. The present paper analyzes linear models with large or infinite dimension under endogeneity in the presence of instrumental variables. Since the dimension of the endogenous variables is large, so is the number of instrumental variables. Our study is therefore tightly connected to the recent studies on estimation in the presence of “many instrumental variables” (Morimune, 1983; Carrasco & Florens, 2000; Chao & Swanson, 2005; Hansen, Hausman, & Newey, 2008; Swanson, Chao, Hausman, Newey, & Woutersen, 2011). In these studies, the dimension of the model is kept fixed (not depending on the sample size) while the dimension of the instrumental vector can be very large or depends on the sample size. Our study below considers the situation where both the dimension of the covariates and the instruments are large.

Estimation with many instrumental variable has a long history in econometrics, see e.g. the above references and the references therein. It is recognized that the empirical performance of inference can be imprecise in this context, because of the presence of weak instruments (see e.g. the survey paper by Stock, Wright, & Yogo, 2002). This issue naturally arises in the high dimensional setting as well. This study provides a new characterization of strong or weak instruments in the statistical model. The strength of an instrumental vector results from an arbitrage among two indices of the statistical model. The first index (denoted β below) is a measure of the regularity of the unknown parameter (which is infinite dimensional) *with respect to* the eigenvalues of the covariance between the endogenous parameters and the instruments. The second index (denoted γ) can be understood as a measure of the signal-to-noise ratio, that is a ratio between the later eigenvalues onto the eigenvalues of the covariance between the instruments and the error term in the linear model. Strong instruments are such that the regularity of the target parameter (β) is large enough compared to the signal-to-noise ratio (γ). Below we show that, under this condition, the penalized least squares estimator of the parameter has the optimal minimax rate of convergence, that has been derived in Chen and Reiss (2011) in a similar context. On the contrary if the instruments are weak, the optimal rate of convergence is not reached.

Because we are interested in the asymptotic behavior of the estimator, this paper ex-

explicitly relates the sampling model to a population model that is infinite dimensional. A convenient framework to make the analysis is to rewrite and generalize model (1.1) using an inner product, that is

$$y = \langle z, \varphi \rangle_{\mathcal{F}} + \varepsilon \quad (1.2)$$

where $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ stands for the inner product between two elements of a Hilbert space $\mathcal{F}$. This model contains the finite dimensional linear model as a particular case if we consider $\mathcal{F} = \mathbb{R}^K$ with a fixed, finite dimension K . However it allows to accommodate high dimensional parameters, for instance when $\mathcal{F}$ is the space of squared summable sequences, ℓ^2 .

The statistical model (1.2) is also appropriate when covariates are not random vectors but random *curves* as it is the case in functional regression. In such a case the Hilbert space $\mathcal{F}$ is for instance the space of real, square integrable functions over $[0, 1]$, denoted $L^2[0, 1]$, equipped with the inner product defined as

$$\langle \varphi_1, \varphi_2 \rangle_{L^2[0,1]} = \int_0^1 \varphi_1(t) \varphi_2(t) dt .$$

Examples of regression with functional covariates can be found in various studies (e.g. Kunitomo, 1980; Ramsay & Silvermann, 2005; Hall & Horowitz, 2007; Ferraty & Vieu, 2006; Cardot & Johannes, 2010; Kokoszka & Horvath, 2012). We provide below an application to growth theory, in which the dependent variable is the annual growth rate per capita of United Kingdom and the independent variable is the annual fertility rate.

Our study starts from the general model (1.2) in which z is endogenous and it supposes that we observe instruments w allowing to identify the parameter φ . The study of inference shows that the problem of estimating φ in this statistical model is an ill-posed inverse problem, and inference thus relies on regularization techniques such as the penalized least squares estimation. Two recent surveys on regularization for ill-posed inverse problems in econometrics on which this paper relies on are Carrasco, Florens, and Renault (2007) and Johannes, Van Bellegem, and Vanhems (2011).

The organization of the paper is as follows. In Section 2 the asymptotic properties of the penalized least-square estimator is studied and the notion of instrument strength is discussed. The estimator can be seen as an extension of the parametric instrumental variable estimator to the high dimensional setting. The two key indices (β and γ) are defined. Section 3 studies the numerical and finite sample properties of the estimator. It shows that how the estimator is computed in practice, using a linear algorithm. This section also contains two discussions on the penalty. The first discussion shows how this penalty can be computed in a data-driven way and the second discussion shows how it is possible to impose smoothness priors such as the existence of derivatives on the solution.

Analogously to the finite-dimensional theory of the efficient estimation by the Generalized Method of Moments (GMM), Section 4 questions the efficiency of the estimator. A generalized version of the estimator is studied, when the model is premultiplied by a data-driven operator. It is proved that for a certain class of operators and under some restrictions

on the statistical model, the premultiplication does not affect the rate of convergence of the estimator. A Central Limit Theorem is then derived for the inner product $\langle \hat{\varphi}, g \rangle_{\mathcal{F}}$, if $\hat{\varphi}$ denoted the generalized penalized least squares estimator and g is a test function in $\mathcal{F}$. The Theorem also provides lower and upper bounds for the asymptotic variance.

Section 5 evaluates the performance of the estimators in an application in growth theory. Recent developments in economics have studied the impact of fertility on growth rate. The dependent variable is the age-specific fertility rate that is the density of birth with respect to mother's age. It is a continuous variable by definition, observed yearly and the framework of functional regression allows to study the marginal impact of the age of fertility onto the growth rate per capita. An appendix collects the technical proofs of results that are stated in the paper.

2 The high dimensional model and the regularized estimator

2.1 Definition of the estimator

We start by defining the statistical model in the two following assumptions¹.

Assumption 1. *y is a real random variable, z is a random vector in a Hilbert space $\mathcal{F}$ and w is a random vector in a Hilbert space $\mathcal{W}$ such that there exists a unique $\varphi \in \mathcal{F}$ that satisfies $\mathbb{E}(\{y - \langle z, \varphi \rangle_{\mathcal{F}}\}w) = 0$. If we set $u = y - \langle z, \varphi \rangle$, we also denote by Ω the covariance operator of uw , that is $\Omega = \mathbb{E}(u^2 \langle w, \bullet \rangle_{\mathcal{W}} w)$.*

Assumption 2. *We observe (y_i, z_i, w_i) for $i = 1, \dots, n$ that are independent and with the same distribution as (y, z, w) .*

In order to derive the normal equations and to find a nonparametric estimator for $\varphi \in \mathcal{F}$ it is useful to translate the above assumptions in a notation with operators. The above assumptions imply that the unknown function $\varphi \in \mathcal{F}$ satisfies:

$$Y_n = K_n \varphi + U_n \tag{2.1}$$

where $U_n = (u_1, \dots, u_n)'$, $Y_n = (y_1, \dots, y_n)'$ and $K_n : \mathcal{F} \rightarrow \mathbb{R}^n$ is a linear operator defined by $K_n \varphi = (\langle z_1, \varphi \rangle_{\mathcal{F}}, \dots, \langle z_n, \varphi \rangle_{\mathcal{F}})'$. By definition, the inner product in $\mathbb{R}^n$ is set to

$$\langle a, b \rangle_{\mathbb{R}^n} = \frac{1}{n} \sum_{i=1}^n a_i b_i$$

for any $a, b \in \mathbb{R}^n$.

¹The assumption includes an identification condition. A primitive condition is found if we consider the covariance operator $\Sigma_{ZW} = \mathbb{E}\{\langle Z, \cdot \rangle_{\mathcal{F}} W\}$. Identification of φ follows if the kernel of this covariance operator is the null set. This can be seen as the infinite dimensional extension of the common rank condition on the covariance matrices appearing in the identification theory of finite dimensional linear models.

We also introduce the operator $W_n^* : \mathbb{R}^n \rightarrow \mathcal{W} : \theta \rightarrow n^{-1} \sum_{i=1}^n \theta_i w_i$, where W_n^* stands for the adjoint of W_n . Applied to model (2.1), we get

$$W_n^* Y_n = W_n^* K_n \varphi + W_n^* U_n$$

which we rewrite

$$r_n = T_n \varphi + \varepsilon_n \tag{2.2}$$

where $r_n = W_n^* Y_n$, $\varepsilon_n = W_n^* U_n$ and where $T_n = W_n^* K_n$ is now an operator from $\mathcal{F}$ onto $\mathcal{W}$.

Getting φ from the last identity is known to be an *ill-posed inverse problem*, in the sense that the (generalized) inverse of T_n is not a bounded operator. Therefore, due to the error term ε_n , a direct inversion of T_n does not lead to a consistent estimation of φ . In order to find a consistent estimator, a regularization of the operator T_n is needed. Several options are possible among the regularization scheme studied in the cited literature (e.g. Engl, Hanke, & Neubauer, 2000). In this paper, we focus on the *Tikhonov regularization*, that is the estimator

$$\hat{\varphi}_{\alpha,n} = (\alpha_n I + T_n^* T_n)^{-1} T_n^* r_n \tag{2.3}$$

where α_n is a sequence of positive numbers converging to zero as n tends to infinity, I denotes the identity operator and T_n^* denotes the adjoint operator of T_n .

2.2 Asymptotic mean square convergence

In this section, we study the asymptotic mean square convergence of the estimator (2.3) under a set of three assumptions that we now introduce. The first assumption is about the limit of the operator T_n when n tends to infinity. We assume that the limit of this operator is well defined and also assume some speed of convergence (in the spectral norm) of T_n to its limit.

Assumption 3. *There exists a compact operator $T : \mathcal{F} \mapsto \mathcal{W}$ such that*

$$\|T_n - T\|^2 := \sup\{\|(T_n - T)g\|^2; g \in \mathcal{F} \text{ of norm } 1\} = \mathcal{O}_p(a_n^2) \tag{2.4}$$

where a_n^2 is sequence decreasing to zero as n tends to infinity. We also denote by λ_j^2 , resp. ϕ_j , the sequence of eigenvalues, resp. eigenfunctions, of the operator $T^* T$ where T^* denotes the adjoint of T .

It is possible to provide primitive conditions in order to derive an explicit rate for a_n^2 . A deeper discussion can be found in Bagdziunas and Van Bellegem (2014), where technical conditions are provided in order to get a rate $a_n^2 = 1/n^{1-\eta}$ for any $\eta > 0$. The conditions relate to the moments of the involved random functions and are found from the general exponential bounds for probability of deviation derived by Mendelson and Pajor (2006).

The next assumption is a smoothness prior imposed on the solution φ . Imposing some smoothness of the solution is a common assumption in nonparametric inference in order to derive the rate of convergence of an estimator. It appears that, in the context of ill-posed inverse problems, a convenient way to express the smoothness assumption is to relate φ to the range of T . The rate of convergence is not solely related to the quality of φ , but to the relative regularity of φ with respect to T . This measure is stated accurately in the following assumption, and then further discussed.

Assumption 4. *We assume the existence of and denote by β the largest positive index such that $\varphi \in \text{Range}(\{T^*T\}^{\beta/2})$.*

Remind that (λ_j^2, ϕ_j) , $j = 1, 2, \dots$, is the sequence of eigenvalues and eigenfunctions of the operator T^*T . The last assumption ensures the existence of a function $\psi \in \mathcal{F}$ such that $\varphi = (T^*T)^{\beta/2}\psi$. In other words, it assumes that $(T^*T)^{-\beta/2}\varphi$ belongs to $\mathcal{F}$. From the singular value decomposition of the self-adjoint operator T^*T , the last constraint is equivalent to assume that

$$\sum_{j \geq 1} \frac{\langle \varphi, \phi_j \rangle}{\lambda_j^\beta} \phi_j \in \mathcal{F}.$$

Since T is a compact operator, $\lambda_j^2 \rightarrow 0$ and this assumption is true either if $\langle \varphi, \phi_j \rangle$ converges fastly to zero or, if not, the index β is sufficiently small. In that sense β measures the relative regularity of φ with respect to (the spectrum of) T^*T .

The next and last assumption is about the relative amplitude of the error term (described by the covariance operator Ω) with respect to the main term of the model (described by T^*T).

Assumption 5. *We assume the existence of and denote by $\gamma > 0$ the largest positive index such that*

$$\sum_{j \geq 1} \frac{\langle \Omega \phi_j, \phi_j \rangle}{\lambda_j^{2\gamma}} < \infty.$$

In order to understand Assumption 5 it is meaningful to consider the particular case where the eigenfunctions ϕ_j are the eigenfunctions of the operator Ω as well (this is an example and not a general assumption of this paper). In this case, $\Omega \phi_j = \omega_j \phi_j$ for all j , where the sequence ω_j are the eigenvalues of Ω , and the assumption becomes $\sum_j (\omega_j / \lambda_j^{2\gamma}) < \infty$. A high value of γ either means that the eigenvalues of Ω converge quickly to zero or the eigenvalues of T do not converge fast to zero. In other words, a high value of the index means that the eigenvalues of the covariance operator between the covariates z and the instruments w are relatively high compared to the eigenvalues of the noise operator Ω . High values of γ therefore corresponds to less degree of ill-posedness in the statistical model.

We are now in position to derive the mean square convergence of the regularized estimator.

Theorem 2.1. *Under assumptions 1 to 5, the regularized estimator (2.3) is such that*

$$\mathbb{E}\|\varphi - \hat{\varphi}_{\alpha,n}\|^2 = \mathcal{O}\left(\alpha_n^{\min(\beta,2)} + \alpha_n^{\min(\beta,2)} \frac{a_n^2}{\alpha_n^2} + \frac{a_n^2 \alpha_n^{\min(\gamma,1)}}{n \alpha_n^3} + \frac{\alpha_n^{\min(\gamma,1)}}{n \alpha_n}\right)$$

The proof of this result is in the line of the proof a more advanced results (Theorem 4.1) and therefore is skipped. If we impose some constraints on α_n and on the indices β and γ , it is possible to have that the first and the fourth terms dominate the rate of convergence. This result is stated in the following corollary.

Corollary 2.1. *Under the assumptions of Theorem 2.1, if $\beta < 2$ and $\gamma < 1$ are such that $\beta + 1 - \gamma > 0$, and if the regularization parameter α_n is such that $a_n^2/\alpha_n^2 = \mathcal{O}(1)$, then*

$$\mathbb{E}\|\varphi - \hat{\varphi}_{\alpha,n}\|^2 = \mathcal{O}\left(\alpha_n^\beta + \frac{\alpha_n^\gamma}{n \alpha_n}\right).$$

In particular, with the choice $\alpha_n \sim n^{-1/(\beta+1-\gamma)}$, the rate of convergence of the regularized estimator is given by

$$\mathbb{E}\|\varphi - \hat{\varphi}_{\alpha,n}\|^2 = \mathcal{O}\left(n^{\frac{-\beta}{\beta+1-\gamma}}\right).$$

This rate is known to be the optimal rate of convergence in the minimax sense over a class of functions satisfying Assumptions 4 and 5 (Chen & Reiss, 2011). The constraints $\beta < 2$ and $\gamma < 1$ are imposed in order to simplify the statement of the result. If they do not hold true, $\min(\beta, 2)$ and $\min(\gamma, 1)$ appear in the statement, and the result is no longer optimal. This phenomenon is known as the “saturation effect” of the Tikhonov regularization. It is a limitation of the Tikhonov regularization method. Optimality can be recovered for $\beta > 2$ and $\gamma > 1$ if another regularization scheme is considered, such as the “iterated Tikhonov regularization” (studied in another context in econometrics e.g. in Johannes et al., 2011).

In the next section, we further discuss the assumptions of this corollary.

2.3 A discussion on instrument strength

Corollary 2.1 shows that minimax optimal rate of convergence is reached under the condition $\beta + 1 - \gamma > 0$. If this condition does not hold true, the regularized estimator is no longer optimal.

The condition $\beta + 1 - \gamma > 0$ imposes constraints on the indices β and γ that we now want to highlight now. If the degree of ill-posedness is high, γ is low and it is not necessary to have a lot of regularity on the solution φ in order to recover the optimal rate of convergence. On the contrary, if the value of γ is high, the ill-posedness is more severe and we can expect to recover γ with some accuracy if φ is regular enough.

The condition is a constraint on the random variables involved in the statistical model. Since it is a sufficient assumption to get the optimal rate of convergence, we suggest to qualify as “strong” the instruments of the corresponding statistical model.

The next assumption formalizes this notion of strong instrument in the functional linear model.

Assumption 6 (Strong instruments). *The instruments of the statistical model defined in Assumptions 1 to 5 are strong, meaning that the indices β and γ are such that $\beta + 1 > \gamma$.*

Below we will work under the assumption that instruments are strong. If the assumption does not hold, other rates of convergence will be found, and the limit theorems we show below are not valid. We leave the study of inference in this context open for future research.

3 Numerical aspects and smooth priors

3.1 Discretized normal equations

In this section we describe how to compute the estimator (2.3) in practice. The estimator involves the operator

$$T_n : \mathcal{F} \rightarrow \mathcal{W} : \varphi \rightarrow \frac{1}{n} \sum_{i=1}^n \langle z_i, \varphi \rangle_{\mathcal{F}} w_i$$

and its adjoint

$$T_n^* : \mathcal{W} \rightarrow \mathcal{F} : \psi \rightarrow \frac{1}{n} \sum_{i=1}^n \langle x_i, \psi \rangle_{\mathcal{W}} z_i .$$

The regularized estimator, $\hat{\varphi}_{\alpha,n}$, satisfies the normal equation

$$(\alpha_n + T_n^* T_n) \varphi = T_n^* r_n$$

which is equivalent, by definition of the operators, to

$$\alpha_n \varphi + \frac{1}{n^2} \sum_{i,j} z_i \langle w_i, w_j \rangle \langle z_j, \varphi \rangle = \frac{1}{n^2} \sum_{i,j} z_i \langle w_i, w_j \rangle y_j \quad (3.1)$$

where the inner products are understood in their appropriate spaces ($\mathcal{W}$ or $\mathcal{F}$). Taking the inner product in $\mathcal{F}$ of all elements in (3.1) with respect to z_l , we obtain for $l = 1, \dots, n$

$$\alpha \zeta_l + \frac{1}{n^2} \sum_{i,j} \langle z_i, z_l \rangle \langle w_i, w_j \rangle \zeta_j = \frac{1}{n^2} \sum_{i,j} \langle z_i, z_l \rangle \langle w_i, w_j \rangle y_j \quad (3.2)$$

where $\zeta_l = \langle \varphi, z_l \rangle$. This linear system is easily solved with respect to $\zeta_1, \dots, \zeta_l$.

In the case where the variables z_i and w_i are curves, the inner products involved in the last linear system must be discretized in practice. To fix this idea, consider that the Hilbert spaces $\mathcal{F}$ and $\mathcal{W}$ contains square integrable functions in $L^2[0, 1]$. For a given (and large) $T > 0$, we replace the inner product $\langle z_i, z_l \rangle$ by

$$\frac{1}{T} \sum_{t=1}^T z_i \left(\frac{t}{T} \right) z_l \left(\frac{t}{T} \right)$$

and similarly for $\langle w_i, w_j \rangle$. If Z denotes the $T \times n$ matrix such that $Z_{ti} = z_i(t/T)$ and if W denotes the $T \times n$ matrix such that $W_{ti} = w_i(t/T)$, the solution of the linear system (3.2) is given by

$$\zeta = \left(\alpha_n I + \frac{1}{n^2} \frac{Z'Z}{T} \frac{W'W}{T} \right)^{-1} \left(\frac{1}{n^2} \frac{Z'Z}{T} \frac{W'W}{T} y \right) \quad (3.3)$$

for a given α_n , and with $\zeta = (\zeta_1, \dots, \zeta_l)'$. The regularized estimator is recovered from equation (3.1), namely

$$\hat{\varphi}_{\alpha,n} = \frac{1}{\alpha n^2} Z \frac{W'W}{T} (y - \zeta) \quad (3.4)$$

where $y = (y_1, \dots, y_n)'$.

The practical implementation of the estimator therefore only involves simple matrix manipulations.

3.2 Illustration

We present in this section a first design of simulation, in which z_i and w_i are random curves in $L^2[0, 1]$. The unknown target function in this design is $\varphi(u) = 16u^3 - 24u^2 + 9u$ for $u \in [0, 1]$. We generate the model

$$\begin{aligned} y_i &= \frac{1}{T} \sum_t z_i \left(\frac{t}{T} \right) \varphi \left(\frac{t}{T} \right) + \sigma_u u_i \\ z_i \left(\frac{t}{T} \right) &= \zeta R w_i \left(\frac{t}{T} \right) + \delta_u u_i + \sigma_v v_i \end{aligned}$$

where v_i and w_i are two brownian bridges, u_i is $N(0, 1)$ and the transform R is

$$R w_i \left(\frac{t}{T} \right) = \frac{1}{T} \sum_s \min \left(\frac{s}{T}, \frac{t}{T} \right) W_i \left(\frac{s}{T} \right).$$

In this design the index δ_u controls for the level of endogeneity and is set to 1 in our simulations. The transform R is a discretized integration,² so that the curves z_i are smoother versions of the brownian bridges w_i plus some noise terms. Figure 1 illustrates this design and shows the curve φ as well as the set of random curves w_i and z_i .

We now illustrate the estimator, based on a sample of $n = 100$ observations. Figure 2 shows the result of the estimation using the linear estimator presented in the previous section with $T = 1000$. Three values of the regularization parameter α_n are studied. As expected, for larger values of α_n the bias of the estimator is higher, while for smaller values of α_n the variance of the estimator is large.

The result is therefore sensitive to the choice of α_n and we will now improve the estimation in two ways. The first way is to find a data-driven rule for the selection of the parameter. The second way is to impose more regularity constraints on the solution. We study these two ideas in the next sections.

²To see this, compare with formula (3.6) below in this text.

Figure 1: Design of simulation

Figure 2: The estimator (in color) and the true curve, for different values of the regularization parameter

3.3 Choice of the regularization parameter

As far as we know, there is no adaptive rule for the selection of the regularization parameter in the context of functional instrumental linear regression. It is one possible extension of this paper to construct an adaptive rule, where “adaptive” means that the resulting rate of convergence is still optimal with the data-driven α_n .

The following argument is based on an argument taken from Fève and Florens (2010) and appears to work well in practice. First, using the same type of mathematical arguments as in the previous section, the following convergence rate can be proved

$$\|T_n \hat{\varphi}_{\alpha,n} - r_n\|^2 = \mathcal{O}_p \left(\frac{\alpha_n^\gamma}{n} + \alpha_n^{1+\beta} \right)$$

if the instruments are strong and if $\gamma \leq 1$ and $\beta \leq 2$. This rate of convergence motivates to look at

$$G(\alpha) := \frac{1}{\alpha} \|T_n \hat{\varphi}_{\alpha,n} - r_n\|^2$$

as a function of α , and which is such that the minimum of G is reached for the optimal α given in Corollary 2.1.

(a) The function to be minimized in order to find a data-driven value of the regularization parameter α (b) The resulting estimator with the data-driven selection of α (in the simulation, $\hat{\alpha} = 6.5 \times 10^{-7}$)

Figure 3: Data-driven selection of the parameter and the resulting estimator from a simulation

This method is illustrated in Figure 3. The figure shows that the curve $G(\alpha)$ has a clear minimum, which can be used to find a data-driven choice of α . The statistical properties of the resulting estimator with data-driven selection of α is a relevant topic for future research, beyond the scope of this paper.

3.4 Imposing a smoothness constraint

The regularized estimator (2.3) also appears to be the function that minimizes the penalized least-square criterion

$$\|T_n \varphi - r_n\|^2 + \alpha \|\varphi\|^2 \quad (3.5)$$

with respect to φ . A common way to impose smoothness constraints on φ is to change the form of the penalty, e.g. to replace the norm of φ by the norm of derivatives of φ .

To this end, we introduce the integral operator M that is such that

$$Mg(t) = \int_0^t g(s)ds$$

for every g in $L^2[0, 1]$. It is easy to show that the adjoint of M is

$$M^*g(t) = \int_t^1 g(s)ds$$

and therefore

$$M^* M g(t) = \int_0^1 \min(s, t) g(s) ds. \quad (3.6)$$

If we want to impose the solution φ to have s existing derivatives, the penalized least-square (3.5) is replaced by

$$\|T_n \varphi - r_n\|^2 + \alpha \|(M^* M)^{-s/2} \varphi\|^2.$$

In order to compute the solution in practice, we can proceed analogously to the above section on discretized normal equation. If we denote by $\tilde{M}$ the triangular $T \times T$ matrix with entries 1 in the lower triangle, then the solution (3.3–3.4) adapted for the new penalty is now

$$\zeta = \left(\alpha_n I + \frac{1}{n^2} Z' \left(\frac{\tilde{M} \tilde{M}'}{T^2} \right)^s Z \frac{W' W}{T} \right)^{-1} \left(\frac{1}{n^2} Z' \left(\frac{\tilde{M} \tilde{M}'}{T^2} \right)^s Z \frac{W' W}{T} y \right) \quad (3.7)$$

and

$$\hat{\varphi}_{\alpha, n} = \frac{1}{\alpha n^2} \left(\frac{\tilde{M} \tilde{M}'}{T^2} \right)^s Z \frac{W' W}{T} (y - \zeta). \quad (3.8)$$

We illustrate this procedure by an example with $s = 1$. Of course, since the solution φ is defined over $[0, 1]$ the existence of $(M^* M)^{-1} \varphi$ implies border conditions, that are here fixed as $\varphi(0) = 0$ and $\varphi'(1) = 0$. For this reason, we now consider a new target function, given by $\varphi(u) = \sin[2\pi(1 - u)^2]$. We consider the same generating process as above for the random curves w_i and z_i . The true function, as well as its estimation from $n = 50$ data are in Figure 4.

This experiment shows that the use of a smoothness prior has a significant impact on the nonparametric estimation. In Figure 4, we also notice that the estimated function of course fulfills the border constraints.

4 Generalized instrumental variable estimation

4.1 Asymptotic mean square convergence of a generalized regularized estimator

In this section, we study a generalized version of the regularized estimator, in which the regression model is premultiplied by the inverse of a data-dependent, self-adjoint operator on $\mathcal{W}$. We denote by H_n this operator in the section. Given that the inverse of an operator is not always defined, we introduce a sequence of positive numbers (ρ_n) converging to zero as n tends to infinity, and define the operator B_n from $\mathcal{W}$ to $\mathcal{W}$ and such that

$$B_n = (\rho_n I + H_n)^{-1} H_n^{1/2}.$$

Figure 4: The figure superimposes the true target function, the Tikhonov regularized estimator (2.3) and the regularized estimator with a smoothness constraint (3.7–3.8) from a single simulation, with a data-driven value of the regularization parameter in both estimators.

The functional regression model (2.2) is premultiplied by the operator B_n , i.e.

$$B_n r_n = B_n T_n \varphi + B_n \varepsilon_n. \quad (4.1)$$

and we now want to study the properties of the generalized estimator

$$\hat{\varphi}_{\alpha,n} = (\alpha_n I + T_n^* B_n^* B_n T_n)^{-1} T_n^* B_n^* B_n r_n. \quad (4.2)$$

In order to get the mean square convergence of the estimator, we need to adapt some of the above assumptions, and to make additional assumptions that are related to the new operators B_n and H_n . The first assumption is about the convergence of H_n that is supposed to have some limit H , and to converge at a certain rate in the spectral norm. The assumption also impose some constraints on the limit operator.

Assumption 7. *There exists an operator $H : \mathcal{W} \rightarrow \mathcal{W}$ such that*

$$\|H_n^{1/2} - H^{1/2}\|^2 = \mathcal{O}_p(h_n^2) \quad (4.3)$$

for some positive sequence h_n decreasing to zero. The operator is such that $\text{tr}(H^{-1/2} \Omega H^{-1/2}) < \infty$ and $H^{-1/2} T$ is bounded, where inverse are understood as generalized inverses.

The next assumption is an extension of the Assumption 4. It measures the prior regularity of φ with respect to the operator $T^*H^{-1}T$ where H^{-1} is a generalized inverse (e.g. Engl et al., 2000). The discussion is analogous to the above comments on Assumption 4.

Assumption 8. *We assume the existence of and denote by β the largest positive index such that $\varphi \in \text{Range}(\{T^*H^{-1}T\}^{\beta/2})$.*

The next assumption is an extension of the Assumption 5 defining the degree of ill-posedness γ .

Assumption 9. *If the operator H satisfies Assumption 7, denote by $\{(\kappa_j, \xi_j, \tilde{\xi}_j); j = 1, 2, 3, \dots\}$ the singular value decomposition of the operator $H^{-1/2}T$. We assume the existence of, and denote by $\gamma > 0$ the largest index such that the condition*

$$\sum_j \frac{|\langle H^{-1/2}\Omega H^{-1/2}\xi_j, \xi_j \rangle|}{\kappa_j^{2\gamma}} < \infty$$

holds true.

In order to derive a rate of convergence, we need an additional technical assumption quantifying the behavior of the limit operator $H^{-1/2}T$ in the singular value decomposition provided by H .

Assumption 10. *If the operator H satisfies Assumption 7, denote by $\{(\nu_j, \psi_j, \tilde{\psi}_j); j = 1, 2, 3, \dots\}$ its singular value decomposition. We assume the existence of, and denote by $\theta > 0$ the largest index such that the condition*

$$\sup_j \sum_k \frac{\langle H^{-1/2}T\psi_j, \psi_k \rangle}{\nu_k^\theta} < \infty$$

is fulfilled.

This last assumption simplifies if we assume a link between the operator T and H , for example by imposing that they have common eigenfunctions. We do not need that stronger assumption to derive and discuss the rate of convergence of the estimator, which is stated in the following theorem.

Theorem 4.1. *Let the statistical model defined by Assumptions 1 and 2. Suppose the operator T of the model satisfies Assumption 3. Consider the generalized regularized estimator (4.2) under Assumptions 7 to 8. Then the mean square error of the estimator (4.2) is such that*

$$\mathbb{E}\|\varphi - \hat{\varphi}_{\alpha,n}\|^2 = \mathcal{O}\left(\frac{1}{\alpha^{1-\min(1,\gamma)}n} + \frac{1}{\alpha} \left\{ \frac{1}{\alpha n} + \alpha^\beta \right\} \left\{ \frac{a_n^2}{\rho} + \frac{h_n^2}{\rho^2} + \rho^{\min(\theta,2)} \right\} + \alpha^\beta\right)$$

Again our discussion will focus on the case where the leading rate is given by the first and the last terms in this rate of convergence. This now requires that the factor depending on the regularization parameter ρ converges fastly enough to zero. This last requirement involves sequences a_n and h_n as well as the index θ .

We illustrate this point in one particular case. We focus on the optimal value of the parameter α (that is $n^{-1/(\beta+1-\gamma)}$, cf Corollary 2.1). Suppose the instruments are strong, and that $\beta < 2$ and $\gamma < 1$. We also suppose that $a_n^2 = h_n^2 = n^{-\epsilon}$ for some $\epsilon > 0$. In that case, we select the optimal regularization parameter $\rho \sim n^{-\epsilon/(\theta+2)}$ and the second term in the rate is negligible if θ is large enough, i.e. such that

$$\frac{\theta}{2+\theta} > \frac{1}{\epsilon} \frac{\beta + \beta\gamma + 1}{(\beta + 1 - \gamma)(\beta + 1)}.$$

Of course, the discussion on the parameters in the general case is more complex. The main message is that the rate of convergence can be altered when we consider the generalized estimator. The optimal rate of convergence is obtained if the eigenvalues of H satisfy Assumption 10 with a sufficiently large index θ . How large it is depends on a mixture of indices in the statistical model (a_n , h_n , etc.). If the operator H has been well selected, the optimal rate of convergence of the estimator can be recovered. We summarize this discussion in the following corollary.

Corollary 4.1. *Under the assumptions of Theorem 4.1, if the instruments are strong and if*

$$\frac{1}{\alpha} \left\{ \frac{1}{\alpha n} + \alpha^\beta \right\} \left\{ \frac{a_n^2}{\rho} + \frac{h_n^2}{\rho^2} + \rho^{\min(\theta, 2)} \right\}$$

is negligible with respect to $1/(\alpha^{1-\min(1,\gamma)}n)$, then the optimal rate of convergence of the generalized regularized estimator is obtained with $\alpha_n \sim n^{-1/(\beta+1-\gamma)}$ and is given by

$$\mathbb{E} \|\varphi - \hat{\varphi}_{\alpha,n}\|^2 = \mathcal{O} \left(n^{\frac{-\beta}{\beta+1-\gamma}} \right).$$

One difficulty for the interpretation of this result is that the indices β and γ now depends on the choice of H . In the next section we further discuss this phenomenon.

4.2 Invariance of the convergence rate

The fact that the rate of convergence depends on the choice of H is not a desirable property of the generalized estimator. In this section, we address the following question: Suppose the assumptions of Corollary 4.1 are fulfilled, that is we are working with a generalized estimator that has the optimal rate of convergence, given H . We now consider the class of H leading to that property (with various β and γ) and ask whether the rate of convergence is identical or not for each choice of H in this class.

We are not able to answer that question in general since the class of all possible H is too large. So we restrict the discussion to a relevant class of operators H appearing in

Assumption 7 of the generalized estimator. From now on, we will assume that H takes the form Ω^δ for some $\delta \geq 0$ (recall that the operator Ω is the covariance operator in the model, cf Assumption 1). If H takes this form, we show below that the rate of convergence does not depend on δ in very general settings.

The first setting in which we will show that property is given by the setting of Hall and Horowitz (2005), who consider functional exogenous regression as well.

Assumption 11 (Hall and Horowitz setting). *All variables are compactly supported on $[0, 1]$, the eigenfunctions of the operators Σ and T are given by the sinus functions $e_j(u) = \dots$. The eigenvalues of T are polynomially decreasing at the rate j^{-k} for some $k > 0$. The eigenvalues of Ω are polynomially decreasing at the rate j^{-2b} for some $b > 0$. The Fourier coefficients of φ , $\langle \varphi, e_j \rangle$ are polynomially decreasing at the rate j^{-c} for some $c > 0$.*

Proposition 4.1. *In the setting of Corollary 4.1 and if the instruments are strong, the mean square error of the generalized estimator (4.2) does not depend on the value of δ . This rate of convergence is given by*

$$\mathbb{E} \|\varphi - \hat{\varphi}_{\alpha, n}\|^2 = n^{-\frac{1-2c}{2(k-b)-2c+2}}. \quad (4.4)$$

Proof. We first have to write the rate of Corollary 4.1 with respect to the parameters of Assumption 11 (k , b and c). Consider the index β in Assumption 8. The assumption implies that $(\Omega^{-\delta/2}T)^{-\beta}\varphi$ belongs to $\mathcal{F}$. In particular, it implies that the series $\sum_j |\langle \varphi, e_j \rangle|^2 / \lambda_j^{2\beta}$ is convergent. Note that, under Assumption 8, $\lambda_j \asymp j^{\delta b} / j^k$ and $|\langle \varphi, e_j \rangle|^2 \asymp 1 / j^{2c}$. Therefore, Assumption 8 requires that the series $\sum_j j^{2k\beta} / j^{2(\beta\delta b + c)}$ is convergent. This constraints β to be at most $(1 - 2c) / (2k - 2\delta b)$.

Similarly we find γ . From Assumption 9, γ is the largest number such that the series

$$\sum_j \frac{j^{-2b(1-\delta)}}{j^{2\gamma\delta b - 2\gamma k}} \quad (4.5)$$

is convergent. This imposes to take γ at most equal to $(b(1 - \delta) + 1) / (k - \delta b)$. Replacing β and γ in the rate found in Corollary 4.1 leads to the result. $\square$

Since the setting of Hall and Horowitz (2005) is very specific, one could ask if the invariance result is a general property of the estimator, or if the result only appears in this setting. In the next proposition, we generalize the result to a milder setting. This setting is written in terms of the Hilbert scales, a general mathematical setting enlarging the setting of Hall and Horowitz (2005). We refer to Johannes et al. (2011) for a discussion of Hilbert scales that are defined in the next assumption.

Assumption 12 (Compatible Hilbert scale setting). *Suppose $\mathcal{L} : \mathcal{F} \rightarrow \mathcal{F}$ is an unbounded self adjoint and strictly positive operator. Consider the set of spaces $\mathcal{H}_q = \text{Dom}(\mathcal{L}^{q/2})$ for $q \in \mathbb{R}$, called a Hilbert scale generated by $\mathcal{L}$. We assume that the operators $\mathcal{L}$, Ω and T have the same eigenfunctions as such that³*

³We write $a \asymp b$ if there exists two positive constants $\underline{c}$ and $\bar{c}$ that are such that $\underline{c}a \leq b \leq \bar{c}a$.

- The solution φ belongs to $\mathcal{H}_p$ for some $p > 0$.
- $\|Tg\| \asymp \|\mathcal{L}^{-k}g\|$
- $\|\Omega g\| \asymp \|\mathcal{L}^{-2b}g\|$

for all g such that the norms exist.

Proposition 4.2. *In the setting of Corollary 4.1 and if the instruments are strong, the mean square error of the estimator 4.2 does not depend on the value of δ .*

Proof. First, remind that the eigenvalues of the operator T are denoted by λ_j . Let us denote by ω_j the eigenvalues of Ω . Assumption 12 implies that $\omega_j \asymp \lambda_j^{2b/k}$ (to see this, write the assumption for the particular g equals to the eigenfunctions). Moreover, the eigenvalues of $\Omega^{-\delta/2}T$, that are denoted by κ_j , are such that $\kappa_j \asymp \lambda_j/\omega_j^{\delta/2}$. This implies that $\|\Omega^{-\delta/2}T\| \asymp \|\mathcal{L}^{-a(\delta)}g\|$ with $a(\delta) = k - b\delta$.

The regularity assumption on φ given by Assumption 8 implies that φ belongs to $\varphi = \mathcal{L}^{-a\beta}\psi$ for some ψ such that the expression is well defined. This with Assumption 12 implies that $\beta = \beta(\delta) = p/(2a_\delta)$. When $\delta = 0$, the largest γ found in Assumption 9 is denoted by γ_0 and is such that the series

$$\sum_j \frac{\omega_j}{\kappa_j^{2\gamma_0}} \asymp \sum_j \frac{\omega_j}{\lambda_j^{2\gamma_0}} \asymp \sum_j \lambda_j^{(2b/k)-2\gamma_0} \quad (4.6)$$

converges. In such a case, $a = k$ and the rate found in Corollary 4.1 is $n^{-\frac{p}{p+k-\gamma_0k}}$.

If $\delta > 0$, the largest γ found in Assumption 9 is now such that the series

$$\sum_j \frac{\omega_j^{1-\delta+\gamma\delta}}{\lambda_j^{2\gamma}} \asymp \sum_j \lambda_j^{(2b/k)(1-\delta+\gamma\delta)-2\gamma} \quad (4.7)$$

is convergent. This implies that $\gamma_0 = (b/k)(1 - \delta + \gamma\delta) - \gamma$ or $\gamma = (\delta b - \gamma_0 k)/(\delta b - k)$. The resulting rate is $n^{-\frac{p}{p+k-\delta b-\gamma(k-\delta b)}} = n^{-\frac{p}{p+k-\gamma_0 k}}$. This proves the result. $\square$

4.3 Central Limit Theorem

We close this theoretical study of the generalized estimator by a statement of a Central limit Theorem on the inner product $\langle \hat{\varphi}_{\alpha,n}, g \rangle$, where g is a fixed function in $\mathcal{F}$. As in the previous section, we restrict the class of generalized estimators to be such that H takes the form Ω^δ for some $\delta \geq 0$.

Before stating the theorem, we introduce a notation for the regularized solution, φ_α which is such that

$$\varphi_\alpha = (\alpha_n I + T^* H^{-1/2} T)^{-1} T^* H^{-1/2} T \varphi \quad (4.8)$$

Theorem 4.2. *Under the assumptions of Corollary 4.1 and under Assumption 12, if the instruments are strong, assume that variables u and w are such that $\mathbb{E}\|uw\|^\nu$ is finite for some $\nu > 2$. If moreover $b \geq 0$ and $\alpha_n^{k/(2k-2b\delta)}\sqrt{n} \rightarrow 0$ as n tends to infinity, then the generalized regularized estimator satisfies the central limit theorem*

$$\langle \hat{\varphi}_{\alpha,n} - \varphi_\alpha, g \rangle \quad \text{is} \quad AN(0, \beta_n^2) \quad \text{as } n \rightarrow \infty,$$

where β_n^2 is such that

$$C_1 \sum_{j: \alpha_n \leq \kappa_j^2} \kappa_j^{-2\frac{k-b}{k-b\delta}} \leq n\beta_n^2 \leq C_2 \alpha^{-\frac{k-b}{k-b\delta}}$$

for n sufficiently large, where C_1 and C_2 are positive constants.

5 Empirical application: The impact of fertility on economic growth

This section shows how the functional linear instrumental regression (FLIR) function sheds another light on the relationship between fertility and economic growth. In the context of economic development the role of fertility is of a large interest both in the theoretical and empirical literature (e.g. Barro & Becker, 1989; Blackburn & Cipriani, 2002; Ashraf, Weil, & Wilde, 2013).

In our analysis we use the FLIR to quantify the relationship between the fertility rate and the growth rate in the United Kingdom. Our dependent variable (y) is given by the annual growth of GDP per capita from 1966 to 2012 and expressed as percentages, see Figure 5(a). Data are taken from the World Bank database and the time period is limited by the availability of fertility data in the UK. The age specific fertility rate is defined yearly as the density of birth with respect to mother's ages. Data over the time period are available from the UK Office of National Statistics. The fertility rate is provided for each mother's age from 15 years old until 44 years old, and for the category of mothers aged 45 or more. Therefore, each year the fertility rate is given for $T = 31$ specific ages or age category, giving a natural discretization of the fertility density. Figure 5(b) superimposes yearly age-specific fertility rates over the time period considered. The figure shows that more recent fertility rates have a higher mode compared to earlier fertility rate, demonstrating a clear shift in mother's ages over the time period.

Fertility rate is an endogenous functional regressor for the growth rate, since other omitted variables related to human capital are also influential on growth. A common instrument for the fertility rate found in the empirical literature is given by the rate of twin births or multiple births, see e.g. Bronars and Grogger (1994); Rosenzweig and Wolpin (1980); Angrist, Lavy, and Schlosser (2010). Yearly multiple birth fertility data are available from the UK Office of National Statistics only for ranges of mothers age given by

$[0, 14]$, $[15, 19]$, $[20, 24]$, $\dots$, $[40, 44]$, $[45, \infty[$. We use those data as instruments and, given the available information, we construct step functions that are constant within a range of ages. Figure 5(c) shows the instrumental curves superimposed. Observe, first, that the irregularity of the curves are not in contradiction with the above theory or simulations, where the instrumental curves show less regularity than the regressing curves. Note also that oldest mothers show the biggest variation in the multiple birth fertility rate, while the youngest mothers show low variation. This effect is well documented and is due to an improvement in the technology and availability of assisted production, see e.g. Sutcliffe and Ludwig (2007). A last point is that, due to a registrars strike in 1981, the availability of birth rate is very limited in 1981. Data for the year 1981 has therefore been excluded from our empirical analysis.

Figure 5: Data are from the United Kingdom, yearly over the time period 1966-2012 (Source: World Bank and the UK Office of National Statistics)

In this context the economic meaning of the FLIR, $\varphi(t)$, is the marginal effect of the age-specific fertility around the age t , to economic growth. Figure 6 shows two estimators. The solid line is the estimator (2.3). The dotted line shows the estimator with a smoothness constraint (3.5). The constraint is set up as in Section 3 and is not in contradiction with the common economic sense. In particular $\varphi(0) = 0$ is obvious, and $\varphi'(45) = 1$ models a decreasing of the impact of fertility from older mothers on economic growth. Empirically, the unconstrained curve is very close to the curve under constraint, which is another sign for the relevance of the constraints.

Three aspects of the estimated curves are of interest. First the maximal, positive effect is given around the age of 25, meaning that the biggest positive effect of fertility on economic growth is found around that age. Then, the biggest negative impact appears around the age of 18. Finally, what happens after the age of 30 has a fewer impact on growth. Given the importance in discussions nowadays on demographic changes and shifts in the age-specific fertility curves, this empirical result and our methodology provide new ways to analyze the impact of those changes on the economy.

To close the discussion we mention that the use of multiple birth as instruments, although it is common and well-established in the literature, is not without criticism, see

Figure 6: The functional instrumental regression function, ϕ , is a function of mother's age and quantifies the marginal importance of the age of fertility on economic growth. The solid is the Tikhonov estimator with $\alpha = 0.15$, the dotted line is the estimator under smoothness constraint with $\alpha = 0.0015$.

Braakmann and Wildman (2014). Other models linking fertility and growth might provide less controversial instruments, such as dynamic models with differential fertility (De la Croix & Doepke, 2003). Differential fertility compares fertility rates between higher educated and lower educated families, conditionally on mother's age. Availability of the data over long periods of time is problematic now, but public data are under construction and the information will be available in the near future. Given the general trend of big data availability, in particular in the field of demography and growth, the FLIR approach can be seen as a very useful tool in modern regression analysis.

APPENDIX

A Lemmas and proofs

In what follows, we write $A \lesssim B$ when there exists a positive, finite constant c that does not depend on A, B and is such that $A \leq cB$. We also write L^{-1} for the generalized inverse of any operator L , if it exists.

The following gathers useful results on the norm of (a function of) operators. A formal proof of the lemma can be found in the appendix of (Florens, Johannes, & Van Bellegem, 2011).

Lemma A.1. *Let $K : \mathbb{H} \rightarrow \mathbb{G}$ be a linear operator defined between the two Hilbert spaces $\mathbb{H}$ and $\mathbb{G}$, and let K^* be the adjoint operator of K . Then, for all $\alpha > 0$, the following bounds on the operator norm hold true:*

$$\|\alpha(\alpha I + K^*K)^{-1}(K^*K)^\gamma\| \leq \begin{cases} \alpha^\gamma & \text{if } 0 < \gamma \leq 1 \\ \|K^*K\|^{\gamma-1}\alpha & \text{if } \gamma > 1 \end{cases}, \quad (\text{A.1})$$

$$\|(\alpha I + K^*K)^{-1}K^*\| = \|K(\alpha I + K^*K)^{-1}\| \lesssim 1/\sqrt{\alpha}, \quad (\text{A.2})$$

$$\|(\alpha I + K^*K)^{-1}\| \leq 1/\alpha, \quad (\text{A.3})$$

$$\|K(\alpha I + K^*K)^{-1}K^*\| \leq 1, \quad (\text{A.4})$$

$$\|K[I - (\alpha I + K^*K)^{-1}K^*K]\| \lesssim \sqrt{\alpha}, \quad (\text{A.5})$$

$$\|I - (\alpha I + K^*K)^{-1}K^*K\| \leq 1. \quad (\text{A.6})$$

A.1 Proof of Theorem 4.1

Lemma A.2. *Let $H_n : \mathcal{W} \rightarrow \mathcal{W}$ be a data-dependent operator. Suppose there exists a compact operator $H : \mathcal{W} \rightarrow \mathcal{W}$ such that*

$$\|H_n^{1/2} - H^{1/2}\|^2 = \mathcal{O}_p(h_n^2) \quad (\text{A.7})$$

for some decreasing sequence h_n^2 . Let $B_n = (\rho + H_n)^{-1}H_n^{1/2}$ where $\rho > 0$ may depend on n , and $B = H^{-1/2}$. Suppose the operators T_n and T satisfy Assumption 3. Then

$$\|B_n T_n - B T\|^2 = \mathcal{O}_p\left(\frac{a_n^2}{\rho} + \frac{h_n^2}{\rho^2} + \rho^{\min(\theta, 2)}\right)$$

Proof. Denote $B_\rho = (\rho I + H)^{-1}H^{1/2}$ and consider the expansion

$$B_n T_n - B T = B_\rho(T_n - T) + (B_n - B_\rho)T + (B_\rho - B)T + (B_n - B_\rho)(T_n - T). \quad (\text{A.8})$$

In the following we control the square norm of each of the four terms separately. The first term is such that

$$\|B_\rho(T_n - T)\|^2 \leq \|B_\rho\|^2 \|T_n - T\|^2 = \mathcal{O}(a_n^2/\rho)$$

from Assumption 3 and from (A.2).

From the definition of B_n and B_ρ the second term is such that

$$\begin{aligned}(B_n - B_\rho)T &= \{(\rho + H_n)^{-1}H_n^{1/2} - (\rho + H)^{-1}H^{1/2}\}T \\ &= (\rho + H_n)^{-1}(H_n^{1/2} - H^{1/2})T + (\rho + H_n)^{-1}(H - H_n)(\rho + H)^{-1}H^{1/2}T\end{aligned}$$

leading to two terms that we control separately. The squared norm of the first part is such that

$$\|(\rho + H_n)^{-1}(H_n^{1/2} - H^{1/2})T\|^2 = \mathcal{O}_p(h_n^2/\rho^2)$$

using (A.7) and (A.3). The second part is such that

$$\|(\rho + H_n)^{-1}(H - H_n)(\rho + H)^{-1}HH^{-1/2}T\|^2 = \mathcal{O}_p(h_n^2/\rho^2)$$

using (A.7), (A.1), (A.3) and also that $H^{-1/2}T$ is bounded.

The third term of (A.8) is such that

$$(B_\rho - B)T = -\rho(\rho + H)^{-1}H^{-1/2}T$$

and therefore, if $\{(\nu_j, \psi_j, \tilde{\psi}_j)_j\}$ denotes the Singular Value Decomposition of the compact operator H ,

$$\begin{aligned}\|(B_\rho - B)T\|^2 &= \sup_{\|g\|=1} \sum_j \frac{\rho^2}{(\rho + \nu_j)^2 \nu_j} |\langle Tg, \psi_j \rangle|^2 \\ &\leq \rho^2 \left(\sup_j \frac{\nu_j^\theta}{(\rho + \nu_j)^2} \right) \left(\sup_{\|g\|=1} \sum_i \frac{|\langle H^{-1/2}Tg, \psi_i \rangle|^2}{\nu_i^\theta} \right) \\ &= \mathcal{O}(\rho^{\min(\theta, 2)})\end{aligned}$$

from (A.1) and for some $\theta > 0$ given in Assumption 10.

The control of the fourth term of (A.8) is analogous and we skip the details. $\square$

Proof of the Theorem. Denote $S_n = (\alpha I + T_n^* B_n^* B_n T_n)^{-1} T_n^* B_n^*$. With T defined in Assumption 3, denote also $S = (\alpha I + T^* B^* B T)^{-1} T^* B^*$.

The proof is based on the decomposition

$$\hat{\varphi}_{\alpha, n} - \varphi = A_n^{(1)} + A_n^{(2)} + A_n^{(3)}$$

where $A_n^{(1)}$, $A_n^{(2)}$ and $A_n^{(3)}$ are, respectively, the stochastic term, the estimation bias term and the regularization bias term that are given by

$$\begin{aligned}A_n^{(1)} &= S_n B_n \varepsilon_n \\ A_n^{(2)} &= S B T \varphi - \varphi \\ A_n^{(3)} &= S_n B_n T_n \varphi - S B T \varphi.\end{aligned}$$

Note that the three terms depend on the regularization parameters α and ρ . Below we control the square norm of each term asymptotically.

Bound for $A_n^{(1)}$ (Stochastic term). It is useful to decompose this term as follows:

$$A_n^{(1)} = SB\varepsilon_n + (S_n - S)B\varepsilon_n + S_n(B_n - B)\varepsilon_n. \quad (\text{A.9})$$

Remind that the singular value decomposition of BT has been denoted by $\{(\kappa_j, \xi_j, \tilde{\xi}_j), j = 1, 2, \dots\}$. Therefore the first term is such that

$$\mathbb{E}\|SB\varepsilon_n\|^2 = \sum_j \frac{\kappa_j^2}{(\alpha + \kappa_j^2)^2} \mathbb{E}|\langle B\varepsilon_n, \xi_j \rangle|^2.$$

Since the covariance operator of ε_n is $n^{-1}\Omega$ we can write $\mathbb{E}|\langle B\varepsilon_n, \xi_j \rangle|^2 = n^{-1}|\langle B\Omega B^* \xi_j, \xi_j \rangle|$. Using Assumption 9 can write

$$\mathbb{E}\|SB\varepsilon_n\|^2 \lesssim \frac{1}{n} \sup_j \frac{\kappa_j^{2(1+\gamma)}}{(\alpha + \kappa_j^2)^2} \leq \frac{1}{\alpha^{1-\min(\gamma,1)} n}.$$

We now study the second term of (A.9). With $L_n = T_n^* B_n^* B_n T_n$ and $L = T^* B^* B T$ this term writes

$$\begin{aligned} & (S_n - S)B\varepsilon_n \\ &= \{(\alpha I + L_n)^{-1} T_n^* B_n^* - (\alpha I + L)^{-1} T^* B^*\} B\varepsilon_n \\ &= (\alpha I + L_n)^{-1} (T_n^* B_n^* - T^* B^*) B\varepsilon_n + \{(\alpha I + L_n)^{-1} - (\alpha I + L)^{-1}\} T^* B^* B\varepsilon_n \end{aligned}$$

Using (A.3), Lemma A.2 and using that $\mathbb{E}\|B\varepsilon_n\|^2 \leq 1/n$, the first of these two terms has an expected squared norm such that

$$\mathbb{E}\|(\alpha I + L_n)^{-1} (T_n^* B_n^* - T^* B^*) B\varepsilon_n\|^2 \leq \mathcal{O}_p \left(\frac{1}{\alpha^2 n} \left\{ \frac{a_n^2}{\rho} + \frac{h_n^2}{\rho^2} + \rho^{\min(\theta,2)} \right\} \right)$$

By definition of L_n and L , the second term is such that

$$\begin{aligned} & \{(\alpha I + L_n)^{-1} - (\alpha I + L)^{-1}\} T^* B^* B\varepsilon_n \\ &= (\alpha I + L_n)^{-1} (T^* B^* B T - T_n^* B_n^* B_n T_n) (\alpha I + L)^{-1} T^* B^* B\varepsilon_n \\ &= (\alpha I + L_n)^{-1} T_n^* B_n^* (B T - B_n T_n) (\alpha I + L)^{-1} T^* B^* B\varepsilon_n \\ &\quad + (\alpha I + L_n)^{-1} (T^* B^* - T_n^* B_n^*) B T (\alpha I + L)^{-1} T^* B^* B\varepsilon_n \end{aligned}$$

Using similar arguments as before, the first of these two terms has an expected squared norm bounded by

$$\begin{aligned} & \mathbb{E}\|(\alpha I + L_n)^{-1} T_n^* B_n^*\|^2 \|B T - B_n T_n\|^2 \|(\alpha I + L)^{-1} T^* B^* B\varepsilon_n\|^2 \\ & \lesssim \frac{1}{\alpha} \left(\frac{a_n^2}{\rho} + \frac{h_n^2}{\rho^2} + \rho^{\min(\theta,2)} \right) \frac{1}{n \alpha^{1-\min(1,\gamma)}}. \end{aligned}$$

while the second term leads to the lower rate of convergence.

$$\frac{1}{\alpha^2} \left(\frac{a_n^2}{\rho} + \frac{h_n^2}{\rho^2} + \rho^{\min(\theta, 2)} \right) \frac{1}{n}$$

We now bound the last term of (A.9). This term is such that

$$\begin{aligned} S_n(B_n - B)\varepsilon_n &= S_n(\rho + H_n)^{-1} \{H_n^{1/2} - (\rho I + H_n)H^{-1/2}\}\varepsilon_n \\ &= -\rho S_n(\rho + H_n)^{-1} H^{-1/2} \varepsilon_n + S_n(\rho + H_n)^{-1} H_n^{1/2} \{H^{1/2} - H_n^{1/2}\} H^{-1/2} \varepsilon_n \end{aligned}$$

leading to two terms. Using Lemma A.1 as above, the expected square norm of the first term is bounded by $\alpha^{-1} \mathbb{E} \|H^{-1/2} \varepsilon_n\|^2 = (\alpha n)^{-1} \text{tr}(H^{-1/2} \Omega H^{-1/2})$. This rate is dominated by the rate of convergence of the expected square norm of the second term that is $(\alpha \rho n)^{-1} h_n^2 \text{tr}(H^{-1/2} \Omega H^{-1/2})$.

The final rate of convergence in probability of the stochastic term is

$$\frac{1}{n\alpha^{1-\min(1, \gamma)}} + \frac{1}{\alpha^2 n} \left(\frac{a_n^2}{\rho} + \frac{h_n^2}{\rho^2} + \rho^{\min(\theta, 2)} \right).$$

Bound for $A_n^{(2)}$ (Regularization bias). We can bound the regularization bias using the Assumption 8:

$$\begin{aligned} \|A_n^{(2)}\|^2 &= \|(\alpha I + T^* B^* B T)^{-1} T^* B^* B T \varphi - \varphi\|^2 \\ &= \|\alpha(\alpha I + T^* B^* B T)^{-1} \varphi\|^2 \\ &= \|\alpha(\alpha I + T^* B^* B T)^{-1} (T^* B^* B T)^{\beta/2} g\|^2 \end{aligned}$$

for some function g in $\mathcal{F}$,

$$\begin{aligned} &\leq \sup_{\lambda > 0} \frac{\alpha^2 \lambda^{2\beta}}{(\alpha + \lambda^2)^2} \\ &\leq \alpha^{\min(\beta, 2)} \end{aligned}$$

Bound for $A_n^{(3)}$ (Estimation bias). If we denote $L_n = T_n^* B_n^* B_n T_n$ and $L = T^* B^* B T$ as above in this proof, the estimation bias term writes

$$(\alpha I + L_n)^{-1} L_n \varphi - (\alpha I + L)^{-1} L \varphi = \alpha(\alpha I + L_n)^{-1} (L_n - L) L^{-1} (\alpha I + L)^{-1} L \varphi \quad (\text{A.10})$$

Using the identity

$$L_n - L = T_n^* B_n^* (B_n T_n - B T) + (T_n^* B_n^* - T^* B^*) B T$$

the estimation bias term (A.10) leads to two terms. With $\varphi = (B T)^\beta g$ for some $g \in \mathcal{F}$, the first term is

$$\alpha(\alpha I + L_n)^{-1} T_n^* B_n^* (B_n T_n - B T) L^{-1} (\alpha I + L)^{-1} L (B T)^\beta g$$

and the expectation of its square norm has rate

$$\frac{1}{\alpha^{1-\beta}} \left(\frac{1}{\rho n} + \frac{h_n^2}{\rho} + \rho^{\min(\theta, 2)} \right). \quad (\text{A.11})$$

The second term is treated analogously and leads to the same rate of convergence. The concatenation of all rates of convergence gives the final result. $\square$

A.2 Proof of Theorem 4.2

As in previous proofs, we denote $S_n = (\alpha I + T_n^* B_n^* B_n T_n)^{-1} T_n^* B_n^*$. With T defined in Assumption 3, denote also $S = (\alpha I + T^* B^* B T)^{-1} T^* B^*$ with $B = \Omega^{-\delta/2}$. The inner product of interest is such that

$$\langle \hat{\varphi}_{\alpha, n} - \varphi_\alpha, g \rangle = \langle (S_n B_n T_n - S B T) \varphi, g \rangle + \langle S_n B_n \varepsilon_n, g \rangle$$

A Cauchy Schwarz inequality on the first term and the control of $A_n^{(3)}$ in the proof of Theorem 4.1 show that the first term converges in probability to zero with rate given by (A.11). The second term is itself decomposed as follows:

$$\langle S_n B_n \varepsilon_n, g \rangle = \langle S B \varepsilon_n, g \rangle + \langle (S_n - S) B \varepsilon_n, g \rangle + \langle S_n (B_n - B) \varepsilon_n, g \rangle$$

In the proof of Theorem 4.1 the control of $A_n^{(1)}$ shows that the two last terms converge in probability under the assumptions of Theorem 4.2.

We now treat the term leading the asymptotic distribution which is $\langle S B \varepsilon_n, g \rangle$. By definition of ε_n we can write

$$\langle S B \varepsilon_n, g \rangle = \sum_{i=1}^n X_{i, n}$$

where

$$X_{i, n} = \frac{1}{n} \langle u_i w_i, B^* S^* g \rangle$$

is a zero-mean random variable and such that

$$\beta_n^2 := \text{Var} \left(\sum_{i=1}^n X_{i, n} \right) = \frac{1}{n} \|\Omega^{1/2} B^* S^* g\|^2.$$

We will now show that $X_{i, n}$ fulfills sufficient assumptions to apply the Lindeberg-Feller Central Limit Theorem (e.g. Serfling, 1980). The condition we are going to check is $\mathbb{E} \sum_i |X_{i, n}|^\nu = o(\beta_n^\nu)$ for some $\nu > 2$. Using the assumptions of the theorem and by definition of S we can write

$$\begin{aligned} \mathbb{E} \sum_i |X_{i, n}|^\nu &\leq \frac{1}{n^\nu} \sum_i \mathbb{E} |\langle u_i w_i; B^* S^* g \rangle|^\nu \\ &\lesssim \frac{1}{n^\nu} \|\Omega^{-\delta/2} (\alpha_n I_n + T^* \Omega^{-\delta} T)^{-1} T^* \Omega^{-\delta/2} g\|^\nu. \end{aligned}$$

In the proof of Proposition 4.2 we showed that Assumption 12 implies $\|\Omega^{-\delta/2}T\| \asymp \|\mathcal{L}^{-a}g\|$ with $a = k - b\delta$. If we denote by (a_i, v_i) the singular value decomposition of the operator $\mathcal{L}$ appearing in Assumption 12 the last expression is therefore bounded by

$$\frac{1}{n^\nu} \left(\sum_j \frac{a_j^{b\delta-a}}{\alpha_n + a_j^{-2a}} \langle g, v_j \rangle v_j \right)^\nu \lesssim \frac{\alpha_n^{\nu \min\{1, (a-b\delta)/(2a)\}}}{n^\nu \alpha_n^\nu}$$

using (A.1). Similarly the variance term is such that

$$\beta_n^2 = \frac{1}{n} \sum_j \left(\frac{a_j^{b(\delta-1)-a}}{\alpha_n + a_j^{-2a}} \right)^2 \langle g, v_j \rangle^2 \quad (\text{A.12})$$

$$\begin{aligned} &= \frac{1}{n} \sum_{j: \alpha_n > a_j^{-2a}} \left(\frac{a_j^{b(\delta-1)-a}}{\alpha_n + a_j^{-2a}} \right)^2 \langle g, v_j \rangle^2 + \frac{1}{n} \sum_{j: \alpha_n \leq a_j^{-2a}} \left(\frac{a_j^{b(\delta-1)-a}}{\alpha_n + a_j^{-2a}} \right)^2 \langle g, v_j \rangle^2 \\ &\geq \frac{1}{4\alpha_n^2 n} \sum_{j: \alpha_n > a_j^{-2a}} \left(a_j^{b(\delta-1)-a} \right)^2 \langle g, v_j \rangle^2 + \frac{1}{4n} \sum_{j: \alpha_n \leq a_j^{-2a}} \left(a_j^{b(\delta-1)+a} \right)^2 \langle g, v_j \rangle^2 \\ &\geq \frac{1}{4n} \sum_{j: \alpha_n \leq a_j^{-2a}} \left(a_j^{b(\delta-1)+a} \right)^2 \langle g, v_j \rangle^2 \end{aligned} \quad (\text{A.13})$$

where the set $\{j : \alpha_n \leq a_j^{-2a}\}$ is growing when $\alpha_n \rightarrow 0$. As a consequence, the ratio $\mathbb{E} \sum_i |X_{i,n}|^\nu / \beta_n^\nu$ which is bounded from above up to a constant by

$$\frac{\alpha_n^{\nu \min\{1, (a-b\delta)/(2a)\}}}{n^\nu \alpha_n^\nu} \frac{n^{\nu/2}}{\sum_{j: \alpha_n \leq a_j^{-2a}} \left(a_j^{b(\delta-1)+a} \right)^2 \langle g, v_j \rangle^2}$$

converges to zero provided that $\sqrt{n} \alpha^{1-(a-b\delta)/(2a)}$ as stated in the theorem.

From Assumption 12, note that $a_j = \kappa_j^{-1/a}$ where κ_j are the eigenvalues of $\Sigma^{-\delta/2}T$. Therefore (A.13) gives the lower bound stated for $n\beta_n^2$ in the theorem. It remains to show the upper bound. Given (A.12) and inequality (A.1),

$$\beta_n^2 \leq \frac{\alpha_n^{\min\{2, 1-b(\delta-1)/a\}}}{\alpha_n^2 n}.$$

If $b \geq 0$, the minimum involved in the last expression is given by $1 - b(\delta - 1)/a$ and the result is proved.

References

- Angrist, J., Lavy, V., & Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4), 773-824.
- Ashraf, Q. H., Weil, D. N., & Wilde, J. (2013). The effect of fertility reduction on economic growth. *Population and Development Review*, 39(1), 97-130.

- Bagdziunas, R., & Van Bellegem, S. (2014). *Non linear estimation in endogenous functional linear regression models* (Discussion paper). CORE, Université catholique de Louvain.
- Barro, R. J., & Becker, G. S. (1989). Fertility choice in a model of economic growth. *Econometrica*, 57(2), 481-501.
- Belloni, A., & Chernozhukov, V. (2009, May). *L1-penalized quantile regression in high-dimensional sparse models* (CeMMAP working papers No. CWP10/09). Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2011, December). *Inference for high-dimensional sparse econometric models* (CeMMAP working papers No. CWP41/11). Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- Blackburn, K., & Cipriani, G. P. (2002). A model of longevity, fertility and growth. *Journal of Economic Dynamics and Control*, 26(2), 187-204.
- Braakmann, N., & Wildman, J. (2014). *Fertility treatments and the use of twin births as an instrument for fertility* (MRPA Paper No. 54106). Munich.
- Bronars, S. G., & Grogger, J. (1994). The economic consequences of unwed motherhood: Using twin births as a natural experiment. *American Economic Review*, 84(5), 1141-1156.
- Cardot, H., & Johannes, J. (2010). Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis*, 101, 395-408.
- Carrasco, M., & Florens, J.-P. (2000). Generalization of gmm to a continuum of moment conditions. *Econometric Theory*, 16(06), 797-834.
- Carrasco, M., Florens, J.-P., & Renault, E. (2007). Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In J. Heckman & E. Leamer (Eds.), *Handbook of econometrics* (Vol. 6B, pp. 5633-5746). Elsevier.
- Chao, J. C., & Swanson, N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica*, 73(5), 1673-1692.
- Chen, X., & Reiss, M. (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27, 497-521.
- De la Croix, D., & Doepke, M. (2003). Inequality and growth: why differential fertility matters. *American Economic Review*, 93, 1091-1113.
- Engl, H. W., Hanke, M., & Neubauer, A. (2000). *Regularization of inverse problems*. Dordrecht: Kluwer Academic.
- Fan, J., Fan, Y., & Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1), 186-197.
- Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis*. New York: Springer.
- Fève, F., & Florens, J.-P. (2010). The practice of non-parametric estimation by solving inverse problems: the example of transformation models. *The Econometrics Journal*, 13, S1-S27.
- Florens, J.-P., Johannes, J., & Van Bellegem, S. (2011). Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, 27, 472-496.
- Hall, P., & Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of

- instrumental variables. *Annals of Statistics*, 33, 2904–2929.
- Hall, P., & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35, 70–91.
- Hansen, C., Hausman, J., & Newey, W. (2008). Estimation with many instrumental variables. *Journal of Business and Economic Statistics*, 26, 398–422.
- Johannes, J., Van Bellegem, S., & Vanhems, A. (2011). Convergence rates for ill-posed inverse problems with an unknown operator. *Econometric Theory*, 27, 522–545.
- Kokoszka, P., & Horvath, L. (2012). *Inference for functional data with applications*. New York: Springer.
- Kunitomo, N. (1980). Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association*, 75, 693–700.
- Mendelson, S., & Pajor, A. (2006). On singular values of matrices with independent rows. *Bernoulli*, 12, 761–773.
- Morimune, K. (1983). Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica*, 51, 821–841.
- Ramsay, J., & Silvermann, B. W. (2005). *Functional data analysis*. New York: Springer.
- Rosenzweig, M. R., & Wolpin, K. I. (1980). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica*, 48(1), 227–240.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons.
- Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4), 518–29.
- Sutcliffe, A. G., & Ludwig, M. (2007). Outcome of assisted reproduction. *The Lancet*, 370(9584), 351–359.
- Swanson, N. R., Chao, J. C., Hausman, J. A., Newey, W. K., & Woutersen, T. (2011, May). *Instrumental variable estimation with heteroskedasticity and many instruments* (Departmental Working Papers No. 201111). Rutgers University, Department of Economics.

Recent titles

CORE Discussion Papers

- 2014/14 Jean-François CARPANTIER and Arnaud DUFAYS. Specific Markov-switching behaviour for ARMA parameters.
- 2014/15 Federico GRIGIS DE STEFANO. Strategic stability of equilibria: the missing paragraph.
- 2014/16 Claudio TELHA and Mathieu VAN VYVE. Efficient approximation algorithms for the economic lot-sizing in continuous time.
- 2014/17 Yukai YANG. Testing constancy of the error covariance matrix in vector models against parametric alternatives using a spectral decomposition.
- 2014/18 Koen DECANCO, Marc FLEURBAEY and Erik SCHOKKAERT. Inequality, income, and well-being.
- 2014/19 Paul BELLEFLAMME and Martin PEITZ. Digital piracy: an update.
- 2014/20 Eva-Maria SCHOLZ. Licensing to vertically related markets.
- 2014/21 N. Baris VARDAR. Optimal energy transition and taxation of non-renewable resources.
- 2014/22 Benoît DECERF. Income poverty measures with relative poverty lines.
- 2014/23 Antoine DEDRY, Harun ONDER and Pierre PESTIEAU. Aging, social security design and capital accumulation.
- 2014/24 Biung-Ghi JU and Juan D. MORENO-TERNERO. Fair allocation of disputed properties.
- 2014/25 Nguyen Thang DAO. From agriculture to manufacture: How does geography matter ?
- 2014/26 Xavier Y. WAUTHY. From Bertrand to Cournot via Kreps and Scheinkman: a hazardous journey.
- 2014/27 Gustavo BERGANTIÑOS and Juan MORENO-TERNERO. The axiomatic approach to the problem of sharing the revenue from bundled pricing.
- 2014/28 Jean HINDRIKS and Yukihiro NISHIMURA. International tax leadership among asymmetric countries.
- 2014/29 Jean HINDRIKS and Yukihiro NISHIMURA. A note on equilibrium leadership in tax competition models.
- 2014/30 Olivier BOS and Tom TRUYTS. Auctions with prestige motives.
- 2014/31 Juan D. MORENO-TERNERO and Lars P. ØSTERDAL . Normative foundations for equity-sensitive population health evaluation functions.
- 2014/32 P. Jean-Jacques HERINGS, Ana MAULEON and Vincent VANNETELBOSCH. Stability of networks under Level-K farsightedness.
- 2014/33 Lionel ARTIGE, Laurent CAVENAILE and Pierre PESTIEAU. The macroeconomics of PAYG pension schemes in an aging society.
- 2014/34 Tanguy KEGELART and Mathieu VAN VYVE. A conic optimization approach for SKU rationalization.
- 2014/35 Ulrike KORNEK, Kei LESSMANN and Henry TULKENS. Transferable and non transferable utility implementations of coalitional stability in integrated assessment models.
- 2014/36 Ibrahim ABADA, Andreas EHRENMANN and Yves SMEERS. Endogenizing long-term contracts in gas market models.
- 2014/37 Julio DAVILA. Output externalities on total factor productivity.
- 2014/38 Diane PIERRET. Systemic risk and the solvency-liquidity nexus of banks.
- 2014/39 Paul BELLEFLAMME and Julien JACQMIN. An economic appraisal of MOOC platforms: business models and impacts on higher education.
- 2014/40 Marie-Louise LEROUX, Pierre PESTIEAU and Grégory PONTIERE. Longévité différentielle et redistribution: enjeux théoriques et empiriques.
- 2014/41 Chiara CANTA, Pierre PESTIEAU and Emmanuel THIBAUT. Long term care and capital accumulation: the impact of the State, the market and the family.
- 2014/42 Gilles GRANDJEAN, Marco MANTOVANI, Ana MAULEON and Vincent VANNETELBOSCH. Whom are you talking with ? An experiment on credibility and communication structure.
- 2014/43 Julio DAVILA. The rationality of expectations formation.

Recent titles

CORE Discussion Papers - continued

- 2014/44 Florian MAYNERIS, Sandra PONCET and Tao ZHANG. The cleaning effect of minimum wages. Minimum wages, firm dynamics and aggregate productivity in China.
- 2014/45 Thierry BRECHET, Natali HRITONENKOVA and Yuri YATSENKO. Domestic environmental policy and international cooperation for global commons.
- 2014/46 Mathieu PARENTI, Philip USHCHEV and Jacques-François THISSE. Toward a theory of monopolistic competition.
- 2014/47 Takatoshi TABUCHI, Jacques-François THISSE and Xiwei ZHU. Does technological progress affect the location of economic activity?
- 2014/48 Paul CASTANEDA DOWER, Victor GINSBURGH and Shlomo WEBER. Colonial legacy, linguistic disenfranchisement and the civil conflict in Sri Lanka.
- 2014/49 Victor GINSBURGH, Jacques MELITZ and Farid TOUBAL. Foreign language learnings: An econometric analysis.
- 2014/50 Koen DECANQ and Dirk NEUMANN. Does the choice of well-being measure matter empirically? An illustration with German data.
- 2014/51 François MANIQUET. Social ordering functions.
- 2014/52 Ivar EKELAND and Maurice QUEYRANNE. Optimal pits and optimal transportation.
- 2014/53 Luc BAUWENS, Manuela BRAIONE and Giuseppe STORTI. Forecasting comparison of long term component dynamic models for realized covariance matrices.
- 2014/54 François MANIQUET and Philippe MONGIN. Judgment aggregation theory can entail new social choice results.
- 2014/55 Pasquale AVELLA, Maurizio BOCCIA and Laurence A. WOLSEY. Single-period cutting planes for inventory routing problems.
- 2014/56 Jean-Pierre FLORENS and Sébastien VAN BELLEGEM. Instrumental variable estimation in functional linear models.

Books

- V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
- I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
- W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
- L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
- J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
- M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
- J. HINDRIKS (2012), *Gestion publique*. De Boeck.
- M. FUJITA and J.F. THISSE (2013), *Economics of agglomeration: cities, industrial location, and globalization*. (2nd edition). Cambridge University Press.
- J. HINDRIKS and G.D. MYLES (2013). *Intermediate public economics*. (2nd edition). MIT Press.
- J. HINDRIKS, G.D. MYLES and N. HASHIMZADE (2013). *Solutions manual to accompany intermediate public economics*. (2nd edition). MIT Press.

CORE Lecture Series

- R. AMIR (2002), Supermodularity and complementarity in economics.
- R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
- A. SHAPIRO (2010), Stochastic programming: modeling and theory.