

2013s-21

# **Efficient estimation with many weak instruments using regularization techniques**

*Marine Carrasco, Guy Tchuente*

---

**Série Scientifique**  
*Scientific Series*

---

**Montréal**  
**Juillet 2013**

© 2013 *Marine Carrasco, Guy Tchuente*. Tous droits réservés. *All rights reserved*. Reproduction partielle permise avec citation du document source, incluant la notice ©.  
*Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*



Centre interuniversitaire de recherche en analyse des organisations

## **CIRANO**

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère du Développement économique et régional et de la Recherche, de même que des subventions et mandats obtenus par ses équipes de recherche.

*CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère du Développement économique et régional et de la Recherche, and grants and research mandates obtained by its research teams.*

### **Les partenaires du CIRANO**

#### **Partenaire majeur**

Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie

#### **Partenaires corporatifs**

Autorité des marchés financiers  
Banque de développement du Canada  
Banque du Canada  
Banque Laurentienne du Canada  
Banque Nationale du Canada  
Banque Scotia  
Bell Canada  
BMO Groupe financier  
Caisse de dépôt et placement du Québec  
Fédération des caisses Desjardins du Québec  
Financière Sun Life, Québec  
Gaz Métro  
Hydro-Québec  
Industrie Canada  
Investissements PSP  
Ministère des Finances et de l'Économie  
Power Corporation du Canada  
Rio Tinto Alcan  
State Street Global Advisors  
Transat A.T.  
Ville de Montréal

#### **Partenaires universitaires**

École Polytechnique de Montréal  
École de technologie supérieure (ÉTS)  
HEC Montréal  
Institut national de la recherche scientifique (INRS)  
McGill University  
Université Concordia  
Université de Montréal  
Université de Sherbrooke  
Université du Québec  
Université du Québec à Montréal  
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

*This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.*

ISSN 1198-8177

Partenaire financier

Enseignement supérieur,  
Recherche, Science  
et Technologie  
Québec 

# Efficient estimation with many weak instruments using regularization techniques<sup>\*</sup>

*Marine Carrasco<sup>†</sup>, Guy Tchuente<sup>‡</sup>*

## **Résumé / Abstract**

The problem of weak instruments is due to a very small concentration parameter. To boost the concentration parameter, we propose to increase the number of instruments to a large number or even up to a continuum. However, in finite samples, the inclusion of an excessive number of moments may be harmful. To address this issue, we use regularization techniques as in Carrasco (2012) and Carrasco and Tchuente (2013). We show that normalized regularized 2SLS and LIML are consistent and asymptotically normally distributed. Moreover, their asymptotic variances reach the semiparametric efficiency bound unlike most competing estimators. Our simulations show that the leading regularized estimators (LF and T of LIML) work very well (is nearly median unbiased) even in the case of relatively weak instruments.

**Mots clés/Keywords** : Many weak instruments, LIML, 2SLS, regularization methods.

---

<sup>\*</sup> Carrasco gratefully acknowledges financial support from SSHRC.

<sup>†</sup> University of Montreal, CIREQ, CIRANO, email: [marine.carrasco@umontreal.ca](mailto:marine.carrasco@umontreal.ca).

<sup>‡</sup> University of Montreal, CIREQ, email: [guy.tchuente.nguembu@umontreal.ca](mailto:guy.tchuente.nguembu@umontreal.ca).

# 1 Introduction

The problem of weak instruments or weak identification has recently received considerable attention in both theoretical and applied econometrics<sup>1</sup>. Empirical examples include Angrist and Krueger (1991) who measure return to schooling, Eichenbaum, Hansen, and Singleton (1988) who consider consumption asset pricing models. Theoretical literature on weak instruments includes papers by Staiger and Stock (1997), Zivot, Startz, and Nelson (1998), Guggenberger and Smith (2005), Chao and Swanson (2005), Han and Phillips (2006), Hansen, Hausman, and Newey (2008), and Newey and Windmeijer (2009) among others<sup>2</sup>. Staiger and Stock (1997) proposed an asymptotic framework with local-to-zero parametrization of the coefficients of the instruments in the first-stage regression. They show that with the number of instruments fixed, the two-stage least squares (2SLS) and limited information maximum likelihood (LIML) estimators are not consistent and converge to nonstandard distributions. Subsequent work focused on situations where the number of instruments is large, using an asymptotic framework that lets the number of instruments go to infinity as a function of sample size. In these settings, the use of many moments can improve estimator accuracy. Unfortunately, usual Gaussian asymptotic approximation can be poor and IV estimators can be biased.

Carrasco (2012) and Carrasco and Tchuente (2013) proposed respectively regularized versions of 2SLS and LIML estimators for many strong instruments. The regularization permits to address the singularity of the covariance matrix resulting from many instruments. These papers use three regularization methods borrowed from inverse problem literature. The first estimator is based on Tikhonov (ridge) regularization, the second estimator is based on an iterative method called Landweber-Fridman (LF), the third estimator is based on the principal components associated with the largest eigenvalues. We extend these previous works to allow for the presence of a

---

<sup>1</sup>Hahn and Hausman (2003) define weak instruments, by two features: (i) two-stage least squares (2SLS) analysis is badly biased toward the ordinary least-squares (OLS) estimate, and alternative unbiased estimators such as limited-information maximum likelihood (LIML) may not solve the problem; and (ii) the standard (first-order) asymptotic distribution does not give an accurate framework for inference. Weak instrument may also be an important cause of the finding that the often-used test of over identifying restrictions (OID test) rejects too often.

<sup>2</sup>Section 4 discusses related literature in more details.

large number of weak instruments or weak identification. We consider a linear model with homoskedastic error and allow for weak identification as in Hansen, Hausman, and Newey (2008) and Newey and Windmeijer (2009). This specification helps us to have different types of weak instruments sequences, including the many instruments sequence of Bekker (1994) and the many weak instruments sequence of Chao and Swanson (2005). We impose no condition on the number of moment conditions since our framework allows for an infinite countable or even a continuum of instruments. The advantage of regularization is that all available moments can be used without discarding any a priori. We show that regularized 2SLS and LIML are consistent in the presence of many weak instruments. If properly normalized, the regularized 2SLS and LIML are asymptotically normal and reach the semiparametric efficiency bounds. Therefore, their asymptotic variance is smaller than that of Hansen, Hausman, and Newey (2008) and Newey and Windmeijer (2009). All these methods involve a regularization parameter, which is the counterpart of the smoothing parameter that appears in the nonparametric literature. A data driven method was developed in Carrasco (2012) and Carrasco and Tchuente (2013) to select the best regularization parameter when the instruments are strong. We use these methods in our simulations for selecting the regularization parameter when the instruments are weak but we do not prove that these methods are valid in this case.

A Monte-carlo experiment shows that the leading regularized estimators (LF and T LIML) perform very well (are nearly median unbiased) even in the case of weak instruments.

The paper is organized as follows. Section 2 introduces four regularization methods we consider and the associated estimators. Section 3 derives the asymptotic properties of the estimators. Section 4 discusses efficiency and related results. Section 5 presents Monte Carlo experiments. Section 6 concludes. The proofs are collected in Appendix.

## 2 Presentation of the regularized 2SLS and LIML estimators

This section presents the weak instruments setup used in this paper and the estimators. Estimators studied here are the regularized 2SLS and LIML estimators introduced in Carrasco (2012) and Carrasco and Tchuente (2013). They can be used with many or even a continuum of instruments. This work extends previous works by allowing for weak instruments as in Hansen, Hausman, and Newey (2008).

Our model is inspired from Hausman, Newey, Woutersen, Chao, and Swanson (2012). The model is

$$\begin{cases} y_i = W_i' \delta_0 + \varepsilon_i, \\ W_i = \gamma_i + u_i, \end{cases}$$

$i = 1, 2, \dots, n$ . The vector of interest  $\delta_0$  is  $p \times 1$ .

$E(u_i|x_i) = E(\varepsilon_i|x_i) = 0$ ;  $E(\varepsilon_i^2|x_i) = \sigma_\varepsilon^2$ .  $y_i$  is a scalar and  $x_i$  is a vector of exogenous variables. Some rows of  $W_i$  may be exogenous, with the corresponding rows of  $u_i$  being zero.  $\gamma_i = E(W_i|x_i)$  is a  $p \times 1$  vector of reduced form values with  $E(\gamma_i \varepsilon_i) = 0$ .  $\gamma_i$  is the optimal instrument which is typically unknown. The estimation is based on a sequence of instruments  $Z_i = Z(\tau; x_i)$ .  $\tau \in S$  may be an integer or an index taking its values in an interval. Examples of  $Z_i$  are the following.

- $S = \{1, 2, \dots, L\}$  thus we have  $L$  instruments.
- $Z_{ij} = (x_i)^{j-1}$  with  $j \in S = \mathbb{N}$  infinite countable instruments.
- $Z_i = Z(\tau; x_i) = \exp(i\tau' x_i)$  where  $\tau \in S = \mathbb{R}^{\dim(x_i)}$  continuum of moment

Let  $\pi$  be a positive measure on  $S$ . We denote  $L^2(\pi)$  the Hilbert space of square integrable functions with respect to  $\pi$ .

This model allows for  $\gamma_i$  to be a linear or a non linear combination of  $Z_i$ . The model also allows for  $\gamma_i$  to approximate the reduced form. For example, we could let  $\gamma_i$  be a vector of unknown functions of a vector  $x_i$  of underlying instruments.

The estimate  $\delta$  is based on the orthogonality condition.

$$E[(y_i - W_i' \delta) Z_i] = 0.$$

The extension of the generalized method of moments described in Carrasco and Florens (2000), Carrasco and Florens (2012), and Carrasco (2012) gives the regularized 2SLS.

$$\text{Let } W = \begin{pmatrix} W'_1 \\ W'_2 \\ \cdot \\ \cdot \\ W'_n \end{pmatrix} n \times p \text{ and } u = \begin{pmatrix} u'_1 \\ u'_2 \\ \cdot \\ \cdot \\ u'_n \end{pmatrix} n \times p.$$

We define the covariance operator  $K$  of the instruments as

$$K : L^2(\pi) \rightarrow L^2(\pi)$$

$$(Kg)(\tau_1) = \int E(Z(\tau_1; x_i) \overline{Z(\tau_2; x_i)}) g(\tau_2) \pi(\tau_2) d\tau_2$$

where  $\overline{Z(\tau_2; x_i)}$  denotes the complex conjugate of  $Z(\tau_2; x_i)$ .

$K$  is assumed to be a Hilbert-Schmidt operator. Let  $\lambda_j$  and  $\phi_j$ ,  $j = 1, 2, \dots$  be respectively, the eigenvalues (ranked in decreasing order) and orthonormal eigenfunctions of  $K$ .  $K$  can be estimated by  $K_n$  defined as:

$$K_n : L^2(\pi) \rightarrow L^2(\pi)$$

$$(K_n g)(\tau_1) = \int \frac{1}{n} \sum_{i=1}^n Z(\tau_1; x_i) \overline{Z(\tau_2; x_i)} g(\tau_2) \pi(\tau_2) d\tau_2.$$

If the number of moment conditions is infinite then the inverse of  $K_n$  needs to be regularized because it is not continuous. By definition (see Kress, 1999, page 269), a regularized inverse of an operator  $K$  is

$$R_\alpha : L^2(\pi) \rightarrow L^2(\pi)$$

such that  $\lim_{\alpha \rightarrow 0} R_\alpha K \varphi = \varphi$ ,  $\forall \varphi \in L^2(\pi)$ .

Three different types of regularization schemes are considered: Tikhonov (T), Landwerber Fridman (LF), Spectral cut-off (SC) or Principal Components (PC). They are defined as follows:

### 1. Tikhonov(T)

This regularization scheme is related to the ridge regression.

$$(K^\alpha)^{-1} = (K^2 + \alpha I)^{-1}K$$
$$(K^\alpha)^{-1}r = \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \alpha} \langle r, \phi_j \rangle \phi_j$$

where  $\alpha > 0$  is the regularization parameter. A fixed  $\alpha$  would result in a loss of efficiency. For the estimator to be asymptotically efficient,  $\alpha$  has to go to zero at a certain rate which will be determined later on.

### 2. Landweber Fridman (LF)

This method of regularization is iterative. Let  $0 < c < 1/\|K\|^2$  where  $\|K\|$  is the largest eigenvalue of  $K$  (which can be estimated by the largest eigenvalue of  $K_n$ ).  $\hat{\varphi} = (K^\alpha)^{-1}r$  is computed using the following algorithm:

$$\begin{cases} \hat{\varphi}_l = (1 - cK^2)\hat{\varphi}_{l-1} + cKr, & l=1,2,\dots,\frac{1}{\alpha} - 1; \\ \hat{\varphi}_0 = cKr, \end{cases}$$

where  $\frac{1}{\alpha} - 1$  is some positive integer. We also have

$$(K^\alpha)^{-1}r = \sum_{j=1}^{\infty} \frac{[1 - (1 - c\lambda_j^2)^{\frac{1}{\alpha}}]}{\lambda_j} \langle r, \phi_j \rangle \phi_j.$$

### 3. Spectral cut-off (SC)

This method consists in selecting the eigenfunctions associated with the eigenvalues greater than some threshold. The aim is to select those who have greater contribution.

$$(K^\alpha)^{-1}r = \sum_{\lambda_j^2 \geq \alpha} \frac{1}{\lambda_j} \langle r, \phi_j \rangle \phi_j$$

for  $\alpha > 0$ .

This method is similar to principal components (PC) which consists in using the



first eigenfunctions:

$$(K^\alpha)^{-1}r = \sum_{j=1}^{1/\alpha} \frac{1}{\lambda_j} \langle r, \phi_j \rangle \phi_j$$

where  $\frac{1}{\alpha}$  is some positive integer. It is equivalent to projecting on the first principal components of  $W$ . Interestingly, this approach is used in factor models where  $W_i$  is assumed to depend on a finite number of factors (see Bai and Ng (2002), Stock and Watson (2002)) As the estimators based on PC and SC are identical, we will use PC and SC interchangeably.

These regularized inverses can be rewritten in common notation as:

$$(K^\alpha)^{-1}r = \sum_{j=1}^{\infty} \frac{q(\alpha, \lambda_j^2)}{\lambda_j} \langle r, \phi_j \rangle \phi_j$$

where for T:  $q(\alpha, \lambda_j^2) = \frac{\lambda_j^2}{\lambda_j^2 + \alpha}$ ,

for LF:  $q(\alpha, \lambda_j^2) = [1 - (1 - c\lambda_j^2)^{1/\alpha}]$ ,

for SC:  $q(\alpha, \lambda_j^2) = I(\lambda_j^2 \geq \alpha)$ , for PC  $q(\alpha, \lambda_j^2) = I(j \leq 1/\alpha)$ .

In order to compute the inverse of  $K_n$  we have to choose the regularization parameter  $\alpha$ . Let  $(K_n^\alpha)^{-1}$  be the regularized inverse of  $K_n$  and  $P^\alpha$  a  $n \times n$  matrix defined as in Carrasco (2012) by  $P^\alpha = T(K_n^\alpha)^{-1}T^*$  where

$$T : L^2(\pi) \rightarrow \mathbb{R}^n$$

$$Tg = \begin{pmatrix} \langle Z_1, g \rangle \\ \langle Z_2, g \rangle \\ \cdot \\ \cdot \\ \langle Z_n, g \rangle \end{pmatrix}$$

and

$$T^* : \mathbb{R}^n \rightarrow L^2(\pi)$$

$$T^*v = \frac{1}{n} \sum_{i=1}^n Z_i v_i$$

such that  $K_n = T^*T$  and  $TT^*$  is an  $n \times n$  matrix with typical element  $\frac{\langle Z_i, Z_j \rangle}{n}$ . Let  $\hat{\phi}_j$ ,  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots > 0$ ,  $j = 1, 2, \dots$  be the orthonormalized eigenfunctions and eigenvalues of  $K_n$ .  $\hat{\lambda}_j$  are consistent estimators of  $\lambda_j$  the eigenvalues of  $TT^*$ . We then have  $T\hat{\phi}_j = \sqrt{\lambda_j}\psi_j$  and  $T^*\psi_j = \sqrt{\lambda_j}\hat{\phi}_j$ . For  $v \in \mathbf{R}^n$ ,  $P^\alpha v = \sum_{j=1}^{\infty} q(\alpha, \lambda_j^2) \langle v, \psi_j \rangle \psi_j$ . It follows that for any vectors  $v$  and  $w$  of  $\mathbf{R}^n$ :

$$\begin{aligned} v'P^\alpha w &= v'T(K_n^\alpha)^{-1}T^*w \\ &= \left\langle (K_n^\alpha)^{-1/2} \sum_{i=1}^n Z_i(\cdot) v_i, (K_n^\alpha)^{-1/2} \frac{1}{n} \sum_{i=1}^n Z_i(\cdot) w_i \right\rangle. \end{aligned} \quad (1)$$

Consider regularized k-class estimators defined as follows:

$$\hat{\delta}_\nu = (W'(P^\alpha - \nu I_n)W)^{-1}W'(P^\alpha - \nu I_n)y.$$

where  $\nu$  is either a constant term or a random variable. We are interested in the case where  $\nu = 0$  corresponding to regularized 2SLS studied in Carrasco (2012):

$$\hat{\delta} = (W'P^\alpha W)^{-1}W'P^\alpha y$$

and in the case  $\nu = \nu_\alpha = \min_{\delta} \frac{(y - W\delta)'P^\alpha(y - W\delta)}{(y - W\delta)'(y - W\delta)}$  corresponding to the regularized LIML studied in Carrasco and Tchuente (2013). We denote  $\hat{\delta}$  the 2SLS estimators and  $\hat{\delta}_L$  the LIML estimators.

We study both 2SLS and LIML because LIML may have some advantages over 2SLS. For example when the number of instruments,  $L$ , increases with the sample size,  $n$ , so that  $L/n \rightarrow c$  (with  $c$  constant), the standard 2SLS estimator is not consistent whereas standard LIML estimator is consistent.

### 3 Asymptotic properties

In Carrasco (2012) and Carrasco and Tchuente (2013), modified IV estimators based on different regularization schemes were proposed. These estimators improved the small sample properties of IV estimators when the number of instruments is very large or infinite. The focus was on strong instruments. They found that regularized 2SLS and LIML estimators are asymptotically normal and attain the semiparametric efficiency bound. We extend Carrasco (2012) and Carrasco and Tchuente (2013) results to the case of weak instruments.

**Assumption 1:**

- i) There exists a  $p \times p$  matrix  $S_n = \tilde{S}_n \text{diag}(\mu_{1n}, \dots, \mu_{pn})$  such that  $\tilde{S}_n$  is bounded, the smallest eigenvalue of  $\tilde{S}_n \tilde{S}_n'$  is bounded away from zero; for each  $j$ , either  $\mu_{jn} = \sqrt{n}$  (strong identification) or  $\mu_{jn}/\sqrt{n} \rightarrow 0$  (weak identification),  

$$\mu_n = \min_{1 \leq j \leq p} \mu_{jn} \rightarrow \infty \text{ and } \alpha \rightarrow 0.$$
- ii) There exists a function  $f_i = f(x_i)$  such that  $\gamma_i = S_n f_i / \sqrt{n}$  and  $\mu_n S_n^{-1} \rightarrow S_0$ .  

$$\sum_{i=1}^n \|f_i\|^4 / n^2 \rightarrow 0, \sum_{i=1}^n f_i f_i' / n \text{ is bounded and uniformly nonsingular.}$$

These conditions allow for many weak instruments. If  $\mu_{jn} = \sqrt{n}$  this leads to asymptotic theory like in Kunitomo (1980), Morimune (1983), and Bekker (1994), but here we use regularization parameter instead of having an increasing sequence of instruments. For  $\mu_n^2$  growing slower than  $n$  the convergence rate will be slower than  $\sqrt{n}$ , leading to an asymptotic approximation like that of Chao and Swanson (2005). This is the case where we have many instruments without strong identification. Assumption 1 also allows for some components of the reduced form to give only weak identification (corresponding to  $\mu_{jn}/\sqrt{n} \rightarrow 0$  which allows the concentration parameter to grow slower than  $\sqrt{n}$ ), and other components (corresponding to  $\mu_{jn} = \sqrt{n}$ ) to give strong identification for some coefficients of the reduced form. In particular, this condition allows for fixed constant coefficients in the reduced form. This specification of weak instruments can also be viewed as a generalization of Chao and Swanson (2007). This specification of weak identification is different of Antoine and Lavergne (2012) whose identification strength is directly defined through the conditional moments that flatten

as the sample size increases. To illustrate Assumption 1 let us consider the following example.

**Example 1:** Assume that  $p = 2$ ,  $\tilde{S}_n = \begin{pmatrix} 1 & 0 \\ \pi_{21} & 1 \end{pmatrix}$ , and  $\mu_{jn} = \begin{cases} \sqrt{n}, & j = 1 \\ \mu_n, & j = 2 \end{cases}$ .

Then for  $f(x_i) = (f'_{1i}, f'_{2i})'$  the reduced form is

$$\gamma_i = \begin{pmatrix} f_{1i} \\ \pi_{21}f_{1i} + \frac{\mu_n}{\sqrt{n}}f_{2i} \end{pmatrix}.$$

We also have

$$\mu_n S_n^{-1} \rightarrow S_0 = \begin{pmatrix} 0 & 0 \\ -\pi_{21} & 1 \end{pmatrix}.$$

**Proposition 1.** (*Asymptotic properties of regularized 2SLS with many weak instruments*)

Assume  $\{y_i; W_i; x_i\}$  are iid,  $E(\varepsilon_i^2|X) = \sigma_\varepsilon^2$ ,  $K$  is compact,  $\alpha$  goes to zero and  $n$  goes to infinity. Moreover,  $\gamma_a$  belongs to the closure of the linear span of  $\{Z(\cdot; x)\}$  for  $a = 1, \dots, p$ ,  $E(Z(\cdot, x_i)f_{ia})$  belong to the range of  $K$  and Assumption 1 is satisfied. Then, the  $T$ ,  $LF$ , and  $SC$  estimators of 2SLS satisfy:

1. *Consistency:*  $S'_n(\hat{\delta} - \delta_0)/\mu_n \rightarrow 0$  in probability as  $n$ ,  $n\alpha^{\frac{1}{2}}$  and  $\mu_n$  go to infinity. This implies  $(\hat{\delta} - \delta_0) \rightarrow 0$ .

2. *Asymptotic normality:* moreover, if each element of  $E(Z(\cdot; x_i)W_i)$  belongs to the range of  $K$ , then

$$S'_n(\hat{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\varepsilon^2 [E(f_i f'_i)]^{-1}\right)$$

as  $n$ ,  $n\alpha$  and  $\mu_n$  go to infinity; where  $E(f_i f'_i)$  is the  $p \times p$  matrix.

**Proof** In Appendix.

This result shows that the three estimators have the same asymptotic distribution. Instead of restricting the number of instruments, we have a regularization parameter which goes to zero. This insures us that all available and valid instruments are used in an efficient way even if there are weak. To obtain consistency, the condition on  $\alpha$  is  $n\alpha^{\frac{1}{2}}$  go infinity, whereas for the asymptotic normality, we need  $n\alpha$  go to infinity. This means that  $\alpha$  is allowed to go to zero at a slower rate. However this rate does not depend on the weakness of the instruments.

Interestingly, our regularized 2SLS estimators reach the semiparametric efficiency bound. This result will be further discussed in Section 4.

We are now going to derive the asymptotic properties of the regularized LIML with many weak instruments.

**Proposition 2.** (*Asymptotic properties of regularized LIML with many weak instruments*)

Assume  $\{y_i; W_i; x_i\}$  are iid,  $E(\varepsilon_i^2|X) = \sigma_\varepsilon^2$ ,  $E(\varepsilon_i^4|X) < \infty$ ,  $E(u_{bi}^4|X) < \infty$ ,  $K$  is compact,  $\alpha$  goes to zero and  $n$  goes to infinity. Moreover,  $\gamma_a$  belongs to the closure of the linear span of  $\{Z(\cdot; x)\}$  for  $a = 1, \dots, p$ ,  $E(Z(\cdot, x_i)f_{ia})$  belong to the range of  $K$  and Assumption 1 is satisfied. Then, the T, LF, and SC estimators of LIML with weak instruments satisfy:

1. Consistency:  $S'_n(\hat{\delta}_L - \delta_0)/\mu_n \rightarrow 0$  in probability as  $n$ ,  $\mu_n$  and  $\mu_n^2\alpha$  go to infinity. This implies  $(\hat{\delta} - \delta_0) \rightarrow 0$ .

2. Asymptotic normality: moreover, if each element of  $E(Z(\cdot; x_i)W_i)$  belongs to the range of  $K$ , then

$$S'_n(\hat{\delta}_L - \delta_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\varepsilon^2 [E(f_i f_i')]^{-1}\right)$$

as  $n$ ,  $\mu_n$  and  $\mu_n^2\alpha$  go to infinity where  $E(f_i f_i')$  is the  $p \times p$  matrix.

**Proof** In Appendix.

Interestingly we obtain the same asymptotic distribution as in the many strong instruments case (with a slower rate of convergence). We also find the same speed of convergence as in Newey and Windmeijer (2009) and Hansen, Hausman, and Newey (2008). For the consistency and asymptotic normality,  $\mu_n^2\alpha$  needs to go to infinity, which means that the regularization parameter should go to zero at a slower rate than the concentration parameter. The asymptotic variance of regularized LIML corresponds to the lower bound and is smaller than that obtained in Hansen, Hausman, and Newey (2008).

**Example 1:**(cont)

$$S'_n(\hat{\delta} - \delta_0) = \begin{pmatrix} \sqrt{n}[(\hat{\delta}_1 - \delta_{01}) + \pi_{21}(\hat{\delta}_2 - \delta_{02})] \\ \mu_n(\hat{\delta}_2 - \delta_{02}) \end{pmatrix} \text{ is jointly asymptotically normal.}$$

The linear combination  $(\hat{\delta}_1 - \delta_{01}) + \pi_{21}(\hat{\delta}_2 - \delta_{02})$  converges at rate  $\sqrt{n}$ . This is the co-

efficient of  $f_{i1}$  in the reduced form equation for  $y_i$ . And the estimator of the coefficients  $\delta_2$  of  $W_{i2}$  variables converges at rate  $\frac{1}{\mu_n}$ .

Now as in Newey and Windmeijer (2009) we consider a t-ratios for a linear combination  $c'\delta$  of the parameter of interest. The following proposition is a corollary of Proposition 2.

**Proposition 3.** *Under the assumptions of Proposition 2 and if we assume that there exist  $r_n$ ,  $c$  and  $c^* \neq 0$  such that  $r_n S_n^{-1} c \rightarrow c^*$  and  $S_n' \hat{\Phi} S_n / n \rightarrow \Phi$  in probability with  $\Phi = \sigma_\varepsilon^2 [E(f_i f_i')]^{-1}$ .*

Then,

$$\frac{c'(\hat{\delta}_L - \delta_0)}{\sqrt{c' \hat{\Phi} c}} \xrightarrow{d} \mathcal{N}(0, 1)$$

as  $n$  and  $\mu_n^2 \alpha$  go to infinity.

This result allows us to form confidence intervals and test statistics for a single linear combination of parameters in the usual way.

## 4 Efficiency and Related Literature

### 4.1 Efficiency

If the optimal instrument  $\gamma_i$  were known, the estimator would be solution of

$$\frac{1}{n} \sum_{i=1}^n \gamma_i (y_i - W_i' \delta) = 0.$$

Hence,

$$\begin{aligned} \hat{\delta} &= \left( \sum_{i=1}^n \gamma_i W_i' \right)^{-1} \sum_{i=1}^n \gamma_i y_i, \\ \hat{\delta} - \delta_0 &= \left( \sum_{i=1}^n \gamma_i W_i' \right)^{-1} \sum_{i=1}^n \gamma_i \varepsilon_i \\ &= \left( S_n \frac{\sum_{i=1}^n f_i f_i'}{n} S_n' + S_n \frac{\sum_{i=1}^n f_i u_i}{\sqrt{n}} \right)^{-1} S_n \frac{\sum_{i=1}^n f_i \varepsilon_i}{\sqrt{n}}, \end{aligned}$$

$$S_n \left( \hat{\delta} - \delta_0 \right) = \left( \frac{\sum_{i=1}^n f_i f_i'}{n} + \frac{\sum_{i=1}^n f_i u_i}{\sqrt{n}} S_n'^{-1} \right)^{-1} \frac{\sum_{i=1}^n f_i \varepsilon_i}{\sqrt{n}} \\ \xrightarrow{d} \mathcal{N} \left( 0, \sigma_\varepsilon^2 [E(f_i f_i')]^{-1} \right).$$

So the lowest asymptotic variance that can be obtained is  $\sigma_\varepsilon^2 [E(f_i f_i')]^{-1}$ . We will refer to this as the semiparametric efficiency bound.

Now, if the  $\gamma_i$  is unknown but estimated by a consistent nonparametric estimator  $\hat{\gamma}_i$ . The resulting estimator

$$\hat{\delta} = \left( \sum_{i=1}^n \hat{\gamma}_i W_i' \right)^{-1} \sum_{i=1}^n \hat{\gamma}_i y_i,$$

will satisfy  $S_n \left( \hat{\delta} - \delta_0 \right) \xrightarrow{d} \mathcal{N} \left( 0, \sigma_\varepsilon^2 [E(f_i f_i')]^{-1} \right)$  regardless of the convergence rate of  $\hat{\gamma}_i$ . This can be proved using the same arguments as in Newey (1990).

In Carrasco (2012, Section 2.4), it was shown that the regularized 2SLS estimator coincides with a 2SLS estimator that uses a specific nonparametric estimator of  $\gamma_i$ . This explains why for the regularized 2SLS estimator, the conditions on  $\alpha$  are not related to  $\mu_n$ . On the contrary, LIML can not be written as such a 2SLS estimator. This may explain why in the case of LIML, the rate of convergence of  $\alpha$  depends on how weak the instruments are.

## 4.2 Related Literature

In the literature on many weak instruments the asymptotic behavior of estimator depends on the relation between the number of moment conditions and sample size. For the CUE, the number of moments  $L$  and the sample size  $n$  needs to satisfy  $L^2/n \rightarrow 0$  to have consistency. Asymptotic normality, requires  $L^3/n \rightarrow 0$ . Under homoskedasticity, Stock and Yogo (2005) require  $L^2/n \rightarrow 0$ . Hansen, Hausman, and Newey (2008) allowed  $L$  to grow at the same rate as  $n$ , but restricted  $L$  to grow slower than the square of the concentration parameter, for the consistency of LIML and FULL. Andrews and Stock (2006) require  $L^3/n \rightarrow 0$  when normality is not imposed.

Caner and Yildiz (2012) in a recent work consider a Continuous Updating Estimator (CUE) with many weak moments under nearly singular design. They show that the

nearly singular design affects the form of asymptotic covariance matrix of the estimator compared to that of Newey and Windmeijer (2009). Our work is also related to Hausman, Lewis, Menzel, and Newey (2011) who propose the Regularized CUE (RCUE) as a solution in presence of many weak instruments. The RCUE solves a modification of the first-order conditions for the CUE estimator and is shown to be asymptotically equivalent to CUE under many weak moments asymptotics.

Belloni, Chen, Chernozhukov, and Hansen (2012) propose to use an alternative regularization named lasso in the IV context. This regularization imposes a  $l_1$  type penalty on the first stage coefficient. Assuming that the first stage equation is approximately sparse, they show that the postlasso estimator reaches the asymptotic efficiency bound.

Just as 2SLS is not consistent if  $L$  is too large relative to  $n$ , LIML estimator is not feasible if  $L > N$  because the matrix  $Z'Z$  is not invertible. Therefore, some form of regularization needs to be implemented to obtain consistent estimators when the number of instruments is really large. Regularization has also the advantage to deliver an asymptotically efficient estimator.

Table 1 gives an overview of the assumptions used in the main papers on many weak instruments.

Table 1: Comparison of different IV asymptotics

	Number of instruments	Extra assumptions
Conventional	Fixed $L$	
Phillips (1989)	Fixed $L$ , $Cov(W, x) = 0$	
Staiger and Stock (1997)	Fixed $L$ , $Cov(W, x) = O(n^{-1/2})$	
Bekker (1994)	$L/n \rightarrow c < 1$ , $\mu_n^2 = O(n)$	
Han and Phillips (2006)	$L \rightarrow \infty$ and $\frac{L}{nc_n} \rightarrow c$ $c_n \mu_n$ constant or zero	
Chao and Swanson (2005)	$\frac{L}{\mu_n^2} \rightarrow 0$ or $\frac{L^{1/2}}{\mu_n^2} \rightarrow 0$	
Hansen et al. (2008)	(I) $\frac{L}{\mu_n^2}$ bounded or (II) $\frac{L}{\mu_n^2} \rightarrow \infty$	$\sum z_i z_i' / n$ nonsingular
Newey and Windmeijer (2009)	$L \rightarrow \infty$ and $\frac{L}{\mu_n^2}$ bounded	
Carrasco (2012)	No condition on $L$ , possibly continuum strong instruments	Compactness of covariance matrix
Belloni et al. (2012)	$\log(L) = o(n^{1/3})$ , strong instruments	Approximately sparse first stage equation



## 5 Monte Carlo study

We now carry out a Monte Carlo simulation for the simple linear IV model where the disturbances and instruments have a Gaussian distribution as in Newey and Windmeijer (2009). The parameters of this experiment are the correlation coefficient  $\rho$  between the structural and reduced form errors, the concentration parameter ( $CP$ ), and the number of instruments  $L$

The data generating process is given by:

$$y_i = W_i' \delta_0 + \varepsilon_i,$$

$$W_i = x_i' \pi + u_i,$$

$$\varepsilon_i = \rho u_i + \sqrt{1 - \rho^2} v_i,$$

$$u_i \sim \mathcal{N}(0, 1), v_i \sim \mathcal{N}(0, 1), x_i \sim \mathcal{N}(0, I_L)$$

$$\pi = \sqrt{\frac{CP}{Ln}} \iota_L$$

where  $\iota_L$  is an  $L$ -vector of ones. The sample size is  $n = 500$ ; the number of instruments  $L$  equals to 15 and 30; value of  $\rho = 0.5$ ;  $\delta_0 = 0.1$ ; values of  $CP$  equals to 4, 35 or 65.

The estimators we proposed in this paper depend on a regularization (smoothing) parameter  $\alpha$  that needs to be selected. In the simulations, we use a data-driven method for selecting  $\alpha$  based on an expansion of the MSE and proposed in Carrasco (2012) and Carrasco and Tchuente (2013). These selection criteria were derived assuming strong instruments and may not be valid in presence of weak instruments. Providing a robust to weak instruments selection procedure is beyond the scope of this paper.

We report the median bias (Med.bias), the median of the absolute deviation of the estimator from the true value (Med.abs), the difference between the 0.1 and 0.9 quantiles(dis) of the distribution of each estimator, and the coverage rate(Cov.) of a nominal 95% confidence interval for unfeasible instrumental variable regression(IV), regularized two-stage least squares (T2SLS (Tikhonov), L2SLS (Landweber Fridman), P2SLS (Principal component)), LIML and regularized LIML (TLIML (Tikhonov), LLIML (Landweber Fridman), PLIML(Principal component)) and Donald and Newey's (2001)

2SLS and LIML (D2SLS and DLIML). For confidence intervals we compute the coverage probabilities using the following estimate of asymptotic variance as in Donald and Newey (2001) and Carrasco (2012).

$$\hat{V}(\hat{\delta}) = \frac{(y - W\hat{\delta})'(y - W\hat{\delta})}{n} \left( \hat{W}'W^{-1} \right)^{-1} \hat{W}'\hat{W} \left( W'\hat{W} \right)^{-1}$$

where  $\hat{W} = P^\alpha W$ .

Table 2: Simulations results with  $C_p = 4, 8, 35$  and  $65$  ;  $n = 500$ ;  $L = 15$  and  $30$ ; 1000 replications.

			T2SLS	L2SLS	P2SLS	D2SLS	TLIML	LLIML	PLIML	DLIML	IV
L=15	C <sub>p</sub> =4	Med.bias	0.3918	0.3876	0.3567	0.4093	0.1443	0.1378	0.3497	0.3907	0.0212
		Med.abs	0.3925	0.3883	0.5334	0.5633	0.5649	0.5644	0.7021	0.5208	0.3476
		Disp	0.5590	0.5806	2.6328	2.4639	3.2294	2.9889	3.5783	1.8474	1.5407
		Cov	0.5040	0.5330	0.7750	0.7750	0.4500	0.4660	0.9010	0.7130	0.9520
	C <sub>p</sub> =8	Med.bias	0.3248	0.3189	0.3036	0.3385	0.0381	0.0383	0.2471	0.2856	0.0031
		Med.abs	0.3265	0.3205	0.3917	0.4080	0.3722	0.3698	0.6087	0.3864	0.2506
		Disp	0.5044	0.5388	1.5854	1.7167	1.9193	1.9012	3.1534	1.3395	1.0210
		Cov	0.5630	0.5900	0.7420	0.7380	0.5690	0.5790	0.9240	0.7210	0.9500
	C <sub>p</sub> =35	Med.bias	0.1421	0.1409	0.1584	0.1940	0.0003	0.0013	0.0669	0.0383	0.0012
		Med.abs	0.1519	0.1524	0.1938	0.2123	0.1337	0.1357	0.2734	0.1316	0.1214
		Disp	0.3426	0.3656	0.4974	0.4297	0.5348	0.5448	1.4268	0.4907	0.4551
		Cov	0.7680	0.7810	0.7770	0.7430	0.8580	0.8680	0.9340	0.8640	0.9540
	C <sub>p</sub> =65	Med.bias	0.0876	0.0865	0.0953	0.1198	-0.0010	0.0025	0.0295	0.0061	0.0009
		Med.abs	0.1012	0.1051	0.1222	0.1343	0.0927	0.0905	0.1533	0.0920	0.0881
		Disp	0.2796	0.2932	0.3359	0.3039	0.3482	0.3505	0.7478	0.3483	0.3240
		Cov	0.8350	0.8300	0.8340	0.8120	0.9020	0.8960	0.9330	0.8980	0.9550
L=30	C <sub>p</sub> =4	Med.bias	0.4493	0.4401	0.3622	0.4197	0.2147	0.1825	0.3792	0.4156	0.0280
		Med.abs	0.4493	0.4401	0.5737	0.6374	0.6517	0.6559	0.6629	0.5265	0.3822
		Disp	0.4037	0.4378	2.8445	3.2928	3.7483	3.6305	2.9905	1.9820	1.5888
		Cov	0.2010	0.2680	0.7870	0.7780	0.2850	0.3180	0.8860	0.7010	0.9600
	C <sub>p</sub> =8	Med.bias	0.4019	0.3893	0.3260	0.3802	0.0618	0.0800	0.2915	0.3591	-0.0096
		Med.abs	0.4019	0.3893	0.4665	0.5419	0.4413	0.4568	0.5940	0.4513	0.2624
		Disp	0.3857	0.4262	2.4101	2.6333	2.4090	2.2558	2.7095	1.4976	1.0184
		Cov	0.2470	0.3270	0.7530	0.7180	0.3790	0.4010	0.9000	0.6960	0.9540
	C <sub>p</sub> =35	Med.bias	0.2316	0.2174	0.2211	0.2681	0.0028	0.0078	0.1054	0.1243	-0.0046
		Med.abs	0.2316	0.2176	0.2617	0.2912	0.1530	0.1558	0.3991	0.1827	0.1203
		Disp	0.2938	0.3195	0.6125	0.6097	0.6115	0.6275	2.0311	0.6385	0.4363
		Cov	0.4810	0.5580	0.6870	0.6270	0.7190	0.7340	0.9500	0.7180	0.9580
	C <sub>p</sub> =65	Med.bias	0.1586	0.1466	0.1582	0.2026	0.0038	0.0037	0.0558	0.0335	-0.0032
		Med.abs	0.1586	0.1475	0.1868	0.2186	0.1023	0.1027	0.2440	0.1028	0.0883
		Disp	0.2545	0.2710	0.4219	0.4517	0.3918	0.3989	1.2399	0.3962	0.3156
		Cov	0.6120	0.6610	0.7370	0.6750	0.8330	0.8320	0.9410	0.8270	0.9590

Table 2 reports simulation results. We use different strength (measured by the concentration parameter) of instruments and number of instruments. We investigate the case of very weak instruments for example, when  $CP = 8$  and  $L = 30$ , the first stage F-statistic equals  $\frac{CP}{L} = 0.26$ .

We observe that

(a) The performances of the regularized estimators increase with the strength of instruments but decrease with the number of instruments. Providing regularization parameter selection procedure robust to weak instruments would certainly improve these results.

(b) The bias of regularized LIML is quite a bit smaller than that of regularized 2SLS.

(c) The bias of our regularized estimators are smaller than those of the corresponding Donald and Newey's estimators.

(d) LF LIML and T-LIML estimators have very low median bias even in the case of relatively weak instruments (CP=8).

(e) The coverage of our estimators deteriorates when the instruments are weak.

## 6 Conclusion

This paper illustrates the usefulness of regularization techniques for estimation in many weak instruments framework. We derived the properties of the regularized 2SLS and LIML estimators in the presence of many or a continuum of moments that may be weak. We show that if well normalized the regularized 2SLS and LIML are consistent and reach the semiparametric efficiency bound. Our simulations show that the leading regularized estimators (LF and T of LIML) perform well.

In this work we restricted our investigation to 2SLS and LIML with weak instruments. It would be interesting, for future research to study the behavior of regularized version of other k-class estimators, such as FULL or B2SLS or other estimators as GMM or GEL, in presence of many weak instruments. This will help us to have results that can be compared with those of Newey and Windmeijer (2009) and Hansen, Hausman, and Newey (2008). Another topic of interest is the use of our regularization tools to provide version of robust tests for weak instruments as AR, LM, CLR or KLM tests, that can be used with a large number or a continuum of moment conditions.

## References

- ANDREWS, D. W., AND J. H. STOCK (2006): “Inference with Weak Instruments,” in *Advances in Economics and Econometrics*, ed. by R. Blundell, W. Newey, and T. Persson, vol. 3. Cambridge University Press.
- ANGRIST, J. D., AND A. B. KRUEGER (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?,” *The Quarterly Journal of Economics*, 106(4), 979–1014.
- ANTOINE, B., AND P. LAVERGNE (2012): “Conditional Moment Models under Semi-Strong Identification,” Discussion Papers dp11-04, Department of Economics, Simon Fraser University.
- BAI, J., AND S. NG (2002): “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70(1), 191–221.
- BEKKER, P. A. (1994): “Alternative Approximations to the Distributions of Instrumental Variable Estimators,” *Econometrica*, 62(3), 657–81.
- BELLONI, A., D. CHEN, V. CHERNOZHUKOV, AND C. HANSEN (2012): “Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain,” *Econometrica*, 80(6), 2369–2429.
- CANER, M., AND N. YILDIZ (2012): “CUE with many weak instruments and nearly singular design,” *Journal of Econometrics*, 170(2), 422–441.
- CARRASCO, M. (2012): “A regularization approach to the many instruments problem,” *Journal of Econometrics*, 170(2), 383–398.
- CARRASCO, M., AND J.-P. FLORENS (2000): “Generalization Of Gmm To A Continuum Of Moment Conditions,” *Econometric Theory*, 16(06), 797–834.
- (2012): “On the Asymptotic Efficiency of GMM,” *forthcoming in Econometric Theory*.

- CARRASCO, M., AND G. TCHUENTE (2013): “Regularized LIML for many instruments,” Discussion paper, University of Montreal.
- CHAO, J. C., AND N. R. SWANSON (2005): “Consistent Estimation with a Large Number of Weak Instruments,” *Econometrica*, 73(5), 1673–1692.
- (2007): “Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction,” *Journal of Econometrics*, 137(2), 515–555.
- EICHENBAUM, M. S., L. P. HANSEN, AND K. J. SINGLETON (1988): “A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty,” *The Quarterly Journal of Economics*, 103(1), 51–78.
- GUGGENBERGER, P., AND R. J. SMITH (2005): “Generalized Empirical Likelihood Estimators And Tests Under Partial, Weak, And Strong Identification,” *Econometric Theory*, 21(04), 667–709.
- HAHN, J., AND J. HAUSMAN (2003): “Weak Instruments: Diagnosis and Cures in Empirical Econometrics,” *American Economic Review*, 93(2), 118–125.
- HAN, C., AND P. C. B. PHILLIPS (2006): “GMM with Many Moment Conditions,” *Econometrica*, 74(1), 147–192.
- HANSEN, C., J. HAUSMAN, AND W. NEWEY (2008): “Estimation With Many Instrumental Variables,” *Journal of Business & Economic Statistics*, 26, 398–422.
- HAUSMAN, J., R. LEWIS, K. MENZEL, AND W. NEWEY (2011): “Properties of the CUE estimator and a modification with moments,” *Journal of Econometrics*, 165(1), 45 – 57, Moment Restriction-Based Econometric Methods.
- HAUSMAN, J. A., W. K. NEWEY, T. WOUTERSEN, J. C. CHAO, AND N. R. SWANSON (2012): “Instrumental variable estimation with heteroskedasticity and many instruments,” *Quantitative Economics*, 3(2), 211–255.

- 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(371), 693–700.
- MORIMUNE, K. (1983): “Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability Is Large Compared with the Sample Size,” *Econometrica*, 51(3), 821–41.
- NEWKEY, W. K. (1990): “Efficient Estimation of Models with Conditional Moment Restrictions,” in *Handbook of Statistics*, ed. by G. Maddala, C. Rao, and H. Vinod, vol. 11, pp. 419–454. Elsevier.
- NEWKEY, W. K., AND F. WINDMEIJER (2009): “Generalized Method of Moments With Many Weak Moment Conditions,” *Econometrica*, 77(3), 687–719.
- STAIGER, D., AND J. H. STOCK (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65(3), 557–586.
- STOCK, J. H., AND M. W. WATSON (2002): “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97(460), pp. 1167–1179.
- STOCK, J. H., AND M. YOGO (2005): “Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments,” in *Identification and Inference for Econometric Models*, ed. by D. W. K. Andrews, and J. H. Stock, pp. 109–120. Cambridge University Press.
- ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): “Valid Confidence Intervals and Inference in the Presence of Weak Instruments,” *International Economic Review*, 39(4), 1119–46.

## A Proofs

### Proof of Proposition 1:

We first prove the consistency of our estimator.

Let  $g_n = \frac{1}{n} \sum_{i=1}^n Z_i W_i = S_n \left[ \frac{1}{n} \sum_{i=1}^n Z_i f_i \right] / \sqrt{n} + \frac{1}{n} \sum_{i=1}^n Z_i u_i = S_n g_{n1} / \sqrt{n} + g_{n2}$  (remember that  $g_n$  is a function indexed by  $\tau$  and  $Z_i$  is also a function of  $\tau$ , such a representation can handle both countable and continuum of instruments). Note<sup>3</sup> that  $g_{n2} = \frac{1}{n} \sum_{i=1}^n Z_i u_i = o_p(1)$ ,  $\sqrt{n} g_{n2} = O_p(1)$  and  $S_n / \sqrt{n}$  is bounded from assumption 1(i).

$$\hat{\delta} - \delta_0 = (W' P^\alpha W)^{-1} W' P^\alpha \varepsilon$$

We have  $S_n'(\hat{\delta} - \delta_0) / \mu_n = [S_n^{-1} W' P^\alpha W S_n^{-1}]^{-1} [S_n^{-1} W' P^\alpha \varepsilon / \mu_n]$  and by construction of  $P^\alpha$ :

$$\begin{aligned} W' P^\alpha W &= n \left\langle (K_n^\alpha)^{-1/2} g_n, (K_n^\alpha)^{-1/2} g_n' \right\rangle \\ &= S_n \left\langle (K_n^\alpha)^{-1/2} g_{n1}, (K_n^\alpha)^{-1/2} g_{n1}' \right\rangle S_n' \\ &\quad + S_n \left\langle (K_n^\alpha)^{-1/2} g_{n1}, (K_n^\alpha)^{-1/2} g_{n2}' \right\rangle \sqrt{n} \\ &\quad + \left\langle (K_n^\alpha)^{-1/2} g_{n2}, (K_n^\alpha)^{-1/2} g_{n1}' \right\rangle S_n' \sqrt{n} \\ &\quad + \left\langle (K_n^\alpha)^{-1/2} g_{n2}, (K_n^\alpha)^{-1/2} g_{n2}' \right\rangle n. \end{aligned}$$

$$\begin{aligned} S_n^{-1} W' P^\alpha W S_n^{-1'} &= \left\langle (K_n^\alpha)^{-1/2} g_{n1}, (K_n^\alpha)^{-1/2} g_{n1}' \right\rangle \\ &\quad + \left\langle (K_n^\alpha)^{-1/2} g_{n1}, (K_n^\alpha)^{-1/2} \sqrt{n} g_{n2}' \right\rangle S_n^{-1'} \\ &\quad + S_n^{-1} \left\langle (K_n^\alpha)^{-1/2} \sqrt{n} g_{n2}, (K_n^\alpha)^{-1/2} g_{n1}' \right\rangle \\ &\quad + S_n^{-1} \left\langle (K_n^\alpha)^{-1/2} \sqrt{n} g_{n2}, (K_n^\alpha)^{-1/2} \sqrt{n} g_{n2}' \right\rangle S_n^{-1'}. \end{aligned}$$

Hence,

$$S_n^{-1} [W' P^\alpha W] S_n^{-1'} = \left\langle (K_n^\alpha)^{-\frac{1}{2}} g_{n1}, (K_n^\alpha)^{-\frac{1}{2}} g_{n1}' \right\rangle + o_p(1).$$

---

<sup>3</sup>Let  $g$  and  $h$  be two  $p$  vectors of functions of  $L^2(\pi)$ . By a slight abuse of notation,  $\langle g, h' \rangle$ ; denotes the matrix with elements  $\langle g_a, h_b \rangle$   $a, b = 1, \dots, p$

At this stage, we can apply the same proof as that of Proposition 1 of Carrasco (2012) which shows that

$$\langle (K_n^\alpha)^{-\frac{1}{2}} g_{n1}, (K_n^\alpha)^{-\frac{1}{2}} g'_{n1} \rangle \rightarrow \langle g_1, g'_1 \rangle_K$$

in probability as  $n$  and  $n\alpha^{\frac{1}{2}}$  go to infinity, with  $\langle g_1, g'_1 \rangle_K$  a  $p \times p$  matrix with  $(a, b)$  element  $\langle K^{-\frac{1}{2}} E(Z(\cdot, x_i) f_{ia}), K^{-\frac{1}{2}} E(Z(\cdot, x_i) f_{ib}) \rangle$  which is assumed to be nonsingular.

$$\begin{aligned} \frac{S_n^{-1} W' P^\alpha \varepsilon}{\mu_n} &= \frac{n S_n^{-1}}{\mu_n} \left\langle (K_n^\alpha)^{-1/2} g_n, (K_n^\alpha)^{-1/2} \frac{1}{n} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \\ &= \frac{1}{\mu_n} \left\langle (K_n^\alpha)^{-1/2} g_{n1}, (K_n^\alpha)^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \\ &\quad + \frac{S_n^{-1}}{\mu_n} \left\langle (K_n^\alpha)^{-1/2} \sqrt{n} g_{n2}, (K_n^\alpha)^{-1/2} \frac{\sqrt{n}}{n} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \\ &= o_p(1). \end{aligned}$$

because  $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i = O_p(1)$ . This proves the consistency of the regularized 2SLS.

For the asymptotic normality we write

$$S'_n(\hat{\delta} - \delta_0) = [S_n^{-1} W' P^\alpha W S_n'^{-1}]^{-1} [S_n^{-1} W' P^\alpha \varepsilon]$$

We then have

$$\begin{aligned} S_n^{-1} W' P^\alpha \varepsilon &= n S_n^{-1} \left\langle (K_n^\alpha)^{-1} g_n, \frac{1}{n} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \\ &= \left\langle (K_n^\alpha)^{-1/2} g_{n1}, (K_n^\alpha)^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \\ &\quad + S_n^{-1} \left\langle (K_n^\alpha)^{-1/2} \sqrt{n} g_{n2}, (K_n^\alpha)^{-1/2} \frac{\sqrt{n}}{n} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \\ &= \left\langle (K_n^\alpha)^{-1/2} g_{n1}, (K_n^\alpha)^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \\ &\quad + o_p(1). \end{aligned}$$



Moreover,

$$\left\langle (K_n^\alpha)^{-1/2} g_{n1}, (K_n^\alpha)^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \quad (2)$$

$$= \left\langle (K_n^\alpha)^{-1} g_{n1} - K^{-1} g_1, \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \quad (3)$$

$$+ \left\langle K^{-1} g_1, \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle.$$

The first term is negligible since

$$\left\langle (K_n^\alpha)^{-1} g_{n1} - K^{-1} g_1, \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \leq \| (K_n^\alpha)^{-1} g_{n1} - K^{-1} g_1 \| \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\| = o_p(1) O_p(1).$$

By the functional central limit theorem, we obtain the following result

$$\left\langle K^{-1} g_1, \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \varepsilon_i \right\rangle \rightarrow \mathcal{N}(0, \sigma_\varepsilon^2 \langle g_1, g_1' \rangle_K)$$
 as  $n$  and  $n\alpha$ ,  $\mu_n$  go to infinity.

We then apply the continuous mapping theorem and Slutsky lemma to show that

$$S'_n(\hat{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2 \langle g_1, g_1' \rangle_K^{-1}).$$

By assumption,  $g_{1a} = E(Z(\cdot, x_i) f_{ia})$  belong to the range of  $K$ . Let  $L^2(Z)$  be the closure of the space spanned by  $\{Z(x, \tau), \tau \in I\}$  and  $g_1$  is an element of this space. If  $f_i \in L^2(Z)$  we can compute the inner product in the RKHS and show that

$$\langle g_{1a}, g_{1b} \rangle_K = E(f_{ia} f_{ib}).$$

This can be seen by applying Theorem 6.4 of Carrasco, Florens, and Renault (2007).

It follows that

$$S'_n(\hat{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\varepsilon^2 [E(f_i f_i')]^{-1}\right)$$

This completes the proof of Proposition 1.

### Proof of Proposition 2:

To prove this proposition we need some lemmas. The first lemma corresponds to lemma A0 of Hansen, Hausman, and Newey (2008).

**Lemma 1:** Under assumption 1 if  $\|S'_n(\hat{\delta}_L - \delta_0)/\mu_n\|^2 / (1 + \|\hat{\delta}_L\|^2) \rightarrow 0$  then  $\|S'_n(\hat{\delta}_L - \delta_0)/\mu_n\| \rightarrow 0$ .

**proof:** The proof of this lemma is the same as in Hansen, Hausman, and Newey (2008).

**Lemma 2:** Let us assume that there exists a constant  $C$  such that  $E(\|\varepsilon_i\|^4|X) \leq C$  and  $E(\|u_{ai}\|^4|X) \leq C$  for all  $i$ . Then,

$$\text{Var}(\varepsilon' P^\alpha u_a) \leq C \left( \sum_j q_j^2 \right),$$

$$\begin{aligned} \varepsilon' P^\alpha u_a - E(\varepsilon' P^\alpha u_a | X) &= O\left(\left(\sum_j q_j^2\right)^{\frac{1}{2}}\right), \\ \frac{\varepsilon' P^\alpha \varepsilon}{\mu_n^2} &= O_p\left(\frac{1}{\alpha \mu_n^2}\right) = o_p(1). \end{aligned}$$

**proof:**

For notational simplicity we suppress the conditioning on  $X$ . Let  $E(\varepsilon_i^2) = \sigma_\varepsilon^2$ ,  $E(\varepsilon_i u_{ai}) = \sigma_{\varepsilon u_a}$  and  $E(u'_{ai} u_{ai}) = \sigma_{u_a}^2$ ,

$$\text{Var}(\varepsilon' P^\alpha u_a) = E(\varepsilon' P^\alpha u_a u'_a P^\alpha \varepsilon) - E(\varepsilon' P^\alpha u_a) E(u'_a P^\alpha \varepsilon).$$

Using the spectral decomposition of  $P^\alpha$ , we have

$$\begin{aligned} E(\varepsilon' P^\alpha u_a u'_a P^\alpha \varepsilon) &= \frac{1}{n^2} \sum_{j,l} q_j q_l E\left\{(\varepsilon' \psi_l)(u'_a \psi_l)'(\varepsilon' \psi_j)(u'_a \psi_j)\right\} \\ &= \frac{1}{n^2} \sum_{j,l} q_j q_l E\left\{\sum_i \varepsilon_i u'_{ai} \psi_{li}^2 \sum_b \varepsilon_b u_{ab} \psi_{jb}^2\right. \\ &\quad + \sum_c \varepsilon_c u'_{ac} \psi_{lc} \psi_{jc} \sum_d \varepsilon_d u_{ad} \psi_{jd} \psi_{ld} \\ &\quad \left. + \sum_c \varepsilon_c^2 \psi_{lc} \psi_{jc} \sum_d u'_{ad} u_{ad} \psi_{jd} \psi_{ld}\right\} \\ &= \left(\sum_j q_j\right)^2 \sigma'_{\varepsilon u_a} \sigma_{\varepsilon u_a} + (\sigma'_{\varepsilon u_a} \sigma_{\varepsilon u_a} + \sigma_\varepsilon^2 \sigma_{u_a}^2) \sum_j q_j^2 \end{aligned}$$

by the fact that  $(u_{ai}, \varepsilon_i)$  are independent across  $i$  and the eigenvectors are orthonormal.

$$\begin{aligned} E(\varepsilon' P^\alpha u_a) &= \frac{1}{n} \sum_l q_l E\left\{\left(\sum_k u'_{ak} \psi_{lk}\right) \left(\sum_i \varepsilon_i \psi_{li}\right)\right\} \\ &= \frac{1}{n} \sum_l q_l n \sigma'_{\varepsilon u_a} \\ &= \sigma'_{\varepsilon u_a} \left(\sum_j q_j\right). \end{aligned}$$

Thus

$$\text{Var}(\varepsilon' P^\alpha u_a) = (\sigma'_{\varepsilon u_a} \sigma_{\varepsilon u_a} + \sigma_\varepsilon^2 \sigma_{u_a}^2) \sum_j q_j^2 \leq C \left( \sum_j q_j^2 \right).$$

The second conclusion follows by Markov inequality.

$$\begin{aligned} E(\varepsilon' P^\alpha \varepsilon) &= \text{tr}(P^\alpha E(\varepsilon \varepsilon')) \\ &= \sigma_\varepsilon^2 \left( \sum_j q_j \right) = O_p(1/\alpha). \end{aligned}$$

Using the result for  $\varepsilon' P^\alpha u_a$  with  $\varepsilon$  in place of  $u_a$ , we obtain

$$\text{Var}(\varepsilon' P^\alpha \varepsilon) \leq C \left( \sum_j q_j^2 \right).$$

It follows that  $(\varepsilon' P^\alpha \varepsilon - E(\varepsilon' P^\alpha \varepsilon)) / \mu_n^2 = O_p \left( \left( \sum_j q_j^2 \right)^{1/2} / \mu_n^2 \right) = o_p \left( \sum_j q_j / \mu_n^2 \right)$ .

Hence, the third equality holds.

**Lemma 3:** Let  $\hat{A} = \frac{f' P^\alpha f}{n}$  and  $\hat{B} = \frac{\bar{W}' \bar{W}}{n}$  with  $\bar{W} = [y, W]$ , there exist two constants  $C$  and  $C'$  such that  $\hat{A} \geq C I_p$  and  $\|\hat{B}\| \leq C'$ .

**Proof:** By the definition of  $P^\alpha$ , we have (see Equation (1)):

$$\hat{A} = \frac{f' P^\alpha f}{n} = \langle (K_n^\alpha)^{-\frac{1}{2}} f_n, (K_n^\alpha)^{-\frac{1}{2}} f_n \rangle$$

with

$$f_n = \frac{1}{n} \sum_i Z_i f_i.$$

By Lemma 5(i) of Carrasco (2012) and the law of large numbers,

$$\frac{f' P^\alpha f}{n} = \frac{f' f}{n} + o_p(1) = E(f'_i f_i) + o_p(1)$$

as  $\alpha$  goes to zero. Because  $E(f'_i f_i)$  is positive definite, there exists a constant  $C$  such that

$$\hat{A} \geq C I_p$$

with probability one.

We have  $\bar{W} = [y, W] = WD_0 + \varepsilon e$  where  $D_0 = [\delta_0, I]$ ,  $\delta_0$  is the true value of the parameter and  $e$  is the first unit vector.

$$\begin{aligned}\hat{B} &= \frac{\bar{W}'\bar{W}}{n} \\ &= D_0'S_n \frac{f'f}{n} S_n' D_0/n + D_0'S_n \frac{f'u}{n} D_0/\sqrt{n} + D_0'S_n \frac{f'\varepsilon}{n} e/\sqrt{n} \\ &+ D_0' \frac{u'f}{n} S_n' D_0/\sqrt{n} + D_0' \frac{u'u}{n} D_0 + D_0' \frac{u'\varepsilon}{n} e \\ &+ e' \frac{\varepsilon'f}{n} S_n' D_0/\sqrt{n} + e' \frac{\varepsilon'u}{n} D_0 + e' \frac{\varepsilon'\varepsilon}{n} e\end{aligned}$$

Using the law of large numbers, we can conclude that  $\|\hat{B}\| \leq C'$ , where  $C'$  is a constant, with probability one.

### Proof of consistency

Let us consider

$$\begin{aligned}\hat{Q}(\delta) &= \frac{(y - W\delta)' P^\alpha (y - W\delta) / \mu_n^2}{(y - W\delta)' (y - W\delta) / n} \\ \hat{\delta}_L &= \operatorname{argmin} Q(\delta)\end{aligned}$$

For  $\delta = \delta_0$ ,  $\hat{Q}(\delta_0) = \frac{\varepsilon' P^\alpha \varepsilon / \mu_n^2}{\varepsilon' \varepsilon / n}$ . With probability one  $\varepsilon' \varepsilon / n > C$  and by lemma 2

$$\varepsilon' P^\alpha \varepsilon / \mu_n^2 = o_p(1).$$

Hence  $\hat{Q}(\delta_0) = o_p(1)$ .

Since  $0 \leq \hat{Q}(\hat{\delta}_L) \leq \hat{Q}(\delta_0)$  it is easy to see that  $\hat{Q}(\hat{\delta}_L) = o_p(1)$ .

Let us show that

$$\mu_n^{-2} (y - W\delta)' P^\alpha (y - W\delta) \geq C \|S_n'(\hat{\delta}_L - \delta_0) / \mu_n\|^2.$$

Let  $D(\delta) = \mu_n^{-2} (y - W\delta)' P^\alpha (y - W\delta)$ ,  $D = \mu_n^{-2} (1, -\delta') \bar{W}' P^\alpha \bar{W} (1, -\delta)'$  and it is possible to proof that  $D(\delta) = \mu_n^{-2} (1, -\delta') D_0' S_n \frac{f' P^\alpha f}{n} S_n' D_0 (1, -\delta)' + o_p(1)$ . It follows from lemma 3 that

$$D(\hat{\delta}_L) \geq C \|S_n'(\hat{\delta}_L - \delta_0) / \mu_n\|^2.$$

We also have that  $(y - W\delta)'(y - W\delta)/n = (1, -\delta')\hat{B}(1, -\delta')^2$ . Hence,

$$\frac{\|S'_n(\hat{\delta}_L - \delta_0)/\mu_n\|^2}{(1 + \|\hat{\delta}_L\|^2)} \leq C\hat{Q}(\hat{\delta}_L)$$

Then by Lemma 1 we have  $S'_n(\hat{\delta}_L - \delta_0)/\mu_n \rightarrow 0$  in probability as  $n$  and  $\mu_n^2\alpha$  go to infinity. This proves the consistency of LIML with many weak instruments.

Now let us prove the asymptotic normality.

**Proof of asymptotic normality:**

Denote  $A(\delta) = (y - W\delta)'P^\alpha(y - W\delta)/2$ ,  $B(\delta) = (y - W\delta)'(y - W\delta)$  and

$$\Lambda(\delta) = \frac{A(\delta)}{B(\delta)}$$

We know that the LIML is  $\hat{\delta}_L = \operatorname{argmin}\Lambda(\delta)$ .

We calculate the gradient and Hessian  $\Lambda_\delta(\delta) = B(\delta)^{-1}[A_\delta(\delta) - \Lambda(\delta)B_\delta(\delta)]$

$\Lambda_{\delta\delta}(\delta) = B(\delta)^{-1}[A_{\delta\delta}(\delta) - \Lambda(\delta)B_{\delta\delta}(\delta)] - B(\delta)^{-1}[B_\delta(\delta)\Lambda'_\delta(\delta) - \Lambda_\delta(\delta)B'_\delta(\delta)]$ .

Then by a standard mean-value expansion of the first-order conditions  $\Lambda_\delta(\tilde{\delta}) = 0$ , we have with probability one.

$$S'_n(\hat{\delta}_L - \delta_0) = -[S_n^{-1}\Lambda_{\delta\delta}(\tilde{\delta})S_n^{-1}]^{-1}[S'_n\Lambda_\delta(\delta_0)]$$

where  $\tilde{\delta}$  is the mean-value. By the consistency of  $\hat{\delta}_L$ ,  $\tilde{\delta} \rightarrow \delta_0$ .

It then follows that

$$\begin{aligned} B_\delta(\tilde{\delta})/n &= -2 \sum_i W_i \tilde{\varepsilon}_i / n, \\ &= -2 \sum_i (\gamma_i + u_i) \tilde{\varepsilon}_i / n \\ &= -2S_n / \sqrt{n} (\sum_i f_i \tilde{\varepsilon}_i / n) - 2(\sum_i u_i \tilde{\varepsilon}_i / n) \\ &= -2\sigma_{u\varepsilon} + o_p(1) \end{aligned}$$

under the assumption that  $S_n/\sqrt{n}$  is bounded, with  $\tilde{\varepsilon}_i = (y_i - W'_i\tilde{\delta})$  and  $\sigma_{u\varepsilon} = E(u_i\varepsilon_i)$ .

$$B(\tilde{\delta})/n \xrightarrow{P} \sigma_\varepsilon^2, \quad B_\delta(\tilde{\delta})/n \xrightarrow{P} -2\sigma_{u\varepsilon}$$

$$\Lambda(\delta) = \frac{(y - W\delta)'P^\alpha(y - W\delta)/2n}{(y - W\delta)'(y - W\delta)/n}$$

For  $\delta = \delta_0$ ,  $\Lambda(\delta_0) = \frac{\varepsilon'P^\alpha\varepsilon/2n}{\varepsilon'\varepsilon/n}$ . With probability one  $\varepsilon'\varepsilon/n > C$  and with lemma 2 and  $\mu_n^2 \leq n$ ,

$$\varepsilon'P^\alpha\varepsilon/n = o_p(1).$$

We have  $\Lambda(\delta_0) = o_p(1)$ . Therefore,  $\Lambda(\tilde{\delta}) \xrightarrow{P} 0$ . By the first order condition, we also have

$$\Lambda_\delta(\tilde{\delta}) \xrightarrow{P} 0.$$

$$B_{\delta\delta}(\tilde{\delta}) = 2W'W/n \xrightarrow{P} 2E(W_iW_i'), \quad A_{\delta\delta}(\tilde{\delta})/n = W'^\alpha W/n.$$

We can then conclude that  $\Lambda_{\delta\delta}(\tilde{\delta}) = nB^{-1}(\tilde{\delta})[A_{\delta\delta}(\tilde{\delta})/n] + o_p(1)$ . Hence

$$\begin{aligned} n\tilde{\sigma}_\varepsilon^2\Lambda_{\delta\delta}(\tilde{\delta}) &= W'P^\alpha W \\ &= S_n \langle (K_n^\alpha)^{-\frac{1}{2}}g_{n1}, (K_n^\alpha)^{-\frac{1}{2}}g'_{n1} \rangle S_n' + o_p(1) \\ &= S_n H S_n' + o_p(1) \end{aligned}$$

with  $H = E(f(x_i)f(x_i)')$  and  $\tilde{\sigma}_\varepsilon^2 = (y - W\tilde{\delta})'(y - W\tilde{\delta})/n$ .

Hence

$$n\tilde{\sigma}_\varepsilon^2 S_n^{-1} \Lambda_{\delta\delta}(\tilde{\delta}) S_n^{-1'} = H + o_p(1)$$

Let  $\hat{\phi} = \frac{W'\varepsilon}{\varepsilon'\varepsilon}$ ,  $\phi = \frac{\sigma_{u\varepsilon}}{\sigma_\varepsilon^2}$  and  $v = u - \varepsilon\phi'$ . It is useful to remark that  $v'P^\alpha\varepsilon = O_p(1/\sqrt{\alpha})$  using Lemma 2 with  $v$  in place of  $u$  and  $E(u_i v_i) = 0$ . Moreover,  $\hat{\phi} - \phi = O_p(1/\sqrt{n})$  by the central limit theorem and the delta method. Hence,  $(\hat{\phi} - \phi)\varepsilon'P^\alpha\varepsilon = O_p(1/\alpha\sqrt{n})$ .

Furthermore,  $f'(I - P^\alpha)\varepsilon/\sqrt{n} = O_p(\Delta_\alpha^2) = o_p(1)$  by Lemma 5(ii) Carrasco (2012) with  $\Delta_\alpha = \text{tr}(f'(I - P^\alpha)^2 f/n) = O_p(\alpha^{\min(\beta, 2)}) = o_p(1)$ .

$$\begin{aligned} -n\tilde{\sigma}_\varepsilon^2 S_n^{-1} \Lambda_\delta(\delta_0) &= S_n^{-1} (W'P^\alpha\varepsilon - \varepsilon'P^\alpha\varepsilon \frac{W'\varepsilon}{\varepsilon'\varepsilon}) \\ &= f'\varepsilon/\sqrt{n} - f'(I - P^\alpha)\varepsilon/\sqrt{n} + S_n^{-1}v'P^\alpha\varepsilon - S_n^{-1}(\hat{\phi} - \phi)\varepsilon'P^\alpha\varepsilon \\ &= f'\varepsilon/\sqrt{n} + o_p(1) + S_n^{-1}O_p(1/\sqrt{\alpha}) + S_n^{-1}O_p(1/\alpha\sqrt{n}) \\ &= f'\varepsilon/\sqrt{n} + o_p(1) \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2 H) \end{aligned}$$

as  $n, \alpha\mu_n^2$  go to infinity under the assumption  $\mu_n S_n^{-1} \rightarrow S_0$ .

The conclusion follows from the use of Slutsky theorem.