Essays in dynamic panel data models and Labor supply

par
Ada Kolobadia Nayihouba

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

Août, 2019

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :
Essays in dynamic panel data models and Labor supply

présentée par :
Ada Kolobadia Nayihouba

a été évaluée par un jury composé des personnes suivantes :

Benoit Perron,	president du jury
Marine Carrasco,	directrice de recherche
Raphael Godefroy,	membre du jury
Ryo Okui,	examinateur externe
Manuel Morales,	représentant du doyen de la FAS

Thèse acceptée le 29 Août 2019

à mes frères et à mes parents !

Remerciements

Acknowledgments

Je remercie infiniment Marine Carrasco, ma directrice de recherche pour sa très grande disponibilité ainsi que ses conseils et encouragements tout au long de cette thèse. J'ai beaucoup appris de sa rigueur dans la recherche et je suis très heureux d'avoir travaillé avec elle sur un sujet de recherche aussi passionnant.
Je suis aussi reconnaissant à Benoit Perron et à Godefroy Raphael qui ont bien voulu accepter de faire partir de mon comité de thèse depuis ma deuxième année de thèse. Leurs conseils et aides multiformes m'ont été très précieux.
Mes remerciements vont également aux enseignants, chercheurs et au personel administratif du département des sciences économiques et du CIREQ. Cette thèse a été réalisée avec le soutien financier du département et du CIREQ.
Je tiens aussi à remercier mes amis et collègues avec qui j'ai partagé cette étape importante de mon parcours académique. Merci à Soumouaga, Téwou, Karim, David, Johnatan, Louphou, Marius, Hervé, Alex, Désiré, Doukhali, William, Vincent ainsi qu'à tous les autres collègues du CIREQ et du département de sciences économiques.
Je remercie également toutes les personnes dont j'ai fait la connaissance dans le cadre de mon séjour à Montréal et qui m'ont beaucoup soutenu et aidé.
Enfin j'ai une pensée pour mes parents et mes frères pour le soutien et les encouragements tout au long de mes études.

Résumé

Cette thèse est organisée en trois chapitres. Les deux premiers proposent une approche régularisée pour l'estimation du modèle de données de panel dynamique : l'estimateur GMM et l'estimateur LIML. Le dernier chapitre de la thèse est une application de la méthode de régularisation à l'estimation des élasticités de l'offre de travail en utilisant des modèles de pseudo-données de panel.

Dans un modèle de panel dynamique, le nombre de conditions de moments augmente rapidement avec la dimension temporelle du panel conduisant à une matrice de covariance des instruments de grande dimension. L'inversion d'une telle matrice pour calculer l'estimateur affecte négativement les propriétés de l'estimateur en échantillon fini. Comme solution à ce problème, nous proposons une approche par la régularisation qui consiste à utiliser une inverse généralisée de la matrice de covariance au lieu de son inverse classique. Trois techniques de régularisation sont utilisées : celle des composantes principales, celle de Tikhonov qui est basée sur le Ridge régression (aussi appelée Bayesian shrinkage) et enfin celle de Landweber Fridman qui est une méthode itérative. Toutes ces techniques introduisent un paramètre de régularisation qui est similaire au paramètre de lissage dans les régressions non paramétriques. Les propriétés en ećhantillon fini de l'estimateur régularisé dépend de ce paramètre qui doit être sélectionné parmis plusieurs valeurs potentielles.

Dans le premier chapitre (co-écrit avec Marine Carrasco), nous proposons l'estimateur GMM régularisé du modèle de panel dynamique. Sous l'hypothèse que le nombre d'individus et de périodes du panel tendent vers l'infini, nous montrons que nos estimateurs sont convergents and assymtotiquement normaux. Nous dérivons une méthode empirique de sélection du paramètre de régularisation basée sur une expansion de second ordre du l'erreur quadratique moyenne et nous démontrons l'optimalité de cette procédure de sélection. Les simulations montrent que la régularisation améliore les propriétés de l 'estimateur GMM classique. Comme application empirique, nous avons analysé l'effet du développement financier sur la croissance économique.

Dans le deuxième chapitre (co-écrit avec Marine Carrasco), nous nous intéressons à l'estimateur LIML régularisé du modèle de données de panel dynamique. L'estimateur LIML est connu pour avoir de meilleures propriétés en échantillon fini que l'estimateur GMM mais son utilisation devient problématique lorsque la dimension temporelle du panel devient large. Nous dérivons les propriétes assymtotiques de l'estimateur LIML régularisé sous l'hypothèse que le nombre d'individus et de périodes du panel tendent vers l'infini. Une procédure empirique de sélection du paramètre de régularisation est aussi proposée. Les bonnes performances de l'estimateur régularisé par rapport au LIML classique (non régularisé), au GMM classique ainsi que le GMM régularisé sont confirmées par des simulations.

Dans le dernier chapitre, je considère l'estimation des élasticités d'offre de travail des hommes canadiens. L'hétérogéneité inobservée ainsi que les erreurs de mesures sur les salaires et les revenus sont connues pour engendrer de l'endogéneité quand on estime les modèles d'offre de travail. Une solution fréquente à ce problème d'endogéneité consiste à régrouper les données sur la base des carastéristiques observables et d' éffectuer les moindres carrées pondérées sur les moyennes des goupes. Il a été démontré que cet estimateur est équivalent à l'estimateur des variables instrumentales sur les données individuelles avec les indicatrices de groupe comme instruments. Donc, en présence d'un grand nombre de groupe, cet estimateur souffre de biais en échantillon fini similaire à celui de l'estimateur des variables instrumentales quand le nombre d'instruments est élevé. Profitant de cette correspondance entre l'estimateur sur les données groupées et l'estimateur des variables instrumentales sur les données individuelles, nous proposons une approche régularisée à l'estimation du modèle. Cette approche conduit à des élasticités substantiellement différentes de ceux qu'on obtient en utilisant l'estimateur sur données groupées.
Mots-clés : modèle de données de panel dynamique, beaucoup de conditions de moments, méthode de régularisation, erreur quadratiqe moyenne, Méthodes des moments généralisés, maximun de vraisemblance à information limité, élasticités d'offre de travail, estimateur sur données groupées.

Abstract

This thesis is organized in three chapters. The first two chapters propose a regularization approach to the estimation of two estimators of the dynamic panel data model : the Generalized Method of Moment (GMM) estimator and the Limited Information Maximum Likelihood (LIML) estimator. The last chapter of the thesis is an application of regularization to the estimation of labor supply elasticities using pseudo panel data models.

In a dynamic panel data model, the number of moment conditions increases rapidly with the time dimension, resulting in a large dimensional covariance matrix of the instruments. Inverting this large dimensional matrix to compute the estimator leads to poor finite sample properties. To address this issue, we propose a regularization approach to the estimation of such models where a generalized inverse of the covariance matrix of the intruments is used instead of its usual inverse. Three regularization schemes are used : Principal components, Tikhonov which is based on Ridge regression (also called Bayesian shrinkage) and finally Landweber Fridman which is an iterative method. All these methods involve a regularization parameter which is similar to the smoothing parameter in nonparametric regressions. The finite sample properties of the regularized estimator depends on this parameter which needs to be selected between many potential values.

In the first chapter (co-authored with Marine Carrasco), we propose the regularized GMM estimator of the dynamic panel data models. Under double asymptotics, we show that our regularized estimators are consistent and asymptotically normal provided that the regularization parameter goes to zero slower than the sample size goes to infinity. We derive a data driven selection of the regularization parameter based on an approximation of the higher-order Mean Square Error and show its optimality. The simulations

confirm that regularization improves the properties of the usual GMM estimator. As empirical application, we investigate the effect of financial development on economic growth.

In the second chapter (co-authored with Marine Carrasco), we propose the regularized LIML estimator of the dynamic panel data model. The LIML estimator is known to have better small sample properties than the GMM estimator but its implementation becomes problematic when the time dimension of the panel becomes large. We derive the asymptotic properties of the regularized LIML under double asymptotics. A data-driven procedure to select the parameter of regularization is proposed. The good performances of the regularized LIML estimator over the usual (not regularized) LIML estimator, the usual GMM estimator and the regularized GMM estimator are confirmed by the simulations.

In the last chapter, I consider the estimation of the labor supply elasticities of Canadian men through a regularization approach. Unobserved heterogeneity and measurement errors on wage and income variables are known to cause endogeneity issues in the estimation of labor supply models. A popular solution to the endogeneity issue is to group data in categories based on observable characteristics and compute the weighted least squares at the group level. This grouping estimator has been proved to be equivalent to instrumental variables (IV) estimator on the individual level data using group dummies as intruments. Hence, in presence of large number of groups, the grouping estimator exhibites a small bias similar to the one of the IV estimator in presence of many instruments. I take advantage of the correspondance between grouping estimators and the IV estimator to propose a regularization approach to the estimation of the model. Using this approach leads to wage elasticities that are substantially different from those obtained through grouping estimators.
Keywords : Dynamic panel data model, many moment conditions, mean square error, regularization methods, GMM, LIML, labor supply elasticities, grouping estimators.

Table des matières

Liste des tableaux

Chapitre 1

A regularization approach to the dynamic panel data model estimation *

In this paper, we propose a regularization approach to the estimation of a dynamic panel data model (DPM) with individual fixed effect. The presence of this last element creates a correlation between the error term of the model and one of the explanatory variable which is the lagged value of the dependent variable. Hence, Generalized Method of Moments (GMM) are widely used to estimate such models with lagged levels dependent variable as instruments. A feature of the DPM is that, if a variable at a certain time period can be used as an instrument, then all the past realizations of that variable can also be used as instruments. Therefore, the number of moment conditions can be very large even if the time dimension is moderately large. Although using many instruments increases asymptotic efficiency of GMM estimator, it has been proved that its finite sample bias also increases with the number of instruments. Therefore, estimation in the presence of many moment conditions involves a variance-bias trade-off also referred to as the many instruments problem. As a solution to the many instruments problem, Carrasco (2012) used regularization to invert the covariance matrix of instruments. In this methodology, the bias is controlled by the choice of a regu-

∗. This chapter is co-authored with my advisor Marine Carrasco (Université de Montréal).

larization parameter and does no longer depend on the number of moment conditions which can then be increased (even infinitely) to improve the efficiency. This paper proposes regularization as a solution to the many instruments problem in the dynamic panel data model.
As in Carrasco (2012) and Carrasco and Tchuente (2015) on cross-sectional data, we compute three regularized estimators based on spectral cut-off, Tikhonov and Landweber Fridman. The spectral cut-off regularization scheme is based on principal components whereas the Tikhonov's one is based on Ridge regression (also called Bayesian shrinkage) and the last one is an iterative method. Our modified estimator using spectral cut-off regularization scheme is similar to the bias correction estimator using principal components proposed by Doran and Schmidt (2006). Our work complements their paper by proposing a data driven method to choose the optimal number of principal components to use in order to improve the finite sample properties of the estimator. The Tikhonov regularization scheme we propose can be considered as the dynamic panel version of the ridge regression. All these methods involve a regularization parameter similar to the smoothing parameter in nonparametric regression. This parameter needs to converge to zero at an appropriate rate to obtain an asymptotically efficient estimator.
We derive the first order asymptotic properties of the modified estimator under double asymptotics following Alvarez and Arellano (2003). Then, we derive the leading term of the MSE in a second order expansion of the regularized estimators when N and T go to infinity. This allows us to propose a data-driven selection of the regularization parameter as minimum of the approximate MSE. We then prove the optimality of the selection in the sense of Li (1986, 1987).

The literature related on many instruments problem is very large. Working on cross sectional models, Donald and Newey (2001) propose to select the number of instruments that minimizes the Mean Square Error (MSE) of the estimator. Okui (2011) introduces a shrinkage parameter to allocate less weight on a subset of instruments. Kuersteiner (2012) proposes a kernel weighted GMM estimator in a time series framework.
A regularization approach to handle many instruments for two stage least-squares estimation is proposed by Carrasco (2012) whereas Carrasco and Tchuente (2015) proposed the regularized version of the Limited Information Maximum Likelihood estimator (LIML). However, even under conditional homoskedasticity assumption, a correlation arises in the DPM framework in the equation linking the endogenous regressor and the optimal instrument

so that results of Carrasco (2012) no longer apply. Moreover, in the dynamic panel data setting, the number of instruments is automatically related to the sample size through the time dimension T.

Several bias corrected estimator have been proposed for DPM (Hahn et al., 2001, Bun and Kiviet, 2006, Alvarez and Arellano, 2003, Kiviet (1995), Hahn and Kuersteiner (2002)). Our methodology complements those methods as regularization provides a partial bias-correction which can be improved by bias-correcting the regularized estimators. In an identical framework as ours, Okui (2009) derived a higher-order expansion of the MSE and proposed to choose the optimal number of moments conditions to minimize an estimated version of this expansion. However, the finite sample bias problem is not completely addressed since his simulations present large bias for the GMM estimator when the autoregressive parameter is close to unity.

The remainder of this paper is organized as follows. Section 1.1 presents the DPM and the classical GMM estimator. Section 1.2 presents regularized estimator whereas Sections 1.3 and 1.4 respectively present first asymptotic properties and high order properties of regularized GMM estimators. A data-driven selection of the regularization parameter is presented in section 1.5. Section 1.6 presents the extension of the model to exogenous covariates and the section 1.7 presents the results of Monte Carlo simulations. An empirical application estimating the effect of financial development on economic growth is discussed in Section 1.8. It appears that the regularization corrects the bias of the usual GMM estimator which seems to underestimate the financial development - economic growth effect.

Throughout the paper, we use the notations I and $I_{\bar{q}}$ respectively for the $N \times N$ and $\bar{q} \times \bar{q}$ identity matrix. The proofs are collected in an online appendix.

1.1 The model

We consider a simple $AR(1)$ model with individual effects described in the following equation : for $i = 1,...,N$, $t = 1,...,T$,

$$y_{it} = \delta y_{i,t-1} + \eta_i + v_{it}, \tag{1.1}$$

where δ is the parameter of interest satisfying $|\delta| < 1$, η_i is the unobserved individual effect, v_{it} the idiosyncratic error with conditional mean zero and variance σ^2 conditionally on $\eta_i, y_{it-1}...y_{i0}$. For simplicity, we assume that y_{i0}

is observed. Moreover, we denote $y_{i,t-1}$ by $x_{i,t}$.
As it is usual in estimating such models, we first transform the model to eliminate the individual effects.Two widely used transformations are the first differences and the foward orthogonal deviation operator. In this paper, we use the latter for theoretical and computational purposes. Indeed, this transformation preserves homoskedasticity and no serial correlation properties of the error term. Let the $(T-1) \times T$ matrix A denote the forward orthogonal deviations operator as used by Arellano and Bover (1995) and define $v_i^* = Av_i$, $x_i^* = Ax_i$, $y_i^* = Ay_i$ where $v_i = (v_{i1}, ..., v_{iT})'$, $x_i = (x_{i1}, ..., x_{iT})'$, $y_i = (y_{i1}, ..., y_{iT})'$. In particular, the $t-th$ element of y_i^* is given by

$$y_{it}^* = c_t[y_{it} - \frac{1}{T-t}(y_{it+1} + ... + y_{iT})]$$

with $c_t^2 = (T-t)/(T-t+1)$.
By multipliying the model by A, equation (1.1) becomes

$$y_{it}^* = \delta x_{it}^* + v_{it}^*$$

We have $E(x_{i,t}^* v_{it}^*) \neq 0$ so that the Ordinary Least Square (OLS) estimator of the transformed model is not consistent for fixed T as N tends to infinity. However, $E(x_{i,t-s} v_{it}^*) = 0$ for $s = 0, ..., t-1$ and $t = 1, ..., T-1$. Then, we consider the GMM estimator of δ based on these moment conditions. The number of moment conditions is $\bar{q} = T(T-1)/2$ which can be very large even if T is moderately large. Let z_{it}=$(x_{i1}, ..., x_{it})'$ and Z_i be the $(T-1) \times \bar{q}$ block diagonal matrix whose t-th block is z'_{it}. The moment conditions are then given by $E(Z'_i v_i^*) = 0$ with $v_i^* = (v_{i1}^*, ..., v_{i,T-1}^*)$. Under the conditional homoscedasticity of v_{it}, the covariance matrix of the orthogonality conditions is σ^2 $E(Z'_i Z_i)$. The GMM estimator of the parameter is given by

$$\hat{\delta} = \left(\sum_{t=1}^{T-1} x_t^{*'} M_t x_t^* \right)^{-1} \left(\sum_{t=1}^{T-1} x_t^{*'} M_t y_t^* \right)$$

with M_t the $N \times N$ matrix $Z_t(Z'_t Z_t)^{-1} Z'_t$ with $Z_t = (z_{1t}, ..., z_{Nt})'$, $x_t^* = (x_{1t}^*, ..., x_{Nt}^*)'$ and y_t^* defined in the same way. Letting x^*=$(x_1^{*'}, ..., x_{T-1}^{*'})'$ and y^*=$(y_1^{*'}, ..., y_{T-1}^{*'})'$, the GMM estimator can also be written as

$$\hat{\delta} = \frac{x^{*'} M y^*}{x^{*'} M x^*}$$

with $M = Z(Z'Z)^{-1}Z'$ and $Z = (Z'_1, ..., Z'_N)'$, a $N(T-1) \times \bar{q}$ matrix.

Even though, it is widely used by empirical researchers, this GMM estimator suffers from poor finite sample properties. Using a simple $AR(1)$, Blundell et Bond (1998) showed that the lagged levels of the dependent variable become weak instruments when the autoregressive parameter gets close to unity or when the variance of the unobserved individual effect increases toward the variance of the idiosyncratic error v_{it}. Moreover, Doran et Schmidt (2006) argue that in presence of many instruments, the marginal contribution of some of them can be small. As a result, many simulations including those in Okui (2009) showed that the GMM estimator of dynamic panel data performs poorly in these settings.
The intuition is that when T is very large, the dimension of the $\bar{q} \times \bar{q}$ matrix $Z'Z$ is large and its condition number (the ratio of its largest over its smallest eigenvalue) is also large. A large condition number indicates that the matrix is ill-conditioned and its inverse is highly unstable. The poor finite sample properties of the GMM estimator arise because inverting $Z'Z$ amplifies the potential sampling errors. We propose to use a regularized inverse of $Z'Z$ instead of the usual inverse $(Z'Z)^{-1}$ to compute the GMM estimator. Regularization can be seen as a way to stabilize the inverse and reduce the variability of the estimated weighting matrix, and consequently improve the finite sample properties of the estimator.

1.2 The regularized estimator

The regularization methods used in this paper are drawn from the literature on inverse problems (Kress, 1999). They are designed to stabilize the inverse of Hilbert-Schmidt operators (operators for which the eigenvalues are square summable). The matrix $Z'Z$ is not Hilbert-Schmidt however we will show in Lemma 1 that $K = E[Z'Z/NT^{3/2}]$ is Hilbert-Schmidt. So the regularization will be applied to $K_N = Z'Z/NT^{3/2}$, the sample counterpart of K.

Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 ... \geq \hat{\lambda}_q \geq 0$ be the eigenvalues of K_N. By spectral decomposition, we have $K_N = P_N D_N P'_N$ with $PP' = I_{\bar{q}}$ where P_N is the matrix of eigenvectors and D the diagonal matrix with eigenvalues $\hat{\lambda}_j$ on the diagonal. Let K_N^α denote the regularized inverse of K_N which is defined as

$$K_N^\alpha = P_N D_N^\alpha P'_N$$

where D_N^α is the diagonal matrix with elements $q(\alpha, \hat{\lambda}_j^2)/\hat{\lambda}_j$.

The positive parameter α is the regularization parameter, a kind of smoothing parameter, and the real function $q(\alpha, \lambda^2)$ depends on the regularization scheme used. As in Carrasco (2012), three regularization schemes will be considered : Tikhonov, spectral cut-off and Landweber Fridman regularization schemes. More details on these schemes can be found in Carrasco et al. (2007). If we let λ be an arbitrary eigenvalue of the matrix K_N, we can define :

1. **Tikhonov regularization (TH) :**
   This regularization scheme is close to the well known ridge regression used in presence of multicolinearity to improve properties of OLS estimators. In Tikhonov regularization scheme, the real function $q(\alpha, \lambda^2)$ is given by

$$q(\alpha, \lambda^2) = \frac{\lambda^2}{\lambda^2 + \alpha}.$$

2. **The spectral cut-off (SC)**
   It consists in selecting the eigenvectors associated with the eigenvalues greater than some threshold.

$$q(\alpha, \lambda^2) = I\{\lambda^2 \geq \alpha\} = \begin{cases} 1 \text{ if } \lambda^2 \geq \alpha, \\ 0 \text{ otherwise.} \end{cases}$$

   Another version of this regularization scheme is Principal Components (PC) which consists in using a certain number of eigenvectors to compute the inverse of the operator. PC and SC are perfectly equivalent, only the definition of the regularization term α differs. In PC, α is the number of principal components. In practice, both methods will give the same estimator so that we will study the properties of SC in detail in this paper.

3. **Landweber Fridman regularization (LF)**
   In this regularization scheme, K_N^α is computed by an iterative procedure with the formula

$$\begin{cases} K_{N,l}^\alpha = (I - cK_N)K_{N,l-1}^\alpha + cK_N, \ l = 1, 2, ..., 1/\alpha - 1, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad K_{N,0}^\alpha = cK_N \end{cases}$$

The constant c must satisfy $0 < c < 1/\lambda_1^2$ where λ_1^2 is the largest eigenvalue of the matrix K_N. Alternatively, we can compute this regularized inverse with

$$q(\alpha, \lambda^2) = 1 - (1 - c\lambda^2)^{\frac{1}{\alpha}}$$

In each regularization scheme, the real valued function $q(\alpha, \lambda^2)$ satisfies $0 \leq q(\alpha, \lambda^2) \leq 1$ and $\lim_{\alpha \to 0} q(\alpha, \lambda^2) = 1$ so that the usual GMM estimator corresponds to a regularized estimator with $\alpha = 0$.
Remark that $M = Z\left(Z'Z\right)^{-1} Z' = ZK_N^{-1}Z'/NT^{3/2}$. Similarly, let us denote the matrix $M^\alpha = ZK_N^\alpha Z'/NT^{3/2}$. The regularized GMM estimator for a given regularization scheme is :

$$\widehat{\delta}^\alpha = \frac{x^{*'} M^\alpha y^*}{x^{*'} M^\alpha x^*}. \tag{1.2}$$

The matrix K_N is a block diagonal matrix with the $t \times t$ matrix $Z_t' Z_t/NT^{3/2}$ at the $t-$th block. Exactly as K_N^{-1}, the regularized inverse K_N^α is also a block diagonal matrix where each block is the regularized inverse of the corresponding block of K_N [1]. So, if we define $M_t^\alpha = Z_t(K_{Nt})^\alpha Z_t'/NT^{3/2}$ with $(K_{Nt})^\alpha$ being the $t - th$ block of the matrix K_N^α, the regularized estimator can be rewritten as :

$$\widehat{\delta}^\alpha = \left(\sum_{t=1}^{T-1} x_t^{*'} M_t^\alpha x_t^*\right)^{-1} \left(\sum_{t=1}^{T-1} x_t^{*'} M_t^\alpha y_t^*\right).$$

1.3 First order asymptotic properties

In this section, we derive the asymptotic properties of the regularized estimator. As in Okui (2009), we make the following assumptions :
Assumption 1 : $\{v_{it}\}$ $(t = 1...T\,;\; i = 1...N)$ are *i.i.d.* across time and individuals and independent of η_i and y_{i0} with $E(v_{it}) = 0$, $var(v_{it}) = \sigma^2$, and $E(v_{it}^4) < \infty$.
Assumption 2 : The initial observation satisfies

$$y_{i0} = \frac{\eta_i}{1-\delta} + w_{i0} \qquad (i = 1, ..., N)$$

1. This holds because regularization transforms only the eigenvalues, not the eigenvectors.

where w_{i0} is independent of η_i and *i.i.d.* with the steady state distribution of the homogeneous process, so that $w_{i0} = \sum_{j=0}^{\infty} \alpha^j v_{i(-j)}$.

Assumption 3 : η_i are *i.i.d.* across individuals with $E(\eta_i) = 0$, var(η_i)=σ_η^2 with $0 < \sigma_\eta^2 < \infty$, and finite fourth order moment.

Moreover, asymptotic properties are derived under the assumption that both N and T go to infinity with $T < N$. Under this restriction the matrix K_N is non singular and so has nonzero eigenvalues.

Let $K = E[Z'Z/NT^{3/2}]$ and $(\lambda_l,\ \phi_l,\ l = 1, 2, ...)$ be the eigenvalues and orthonormal eigenvectors of K. In the inverse problem literature, this matrix is referred to as the operator. In Carrasco (2012) and Carrasco and Tchuente (2015), the operator is assumed to be a trace-class operator which is satisfied if and only if its trace is finite. Here, however, K is not trace class but it is Hilbert-Schmidt which is a slightly weaker condition. In the following lemma, we prove that K is a Hilbert-Schmidt matrix.

Lemma 1. *If assumptions 1-3 are satisfied, then*

(i) $Tr\,(K) = O\left(T^{1/2}\right)$;

(ii) The matrix K is a Hilbert-Schmidt matrix meaning that $tr\,(K^2) = O\,(1)$.

Lemma 1 shows that even though the eigenvalues of K are not summable as T goes to infinity, they are square summable. The Hilbert-Schmidt property is sufficient to derive proofs in our framework. This property is especially useful to establish the order of magnitude of the bias of the regularized estimator.

The following proposition provides the first order asymptotic properties of the regularized estimator.

Proposition 1.

If assumptions 1-3 are satisfied, α the parameter of regularization goes to 0, $\alpha\sqrt{NT}$ goes to infinity and both N and T tend to infinity with $T < N$, then

(i) Consistency : $\widehat{\delta}^{\alpha} \to \delta$ *in probability;*

(ii) Asymptotic normality : $\sqrt{NT}\left(\widehat{\delta}^{\alpha} - \delta\right) \xrightarrow{d} N\left(0, 1 - \delta^2\right)$.

For these properties, we need that α goes to zero slower than $\sqrt{NT}$ goes to infinity. Under similar assumptions, Alvarez and Arellano (2003) proved that the bias expression of the GMM estimator of DPM in the no regularization setting is given by the limit of

$$b_{NT} = \left[\frac{x^{*'}Mx^*}{NT}\right]^{-1}\left[-\frac{\sigma^2}{(1-\delta)}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} t\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]$$

and is of order $\sqrt{T/N}$. Hence, if T/N tends to a positive scalar, the classical GMM estimator is asymptotically biased. In our regularization setting, the bias is given by the limit of

$$b_{NT}^{\alpha} = \left[\frac{x^{*'}M^{\alpha}x^*}{NT}\right]^{-1}\left[-\frac{\sigma^2}{(1-\delta)}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} E[tr[M_t^{\alpha}]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]$$

This bias is of order $1/\alpha\sqrt{NT}$ so that the asymptotic bias of the regularized estimator vanishes under the assumption $\alpha\sqrt{NT}$ goes to ∞. We will take advantage of having an expression for the bias of the regularized estimator to derive a bias corrected regularized estimator. The simulations results on this bias corrected regularized estimator are provided in section 8. Since the asymptotic properties of the regularized estimator presented in this section do not depend on the regularization scheme we need to investigate higher order properties to establish the impact of regularization.

1.4 Higher-order asymptotic properties

In this section, we derive the Nagar (1959)'s decomposition of $E[(\hat{\delta}^{\alpha}-\delta)^2]$ the Mean Square Error (MSE) of our estimators. This type of expansion is used in many papers on instrumental variables such as Carrasco (2012), Donald and Newey (2001) and particularly Okui (2009) who works on a DPM. Moreover, this expansion will guide us in our goal to provide a data-driven method for selecting the regularization parameter.
The Nagar approximation of the MSE is the $\sigma^2 H^{-1} + S(\alpha)$ in the following decomposition :

$$NT(\hat{\delta}^{\alpha} - \delta)^2 = Q + r, \quad E(Q) = \sigma^2 H^{-1} + S(\alpha) + R \tag{1.3}$$

where $(r+R)/S(\alpha) \to 0$ as $N \to \infty, T \to \infty, \alpha \to 0$, and $H = \frac{\sigma^2}{1-\delta^2}\left(\frac{1}{T}\sum_{t=1}^{T}\psi_t^2\right)$.

Proposition 2.

Suppose assumptions 1-3 are satisfied and $E\left(v_{it}^{3}\right)=0$*. If* $N\to\infty$*,* $T\to\infty$*,* $\alpha\to 0$*,* $\alpha\sqrt{NT}\to\infty$*, and* $\alpha\left(\ln T\right)\sqrt{T}\to 0$*, then for the regularized GMM estimator, the decomposition given in* (1.3) *holds with :*

$$S(\alpha)=\frac{(1+\delta)^{2}}{NT}\left\{\sum_{t=1}^{T-1}E\Big(tr[M_{t}^{\alpha}]\Big)\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^{2}+\frac{(1-\delta^{2})^{2}}{\sigma^{2}}\frac{1}{NT}\sum_{t=1}^{T-1}\psi_{t}^{2}E[w_{t-1}'(I-M_{t}^{\alpha})^{2}w_{t-1}]$$

where $w_{it}=y_{it}-\eta_{i}/(1-\delta)$*,* $\phi_{j}=(1-\delta^{j})/(1-\delta)$*, and* $\psi_{t}=c_{t}(1-\delta\phi_{T-t}/(T-t))$*.*

In this decomposition, the first term of $S(\alpha)$ comes from the square of the bias that increases when α goes to zero whereas the second term is from the second-order expansion of the variance that decreases when α goes to zero. We observe the usual bias-variance trade-off that arises when selecting a tuning parameter. A large α will reduce the bias but increase the variance. The rate for the squared bias is $O(1/(\alpha^{2}NT))$ but the rate of the variance term is unknown. Unlike in Carrasco (2012) and in Carrasco and Tchuente (2015), our expression of MSE is unconditional as in Okui (2009) and Kuersteiner (2012). In Okui (2009) the GMM estimator is computed using $\min\{t,K\}$ lags for each period t with K the optimal number of instruments selected to minimize $S(K)$, a criterion similar to our $S(\alpha)$. The expression of $S(K)$ is simplified by $tr[M_{t}^{K}]=min\{t,K\}$ and $w_{t-1}'(I-M_{t}^{K})^{2}w_{t-1}=w_{t-1}'(I-M_{t}^{K})w_{t-1}$ because M_{t}^{K} is a projection matrix. In the present paper, $tr[M_{t}^{\alpha}]$ is random.

In our panel data setting, the bias expression of $S(\alpha)$ is the sum of the bias of each period $H^{-1}Etr[M_{t}^{\alpha}]E[\tilde{v}_{tT}'v_{t}^{*}]$ where H is the asymptotic variance. As the formula (3.14) in the special case of Kuersteiner (2012), this period bias expression is the product of the inverse of H, $Etr[M_{t}^{\alpha}]$ the contribution of the instrument matrix, and $E[\tilde{v}_{tT}'v_{t}^{*}]$ the correlation between the error term v_{t}^{*} and the residual from the reduced-form equation relating x_{it}^{*} to its optimal instrument $\psi_{t}w_{it}$. A difference with Kuersteiner (2012)'s is that the contribution $Etr[M_{t}^{\alpha}]$ depends on t and is not the number of instruments. The regularized estimator using spectral cut-off corresponds to the principal components estimator of Doran and Schmidt (2006) and, through $S(\alpha)$, we complement their paper with a data driven method for selecting the optimal number of principal components.

1.5 Data-driven selection of the regularization parameter

1.5.1 Estimation of the approximate MSE

In Propositon 2, we derived the leading terms of a second-order expansion of the MSE of the regularized estimator. The aim of this section is to select α that minimizes an estimated $S(\alpha)$. First, we introduce some notations :

$$\mathcal{A}(\alpha) = \sum_{t=1}^{T-1} E[tr[M_t^{\alpha}]] \left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} \right)$$

and

$$R(\alpha) = \frac{1}{NT} \sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t^{\alpha})^2 w_{t-1}]$$

so that,

$$S(\alpha) = \frac{(1+\delta)^2}{NT} \mathcal{A}(\alpha)^2 + \frac{(1-\delta^2)^2}{\sigma^2} R(\alpha).$$

Let $\hat{\delta}$ and $\hat{\sigma}^2$ be consistent estimators of δ and σ^2, respectively. Then $S(\alpha)$ can be estimated by

$$\widehat{S}(\alpha) = \frac{(1+\hat{\delta})^2}{NT} \widehat{\mathcal{A}}^2(\alpha) + \frac{(1-\hat{\delta}^2)^2}{\hat{\sigma}^2} \widehat{R}(\alpha)$$

with

$$\widehat{\mathcal{A}}(\alpha) = \sum_{t=1}^{T-1} tr(M_t^{\alpha}) \left(\frac{\widehat{\phi}_{T-t}}{T-t} - \frac{\widehat{\phi}_{T-t+1}}{T-t+1} \right)$$

where

$$\widehat{\phi}_j = \frac{1-\widehat{\delta}^j}{1-\widehat{\delta}}$$

and

$$\widehat{R}(\alpha) = \frac{1}{NT} \sum_{t=1}^{T-1} x_t^{*'} (I - M_t^{\alpha})^2 x_t^*.$$

Note that from Okui (2009, p.3),

$$E_{t-1}\left(x_{it}^*\right) = \psi_t \left(y_{it} - \frac{\eta_i}{1-\delta} \right) = \psi_t w_{it-1}$$

where E_{t-1} denotes the conditional expectation conditional on $(\eta_i, x_{it}, x_{it-1}, ...)$ so that $\widehat{R}(\alpha)$ is an unbiased estimator of $R(\alpha)$. The optimal parameter of regularization is selected by minimizing this estimated $S(\alpha)$

$$\hat{\alpha} = \arg\min_{\alpha \in \mathcal{E}_T} \widehat{S}(\alpha)$$

where $\mathcal{E}_T$ is the index set of α. $\mathcal{E}_T$ is a real compact subset for TH, $\mathcal{E}_T$ is such that $\frac{1}{\alpha} \in \{1, 2, ..., \bar{q}\}$ for PC, and $\mathcal{E}_T$ is such that $\frac{1}{\alpha}$ is a positive integer for LF. Next, we analyse the impact of using an estimated version of $S(\alpha)$ to select α instead of the true and unknow criterion.

1.5.2 Optimality

We wish to establish the optimality of the regularization parameter selection criterion in the following sense

$$\frac{S(\hat{\alpha})}{\inf_{\alpha \in \mathcal{E}_T} S(\alpha)} \xrightarrow{p} 1 \tag{1.4}$$

as $N \to \infty$, $T \to \infty$. It should be noticed that the result (1.4) is not a convergence result of $\hat{\alpha}$. It simply etablishes that using an estimated version of $S(\alpha)$ to choose the regularization parameter is asymptotically equivalent to using the true and unknown value of $S(\alpha)$.

Proposition 3.
Suppose that Assumptions 1-3 are satisfied and $\hat{\delta} \to \delta$, $\hat{\sigma}^2 \to \sigma^2$. If $N \to \infty$, $T \to \infty$ and $\alpha\sqrt{NT} \to \infty$, $\alpha (\ln T) \sqrt{T} \to 0$, $T^3/N (\ln T)^2 \to 0$, then the regularization parameter selection criterion is asymptotically optimal in the sense of (1.4) for Spectral Cut-off and Landweber Fridman regularization schemes provided that $\#\mathcal{E}_T = O(T^2)$ where $\#\mathcal{E}_T$ refers to the number of elements in the set $\mathcal{E}_T$.

Remarks.

1. Proposition 3 proves the optimality for Principal Components and Landweber Fridman regularization schemes which have discrete index set $\mathcal{E}_T$. The condition $\#\mathcal{E}_T = O(T^2)$ is a sufficient condition in the Landweber Fridman regularization scheme and necessarily holds for the principal components case [2]. Rather than imposing a maximum number of iterations, this

2. Recall that $\#\mathcal{E}_T = \bar{q}$ for principal components case.

condition restricts the order of magnitude of the number of elements of the index set $\mathcal{E}_T$. A rigorous proof for the Tikhonov's continous index set requires more complicated material which is beyond the scope of this work. However, optimality could be established for the continous index set case using a discretization of the compact set $\mathcal{E}_T$ and the fact that the regularization function $q(\alpha, \lambda^2)$ of Tikhonov regularization scheme is a real continous function as in Hansen (2007) .

2. Proposition 3 is related to Donald and Newey (2001) optimality result for the selection of the number of instruments in a linear IV model and Carrasco and Tchuente (2015) for the selection of the regularization parameter for the regularized LIML estimator. But their proofs rely on applying results by Li (1986, 1987) on cross-validation for the first stage equation. From $x_{it}^* = \psi_t w_{it} - c_t \tilde{v}_{it}$, the term $-c_t \tilde{v}_{it}$ can be regarded as the error term of the first stage equation since $\psi_t w_{it}$ is considered as the optimal instrument in Okui (2009). However, Li (1986, 1987)'s results do not apply in our framework because of the autocorrelation of this error term. As a result, our proof combines the strategies of Okui (2009) and Kuersteiner (2012).

3. Compared to Okui (2009), our proof is complicated by the fact that the $tr(M_t(\alpha))$ is random in our case while it is deterministic in Okui and the term $\mathcal{A}(\alpha)$, which is known in closed form in Okui, needs to be estimated here. As a result, we need to analyse extra terms and we obtain more stringent conditions on the rate at which T can diverge relatively to N. Our condition is $T^3/N(\ln T)^2 \to 0$ whereas it is $T \ln(T)/N \to 0$ in Okui (2009).

1.6 Introduction of exogenous regressors

In this section, we aim to generalize the model by taking into account exogenous covariates. We are now interested in the following model :

$$y_{it} = \delta y_{i,t-1} + \gamma' m_{it} + \eta_i + v_{it} \tag{1.5}$$

where m_{it} is a L_m dimensional vector of strictly exogenous variables in the sense that $E(m_{it} v_{is}) = 0$ for each t and s. Following Okui (2009), we assume that time-invariant variables f_i that satisfy $E(f_i v_{it}) = 0$ for all t are available and we denote by L_f the dimension of this vector. Even though they are often omitted in proofs, time-invariant variables are widely used in empirical works.

Let us define $\theta = (\delta, \gamma')'$, $x_{it} = (y_{i,t-1}, m_{it}')'$ and denote $y_i = (y_{i1}, ..., y_{iT})'$,

$x_i = (x_{i1}, ..., x_{iT})'$ and $v_i = (v_{i1}, ..., v_{iT})'$. Let A be the matrix of forward orthogonal deviation operator and denote $y_i^* = Ay_i$, $x_i^* = Ax_i$, $v_i^* = Av_i$. The model becomes :

$$y_{it}^* = \theta x_{it}^* + v_{it}^* \tag{1.6}$$

The vector of potential instruments for the endogenous regressor x_{it}^* is the $q_t = (L_f + (T+1)L_m + t)$ dimensional vector $z_{it} = (f_i', m_{i0}', ..., m_{iT}', y_{i0}, ..., y_{i,t-1})'$. In this setting, the total number of instruments is $\bar{q} = \sum_t q_t$. Let us define the following matrix $Z_t = (z_{1t}', ..., z_{Nt}')'$, $x_t^* = (x_{1t}^*, ..., x_{Nt}^*)'$ and $y_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$. If we denote $K_N = Z'Z/NT^{3/2}$ and K_N^α the regularized inverse of K_N given a regularization parameter α, then the regularized GMM estimator of θ is

$$\widehat{\theta^\alpha} = \left(x^{*'} M^\alpha x^*\right)^{-1} \left(x^{*'} M^\alpha y^*\right) \tag{1.7}$$

with $M^\alpha = ZK_N^\alpha Z'/NT^{3/2}$, $Z = (Z_1', ..., Z_N')'$ and Z_i has the same definition as in the model without covariates.

We now make assumptions to derive the second-order expansion of the MSE of $\hat{\theta}$ in this general model. Let $E_Z(a) = E(a|\eta_i, z_{it}, z_{i,t-1}, ...)$ for the random variable a.

Assumption 1' : (i) $\{v_{it}\}$ $(t = 1...T\,;\ i = 1...N)$ are iid across time and individuals and independent of η_i and y_{i0} with $E_Z(v_{it}) = 0$, $E_Z(v_{it}^2) = \sigma^2 < \infty$, $E_Z(v_{it}^3) = 0$, $E_Z(v_{it}^4) < \infty$ and finite moments up to fourth order. (ii) η_i are $i.i.d$ across individuals with $E(\eta_i) = 0$, $var(\eta_i) = \sigma_\eta^2$, and finite fourth order moment.

Assumption 2' : (i) (y_{it}, m_{it}) is a strictly stationary finite-order vector autoregressive (VAR) process conditional on η_i such that the distribution of $\{(y_{it}, m_{it}')', ..., (y_{i,t+s}, m_{i,t+s}')'\}$ conditional on η_i does not depend on the subscript t for all s. (ii) $\{\{m_{it}\}_{t=1}^T\}_{i=1}^N$ is an $i.i.d.$ sequence across individuals with finite fourth-order moments.

These previous two assumptions are from Okui (2009) who also states that $E_Z(x_{it}^*) = \tilde{w}_{i,t-1} = (w_{i,t-1}, m_{it}^*)'$ with

$$w_{i,t-1} = \psi_t(y_{i,t-1} - \mu_i) - \frac{c_t}{T-t}\gamma'(\phi_{T-t} m_{i,t} + ... + \phi_1 m_{i,T-1})$$

where $c_t = \sqrt{\frac{T-t}{T-t+1}}$, $\mu = \frac{\eta}{1-\delta}$ and $\phi_j = \frac{1-\delta^j}{1-\delta}$.

Let $K = E[Z'Z/NT^{3/2}]$ and $(\lambda_l,\ \phi_l,\ l = 1, 2, ...)$ be the eigenvalues and orthonormal eigenvectors of K. The matrix K is assumed to be a Hilbert-Schmidt matrix. Moreover, in the extended model we make an assumption on the growth rate of the eigenvalues of K. If we define $\tilde{W} = (\tilde{w}_1', ..., \tilde{w}_{T-1}')'$ with $\tilde{w}_t = (\tilde{w}_{1t}, ..., \tilde{w}_{N,t})'$, then we impose the following condition :

Assumption 3 : The matrix K is Hilbert-Schmidt and there is a $\beta > 0$ such that

$$\frac{1}{NT} E \sum_{j=1}^{\infty} \left[\frac{< \widetilde{W}_a, \widehat{\psi}_j >^2}{\widehat{\lambda}_j^{2\beta}} \right] < \infty$$

for all N and T, where $\widetilde{W}_a$ is the ath column of $\widetilde{W}$, $\widehat{\psi}_j$ and $\widehat{\lambda}_j$ denote the eigenvectors and eigenvalues of $ZZ'/NT^{3/2}$ and $<, >$ denotes the inner product in $\mathbb{R}^{N(T-1)}$.

Assumption 3 is a source condition similar to Assumption 2(ii) in Carrasco (2012). It requires that the Fourier coefficients $< \widetilde{W}_a, \widehat{\psi}_j >$ decline faster than the eigenvalues $\hat{\lambda}_j$ to a certain power. It allows us to derive the rate of convergence of the MSE. More precisely, under this assumption we have that $E[||\widetilde{W} - M^{\alpha}\widetilde{W}||^2] = O(\alpha^{\beta})$ for PC and LF and $E[||\widetilde{W} - M^{\alpha}\widetilde{W}||^2] = O(\alpha^{min(\beta,2)})$ for TH.

We now prove that under these assumptions, we can isolate the leading terms of a second-order expansion of the MSE of $\hat{\theta}$: $NTE\left[\left(\widehat{\theta} - \theta_0\right)\left(\widehat{\theta} - \theta_0\right)'\right]$.

Proposition 4.
Assume that assumptions 1', 2', and 3 are satisfied. If the parameter of regularization α goes to 0, N and T go to infinity, $\alpha \ln(T)\sqrt{T} \to 0$, and $\alpha^{min\{\beta,1\}}\sqrt{NT} \to \infty$, then the leading terms in the higher-order expansion of the MSE of $\widehat{\theta}$ have the following form :

$$S(\alpha) = H^{-1}\left\{ \frac{\sigma^4}{(1-\delta)^2} \begin{bmatrix} \mathcal{A}(\alpha) & 0 \\ 0 & 0 \end{bmatrix} + \frac{\sigma^2}{NT} \sum_{t=1}^{T-1} E[\tilde{w}_{t-1}'(I - M_t^{\alpha})^2 \tilde{w}_{t-1}] \right\} H^{-1},$$

where

$$\mathcal{A}(\alpha) = \frac{1}{NT} \left[\sum_{t=1}^{T-1} Etr[M_t^{\alpha}] \left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} \right) \right]^2$$

and

$$H=\begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}E\left(w_{it}^{2}\right) & \frac{1}{T}\sum_{t=1}^{T}E\left(w_{it}m_{it}^{*\prime}\right) \\ \frac{1}{T}\sum_{t=1}^{T}E\left(m_{it}^{*}w_{it}\right) & \frac{1}{T}\sum_{t=1}^{T}E\left(m_{it}^{*}m_{it}^{*\prime}\right) \end{bmatrix}.$$

As in the model without covariates, the first part of $S(\alpha)$ is the squared bias and increases as the regularization parameter goes to zero. The second term of $S(\alpha)$ is the second order variance of the regularized estimator and it decreases when α goes to zero.
We observe that the bias term depends only on the bias of the autoregressive coefficient $\hat{\delta}$ and not on $\hat{\gamma}$, whereas the variance term depends on both so that globally the MSE depends also on $\hat{\gamma}$. This is different from Okui (2009). In his Theorem 4, Okui imposes the extra assumption that the subset of instruments z_{it}^{K} includes either y_{it-1} and m_{it}^{*} or linear combinations of these. He shows that, in this case, only the element (1,1) of the matrix $S(\alpha)$ is nonzero so that he can focus on this scalar to select the regularization parameter. Interestingly, it means that only the MSE of $\widehat{\delta}$ matters for selecting α. In contrast, we do not impose this extra assumption and we show that the MSE of $\hat{\gamma}$ also plays a role. Given $S(\alpha)$ is a matrix, α can be selected by minimizing $\ell'\widehat{S}(\alpha)\ell$ for an arbitrary L_m+1 vector ℓ and some estimator $\widehat{S}(\alpha)$ of $S(\alpha)$. In the simulations, we choose ℓ so that $\ell' H^{-1}$ is a vector of ones. For the estimation of $S(\alpha)$, similarly to the model without covariates, $\mathcal{A}(\alpha)$ can be estimated by

$$\hat{\mathcal{A}}(\alpha)=\frac{1}{NT}\left[\sum_{t=1}^{T-1}tr[M_t^{\alpha}]\left(\frac{\hat{\phi}_{T-t}}{T-t}-\frac{\hat{\phi}_{T-t+1}}{T-t+1}\right)\right]^2,$$

the unknown parameters, σ^2 and ϕ_j, are estimated using a preliminary estimation of θ, and $E[w_{t-1}'(I-M_t^{\alpha})^2 w_{t-1}]$ can be estimated by $x_t^{*\prime}(I-M_t^{\alpha})^2 x_t^{*}$ where $x_t^{*}=(u_t^{*},m_t^{*\prime})'$ with $u_t=y_{t-1}$.

1.7 Simulation study

In this section, we present Monte Carlo simulations to illustrate the finite sample properties of the regularized estimators and compare them to others estimators. We investigate on the simple DPM as well as the extension of the model to include covariates.

1.7.1 Model without covariate

We first consider the simple DPM given by the following autoregressive model

$$y_{it} = \delta y_{i,t-1} + \eta_i + v_{it},$$

with $\eta_i \sim i.i.d.N(0, \sigma_\eta^2)$, $y_{i0} \sim i.i.d.N(\eta_i/(1-\delta), \sigma^2/(1-\delta^2))$ and $v_{it} \sim i.i.d.N(0, \sigma^2)$. Each simulation is defined by a choice of vector $(N, T, \sigma^2, \sigma_\eta^2, \delta)$. We consider $N = 50$ and $N = 100$. For each value of N, we simulated samples with $T = 10$ and $T = 25$ and for three values of δ=(0.5,0.75,0.95). The number of replications is 5000 for all cases. Different estimators of the parameter of interest δ are presented. We denote by GMM the usual GMM estimator using all available lags of y_{it} as instruments. GMM5 is the GMM based on the five first instruments at each period. This ad hoc selection of the number of lags to be used as instruments at each period is used in empirical studies to handle the many instruments problem (Roodman, 2009). The estimator proposed in Okui (2009) based on optimal selection of instruments is denoted IVK. Finally the regularized estimators are denoted as TH for Tikhonov, PC for principal components and LF for Landweber-Fridman.
The regularized estimators presented in this section are bias corrected via an iterative procedure. From the bias expression given in Section 1.4, we can derive an expression for the finite sample bias of the regularized estimator

$$b_{NT}^\alpha = -\frac{(1+\delta)}{\sqrt{NT}} \sum_{t=1}^{T-1} E[tr[M_t^\alpha]] \left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} \right)$$

and prove that the bias vanishes when $N \to \infty$, $T \to \infty$, $\alpha \to 0$ and $\alpha\sqrt{NT} \to \infty$. Since the bias is a non-linear function of the parameter of interest, we obtain the bias-corrected regularized estimator through an iterative procedure.
If we denote $Q_T(\alpha) = \sqrt{NT} b_{NT}^\alpha$ then we have

$$\hat{\delta}^\alpha = \delta - \frac{(1+\delta)}{NT} Q_T(\delta)$$

and the bias corrected regularized estimator of δ is given by

$$\hat{\delta}^{\alpha}_{i+1} = \frac{\hat{\delta}^{\alpha}_{i} + Q_T(\hat{\delta}^{\alpha}_{i})/NT}{1 - Q_T(\hat{\delta}^{\alpha}_{i})/NT}$$

The initial value for this iterative procedure is a regularized estimator where the regularization parameter is selected by minimizing $\widehat{S}(\alpha)$ as described in Section 1.5. The GMM estimator is used to estimate unknown parameters in the minimization criteria. Hence, the estimate for the variance σ^2 is given by

$$\widehat{\sigma^2} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T-1} (y^*_{it} - \hat{\delta} x^*_{it})^2$$

where $\hat{\delta}$ denotes the GMM estimator of the parameter of interest δ. For each estimator, we compute the median bias (Med.bias), the median absolute bias (Med.abs), the length of the inter quartile range (Iqr.), the median mean square error (Med.mse), and the coverage probabilities (cov) of the 95 % confidence intervals. The standard error is computed with the formula :

$$\bar{V} = \sqrt{\frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T-1} (y^*_{it} - \hat{\delta} x^*_{it})^2 [x^{*\prime} M^{\alpha 2} x^*]^{-1}}$$

Table A.1 presents the distribution of the condition number of the matrix $Z'Z$. The condition number is defined as the ratio of the largest eigenvalue on the smallest one and is independent of the scaling (so that the condition number of $Z'Z$ is the same as that of K_N). The higher the condition number is, the more ill-conditioned the matrix is and so inverting its eigenvalues can negatively affect the estimator, therefore the need of regularization is higher. We present the min, the first quartile, the mean, the median, the third quartile and the max. The last column gives the dimension of the matrix $Z'Z$ which is the total number of instruments $\bar{q} = 0.5 \times T \times (T-1)$ for each value of T. From Table A.1, we observe that the need of regularization increases with T for a given δ and also increases with δ for a given T. As δ gets closer to 1, the instruments become weak yielding a ill-conditioned matrix.
Table A.2-A.9 in appendix contain simulations for different combinations

of $N, T, \delta, \sigma^2, \sigma_\eta^2$ for the simple model without covariate. For each simulations setting, we provide a table of the properties of the optimal regularization parameter. Table A.3 contains summary statistics for the value of the regularization parameter which minimizes the approximate MSE for $N = 50, \sigma^2 = 1, \sigma_\eta^2 = 1$, and different values of δ and T. We report the mean, the standard error (std), the mode and the three quartiles of the distribution of the regularization parameter. The regularization parameter is the optimal α for TH, the optimal number of iterations for LF and the optimal number of principal components for PC. Noting that the standard GMM estimator corresponds to the Tikhonov estimator with $\alpha = 0$, to the LF estimator with an infinite number of iterations and to the PC using all the principal components, a higher level of regularization will correspond to a larger value of α, a smaller number of principal components and a smaller number of iterations for the LF estimator.
The proportion of PC decreases with T (and hence the number of total instruments) for a given δ. This proportion also decreases as δ gets close to unity for a given T. Hence, as we could expect, a smaller number of principal components is selected as the need of regularization increases through an higher condition number. For TH, for a given value of T, the optimal α increases with δ so then with the need for regularization. But, when δ is fixed, the optimal α decreases as T increases and the need of regularization increases so that settings with large condition numbers are associated with small α. This is because an increase of T has two potential effects on the bias of the GMM estimator : a reduction of the bias as the sample size increases with T and an increase of the bias as the number of instrument increases with T. The small optimal regularization parameter for the TH in large T settings seems to reveal that the first effect dominates the second one. Table A.3 results also show that the optimal number of iterations of LF is high in large T settings for a given δ for LF and in large δ settings for a given T.

Results comparing the six estimators are presented in Table A.2 for $N = 50$. For the LF regularization scheme, the parameter c is set to $0.1/\widehat{\lambda}_1^2$ (where $\widehat{\lambda}_1$ is the largest eigenvalue of K_N) as suggested in Carrasco (2012). The estimators GMM5 and IVK have better properties than the usual GMM estimator as expected. Regularization improves the finite sample properties of the GMM estimator of the DPM model. In almost all combinaisons of T and δ, PC and LF have smaller bias, smaller MSE and better coverage rate than the usual GMM, the GMM5 and the IVK. Moreover, as shown by the Iqr, regularized estimators tend to have less concentrated distribution than the GMM estimator and more concentrated distribution than IVK giving better coverage.

Comparisons between regularized estimators show that PC performs much better than TH and LF in all settings in terms of bias and MSE. The LF regularized estimator provides the best coverage probabilities. TH does not perform as well in this setting as it has higher bais and MSE.

Tables A.4 and A.8 in appendix present simulations results when $\sigma_\eta^2 = 10$ instead of $\sigma_\eta^2 = 1$ in the previous tables. In addition to the case when the autoregressive parameter is close to unity, larger ratio σ_η^2/σ^2 setting is known to be a weak instrument settings where the GMM perform very poorly Blundell et Bond (1998). In comparison to Tables A.2 and A.6, all the estimators perform worse in terms of bias, MSE and coverage rates. However, regularization procedure provides better properties especially when T is large and δ close to unity. To summarize these simulation results, one can notice that regularization procedure proposed in this paper reduces the bias, the MSE and increases the coverage rates of the GMM estimator. Our regularized estimators outperform the GMM5 and the IVK especially in small T when δ is not large and in large T setting when δ is close to unity. Regularized TH and LF provide the largest bias correction whereas LF rather provides better coverage rate.

1.7.2 Model with a strictly exogenous regressor

We now consider the model including a strictly exogenous covariate. The equation is given by :

$$\begin{aligned} y_{it} &= \delta y_{it-1} + \theta m_{it} + \eta_i + v_{it}, |\delta| < 1, \\ m_{it} &= \rho\eta_i + e_{it}, \end{aligned}$$

with $v_{it} \sim i.i.d.\ N(0, \sigma^2)$, $\eta_i \sim i.i.d.\ N(0, \sigma_\eta^2)$, $e_{it} \sim i.i.d.\ N(0, \sigma_e^2)$. Moreover, the initial value of y_{i0} is drawn by

$$y_{i0} \sim i.i.d.N\left(\eta_i \frac{1+\rho\theta}{1-\delta}, \frac{\theta^2\sigma_e^2 + \sigma^2}{1-\delta^2}\right)$$

In this setting, for each period t, m_{i0}...,m_{iT} are potential instruments in addition to the lags of y_{it}. Hence, in comparison to the model without covariate, the total number of instruments increases from $\bar{q} = 0.5 \times T \times (T-1)$ to

$\bar{q} = 0.5 \times T \times (T-1) + (T-1) \times (T+1)$. We present the results with fixed $\rho = 1$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_e^2 = 1$, $N = 50$. The optimal regularization parameter is selected following the procedure described in Section 1.5. Simulation results are presented in Table A.10 for three values of the autoregressive parameter : 0.50 , 0.75 and 0.95 and the fixed value $\theta = 0.95$ for the coefficient of the covariate.
As for the model without covariate, we compare the estimators, GMM, GMM5, IVK, TH, PC, LF, using the median bias (Med.bias), the median absolute bias (Med.abs), the Interquartile range (Iqr), the mean square error (Med.mse) and the coverage probability (cov).
Table A.10 results show a significant reduction of the bias in almost all settings on both the estimate of the autoregressive parameter as well as the exogenous covariate estimate. PC regularisation scheme provides the best properties among all the estimators. In this model with covariate, the regularized estimators perform relatively better than in the model without covariate certainly because the former has much more instruments. It is worth noting that the regularized estimators are not bias-corrected since we did not derive the expression of the bias of the regularized estimator in presence of exogenous regressors.

1.7.3 Model with an endogenous regressor

Even though in the section 1.6 we only considered the case of a strictly exogenous covariates in order to simplify proofs, we now present simulation results with an endogenous covariate. This is obtained by adding a v_{it-1} term in the equation of the covariate.

$$y_{it} = \delta y_{it-1} + \beta m_{it} + \eta_i + v_{it}, |\delta| < 1,$$
$$m_{it} = \rho \eta_i + v_{it-1} + e_{it}, |\rho| < 1$$

with $v_{it} \sim i.i.d.N(0, \sigma^2)$, $\eta_i \sim i.i.d.N(0, \sigma_\eta^2)$, $e_{it} \sim i.i.d.N(0, \sigma_e^2)$. Moreover, the initial value of y_{i0} is drawn by

$$y_{i0} \sim iidN\left(\eta_i \frac{1+\rho\beta}{1-\delta}, \frac{\beta^2\sigma_e^2 + \sigma^2}{1-\delta^2}\right)$$

In the case of a strictly exogenous regressor, all the values (past and future) of the covariate are used as instruments. When the covariate is endogenous, only the past values are valid instruments. As result, there are fewer instruments in comparison with the strictly exogenous case but still more instruments than in the model without covariate. Table A.11 presents simulation results of the DPM extended with an endogenous covariate. The sample size and the values of the parameters are selected to be close to the values obtained in the empirical application in 1.8. Results in Table A.11 show that regularization reduces the bias of the estimators especially for the coefficient of the covariate.

1.8 Empirical application

In this section, we estimate the effect of financial development on economic growth using the DPM. Financial development can exert a causal influence on economic growth by improving information asymmetries and facilitating transactions. As a solution to potential endogeneity of financial development in a growth regression due to reverse causality or mesurement errors, DPM are widely used to evaluate the effect of the exogenous components of financial development on economic growth (see Levine et al. (2000), Beck et al. (2000) among others). This model improves upon the usual purely cross-sectional model by adding more variability through the time-series dimension, taking into account the unobserved country-specific effects and controlling for the potential endogeneity of all explanatory variables.
The most used model in the literature to address the financial development-economic growth question is the following :

$$y_{it} = \delta y_{i,t-1} + \beta_1 FinancialDevelopment_{it} + \beta' X_{it} + \eta_i + \epsilon_{it} \quad (1.8)$$

where y is the log real GDP per capita, X is a matrix of explanatory variables, η an unobserved country-specific effect, ϵ is the error term, and the subscripts i and t represent country and time period. $FinancialDevelopment$ is the indicator of financial development. For ease of interpretation this regression model can be transformed into :

$$y_{it} - y_{i,t-1} = (\delta - 1) y_{i,t-1} + \beta_1 FinancialDevelopment_{it} + \beta' X_{it} + \eta_i + \epsilon_{it}$$

as the $y_{it} - y_{i,t-1}$ can be seen as the variation of the real GDP per capita. In the estimation, many empirical researchers use the growth rate of the real GDP per capita $g_{it} = [exp(y_{i,t}) - exp(y_{i,t-1})]/exp(y_{i,t-1})$ instead of $y_{it} - y_{i,t-1}$.

But using this approximation removes the autoregressive part of the model as g_{it-1} is not included as a regressor. So we will estimate exactly the regression equation (1.8) following our theorical model.
Many indicators are used in the literature to measure financial development. In this paper we combine two of them : domestic credit to private sector as a share of GDP and stock market capitalization as a share of GDP. This approach presents the advantage to take into account both the development of financial markets (stock market capitalization) and the development of financial intermediaries (credit to private sector). Khan and Senhadji (2000) consider that the most exhaustive indicator of financial depth is the one that combines domestic credit to private sector as a share of GDP, stock market capitalization as a share of GDP and bond (private and public) market capitalization as share of GDP. Since this last variable is not available for a lot of countries, we do not consider it in the indicator of financial development. To be more precise, in this paper, we measure financial development by the sum of domestic credit to private sector as a share of GDP and stock market capitalization as a share of GDP.
The data used are from the World Development Indicators of the World Bank Group (2015) and cover the period from 1990-2011. The choice of this period is guided by the availability of the data on financial development. Moreover, we worked on yearly data instead of non-overlapping five-year average data as this reduces the time dimension of the panel. We depart from all the previous papers on the growth-financial development by the estimation the true baseline DPM equation (1.8), the use of more recent yearly data and the use of regularization as estimation technique. The matrix of controls X includes the gross enrolment ratio in secondary education as a control for Human capital, Trade Openness measured by the sum of exports and imports as a percentage of GDP, the inflation rate measured by the variation in the consumer price index and the government size measured by the government spending share of the GDP. Moreover, all other variables are expressed as natural logarithms. The panel unit root tests on the variables in log reject the presence of unit roots. The final sample is an unbalanced panel for 77 countries covering 22 periods from 1990 to 2011 [3]. We also included time dummies to account for time-specific effects [4].

3. To deal with missing values, equations with at least one missing observation have been deleted.

4. Actually we included dummies for each five year period from 1990-2010.

	GMM	TH	PC	LF
δ	0.869	0.737	0.808	0.815
	(0.016)	(0.060)	(0.071)	(0.269)
β_1	0.019	0.040	0.032	0.076
	(0.004)	(0.014)	(0.015)	(0.095)
Optimal parameter		0.003	36	2000

Table 1.1 – Impact of financial development on economic growth

We have a total of $\bar{q} = 517$ instruments so that the covariance matrix $K_N = Z'Z/NT^{3/2}$ has a large condition number (of the order of 10^{19}) and many very small eigenvalues which are a motivation for the use of regularization. In Table (1.1), we report various estimates of the financial development indicator and their standard errors (in brackets). We also report for regularized estimators, the optimal regularization parameter for Tikhonov scheme (TH), the optimal number of Principal Components (PC) and the optimal number of iterations for the Landweber Fridman regularization scheme (LF). These parameters are selected following the procedure described in section 1.5. The GMM estimator using all available instruments (all lagged values of the regressors) is used as preliminary estimator to estimate unknown parameters of $S(\alpha)$. As optimization set E_n, we choose a grid of 21 points between 0.002 and 0.004 for TH whereas the optimal number of iterations for LF is searched from 1000 to 2500. For PC, the number of principal components is selected between 1 and $\bar{q}$.

The regularized estimates of the effect of financial development on economic growth are larger than the GMM one suggesting that these methods provide a bias correction. However, their standard errors are also larger. This illustrates the trade-off between bias and variance. The high bias correction of regularized estimators is reflected in the optimal regularization parameters. Indeed, less than 10 % of the principal components (36 out of 527) are selected whereas a relatively low number of iterations is selected in the case of LF. Analysis in terms of confidence interval shows that none the regularized estimators is included in the confidence intervall of the GMM coefficient. As in the simulations, PC was found to be more reliable than the other regularizations, we are focusing our attention on the PC estimator. According to the PC regularized GMM estimator, the financial development has a positive and significant effect on the economic growth. An increase of 1% of the financial development is associated with an increase in the real GDP per capita of 0.032%.

With an autoregressive coefficient close to unity and a very large number of instruments, this empirical application provides a good illustration of the bias correction power of regularization on the GMM estimation in DPM. It is worth noting that our results may not be directly comparable to those in the literature because we used the most recent data available and estimated the baseline DPM so we did not approximate $y_{it} - y_{i,t-1}$ by the growth rate of the real GDP per capita. Furthermore we worked on yearly data instead of five-year average data.

1.9 Conclusion and further extensions

In dynamic panel data models, the number of moment conditions increase with the sample size so that the GMM estimator has poor finite sample properties. Instead of selecting a subset of moment conditions, we propose a regularization approach based of three ways of inverting the covariance matrix of instruments. All the regularization methods involve a tuning parameter which is selected by a data-driven method based on a higher-order expansion of the MSE under double asymptotic. Simulations show that those estimators outperform the classical GMM estimator especially in weak instruments settings.

There are several possible extensions to this work. To address the poor finite sample problem of the GMM, Blundell and Bond (1998) proposed the system GMM estimator of dynamic panel model which combines moment conditions for the model in first differences with moment conditions for the model in levels. However, even though it is widely used in empirical analysis (Blundell and Bond (2000), Levine et al. (2000) among others) the weak instrument problem in the GMM estimation of dynamic panel data models is not completely resolved by the system GMM estimator. Actually, Bun and Windmeijer (2010) show that the system GMM estimator suffers from the weak instrument problem if the variance ratio of individual effects to the disturbance is large. Then, extending our regularization approach to the system estimator would be of great interest. Another interesting extension of this paper would be to derive the regularized Limited Information Likelihood maximum (LIML). The LIML estimator is known to have smaller bias properties than the GMM estimator. Then applying regularization to the LIML estimator may provide an improved estimator of DPM.

Chapitre 2

Regularized LIML for dynamic panel data models *

This paper focus on the estimation of a linear dynamic panel data model (DPM) using a large number of moment conditions. DPM permits to take into account both the dynamic nature of a problem and the within population heterogeneity. In applied research, these models have been used to study growth models (Caselli et al. (1996), Bond, Hoeffler and Temple (2001), Levine et al. (2000)), Income dynamics (Hirano (2002)), partial adjustment model for employment dynamics (Arellano (2016)) estimation of production functions (Blundell and Bond (2000)), Poverty(Forbes (2000)) and Inequality (Dollar and Kraay (2002)) analysis and many other economic phenomena.
Generalized Method of Moments (GMM) estimators are widely used to estimate DPM with lagged levels dependent variable as instruments. A feature of dynamic panel models is that, if a variable at a certain time period can be used as an instrument, then all the past realizations of that variable can also be used as instruments. The number of moment conditions can then be very large even if the time dimension is moderately large. Although using many instruments increases the asymptotic efficiency of GMM estimator, it has been proved that its finite sample bias also increases with the number of

*. This chapter is co-authored with my advisor Marine Carrasco (Université de Montréal).

instruments. Therefore, estimation in presence of many moment conditions involves a variance-bias trade-off also referred to as the many instruments problem.

As a solution to the many instruments problem, Carrasco (2012) proposed the regularized approach based on the way of inverting the covariance matrix of instruments. The main originality of this estimation procedure is that it does not need to restrict the number of instruments, which may be smaller or larger than the sample size, or even infinite. No instruments are discarded a priori. Following the regularized two-stage least squares (2SLS) proposed in this paper, the regularized version of the Limited Information Maximum Likelihood (LIML) estimator is proposed by Carrasco and Tchuente (2015) in a cross sectional setting. The first application of regularization in a panel framework is proposed by Carrasco and Nayihouba (2019) (CN (2019) henceforth) who derived the regularized GMM estimator for DPM. This paper proposes a regularization approach to the LIML estimator using three dimension reduction schemes. The first scheme is based on principal components, the second one is based on Ridge regression (also called Bayesian shrinkage), finally the last one is an iterative method called Landweber Fridman. All these methods involve a regularization parameter similar to the smoothing parameter in nonparametric regressions. This parameter needs to converge to zero at an appropriate rate to obtain an efficient estimator.

We show that the regularized LIML estimator is consistent and asymptotically normal under double asymptotic as in Alvarez and Arellano (2003). We also derive the leading term of the MSE in a second order expansion of Nagar (1959)'s type for the regularized estimators. Our three estimators involve a regularization or tuning parameter, which needs to be selected in practice. The expansion of the MSE provides a tool for selecting the regularization parameter. Following the same approach as in Carrasco (2012) and Carrasco and Tchuente (2015), we propose a data-driven method for selecting the regularization parameter α based on an approximation of the MSE. We show that this selection method is optimal in the sense of Li (1986, 1987), meaning that the choice of α using the estimated MSE is asymptotically as good as minimizing the true and unknown MSE.

Although the LIML estimator is proven to have better properties than the GMM ones, the existing literature on the many instuments problem in the DPM framewok mainly focus on the latter estimator. Hence, Andrews and Lu (2001) introduce and apply to the DPM, a consistent model and moment selection procedure. But, their method aims to discriminate correct moment

conditions from incorrect ones whereas we assume that all the moment conditions are valid. Under double asymptotic, Okui (2009) proposes to select the optimal number of lags by minimizing an approximation of the MSE based on higher-order asymptotic theory. Our method uses all the available instruments and then is not an instruments selection one. Our approach is closer to the one of Doran and Schmidt (2006) who investigate in a simulation study the use of principal components in panel data models. Several bias corrected estimator have been proposed for dynamic panel data models (Hahn et al. (2001), Bun and Kiviet (2006), Alvarez and Arellano (2003), Kiviet (1995), Hahn and Kuersteiner (2002)). Our methodology complements those estimators since we can apply these bias correction procedures to our regularized estimator to improve the finite sample properties. The very few papers on the LIML estimator for the DPM have focused on deriving its asymptotic properties. These include Alvarez and Arellano (2003), Akashi and Konitomo (2011). Following Donald and Andrews (2001) and Carrasco and Tchuente (2015) on cross-sectional data models, this paper aims to contribute to the literature on the LIML estimator by proposing regularization as a solution to the many instruments problem in the DPM framework.
The remainder of this paper is organized as follows. Section 2.1 presents the DPM and the usual LIML estimator. In the section 2.2, the regularized estimators are presented whereas section 2.3 and 2.4 respectively present first order asymptotic properties and high-order properties of regularized LIML estimators. A data-driven selection of the regularization parameter is presented in section 2.5 and section 2.6 presents the results of Monte Carlo simulations.
Throughout the paper, we use the notations I and $I_{\bar{q}}$ respectively for the $N \times N$ and $\bar{q} \times \bar{q}$ identity matrix.

2.1 The model

We consider a simple $AR(1)$ model with individual effects described in the following equation : for i=1...N, t=1...T,

$$y_{it} = \delta y_{i,t-1} + \eta_i + v_{it}, \tag{2.1}$$

where δ is the parameter of interest satisfying $|\delta| < 1$, η_i is the unobserved individual effect, v_{it} the idiosyncratic error with conditional mean zero and variance σ^2 conditionally on $\eta_i, y_{it-1}...y_{i0}$. For simplicity, we assume that y_{i0}

is observed. Moreover, we denote $y_{i,t-1}$ by $x_{i,t}$.
As it is usual in estimating such models, we first transform the model to eliminate the individual effects. Two widely used transformations are the first differences and the foward orthogonal deviation operator. In this paper, we will use the latter for theoretical and computational purposes. Indeed, this transformation preserves homoskedasticity and no serial correlation properties of the error term. Let the $(T-1) \times T$ matrix A denotes the forward orthogonal deviations operator as used by Arellano and Bover (1995) and define $v_i^*{=}Av_i$, $x_i^*{=}Ax_i$, $y_i^*{=}Ay_i$ where $v_i = (v_{i1}, ..., v_{iT})'$, $x_i = (x_{i1}, ..., x_{iT})'$, $y_i = (y_{i1}, ..., y_{iT})'$. For example the $t-th$ element of v_i^* is given by

$$y_{it}^* = c_t[y_{it} - \frac{1}{T-t}(y_{it+1} + ... + y_{iT})]$$

with $c_t^2 = (T-t)/(T-t+1)$.
By multipliying the model by A, it becomes

$$y_{it}^* = \delta x_{it}^* + v_{it}^* \tag{2.2}$$

We have $E(x_{i,t}^* v_{it}^*) \neq 0$ so that OLS estimator of the transformed model is not consistent. However, $E(x_{i,t-s} v_{it}^*) = 0$ for $s = 0, ..., t-1$ and $t = 1, ..., T-1$. Then, we are interested in the LIML estimator of δ based on these moment conditions. There are $\bar{q} = T(T-1)/2$ moment conditions which can be very large even if T is moderately large. Let $z_{it}{=}(x_{i1}, ..., x_{it})'$ and Z_i be the $(T-1)$ $\times$ $\bar{q}$ block diagonal matrix whose t-th block is z_{it}'. The moment conditions are then given by $E(Z_i' v_i^*) = 0$. Under conditional homoscedasticity of v_{it}, the covariance matrix of the orthogonality conditions is $\sigma^2 E(Z_i' Z_i)$. If we define the $m = N(T-1)$ vectors $y^*{=}(y_1^{*\prime}, ..., y_N^{*\prime})'$ and $x^*{=}(x_1^{*\prime}, ..., x_N^{*\prime})'$ with $y_i^* = (y_{i1}^*, ..., y_{i,T-1}^*)'$ and $x_i^* = (x_{i1}^*, ..., x_{i,T-1}^*)'$, then the LIML estimator of the parameter is given by

$$\hat{\delta} = \underset{a}{\operatorname{argmin}} \frac{(y^* - ax^*)' Z (Z'Z)^{-1} Z' (y^* - ax^*)}{(y^* - ax^*)'(y^* - ax^*)} \tag{2.3}$$

where Z is the $m \times \bar{q}$ matrix define by $Z{=}(Z_1', ..., Z_N')'$. Let us define the matrix $M = Z(Z'Z)^{-1}Z'$

$$\hat{\delta} = \underset{a}{\operatorname{argmin}} \frac{(y^* - ax^*)' M (y^* - ax^*)}{(y^* - ax^*)'(y^* - ax^*)} \tag{2.4}$$

Alvarez and Arellano (2003) show that the LIML estimator can be rewritten in as :

$$\hat{\delta} = \frac{x^{*\prime} M y^{*} - \hat{l} x^{*\prime} y^{*}}{x^{*\prime} M x^{*} - \hat{l} x^{*\prime} x^{*}} \tag{2.5}$$

where $\hat{l}$ is the smallest eigenvalue of the matrix $[W^{*\prime} M W^{*}][W^{*\prime} W^{*}]^{-1}$ and $W^{*} = (y^{*}\ x^{*})$.

As mentioned in Alvarez and Arellano (2003), this estimator is actually a symmetrically normalized estimator of the kind considered by Alonso-Borrego and Arellano (1999). It is asymptotically equivalent to the GMM estimator for fixed T as $N \to \infty$. Moreover, this estimator belongs to the $k - class$ estimators.

The computation of the LIML estimator involves the inverse of the $\bar{q} \times \bar{q}$ matrix $Z'Z$. Even for moderate values of T, $\bar{q}$ can be very large so that the condition number of $Z'Z$ (the ratio of its largest over its smallest eigenvalue) is large. A large condition number indicates that the matrix is ill-conditioned and its inverse is highly unstable. Using an unstable matrix to derive the LIML estimator may amplify the potential sampling errors and lead to poor finite sample properties. We propose to improve the finite sample properties of the LIML estimator by the use of a regularized inverse of $Z'Z$ instead of its usual inverse $(Z'Z)^{-1}$. Regularization can be seen as a way to stabilize the inverse. Unlike moments selection procedures, regularization is a dimension reduction technique using all the available moment conditions. The regularized LIML estimator is presented in the next section.

2.2 The Regularized LIML estimator

The regularization methods used in this paper are drawn from the literature on inverse problems (Kress, 1999). They are designed to stabilize the inverse of Hilbert-Schmidt operators (operators for which the eigenvalues are square summable). The matrix $Z'Z$ is not Hilbert-Schmidt however Lemma 1 of Carrasco and Nayihouba (2019) show that $K = E[Z'Z/NT^{3/2}]$ is Hilbert-Schmidt. So the regularization will be applied to $K_N = Z'Z/NT^{3/2}$, the sample counterpart of K.

Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 ... \geq \hat{\lambda}_q \geq 0$ be the eigenvalues of K_N. By spectral decomposition, we have $K_N = P_N D_N P_N'$ with $P_N P_N' = I_{\bar{q}}$ where P_N is the matrix of eigenvectors and D_N the diagonal matrix with eigenvalues $\hat{\lambda}_j$ on the diagonal. Let K_N^α denote the regularized inverse of K_N which is defined as

$$K_N^\alpha = P_N D_N^\alpha P_N'$$

where D_N^α is the diagonal matrix with elements $q(\alpha, \hat{\lambda}_j^2)/\hat{\lambda}_j$.

The positive parameter α is the regularization parameter, a kind of smoothing parameter, and the real function $q(\alpha, \lambda^2)$ depends on the regularization scheme used. As in Carrasco (2012), three regularization schemes will be considered : Tikhonov, spectral cut-off and Landweber Fridman regularization schemes. More details on these schemes can be found in Carrasco et al. (2007). If we let λ be an arbitrary eigenvalue of the matrix K_N, we can define :

1. **Tikhonov regularization (TH) :**
   This regularization scheme is close to the well known ridge regression used in presence of multicolinearity to improve properties of OLS estimators. In Tikhonov regularization scheme, the real function $q(\alpha, \lambda^2)$ is given by

   $$q(\alpha, \lambda^2) = \frac{\lambda^2}{\lambda^2 + \alpha}.$$

2. **The spectral cut-off (SC)**
   It consists in selecting the eigenvectors associated with the eigenvalues greater than some threshold.

   $$q(\alpha, \lambda^2) = I\{\lambda^2 \geq \alpha\} = \begin{cases} 1 \text{ if } \lambda^2 \geq \alpha, \\ 0 \text{ otherwise.} \end{cases}$$

   Another version of this regularization scheme is Principal Components (PC) which consists in using a certain number of eigenvectors to compute the inverse of the operator. PC and SC are perfectly equivalent, only the definition of the regularization term α differs. In PC, α is the number of principal components. In practice, both methods will give the same estimator so that we will study the properties of SC in detail in this paper.

3. **Landweber Fridman regularization (LF)**
   In this regularization scheme, K_N^α is computed by an iterative procedure

with the formula

$$\left\{ \begin{array}{r} K_{N,l}^{\alpha} = (I - cK_N)K_{N,l-1}^{\alpha} + cK_N,\ l = 1, 2, ..., 1/\alpha - 1, \\ K_{N,0}^{\alpha} = cK_N \end{array} \right.$$

The constant c must satisfy $0 < c < 1/\lambda_1^2$ where λ_1^2 is the square of the largest eigenvalue of the matrix K_N. Alternatively, we can compute this regularized inverse with

$$q(\alpha, \lambda^2) = 1 - (1 - c\lambda^2)^{\frac{1}{\alpha}}$$

In each regularization scheme, the real valued function $q(\alpha, \lambda^2)$ satisfies $0 \leq q(\alpha, \lambda^2) \leq 1$ and $\lim_{\alpha \to 0} q(\alpha, \lambda^2) = 1$ so that the usual GMM estimator corresponds to a regularized estimator with $\alpha = 0$.
Remark that $M = Z\left(Z'Z\right)^{-1} Z' = ZK_N^{-1}Z'/NT^{3/2}$. Similarly, let us denote the matrix $M^{\alpha} = ZK_N^{\alpha}Z'/NT^{3/2}$. The regularized LIML estimator for a given regularization scheme is :

$$\widehat{\delta^{\alpha}} = \underset{a}{\operatorname{argmin}} \frac{(y^* - ax^*)'M^{\alpha}(y^* - ax^*)}{(y^* - ax^*)'(y^* - ax^*)} \tag{2.6}$$

or alternatively

$$\widehat{\delta^{\alpha}} = \frac{x^{*\prime}M^{\alpha}y^* - \hat{\wedge}x^{*\prime}y^*}{x^{*\prime}M^{\alpha}x^* - \hat{\wedge}x^{*\prime}x^*} \tag{2.7}$$

where $\hat{\wedge} = \wedge(\hat{\delta}^{\alpha})$ and

$$\wedge(\delta) = \frac{(y^* - \delta x^*)'M^{\alpha}(y^* - \delta x^*)}{(y^* - \delta x^*)'(y^* - \delta x^*)} \tag{2.8}$$

The matrix K_N is a block diagonal matrix with the $t \times t$ matrix $Z_t'Z_t/NT^{3/2}$ at the $t-$th block. Exactly as K_N^{-1}, the regularized inverse K_N^{α} is also a block diagonal matrix where each block is the regularized inverse of the corresponding block of K_N †. So, if we define $M_t^{\alpha} = Z_t(K_{Nt})^{\alpha}Z_t'/NT^{3/2}$ with $(K_{Nt})^{\alpha}$ being the $t - th$ block of the matrix K_N^{α}, the regularized estimator can be rewritten as :

$$\widehat{\delta^{\alpha}} = \left(\sum_{t=1}^{T-1} x_t^{*\prime}M_t^{\alpha}x_t^* - \hat{\wedge}\sum_{t=1}^{T-1} x_t^{*\prime}x_t\right)^{-1}\left(\sum_{t=1}^{T-1} x_t^{*\prime}M_t^{\alpha}y_t^* - \hat{\wedge}\sum_{t=1}^{T-1} x_t^{*\prime}y_t\right) \tag{2.9}$$

†. This holds because regularization transforms only the eigenvalues, not the eigenvectors.

2.3 Asymptotic properties of the regularized LIML

In this section, we derive the asymptotic properties of the regularized estimator. As in Alvarez and Arellano (2003), we make the following assumptions :

Assumption 1 : $\{v_{it}\}$ $(t = 1, ..., T\,;\, i = 1, ..., N)$ are *i.i.d.* across time and individuals and independent of η_i and y_{i0} with $E(v_{it}) = 0$, $var(v_{it}) = \sigma^2$, and $E(v_{it}^4) < \infty$.

Assumption 2 : The initial observation satisfies

$$y_{i0} = \frac{\eta_i}{1-\delta} + w_{i0} \qquad (i = 1, ..., N)$$

where w_{i0} is independent of η_i and *i.i.d.* with the steady state distribution of the homogeneous process, so that $w_{i0} = \sum_{j=0}^{\infty} \alpha^j v_{i(-j)}$.

Assumption 3 : η_i are *i.i.d.* across individuals with $E(\eta_i) = 0$, var(η_i)=σ_η^2 with $0 < \sigma_\eta^2 < \infty$, and finite fourth order moment.

Moreover, asymptotic properties are derived under the assumption that both N and T go to infinity with $T < N$. Under this restriction the matrix K_N is non singular and so has nonzero eigenvalues.

Let $K = E[Z'Z/NT^{3/2}]$ and $(\lambda_l,\, \phi_l,\, l = 1, 2, ...)$ be the eigenvalues and orthonormal eigenvectors of K. In the inverse problem literature, this matrix is referred to as the operator. In Carrasco (2012) and Carrasco and Tchuente (2015), the operator is assumed to be a trace-class operator which is satisfied if and only if its trace is finite. Working on the GMM estimator of the DPM, Carrasco and Nayihouba (2019) proved under the same assumptions that K is not trace class but it is Hilbert-Schmidt ($Tr\,(K) = O\left(T^{1/2}\right)$ and $Tr\,(K^2) = O\,(1)$). These results show that even though the eigenvalues of K are not summable as T goes to infinity, they are square summable. The Hilbert-Schmidt property is sufficient to derive proofs in our framework. This property is especially useful to establish the order of magnitude of the bias of the regularized estimator.

The following proposition provides the first order asymptotic properties of the regularized estimator.

Proposition 1.
Suppose that Assumptions 1–3 hold. Then as both N and T go to infinity and α the parameter of regularization goes to zero, provided that $\alpha\sqrt{NT} \to \infty$, we have
(i) Consistency : $\widehat{\delta}^{\alpha} \to \delta$ *in probability ;*
(ii) Asymptotic normality : $\sqrt{NT}\left(\widehat{\delta}^{\alpha} - \delta\right) \xrightarrow{d} N\left(0, 1-\delta^2\right).$

For these properties to hold, we need that α goes to zero slower than $\sqrt{NT}$ goes to infinity. Carrasco and Nayihouba (2019) proved the consistency and the asymptotic normality of the regularized GMM estimator under the same condition. Alvarez and Arellano (2003) proved, under similar assumptions, that the bias of the LIML estimator is of order $(1+\delta)/(2N-T)$ if $0 \leq \lim T/N \leq 2$, whereas the bias of the GMM estimator is of order $(1+\delta)/N$ provided that $0 \leq \lim T/N \leq \infty$. As long as $N > T$ the bias of the LIML estimator is smaller than the bias of the GMM estimator.
The finite sample bias of the regularized LIML estimator depends on the regularization parameter so that the selection of this parameter is of great importance. To gain a better understanding of the effect of α, we proceed to a higher-order expansion of the mean square error in the next section.

2.4 Mean square error for regularized LIML

In this section, we derive the Nagar's approximation of the mean square error (MSE) of $\hat{\delta}^{\alpha}$. This expansion is used in many papers on IV literature such as Carrasco (2012), Donald and Newey (2001) and particularly Okui (2009) who works on a dynamic panel data model. Moreover, this expansion will guide us in our goal to provide a data-driven method for selecting the regularization parameter. The Nagar approximation of the MSE is the term $\sigma^2 H^{-1} + S(\alpha)$ in the following decomposition :

$$NT(\hat{\delta}^{\alpha} - \delta)^2 = Q + r, \quad E(Q) = \sigma^2 H^{-1} + S(\alpha) + R \tag{2.10}$$

where $(r+R)/S(\alpha) \to 0$ as $N \to \infty$, $T \to \infty$, $\alpha \to 0$, and

$$H = \frac{\sigma^2}{1-\delta^2}\left(\frac{1}{T}\sum_{t=1}^{T}\psi_t^2\right)$$
$$\psi_t = c_t\left(1 - \frac{\delta\phi_{T-t}}{T-t}\right)$$
$$\phi_j = \frac{1-\delta^j}{1-\delta}.$$

Proposition 2.
Suppose assumptions 1-3 are satisfied and $E\left(v_{it}^3\right) = E\left(v_{it}^5\right) = 0$*. If* $N \to \infty$*,* $T \to \infty$*,* $\alpha \to 0$*,* $\alpha\sqrt{NT} \to \infty$*, then for the regularized LIML estimator, the decomposition given in (2.10) holds with :*

$$S(\alpha) = \frac{(1-\delta^2)^2}{\sigma^4}\Bigg\{\frac{\sigma^4}{NT}\sum_{t=1}^{T-1} Etr[(M_t^\alpha)^2]\left[\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)^2\right]$$
$$+ \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t^\alpha)^2 w_{t-1}]\Bigg\}$$

where $w_{it} = y_{it} - \eta_i/(1-\delta)$*.*

The first term of $S(\alpha)$ increases when α goes to zero whereas the second term decreases when α goes to zero. Hence, a large α will reduce the first term and increase the second one. Proposition 2 shows that the MSE dominant term of the regularized LIML estimator, $S(\alpha)$, comes from two variance terms unlike for the regularized GMM estimator of the DPM. Indeed, under similar assumptions, Carrasco and Nayihouba (2019) derived the following expression of $S(\alpha)$ for the regularized GMM estimator :

$$S_{GMM}(\alpha) = \frac{(1-\delta^2)^2}{\sigma^4}\Bigg\{\left[\frac{\sigma^2}{1-\delta}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} Etr(M_t^\alpha)\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]^2$$
$$+ \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t^\alpha)^2 w_{t-1}]\Bigg\}$$

The first term of $S_{GMM}(\alpha)$ comes from the square of the bias whereas the second term comes from the second-order expansion of the variance. Since

$S(\alpha)$ and $S_{GMM}(\alpha)$ have the same second terms, the difference between them comes from their the first terms. The first term of $S(\alpha)$ is of order $O(1/\alpha NT)$ whereas that of $S_{GMM}(\alpha)$ is of order $O(1/\alpha^2 NT)$. Hence, the approximate MSE of regularized LIML is of smaller order than that of the regularized GMM estimator.

If we define $\tilde{v}_{itT} = (\phi_{T-t}v_{it}+, ..., +\phi_1 v_{i,T-1})/(T-t)$, $u_{it} = -c_t\tilde{v}_{itT}$, $\rho_t = E(u_{it}v^*_{it})/\sigma^2$, $\epsilon_{it} = u_{it} - \rho_t v^*_{it}$, then

$$S(\alpha) = \frac{(1-\delta^2)^2}{\sigma^4}\left\{\frac{1}{NT}\sum_{t=1}^{T-1} Etr[(M_t^\alpha)^2]\sigma^2 E(\epsilon_{it}^2) + \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t^\alpha)^2 w_{t-1}]\right\}$$

which is similar to the expression of $S(\alpha)$ derived in Carrasco and Tchuente (2015) for the regularized LIML estimator in a cross-sectional setting. The ϵ_{it}, u_{it}, v^*_{it}, $\psi_t w_{it}$ in our notation correspond respectively to v_i, u_i, ϵ_i,f_i in their notation. Moreover, our matrix M^α is denoted by P^α in their paper. Because $E[v_i\epsilon_i] = 0$ in the cross-sectional setting, $S(\alpha)$ comes from two variance terms. We obtain the same result as we have $E(\epsilon_{it}v^*_{it}) = 0$. However, as in Carrasco and Nayihouba (2019), our expression of $S(\alpha)$ is unconditionnal unlike in Carrasco (2012) and Carrasco and Tchuente (2015).

2.5 Data-driven selection of the regularization parameter

2.5.1 Estimation of the approximate MSE

In this section, we show how to select the regularization parameter α in order to minimize the MSE. Let us introduce the following notations

$$\mathcal{A}(\alpha) = \sum_{t=1}^{T-1}\left[\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)^2\right]E[tr(M_t^\alpha)^2]$$

and

$$R(\alpha) = \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t^\alpha)^2 w_{t-1}]$$

so that,

$$S(\alpha) = \frac{(1-\delta^2)^2}{NT}\mathcal{A}(\alpha) + \frac{(1-\delta^2)^2}{\sigma^2}R(\alpha).$$

Let $\hat{\delta}$ and $\hat{\sigma}^2$ be consistent estimators of δ and σ^2, respectively. Then $S(\alpha)$ can be estimated by

$$\widehat{S}(\alpha) = \frac{(1-\hat{\delta}^2)^2}{NT}\widehat{\mathcal{A}}(\alpha) + \frac{(1-\hat{\delta}^2)^2}{\hat{\sigma}^2}\widehat{R}(\alpha)$$

with

$$\widehat{\mathcal{A}}(\alpha) = \sum_{t=1}^{T-1} tr[(M_t^\alpha)^2]\left[\frac{[\widehat{\phi}_{T-t}^2 + ... + \widehat{\phi}_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\hat{\delta})^2}\left(\frac{\widehat{\phi}_{T-t}}{T-t} - \frac{\widehat{\phi}_{T-t+1}}{T-t+1}\right)^2\right]$$

where

$$\widehat{\phi}_j = \frac{1-\widehat{\delta}^j}{1-\widehat{\delta}}$$

and

$$\widehat{R}(\alpha) = \frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*'}(I - M_t^\alpha)^2 x_t^*.$$

Note that from Okui (2009, p.3),

$$E_{t-1}\left(x_{it}^*\right) = \psi_t\left(y_{it} - \frac{\eta_i}{1-\delta}\right) = \psi_t w_{it-1}$$

where E_{t-1} denotes the conditional expectation conditional on $(\eta_i, x_{it}, x_{it-1}, ...)$ so that $\widehat{R}(\alpha)$ is an unbiased estimator of $R(\alpha)$. The optimal regularization parameter is selected by minimizing this estimated $S(\alpha)$

$$\hat{\alpha} = \arg\min_{\alpha \in \mathcal{E}_T} \widehat{S}(\alpha)$$

where $\mathcal{E}_T$ is the index set of α. $\mathcal{E}_T$ is a real compact subset for TH, $\mathcal{E}_T$ is such that $\frac{1}{\alpha} \in \{1, 2, ..., \bar{q}\}$ for PC, and $\mathcal{E}_T$ is such that $\frac{1}{\alpha}$ is a positive integer for Landweber Fridman. Next, we analyse the impact of using an estimated version of $S(\alpha)$ to select α instead of the true and unknown criterion.

2.5.2 Optimality

We wish to establish the optimality of the regularization parameter selection criterion in the following sense (see Li, 1986, 1987)

$$\frac{S(\hat{\alpha})}{\inf_{\alpha \in \mathcal{E}_T} S(\alpha)} \xrightarrow{p} 1 \qquad (2.11)$$

as $N \to \infty$, $T \to \infty$. The result (2.11) etablishes that using an estimated version of $S(\alpha)$ to choose the regularization parameter is asymptotically equivalent to use the true and unknown value of $S(\alpha)$. Hence, it is not a convergence result of $\hat{\alpha}$.

Proposition 3.
Suppose that Assumptions 1-3 are satisfied and $\hat{\delta} \to \delta$, $\hat{\sigma}^2 \to \sigma^2$. If $N \to \infty$, $T \to \infty$ and $\alpha\sqrt{NT} \to \infty$, $T^3/N\,(\ln T)^2 \to 0$, then the regularization parameter selection criterion is asymptotically optimal in the sense of (2.11) for Spectral cut-off and Landweber Fridman regularization schemes provided that $\#\mathcal{E}_T = O(T^2)$ where $\#\mathcal{E}_T$ refers to the number of elements in the set $\mathcal{E}_T$.

Carrasco and Tchuente (2015) apply regularization in cross-sectional data and use Li (1986, 1987) to establish the optimality of their selection rule. From $x^*_{it} = \psi_t w_{it} - c_t \tilde{v}_{it}$, the term $-c_t \tilde{v}_{it}$ can be seen as the error of the second stage equation since $\psi_t w_{it} = E_{t-1}(x^*_{it})$. Then, Li (1986, 1987)'s results do not work in our framework because of the autocorrelation of this error term. Consequently, our proof combines the strategies of Kuersteiner (2012) and Okui (2009).
Proposition 3 proves optimality for Spectral cut-off and Landweber Fridman regularization schemes which have discrete index set $\mathcal{E}_T$. The condition $\#\mathcal{E}_T = O(T^2)$ is a sufficient condition in the Landweber Fridman regularization scheme and it automatically holds for the principal components case.[‡] Rather than imposing a maximum number of iterations, this condition restricts the order of magnitude of the number of elements of the index set $\mathcal{E}_T$. A rigorous proof for the Tikhonov's continous index set requires more complicated material which is beyond the scope of this work. However, optimality could be established by approximating the continous index set by a discrete one.

2.6 Simulation study

In this section, we present Monte Carlo simulations to illustrate the performance of the regularized LIML estimator and compare them to other estimators such as the usual not regularized GMM and LIML estimators, the regularized GMM estimator and the one presented in Okui (2009). In

‡. Recall that $\#\mathcal{E}_T = \bar{q}$ for principal components case.

our simulations, we consider the following autoregressive model :

$$y_{it} = \delta y_{i,t-1} + \eta_i + v_{it},$$

with $\eta_i \sim iidN(0, \sigma^2_\eta)$, $y_{i0} \sim iidN(\eta_i/(1-\delta), \sigma^2/(1-\delta^2))$ and $v_{it} \sim iidN(0, \sigma^2)$. Each simulation corresponds to a choice of vector $(N, T, \sigma^2, \sigma^2_\eta, \delta)$. We consider $N = 50$ and $N = 100$. For each values of N, we simulated for $T = 10$, $T = 25$ for three values of δ=(0.5,0.75,0.95). The number of replications is 5000 for all cases. Nine estimators of the parameter of interest are presented. We denote by GMM, the GMM estimator using all available lags of y_{it} as instruments. IVK is the estimator when the instruments are selected by the selection procedure proposed by Okui (2009). Finally the regularized LIML and GMM estimators are denoted as TGMM and TLIML if the Tikhonov regularization scheme is used, PGMM and PLIML for principal components which is a variant of spectral cut-off that has an easy interpretation and LGMM, LLIML for Landweber-Friedman regularization scheme.
In order to select the regularization parameter α, we minimized the estimated version of $S(\alpha)$ given in the previous section. As a consistent estimator of δ, we used the GMM estimator for TGMM, PGMM, LGMM and the usual LIML estimator for TLIML, PLIML and LLIML. The variance estimates $\widehat{\sigma^2}$ is given by :

$$\widehat{\sigma^2} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T-1} (y^*_{it} - \hat{\delta} x^*_{it})^2$$

For each estimator, we compute the median bias (Med.bias), the median absolute bias (Med.abs), the length of the inter quantile range (Iqr.), the median mean square error (Med.mse), and the coverage probabilities (Cov) of the 95 % confidence intervals. The estimate of the covariance matrix is computed with the formula :

$$V(\bar{\widehat{\delta^\alpha}}) = \widehat{\sigma^2}[x^{*\prime}(M^\alpha - \hat{\wedge} I_m)x^*]^{-1}[x^{*\prime}(M^\alpha - \hat{\wedge} I_m)^2][x^{*\prime}(M^\alpha - \hat{\wedge} I_m)x^*]^{-1}$$

where I_m is the $m \times m$ identity matrix.
Tables B.1-B.4 in appendix present the simulations results for different combinations of N, T, δ, σ^2 and σ^2_η. For each simulations setting, we provide a table of the properties of the optimal regularization parameter.
Table B.2 contains summary statistics for the value of the regularization parameter which minimizes the approximate MSE for $N = 50, \sigma^2 = 1, \sigma^2_\eta = 1$ and different values of δ and T. We report the mean, the standard deviation

(std), the mode and the three quartiles of the distribution of the regularization parameter. The regularization parameter is the optimal α for TH, the optimal number of iterations for LF and the optimal number of principal components for PC. Noting that the standard GMM and LIML estimators (not regularized) correspond to the Tikhonov estimator with $\alpha = 0$, to the LF estimator with an infinite number of iterations and to the PC using all the principal components, a higher level of regularization will correspond to a larger value of α, a smaller number of principal components and a smaller number of iterations for the LF estimator.
The bias reduction of regularization on the LIML estimator is higher when $T = 25$ for a given δ and for large δ when T is fixed. So, the optimal proportion of PC selected decreases when δ increases for a given T and increases with T as δ increases. For TH, the optimal regularization parameter is higher in $\delta = 0.95$ and $T = 25$ settings. The optimal number of iterations of LF is smaller in those latter settings.
The simulations results for $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ are presented in Table B.1 in appendix. The parameter c for LF is set to $0.95/\widehat{\lambda}_1^2$ (where $\widehat{\lambda_1}$ is the largest eigenvalue of K_N.)
The usual (not regularized LIML estimator has better properties than the usual GMM estimator except when δ is close to unity. When $\delta = 0.95$, the usual GMM dominates the LIML estimator in terms of bias and MSE.
Regularisation improves the small sample properties of the LIML and the GMM estimators as the regularized estimators have smaller bias, smaller MSE and higher coverage probabilities than the not regularized estimators.
The IVK dominates the usual GMM for each value of T and δ with respect to all criterion. The usual LIML has smaller bias than the IVK when $T = 10$ for $\delta = 0.50$ and $\delta = 0.75$. For these values of δ, regularized LIML estimators TLIML and LLIML have smaller bias, smaller MSE that IVK. For $T = 25$ and $\delta \neq 0.95$, IVK dominates the LIML estimator. However, regularized estimators TLIML and LLIML have smaller bias, and higher coverage rates than IVK.
Table 1 also shows that the LIML estimaot has very poor properties when $\delta = 0.95$. Even tough regularisation improves the properties of the LIML estimators for this value of δ, the regularized LIML estimators are dominated by all the others estimators in terms of bias, MSE and coverage probabilities.
Comparison between regularized LIML estimators shows that TLIML and

LLIML have the best coverage rates. In terms of bias and MSE, PLIML dominates for $T = 10$ whereas the two other dominate for $T = 25$.
Tables B.3 and B.4 show the simulations results when $N = 100$. In Table B.3 all the estimators have better properties as the number of observations increased in comparison to the previous tables. Regularization also provides better improvements with an important increase of coverage probabilities as well as an important reduction of the MSE. In this setting, the LIML estimator is also dominated by the GMM and the IVK for $\delta = 0.95$ but when $T = 25$, regularized estimators LIML estimators TLIML and PLIML have the smallest bias, the smallest mean square error and the highest coverage rate.
It should be noted that the simulations here are not directly comparable with those in Tables 2 to 9 in Carrasco and Nayihouba (2019) because there the regularized GMM estimators were bias-corrected.

2.7 Conclusion

Altough the LIML estimator is proved to have better properties than the GMM estimator, it also exhibits a bias in presence of the many instruments in the DPM. To improve the properties of the LIML estimator of the DPM, we propose a regularization approach based of the way of computing the inverse of the covariance matrix of the instruments. This technique depends neither on the ranking of the instruments nor on the dimension of the covariance matrix.
We prove the consistency and the asymptotic normality of the regularized estimators. We also derive a data driven way to select the regularization parameter induced by the regularization procedure. Our simulations results show that regularization improves the properties of the classical LIML estimator of DPM. This domination of regularized estimators over classical LIML estimator is higher when T is large (where the classical LIML estimator suffers from many instrument bias) and when the autoregressive parameter is close to unity (presence of weak instrument bias). Moreover, the simulations show that the regularized LIML perform better than regularized GMM for small δ but not for large δ.
In this paper, we introduced the regularized LIML estimator in a very simple version of the DPM : *iid* error terms and only the lag of the dependent variable as covariate. In order to be more realistic and useful for applied re-

search, the model should be extended to include exogenous or predetermined covariates and also relax the *iid* assumption on the error term.

Chapitre 3

A regularization approach to estimating labor elasticities

The estimation of labor supply elasticities rises endogeneity issues because of measurement errors on wages and income variables, omitted bias and unobserved heterogeneity. To take into account these issues, several estimation techniques are used in the literature including instrumental variables estimation, Maximum Likelihood Estimator (MLE) especially in presence of progressive taxation.
A popular solution to the endogeneity problem in the labor supply literature is grouping estimators. This technique consists in dividing the sample into groups and then carrying out the estimation at the group level. The idea behind grouping data and estimating at the groups level is that measurement errors as well as the effect of unobserved heterogeneity are assumed to cancel out as the group size gets larger (Angrist, 1991). A necessary condition for this assumption to hold is to have homogenous groups (so a large number of groups) as well as large number of observations within groups. An advantage of grouping estimators is that they don't rely on the exclusion assumption required for the traditional instrumental variables estimation and that can be hard to prove sometimes. Moreover, unlike instrumental variables estimators, grouping estimators can be computed when only aggregated data are available.

The simplest and most widely used grouping estimator is the EWALD (Efficient Wald) estimator which is the weighted least squares of group means. Angrist (1991) proved that the EWALD is equivalent to the two stage least squares (2SLS) estimator of the model at the individual level using groups dummies as instruments. Hence, in presence of large number of groups, the EWALD estimator suffers from finite sample bias that arises when estimating instrumental variables estimators in presence of many instruments (see Angrist and Krueger (1991)).
In this paper, we estimate the labor supply elasticity using group dummies as instruments and use a regularization approach to handle the many instruments problem. This approach proposed by Carrasco (2012) is based on the way to invert the covariance matrix of the instruments. Indeed, in presence of many instruments the covariance matrix of the instruments is of large dimension and inverting it to compute the instrumental variables estimator can lead to inaccurate results. Following Carrasco (2012), we use three regularization schemes : Tikhonov (TH), Principal components (PC) and Landweber Fridman (LF). Each regularization scheme introduces a regularization parameter which is similar to the smoothing parameter on the nonparametric regression.
The originality of the regularization approach is that there is no restriction on the number of instruments which can be very large or even infinite. All the available instruments are used in the estimation. Details on the theorical properties as well as simulations results on the performance of the regularized estimators can be found in Carrasco (2012) for the 2SLS estimator, in Carrasco and Tchuente (2015) for the regularized LIML (Limited Information Maximum Likelihood), Carrasco and Doukali (2016) for the regularized Jackknife, Carrasco and Nayihouba (2019a) for the regularized GMM estimator for dynamic panel data models, Carrasco and Nayihouba (2019b) for the regularized LIML for dynamic panel data models.
Grouping estimators are very popular in labor supply literature. Dostie and Krogman (2012) used survey data to estimate the labor elasticities of Canadian women. Their grouping estimator is based on 48 groups defined by the interaction of husbands' age (3 categories), combined education (4 categories), and geographic location (4 categories). Morissette and Hou (2008) used population census data and then have larger number of groups. They focus on the link between Canadian women labor supply and their husband wages. They grouped the data in 300 groups based on the interaction of province indicators, three husbands' age categories (25-34 ; 35-44 ; 45-54), and spouses'

educational attainment (10 categories). The same categories are used in Blau and Kahn (2007). Devereux (2007a) constructed 1296 categories defined by the interaction of 12 husband types, 12 wife types, and 9 census regions to analyse the link between changes in relative wages and family labor supply in the US. Husband's and wife's types are defined on their education level and age. Finally, Devereux (2007b) estimates the intertemporal labor supply elasticity of men using grouping estimators on 432 groups. The categories are defined using birth cohort and census year.
Devereux (2007b) first linked the grouping estimator literature and the instrumental variables literature by proving the correspondance of grouping estimators to instrumental variables estimators. We somewhat extend the connexion between the two literatures by applying regularization, a procedure developed in the instrumental variables literature, as a solution to the finite sample properties of the grouping estimator. Following Devereux (2007b) the groups are based on the cohort of birth, the province and the census year. Moreover, population census data are used to take advantage of the large sample size necessary to identification assumption of the grouping estimator.
Using the regularization approach leads to wage elasticities that range from 0.10 to 0.12 whereas grouping estimator leads to an elasticity of 0.08. These results suggest that the grouping estimator tends to under estimate the wage elasticity.
The next section of the paper presents the labor supply model estimated. The data set used for estimation is presented in section 3.2. A simulation study that compares the grouping estimator and the regularized estimators is presented in section 3.3 whereas the estimation results are given in section 3.4.

3.1 The model

We consider the model used in Devereux (2007b) and widely used in labor supply studies using microdata. For an individual i, we have

$$Y_i = \beta X_i + \epsilon_i = \beta_0 + \beta_1 Logw_i + \beta_2 I_i + \beta_3 X_{2i} + \epsilon_i \tag{3.1}$$

$i = 1, ..., n$ where Y_i is the hours worked, $Logw_i$ is the log of the wage rate, I_i is the non labor income, X_{2i} is a vector of control variables and ϵ_i an error term.

The OLS estimator of equation (3.1) is biased because of measurement error on wage rate and non labor income variables as well as unobserved factors that are correlated to wage and hours worked. A popular alternative to the OLS estimator is to group the data and compute weighted least squares at the group level. This grouping estimator, referred to as the efficient Wald estimator (EWALD) is given by

$$\beta^{EWALD} = \left(\sum_{g=1}^{G} n_g \bar{x}_g \bar{x}_g'\right)^{-1} \left(\sum_{g=1}^{G} n_g \bar{x}_g \bar{y}_g\right)^{-1} \tag{3.2}$$

where G is the number of groups and n_g is the number of observations in group g.
Even though this estimator is consistent as n (the total sample size) goes to infinity, it is biased in finite sample.
Angrist (1991) proved that this grouping estimator corresponds to the 2SLS estimator of the model estimated at the individual level where groups dummies are used as instruments. If we denote by Z the instrument matrix, the 2SLS estimator is

$$\beta^{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) \tag{3.3}$$

where $Y = (Y_1, ..., Y_n)$ and $X = (X_1, ..., X_n)$.
When G the number of groups is very large, the matrix $K_n = Z'Z/n$ is large dimensional and inverting it impacts negatively the properties of the estimator. This problem arises because of very small eigenvalues which once inversed lead to very high values.
From the eigenvalue–eigenvector decomposition we can write

$$K_n = P_n' D_n P_n$$

with P_n and D_n are real matrix such that $P_n' P_n = I_G$ the $G \times G$ identity matrix and D_n the diagonal matrix of the eigenvalues of K_n. Given this decomposition, the inverse of K_n is

$$K_n^{-1} = P_n' D_n^{-1} P_n$$

with the diagonal elements of D_n^{-1} are the inverse of the eigenvalues of K_n.
When the eigenvalues of the K_n are small, inverting them leads to very large values that can negatively impact the properties of the estimators. For a given

eigenvalue λ of K_n, the regularization approach smoothes the procedure of inversion by the use of $q(\alpha, \lambda^2)/\lambda$ as diagonal elements of D_n^{-1} instead of $1/\lambda$. The real value function $q(\alpha, \lambda^2)$ depends on the regularization scheme and α is a smoothing parameter also referred to as the regularization parameter. Put differently, instead of the classical inverse $K_n^{-1} = P' D_n^{-1} P$, regularization approach proposes to use the regularized inverse defined by

$$K_n^{\alpha} = P_n' D_n^{\alpha} P_n$$

where D_n^{α} is the diagonal matrix of $q(\alpha^2, \lambda)/\lambda$ instead of $1/\lambda$ for each eigenvalue of K_n.*

As in Carrasco (2012), we consider three regularization schemes that differs from each other by the definition of the function $q(\alpha, \lambda)$.

1. **Tikhonov regularization**
   This regularization scheme is close to the well known ridge regression used in presence of multicolinearity to improve properties of OLS estimator. In Tikhonov regularization scheme, the real function $q(\alpha, \lambda)$ is given by

$$q(\alpha, \lambda) = \frac{\lambda}{\lambda + \alpha}.$$

2. **The spectral Cut-off**
   It consists in selecting the eigenvectors associated with the eigenvalues greater than some threshold.

$$q(\alpha, \lambda) = I\{|\lambda| \geq \alpha\} = \begin{cases} 1 \; if \; |\lambda| \geq \alpha, \\ 0 \; otherwise. \end{cases}$$

   An other variante of this regularization scheme is Principal Components which consists to use a certain number of eigenvalues to compute the inverse of the diagonal matrix D_n.

3. **Landweber Fridman regularization**
   In this regularization scheme, K_n^{α} is computed with

$$q(\alpha, \lambda) = \frac{1 - (1 - c\lambda)^{\frac{1}{\alpha}}}{\lambda}.$$

*. To put simply we have $D_n = diag[\lambda_1, \lambda_2, ..., \lambda_{\bar{q}}]$, $D_n^{-1} = diag[1/\lambda_1, 1/\lambda_2, ..., 1/\lambda_{\bar{q}}]$ and $D_n^{\alpha} = diag[q(\alpha, \lambda_1^2)/\lambda_1, q(\alpha, \lambda_2^2)/\lambda_2, ..., q(\alpha, \lambda_{\bar{q}}^2)/\lambda_{\bar{q}}]$.

The constant c must satisfy $0 < c < 1/\|K_n\|^2$ where $\|K_n\|$ is the largest eigenvalue of the matrix K_n. Alternatively, K_n^α can be computed by iteration from

$$\begin{cases} K_{n,l}^\alpha = (I - cK_n)K_{n,l-1}^\alpha + cK_n, l = 1, 2, 3.........., 1/\alpha - 1, \\ \hfill K_{n,0}^\alpha = cK_n \end{cases}$$

More details on these schemes can be found in Carrasco et al.(2007) . In each regularization scheme, the real values function $q(\alpha, \lambda)$ satisfies $0 \leq q(\alpha, \lambda) \leq 1$ and $\lim_{\alpha \to 0} q(\alpha, \lambda) = 1$. The previous inequality says that for a given eigenvalue, its regularized inverse is smaller than its classical inverse $(1/\lambda)$. Moreover, when $\alpha = 0$ both regularized and classical inverse are equivalent.
Using the regularized inverse K_n^α instead of the classical inverse $(Z'Z)^{-1}$, the regularized estimator is

$$\widehat{\beta^\alpha} = (X'ZK_n^\alpha Z'X)^{-1}(X'ZK_n^\alpha Z'Y) \tag{3.4}$$

Carrasco (2012) proved that the regularized 2SLS is consistent as the number of observations n goes to infinity and $\alpha \to 0$ provided that $n\alpha^{\frac{1}{2}} \to \infty$. Moreover, as $n \to \infty$ the regularized estimator is asymptotically normal if $n\alpha \to \infty$.
The finite sample properties of the regularized estimators depends on the regularization parameter α. Carrasco (2012) provides a data-driven procedure to select the regularization parameter for the regularized 2SLS estimator. This procedure is done in two stages : a criteria to minimize is firstly derived through a second order expansion of the mean square error of estimator. Then the optimal regularization paramater is selected to minimize an estimated version of the criterion where unknown parameters are replaced by their consistent estimates. This two stages selection procedure is proved to be optimal.
The estimated version of the minimisation criterion is given :

$$\widehat{S(\alpha)} = \hat{\sigma}_{u\epsilon}\frac{\left(tr[P_n^\alpha]\right)^2}{n} + \hat{\sigma}_\epsilon^2\widehat{R(\alpha)} \tag{3.5}$$

where

$$\begin{aligned}P_n^\alpha &= Z'K_n^\alpha Z/n \\ \hat{\sigma}_\epsilon^2 &= (Y - X\hat{\beta})'(Y - X\hat{\beta})/n \\ \hat{\sigma}_{u\epsilon} &= e'X'(Y - X\hat{\beta})/n\end{aligned}$$

with $\hat{\beta}$ a preliminary consistent estimator of β, e a column vector of ones, I_n is the identity matrix and $\widehat{R(\alpha)}$ a criteria of goodness of fit. In this paper, we use the generalized cross validation criteria (Craven and Wahba,1978) given by :

$$\widehat{R(\alpha)} = \frac{1}{n}\frac{e'X'(I_n - P_n^\alpha)(I_n - P_n^\alpha)Xe}{\left(1 - \frac{tr[P_n^\alpha]}{n}\right)^2}$$

The $\widehat{S(\alpha)}$ has a bias related term (the one with $tr[P_n^\alpha]$) which decreases as α increases and a variance related term which increases with α. The optimal regularized parameter is chosen over an optimisation set E_n that depends on the regularization scheme. E_n is a compact subset of $[0, 1]$ for TH, E_n is such $1/\alpha \in \{1, 2, ..., G\}$ and E_n is such $1/\alpha$ is positive integer for LF.

3.2 The data

This paper uses data from Canadian population census of 2011, 2001, 1991 and 1981. Census data take advantage over surveys data sets on grouping estimators as they provide large sample size within the groups. The sample includes Canadian men engaged in salary work aged 45 to 60 in 2011, 35 to 50 in 2001, 25 to 40 in 1991 and 25 to 30 in 1981. Hours of work are measured by the annual hours worked. Wage rate is measured by the hourly wage, calculated by dividing the annual earned income by the total hours worked during the year. We compute the individual's non labor income by subtracting his earned income from his total income. Both total and earned incomes are directly provided in the data. Hence, the individual's non labor income includes income sources such as rents, interest, dividends and his welfare income. This later variable represents all the transfers received by a person during the previous year from government. The final sample excludes individuals who did not work the last year, individuals with wage rate less than 2 $ or greater than 200 $ (in 2010 $ CAD). Descriptive statistics for the variables used in the estimation are presented in Table C.1 in appendix. Following Devereux (2007b), we define a cohort birth by year and province of birth. There are 540 groups for the four census years. The number of observations within the groups varies from 13 to 1910 and averages to 406. These groups sizes are comparable to many of those existing in grouping estimators applied to labor supply literature. For example, the average group size is 129 in Dostie and Krogman (2012), 678 in Morisette and Hou (2008), 843 in Devereux (2004). Theoretically at least two observations are needed in each group in order to be able to derive the variance of the grouping estimator.

3.3 Simulation study

In this section we compare the grouping estimator EWALD to the regularized estimators using a monte carlo simulation. We use the same simulation design as in Devereux (2007b). For an individual i, the hours of work H, the log of the wage rate w and the non labor income y are generated by

$$
\begin{aligned}
H_{ict} &= H_{ct} + z_{ict} \\
w_{ict} &= w_{ct} + v_{ict} \\
y_{ict} &= y_{ct} + e_{ict}
\end{aligned}
$$

The z, v and e are drawn from a trivariate normal with the covariance matrix estimated by the within-groups variances and covariances in the data. If we denote by $\bar{w}$ and $\bar{y}$ the means of w and y respectively over the whole sample, then w_{ct} and y_{ct} are given by

$$
\begin{aligned}
w_{ct} &= 0.75\bar{w}_{ct} + 0.25\bar{w} \\
y_{ct} &= 0.75\bar{y}_{ct} + 0.25\bar{y}
\end{aligned}
$$

where $\bar{w}_{ct}$ and $\bar{y}_{ct}$ are the means of w and y for cohort c and year t. The dependent variable H_{ct} is generated as follow :

$$H_{ct} = \alpha + \beta w_{ct} + \delta y_{ct} + f_c + f_t$$

where f_c represents cohort fixed effect and f_t the year fixed effect.
The parameters α, β, δ, f_c and f_t are estimated by the EWALD estimates

of H on w and y. For computational purpose, this simulation study is based on the 2011, 2001 and 1991 census data. Hence, a total sample size of $n = 200000$ and 480 groups (160 cohorts and 3 years) are considered. The groups sizes in the generated data are set be equal to those in the actual data.

Table 3.1 presents the simulations results for the coefficient β (the coefficient of the wage variable) after 1000 replications for the EWALD and the three regularized estimators presented in section 3.1. The TH column refers to the regularized estimator using the Tikhonov regularization scheme, PC is the one using principal components and LF is the estimator using Landweber Fridman regularized scheme. For each estimator, we report the median bias (Med.bias), the mean absolute bias (Med.abs), the inter quartile range (Iqr), the median mean square error (Med.mse) and the 95 % coverage rate (Cov.rate). The variance of the regularized estimators is calculated with the same formula as in Carrasco (2012) :

$$\widehat{V\{\hat{\beta}^{\alpha}\}} = \frac{(Y - X\hat{\beta}^{\alpha})(Y - X\hat{\beta}^{\alpha})'}{n}(\widehat{X}'X)^{-1}(\widehat{X}'\widehat{X})(X'\widehat{X})^{-1} \tag{3.6}$$

where $\widehat{X} = P_n^{\alpha} X$.
Conventional instrumental variables standard errors are used to calculate the coverage rates of the grouping estimator. This is equivalent to using formula (3.6) with $\alpha = 0$. For the regularized estimators, the regularization parameter is selected to minimize the criterion given in (3.5). The EWALD estimates are used to estimate the unknown parameters. The optimal regularization parameter for TH is searched between 0.005 and 0.9 with 0.001 increment. The range of the values of the number of iterations for LF is from 100 to 9000 and the number of principal components is selected between 1 and the numbers of instruments.

	EWALD	**TH**	**PC**	**LF**
Med.bias	-0.0350	-0.0182	-0.0232	-0.0117
Med.abs	0.0353	0.0210	0.0247	0.0214
Iqr	0.0218	0.0238	0.0241	0.0316
Med.mse	0.0012	0.0004	0.0005	0.0003
Cov	0.4210	0.8150	0.9420	0.9320

TABLE 3.1 – Simulations for 1000 replications.

Table 3.1 results show that all the regularized estimators have better properties than the grouping estimator EWALD. Indeed, the regularized estimators have smaller bias, smaller mean square error and larger coverage rate than the EWALD estimator.

3.4 Estimation results

Table 3.2 presents the coefficient of Log of wage variable obtained from several estimators. In addition to the wage variable, all the estimated models include additional covariates which are the non labor income and cohort and year dummies.

The first column presents the results of the ordinary least squares (OLS) estimator of the model at the individual level. This estimator suggests a negative wage elasticity of 0.12. This estimator does not account for the several sources of endogeneity coming from the measurement errors on wage (computed by dividing the annual income by the annual hours of work) as well as the effect of unobserved heterogeneity driving simultaneously the labor supply and the individual's wage.

Column 2 presents the results of the grouping estimator. This estimator is the 2SLS estimator on the individual level data using the groups dummies as instruments. In the grouping literature, EWALD estimator corresponds to the weighted least squares on the groups mean variables. The results of EWALD suggest a wage elasticity of 0.08.

The regularized estimators are presented in the last three columns. The TH column refers to the regularized estimator using the Tikhonov regularization

scheme, PC is the one using principal components and LF is the estimator using Landweber Fridman regularized scheme. All these regularized estimators induce a regularization parameter which is the optimal α for TH, the optimal number of principal components for PC and the optimal number of iterations for LF. The optimal regularization parameter is selected to minimize the criterion given in (3.5).
Following Table 3.2, the TH estimator is obtained at a value of 0.016 for the regularization parameter, for PC 375 out of 540 principal components are selected whereas 6000 iterations are selected to estimate LF. In order to have an idea of how big or small these parameters are, one should keep in mind that when the regularization parameter is equal to zero, the TH is exactly the EWALD estimator or the 2SLS estimator in the instrumental variables literature. Using a second order expansion of the MSE, Carrasco (2012) proved that the bias of the regularized estimator TH is of order $1/\alpha\sqrt{n}$ whereas its variance is of order $\alpha^{min(2,\beta)}$ where $\beta \geq 1/2$. Hence, as the parameter of regularization increases the bias of the regularized estimator decreases whereas its variance increases. Similarly, the EWALD estimator corresponds to the PC where all the principal components are used. It also corresponds to the LF with the number of the iterations tends to infinity. So as the number of principal components (in the case of PC) and the number iterations (in the case of LF) decreases, the bias of regularized estimators decreases and the variance increases.

	OLS	**EWALD**	**TH**	**PC**	**LF**
Ln(wage)	-0.120	0.083	0.104	0.120	0.110
	(0.002)	(0.012)	(0.013)	(0.014)	(0.013)
α^*			0.016	375	6000
Observations	219316	219316	219316	219316	219316

Table 3.2 – Wage elasticities from OLS, EWALD, TH, PC and LF.

The wage elasticities provided by the regularized estimators are larger than those of the grouping estimator. This results is consistent with those in Table 3.1 which shows that the bias of EWALD is negative and larger than those

of the regularized estimators. In comparison with the EWALD estimator, the regularized estimators have larger standard errors. Hence, regularization reduces the bias of the EWALD estimator at the expense of an increase of the variance. This is the traditional bias-variance trade-off that arises in instrumental variables estimation in presence of many instruments.
The coefficient of the non labor income variable (not showed in the results table) is only significant in the OLS model and not in any of the others models.
Taxation is also known to be another source of endogeneity in labor supply models. In presence of progressive taxation the after tax wage depends on the marginal tax rate which depends on the taxable income which in turn is function of the hours worked. In this study we do not account for taxation as census data do not provide enough details to compute correctly the marginal tax rates. Labor or income dynamics surveys usually provide enough information to compute tax rates but using these data for grouping estimators may be problematic because of small sample size within-groups.

3.5 Conclusion

In this paper, we estimate the labor supply elasticities of Canadian men using population census 2011, 2001, 1991 and 1981. We take advantage of the equivalence between grouping estimators and the tradition instrumental variables estimators to apply a regularization approach. This technique was developed in the instrumental literature to improve the finite sample properties of the 2SLS estimator in presence of many insruments. The wage elasticities obtained by the regularized estimator are larger than those of the EWALD. These results suggest that the use of regularization in the grouping estimator setting can be a good alternative to the traditional instrumental variables approach where the exclusion assumption is sometimes hard to justify.
The Errors in Variables Estimator (EVE) and the Unbaised Errors in Variables Estimator (UEVE) are others grouping estimators proposed as solutions to the poor performances of the EWALD in finite sample. Devereux (2007) proved that the EVE is exactly equivalent to the Jackknife Instrumental Variables Estimator (JIVE) suggested by Phillips and Hale (1977) and popularized by Angrist, Imbens, and Krueger (1999) and that the UEVE can be seen as a k-class estimators. Hence, possible extensions of this study

would be to compare the finite sample properties UEVE and EVE to the regularized LIML proposed by Carrasco and Tchuente (2015) and the regularized Jackknife instrumental varibale estimator (RJIVE) proposed by Carrasco and Doukali (2016).

Bibliographie

[1] Paul J. Devereux (a). Changes in relative wages and family labor supply. *The Journal of Human Resources*, 39(3) :696–722, 2004.

[2] Kentaro Akashi and Naoto Kunitomo. Some properties of the liml estimator in a dynamic panel structural equation. *Journal of Econometrics*, 166(2) :167 – 183, 2012.

[3] César Alonso-Borrego and Manuel Arellano. Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business & Economic Statistics*, 17(1) :36–49, 1999.

[4] Javier Alvarez and Manuel Arellano. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4) :1121–1159, 2003.

[5] Donald Andrews and Biao Lu. Consistent model and moment selection procedures for gmm estimation with application to dynamic panel data models. *Journal of Econometrics*, 101(1) :123–164, 2001.

[6] Joshua D. Angrist. Grouped-data estimation and testing in simple labor-supply models. *Journal of Econometrics*, 47(2) :243 – 266, 1991.

[7] Joshua D. Angrist and Alan B. Krueger. Does compulsory school attendance affect schooling and earnings ? *The Quarterly Journal of Economics*, 106(4) :979–1014, 1991.

[8] Manuel Arellano. Modelling optimal instrumental variables for dynamic panel data models. *Research in Economics*, 70(2) :238 – 261, 2016.

[9] Manuel Arellano and Olympia Bover. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1) :29 – 51, 1995.

[10] Paul J Devereux (b). Improved errors-in-variables estimators for grouped data. *Journal of Business & Economic Statistics*, 25(3) :278–287, 2007.

[11] Thorsten Beck, Ross Levine, and Norman Loayza. Finance and the sources of growth. *Journal of Financial Economics*, 58(1–2) :261 – 300, 2000. Special Issue on International Corporate Governance.

[12] Francine D. Blau and Lawrence M. Kahn. Changes in the labor supply behavior of married women : 1980–2000. *Journal of Labor Economics*, 25(3) :393–438, 2007.

[13] Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1) :115 – 143, 1998.

[14] Richard Blundell and Stephen Bond. Gmm estimation with persistent panel data : an application to production functions. *Econometric Reviews*, 19(3) :321–340, 2000.

[15] Stephen Bond, Anke Hoeffler, and Jonathan Temple. GMM Estimation of Empirical Growth Models. Economics Papers 2001-W21, Economics Group, Nuffield College, University of Oxford, 2001.

[16] Maurice J. G. Bun and Frank Windmeijer. The weak instrument problem of the system gmm estimator in dynamic panel data models. *Econometrics Journal*, 13(1) :95–126, 2010.

[17] Maurice J.G. Bun and Jan F. Kiviet. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics*, 132(2) :409–444, June 2006.

[18] Marine Carrasco. A regularization approach to the many instruments problem. *Journal of Econometrics*, 170(2) :383–398, 2012.

[19] Marine Carrasco and Mohamed Doukali. Efficient estimation using regularized jackknife iv estimator. *Annals of Economics and Statistics*, (128) :109–149, 2017.

[20] Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Chapter 77 linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. volume 6, Part B of *Handbook of Econometrics*, pages 5633 – 5751. Elsevier, 2007.

[21] Marine Carrasco and Ada Nayihouba (a). A regularization approach to the dynamic panel data model estimation. *Unpublished manuscript*, 2019.

[22] Marine Carrasco and Ada Nayihouba (b). Regularized liml for dynamic panel data models. *Unpublished manuscript*, 2019.

[23] Marine Carrasco and Guy Tchuente. Regularized liml for many instruments. *Journal of Econometrics*, 186(2) :427 – 442, 2015.

[24] Marine Carrasco and Guy Tchuente. Regularized LIML for many instruments. *Journal of Econometrics*, 186(2) :427–442, 2015.

[25] Francesco Caselli, Gerardo Esquivel, and Fernando Lefort. Reopening the Convergence Debate : A New Look at Cross-Country Growth Empirics. *Journal of Economic Growth*, 1(3) :363–89, 1996.

[26] Peter Craven and Grace Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4) :377–403, Dec 1978.

[27] David Dollar and Aart Kraay. Growth is good for the poor. *Journal of Economic Growth*, 7(3) :195–225, Sep 2002.

[28] Stephen G Donald and Whitney K Newey. Choosing the Number of Instruments. *Econometrica*, 69(5) :1161–91, 2001.

[29] Howard E. Doran and Peter Schmidt. Gmm estimators with improved finite sample properties using principal components of the weighting matrix, with an application to the dynamic panel data model. *Journal of Econometrics*, 133(1) :387 – 409, 2006.

[30] Benoit Dostie and Lene Kromann. Labour Supply and Taxes : New Estimates of the Responses of Wives to Husbands' Wages. IZA Discussion Papers 6392, Institute of Labor Economics (IZA), February 2012.

[31] Kristin J. Forbes. A Reassessment of the Relationship between Inequality and Growth. *American Economic Review*, 90(4) :869–887, 2000.

[32] C.W. Groetsch. *Inverse problems in the mathematical sciences.* Theory and Practice of Applied Geophysics Series. Vieweg, 1993.

[33] Hausman J. Kuersteiner G. Hahn, J. Bias corrected instrumental variables estimation for dynamic panel models with fixed effects. *Unpublished manuscript.*, 2001.

[34] Jinyong Hahn and Guido Kuersteiner. Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both "n" and "T" Are Large. *Econometrica*, 70(4) :1639–1657, 2002.

[35] Bruce E. Hansen. Least Squares Model Averaging. *Econometrica*, 75(4) :1175–1189, 07 2007.

[36] Keisuke Hirano. Semiparametric bayesian inference in autoregressive panel data models. *Econometrica*, 70(2) :781–799, 2002.

[37] Guido Imbens, Joshua Angrist, and Alan Krueger. Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1) :57–67, 1999.

[38] Mr Mohsin S Khan and Mr A Senhadji Semlali. *Financial development and economic growth : an overview.* Number 0-209. International Monetary Fund, 2000.

[39] Jan F. Kiviet. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, 68(1) :53 – 78, 1995.

[40] Rainer Kress. Linear integral equations. 1999.

[41] Guido M. Kuersteiner. Kernel-weighted gmm estimators for linear time series models. *Journal of Econometrics*, 170(2) :399 – 421, 2012. Thirtieth Anniversary of Generalized Method of Moments.

[42] Ross Levine, Norman Loayza, and Thorsten Beck. Financial intermediation and growth : Causality and causes. *Journal of Monetary Economics*, 46(1) :31 – 77, 2000.

[43] Ross Levine, Norman Loayza, and Thorsten Beck. Financial intermediation and growth : Causality and causes. *Journal of Monetary Economics*, 46(1) :31–77, 2000.

[44] Ker-Chau Li. Asymptotic optimality of cl and generalized cross-validation in ridge regression with application to spline smoothing. *The Annals of Statistics*, 14(3) :pp. 1101–1112, 1986.

[45] Ker-Chau Li. Asymptotic optimality for cp, cl, cross-validation and generalized cross-validation : Discrete index set. *The Annals of Statistics*, 15(3) :pp. 958–975, 1987.

[46] René Morissette and Feng Hou. Does the labour supply of wives respond to husbands' wages ? canadian evidence from micro data and grouped data. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 41(4) :1185–1210, 2008.

[47] A. L. Nagar. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica*, 27(4) :pp. 575–595, 1959.

[48] Ryo Okui. The optimal choice of moments in dynamic panel data models. *Journal of Econometrics*, 151(1) :1–16, 2009.

[49] Ryo Okui. Instrumental variable estimation in the presence of many moment conditions. *Journal of Econometrics*, 165(1) :70–86, 2011.

[50] G. D. A. Phillips and C. Hale. The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review*, 18(1) :219–228, 1977.

[51] David Roodman. A note on the theme of too many instruments*. *Oxford Bulletin of Economics and Statistics*, 71(1) :135–158, 2009.

Annexes

Appendices

A Chapter 1

A.1 Simulations results.

	Min	q1	Mean	Median	q3	max	$\bar{q}$
$\delta = 0.20$							
T= 5	6.2	10.7	13.4	12.8	15.5	32.1	10
T= 10	21.5	34.7	42.7	41.5	49.7	93.3	45
T= 25	143.2	282.4	359.9	343.6	421.3	959.1	300
T= 50	17418.9	68429.4	734508.6	137592.7	310565.6	92594681.5	1225
$\delta = 0.50$							
T= 5	15.3	35.0	44.5	42.9	52.0	131.1	10
T= 10	56.3	119.5	148.4	141.9	173.9	335.3	45
T= 25	343.2	892.4	1176.6	1119.4	1361.3	3137.3	300
T= 50	33789.1	217795.9	2097168.8	428699.4	970051.8	356480210.5	1225
$\delta = 0.90$							
T= 5	626.2	1183.7	1509.0	1440.7	1751.9	3901.3	10
T= 10	2168.4	4245.2	5316.8	5067.7	6214.7	13130.8	45
T= 25	16422.2	32971.4	42706.8	40619.7	49839.7	108698.3	300
T= 50	1124835.1	7618036.7	242677244.8	14558610.1	36715676.3	178194140502.4	1225

TABLE A.1 – Properties of the condition number with N=50, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 1000 replications.

		GMM	GMM5	IVK	TH	PC	LF
$\delta = 0.50$							
	Med.bias	-0.0626	-0.0456	-0.0383	-0.0336	-0.0248	-0.0307
T= 10	Med.abs	0.0696	0.0616	0.0635	0.0586	0.0579	0.0581
	Iqr	0.1029	0.1033	0.1124	0.1070	0.1091	0.1072
	Med.mse	0.0048	0.0038	0.0040	0.0034	0.0033	0.0034
	Cov	0.8538	0.9006	0.9148	0.9676	0.9268	0.9538
	Med.bias	-0.0398	-0.0121	-0.0144	-0.0086	-0.0017	-0.0069
T= 25	Med.abs	0.0401	0.0261	0.0265	0.0249	0.0252	0.0251
	Iqr	0.0442	0.0499	0.0487	0.0471	0.0505	0.0485
	Med.mse	0.0016	0.0007	0.0007	0.0006	0.0006	0.0006
	Cov	0.7654	0.9294	0.9234	0.9868	0.9424	0.9754
$\delta = 0.75$							
	Med.bias	-0.1273	-0.1009	-0.0876	-0.0874	-0.0747	-0.0846
T= 10	Med.abs	0.1276	0.1034	0.1122	0.0989	0.0928	0.0970
	Iqr	0.1324	0.1334	0.1790	0.1481	0.1507	0.1495
	Med.mse	0.0163	0.0107	0.0126	0.0098	0.0086	0.0094
	Cov	0.6914	0.7962	0.8656	0.9388	0.8822	0.9214
	Med.bias	-0.0638	-0.0274	-0.0267	-0.0259	-0.0164	-0.0242
T= 25	Med.abs	0.0638	0.0333	0.0344	0.0329	0.0306	0.0328
	Iqr	0.0449	0.0552	0.0583	0.0538	0.0567	0.0548
	Med.mse	0.0041	0.0011	0.0012	0.0011	0.0009	0.0011
	Cov	0.4748	0.8860	0.8928	0.9774	0.9316	0.9624
$\delta = 0.95$							
	Med.bias	-0.3411	-0.3024	-0.3176	-0.3046	-0.2888	-0.3193
T= 10	Med.abs	0.3411	0.3024	0.3219	0.3069	0.2947	0.3235
	Iqr	0.1955	0.2092	0.3381	0.3090	0.3283	0.3926
	Med.mse	0.1163	0.0914	0.1036	0.0942	0.0868	0.1046
	Cov	0.1900	0.3506	0.7434	0.8430	0.7720	0.8934
	Med.bias	-0.1380	-0.0951	-0.1084	-0.0948	-0.0806	-0.0962
T= 25	Med.abs	0.1380	0.0951	0.1088	0.0953	0.0819	0.0971
	Iqr	0.0575	0.0719	0.1188	0.1039	0.1071	0.1284
	Med.mse	0.0191	0.0090	0.0118	0.0091	0.0067	0.0094
	Cov	0.0148	0.4852	0.7380	0.8848	0.8224	0.8998

TABLE A.2 – Simulations results with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 5000 replications.

		Mean	std	Mode	q1	Median	q3
$\delta = 0.50$							
	TH	0.00054	0.00020	0.00012	0.00040	0.00050	0.00064
T= 10	PC	29.794	4.329	30.000	27.000	30.000	33.000
	LF	2405.839	1330.126	2286.000	1455.000	2000.000	2909.000
	TH	0.00008	0.00002	0.00004	0.00007	0.00008	0.00009
T= 25	PC	137.492	16.972	135.000	125.000	135.000	150.000
	LF	6220.263	2808.828	5333.000	4267.000	5818.000	8000.000
$\delta = 0.75$							
	TH	0.00550	0.05349	0.00016	0.00107	0.00165	0.00268
T= 10	PC	20.757	4.856	21.000	18.000	21.000	24.000
	LF	14185.296	13041.976	10667.000	5333.000	10667.000	18286.000
	TH	0.00037	0.00011	0.00013	0.00029	0.00035	0.00043
T= 25	PC	86.896	13.353	85.000	75.000	85.000	95.000
	LF	24279.314	12967.227	21333.000	16000.000	21333.000	32000.000
$\delta = 0.95$							
	TH	0.37266	0.45504	0.45000	0.00601	0.03849	0.99033
T= 10	PC	11.636	5.375	9.000	9.000	9.000	15.000
	LF	91501.814	578538.647	8000.000	302.000	8000.000	8000.000
	TH	0.15202	0.32676	0.45000	0.00365	0.00845	0.03550
T= 25	PC	40.753	15.385	25.000	25.000	40.000	50.000
	LF	435107.242	1032787.507	8000.000	8000.000	8000.000	512000.000

TABLE A.3 – Properties of the distribution of the regularization parameters with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 5000 replications.

		GMM	GMM5	IVK	TH	PC	LF
$\delta = 0.50$							
T= 10	Med.bias	-0.0721	-0.0526	-0.0482	-0.0448	-0.0331	-0.0423
	Med.abs	0.0769	0.0666	0.0699	0.0634	0.0632	0.0637
	Iqr	0.1079	0.1129	0.1192	0.1135	0.1172	0.1158
	Med.mse	0.0059	0.0044	0.0049	0.0040	0.0040	0.0041
	Cov	0.8500	0.9052	0.9100	0.9734	0.9328	0.9604
T= 25	Med.bias	-0.0417	-0.0152	-0.0172	-0.0112	-0.0039	-0.0098
	Med.abs	0.0419	0.0281	0.0277	0.0260	0.0264	0.0262
	Iqr	0.0452	0.0527	0.0504	0.0507	0.0523	0.0516
	Med.mse	0.0018	0.0008	0.0008	0.0007	0.0007	0.0007
	Cov	0.7558	0.9398	0.9280	0.9898	0.9482	0.9806
$\delta = 0.75$							
T= 10	Med.bias	-0.1529	-0.1246	-0.1411	-0.1141	-0.0958	-0.1120
	Med.abs	0.1530	0.1257	0.1505	0.1205	0.1098	0.1191
	Iqr	0.1410	0.1448	0.1907	0.1559	0.1633	0.1580
	Med.mse	0.0234	0.0158	0.0227	0.0145	0.0121	0.0142
	Cov	0.6388	0.7644	0.8460	0.9396	0.8610	0.9186
T= 25	Med.bias	-0.0654	-0.0301	-0.0321	-0.0276	-0.0175	-0.0259
	Med.abs	0.0654	0.0358	0.0385	0.0335	0.0310	0.0331
	Iqr	0.0459	0.0569	0.0590	0.0552	0.0584	0.0568
	Med.mse	0.0043	0.0013	0.0015	0.0011	0.0010	0.0011
	Cov	0.4708	0.8864	0.8872	0.9768	0.9306	0.9600
$\delta = 0.95$							
T= 10	Med.bias	-0.3450	-0.3077	-0.3405	-0.3205	-0.3034	-0.3525
	Med.abs	0.3450	0.3077	0.3418	0.3211	0.3080	0.3538
	Iqr	0.1953	0.2061	0.3170	0.2957	0.3251	0.3973
	Med.mse	0.1190	0.0947	0.1168	0.1031	0.0949	0.1251
	Cov	0.1754	0.3414	0.7362	0.8366	0.7710	0.9022
T= 25	Med.bias	-0.1399	-0.0968	-0.1157	-0.1019	-0.0880	-0.1408
	Med.abs	0.1399	0.0968	0.1162	0.1020	0.0897	0.1412
	Iqr	0.0574	0.0724	0.1174	0.1044	0.1105	0.1896
	Med.mse	0.0196	0.0094	0.0135	0.0104	0.0080	0.0199
	Cov	0.0128	0.4802	0.7394	0.8842	0.8210	0.8966

TABLE A.4 – Simulations results with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 10$ for 5000 replications.

		Mean	std	Mode	q1	Median	q3
$\delta = 0.50$							
	TH	0.00054	0.00021	0.00011	0.00039	0.00050	0.00064
T= 10	PC	29.815	4.367	30.000	27.000	30.000	33.000
	LF	Inf	NaN	170667.000	128000.000	170667.000	256000.000
	TH	0.00009	0.00002	0.00004	0.00007	0.00008	0.00009
T= 25	PC	138.217	17.054	135.000	125.000	135.000	150.000
	LF	577781.303	262991.985	512000.000	390095.000	512000.000	682667.000
$\delta = 0.75$							
	TH	0.00344	0.03671	0.00009	0.00098	0.00150	0.00236
T= 10	PC	21.211	4.954	21.000	18.000	21.000	24.000
	LF	Inf	NaN	1024000.000	512000.000	1024000.000	1638400.000
	TH	0.00036	0.00011	0.00011	0.00028	0.00035	0.00042
T= 25	PC	87.302	13.221	85.000	80.000	85.000	95.000
	LF	2300796.828	1223621.596	2048000.000	1365333.000	2048000.000	2730667.000
$\delta = 0.95$							
	TH	0.35606	0.45028	0.45000	0.00504	0.03048	0.98778
T= 10	PC	11.872	5.668	9.000	9.000	9.000	15.000
	LF	5521.438	3632.648	8000.000	327.000	8000.000	8000.000
	TH	0.13835	0.31678	0.45000	0.00359	0.00791	0.02828
T= 25	PC	40.833	15.371	25.000	25.000	40.000	50.000
	LF	4354.137	3454.176	8000.000	1143.000	2667.000	8000.000

Table A.5 – Properties of the distribution of the regularization parameters with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 10$ for 5000 replications.

		GMM	GMM5	IVK	TH	PC	LF
$\delta = 0.50$							
	Med.bias	-0.0333	-0.0230	-0.0204	-0.0173	-0.0123	-0.0160
T= 10	Med.abs	0.0424	0.0404	0.0410	0.0389	0.0383	0.0390
	Iqr	0.0733	0.0737	0.0775	0.0732	0.0758	0.0729
	Med.mse	0.0018	0.0016	0.0017	0.0015	0.0015	0.0015
	Cov	0.9098	0.9288	0.9320	0.9694	0.9410	0.9588
	Med.bias	-0.0221	-0.0078	-0.0102	-0.0056	-0.0011	-0.0046
T= 25	Med.abs	0.0236	0.0188	0.0186	0.0172	0.0172	0.0173
	Iqr	0.0326	0.0365	0.0345	0.0339	0.0348	0.0347
	Med.mse	0.0006	0.0004	0.0003	0.0003	0.0003	0.0003
	Cov	0.8456	0.9434	0.9376	0.9864	0.9530	0.9766
$\delta = 0.75$							
	Med.bias	-0.0820	-0.0649	-0.0559	-0.0536	-0.0467	-0.0514
T= 10	Med.abs	0.0836	0.0704	0.0781	0.0679	0.0638	0.0670
	Iqr	0.0995	0.1041	0.1276	0.1107	0.1098	0.1112
	Med.mse	0.0070	0.0050	0.0061	0.0046	0.0041	0.0045
	Cov	0.7796	0.8528	0.8896	0.9428	0.8938	0.9276
	Med.bias	-0.0371	-0.0155	-0.0163	-0.0135	-0.0083	-0.0122
T= 25	Med.abs	0.0371	0.0228	0.0233	0.0214	0.0205	0.0212
	Iqr	0.0346	0.0405	0.0414	0.0395	0.0399	0.0403
	Med.mse	0.0014	0.0005	0.0005	0.0005	0.0004	0.0004
	Cov	0.6732	0.9170	0.9136	0.9710	0.9380	0.9592
$\delta = 0.95$							
	Med.bias	-0.3065	-0.2705	-0.2885	-0.2606	-0.2464	-0.2923
T= 10	Med.abs	0.3065	0.2705	0.2903	0.2625	0.2543	0.2964
	Iqr	0.1921	0.1951	0.3301	0.2839	0.3043	0.4002
	Med.mse	0.0939	0.0732	0.0843	0.0689	0.0647	0.0879
	Cov	0.2344	0.4048	0.7692	0.8534	0.7896	0.9086
	Med.bias	-0.1189	-0.0780	-0.0911	-0.0717	-0.0611	-0.0794
T= 25	Med.abs	0.1189	0.0780	0.0925	0.0723	0.0627	0.0815
	Iqr	0.0532	0.0650	0.1100	0.0821	0.0850	0.1262
	Med.mse	0.0141	0.0061	0.0086	0.0052	0.0039	0.0066
	Cov	0.0324	0.5588	0.7634	0.8820	0.8112	0.9130

TABLE A.6 – Simulations results with $N = 100$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 5000 replications.

		Mean	std	Mode	q1	Median	q3
$\delta = 0.50$							
	TH	0.00032	0.00008	0.00010	0.00026	0.00031	0.00036
T= 10	PC	33.089	3.244	33.000	30.000	33.000	36.000
	LF	3475.854	1274.336	3200.000	2667.000	3200.000	4000.000
	TH	0.00005	0.00001	0.00003	0.00004	0.00005	0.00005
T= 25	PC	167.610	16.523	175.000	155.000	170.000	180.000
	LF	10265.839	3236.181	10667.000	8000.000	9846.000	11636.000
$\delta = 0.75$							
	TH	0.00120	0.00081	0.00019	0.00072	0.00102	0.00146
T= 10	PC	23.418	4.182	21.000	21.000	21.000	27.000
	LF	20999.097	12820.196	16000.000	10667.000	18286.000	25600.000
	TH	0.00020	0.00005	0.00009	0.00017	0.00020	0.00023
T= 25	PC	103.970	12.934	100.000	95.000	105.000	110.000
	LF	42473.371	16004.677	42667.000	32000.000	42667.000	51200.000
$\delta = 0.95$							
	TH	0.29746	0.42469	0.45000	0.00457	0.01843	0.92724
T= 10	PC	12.241	5.612	9.000	9.000	9.000	15.000
	LF	26886.475	501363.427	8000.000	372.000	8000.000	8000.000
	TH	0.04658	0.18151	0.45000	0.00239	0.00423	0.00855
T= 25	PC	47.785	13.911	50.000	35.000	50.000	60.000
	LF	407484.865	933804.334	8000.000	8000.000	8000.000	512000.000

TABLE A.7 – Properties of the distribution of the regularization parameters with $N = 100$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 5000 replications.

		GMM	GMM5	IVK	TH	PC	LF
$\delta = 0.50$							
	Med.bias	-0.0391	-0.0279	-0.0265	-0.0241	-0.0184	-0.0225
T= 10	Med.abs	0.0486	0.0449	0.0461	0.0439	0.0438	0.0440
	Iqr	0.0800	0.0821	0.0856	0.0816	0.0827	0.0820
	Med.mse	0.0024	0.0020	0.0021	0.0019	0.0019	0.0019
	Cov	0.9028	0.9260	0.9310	0.9696	0.9402	0.9586
	Med.bias	-0.0226	-0.0080	-0.0101	-0.0056	-0.0009	-0.0041
T= 25	Med.abs	0.0244	0.0197	0.0191	0.0181	0.0184	0.0182
	Iqr	0.0336	0.0377	0.0356	0.0358	0.0368	0.0359
	Med.mse	0.0006	0.0004	0.0004	0.0003	0.0003	0.0003
	Cov	0.8496	0.9448	0.9366	0.9870	0.9510	0.9762
$\delta = 0.75$							
	Med.bias	-0.0949	-0.0761	-0.0871	-0.0699	-0.0591	-0.0677
T= 10	Med.abs	0.0965	0.0814	0.1004	0.0787	0.0756	0.0771
	Iqr	0.1091	0.1107	0.1419	0.1175	0.1223	0.1190
	Med.mse	0.0093	0.0066	0.0101	0.0062	0.0057	0.0059
	Cov	0.7630	0.8474	0.8804	0.9516	0.8898	0.9322
	Med.bias	-0.0390	-0.0173	-0.0192	-0.0149	-0.0094	-0.0140
T= 25	Med.abs	0.0390	0.0247	0.0255	0.0226	0.0215	0.0223
	Iqr	0.0354	0.0434	0.0425	0.0404	0.0419	0.0411
	Med.mse	0.0015	0.0006	0.0006	0.0005	0.0005	0.0005
	Cov	0.6548	0.9164	0.9050	0.9754	0.9360	0.9612
$\delta = 0.95$							
	Med.bias	-0.3258	-0.2888	-0.3212	-0.2958	-0.2836	-0.3389
T= 10	Med.abs	0.3258	0.2888	0.3223	0.2961	0.2901	0.3403
	Iqr	0.1921	0.2000	0.3040	0.2733	0.3012	0.3984
	Med.mse	0.1062	0.0834	0.1039	0.0877	0.0842	0.1158
	Cov	0.2052	0.3630	0.7516	0.8376	0.7618	0.9112
	Med.bias	-0.1220	-0.0841	-0.1082	-0.0790	-0.0685	-0.1420
T= 25	Med.abs	0.1220	0.0841	0.1086	0.0795	0.0693	0.1426
	Iqr	0.0551	0.0666	0.1089	0.0833	0.0863	0.1920
	Med.mse	0.0149	0.0071	0.0118	0.0063	0.0048	0.0203
	Cov	0.0260	0.5326	0.7394	0.8872	0.8104	0.8950

Table A.8 – Simulations results with $N = 100$, $\sigma^2 = 1$, $\sigma_\eta^2 = 10$ for 5000 replications.

		Mean	std	Mode	q1	Median	q3
$\delta = 0.50$							
	TH	0.00032	0.00008	0.00011	0.00026	0.00031	0.00036
T= 10	PC	33.160	3.295	33.000	30.000	33.000	36.000
	LF	Inf	NaN	256000.000	204800.000	256000.000	341333.000
	TH	0.00006	0.00004	0.00003	0.00004	0.00005	0.00005
T= 25	PC	167.226	16.337	170.000	155.000	165.000	180.000
	LF	Inf	NaN	1024000.000	744727.000	910222.000	1092267.000
$\delta = 0.75$							
	TH	0.00130	0.01223	0.00015	0.00069	0.00097	0.00135
T= 10	PC	23.723	4.329	21.000	21.000	24.000	27.000
	LF	Inf	NaN	2048000.000	1024000.000	1638400.000	2340571.000
	TH	0.00020	0.00005	0.00009	0.00017	0.00020	0.00023
T= 25	PC	103.911	12.851	100.000	95.000	105.000	115.000
	LF	3965736.310	1531196.863	4096000.000	2730667.000	4096000.000	4681143.000
$\delta = 0.95$							
	TH	0.27224	0.41269	0.45000	0.00409	0.01478	0.72823
T= 10	PC	12.484	5.869	9.000	9.000	9.000	15.000
	LF	6092.786	3354.498	8000.000	8000.000	8000.000	8000.000
	TH	0.03997	0.16696	0.45000	0.00228	0.00403	0.00780
T= 25	PC	48.557	14.235	50.000	40.000	50.000	60.000
	LF	5070.820	3519.871	8000.000	1231.000	8000.000	8000.000

Table A.9 – Properties of the distribution of the regularization parameters with $N = 100$, $\sigma^2 = 1$, $\sigma_\eta^2 = 10$ for 5000 replications.

		GMM	GMM5	IVK	TH	PC	LF	GMM	GMM5	IVK	TH	PC	LF
				$\delta = 0.50$					$\theta = 0.95$				
T= 10	Med.bias	-0.037	-0.050	-0.049	-0.024	-0.021	-0.022	-0.006	-0.010	-0.006	-0.004	0.003	-0.003
	Med.abs	0.039	0.054	0.053	0.032	0.031	0.031	0.033	0.081	0.082	0.035	0.042	0.037
	Iqr	0.050	0.078	0.075	0.052	0.054	0.053	0.065	0.165	0.165	0.071	0.083	0.073
	Med.mse	0.001	0.003	0.003	0.001	0.001	0.001	0.001	0.007	0.007	0.001	0.002	0.001
	Cov	0.509	0.861	0.849	0.480	0.402	0.454	0.714	0.949	0.953	0.652	0.451	0.611
T= 25	Med.bias	-0.032	-0.025	-0.026	-0.015	-0.013	-0.014	-0.003	-0.003	0.000	-0.002	0.002	-0.002
	Med.abs	0.032	0.030	0.027	0.018	0.018	0.018	0.019	0.046	0.031	0.024	0.030	0.025
	Iqr	0.026	0.049	0.032	0.029	0.030	0.029	0.038	0.091	0.061	0.047	0.059	0.050
	Med.mse	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.002	0.001	0.001	0.001	0.001
	Cov	0.202	0.900	0.815	0.320	0.269	0.298	0.625	0.951	0.949	0.476	0.305	0.438
				$\delta = 0.75$					$\theta = 0.95$				
T= 10	Med.bias	-0.052	-0.089	-0.074	-0.037	-0.031	-0.036	-0.013	-0.027	-0.006	-0.009	-0.003	-0.008
	Med.abs	0.052	0.089	0.081	0.039	0.036	0.038	0.032	0.082	0.111	0.036	0.048	0.038
	Iqr	0.047	0.086	0.110	0.052	0.055	0.053	0.063	0.161	0.222	0.072	0.095	0.075
	Med.mse	0.003	0.008	0.007	0.001	0.001	0.001	0.001	0.007	0.012	0.001	0.002	0.001
	Cov	0.329	0.706	0.849	0.450	0.359	0.428	0.701	0.939	0.954	0.706	0.451	0.665
T= 25	Med.bias	-0.038	-0.035	-0.031	-0.018	-0.014	-0.017	-0.005	-0.006	0.000	-0.003	0.000	-0.003
	Med.abs	0.038	0.036	0.032	0.019	0.018	0.019	0.020	0.045	0.042	0.026	0.035	0.027
	Iqr	0.022	0.045	0.038	0.027	0.029	0.028	0.039	0.091	0.083	0.051	0.070	0.054
	Med.mse	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.002	0.002	0.001	0.001	0.001
	Cov	0.063	0.800	0.797	0.289	0.245	0.280	0.607	0.952	0.950	0.545	0.301	0.501
				$\delta = 0.95$					$\theta = 0.95$				
T= 10	Med.bias	-0.074	-0.170	-0.153	-0.066	-0.057	-0.083	-0.029	-0.071	-0.055	-0.025	-0.016	-0.018
	Med.abs	0.074	0.170	0.156	0.066	0.061	0.095	0.038	0.097	0.176	0.044	0.064	0.094
	Iqr	0.045	0.109	0.206	0.058	0.078	0.172	0.065	0.165	0.352	0.081	0.127	0.186
	Med.mse	0.005	0.029	0.024	0.004	0.004	0.009	0.001	0.009	0.031	0.002	0.004	0.009
	Cov	0.122	0.308	0.828	0.320	0.272	0.337	0.624	0.905	0.957	0.737	0.433	0.585
T= 25	Med.bias	-0.051	-0.058	-0.050	-0.041	-0.037	-0.043	-0.017	-0.022	-0.010	-0.014	-0.007	-0.014
	Med.abs	0.051	0.058	0.052	0.042	0.042	0.044	0.024	0.049	0.081	0.043	0.087	0.053
	Iqr	0.018	0.040	0.063	0.038	0.064	0.046	0.040	0.091	0.162	0.086	0.173	0.107
	Med.mse	0.003	0.003	0.003	0.002	0.002	0.002	0.001	0.002	0.007	0.002	0.008	0.003
	Cov	0.001	0.409	0.796	0.143	0.206	0.164	0.526	0.942	0.949	0.510	0.313	0.459

TABLE A.10 – Simulations with an exogenous covariate results with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma^2 = 1$, $\rho = 1$ for 5000 replications.

	GMM	GMM5	IVK	TH	PC	LF	GMM	GMM5	IVK	TH	PC	LF
			$\delta = 0.87$					$\theta = 0.02$				
Med.bias	-0.1103	-0.0832	-0.0769	-0.0581	-0.0420	-0.0548	0.0252	0.0245	0.0279	0.0083	0.0031	0.0066
Med.abs	0.1103	0.0832	0.0774	0.0586	0.0500	0.0561	0.0268	0.0347	0.0398	0.0386	0.0626	0.0414
Iqr	0.0463	0.0627	0.0660	0.0636	0.0760	0.0661	0.0354	0.0595	0.0670	0.0769	0.1243	0.0822
Med.mse	0.0122	0.0069	0.0060	0.0034	0.0025	0.0031	0.0007	0.0012	0.0016	0.0015	0.0039	0.0017
Cov	0.0030	0.4208	0.5476	0.9510	0.8012	0.9214	0.4112	0.8798	0.8716	0.9994	0.9192	0.9984

TABLE A.11 – Simulations with an endogenous covariate results with $N = 77$, $T = 21$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma^2 = 1$, $\rho = 1$ for 5000 replications.

A.2 Proofs.

Proof of Lemma 1.

(i)

$$\begin{aligned} tr[K] &= \frac{1}{NT^{3/2}} tr[E[Z'Z]] \\ &= \frac{1}{NT^{3/2}} \sum_{i=1}^{N} tr[E[Z_i'Z_i]] \\ &= \frac{1}{T^{3/2}} tr[E[Z_i'Z_i]]. \end{aligned}$$

By construction the matrix $E[Z_i'Z_i]$ is a block diagonal matrix which is defined in the following way

$$E[Z_i'Z_i] = Diag[E[z_{i1}z_{i1}'], ..., E[z_{it}z_{it}'], ..., E[z_{iT-1}z_{iT-1}']].$$

For any t, the matrix $E[z_{it}z_{it}']$ is of order $t \times t$ with diagonal elements in the form of $E[x_{is}^2]$ for $s = 1, 2, ...t$. So

$$tr[K] = \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} tE[x_{it}^2].$$

Now using the fact that $x_{it} = y_{it-1}$ and the autoregressive equation defining the DPM model

$$\begin{aligned} E[y_{it-1}^2] = Var[y_{it-1}] &= Var[w_{it-1} - \mu_i] \\ &= Var[w_{it-1}] - 2Cov[w_{it-1}, \mu_i] + Var[\mu_i]. \end{aligned}$$

By Assumption 3, we have $Cov[w_{it-1}, \mu_i] = 0$ because w_{i0} is independent of μ_i so that

$$\begin{aligned} E[y_{it-1}^2] &= Var[w_{it-1}] + Var[\mu_i] \\ &= \frac{\sigma^2}{1-\delta^2} + \frac{\sigma_\eta^2}{(1-\delta)^2} \end{aligned}$$

Now

$$E[x_{it}^2] = E[y_{it-1}^2] = \frac{\sigma^2}{1-\delta^2} + \frac{\sigma_\eta^2}{(1-\delta)^2}.$$

It follows that

$$\begin{aligned} tr[K] &= \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} tE[x_{it}^2] \\ &= \frac{1}{T^{3/2}} E[x_{it}^2] \sum_{t=1}^{T-1} t \\ &= \frac{1}{T^{3/2}} E[x_{it}^2] O(T^2) \\ &= O(T^{1/2}). \end{aligned}$$

(ii) For any symmetric matrix $A = (a_{ij})$ we have

$$tr[A^2] = \sum_{i,j} a_{i,j}^2$$

So the trace of $tr[K^2]$ is given by the sum of the squares of all the elements of the matrix K. By construction, we have

$$K = Diag[K_1, ..., K_t, ...K_{T-1}]$$

where for a given t, $K_t = \sum_i E[z_{it}z_{it}'/NT^{3/2}] = E[z_{it}z_{it}']/T^{3/2}$. Let us denote by $K_{ab,t}$ the element (a, b) of the $t \times t$ matrix K_t. We have

$$\begin{aligned} K_{ab,t} &= \frac{1}{T^{3/2}} E[x_{ia}x_{ib}] \\ &= \frac{1}{T^{3/2}} \left(\frac{\sigma^2}{(1-\delta^2)} \delta^{|a-b|} + \frac{\sigma_\eta^2}{(1-\delta)^2} \right) \end{aligned}$$

with $1 \le a, b \le t$.
We now calculate $tr[K_t^2]$ by summing the squares of the elements of K_t.

$$\begin{aligned} tr[K_tK_t] &= \sum_{a,b=1}^{t} K_{ab,t}^2 \\ &= \sum_{a=1}^{t} K_{aa,t}^2 + \sum_{a\neq b=1}^{t} K_{ab,t}^2. \end{aligned}$$

We have

$$\begin{aligned}
\sum_{a=1}^{t} K_{aa,t}^2 &= \sum_{a=1}^{t} \frac{1}{T^3}\left(\frac{\sigma^2}{(1-\delta^2)} + \frac{\sigma_\eta^2}{(1-\delta)^2}\right)^2 \\
&= \frac{t}{T^3}\left(\frac{\sigma^2}{(1-\delta^2)} + \frac{\sigma_\eta^2}{(1-\delta)^2}\right)^2 \\
&= O\left(\frac{t}{T^3}\right)
\end{aligned}$$

$$\begin{aligned}
\sum_{a\neq b=1}^{t} K_{ab,t}^2 &= \frac{1}{T^3}\sum_{a\neq b=1}^{t}\left(\frac{\sigma^2}{(1-\delta^2)}\delta^{|a-b|} + \frac{\sigma_\eta^2}{(1-\delta)^2}\right)^2 \\
&= \frac{1}{T^3}\frac{\sigma^4}{(1-\delta^2)^2}\sum_{a\neq b=1}^{t}\delta^{2(a-b)} + \frac{2}{T^3}\frac{\sigma^2}{(1-\delta^2)}\frac{\sigma_\eta^2}{(1-\delta)^2}\sum_{a\neq b=1}^{t}\delta^{|a-b|} \\
&+ \frac{1}{T^3}\sum_{a\neq b=1}^{t}\frac{\sigma_\eta^4}{(1-\delta)^4}
\end{aligned}$$

But

$$\begin{aligned}
\sum_{a\neq b=1}^{t}\delta^{2(a-b)} &= 2\sum_{a=2}^{t}\sum_{b=1}^{a-1}\delta^{2b} \\
&= 2\sum_{a=2}^{t}\left[\frac{1-\delta^{2a}}{(1-\delta^2)} - 1\right] \\
&= \frac{2}{(1-\delta^2)}\sum_{a=2}^{t}\left[\delta^2 - \delta^{2a}\right] \\
&= \frac{2}{(1-\delta^2)}\left[\sum_{a=2}^{t}\delta^2 - \sum_{a=2}^{t}\delta^{2a}\right] \\
&= \frac{2}{(1-\delta^2)}\left[\delta^2(t-1) - \sum_{a=2}^{t}\delta^{2a}\right] \\
&= \frac{2}{(1-\delta^2)}\left[\delta^2(t-1) - \Big(\frac{1-\delta^{2t+2}}{(1-\delta^2)} - 1 - \delta^2\Big)\right] \\
&= O\big(t\big)
\end{aligned}$$

Similarly we have

$$\sum_{a\neq b=1}^{t}\delta^{(a-b)} = O\big(t\big)$$

Moreover
$$\sum_{a\neq b=1}^{t}\frac{\sigma_\eta^4}{(1-\delta)^4}=O\Big(t^2\Big)$$
From the three last results we have
$$\sum_{a\neq b=1}^{t}K_{ab,t}^2=O\Big(\frac{t}{T^3}\Big)$$

and then
$$\begin{aligned}tr[K_t^2] &= \sum_{a,b=1}^{t}K_{ab,t}^2\\ &= O\Big(\frac{t}{T^3}\Big)+O\Big(\frac{t^2}{T^3}\Big)\\ &= O\Big(\frac{t^2}{T^3}\Big).\end{aligned}$$

And finally,
$$tr[K^2]=\sum_{t=1}^{T-1}tr[K_tK_t]=\sum_{t=1}^{T-1}O\Big(\frac{t^2}{T^3}\Big)=O(1).$$

Preliminary results for the proof of Proposition 1

We begin by three lemmas which establish some preliminary useful results. We essentially show how to adapt some results of Alvarez and Arellano (2003)[AA(2003) hereafter] in our case. We denote by $E_t(.)$ the expectation conditional on η_i and $\Big\{v_{i(t-j)}\Big\}_{j=1}^{\infty}$.

Lemma 2. Let us denote by $d_t(\alpha)$ the $N\times 1$ vectors containing the diagonal elements of M_t^α, κ_3 and κ_4 be the third and fourth-order cumulants of v_{it}. Under assumptions 1-3 :

(*i*) $tr(M_t^\alpha)\leq t$,

(*ii*) $Var(v_t'M_t^\alpha v_t)\leq(2\sigma^4+\kappa_4)tr[M_t^{\alpha 2}]\leq(2\sigma^4+\kappa_4)t$,

(*iii*) $Var(v_t'M_t^\alpha v_{t+j})=\sigma^4 tr(M_t^\alpha)\leq\sigma^4 t$, for $j>0$,

(iv) $Cov(v_t' M_t^\alpha v_{t+j}, v_{t+j}' M_{t+j}^\alpha v_{t+j}) \leq \kappa_3 E\left(d_{t+j}(\alpha)' M_t^\alpha v_t\right) \leq \kappa_3 \sigma \sqrt{t+j}\,(tr\,(M_t^{\alpha 2}))^{1/2}$, for $j > 0$.

Proof of Lemma 2.

(i) The $t \times t$ symmetric matrix $Z_t' Z_t / NT^{3/2}$ can be decomposed as $P_t D_t P_t'$ with $P_t P_t' = I_t$ the t-dimensional identity matrix and $D_t = diag(\lambda_1^t, \lambda_2^t, ..., \lambda_t^t)$. The regularized inverse of D_t is $D_t(\alpha)$=$diag(\frac{q(\alpha,\lambda_1^t)}{\lambda_1^t}, ..., \frac{q(\alpha,\lambda_t^t)}{\lambda_t^t})$. If we denote by $(Z_t' Z_t / NT^{3/2})^\alpha$ the regularized inverse of $(Z_t' Z_t / NT^{3/2})$, then

$$\begin{aligned} tr(M_t^\alpha) &= tr[Z_t (Z_t' Z_t / NT^{3/2})^\alpha Z_t'] / NT^{3/2} = tr[P_t D_t P_t' P_t D(\alpha) P_t'] \\ &= tr[D_t D_t(\alpha)] = \sum_{l=1}^{t} q(\alpha, \lambda_l^{t2}). \end{aligned}$$

The result follows from $0 \leq q(\alpha, \lambda_l^{t2}) \leq 1$.

(ii)

$$\begin{aligned} E_t(v_t' M_t^\alpha v_t v_t' M_t^\alpha v_t) &= \sum_i \sum_j \sum_k \sum_l m(\alpha)_{ij}^t m(\alpha)_{kl}^t E_t(v_{it} v_{jt} v_{kt} v_{lt}) \\ &= (3\sigma^4 + \kappa_4) d_t'(\alpha) d_t(\alpha) + \sigma^4 \sum_i \sum_{k \neq i} m(\alpha)_{ii}^t m(\alpha)_{kk}^t \\ &+ 2\sigma^4 \sum_i \sum_{j \neq i} m(\alpha)_{ij}^t m(\alpha)_{ij}^t \\ &= \kappa_4 d_t'(\alpha) d_t(\alpha) + \sigma^4 tr(M_t^\alpha) tr(M_t^\alpha) + 2\sigma^4 tr(M_t^\alpha M_t^\alpha) \end{aligned}$$

where $m(\alpha)_{ij}^t$ is the (i, j) element of the matrix M_t^α. Moreover,

$$E_t(v_t' M_t^\alpha v_t) = tr[M_t^\alpha E_t(v_t v_t')] = \sigma^2 tr(M_t^\alpha).$$

So that,

$$\begin{aligned} var_t(v_t' M_t^\alpha v_t) &= E_t(v_t' M_t^\alpha v_t v_t' M_t^\alpha v_t) - E_t(v_t' M_t^\alpha v_t) E_t(v_t' M_t^\alpha v_t) \\ &= \kappa_4 d_t'(\alpha) d_t(\alpha) + 2\sigma^4 tr(M_t^\alpha M_t^\alpha). \end{aligned}$$

By definition $d_t'(\alpha) d_t(\alpha)$= $\sum_i m(\alpha)_{ii}^{t2} = tr(M_t^{\alpha 2}) \leq t$ so that $var_t(v_t' M_t^\alpha v_t)$ $\leq (\kappa_4 + 2\sigma^4) t$ and the result follows by the law of total variance.

(*iii*) By the law of iterated expectations, the expectation of $v_t' M_t^\alpha v_{t+j}$ is null for $j > 0$, so that $Var(v_t' M_t^\alpha v_{t+j}) = E(v_t' M_t^\alpha v_{t+j} v_{t+j}' M_t^\alpha v_t)$. Conditioning on t, it follows that

$$\begin{aligned} E_t(v_t' M_t^\alpha v_{t+j} v_{t+j}' M_t^\alpha v_t) &= E_t[tr(M_t^\alpha v_{t+j} v_{t+j}' M_t^\alpha v_t v_t')] \\ &= tr[M_t^\alpha E_t(v_{t+j} v_{t+j}') M_t^\alpha E_t(v_t v_t')] \\ &= \sigma^4 tr[M_t^{\alpha 2}] \\ &\leq \sigma^4 t. \end{aligned}$$

The result of (*iii*) follows by taking the expectation of both sides of the inequality.

(*iv*) $Cov(v_t' M_t^\alpha v_{t+k}, v_{t+k}' M_{t+k}^\alpha v_{t+k}) = E(v_{t+k}' M_{t+k}^\alpha v_{t+k} v_{t+k}' M_t^\alpha v_t)$

$$\begin{aligned} E_{t+k}(v_{t+k}' M_{t+k}^\alpha v_{t+k} v_{t+k}' M_t^\alpha v_t) &= E_{t+k}(v_{t+k}' M_{t+k}^\alpha v_{t+k} v_{t+k}') M_t^\alpha v_t \\ &= \sum_l \sum_i \sum_j m(\alpha)_{ij}^{t+k} E_{t+k}(v_{it+k} v_{jt+k} v_{lt+k}) M_t^\alpha v_t \\ &= \kappa_3 d_{t+k}'(\alpha) M_t^\alpha v_t \end{aligned}$$

where the last equality comes from $E_{t+k}(v_{it+k} v_{jt+k} v_{lt+k}) = \kappa_3$ if $l = i = j$ and 0 otherwise. We have just proved that $E(v_{t+k}' M_{t+k}^\alpha v_{t+k} v_{t+k}' M_t^\alpha v_t) = E(\kappa_3 d_{t+k}'(\alpha) M_t^\alpha v_t)$.
Moreover, by Cauchy-Schwarz inequality,

$$(d_{t+k}'(\alpha) M_t^\alpha v_t)^2 \leq (d_{t+k}'(\alpha) d_{t+k}(\alpha))(v_t' M_t^{\alpha 2} v_t)$$

Since $d_{t+k}'(\alpha) d_{t+k}(\alpha) \leq tr[M_{t+k}^{\alpha 2}] \leq t + k$ and $E(v_t' M_t^{\alpha 2} v_t) \leq \sigma^2 tr[M_t^{\alpha 2}]$ $\leq \sigma^2 t$, by taking expectation of the previous inequality, we have $E[(d_{t+k}'(\alpha) M_t^\alpha v_t)^2] \leq [(t+k)\sigma]^2 tr[M_t^{\alpha 2}]$. The result (*iv*) follows by noting that $[E(d_{t+k}'(\alpha) M_t^\alpha v_t)]^2 \leq E\left[(d_{t+k}'(\alpha) M_t^\alpha v_t)^2\right]$.

Lemma 3. Let $\tilde{v}_{tT} = \frac{1}{T-t}(\phi_{T-t} v_t + ... + \phi_1 v_{T-1})$ and $\phi_j = \frac{1-\delta^j}{1-\delta}$. If $N \to \infty$, $T \to \infty$ and $\alpha \to 0$, then

(*i*)

$$\frac{1}{NT} \sum_{t=1}^{T-1} E(w_{t-1}'[M_t - M_t^\alpha] w_{t-1}) = o(1),$$

(ii)

$$\frac{1}{NT}\sum_{t=1}^{T-1} E(w'_{t-1}[M_t - (M_t^\alpha)^2]w_{t-1}) = o(1).$$

(iii) Let us define $\bar{v}_{tT}$ =$(v_t + ... + v_T)/(T - t + 1)$. If $\ln T/\alpha NT \to 0$, then $Var\,(\Upsilon^\alpha_{21NT}) \to 0$ and $Var\,(\Upsilon^\alpha_{22NT}) \to 0$ where

$$\Upsilon^\alpha_{21NT} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} \tilde{v}'_{tT} M_t^\alpha v_t, \quad \text{(A.1)}$$

$$\Upsilon^\alpha_{22NT} = -\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} \tilde{v}'_{tT} M_t^\alpha \bar{v}_{tT}. \quad \text{(A.2)}$$

Moreover,

$$V\left[\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t\tilde{v}'_{tT} M_t^\alpha v_t^*\right] = O\left(\frac{\ln T}{\alpha NT}\right)$$

Proof of Lemma 3.

(i) Let us define $W = (w'_0, ..., w'_{T-2})'$, then

$$\frac{1}{NT}\sum_{t=1}^{T-1} w'_{t-1}[M_t - M_t^\alpha]w_{t-1} = \frac{1}{NT}W'Z[K_N^{-1} - K_N^\alpha]Z'W/NT^{3/2}.$$

By eigenvalues-eigenvectors decomposition, we can write $K_N^{-1} = P'_N D_N^{-1} P_N$ and $K_N^\alpha = P'_N D_N^\alpha P_N$ with $D_N^\alpha = diag[\frac{\widehat{q}_1}{\widehat{\lambda}_1}......\frac{\widehat{q}_{\bar{q}}}{\widehat{\lambda}_{\bar{q}}}]$ where $\frac{\widehat{q}_l}{\widehat{\lambda}_l}$ is a notation for $q(\alpha, \hat{\lambda}_l^2)/\hat{\lambda}_l$. Let $U_N = P_N Z'W/\sqrt{N}T^{3/4}$ a $\bar{q} \times 1$ vector, then

$$\begin{aligned}
\frac{1}{NT}W'Z[K_N^{-1} - K_N^\alpha]Z'W/NT^{3/2} &= \frac{1}{NT}W'ZP'_N[D_N^{-1} - D_N^\alpha]P_{Nn}Z'W/NT^{3/2} \\
&= \frac{1}{NT}U'_N[D_N^{-1} - D_N^\alpha]U_n \\
&= \frac{1}{NT}\sum_{l=1}^{\bar{q}}(1 - \widehat{q}_l)\frac{U_{N,l}^2}{\widehat{\lambda}_l} \\
&\le sup_{\hat{\lambda}_l}(1 - \widehat{q}_l)\frac{1}{NT}\sum_{l=1}^{\bar{q}}\frac{U_{N,l}^2}{\widehat{\lambda}_l} \\
&\le sup_{\widehat{\lambda}_l}(1 - \widehat{q}_l)\frac{1}{NT}W'ZK_N^{-1}Z'W/NT^{3/2}.
\end{aligned}$$

As $\hat{q}_l$ is between 0 and 1 so that $sup_{\widehat{\lambda}_l}(1-\hat{q}_l)$ is bounded by 1. Moreover,

$$\frac{1}{NT}E\left(W'ZK_N^{-1}Z'W/NT^{3/2}\right) = \frac{1}{NT}E\left(\sum_{t=1}^{T-1} w'_{t-1}M_t w_{t-1}\right) < \infty.$$

We have just proved that

$$\frac{1}{NT}E\left[\sum_{t=1}^{T-1} w'_{t-1}[M_t - M_t^{\alpha}]w_{t-1}\right] < \infty.$$

Following Groetsch (1993), we may in passing to the limit as $\alpha \to 0$, interchange the limit and summation, giving

$$\lim_{\alpha\to 0}\frac{1}{NT}E\left[\sum_{t=1}^{T-1} w'_{t-1}[M_t - M_t^{\alpha}]w_{t-1}\right] = 0.$$

(ii) The proof of this result uses the same argument as before noting that

$$E\left(\frac{1}{NT}\sum_{t=1}^{T-1} w'_{t-1}[M_t - M_t^{\alpha 2}]w_{t-1}\right) \leq E\left(\sup_{\hat{\lambda}_l}(1-\widehat{q}_l^2)\frac{1}{NT}\sum_{t=1}^{T-1} w'_{t-1}M_t w_{t-1}\right) < \infty.$$

(iii)

$$Var(\Upsilon_{21NT}^{\alpha}) = \frac{1}{NT}Var\left[\sum_{t=1}^{T-1}\frac{1}{T-t}v'_tM_t^{\alpha}(\phi_{T-t}v_t + ... + \phi_1 v_{T-1})\right] = a_{0NT}^{\alpha} + a_{1NT}^{\alpha}.$$

where a_{0NT}^{α} and a_{1NT}^{α} have the same form as a_{0NT} and a_{1NT} of AA(2003) but with M_t^{α} instead of M_t. First, consider a_{0NT}^{α}.

$$a_{0NT}^{\alpha} = \frac{1}{NT}\sum_{t=1}^{T-1}\frac{1}{(T-t)^2}[\phi_{T-t}^2 Var(v'_tM_t^{\alpha}v_t) + ... + \phi_1^2 Var(v'_tM_t^{\alpha}v_{T-1})].$$

Using Lemma 2 (i)-(iii), we can note that

$$Var_t(v'_tM_t^{\alpha}v_t) \leq (\kappa^4 + 2\sigma^4)tr[M_t^{\alpha 2}]$$

so that

$$Var(v'_tM_t^{\alpha}v_t) = EVar_t(v'_tM_t^{\alpha}v_t) + VE_t(v'_tM_t^{\alpha}v_t) \leq (\kappa^4 + 2\sigma^4)Etr[M_t^{\alpha 2}].$$

Moreover, for $j > 0$, we have

$$Var(v_t' M_t^\alpha v_{t+j}) = \sigma^4 Etr[M_t^{\alpha 2}].$$

Hence,

$$\begin{aligned}
a_{0NT}^\alpha &= \frac{1}{NT}\sum_{t=1}^{T-1}\frac{1}{1-\delta^2}\frac{Etr[M_t^{\alpha 2}]}{(T-t)^2}[\kappa_4 + 2\sigma^4 + (T-t-1)\sigma^4] \\
&\leq \frac{C}{NT}\sum_{t=1}^{T-1}\frac{Etr[M_t^{\alpha 2}]}{(T-t)} \\
&\leq \frac{C}{NT}\sum_{t=1}^{T-1}\frac{Etr[M_t^{\alpha}]}{(T-t)} \\
&\leq \frac{C}{NT}\sum_{t=1}^{T-1} Etr[M_t^{\alpha}]
\end{aligned}$$

for some constant $C > 0$ and we can conclude that

$$a_{0NT}^\alpha = O\left(\frac{1}{\alpha NT}\right).$$

Now looking at a_{1NT}^α, we have

$$a_{1NT}^\alpha = \frac{2}{NT}\sum_{t=1}^{T-2}\left[\sum_{j=1}^{T-t-1}\frac{\phi_{T-t-j}^2 cov(v_t' M_t^\alpha v_{t+j}, v_{t+j}' M_{t+j}^\alpha v_{t+j})}{(T-t-j)(T-t)}\right]$$

Using Lemma 2 (iv), we have

$$\begin{aligned}
|a_{1NT}^\alpha| &= \left|\frac{2}{NT}\sum_{t=1}^{T-2}\left[\sum_{jk=1}^{T-t-1}\frac{\phi_{T-t-j}^2 cov(v_t' M_t^\alpha v_{t+j}, v_{t+j}' M_{t+j}^\alpha v_{t+j})}{(T-t-j)(T-t)}\right]\right| \\
&\leq \frac{1}{(1-\delta)^2}\frac{2}{NT}\left|\sum_{t=1}^{T-2}\frac{1}{(T-t)}\left[\sum_{j=1}^{T-t-1}\frac{\kappa_3 E(d_{t+j}(\alpha) M_t^\alpha v_t)}{T-t-j}\right]\right| \\
&\leq \frac{\sigma^2}{(1-\delta)^2}\frac{2\kappa_3}{NT}\sum_{t=1}^{T-2}\frac{\sqrt{EtrM_t^{\alpha 2}}}{(T-t)}\left[\sum_{j=1}^{T-t-1}\frac{\sqrt{t+j}}{T-t-j}\right] \\
&\leq \frac{\kappa_3\sigma}{NT}\sum_{t=1}^{T-2} Etr[M_t^\alpha] O(\ln T)
\end{aligned}$$

so that

$$a_{1NT}^{\alpha} = O\bigg(\ln T/(\alpha NT) \bigg)$$

This allows us to conclude that

$$Var(\Upsilon_{21NT}^{\alpha}) = O\bigg(\ln T/(\alpha NT) \bigg)$$

We now look at the term Υ_{22NT}^{α}.

$$Var(\Upsilon_{22NT}^{\alpha}) = b_{0NT}^{\alpha} + b_{1NT}^{\alpha}.$$

where using arguments similar to those of AA (2003) (A72) and (A73) and Okui (2009),

$$\begin{aligned} b_{0NT}^{\alpha} &= \frac{1}{NT} \sum_{t=1}^{T-1} Var(\tilde{v}_{tT}' M_t^{\alpha} \bar{v}_{tT}) \\ &= O\bigg(\frac{1}{NT} \sum_{t=1}^{T-1} \frac{Etr(M_t^{\alpha 2})}{(T-t)^2} \bigg) \\ &= O\bigg(\frac{1}{\alpha NT} \bigg) \end{aligned}$$

and

$$\begin{aligned} |b_{1NT}^{\alpha}| &\leq \frac{2}{NT} \sum_{s} \sum_{t>s} |cov(\tilde{v}_{tT}' M_t^{\alpha} \bar{v}_{tT}, \tilde{v}_{sT}' M_s^{\alpha} \bar{v}_{sT})| \\ &= \frac{1}{NT} O\left(\sum_{s} \sum_{t>s} \frac{(Etr[M_t^{\alpha 2}])^{1/2}}{T-t} \frac{(Etr[M_s^{\alpha 2}])^{1/2}}{T-s} \right) \end{aligned}$$

But, for $s < t$

$$Etr[M_t^{\alpha 2}] \leq Etr[M_s^{\alpha 2}]$$

so that

$$\begin{aligned} |b_{1NT}^{\alpha}| &\leq \frac{C}{NT} \sum_{t=1}^{T-1} \frac{Etr[M_t^{\alpha 2}]}{T-t} \sum_{s=1}^{T-1} \frac{1}{T-s} \\ &= O\bigg(\frac{\ln(T)}{\alpha NT} \bigg) \end{aligned}$$

and finally

$$Var(\Upsilon^{\alpha}_{22NT}) = O\bigg(\frac{\ln(T)}{\alpha NT}\bigg).$$

To end the proof of (iii), we note that

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t\tilde{v}'_{tT}M^{\alpha}_t v^*_t = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\tilde{v}'_{tT}M^{\alpha}_t v_t - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\tilde{v}'_{tT}M^{\alpha}_t\bar{v}_{tT}$$

because $v^*_t = (v_t - \overline{v}_{tT})/c_t$. Hence,

$$Var\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t\tilde{v}'_{tT}M^{\alpha}_t v^*_t\right) = Var(\Upsilon^{\alpha}_{21NT})+Var(\Upsilon^{\alpha}_{22NT})+2Cov(\Upsilon^{\alpha}_{21NT},\Upsilon^{\alpha}_{22NT})$$

Using Cauchy-Schwarz inequality, we have

$$\begin{aligned} Var\bigg[\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t\tilde{v}'_{tT}M^{\alpha}_t v^*_t\bigg] &\leq Var(\Upsilon^{\alpha}_{21NT}) + Var(\Upsilon^{\alpha}_{22NT}) \\ &\quad + 2\bigg(Var(\Upsilon^{\alpha}_{21NT})\bigg)^{1/2}\bigg(Var(\Upsilon^{\alpha}_{22NT})\bigg)^{1/2} \\ &= O\bigg(\frac{\ln(T)}{\alpha NT}\bigg). \end{aligned}$$

and provided that $\ln(T)/\alpha NT \to 0$, (iii) holds.

Lemma 4. For a matrix A, let us define the norm $||A||^2 = tr(AA')$. If Assumptions 1-3 are satisfied, then

(i) $||K_N - K|| = O_p(1/\sqrt{N})$,

(ii) $Etr[M^{\alpha}] = O(1/\alpha)$

(iii)

$$\begin{aligned} \mu^{\alpha}_{NT} &\equiv \frac{1}{\sqrt{NT}}E\left[\sum_{t=1}^{T-1} c_t\tilde{v}'_{tT}M^{\alpha}_t v^*_t\right] \\ &= \frac{\sigma^2}{1-\delta}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} trEM^{\alpha}_t\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \\ &= O\bigg(\frac{1}{\alpha\sqrt{NT}}\bigg). \end{aligned}$$

Proof of Lemma 4.

(i) Let us define $K_{N,t}$ and K_t as the $t-th$ block of the matrix K_N and K.

$$E||K_N - K||^2 = Etr[(K_N - K)^2] \\ = E\sum_{t=1}^{T-1} tr[(K_{Nt} - K_t)^2].$$

But for a given t, $tr[(K_{Nt} - K_t)^2]$ is the sum of the squares of the elements of $(K_{Nt} - K_t)$. By definition of the matrix K_{Nt} and K_t, the (a, b) element of $K_{Nt} - K_t$ is

$$\frac{\sum_i x_{ia}x_{ib}}{NT^{3/2}} - \frac{\sum_i E[x_{ia}x_{ib}]}{NT^{3/2}}$$

Hence,

$$\begin{aligned} E||K_N - K||^2 &= Etr[(K_N - K)^2] \\ &= E\sum_{t=1}^{T-1} tr[(K_{Nt} - K_t)^2] \\ &= \sum_{t=1}^{T-1}\sum_{a,b}^{t} E\left[\frac{\sum_i x_{ia}x_{ib}}{NT^{3/2}} - \frac{\sum_i E[x_{ia}x_{ib}]}{NT^{3/2}}\right]^2 \\ &= \frac{1}{N^2T^3}\sum_{t=1}^{T-1}\sum_{a,b}^{t} Var\left[\sum_i x_{ia}x_{ib}\right] \\ &= \frac{1}{NT^3}\sum_{t=1}^{T-1}\sum_{a,b}^{t} Var\left[x_{ia}x_{ib}\right]. \end{aligned}$$

But

$$\begin{aligned} Var\left[x_{ia}x_{ib}\right] &\leq E[x_{ia}^2x_{ib}^2] \\ &\leq E[x_{ia}^4]^{1/2}E[x_{ib}^4]^{1/2}. \end{aligned}$$

We now prove that $E[x_{ia}^4] < \infty$. From $x_{ia} = y_{i,a-1} = w_{i,a-1} + \mu_i$ with $\mu_i = \eta_i/(1-\delta)$ we have

$$\begin{aligned} E[x_{ia}^4] &= E[(w_{i,a-1} + \mu_i)^4] \\ &= E[w_{i,a-1}^4 + 4w_{i,a-1}^3\mu_i + 6w_{i,a-1}^2\mu_i^2 + 4w_{i,a-1}\mu_i^3 + \mu_i^4] \\ &= E[w_{i,a-1}^4] + 4E[w_{i,a-1}^3\mu_i] + 6E[w_{i,a-1}^2\mu_i^2] + 4E[w_{i,a-1}\mu_i^3] + E[\mu_i^4]. \end{aligned}$$

- $E[w_{i,a-1}^4]$ is bounded because $w_{i,a-1}$ is an $AR(1)$ and we have from Assumption 1 that $E[v_{it}^4] < \infty$.
- $E[\mu_i^4]$ is bounded from assumption 3 (η_i has fourth moments).
- $E[w_{i,a-1}^2 \mu_i^2]$ is bounded from Cauchy-Schwarz's inequality and the fact that $E[w_{i,a-1}^4]$ and $E[\mu^4]$ are bounded.
- As an AR(1), $w_{i,a-1}$ can be written as the sum of the v_{it}. From Assumption 1, η_i is independent of the v_{it} so that $E[w_{i,a-1}^3 \mu_i] = E[w_{i,a-1}\mu_i^3] = 0$.

We have just proved that $E[x_{ia}^4] < \infty$. Hence,

$$\begin{aligned}
E||K_N - K||^2 &= \frac{1}{NT^3} \sum_{t=1}^{T-1} \sum_{a,b}^{t} Var\left[x_{ia}x_{ib}\right] \\
&\leq \frac{1}{NT^3} \sum_{t=1}^{T-1} \sum_{a,b}^{t} E[x_{ia}^4]^{1/2} E[x_{ib}^4]^{1/2} \\
&\leq \frac{E[x_{ia}^4]}{NT^3} \sum_{t=1}^{T-1} \sum_{a,b}^{t} 1 \\
&\leq \frac{E[x_{ia}^4]}{NT^3} \sum_{t=1}^{T-1} t^2 = O\left(\frac{1}{N}\right).
\end{aligned}$$

(ii)

$$\begin{aligned}
Etr[M^\alpha] &= Etr[Z[Z'Z/NT^{3/2}]^\alpha Z']/NT^{3/2} \\
&= Etr['[Z'Z/NT^{3/2}]^\alpha][Z'Z/NT^{3/2}] \\
&= Etr[K_N^\alpha K_N] \\
&= E\left[\sum_{j=1}^{\bar{q}} q_j(\alpha, \hat{\lambda}_j^2)\right]
\end{aligned}$$

where $\widehat{\lambda}_j$ are the eigenvalues of the matrix K_N. From Kress (1999) and Carrasco et al. (2007, Section3.3), we have that for the three regularizations, $q(\alpha, \lambda^2) \leq C\lambda^2/\alpha$ for some positive constant C. Then we have

$$E\left[\sum_{j=1}^{\bar{q}} q_j(\hat{\lambda}_j^2,\alpha)\right] \leq \frac{C}{\alpha}E\left[\sum_{j=1}^{\bar{q}} \hat{\lambda}_j^2\right]$$
$$\leq \frac{C}{\alpha}Etr[K_N^2]$$
$$\leq \frac{C}{\alpha}E||K_N||^2.$$

We now show that $E||K_N||^2$ is bounded.

$$E||K_N||^2 = E||K_N - K + K||^2$$
$$\leq 2E||K_N - K||^2 + 2E||K||^2 = O\Big(1/N\Big) + O\Big(1\Big) = O\Big(1\Big).$$

where $E||K_N - K||^2 = O(1/N)$ comes from Lemma 1 (i) and $E||K||^2 = O(1)$ comes from Lemma 1 (ii).

(iii) By the law of iterated expectations and Equation (A47) in AA (2003), we have

$$E(c_t\tilde{v}_{tT}'M_t^{\alpha}v_t^*) = E(tr[M_t^{\alpha}v_t^*c_t\tilde{v}_{tT}'])$$
$$= tr(E[M_t^{\alpha}c_tv_t^*\tilde{v}_{tT}'])$$
$$= tr(E[M_t^{\alpha}c_tE_t(v_t^*\tilde{v}_{tT}')])$$
$$= \frac{\sigma^2 tr[E(M_t^{\alpha})]}{1-\delta}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right).$$

Hence

$$E\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t\tilde{v}_{tT}'M_t^{\alpha}v_t^*\right) = \frac{\sigma^2}{1-\delta}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} tr(E[M_t^{\alpha}])\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right).$$

But

$$\left|\frac{\sigma^2}{1-\delta}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} E[tr[M_t^{\alpha}]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right| \leq \frac{\sigma^2}{1-\delta}\frac{1}{\sqrt{NT}}\left|\sum_{t=1}^{T-1} Etr[M_t^{\alpha}]\right|$$
$$\leq \frac{\sigma^2}{1-\delta}\frac{1}{\sqrt{NT}}\left|Etr[M^{\alpha}]\right|$$
$$\leq \frac{\sigma^2}{1-\delta}\frac{1}{\sqrt{NT}}\left|Etr[M^{\alpha}]\right|.$$

The result (iii) follows from (ii).

Proof of Proposition 1.
Proof of consistency.

$$\hat{\delta}^{\alpha} - \delta = \left(\sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} v_t^*\right)\left(\sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} x_t^*\right)^{-1}.$$

According to Equation (A42) of AA(2003), we can decompose the numerator as

$$\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} v_t^* = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \psi_t w_{t-1}' M_t^{\alpha} v_t^* - \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} c_t \tilde{v}_{tT}' M_t^{\alpha} v_t^* \quad \text{(A.3)}$$

using $w_{t-1} = y_{t-1} - \mu$ with $\mu = \eta/(1-\delta)$ and $c_t = \sqrt{(T-t)/(T-t+1)}$,

$$x_t^* = \psi_t w_{t-1} - c_t \tilde{v}_{tT},$$

$$\psi_t = c_t(1 - \frac{\delta \phi_{T-t}}{T-t}),$$

$$\tilde{v}_{tT} = \frac{(\phi_{T-t} v_t + ... + \phi_1 v_{T-1})}{T-t},$$

and

$$\phi_j = \frac{1-\delta^j}{1-\delta}.$$

The expectation of the first term of the right side of (A.3) is null so that

$$E\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} v_t^*\right) = -E\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} c_t \tilde{v'}_{tT} M_t^{\alpha} v_t^*\right). \quad \text{(A.4)}$$

It follows from Lemma 4 (iii) that

$$E\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} v_t^*\right) = O\left(\frac{1}{\alpha\sqrt{NT}}\right).$$

which is $o(1)$ if $\alpha\sqrt{NT} \to \infty$. We now look at the variance of $(x^{*'} M^{\alpha} v^*)/\sqrt{NT}$. Following the decomposition (A49) in AA(2003) we can write :

$$\frac{1}{\sqrt{NT}} x^{*'} M^{\alpha} v^* = \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{t=T-1} w_{t-1}' M_t^{\alpha} v_t - \Upsilon_{11NT}^{\alpha} - \Upsilon_{12NT}^{\alpha}\right) - \left(\Upsilon_{21NT}^{\alpha} - \Upsilon_{22NT}^{\alpha}\right) \quad \text{(A.5)}$$

where Υ^{α}_{21NT} and Υ^{α}_{22NT} are defined in Equations (A.1) and (A.2) respectively and

$$\Upsilon^{\alpha}_{11NT} = \tfrac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} w'_{t-1} M^{\alpha}_t \bar{v}_{tT},$$
$$\Upsilon^{\alpha}_{12NT} = \tfrac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \frac{c_t \delta \phi_{T-t}}{T-t} w'_{t-1} M^{\alpha}_t v^*_t.$$

We have

$$Var\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} w'_{t-1} M^{\alpha}_t v_t\right) = \frac{1}{NT} \sum_{t=1}^{T-1} var(w'_{t-1} M^{\alpha}_t v_t) = \frac{\sigma^2}{NT} \sum_{t=1}^{T-1} E(w'_{t-1} M^{\alpha 2}_t w_{t-1}).$$

$$= \frac{\sigma^2}{NT} \sum_{t=1}^{T-1} E(w'_{t-1}(M^{\alpha 2}_t - M_t) w_{t-1}) + \frac{\sigma^2}{NT} \sum_{t=1}^{T-1} E(w'_{t-1} M_t w_{t-1}).$$

From AA(2003), $\frac{\sigma^2}{NT} \sum_{t=1}^{T-1} E(w'_{t-1} M_t w_{t-1}) \longrightarrow \frac{\sigma^4}{(1-\delta^2)}$.

By Lemma 3 (ii), $\frac{\sigma^2}{NT} \sum_{t=1}^{T-1} E(w'_{t-1}[M^{\alpha}_t M^{\alpha}_t - M_t] w_{t-1}) = o(1)$ and this allows us to conclude that $Var(\frac{1}{\sqrt{NT}} \sum_{t=1}^{t=T-1} w'_{t-1} M^{\alpha}_t v^*_t)$ converges to $\frac{\sigma^4}{(1-\delta^2)}$.

Now we give the order of magnitude of Υ^{α}_{11NT}, Υ^{α}_{12NT}, Υ^{α}_{21NT}, and Υ^{α}_{22NT}.

$$Var(\Upsilon^{\alpha}_{11NT}) = \frac{1}{NT} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} E(w'_{t-1} M^{\alpha}_t \bar{v}_{tT} \bar{v}_{sT} M^{\alpha}_s w_{s-1}).$$

For $t \geq s$,

$$E(w'_{t-1} M^{\alpha}_t E_t(\bar{v}_{tT} \bar{v}_{sT}) M^{\alpha}_s w'_{s-1}) = \frac{\sigma^2}{T-s+1} E(w'_{t-1} M^{\alpha}_t M^{\alpha}_s w_{s-1}).$$

$$\begin{aligned} E(w'_{t-1} M^{\alpha}_t M^{\alpha}_s w_{s-1}) &\leq [E(w'_{t-1} M^{\alpha 2}_t w_{t-1})]^{\frac{1}{2}} [E(w'_{s-1} M^{\alpha 2}_s w_{s-1})]^{\frac{1}{2}} \\ &\leq [E(w'_{t-1} M_t w_{t-1})]^{\frac{1}{2}} [E(w'_{s-1} M_s w_{s-1})]^{\frac{1}{2}} \\ &\leq [E(w'_0 M_1 w_0)]^{\frac{1}{2}} [E(w'_0 M_1 w_0)]^{\frac{1}{2}} \\ &\leq E(w'_0 M_1 w_0). \end{aligned}$$

By similar calculations as in AA(2003), we have that Var($\Upsilon^{\alpha}_{11NT}) \longrightarrow 0$.
Next, following (A60) from AA(2003), we have

$$\begin{aligned}
Var(\Upsilon^{\alpha}_{12NT}) &= \frac{1}{NT} var(\sum_{t=1}^{T-1} \frac{c_t \delta \phi_{T-t}}{T-t} w'_{t-1} M^{\alpha}_t v^*_t) \\
&= \frac{1}{NT} \sum_{t=1}^{T-1} \frac{c_t^2 \delta^2 \phi^2_{(T-t)}}{(T-t)(T-t+1)} var(w'_{t-1} M^{\alpha}_t v^*_t) \\
&= \frac{\sigma^2}{NT} \sum_{t=1}^{T-1} \frac{c_t^2 \delta^2 \phi^2_{(T-t)}}{(T-t)(T-t+1)} E(w'_{t-1} M^{\alpha 2}_t w_{t-1}) \\
&\leq \frac{\sigma^2}{NT} \sum_{t=1}^{T-1} \frac{c_t^2 \delta^2 \phi^2_{(T-t)}}{(T-t)(T-t+1)} E(w'_{t-1} M_t w_{t-1}) \longrightarrow 0.
\end{aligned}$$

The last inequality comes from the fact that $M_t - M^{\alpha}_t M^{\alpha}_t$ is non-negative definite so that $E(w'_{t-1} M^{\alpha}_t M^{\alpha}_t w_{t-1}) \leq E(w'_{t-1} M_t w_{t-1})$. Moreover, from Lemma 3(iii), the variance of $\Upsilon^{\alpha}_{21NT} - \Upsilon^{\alpha}_{22NT}$ goes to 0 if $\ln T/\alpha NT \to 0$.
Summing up, we have that $Var(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} w'_{t-1} M^{\alpha}_t v^*_t)$ goes to $\frac{\sigma^4}{(1-\delta^2)}$, and each of Υ^{α}_{11NT}, Υ^{α}_{12NT}, Υ^{α}_{21NT}, Υ^{α}_{22NT} have variance going to zero, so that the variance of $\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} x^{*'}_t M^{\alpha}_t v^*_t$ converges to $\frac{\sigma^4}{(1-\delta^2)}$ as N and T go to infinity, α goes to zero and $\ln(T)/\alpha NT \to 0$. The expectation of $(x^{*'} M^{\alpha} v^*)/\sqrt{NT}$ goes to zero and its variance has a finite limit so that $(x^{*'} M^{\alpha} v^*)/\sqrt{NT}$ converges in mean square to zero and then in probability.
Turning to the denominator, we have :

$$\begin{aligned}
\frac{1}{NT} x^{*'} M^{\alpha} x^* &= \frac{1}{NT} \sum_{t=1}^{T-1} x^{*'}_t M^{\alpha}_t x^*_t \\
&= \frac{1}{NT} \sum_{t=1}^{T-1} \psi_t^2 w'_{t-1} M^{\alpha}_t w_{t-1} - \frac{2}{NT} \sum_{t=1}^{T-1} c_t \psi_t w'_{t-1} M^{\alpha}_t \tilde{v}_{tT} + \frac{1}{NT} \sum_{t=1}^{T-1} c_t^2 \tilde{v'}_{tT} M^{\alpha}_t \tilde{v}_{tT}.
\end{aligned}$$

We can write the first term in the following way :

$$\frac{1}{NT} \sum_{t=1}^{T-1} \psi_t^2 w'_{t-1} M^{\alpha}_t w_{t-1} = \frac{1}{NT} \sum_{t=1}^{T-1} \psi_t^2 w'_{t-1} M_t w_{t-1} - \frac{1}{NT} \sum_{t=1}^{T-1} \psi_t^2 w'_{t-1} (M_t - M^{\alpha}_t) w_{t-1}$$

From Lemma C2 of AA(2003) and $\psi_t^2 = O\left(1 - 1/(T-t)\right)$, when T goes to infinity and regardless of whether N goes to infinity or not, we have

$$\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}M_t w_{t-1} \xrightarrow{m.s.} \frac{\sigma^2}{(1-\delta^2)}.$$

By Lemma 3(i), $\frac{1}{NT}\sum_{t=1}^{T-1} w'_{t-1}[M_t - M_t^\alpha]w_{t-1} = o_p(1)$. As a result, similarly to AA(2003), we have that the limit of $\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}M_t^\alpha w_{t-1}$ is $\frac{\sigma^2}{(1-\delta^2)}$.

$\frac{2}{NT}\sum_{t=1}^{T-1} c_t\psi_t w'_{t-1}M_t^\alpha \tilde{v}_{tT}$ is identical to Υ_{11NT}^α and is $o_p(1)$.

Looking at $(\sum_{t=1}^{T-1} c_t^2 \tilde{v'}_{tT} M_t^\alpha \tilde{v}_{tT})/NT$ and using the fact that $E[c_t^2\tilde{v}_{itT}]$ is bounded, we have that

$$\begin{aligned} E\left(\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2\tilde{v'}_{tT}M_t^\alpha\tilde{v}_{tT}\right) &= \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 E\{tr[M_t^\alpha]E_t(\tilde{v'}_{tT}\tilde{v}_{tT})\} \\ &= \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 E\Big[tr[M_t^\alpha]E_t(\tilde{v}_{itT}^2)\Big] \\ &\le \frac{C}{NT}E\left(\sum_{t=1}^{T-1}[tr(M_t^\alpha)]\right) = O\left(\frac{1}{\alpha NT}\right) \end{aligned}$$

where the last equality comes from Lemma 4 (ii). By Markov's inequality,

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2\tilde{v'}_{tT}M_t^\alpha\tilde{v}_{tT} = O\left(\frac{1}{\alpha NT}\right) \tag{A.6}$$

which is $o(1)$ if $\alpha\sqrt{NT}\to\infty$.

This ends the proof that $(x^{*'}M^\alpha x^*)/NT$ tends to $\frac{\sigma^2}{(1-\delta^2)}$ in probability, hence this term is bounded. Summing up, we have that $(x^{*'}M^\alpha v^*)/NT$ converges to 0 in probability and $(x^{*'}M^\alpha x^*)/NT$ is bounded so that the regularized estimator is consistent.

Proof of the asymptotic normality.
From (A.4) and Lemma 4(iii), we have

$$\mu_{NT}^{\alpha} = E((x^{*'} M^{\alpha} v^{*})/\sqrt{NT}) = \frac{\sigma^2}{1-\delta^2}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} E[tr(M_t^{\alpha})]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right).$$

From (A.5) and since the variances of Υ_{11NT}^{α}, Υ_{12NT}^{α}, Υ_{22NT}^{α} and Υ_{21NT}^{α} go to zero, we obtain

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} v_t^{*} - \mu_{NT}^{\alpha} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}' M_t^{\alpha} v_t + o_p(1).$$

The first term of the right hand side can be rewritten as

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}' M_t^{\alpha} v_t = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}' M_t v_t - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}'[M_t - M_t^{\alpha}] v_t$$

Let us denote $h = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}'[M_t - M_t^{\alpha}]v_t$. By the law of iterated expectations, $E(h) = 0$. $Var(h) = \frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E(w_{t-1}'[M_t - M_t^{\alpha}]^2 w_{t-1})$. By Lemma 3 (i), we have $Var(h) = o(1)$ so that $h = o_p(1)$.
From AA(2003), $\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}' M_t v_t \xrightarrow{d} N(0, \frac{\sigma^2}{1-\delta^2})$ and we proved that $(x^{*'} M^{\alpha} x^{*})/NT$ tends to $\frac{\sigma^2}{(1-\delta^2)}$ in probability, so that by Slutsky's theorem

$$\left(\frac{x^{*'} M^{\alpha} x^{*}}{NT}\right)^{-1}\left[\frac{1}{\sqrt{NT}} x^{*'} M^{\alpha} v^{*} - \mu_{NT}^{\alpha}\right] \xrightarrow{d} N(0, 1-\delta^2)$$

or

$$\sqrt{NT}(\hat{\delta}^{\alpha} - \delta) - \left(\frac{x^{*'} M^{\alpha} x^{*}}{NT}\right)^{-1}\mu_{NT}^{\alpha} \xrightarrow{d} N(0, 1-\delta^2)$$

From Lemma 4 (iii), $\mu_{NT}^{\alpha} = o(1)$, hence the bias vanishes and this ends the proof of asymptotic normality.

Preliminary results for the proof of Proposition 2
Let

$$\Delta_{\alpha} = \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t^{\alpha})^2 w_{t-1}].$$

Lemma 5. *If assumptions 1-3 are satisfied, then*

(i) $\Delta_\alpha = o(1)$,

(ii) $\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \psi_t w'_{t-1}(I - M_t^\alpha)v_t^* = O_p(\Delta_\alpha^{1/2})$,

(iii)

$$V\left(\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 [w'_{t-1} M_t^\alpha w_{t-1}]\right) = O\left(\frac{1}{NT}\right).$$

(iv)

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t \psi_t w'_{t-1} M_t^\alpha \tilde{v}_{tT} = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

(v) $H = \sum_{t=1}^{T-1} \psi_t^2 E[w'_{t-1}w_{t-1}]/NT = \frac{\sigma^2}{1-\delta^2} \frac{\sum_{t=1}^{T-1}\psi_t^2}{T} = O(1)$ and $h = \sum_{t=1}^{T-1} \psi_t w'_{t-1} v_t^* / \sqrt{NT} = O_p(1)$.

Proof of Lemma 5.

(i) Noting that $\psi_t^2 \leq 1$, this term can be omitted in the proof.

$$\begin{aligned}
\frac{1}{NT}\sum_{t=1}^{T-1} E[w'_{t-1}(I - M_t^\alpha)^2 w_{t-1}] &= \frac{1}{NT}\sum_{t=1}^{T-1} E[w'_{t-1}(I - 2M_t^\alpha + M_t^{\alpha 2})w_{t-1}] \\
&= \frac{1}{NT}\sum_{t=1}^{T-1} E[w'_{t-1}(I - M_t)w_{t-1}] \\
&+ 2\frac{1}{NT}\sum_{t=1}^{T-1} E[w'_{t-1}(M_t - M_t^\alpha)w_{t-1}] \\
&- \frac{1}{NT}\sum_{t=1}^{T-1} E[w'_{t-1}(M_t - M_t^{\alpha 2})w_{t-1}].
\end{aligned}$$

From Equation (A86) in AA(2003), we have

$$\frac{1}{NT}\sum_{t=1}^{T-1} E[w'_{t-1}(I - M_t)w_{t-1}] = O\left(\frac{log(T)}{T}\right) = o(1).$$

The last two terms are also $o(1)$ using results from Lemma 3(i) and (ii).

(*ii*) The expectation of the term is 0 and its variance is

$$Var\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w'_{t-1}(I-M_t^{\alpha})v_t^*\right)=\sigma^2\Delta_\alpha$$

and the result follows from Markov's inequality.

(*iii*) From Equations (A40) and (A41) of AA(2003), we have $Var\left(\sum_{t=1}^{T-1} w'_{t-1}M_t w_{t-1}/NT\right)=$ $O\left(1/NT\right)$. We can use the same proof as in AA to establish our result given that $\psi_t \leq 1$ and M_t^{α} has eigenvalues smaller than or equal to 1.

(*iv*)

$$E\left(\frac{1}{NT}\sum_{t=1}^{T-1}c_t\psi_t w'_{t-1}M_t^{\alpha}\tilde{v}_{tT}\right)=0.$$

Now for the variance, note that

$$Var\left(\frac{1}{NT}\sum_{t=1}^{T-1}c_t\psi_t w'_{t-1}M_t^{\alpha}\tilde{v}_{tT}\right)=\frac{1}{N^2T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}c_t^2\psi_t^2 E[w'_{t-1}M_t^{\alpha}\tilde{v}_{tT}\tilde{v}'_{sT}M_s^{\alpha}w_{s-1}]$$

For $t \geq s$,

$$E(w'_{t-1}M_t^{\alpha}\tilde{v}_{tT}\tilde{v}'_{sT}M_s^{\alpha}w_{s-1})=E(w'_{t-1}M_t^{\alpha}E_t(\tilde{v}_{tT}\tilde{v}'_{sT})M_s^{\alpha}w_{s-1})$$

But

$$E_t(\tilde{v}_{tT}\tilde{v}'_{sT})=\frac{\sigma^2}{(T-t)(T-s)}[\phi_{T-s}^2+....+\phi_1^2]\leq\frac{\sigma^2}{(T-t)}$$

so that

$$E(w'_{t-1}M_t^{\alpha}\tilde{v}_{tT}\tilde{v}'_{sT}M_s^{\alpha}w_{s-1})\leq\frac{\sigma^2}{(T-t)}E(w'_{t-1}M_t^{\alpha}M_s^{\alpha}w_{s-1})$$

Now by Cauchy-Schwarz inequality's inequality

$$E(w'_{t-1}M_t^{\alpha}M_s^{\alpha}w_{s-1})\leq[Ew'_{t-1}M_t^{\alpha 2}w'_{t-1}]^{1/2}[Ew'_{s-1}M_s^{\alpha 2}w'_{s-1}]^{1/2}$$

Then,

$$
\begin{aligned}
E(w'_{t-1}M_t^{\alpha}\tilde{v}_{tT}\tilde{v}'_{sT}M_s^{\alpha}w_{s-1}) &\leq \frac{\sigma^2}{(T-t)}[Ew'_{t-1}M_t^{\alpha}M_t^{\alpha}w'_{t-1}]^{1/2}[Ew'_{s-1}M_s^{\alpha}M_s^{\alpha}w'_{s-1}]^{1/2} \\
&\leq \frac{\sigma^2}{(T-t)}[Ew'_{t-1}M_t^{\alpha}w'_{t-1}]^{1/2}[Ew'_{s-1}M_s^{\alpha}w'_{s-1}]^{1/2} \\
&\leq \frac{\sigma^2}{(T-t)}[Ew'_{t-1}M_t w'_{t-1}]^{1/2}[Ew'_{s-1}M_s w'_{s-1}]^{1/2} \\
&\leq \frac{\sigma^2}{(T-t)}E(w'_0 M_1 w_0) \leq \frac{\sigma^2 N}{(T-t)}E(w_{i0}^2).
\end{aligned}
$$

Hence

$$
\begin{aligned}
Var\left(\frac{1}{NT}\sum_{t=1}^{T-1} c_t\psi_t w'_{t-1}M_t^{\alpha}\tilde{v}_{tT}\right) &\leq \frac{\sigma^2}{(NT^2)}E(w_{i0}^2)\{(\frac{1}{T-1}) + ... + \frac{1}{2} \\
&+ \frac{2(T-2)}{T-1} + ... + \frac{2}{1}\} \\
&= O(\frac{T}{NT^2}) \ = O(\frac{1}{NT}).
\end{aligned}
$$

so that (iv) holds by Markov's Inequality.

(v)

$$
\begin{aligned}
H &= \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}w_{t-1}] \\
&= \frac{1}{T}\sum_{t=1}^{T-1}\psi_t^2 E[w_{i,t-1}^2] \\
&= \frac{\sigma^2}{1-\delta^2}\frac{1}{T}\sum_{t=1}^{T-1}\psi_t^2
\end{aligned}
$$

and the result follows from the fact $\sum_{t=1}^{T-1}\psi_t^2/T \to 1$.
Regarding h, we have $E(h) = 0$ and $Var(h) = \sigma^2 H$ so that $h = O_p(1)$ since $H = O(1)$.

For completeness, we reproduce here a lemma from Okui (2009) which is essential to derive the higher-order expansion of the MSE.

Lemma 6. (Lemma 2 of Okui (2009))

Let $\rho_\alpha = trS(\alpha)$. Suppose that an estimator $\hat{\delta}$ has a decomposition $\sqrt{NT}\left(\hat{\delta}-\delta\right) = \widehat{H}^{-1}\hat{h},\ \hat{h} = h + T^h + Z^h,\ \widehat{H} = H + T^H + Z^H,$

$\left(h+T^h\right)\left(h+T^h\right)' - hh'H^{-1}T^{H\prime} - T^H H^{-1}hh' = \hat{A} + Z^A,$

such that $T^h = o_p(1),\ h = O_p(1),\ H = O_p(1),$ *the determinant of* H *is bounded away from zero with probability approaching 1,* $\rho_\alpha = o_p(1)$,

$\left\|T^H\right\|^2 = o_p(\rho_\alpha),\ \left\|T^h\right\|\left\|T^H\right\| = o_p(\rho_\alpha),\ \left\|Z^h\right\| = o_p(\rho_\alpha),\ \left\|Z^H\right\| = o_p(\rho_\alpha),$ $\left\|Z^A\right\| = o_p(\rho_\alpha),$

$E\left(\hat{A}\right) = \sigma^2 H + HS(\alpha)H + o_p(\rho_\alpha).$

Then, the decomposition (1.3) *holds for* $\hat{\delta}$.

Proof of Proposition 2.

Let $\rho_\alpha = S(\alpha)$. Notice that

$$\rho_\alpha \geq \frac{(1+\delta)^2}{NT}\left\{\sum_{t=1}^{T-1} Etr[M_t^\alpha]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2 \tag{A.7}$$

and

$$\rho_\alpha \geq \frac{(1-\delta^2)^2}{\sigma^2}\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t^\alpha)^2 w_{t-1}]$$

First, we establish the rate of the RHS of (A.7). Because

$$\phi_j = \frac{1-\delta^j}{1-\delta} \leq \frac{1}{1-\delta}$$

we have $0 \leq \frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} \leq \frac{2}{1-\delta}$ and

$$\begin{aligned}
&\frac{(1+\delta)^2}{NT}\left\{\sum_{t=1}^{T-1} Etr[M_t^\alpha]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2 \\
\leq\ & \frac{(1+\delta)^2}{NT(1-\delta)^2}\left(\sum_{t=1}^{T-1} Etr[M_t^\alpha]\right)^2 \\
=\ & \frac{(1+\delta)^2}{NT(1-\delta)^2}\left(Etr[M^\alpha]\right)^2 \\
=\ & O(1/(\alpha^2 NT))
\end{aligned}$$

by Lemma 4 (ii). Hence, a term that is $o(1/(\alpha^2 NT))$ is necessarily $o(\rho_\alpha)$. Moreover since

$$\Delta_\alpha \geq \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]$$

and from Okui (2009), $logT/T = o(\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]/NT)$ so that $o(logT/T) = o(\rho_\alpha)$. To prove Proposition 2, we use Lemma 6 and

$$\sqrt{NT}(\hat{\delta}^\alpha - \delta) = \left(\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*'} M_t^\alpha x_t^*\right)^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*'} M_t^\alpha v_t^*\right).$$

As in Okui (2009), the numerator can be written in the following way :

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*'} M_t^\alpha v_t^* = h + T_1^h + T_2^h$$

where

$$\begin{aligned}
h &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w_{t-1}' v_t^* \\
T_1^h &= -\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w_{t-1}'(I-M_t^\alpha)v_t^* = O_p(\Delta_\alpha^{1/2}), \\
T_2^h &= -\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t \tilde{v}_{tT}' M_t^\alpha v_t^* = O_p\left(\frac{1}{\alpha\sqrt{NT}}\right)
\end{aligned}$$

where the rate for T_2^h follows from Lemmas 3(iii) and 4(iii). Moreover

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*'} M_t^\alpha x_t^* = H + T^H + \sum_{j=1}^{3} Z_j^H$$

with

$$H = \frac{\sigma^2}{1-\delta^2}\frac{1}{T}\sum_{t=1}^{T-1}\psi_t^2$$

$$T^H = -\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I - M_t^\alpha)w_{t-1}] = O_p(\Delta_\alpha),$$

$$Z_1^H = \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w'_{t-1}w_{t-1}] - H = O_p(1/\sqrt{NT}),$$

$$Z_2^H = -2\frac{1}{NT}\sum_{t=1}^{T-1}c_t\psi_t w'_{t-1}M_t^\alpha \tilde{v}_{tT} = O_p(1/\sqrt{NT}),$$

$$Z_3^H = \frac{1}{NT}\sum_{t=1}^{T-1}c_t^2\tilde{v}'_{tT}M_t^\alpha \tilde{v}_{tT} = O_p(1/\alpha NT).$$

By $1/\sqrt{NT} = o(logT/T)$ and $1/(\alpha NT) = o(1/(\alpha^2 NT)$ we have that Z_j^H are $o_p(\rho_\alpha)$ for $j = 1, 2, 3$ so that $\| \sum_{j=1}^{3} Z_j^H \| = o_p(\rho_\alpha)$ by triangular inequality. Moreover, we have $||T^H||||T_1^h|| = O(\Delta_\alpha/\alpha\sqrt{NT}) = o_p(\rho_\alpha)$ and $||T^H||||T_2^h|| = O(\Delta_\alpha^{3/2}) = o_p(\rho_\alpha)$ so that we can conclude that $||T^H||||T_1^h + T_2^h|| = o_p(\rho_\alpha)$. We now apply Lemma 6 with $Z^A = 0$ and

$$\begin{aligned}\widehat{A} &= (h + T_1^h + T_2^h)^2 - 2h^2H^{-1}T^H \\ &= h^2 + (T_1^h)^2 + (T_2^h)^2 + 2hT_1^h + 2hT_2^h + 2T_1^hT_2^h - 2h^2H^{-1}T^H.\end{aligned}$$

Lemma 6 states that $S(\alpha)$ satisfies $E\left(\widehat{A}\right) = \sigma^2 H + HS(\alpha)H + o_p(\rho_\alpha)$. To calculate the expectation of $\widehat{A}$, we need to compute the expectation of each term. By the third moment condition and the independence assumption both on the error term v_{it}, we can show that $E(hT_2^h) = E(T_1^hT_2^h) = 0$. It can easily be proved that

$$\begin{aligned}&E(h^2) = \sigma^2 H, \\ &E\{(T_1^h)^2\} = \sigma^2\Delta_\alpha, \\ &E(h^2H^{-1}T^H) = E(hT_1^h) = \sigma^2 T^H.\end{aligned}$$

By Lemma 3 (iii) and Lemma 4 (iii), we have

$$\begin{aligned}
E\{(T_2^h)^2\} &= (E(T_2^h))^2 + var(T_2^h) \\
&= \frac{1}{NT}\frac{\sigma^4}{(1-\delta)^2}\left\{\sum_{t=1}^{T-1} E[tr(M_t^\alpha)]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2 \\
&+ O(\frac{(logT)^2}{N}) \\
&= \frac{1}{NT}\frac{\sigma^4}{(1-\delta)^2}\left\{\sum_{t=1}^{T-1} E[tr(M_t^\alpha)]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2 \\
&+ o_p(\rho_\alpha)
\end{aligned}$$

where the third equality comes from the fact that $\ln(T)^2/N = o_p\left(1/\left(\alpha^2 NT\right)\right) = o_p(\rho_\alpha)$ provided $\alpha \ln(T)\sqrt{T} \to 0$. Finally

$$\begin{aligned}
E(\hat{A}) = &\frac{1}{NT}\frac{\sigma^4}{(1-\delta)^2}\left\{\sum_{t=1}^{T-1} E[tr(M_t^\alpha)]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2 \\
&+ \sigma^2 H + \sigma^2\Delta_\alpha + o_p(\rho_\alpha).
\end{aligned}$$

And therefore

$$\begin{aligned}
S(\alpha) &= (\lim_{T\to\infty} H)^{-2}\left\{\frac{1}{NT}\frac{\sigma^4}{(1-\delta)^2}\left\{\sum_{t=1}^{T-1} E[tr(M_t^\alpha)]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2 + \sigma^2\Delta_\alpha\right\} \\
&= \frac{(1+\delta)^2}{NT}\left\{\sum_{t=1}^{T-1} E[tr(M_t^\alpha)]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2 \\
&\quad + \frac{(1-\delta^2)^2}{\sigma^2}\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t^\alpha)^2 w_{t-1}].
\end{aligned}$$

using the fact that $\lim_{T\to\infty} H = \sigma^2/(1-\delta^2)$. This ends the proof of Proposition 2.

Preliminary results for the proof of Proposition 3

The following lemma will be used in the proof of Proposition 3.

Lemma 7.

(*i*) $tr[K_N^2] - tr[K^2] = O_p(1/\sqrt{N})$,

$(ii)\ E\left[\left[\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)-E[tr(M_t^\alpha)]\right)\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)\right]^2\right]=O\left(\frac{1}{\alpha^2 N}\right).$

Proof of Lemma 7.

(i)

$$\begin{aligned} tr[K_N^2]-tr[K^2] &= ||K_N||^2-||K||^2 \\ &=(||K_N||+||K||)(||K_N||-||K||) \end{aligned}$$

From Lemma 1(ii) and Lemma 4(i), we have $||K_N||+||K||=O_p(1)$. Moreover,

$$||K_N||-||K||\leq ||K_N-K||=O_p\left(1/\sqrt{N}\right)$$

by Lemma 4(i) so that we have $tr[K_N^2]-tr[K^2]=O(1/\sqrt{N})$.

(ii) As $0\leq \frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\leq C$, it is sufficient to study the term $E\left[\left[\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)-E[tr(M_t^\alpha)]\right)\right]^2\right]$. We have

$$\begin{aligned} & E\left[\left[\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)-E[tr(M_t^\alpha)]\right)\right]^2\right] \\ =\ & E\left[\left[\sum_{t=1}^{T-1}\Big(tr(Z_t\left(Z_t'Z_t\right)^\alpha Z_t')-E[tr(Z_t\left(Z_t'Z_t\right)^\alpha Z_t')]\Big)\right]^2\right] \\ =\ & E\left[\left[\sum_{t=1}^{T-1}\Big(tr((Z_t'Z_t)^\alpha Z_t'Z_t)-E[tr((Z_t'Z_t)^\alpha Z_t'Z_t)]\Big)\right]^2\right] \\ =\ & E\left[\left[tr(K_N^\alpha K_N)-trE(K_N^\alpha K_N)\right]^2\right] \\ =\ & E\left[\left\|K_N^\alpha K_N-E(K_N^\alpha K_N)\right\|^2\right] \end{aligned}$$

Moreover,

$$\begin{aligned} & E\left[\left\|K_N^\alpha K_N-E(K_N^\alpha K_N)\right\|^2\right] \\ =\ & E\left[\left\|K_N^\alpha\left(K_N-K\right)+\left(K_N^\alpha-E\left(K_N^\alpha\right)\right)K+E\left(K_N^\alpha\left(K-K_N\right)\right)\right\|^2\right] \\ \leq\ & 3E\left\|K_N^\alpha\left(K_N-K\right)\right\|^2+3E\left\|\left(K_N^\alpha-E\left(K_N^\alpha\right)\right)K\right\|^2+3\left\|E\left(K_N^\alpha\left(K-K_N\right)\right)\right\|^2. \end{aligned}$$

We have

$$
\begin{aligned}
E\left\|K_N^\alpha\left(K_N-K\right)\right\|^2 &= E\left\|K_N^\alpha\right\|^2\left\|K_N-K\right\|^2 \leq \frac{C}{\alpha^2}E\left\|K_N-K\right\|^2 = O\left(\frac{1}{\alpha^2 N}\right), \\
E\left\|\left(K_N^\alpha-E\left(K_N^\alpha\right)\right)K\right\|^2 &\leq CE\left\|K_N^\alpha-E\left(K_N^\alpha\right)\right\|^2 = CE\left\|K_N^\alpha-K^\alpha\right\|^2 = O\left(\frac{1}{N}\right), \\
\left\|E\left(K_N^\alpha\left(K-K_N\right)\right)\right\|^2 &= \left\|E\left(\left(K_N^\alpha-K^\alpha\right)\left(K-K_N\right)\right)\right\|^2 = o\left(\frac{1}{N}\right).
\end{aligned}
$$

The result follows.

Proof of Proposition 3.

We want to prove that

$$\frac{S(\hat{\alpha})}{\inf_{\alpha\in\mathcal{E}_T} S(\alpha)} \xrightarrow{P} 1$$

where $\mathcal{E}_T$ is the parameter set for a given regularization scheme. By Lemma A9 of Donald and Newey (2001), it is sufficient to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{S}(\alpha)-S(\alpha)}{S(\alpha)}\right| = o_p(1).$$

Using the fact that, $\frac{\left(1-\delta^2\right)^2}{\sigma^2}R\left(\alpha\right) \leq S\left(\alpha\right)$ and $\frac{(1+\delta)^2}{NT}\mathcal{A}\left(\alpha\right)^2 \leq S\left(\alpha\right)$, we have for some constant C

$$
\begin{aligned}
\frac{1}{C}\left|\frac{\widehat{S}(\alpha)-S(\alpha)}{S(\alpha)}\right| \leq \frac{(1+\hat{\delta})^2}{(1+\delta)^2}\left|\frac{\widehat{\mathcal{A}}\left(\alpha\right)^2-\mathcal{A}\left(\alpha\right)^2}{\mathcal{A}\left(\alpha\right)^2}\right| + \left|\frac{(1+\hat{\delta})^2-(1+\delta)^2}{(1+\delta)^2}\right| \\
+\frac{(1-\hat{\delta}^2)^2/\hat{\sigma}^2}{(1-\delta^2)^2/\sigma^2}\left|\frac{\widehat{R}(\alpha)-R(\alpha)}{R(\alpha)}\right| + \left|\frac{(1-\hat{\delta}^2)^2/\hat{\sigma}^2-(1-\delta^2)^2/\sigma^2}{(1-\delta^2)^2/\sigma^2}\right|.
\end{aligned}
$$

By the consistency of $\hat{\delta}$ and $\hat{\sigma}^2$, we just need to prove that :

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{\mathcal{A}}\left(\alpha\right)^2-\mathcal{A}\left(\alpha\right)^2}{\mathcal{A}\left(\alpha\right)^2}\right| = o_p(1),$$

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{R}(\alpha)-R(\alpha)}{R(\alpha)}\right| = o_p(1).$$

For the first equality, we have

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{\mathcal{A}}\left(\alpha\right)^2-\mathcal{A}\left(\alpha\right)^2}{\mathcal{A}\left(\alpha\right)^2}\right| = \sup_{\mathcal{E}_T}\left|\frac{\widehat{\mathcal{A}}\left(\alpha\right)-\mathcal{A}\left(\alpha\right)}{\mathcal{A}\left(\alpha\right)}\right|\left|\frac{\widehat{\mathcal{A}}\left(\alpha\right)+\mathcal{A}\left(\alpha\right)}{\mathcal{A}\left(\alpha\right)}\right|.$$

Moreover

$$\left|\frac{\widehat{\mathcal{A}}(\alpha)+\mathcal{A}(\alpha)}{\mathcal{A}(\alpha)}\right| \leq 2+\left|\frac{\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha)}{\mathcal{A}(\alpha)}\right|.$$

So it is sufficient to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha)}{\mathcal{A}(\alpha)}\right| = o_p(1).$$

$$\begin{aligned}\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha) &= \sum_{t=1}^{T-1} tr(M_t^\alpha)\left(\frac{\widehat{\phi}_{T-t}}{T-t}-\frac{\widehat{\phi}_{T-t+1}}{T-t+1}\right)-\sum_{t=1}^{T-1} E[tr[M_t^\alpha]]\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)\\ &= \sum_{t=1}^{T-1} tr(M_t^\alpha)\left(\widehat{\nu}_t-\nu_t\right)+\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)-E[tr(M_t^\alpha)]\right)\nu_t\end{aligned}$$

where $\nu_t = \frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}$ and $\widehat{\nu}_t = \frac{\widehat{\phi}_{T-t}}{T-t}-\frac{\widehat{\phi}_{T-t+1}}{T-t+1}$. We will use the following result (see Okui (2008, p.13) : For a random sequence $\{a_k\}_k$, $\sum_k E(a_k^2) = o(1)$ implies that $\sup_k a_k = o_p(1)$.

$$\begin{aligned} E\left\{\left[\sum_{t=1}^{T-1} tr(M_t^\alpha)\left(\widehat{\nu}_t-\nu_t\right)\right]^2\right\} &= E\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)\right)^2\left(\widehat{\nu}_t-\nu_t\right)^2 \\ &\quad +\sum_{t\neq s} E\left[tr\left(M_t^\alpha\right)\left(\widehat{\nu}_t-\nu_t\right)tr(M_s^\alpha)\left(\widehat{\nu}_s-\nu_s\right)\right]. \end{aligned} \tag{A.8}$$

Using $trM_t^\alpha \leq C/\alpha$ and, by the consistency of $\hat{\delta}$, $E\left(\widehat{\nu}_t-\nu_t\right)^2 = O\left(1/NT\right)$, we have

$$E\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)\right)^2\left(\widehat{\nu}_t-\nu_t\right)^2 = O\left(\frac{1}{\alpha^2 N}\right).$$

By Cauchy-Schwarz's inequality, the second term of the RHS of (A.8) is also $O(1/N\alpha^2)$. By Lemma 7(ii),

$$E\left[\left[\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)-E[tr(M_t^\alpha)]\right)\nu_t\right]^2\right] = O\left(\frac{1}{\alpha^2 N}\right)$$

We previously established that

$$\mathcal{A}(\alpha) = \sum_{t=1}^{T-1} Etr[M_t^\alpha]\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right) = O(1/\alpha).$$

Hence for any $\alpha \in \mathcal{E}_T$ (which is discrete and finite for SC and LF) we have

$$E\left|\frac{\widehat{\mathcal{A}}(\alpha) - \mathcal{A}(\alpha)}{\mathcal{A}(\alpha)}\right|^2 = O\left(\frac{1}{N}\right).$$

Now summing up over the elements of $\mathcal{E}_T$, we obtain

$$\sum_{\alpha \in \mathcal{E}_T} O\left(\frac{1}{T}\right) = O\left(\frac{T^2}{N}\right)$$

because the cardinal of $\mathcal{E}_T$ is equal to T^2. Hence, $\sup_\alpha \frac{\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha)}{\mathcal{A}(\alpha)} = o_p(1)$ provided $T^2/N \to 0$ (which is true under the condition $T^3/(N\ln(T)^2) \to 0$).

Now, we want to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{R}(\alpha) - R(\alpha)}{R(\alpha)}\right| = o_p(1). \tag{A.9}$$

We first consider the spectral cut-off regularization scheme
In this case, $(I - M_t^\alpha)^2 = (I - M_t^\alpha)$ so that

$$R(\alpha) = \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t^\alpha)w_{t-1}]$$

and

$$\widehat{R}(\alpha) = \frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*'}(I - M_t^\alpha)x_t^*.$$

Following Okui (2009), we consider the following version of estimated of $\widehat{R}(\alpha)$:

$$\widetilde{R}(\alpha) = \frac{1}{NT}\sum_{t=1}^{T-1}(\psi_t^2 w_{t-1}'w_{t-1} - x_t^{*'}M_t^\alpha x_t^*).$$

Indeed, the difference between $\widetilde{R}(\alpha)$ and $\widehat{R}(\alpha)$ does not depend on α so that maximizing the criterion with using $\widetilde{R}(\alpha)$ instead of $\widehat{R}(\alpha)$ gives the same result. By (A43) of AA(2003), $x_t^* = \psi_t w_{t-1} - c_t\widetilde{v}_{tT}$, so that

$$\begin{aligned}\widetilde{R}(\alpha) - R(\alpha) = &\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w_{t-1}'(I - M_t^\alpha)w_{t-1} - E[w_{t-1}'(I - M_t^\alpha)w_{t-1}]] \\ &+ \frac{2}{NT}\sum_{t=1}^{T-1}\psi_t c_t w_{t-1}'M_t^\alpha\widetilde{v}_{tT} - \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2\widetilde{v}'_{tT}M_t^\alpha\widetilde{v}_{tT}.\end{aligned}$$

Hence to prove (A.9), we have to prove that :

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}(I-M_t^\alpha)w_{t-1}-\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t^\alpha)w_{t-1}}{R(\alpha)}]\right|=o_p(1),$$

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t c_t w'_{t-1}M_t^\alpha\tilde{v}_{tT}}{R(\alpha)}\right|=o_p(1),$$

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\ c_t^2\tilde{v}'_{tT}M_t^\alpha\tilde{v}_{tT}}{R(\alpha)}\right|=o_p(1).$$

Noting that $w'_{t-1}(I-M_t^\alpha)w_{t-1}\geq w'_{t-1}(I-M_t)w_{t-1}$

$$\begin{aligned}
&\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}(I-M_t^\alpha)w_{t-1}-\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t^\alpha)w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t^\alpha)w_{t-1}]}\right|\\
&\leq\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}(I-M_t^\alpha)w_{t-1}-\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t^\alpha)w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t)w_{t-1}]}\right|\\
&\leq\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}w_{t-1}-\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t)w_{t-1}]}\right|\\
&+\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}M_t^\alpha w_{t-1}-\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}M_t^\alpha w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t)w_{t-1}]}\right|. \qquad (A.10)
\end{aligned}$$

Now we want to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w_{t-1}' w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}' w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]}\right| = o_p(1).$$

Note that this term does not depend on α. Moreover, from Okui (2009) in the analysis of his term Z_1^H, we have

$$\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w_{t-1}' w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}' w_{t-1}] = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Moreover from AA(2003)

$$\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}] = O_p\left(\frac{logT}{T}\right)$$

so that we can conclude that

$$\sup_{\mathcal{E}_T}\left|\frac{\sum_{t=1}^{T-1}\psi_t^2 w_{t-1}' w_{t-1}/(NT) - \sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}' w_{t-1}]/(NT)}{\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]/(NT)}\right| = O_p\left(\frac{\sqrt{T}}{\sqrt{N}\ln(T)}\right) = o_p(1). \tag{A.11}$$

We now turn our attention to (A.10). This terms depends on α. From Lemma 5(iii), we have

$$E\left[\left(\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w_{t-1}' M_t^\alpha w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}' M_t^\alpha w_{t-1}]\right)^2\right] = O\left(\frac{1}{NT}\right).$$

Summing over the elements of $\mathcal{E}_T$, we get

$$\begin{aligned}\sum_{\mathcal{E}_T} E\left[\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w_{t-1}' M_t^\alpha w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}' M_t^\alpha w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]}\right]^2 &= O\left(T^2\right)\frac{O\left(\frac{1}{NT}\right)}{O\left(\frac{(\ln T)^2}{T^2}\right)} \\ &= O\left(\frac{T^3}{N\ln(T)^2}\right).\end{aligned}$$

Then, we conclude that

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w_{t-1}'(I-M_t^\alpha)w_{t-1}-\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t^\alpha)w_{t-1}]}{R(\alpha)}\right|=o_p(1)$$

We now consider the proof of

$$\sup_{\mathcal{E}_T}\left|\frac{\sum_{t=1}^{T-1}\psi_t c_t w_{t-1}' M_t^\alpha \tilde{v}_{tT}/(NT)}{\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]/(NT)}\right|=o_p(1).$$

We have

$$E\left[\left(\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t c_t w_{t-1}' M_t^\alpha \tilde{v}_{tT}\right)^2\right]=O\left(\frac{1}{NT}\right)$$

by the proof of Lemma 5(iv). We obtain

$$\sum_{\mathcal{E}_T} E\left[\left(\frac{\sum_{t=1}^{T-1}\psi_t c_t w_{t-1}' M_t^\alpha \tilde{v}_{tT}/(NT)}{\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]/(NT)}\right)^2\right]=O\left(\frac{T^3}{N\ln(T)^2}\right)=o\left(1\right).$$

Then we can conclude

$$\sup_{\mathcal{E}_T}\left|\frac{\sum_{t=1}^{T-1}\psi_t c_t w_{t-1}' M_t^\alpha \tilde{v}_{tT}/(NT)}{\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]/(NT)}\right|=o_p(1).$$

Now, we want to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 \tilde{v}'_{tT} M_t^\alpha \tilde{v}_{tT}}{\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]/(NT)}\right|=o_p(1).$$

Following Okui (2009), we can major this term as follows

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2\tilde{v}'_{tT}M_t^\alpha\tilde{v}_{tT}}{\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t)w_{t-1}]/(NT)}\right| \leq \left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2\tilde{v}'_{tT}M_t\tilde{v}_{tT}}{\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t)w_{t-1}]/(NT)}\right| = O_p\left(\frac{T}{N}\right)$$

which is $o_p(1)$ under the assumption that $T/N \to 0$. This ends the proof of

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{R(\alpha)}-R(\alpha)}{R(\alpha)}\right| = o_p(1).$$

for the spectral cut-off regularization scheme.

We now consider the Landweber Fridman regularization scheme

The particularity here is that the matrix $I - M_t^\alpha$ is no longer idempotent. However, we have

$$(I-M_t^\alpha)^2 = I - 2M_t^\alpha + M_t^\alpha M_t^\alpha = I - \widetilde{M}_t^\alpha$$

where $\widetilde{M}_t^\alpha = 2M_t^\alpha - M_t^\alpha M_t^\alpha$. As in the case of spectral cut-off regularization scheme, let us define

$$\widetilde{R}(\alpha) = \frac{1}{NT}\sum_{t=1}^{T-1}(\psi_t^2 w'_{t-1}w_{t-1} - x_t^{*'}\widetilde{M}_t^\alpha x_t^*).$$

Since the difference between $\widetilde{R}(\alpha)$ and $\widehat{R}(\alpha)$ does not depend on α, we can prove optimality using $\widetilde{R}(\alpha)$ instead of $\widehat{R}(\alpha)$. Hence we have to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\widetilde{R}(\alpha)-R(\alpha)}{R(\alpha)}\right| = o_p(1).$$

Noting that

$$\begin{aligned}\widetilde{R}(\alpha)-R(\alpha) &= \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w'_{t-1}(I-\widetilde{M}_t^\alpha)w_{t-1} - E[w'_{t-1}(I-\widetilde{M}_t^\alpha)w_{t-1}]] \\ &+ \frac{2}{NT}\sum_{t=1}^{T-1}\psi_t c_t w'_{t-1}\widetilde{M}_t^\alpha\tilde{v}_{tT} - \frac{1}{NT}\sum_{t=1}^{T-1}c_t^2\tilde{v}'_{tT}\widetilde{M}_t^\alpha\tilde{v}_{tT},\end{aligned}$$

we have to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}(I-\widetilde{M}_t^\alpha)w_{t-1}-\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-\widetilde{M}_t^\alpha)w_{t-1}]}{\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t)w_{t-1}]/(NT)}\right|=o_p(1),$$

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{2}{NT}\sum_{t=1}^{T-1}\psi_t c_t w'_{t-1}\widetilde{M}_t^\alpha \tilde{v}_{tT}}{\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t)w_{t-1}]/(NT)}\right|=o_p(1),$$

$$\sup_{\mathcal{E}_T}\left|\frac{\frac{1}{NT}\sum_{t=1}^{T-1}c_t^2 \tilde{v}'_{tT}\widetilde{M}_t^\alpha \tilde{v}_{tT}}{\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I-M_t)w_{t-1}]/(NT)}\right|=o_p(1).$$

Since $\widetilde{M}_t^\alpha \leq 2M_t^\alpha \leq 2M_t$, we can apply the same strategy as in the case of spectral cut off regularization scheme provided that $\#\mathcal{E}_T = O(T^2)$ with $\#\mathcal{E}_T$ being the number of elements in the parameter set $\mathcal{E}_T$. Imposing that $\#\mathcal{E}_T = O(T^2)$ is a sufficient condition to have optimality in the Landweber Fridman regularization scheme with no need to impose a condition on the maximum number of iterations.
Summing up, we proved that our procedure of selection of regularization parameter α is optimal under the assumption $\#\mathcal{E}_T = O(T^2)$ for the Landweber Fridman regularization scheme.

The following lemma will be used in the proof of Proposition 4.

Lemma 8. *If assumptions 1', 2' and 3 are satisfied, then*

(i)

$$\begin{aligned}E\left[\left[\sum_{t=1}^{T-1}c_t\tilde{v}'_{tT}M_t^\alpha v_t^*\right]^2\right] &= \frac{\sigma^4}{(1-\delta)^2}E\left[\sum_{t=1}^{T-1}tr[M_t^\alpha]\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)\right]^2+o\left(\left(\sum_{t=1}^{T-1}tr[M_t^\alpha]\right)^2\right)\\ &= O\left(\ln T/(\alpha NT)\right).\end{aligned}$$

(ii)

$$\frac{1}{\sqrt{NT}}E\left[\sum_{t=1}^{T-1} c_t \tilde{v}'_{tT} M_t^{\alpha} v_t^*\right] = \frac{\sigma^2}{(1-\delta)}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} Etr[M_t^{\alpha}]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) = O\left(\frac{1}{\alpha\sqrt{NT}}\right).$$

(iii) Let Δ_α be defined as

$$\Delta_\alpha = \frac{1}{NT}tr\left[\sum_{t=1}^{T-1} E[\tilde{w}'_{t-1}(I-M_t^{\alpha})^2\tilde{w}_{t-1}]\right].$$

Then

$$\Delta_\alpha = \begin{cases} O(\alpha^\beta) & \text{for SC, LF} \\ O(\alpha^{min(\beta,2)}) & \text{for TH} \end{cases}$$

Proof of Lemma 8.

(i)-(ii) These results can be established using proofs similar to those of Lemmas 3(iii) and 4(iii).

(iii)

$$\begin{aligned}
\Delta_\alpha &= \frac{1}{NT}tr\left[\sum_{t=1}^{T-1} E[\tilde{w}'_{t-1}(I-M_t^{\alpha})^2\tilde{w}_{t-1}]\right] \\
&= \frac{1}{NT}tr\left[E[\widetilde{W}'(I-M^{\alpha})^2\widetilde{W}]\right] \\
&= \frac{1}{NT}E\sum_a E[\widetilde{W}'_a(I-M^{\alpha})^2\widetilde{W}_a] \\
&= \frac{1}{NT}E\sum_a\sum_j(1-\hat{q}_j)^2 < \widetilde{W}_a, \hat{\psi}_j >^2 \\
&\leq \frac{1}{NT}E\sup_{\hat{\lambda}_j}\left[\hat{\lambda}_j^{2\beta}(1-\hat{q}_j)^2\right]\sum_a\sum_j\frac{1}{\hat{\lambda}_j^{2\beta}} < \widetilde{W}_a, \hat{\phi}_j >^2 .
\end{aligned}$$

It follows from Carrasco et al. (2007, Proposition 3.11) that the term $sup_{\hat{\lambda}_j}\hat{\lambda}_j^{2\beta}(1-\hat{q}_j)^2 < C\alpha^\beta$ for SC and LF and $C\alpha^{\min(\beta,1)}$ for Tikhonov for some constant $C > 0$. Moreover, the sum

$$\frac{1}{NT}E\sum_a\sum_j\frac{1}{\hat{\lambda}_j^{2\beta}} < \widetilde{W}_a, \hat{\phi}_j >^2$$

is finite by Assumption 3. Hence, the rate of Δ_α follows.

Proof of the proposition 4.

Let $\rho_\alpha = trace(S(\alpha))$. It follows from Lemma 8 that a term is $o_p(\rho_\alpha)$ if it is either $o_p(1/\alpha^2 NT)$ or $o_p(\alpha^\beta)$.

Recall that x_t is a $N \times (L_m + 1)$ matrix with $x_t = (y_{t-1}, m_t) \equiv (u_t, m_t)$. $x_t^* = (u_t^*, m_t^*)$ with $u_t^* = w_{t-1} - c_t \widetilde{v}_{tT}$, $E_Z(u_t^*) = w_{t-1}$, $E_Z(x_t^*) = (w_{t-1}, m_t^*) \equiv \widetilde{w}_{t-1}$, and $x_t^* - E_Z(x_t^*) = (-c_t \widetilde{v}_{tT}, 0)$.

First, we note that

$$\sqrt{NT}(\widehat{\theta}^\alpha - \theta) = \left(\frac{1}{NT} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha x_t^* \right)^{-1} \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha v_t^* \right).$$

Following Okui (2009), we have the following decomposition

$$\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha v_t^* = h + T_1^h + T_2^h$$

where

$$h = \begin{bmatrix} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} w_{t-1}' v_t^* \\ \\ \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} m_t^* v_t^* \end{bmatrix},$$

$$T_1^h = - \begin{bmatrix} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} w_{t-1}' (I - M_t^\alpha) v_t^* \\ \\ \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} m_t^* (I - M_t^\alpha) v_t^* \end{bmatrix},$$

$$T_2^h = - \begin{bmatrix} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} c_t \tilde{v}_{tT}' M_t^\alpha v_t^* \\ \\ 0 \end{bmatrix}.$$

Now consider the denominator.

$$\begin{aligned}
& \frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} x_t^{*} \\
= \quad & \frac{1}{NT}\sum_{t=1}^{T-1} \left(x_t^{*} - E_Z\left(x_t^{*}\right)\right)' M_t^{\alpha} \left(x_t^{*} - E_Z\left(x_t^{*}\right)\right) \text{ (term } Z_4^H\text{)} \\
& + \frac{1}{NT}\sum_{t=1}^{T-1} E_Z\left(x_t^{*}\right)' M_t^{\alpha} E_Z\left(x_t^{*}\right) \\
& + \frac{2}{NT}\sum_{t=1}^{T-1} E_Z\left(x_t^{*}\right)' M_t^{\alpha} \left(x_t^{*} - E_Z\left(x_t^{*}\right)\right) \text{ (term } Z_3^H\text{)}
\end{aligned}$$

$$\begin{aligned}
= \quad & Z_3^H + Z_4^H + H \\
& + \frac{1}{NT}\sum_{t=1}^{T-1} E_Z\left(x_t^{*}\right)' E_Z\left(x_t^{*}\right) - H \text{ (term } Z_1^H\text{)} \\
& - \left[\frac{1}{NT}\sum_{t=1}^{T-1} \left\{E_Z\left(x_t^{*}\right)' \left(I - M_t^{\alpha}\right) E_Z\left(x_t^{*}\right) - E\left[E_Z\left(x_t^{*}\right)' \left(I - M_t^{\alpha}\right) E_Z\left(x_t^{*}\right)\right]\right\}\right] \text{ (term } Z_2^H\text{)} \\
& + \frac{1}{NT}\sum_{t=1}^{T-1} E\left\{E_Z\left(x_t^{*}\right)' \left(I - M_t^{\alpha}\right) E_Z\left(x_t^{*}\right)\right\} \text{ (term } T^H\text{)}
\end{aligned}$$

So that

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} x_t^{*} = H + T^H + \sum_{j=1}^{4} Z_j^H$$

where

$$H = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T-1} E(w_{it}^2) & \frac{1}{T}\sum_{t=1}^{T-1} E(w_{it} m_{it}^{*'}) \\ \frac{1}{T}\sum_{t=1}^{T-1} E(m_{it}^* w_{it}) & \frac{1}{T}\sum_{t=1}^{T-1} E(m_{it}^* m_{it}^{*'}) \end{bmatrix},$$

$$T^H = -\begin{bmatrix} \frac{1}{NT}\sum_{t=1}^{T-1} E[w_{t-1}'(I - M_t^\alpha)w_{t-1}] & \frac{1}{NT}\sum_{t=1}^{T-1} E[w_{t-1}'(I - M_t^\alpha)m_t^*] \\ \frac{1}{NT}\sum_{t=1}^{T-1} E[m_t^{*'}(I - M_t^\alpha)w_{t-1}] & \frac{1}{NT}\sum_{t=1}^{T-1} E[m_t^{*'}(I - M_t^\alpha)m_t^*] \end{bmatrix},$$

$$Z_1^H = \begin{bmatrix} \frac{1}{NT}\sum_{t=1}^{T-1} w_{t-1}' w_{t-1} & \frac{1}{NT}\sum_{t=1}^{T-1} w_{t-1}' m_t^* \\ \frac{1}{NT}\sum_{t=1}^{T-1} m_t^{*'} w_t & \frac{1}{NT}\sum_{t=1}^{T-1} m_t^{*'} m_t^* \end{bmatrix} - H,$$

$$Z_2^H = -\begin{bmatrix} Z_{2,11}^H & Z_{2,12}^H \\ Z_{2,21}^H & Z_{2,22}^H \end{bmatrix},$$

$$Z_{2,11}^H = \frac{1}{NT}\sum_{t=1}^{T-1}[w_{t-1}'(I - M_t^\alpha)w_{t-1} - E\{w_{t-1}'(I - M_t^\alpha)w_{t-1}\}],$$

$$Z_{2,21}^H = \frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*'}(I - M_t^\alpha)w_{t-1} - E\{m_t^{*'}(I - M_t^\alpha)w_{t-1}\}],$$

$$Z_{2,12}^H = \frac{1}{NT}\sum_{t=1}^{T-1}[w_{t-1}'(I - M_t^\alpha)m_t^* - E\{w_{t-1}'(I - M_t^\alpha)m_t^*\}],$$

$$Z_{2,22}^H = \frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*'}(I - M_t^\alpha)m_t^* - E\{m_t^{*'}(I - M_t^\alpha)m_t^*\}],$$

$$Z_3^H = -2\begin{bmatrix} \frac{1}{NT}\sum_{t=1}^{T-1} c_t w_{t-1}' M_t^\alpha \tilde{v}_{tT} & 0 \\ \frac{1}{NT}\sum_{t=1}^{T-1} c_t m_t^{*'} M_t^\alpha \tilde{v}_{tT} & 0 \end{bmatrix},$$

$$Z_4^H = \begin{bmatrix} \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 \tilde{v}_{tT}' M_t^\alpha \tilde{v}_{tT} & 0 \\ 0 & 0 \end{bmatrix}.$$

where

$$\tilde{v}_{tT} = \frac{(\phi_{T-t}v_t + ... + \phi_1 v_{T-1})}{T-t}.$$

The terms h, H and Z_1^H do not depend on the matrix M_t^α so that we can use their order given in Okui (2009). We then have that $H = O(1)$, $h = o_p(1)$ and $Z_1^H = O_p(1/\sqrt{NT}) = o(\rho_\alpha)$ provided that $\alpha^\beta\sqrt{NT} \to \infty$.
Using $\tilde{w}_{t-1} = (w_{t-1}, m_t^*)$, the term T_1^h can be rewritten as

$$T_1^h = \frac{1}{NT}\sum_{t=1}^{T-1}\tilde{w}'_{t-1}(I - M_t^\alpha)v_t^*.$$

We have then $E(T_1^h) = 0$ and $V(T_1^h) = \sigma^2\Delta_\alpha$ so that $T_1^h = O_p(\Delta_\alpha^{1/2})$ by Markov's inequality. Since from Lemma 8, we have $\Delta_\alpha = O_p(\alpha^\beta)$, we can conclude that $T_1^h = o_p(1)$.
Regarding the term T_2^h, we can use the same strategy as in the model without covariates. We have from Lemma 8 that $E\left(T_2^h\right) = O\left(1/\alpha\sqrt{NT}\right)$ and $V\left(T_2^h\right) = O\left(\frac{(\ln T)^2}{N}\right)$ so that $T_2^h = o_p\left(1\right)$ provided that $\alpha\sqrt{NT} \to \infty$.

Next, we consider T^H.

$$T^H = -\frac{1}{NT}\sum_{t=1}^{T-1}E\{w'_{t-1}(I - M_t^\alpha)w_{t-1}\} = O_p(\Delta_\alpha).$$

We now look at the term Z_2^H. In the same way as in the model without covariates, we can prove that diagonal elements of Z_2^H are $O_p(1/\sqrt{NT})$. For the other terms, we have

$$Z_{2,21}^H = \frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*'}(I - M_t^\alpha)\tilde{w}_{t-1} - E\{m_t^{*'}(I - M_t^\alpha)\tilde{w}_{t-1}\}],$$

For a given column k of the exogenous covariates, we can write

$$\begin{aligned} Z_{2,21}^{H,k} &= \frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}(I - M_t^\alpha)\tilde{w}_{t-1} - E\{m_t^{*',k}(I - M_t^\alpha)\tilde{w}_{t-1}\}], \\ &= \frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}\tilde{w}_{t-1} - E\{m_t^{*',k}\tilde{w}_{t-1}\}] \\ &+ \frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}M_t^\alpha\tilde{w}_{t-1} - E\{m_t^{*',k}M_t^\alpha\tilde{w}_{t-1}\}]. \end{aligned}$$

From Okui (2009),

$$\frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}\tilde{w}_{t-1} - E\{m_t^{*',k}\tilde{w}_{t-1}\}] = O_p(1/\sqrt{NT}).$$

Moreover,

$$\begin{aligned}
&\frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}M_t^{\alpha}\tilde{w}_{t-1} - E\{m_t^{*',k}M_t^{\alpha}\tilde{w}_{t-1}\}] \\
&= \frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}M_t^{\alpha}\tilde{w}_{t-1} - \frac{1}{2}E[\tilde{w'}_{t-1}\tilde{w}_{t-1}] - \frac{1}{2}E[m_t^{*',k}m_t^{*,k}]] \\
&- \frac{1}{NT}E\left[\sum_{t=1}^{T-1}[m_t^{*',k}M_t^{\alpha}\tilde{w}_{t-1} - \frac{1}{2}E[\tilde{w'}_{t-1}\tilde{w}_{t-1}] - \frac{1}{2}E[m_t^{*',k}m_t^{*,k}]]\right].
\end{aligned}$$

Now using Cauchy Schwarz inequality,

$$\begin{aligned}
m_t^{*',k}M_t^{\alpha}\tilde{w}_{t-1} &\leq (\tilde{w'}_{t-1}M_t^{\alpha}M_t^{\alpha}\tilde{w}_{t-1})^{1/2}(m_t^{*',k}m_t^{*,k})^{1/2} \\
&\leq (\tilde{w'}_{t-1}\tilde{w}_{t-1})^{1/2}(m_t^{*',k}m_t^{*,k})^{1/2} \\
&\leq \frac{1}{2}(\tilde{w'}_{t-1}\tilde{w}_{t-1}) + \frac{1}{2}(m_t^{*',k}m_t^{*,k})
\end{aligned}$$

Hence,

$$\begin{aligned}
&\frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}M_t^{\alpha}\tilde{w}_{t-1} - \frac{1}{2}E[\tilde{w'}_{t-1}\tilde{w}_{t-1}] - \frac{1}{2}E[m_t^{*',k}m_t^{*,k}]] \\
&\leq \frac{1}{2}\frac{1}{NT}\sum_{t=1}^{T-1}[\tilde{w'}_{t-1}\tilde{w}_{t-1} - E[\tilde{w'}_{t-1}\tilde{w}_{t-1}]] + \frac{1}{2}\frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}m_t^{*,k} - E[m_t^{*',k}m_t^{*,k}]] \\
&= O_p(1/\sqrt{NT}) + O_p(1/\sqrt{NT}) = O_p(1/\sqrt{NT}).
\end{aligned}$$

We have just proved that elements with the same form as $Z_{21}^{H,k}$ are $O_p(1/\sqrt{NT})$. The same strategy can be applied to the non diagonal elements of the l_m dimensional matrix Z_{22}^{H} allowing us to conclude that $Z_2^H = O_p(1/\sqrt{NT})$ so that the $Z_2^H = o_p(\rho_\alpha)$ provided that $\alpha^{\beta}\sqrt{NT} \to \infty$.
For the term Z_3^H, we note that

$$E\left(\frac{1}{NT}\sum_{t=1}^{T-1}c_t\tilde{w}'_{t-1}M_t^{\alpha}\tilde{v}_{tT}\right) = 0$$

and for $\alpha = 0$, Okui (2009) proved that

$$Var\left(\frac{1}{NT}\sum_{t=1}^{T-1} c_t \tilde{w'}_{t-1} M_t^{\alpha} \tilde{v}_{tT}\right) = O\left(\frac{1}{NT}\right)$$

so that

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t \tilde{w'}_{t-1} M_t^{\alpha} \tilde{v}_{tT} = O_p(\frac{1}{\sqrt{NT}}) = o_p(\rho_\alpha).$$

Now looking to the other terms of Z_3^H, they are in form

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t m_t^{l*'} M_t^{\alpha} \tilde{v}_{tT}$$

$l = 1, ..., l_m$. Hence first conditionning on z_t, we can prove in the same way that those terms are $o_p(\rho_\alpha)$.
For Z_4^H, following the same strategy as in the model without covariates, we have

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 \tilde{v}'_{tT} M_t^{\alpha} \tilde{v}_{tT} = O_p(\frac{1}{\alpha NT}).$$

Hence $Z_4^H = o_p(\rho_\alpha)$.
We now apply Lemma 6. Let us define $Z^A = 0$ and

$$\hat{A} = (h + T_1^h + T_2^h)(h + T_1^h + T_2^h)' - hh'H^{-1}T_1^H - T_1^H H^{-1}hh'.$$

Since we want to calculate the expectation of A, we need to calculate the expectation of each term. By the third moment condition and the independence assumption both on the error term v_{it}, we can show that $E(hT_2^{h'}) = E(T_2^h h') = E(T_1^h T_2^{h'}) = E(T_2^h T_1^{h'}) = 0$.
It can easily be proved that $E(hh') = \sigma^2 H$, $E\{hT_1^{h'}\} = E\{T_1^h h'\} =E(hh'H^{-1}T_1^H) = E(T_1^H H^{-1}hh') = \sigma^2 T^H$. Given these equalities,

$$E(\widehat{A}) = \sigma^2 H + E(T_1^h T_1^{h'}) + E(T_2^h T_2^{h'}),$$

$$E(T_1^h T_1^{h'}) = \frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E[\tilde{w}'_{t-1}(I - M_t^{\alpha})^2 \tilde{w}_{t-1}]$$

and

$$E(T_2^h T_2^{h'}) = E(T_2^h)E(T_2^h)' + var(T_2^h)$$

$$= \frac{\sigma^4}{(1-\delta)^2}\begin{bmatrix}\mathcal{A}(\alpha) & 0\\ 0 & 0\end{bmatrix} + o_p(\rho_\alpha)$$

provided $\alpha \ln (T) \sqrt{T} \to 0$. Hence, by Lemma 6, we have

$$E(\widehat{A}) = \sigma^2 H + HS(\alpha)H + o_p(\rho_\alpha)$$

with

$$HS(\alpha)H = \frac{\sigma^4}{(1-\delta)^2} \begin{bmatrix} \mathcal{A}(\alpha) & 0 \\ 0 & 0 \end{bmatrix} + \frac{\sigma^2}{NT} \sum_{t=1}^{T-1} E[\tilde{w}_{t-1}'(I - M_t^\alpha)^2 \tilde{w}_{t-1}].$$

B Chapter 2

B.1 Simulations results.

		GMM	LIML	IVK	TGMM	TLIML	PGMM	PLIML	LGMM	LLIML
$\delta = 0.50$										
T= 10	Med.bias	-0.0622	-0.0254	-0.0377	-0.0524	-0.0231	-0.0442	-0.0211	-0.0503	-0.0228
	Med.abs	0.0683	0.0627	0.0622	0.0658	0.0619	0.0634	0.0604	0.0648	0.0618
	Iqr	0.0992	0.1195	0.1124	0.1068	0.1179	0.1081	0.1187	0.1082	0.1192
	Med.mse	0.0047	0.0039	0.0039	0.0043	0.0038	0.0040	0.0036	0.0042	0.0038
	Cov	0.8590	0.9038	0.9210	0.9482	0.9268	0.8996	0.9132	0.9334	0.9178
T= 25	Med.bias	-0.0396	-0.0257	-0.0151	-0.0271	-0.0144	-0.0219	-0.0220	-0.0260	-0.0146
	Med.abs	0.0400	0.0329	0.0268	0.0323	0.0292	0.0305	0.0315	0.0321	0.0285
	Iqr	0.0456	0.0553	0.0504	0.0495	0.0566	0.0526	0.0557	0.0508	0.0550
	Med.mse	0.0016	0.0011	0.0007	0.0010	0.0009	0.0009	0.0010	0.0010	0.0008
	Cov	0.7572	0.9054	0.9216	0.9664	0.9838	0.9004	0.9186	0.9456	0.9556
$\delta = 0.75$										
T= 10	Med.bias	-0.1350	-0.0595	-0.0960	-0.1052	-0.0578	-0.0900	-0.0523	-0.1059	-0.0545
	Med.abs	0.1352	0.1209	0.1183	0.1133	0.1106	0.1030	0.1059	0.1182	0.1256
	Iqr	0.1334	0.2391	0.1804	0.1507	0.2190	0.1551	0.2153	0.1640	0.2545
	Med.mse	0.0183	0.0146	0.0140	0.0128	0.0122	0.0106	0.0112	0.0140	0.0158
	Cov	0.6730	0.7000	0.8676	0.9170	0.8036	0.8458	0.7710	0.9158	0.8772
T= 25	Med.bias	-0.0625	-0.0440	-0.0264	-0.0352	-0.0189	-0.0272	-0.0360	-0.0334	-0.0168
	Med.abs	0.0625	0.0485	0.0348	0.0382	0.0339	0.0345	0.0437	0.0378	0.0381
	Iqr	0.0455	0.0764	0.0582	0.0557	0.0657	0.0585	0.0740	0.0594	0.0764
	Med.mse	0.0039	0.0024	0.0012	0.0015	0.0011	0.0012	0.0019	0.0014	0.0015
	Cov	0.4892	0.7170	0.8954	0.9512	0.9246	0.8882	0.7588	0.9542	0.9378
$\delta = 0.95$										
T= 10	Med.bias	-0.3336	-1.2295	-0.3194	-0.3080	-0.9338	-0.2991	-0.8060	-0.3219	-0.9243
	Med.abs	0.3336	1.4964	0.3211	0.3100	1.0945	0.3007	0.9693	0.3289	1.2332
	Iqr	0.2034	1.9969	0.3321	0.3111	1.6422	0.3083	1.4702	0.4238	1.7055
	Med.mse	0.1113	2.2394	0.1031	0.0961	1.1980	0.0904	0.9395	0.1082	1.5207
	Cov	0.1996	0.1300	0.7444	0.8330	0.3626	0.7166	0.3356	0.9102	0.7904
T= 25	Med.bias	-0.1362	-1.3211	-0.1069	-0.0986	-0.2386	-0.0946	-0.5943	-0.1216	-0.8887
	Med.abs	0.1362	1.4588	0.1072	0.0988	0.3818	0.0947	0.7421	0.1228	1.1555
	Iqr	0.0568	1.8260	0.1197	0.1061	1.3659	0.1050	1.3985	0.1738	1.7890
	Med.mse	0.0186	2.1281	0.0115	0.0098	0.1458	0.0090	0.5508	0.0151	1.3351
	Cov	0.0156	0.0618	0.7380	0.8710	0.3268	0.6974	0.2132	0.9378	0.6266

Table B.1 – Simulations results with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 5000 replications.

		Mean	Std	Mode	q1	Median	q3
$\delta = 0.50$							
T= 10	TGMM	0.0005	0.0002	0.0005	0.0004	0.0005	0.0007
	PGMM	29.3	4.1	29.0	27.0	29.0	32.0
	LGMM	2362.0	1285.1	2207.0	1500.0	2056.0	2901.5
	TLIML	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
	PLIML	39.0	1.5	40.0	38.0	40.0	40.0
	LLIML	12611.5	3572.9	15127.0	9349.0	15127.0	15127.0
T= 25	TGMM	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
	PGMM	138.5	17.1	140.0	130.0	140.0	150.0
	LGMM	6064.1	2804.5	5778.0	4054.0	5778.0	9349.0
	TLIML	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
	PLIML	285.8	7.63	290.0	280.0	290.0	290.0
	LLIML	15127.0		15127.0	15127.0	15127.0	15127.0
$\delta = 0.75$							
T= 10	TGMM	0.0055	0.0549	0.0007	0.0010	0.0016	0.0026
	PGMM	20.6	4.6	21.0	17.0	20.0	23.0
	LGMM	9493.9	5384.3	15127.0	4845.0	9349.0	15127.0
	TLIML	0.0497	0.2161	0.0001	0.0001	0.0001	0.0001
	PLIML	36.0	5.6	40.0	34.0	38.0	40.0
	LLIML	12698.0	5785.1	15127.0	15127.0	15127.0	15127.0
T= 25	TGMM	0.0004	0.0001	0.0003	0.0003	0.0004	0.0004
	PGMM	87.4	13.7	90.0	80.0	90.0	100.0
	LGMM	14312.2	2220.2	15127.0	15127.0	15127.0	15127.0
	TLIML	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
	PLIML	278.3	18.8	290.0	270.0	290.0	290.0
	LLIML	14830.8	2304.6	15127.0	15127.0	15127.0	15127.0
$\delta = 0.95$							
T= 10	TGMM	0.3927	0.4660	0.9999	0.0063	0.0456	0.9999
	PGMM	13.5	4.5	10.0	10.0	12.0	15.0
	LGMM	3219.4	6903.3	1.0	13.0	205.0	344.0
	TLIML	0.3041	0.4554	0.0001	0.0001	0.0002	0.9999
	PLIML	25.6	13.2	10.0	10.0	31.0	39.0
	LLIML	1358.0	4684.8	1.0	1.0	2.0	184.0
T= 25	TGMM	0.1555	0.3330	0.9999	0.0037	0.0088	0.0365
	PGMM	53.3	7.0	50.0	50.0	50.0	50.0
	LGMM	10110.4	8012.2	15127.0	121.0	15127.0	15127.0
	TLIML	0.3330	0.4711	0.0001	0.0001	0.0001	0.9999
	PLIML	197.9	111.8	290.0	50.0	280.0	290.0
	LLIML	8103.9	8177.8	15127.0	1.0	15127.0	15127.0

TABLE B.2 – Properties of the distribution of the regularization parameters with N=50, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 5000 replications.

		GMM	LIML	IVK	TGMM	TLIML	PGMM	PLIML	LGMM	LLIML
$\delta = 0.50$										
	Med.bias	-0.0337	-0.0121	-0.0223	-0.0291	-0.0111	-0.0251	-0.0096	-0.0278	-0.0109
T= 10	Med.abs	0.0439	0.0416	0.0419	0.0429	0.0415	0.0417	0.0413	0.0428	0.0417
	Iqr	0.0747	0.0818	0.0806	0.0776	0.0818	0.0781	0.0823	0.0776	0.0820
	Med.mse	0.0019	0.0017	0.0018	0.0018	0.0017	0.0017	0.0017	0.0018	0.0017
	Cov	0.8978	0.9248	0.9232	0.9512	0.9410	0.9204	0.9324	0.9384	0.9318
	Med.bias	-0.0218	-0.0112	-0.0090	-0.0158	-0.0077	-0.0129	-0.0099	-0.0154	-0.0073
T= 25	Med.abs	0.0234	0.0190	0.0183	0.0206	0.0193	0.0198	0.0188	0.0206	0.0187
	Iqr	0.0324	0.0354	0.0351	0.0347	0.0374	0.0355	0.0355	0.0347	0.0368
	Med.mse	0.0005	0.0004	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
	Cov	0.8390	0.9266	0.9312	0.9780	0.9882	0.9106	0.9322	0.9416	0.9624
$\delta = 0.75$										
	Med.bias	-0.0816	-0.0270	-0.0546	-0.0622	-0.0228	-0.0525	-0.0237	-0.0588	-0.0192
T= 10	Med.abs	0.0838	0.0672	0.0777	0.0715	0.0646	0.0681	0.0640	0.0709	0.0689
	Iqr	0.0992	0.1291	0.1268	0.1090	0.1267	0.1110	0.1248	0.1121	0.1359
	Med.mse	0.0070	0.0045	0.0060	0.0051	0.0042	0.0046	0.0041	0.0050	0.0047
	Cov	0.7856	0.8464	0.8922	0.9348	0.8908	0.8872	0.8720	0.9396	0.9396
	Med.bias	-0.0370	-0.0178	-0.0160	-0.0211	-0.0094	-0.0160	-0.0149	-0.0189	-0.0076
T= 25	Med.abs	0.0371	0.0243	0.0229	0.0248	0.0219	0.0230	0.0234	0.0249	0.0240
	Iqr	0.0347	0.0437	0.0410	0.0394	0.0422	0.0411	0.0431	0.0431	0.0482
	Med.mse	0.0014	0.0006	0.0005	0.0006	0.0005	0.0005	0.0005	0.0006	0.0006
	Cov	0.6616	0.8662	0.9174	0.9620	0.9610	0.9156	0.8806	0.9742	0.9752
$\delta = 0.95$										
	Med.bias	-0.3105	-0.8503	-0.2941	-0.2766	-0.6223	-0.2629	-0.5158	-0.2981	-0.6905
T= 10	Med.abs	0.3105	1.2083	0.2974	0.2781	0.8514	0.2651	0.7366	0.3054	1.0954
	Iqr	0.1934	1.9015	0.3196	0.2863	1.5497	0.2787	1.3636	0.4076	1.7095
	Med.mse	0.0964	1.4599	0.0884	0.0773	0.7249	0.0703	0.5426	0.0933	1.1998
	Cov	0.2260	0.1620	0.7550	0.8348	0.3902	0.7284	0.3706	0.9092	0.7978
	Med.bias	-0.1179	-0.3631	-0.0886	-0.0724	-0.0671	-0.0661	-0.1447	-0.1036	-0.2292
T= 25	Med.abs	0.1179	0.5368	0.0898	0.0733	0.1102	0.0670	0.1939	0.1054	0.5534
	Iqr	0.0547	1.3733	0.1103	0.0815	0.3715	0.0840	0.8364	0.1630	1.6616
	Med.mse	0.0139	0.2882	0.0081	0.0054	0.0121	0.0045	0.0376	0.0111	0.3063
	Cov	0.0330	0.1558	0.7582	0.8708	0.4744	0.7524	0.3284	0.9394	0.6366

Table B.3 – Simulations results with $N = 100$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 5000 replications.

		Mean	Std	Mode	q1	Median	q3
$\delta = 0.50$							
T= 10	TGMM	0.0003	0.0001	0.0003	0.0002	0.0003	0.0004
	PGMM	32.6	3.1	32.0	31.0	33.0	35.0
	LGMM	3459.1	1273.3	3571.0	2588.0	3191.0	4054.0
	TLIML	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
	PLIML	39.4	1.0	40.0	39.0	40.0	40.0
	LLIML	13406.4	2737.9	15127.0	9349.0	15127.0	15127.0
T= 25	TGMM	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
	PGMM	167.8	16.5	170.0	160.0	170.0	180.0
	LGMM	10182.1	3253.5	9349.0	9349.0	9349.0	15127.0
	TLIML	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
	PLIML	288.9	3.5	290.0	290.0	290.0	290.0
	LLIML	15127.0		15127.0	15127.0	15127.0	15127.0
$\delta = 0.75$							
T= 10	TGMM	0.0012	0.0008	0.0007	0.0007	0.0010	0.0014
	PGMM	23.1	4.1	21.0	21.0	22.0	26.0
	LGMM	13157.8	3624.6	15127.0	9349.0	15127.0	15127.0
	TLIML	0.0013	0.0346	0.0001	0.0001	0.0001	0.0001
	PLIML	37.5	3.1	40.0	36.0	39.0	40.0
	LLIML	14634.5	2837.0	15127.0	15127.0	15127.0	15127.0
T= 25	TGMM	0.0002	0.0000	0.0002	0.0002	0.0002	0.0002
	PGMM	104.1	12.9	100.0	100.0	100.0	110.0
	LGMM	15120.1	200.1	15127.0	15127.0	15127.0	15127.0
	TLIML	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
	PLIML	280.4	14.6	290.0	280.0	290.0	290.0
	LLIML	15099.5	792.7	15127.0	15127.0	15127.0	15127.0
$\delta = 0.95$							
T= 10	TGMM	0.3168	0.4435	0.9999	0.0045	0.0204	0.9999
	PGMM	14.0	4.5	10.0	10.0	13.0	17.0
	LGMM	3549.5	7144.7	1.0	11.0	211.0	356.0
	TLIML	0.2784	0.4455	0.0001	0.0001	0.0001	0.9999
	PLIML	26.3	13.0	10.0	10.0	32.0	39.0
	LLIML	1484.2	4877.8	1.0	1.0	3.0	194.0
T= 25	TGMM	0.0463	0.1843	0.0019	0.0024	0.0042	0.0085
	PGMM	54.7	7.9	50.0	50.0	50.0	60.0
	LGMM	11316.7	7574.7	15127.0	113.0	15127.0	15127.0
	TLIML	0.2057	0.4041	0.0001	0.0001	0.0001	0.0001
	PLIML	217.0	102.2	290.0	80.0	280.0	290.0
	LLIML	8649.8	8008.6	15127.0	1.0	15127.0	15127.0

TABLE B.4 – Properties of the distribution of the regularization parameters with N=100, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$ for 5000 replications.

B.2 Proofs.

Lemma 1 : *If assumptions 1-3 are satisfied, provided that* $\alpha\sqrt{NT} \to \infty$ *we have*

(*i*)

$$\frac{v^{*\prime} M^{\alpha} v^{*}}{\sqrt{NT}} = O_p\left(\frac{1}{\alpha\sqrt{NT}}\right).$$

(*ii*)

$$\frac{v^{*\prime} v^{*}}{NT} - \sigma^2 = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

(*iii*) Let $\wedge_0 = v^{*\prime} M^{\alpha} v^{*}/v^{*\prime} v^{*}$, $\tilde{\wedge} = v^{*\prime} M^{\alpha} v^{*}/\sigma^2 NT$. Then

$$\wedge - \tilde{\wedge} = O_p\left(\frac{1}{\alpha NT}\right) O_p\left(\frac{1}{\sqrt{NT}}\right).$$

(*iv*) Let $\tilde{v}_{itT} = (\phi_{T-t} v_{it} +, ..., +\phi_1 v_{i,T-1})/(T-t)$, $u_{it} = -c_t \tilde{v}_{itT}$, $\rho_t = E(u_{it} v_{it}^{*})/\sigma^2$. Then

$$\sum_{t=1}^{T-1} \left(\frac{E[x^{*\prime} v^{*}]}{\sigma^2 NT} - \rho_t\right) v_t^{*\prime} M_t^{\alpha} v_t^{*} = O_p\left(\frac{1}{T\alpha}\right).$$

Proof of Lemma 1.

(*i*) Let us define $\bar{v}_{tT} = (v_t + ... + v_T)/(T - t + 1)$ so that $v_t^{*} = (v_t - \bar{v}_{tT})/c_t$.

$$
\begin{aligned}
E[v^{*\prime}M^{\alpha}v^{*}] &= \sum_{t=1}^{T} E[v_t^{*\prime}M_t^{\alpha}v_t^{*}] \\
&= \sum_{t=1}^{T} trE[v_t^{*\prime}M_t^{\alpha}v_t^{*}] \\
&= \sum_{t=1}^{T} Etr[v_t^{*\prime}M_t^{\alpha}v_t^{*}] \\
&= \sum_{t=1}^{T} Etr[M_t^{\alpha}v_t^{*}v_t^{*\prime}] \\
&= \sum_{t=1}^{T} Etr[M_t^{\alpha}E_t[v_t^{*}v_t^{*\prime}]] \\
&= \sum_{t=1}^{T} Etr[M_t^{\alpha}\sigma^2 I_N]] \\
&= \sigma^2 \sum_{t=1}^{T} Etr[M_t^{\alpha}] \\
&= \sigma^2 Etr[M^{\alpha}] \\
&= O\left(\frac{1}{\alpha}\right)
\end{aligned}
$$

where the last result comes from Lemma 4 (ii) of CN (2019). We have just proved that

$$
E\left[\frac{v^{*\prime}M^{\alpha}v^{*}}{\sqrt{NT}}\right] = O\left(\frac{1}{\alpha\sqrt{NT}}\right).
$$

To derive the order of magnitude of the variance we use the following decomposition

$$
\frac{v^{*\prime}M^{\alpha}v^{*}}{\sqrt{NT}} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T} c_t^{-2}v_t'M_t^{\alpha}v_t - \frac{2}{\sqrt{NT}}\sum_{t=1}^{T} c_t^{-2}v_t'M_t^{\alpha}\bar{v}_{tT} + \frac{1}{\sqrt{NT}}\sum_{t=1}^{T} c_t^{-2}\bar{v}_{tT}'M_t^{\alpha}\bar{v}_{tT} \tag{B.1}
$$

We have

$$
Var\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T} c_t^{-2}v_t'M_t^{\alpha}v_t\right) = \frac{1}{NT}\sum_{t=1}^{T} c_t^{-4}Var(v_t'M_t^{\alpha}v_t)
$$

From Lemma 2 (ii) of CN (2019)

$$Var(v_t' M_t^\alpha v_t) \leq (2\sigma^4 + \kappa_4) tr([M_t^\alpha]^2)$$

and hence

$$\begin{aligned} Var\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} c_t^{-2} v_t' M_t^\alpha v_t\right) &\leq \frac{(2\sigma^4 + \kappa_4)}{NT} \sum_{t=1}^{T} \left(1 + \frac{1}{T-t}\right)^2 tr([M_t^\alpha]^2) \\ &\leq \frac{2(2\sigma^4 + \kappa_4)}{NT} \sum_{t=1}^{T} tr[M_t^\alpha] \\ &= O\left(\frac{1}{\alpha NT}\right). \end{aligned}$$

The second and third terms of (B.1) are similar to Υ_{21NT}^α and Υ_{22NT}^α defined in CN (2019). So

$$Var\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} v_t' M_t^\alpha \bar{v}_{tT}\right) = O\left(\frac{Ln(T)}{\alpha NT}\right)$$

$$Var\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \bar{v}_{tT}' M_t^\alpha \bar{v}_{tT}\right) = O\left(\frac{Ln(T)}{\alpha NT}\right).$$

Hence

$$\begin{aligned} Var\left(\frac{v^{*\prime} M^\alpha v^*}{\sqrt{NT}}\right) &\leq O\left(\frac{Ln(T)}{\alpha NT}\right) + O\left(\frac{Ln(T)}{\alpha NT}\right) + O\left(\frac{1}{\alpha NT}\right) \\ &\quad 2O\left(\frac{Ln(T)}{\alpha NT}\right) + 2O\left(\frac{\sqrt{Ln(T)}}{\alpha NT}\right) + 2O\left(\frac{\sqrt{Ln(T)}}{\alpha NT}\right) \\ &= O\left(\frac{Ln(T)}{\alpha NT}\right). \end{aligned}$$

Because

$$E\left[\frac{v^{*\prime} M^\alpha v^*}{\sqrt{NT}}\right] = O\left(\frac{1}{\alpha\sqrt{NT}}\right)$$

$$Var\left(\frac{v^{*\prime} M^\alpha v^*}{\sqrt{NT}}\right) = O\left(\frac{Ln(T)}{\alpha NT}\right)$$

we have that if $Ln(T)/\alpha NT \to 0$ then

$$\frac{v^{*\prime}M^{\alpha}v^{*}}{\sqrt{NT}} = O_p\left(\frac{1}{\alpha\sqrt{NT}}\right).$$

But $Ln(T)/\alpha NT \to 0$ because $Ln(T)/\sqrt{NT} \to 0$ and $\alpha\sqrt{NT} \to \infty$.

(ii) We will prove that

$$E\left(\frac{v^{*\prime}v^{*}}{NT} - \sigma^2\right) = 0,$$
$$Var\left(\frac{v^{*\prime}v^{*}}{NT} - \sigma^2\right) = O\left(\frac{1}{NT}\right).$$

By the assumptions on the error term, we have

$$\begin{aligned}
E\left[\frac{v^{*\prime}v^{*}}{NT} - \sigma^2\right] &= E\left[\frac{v^{*\prime}v^{*}}{NT}\right] - \sigma^2 \\
&= \frac{1}{NT}\sum_{t=1}^{T} E[v_t^{*\prime}v_t^{*}] - \sigma^2 \\
&= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i} E[v_{it}^{*}v_{it}^{*}] - \sigma^2 \\
&= \frac{1}{NT}NT\sigma^2 - \sigma^2 \\
&= 0.
\end{aligned}$$

Moreover, because $v_t^{*} = c_t(v_t - \bar{\bar{v}}_t/(T-t))$ with $\bar{\bar{v}}_t = (v_{t+1} + ... + v_T)$, we have

$$\frac{1}{NT}v^{*\prime}v^{*} = \frac{1}{NT}\sum_{t=1}^{T} c_t^2 v_t^{\prime}v_t - \frac{2}{NT}\sum_{t=1}^{T}\frac{c_t^2}{T-t}v_t^{\prime}\bar{\bar{v}}_t + \frac{1}{NT}\sum_{t=1}^{T}\frac{c_t^2}{(T-t)^2}\bar{\bar{v}}_t^{\prime}\bar{\bar{v}}_t$$

We now derive the order of magnitude of the variance of these three terms. From assumption 1, the $\{v_{it}\}$ $(t = 1,...,T;\ i = 1,...,N)$ are

$i.i.d.$ across time and individuals and $E(v_{it}^4) < \infty$. Hence, $Var(v_t'v_t) \leq NE[v_{it}^4]$ and $Cov(v_t'v_t, v_s'v_s) = 0$ for $t \neq s$. Hence,

$$\begin{aligned} Var\left(\frac{1}{NT}\sum_{t=1}^{T} c_t^2 v_t' v_t\right) &= \frac{1}{N^2T^2}\sum_{t=1}^{T} c_t^4 Var(v_t'v_t) \\ &\leq \frac{NE[v_{it}^4]}{N^2T^2}\sum_{t=1}^{T} c_t^4 \\ &\leq \frac{N}{N^2T^2}\sum_{t=1}^{T}\left(\frac{T-t}{T-t+1}\right)^2 = O\left(\frac{1}{NT}\right). \end{aligned}$$

Moreover, for $s = t + a$ with $a > 0$,

$$\begin{aligned} Cov(v_{it}\bar{\bar{v}}_{it}, v_{is}\bar{\bar{v}}_{is}) &= Cov(v_{it}\bar{\bar{v}}_{it}, v_{i,t+a}\bar{\bar{v}}_{i,t+a}) \\ &= E(v_{it}\bar{\bar{v}}_{it}v_{i,t+a}\bar{\bar{v}}_{i,t+a}) \\ &= \sum_{j>0,k>0} E(v_{it}v_{i,t+j}v_{i,t+a}v_{i,t+a+k}) \\ &= 0. \end{aligned}$$

using the independence of the v_{it}. Hence,

$$\begin{aligned} Var\left(\frac{1}{NT}\sum_{t=1}^{T}\frac{c_t^2}{T-t}v_t'\bar{\bar{v}}_t\right) &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\frac{c_t^4}{(T-t)^2}Var(v_t'\bar{\bar{v}}_t) \\ &\leq \frac{N\sigma^4}{N^2T^2}\sum_{t=1}^{T}\frac{c_t^4(T-t)}{(T-t)^2} \\ &\leq \frac{N\sigma^4}{N^2T^2}\sum_{t=1}^{T}\left(\frac{T-t}{T-t+1}\right)^2\frac{1}{T-t} \\ &\leq \frac{N\sigma^4}{N^2T^2}\sum_{t=1}^{T}\frac{T-t}{(T-t+1)^2} \\ &= O\left(\frac{1}{NT}\right). \end{aligned}$$

For the variance of the third term, we have

$$
\begin{aligned}
&Var\left(\frac{1}{NT}\sum_{t=1}^{T}\frac{c_t^2}{(T-t)^2}\bar{\bar{v}}_t'\bar{\bar{v}}_t\right)\\
&=\frac{1}{N^2T^2}\sum_{t=1}^{T}\frac{c_t^4}{(T-t)^4}Var(\bar{\bar{v}}_t'\bar{\bar{v}}_t)+\frac{2}{N^2T^2}\sum_{s}\sum_{t>s}\frac{c_t^2}{(T-t)^2}\frac{c_s^2}{(T-s)^2}Cov(\bar{\bar{v}}_t'\bar{\bar{v}}_t,\bar{\bar{v}}_s'\bar{\bar{v}}_s)\\
&=\frac{1}{N^2T^2}\sum_{t=1}^{T}\frac{c_t^4}{(T-t)^4}Var(\bar{\bar{v}}_t'\bar{\bar{v}}_t)+\frac{2}{N^2T^2}\sum_{s}\sum_{a=1}^{T-s}\frac{c_t^2}{(T-t)^2}\frac{c_s^2}{(T-s)^2}Cov(\bar{\bar{v}}_{s+a}'\bar{\bar{v}}_{s+a},\bar{\bar{v}}_s'\bar{\bar{v}}_s)\\
&=\frac{N}{N^2T^2}\sum_{t=1}^{T}\frac{c_t^4}{(T-t)^4}Var(\bar{\bar{v}}_{it}^2)+\frac{2N}{N^2T^2}\sum_{s}\sum_{a=1}^{T-s}\frac{c_t^2}{(T-t)^2}\frac{c_s^2}{(T-s)^2}Cov(\bar{\bar{v}}_{i,s+a}^2,\bar{\bar{v}}_{is}^2).
\end{aligned}
$$

But for a given t, we have

$$
\begin{aligned}
Var(\bar{\bar{v}}_{it}^2) &\le E(\bar{\bar{v}}_{it}^4)\\
&=\sum_{j,k,l,r}E(v_{i,t+j}v_{i,t+k}v_{i,t+l}v_{i,t+r})\\
&=\sum_{j}E(v_{i,t+j}^4)+3\sum_{j\neq k}E(v_{i,t+j}^2v_{i,t+k}^2)\\
&=(T-t)E(v_{i,t+j}^4)+3(T-t)^2\sigma^4\\
&=O((T-t)^2)
\end{aligned}
$$

so that

$$
\frac{N}{N^2T^2}\sum_{t=1}^{T}\frac{c_t^4}{(T-t)^4}Var(\bar{\bar{v}}_{it}^2)=O\left(\frac{1}{NT}\right).
$$

Moreover, for $a>0$, we can write $\bar{\bar{v}}_{is}=v_{is+1}+...+v_{is+a}+\bar{\bar{v}}_{i,s+a}$ and $Cov(\bar{\bar{v}}_{i,s+a}^2,\bar{\bar{v}}_{is}^2)=O((T-s-a)^2)$. Hence,

$$
\begin{aligned}
\frac{2N}{N^2T^2}\sum_{s}\sum_{a=1}^{T-s}Cov(\bar{\bar{v}}_{i,s+a}^2,\bar{\bar{v}}_{is}^2)&=\frac{2N}{N^2T^2}\sum_{s}\frac{c_s^2}{(T-s)^2}\sum_{a=1}^{T-s}\frac{c_{s+a}^2}{(T-s-a)^2}(T-s-a)^2\\
&=O\left(\frac{LnT}{NT^2}\right)=O\left(\frac{1}{NT}\right).
\end{aligned}
$$

We have just proved that the three variances are $O(1/NT)$. By the

Cauchy-Schwarz inequality we obtain

$$Var\left(\frac{1}{NT}v^{*\prime}v^*\right) = O\left(\frac{1}{NT}\right)$$

so that the result (ii) holds since we proved that $E(v^{*\prime}v^*/NT) = \sigma^2$.

(iii)

$$\wedge_0 - \tilde{\wedge} = \frac{v^{*\prime}M^\alpha v^*}{NT}\left(\frac{1}{v^{*\prime}v^*/NT} - \frac{1}{\sigma^2}\right).$$

From (i), we have

$$\frac{v^{*\prime}M^\alpha v^*}{NT} = \frac{1}{\sqrt{NT}}\frac{v^{*\prime}M^\alpha v^*}{\sqrt{NT}} = O_p\left(\frac{1}{\alpha NT}\right).$$

Moreover, from (ii) we have

$$\frac{v^{*\prime}v^*}{NT} - \sigma^2 = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

By the delta method

$$\frac{1}{v^{*\prime}v^*/NT} - \frac{1}{\sigma^2} = O_p\left(\frac{1}{\sqrt{NT}}\right)$$

and finally

$$\wedge_0 - \tilde{\wedge} = \frac{v^{*\prime}M^\alpha v^*}{NT}\left(\frac{1}{v^{*\prime}v^*/NT} - \frac{1}{\sigma^2}\right) = O_p\left(\frac{1}{\alpha NT}\right)O_p\left(\frac{1}{\sqrt{NT}}\right).$$

(*iv*) We first calculate the expression of $\rho_t = E(u_{it}v_{it}^*)/\sigma^2$.

$$E[u_{it}v_{it}^*] = E[-c_t\tilde{v}_{itT}v_{it}^*] = \frac{-\sigma^2 c_t^2}{T-t}\left[\phi_{T-t} - \frac{\phi_1 + ... + \phi_{T-t-1}}{T-t}\right]$$

From Alvarez and Arellano (2003, page 1144) $\phi_1 + ... + \phi_{j-1} = (j - \phi_j)/(1-\delta)$. So

$$\phi_1 + ... + \phi_{T-t-1} = \frac{T - t - \phi_{T-t}}{1-\delta}$$

and

$$\begin{aligned}
\phi_{T-t} - \frac{\phi_1 + ... + \phi_{T-t-1}}{T-t} &= \phi_{T-t} - \frac{T-t-\phi_{T-t}}{(1-\delta)(T-t)} \\
&= \frac{1-\delta^{T-t}}{1-\delta} - \frac{T-t-\phi_{T-t}}{(1-\delta)(T-t)} \\
&= \frac{1}{1-\delta}\left(1 - \delta^{T-t} - 1 + \frac{\phi_{T-t}}{T-t}\right) \\
&= \frac{1}{1-\delta}\left(\frac{\phi_{T-t}}{T-t} - \delta^{T-t}\right).
\end{aligned}$$

From Alvarez and Arellano (2003, page 1144) $\phi_j = \phi_{j-1} + \delta^{j-1}$ so that

$$\delta^{T-t} = \phi_{T-t+1} - \phi_{T-t}$$

and

$$\begin{aligned}
\frac{\phi_{T-t}}{T-t} - \delta^{T-t} &= \frac{\phi_{T-t}}{T-t} - \phi_{T-t+1} + \phi_{T-t} \\
&= \frac{T-t+1}{T-t}\phi_{T-t} - \phi_{T-t+1}.
\end{aligned}$$

Finally with $c_t^2 = (T-t)/(T-t+1)$

$$\begin{aligned}
E[-c_t\tilde{v}_{it}v_{it}^*] &= \frac{-\sigma^2}{(T-t)}\frac{(T-t)}{(T-t+1)}\frac{1}{(1-\delta)}\left(\frac{T-t+1}{T-t}\phi_{T-t} - \phi_{T-t+1}\right) \\
&= \frac{-\sigma^2}{1-\delta}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)
\end{aligned}$$

so that

$$\rho_t = E(u_{it}v_{it}^*)/\sigma^2 = E(-c_t\tilde{v}_{it}v_{it}^*)/\sigma^2 = \frac{-1}{1-\delta}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right).$$

We now prove (iv).
Because $x_t^* = \psi_t w_{t-1} - c_t\tilde{v}_{tT}$ we have $E(x_t^{*\prime}v_t^*) = E(-c_t\tilde{v}_{tT}'v_t^*)$,

$$\begin{aligned}
\frac{E[x^{*\prime}v^*]}{NT\sigma^2} - \rho_t &= \frac{\sum_t E[-c_t\tilde{v}_{tT}'v_t^*]}{NT\sigma^2} - \rho_t \\
&= \frac{\sum_t\sum_i E[-c_t\tilde{v}_{itT}'v_{it}^*]}{NT\sigma^2} - \rho_t \\
&= \frac{N\sum_t E[-c_t\tilde{v}_{itT}'v_{it}^*]}{NT\sigma^2} - \rho_t \\
&= \frac{N}{NT\sigma^2}\frac{-\sigma^2}{1-\delta}\sum_t\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) + \frac{1}{1-\delta}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \\
&= \frac{-N}{NT(1-\delta)}\sum_t\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) + \frac{1}{1-\delta}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \\
&= \frac{-N}{NT(1-\delta)}\left(\phi_1 - \frac{\phi_T}{T}\right) + \frac{1}{1-\delta}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \\
&= O\left(\frac{1}{T}\right).
\end{aligned}$$

Moreover, from (i)

$$\sum_{t=1}^{T-1} v_t^{*\prime}M_t^\alpha v_t^* = \sqrt{NT}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} v_t^{*\prime}M_t^\alpha v_t^* = \sqrt{NT}O_p\left(\frac{1}{\alpha\sqrt{NT}}\right) = O_p\left(\frac{1}{\alpha}\right).$$

Hence,

$$\sum_{t=1}^{T-1}\left(\frac{E[x^{*\prime}v^*]}{\sigma^2 NT} - \rho_t\right)v_t^{*\prime}M_t^\alpha v_t^* = \left(\frac{1}{T}\right)O_p\left(\frac{1}{\alpha}\right) \qquad = O_p\left(\frac{1}{T\alpha}\right).$$

Lemma 2. Let us define $u_{it} = -c_t\tilde{v}_{itT}$, $\rho_t = E[u_{it}v_{it}^*]/\sigma^2$, $\epsilon_{it} = u_{it} - \rho_t v_{it}^*$. If the assumptions of Proposition 2 are satisfied, then

$$\sum_{t=1}^{T-1} \epsilon_t' M_t^\alpha v_t^* = O_p\left(\frac{1}{\sqrt{\alpha}}\right).$$

Proof of Lemma 2.

By the spectral decomposition of the matrix M_t^α we have $\epsilon_t' M_t^\alpha v_t^* = \sum_i q_{it}(\epsilon_t'\Psi_{it})(v_t^{*\prime}\Psi_{it})$ where Ψ_{it} are the orthonormal eigenvectors of M_t^α and $q_{it} = q(\alpha, \widehat{\lambda_{it}}^2)$.

$$\begin{aligned}
\left(\sum_{t=1}^{T-1} \epsilon_t' M_t^\alpha v_t^*\right)^2 &= \sum_{t,s} \epsilon_t' M_t^\alpha v_t^* \epsilon_s' M_s^\alpha v_s^* \\
&= \sum_{t,s}\sum_{i,j} q_{it}q_{js}(\epsilon_t'\Psi_{it})(v_t^{*\prime}\Psi_{it})(\epsilon_s'\Psi_{js})(v_s^{*\prime}\Psi_{js}). \\
&= \sum_{t,s}\sum_{i,j} q_{it}q_{js}\left(\sum_a \epsilon_{at}\Psi_{ait}\right)\left(\sum_b v_{bt}^*\Psi_{bit}\right)\left(\sum_c \epsilon_{cs}\Psi_{cjs}\right)\left(\sum_d v_{ds}^*\Psi_{djs}\right) \\
&= \sum_{t,s}\sum_{i,j}\sum_{a,b,c,d} q_{it}q_{js}\Psi_{ait}\Psi_{bit}\Psi_{cjs}\Psi_{djs}\epsilon_{at}v_{bt}^*\epsilon_{cs}v_{ds}^*. \\
E\left(\sum_{t=1}^{T-1} \epsilon_t' M_t^\alpha v_t^*\right)^2 &= E\left[\sum_{t,s}\sum_{i,j}\sum_{a,b,c,d} q_{it}q_{js}\Psi_{ait}\Psi_{bit}\Psi_{cjs}\Psi_{djs}E[\epsilon_{at}v_{bt}^*\epsilon_{cs}v_{ds}^*]\right] \\
&= E\left[\sum_{t,s}\sum_{i,j}\sum_{a} q_{it}q_{js}\Psi_{ait}^2\Psi_{ajs}^2 E[\epsilon_{at}v_{at}^*\epsilon_{as}v_{as}^*]\right] \\
&+ E\left[\sum_{t,s}\sum_{i,j}\sum_{a\neq b} q_{it}q_{js}\Psi_{ait}^2\Psi_{bjs}^2 E[\epsilon_{at}v_{at}^*]E[\epsilon_{bs}v_{bs}^*]\right] \\
&+ E\left[\sum_{t,s}\sum_{i,j}\sum_{a\neq b} q_{it}q_{js}\Psi_{ait}\Psi_{ajs}\Psi_{bit}\Psi_{bjs}E[\epsilon_{at}\epsilon_{as}]E[v_{bt}^*v_{bs}^*]\right] \\
&+ E\left[\sum_{t,s}\sum_{i,j}\sum_{a\neq b} q_{it}q_{js}\Psi_{ait}\Psi_{ajs}\Psi_{bit}\Psi_{bjs}E[\epsilon_{at}v_{as}^*]E[\epsilon_{bs}v_{bt}^*]\right].
\end{aligned}$$

We have $E[\epsilon_{at} v^*_{at}] = 0$ for every t, $E[v^*_{bt} v^*_{bs}] = 0$ for $t \neq s$. So that

$$
\begin{aligned}
E\bigg(\sum_{t=1}^{T-1} \epsilon_t' M_t^\alpha v_t^*\bigg)^2 &= E\bigg[\sum_{t,s}\sum_{i,j}\sum_{a} q_{it}q_{js}\Psi_{ait}\Psi_{ait}\Psi_{ajs}\Psi_{ajs}E[\epsilon_{at}v^*_{at}\epsilon_{as}v^*_{as}]\bigg] \\
&+ E\bigg[\sum_{t}\sum_{i,j}\sum_{a\neq b} q_{it}q_{jt}\Psi_{ait}\Psi_{ajt}\Psi_{bit}\Psi_{bjt}E[\epsilon_{at}\epsilon_{at}]E[v^*_{bt}v^*_{bt}]\bigg] \\
&+ E\bigg[\sum_{t,s}\sum_{i,j}\sum_{a\neq b} q_{it}q_{js}\Psi_{ait}\Psi_{ajs}\Psi_{bit}\Psi_{bjs}E[\epsilon_{at}v^*_{as}]E[\epsilon_{bs}v^*_{bt}]\bigg] \\
&= E\bigg[\sum_{t,s}\sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v^*_{at}\epsilon_{as}v^*_{as}]\sum_{a}\Psi_{ait}\Psi_{ait}\Psi_{ajs}\Psi_{ajs}\bigg] \\
&+ E\bigg[\sum_{t}\sum_{i,j} q_{it}q_{jt}E[\epsilon_{at}\epsilon_{at}]E[v^*_{bt}v^*_{bt}]\sum_{a\neq b}\Psi_{ait}\Psi_{ajt}\Psi_{bit}\Psi_{bjt}\bigg] \\
&+ E\bigg[\sum_{t,s}\sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v^*_{as}]E[\epsilon_{bs}v^*_{bt}]\sum_{a\neq b}\Psi_{ait}\Psi_{ajs}\Psi_{bit}\Psi_{bjs}\bigg].
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\sum_{a\neq b}\Psi_{ait}\Psi_{ajt}\Psi_{bit}\Psi_{bjt} &= \sum_{a,b}\Psi_{ait}\Psi_{ajt}\Psi_{bit}\Psi_{bjt} - \sum_{a}\Psi^2_{ait}\Psi^2_{ajt} \\
&= \bigg(\sum_{a}\Psi_{ait}\Psi_{ajt}\bigg)^2 - \bigg(\sum_{a}\Psi^2_{ait}\Psi^2_{ajt}\bigg) \\
\sum_{a\neq b}\Psi_{ait}\Psi_{ajs}\Psi_{bit}\Psi_{bjs} &= \sum_{a,b}\Psi_{ait}\Psi_{ajs}\Psi_{bit}\Psi_{bjs} - \sum_{a}\Psi_{ait}\Psi_{ajs}\Psi_{ait}\Psi_{ajs} \\
&= \bigg(\sum_{a}\Psi_{ait}\Psi_{ajs}\bigg)^2 - \bigg(\sum_{a}\Psi^2_{ait}\Psi^2_{ajs}\bigg).
\end{aligned}
$$

Because $\sum_a \Psi_{ait}\Psi_{ajt} = 1$ if $i = j$ and $\sum_a \Psi_{ait}\Psi_{ajt} = 0$ if $i \neq j$ we have

$$
\begin{aligned}
&E\bigg[\sum_t \sum_{i,j} q_{it}q_{jt}E[\epsilon_{at}\epsilon_{at}]E[v^*_{bt}v^*_{bt}]\sum_{a\neq b}\Psi_{ait}\Psi_{ajt}\Psi_{bit}\Psi_{bjt}\bigg] \\
&= E\bigg[\sum_t \sum_{i,j} q_{it}q_{jt}E[\epsilon_{at}\epsilon_{at}]E[v^*_{bt}v^*_{bt}]\bigg(\sum_a \Psi_{ait}\Psi_{ajt}\bigg)^2\bigg] \\
&- E\bigg[\sum_t \sum_{i,j} q_{it}q_{jt}E[\epsilon_{at}\epsilon_{at}]E[v^*_{bt}v^*_{bt}]\bigg(\sum_a \Psi^2_{ait}\Psi^2_{ajt}\bigg)\bigg] \\
&= E\bigg[\sum_t \sum_i q^2_{it}E[\epsilon_{at}\epsilon_{at}]E[v^*_{bt}v^*_{bt}]\bigg] \\
&- E\bigg[\sum_t \sum_{i,j} q_{it}q_{jt}E[\epsilon_{at}\epsilon_{at}]E[v^*_{bt}v^*_{bt}]\bigg(\sum_a \Psi^2_{ait}\Psi^2_{ajt}\bigg)\bigg]
\end{aligned}
$$

and

$$
\begin{aligned}
&E\bigg[\sum_{t,s} \sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v^*_{as}]E[\epsilon_{bs}v^*_{bt}]\sum_{a\neq b}\Psi_{ait}\Psi_{ajs}\Psi_{bit}\Psi_{bjs}\bigg] \\
&= E\bigg[\sum_{t,s} \sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v^*_{as}]E[\epsilon_{bs}v^*_{bt}]\bigg(\sum_a \Psi_{ait}\Psi_{ajs}\bigg)^2\bigg] \\
&- E\bigg[\sum_{t,s} \sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v^*_{as}]E[\epsilon_{bs}v^*_{bt}]\bigg(\sum_a \Psi^2_{ait}\Psi^2_{ajs}\bigg)\bigg]
\end{aligned}
$$

so that

$$E\bigg(\sum_{t=1}^{T-1}\epsilon_t' M_t^\alpha v_t^*\bigg)^2 = E\bigg[\sum_t\sum_i q_{it}^2 E[\epsilon_{at}\epsilon_{at}]E[v_{bt}^* v_{bt}^*]\bigg]$$

$$- E\bigg[\sum_t\sum_{i,j} q_{it}q_{jt}E[\epsilon_{at}\epsilon_{at}]E[v_{bt}^* v_{bt}^*]\bigg(\sum_a \Psi_{ait}^2\Psi_{ajt}^2\bigg)\bigg]$$

$$+ E\bigg[\sum_{t,s}\sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v_{at}^*\epsilon_{as}v_{as}^*]\sum_a \Psi_{ait}^2\Psi_{ajs}^2\bigg]$$

$$+ E\bigg[\sum_{t,s}\sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v_{as}^*]E[\epsilon_{bs}v_{bt}^*]\bigg(\sum_a \Psi_{ait}\Psi_{ajs}\bigg)^2\bigg]$$

$$- E\bigg[\sum_{t,s}\sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v_{as}^*]E[\epsilon_{bs}v_{bt}^*]\bigg(\sum_a \Psi_{ait}^2\Psi_{ajs}^2\bigg)\bigg].$$

Since the second and the fifth elements of the right hand side term are positive, we will prove that the sum of the third and the fourth elements are negligible with respect to the first term. To do that we note that $\sum_a \Psi_{ait}^2\Psi_{ajs}^2$ and $\sum_a \Psi_{ait}\Psi_{ajs}$ are bounded as the Ψ_{ait}^2 are summable because the eigenvectors of M_t^α are orthonormal. Moreover, because $E[\epsilon_{at}v_{at}^*\epsilon_{as}v_{as}^*]$ and $E[\epsilon_{at}v_{as}^*]E[\epsilon_{bs}v_{bt}^*]$ are bounded, we have

$$\sum_{t,s}\sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v_{at}^*\epsilon_{as}v_{as}^*]\sum_a \Psi_{ait}^2\Psi_{ajs}^2 + \sum_{t,s}\sum_{i,j} q_{it}q_{js}E[\epsilon_{at}v_{as}^*]E[\epsilon_{bs}v_{bt}^*]\bigg(\sum_a \Psi_{ait}\Psi_{ajs}\bigg)^2$$

$$= O_p\bigg(\sum_{t,s}\sum_{i,j} q_{it}q_{js}\bigg) = O_p\bigg(\sum_{l,j} q_l q_j\bigg).$$

Because $\sum_t\sum_i q_{it}^2 = \sum_j q_j^2$ and by Lemma 3 of Carrasco and Tchuente (2015, page 438), the sum of the third and the fourth elements are negligible with respect to the first term. Hence, by $0 \le q_j \le 1$, we have

$$\sum_j q_j^2 \le \sum_j q_j = tr(M^\alpha) = O_p\bigg(\frac{1}{\alpha}\bigg).$$

following Lemma 4 (*ii*) of CN (2019). By the Markov's inequality,

$$\sum_{t=1}^{T-1} \epsilon_t' M_t^{\alpha} v_t^* = O_p\left(\left(\sum_j q_j^2\right)^{1/2}\right) = O_p\left(\frac{1}{\sqrt{\alpha}}\right).$$

The following lemma will be used in the proof of Proposition 1.

Lemma 2 (Lemma A4 of Donald and Newey (2001))

If $\hat{A} \xrightarrow{p} A$ and $\hat{B} \xrightarrow{p} B$. A is positive semi definite and B is positive definite, $\tau_0 = argmin_{\tau_1=1} \frac{\tau' A \tau}{\tau' B \tau}$ exists and is unique (with $\tau = (\tau_1, \tau_2')'$ and $\tau_1 \in R$) then

$$\hat{\tau} = argmin_{\tau_1=1} \frac{\tau' \hat{A} \tau}{\tau' \hat{B} \tau} \xrightarrow{p} \tau_0.$$

Proof of proposition 1

Proof of consistency of the regularized LIML estimator

The regularized LIML estimator of the parameter of interest δ is given by :

$$\begin{aligned}
\hat{\delta} &= \underset{\delta}{\text{argmin}} \frac{(y^* - \delta x^*)' M^{\alpha} (y^* - \delta x^*)}{(y^* - \delta x^*)'(y^* - \delta x^*)} \\
&= \underset{\delta}{\text{argmin}} \frac{(1, -\delta) \hat{A} (1, -\delta)'}{(1, -\delta) \hat{B} (1, -\delta)'}
\end{aligned}$$

where $\hat{A} = W' M^{\alpha} W / NT$, $\hat{B} = W'W/NT$ and $W = [y^*, x^*]$.
Then,

$$\begin{aligned}
\hat{A} &= \frac{W' M^{\alpha} W}{NT} \\
&= \frac{1}{NT} \begin{bmatrix} \sum_{t=1}^{T-1} y_t^{*\prime} M_t^{\alpha} y_t^* & \sum_{t=1}^{T-1} y_t^{*\prime} M_t^{\alpha} x_t^* \\ \sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} y_t^* & \sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} x_t^* \end{bmatrix}
\end{aligned}$$

We now calculate the limit of each term of this matrix. From Okui (2009), $x_t^* = \psi_t w_{t-1} - c_t \tilde{v}_{tT}$ with

$$w_{t-1} = \psi_t (y_{t-1} - \mu)$$

where $c_t = \sqrt{\frac{T-t}{T-t+1}}$, $\mu = \frac{\eta}{1-\delta}$, $\phi_j = \frac{1-\delta^j}{1-\delta}$ and $\tilde{v}_{tT} = \frac{\phi_{T-t} v_t + \ldots + \phi_1 v_{T-1}}{T-t}$.
We have,

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha x_t^* = \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}' M_t^\alpha w_{t-1} - \frac{2}{NT}\sum_{t=1}^{T-1} \psi_t c_t w_{t-1}' M_t^\alpha \tilde{v}_{tT} + \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 \tilde{v'}_{tT} M_t^\alpha \tilde{v}_{tT}.$$

But,

$$\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}' M_t^\alpha w_{t-1} = \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}' M_t w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}' (M_t - M_t^\alpha) w_{t-1}.$$

From Lemma C2 of Alvarez and Arellano (2003) and $\psi_t^2 = O\left(1 - 1/(T-t)\right)$, when T goes to infinity and regardless of whether N goes to infinity or not, we have

$$\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}' M_t w_{t-1} \xrightarrow{m.s.} \frac{\sigma^2}{(1-\delta^2)}.$$

By Lemma 3(i) for CN (2019), $\frac{1}{NT}\sum_{t=1}^{T-1} w_{t-1}'[M_t - M_t^\alpha] w_{t-1} = o_p(1)$. As a result, similarly to Alvarez and Arrelano (2003), we have that the limit of $\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}' M_t^\alpha w_{t-1}$ is $\frac{\sigma^2}{(1-\delta^2)}$.
We have just proved that

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha x_t^* \xrightarrow{m.s.} \frac{\sigma^2}{1-\delta^2}.$$

We now look to the term $\sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha y_t^* / NT$ of the matrix $\hat{A}$.
From $y^* = \delta x^* + v^*$, we have

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha y_t^* = \frac{\delta}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha x_t^* + \frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^\alpha v_t^*.$$

The second term of this equality can be decomposed as follows :

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} v_t^* = \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t w_{t-1}' M_t^{\alpha} v_t^* - \frac{1}{NT}\sum_{t=1}^{T-1} c_t \tilde{v'}_{tT} M_t^{\alpha} v_t^*.$$

Moreover,

$$E\left(\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t w_{t-1}' M_t^{\alpha} v_t^*\right) = 0.$$

and

$$\begin{aligned} var\left(\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t w_{t-1}' M_t^{\alpha} v_t^*\right) &= \frac{1}{N^2T^2}\sum_{t=1}^{T-1} E[\psi_t^2 w_{t-1}'(M_t^{\alpha})^2 w_{t-1}] \\ &\leq \frac{\sigma^2}{N^2T^2}\sum_{t=1}^{T-1} E[\psi_t^2 w_{t-1}' w_{t-1}] \\ &= O(1/NT). \end{aligned}$$

so that

$$\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t w_{t-1}' M_t^{\alpha} v_t^* = O_p(1/\sqrt{NT}).$$

Moreover from Lemma 3 (*iii*) and Lemma 4 (*iii*) of CN (2019), we have

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t \tilde{v'}_{tT} M_t^{\alpha} v_t^* = O_p(1/\alpha NT).$$

provided that $\alpha\sqrt{NT} \to \infty$. These two previous results allow us to conclude that

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} v_t^* = o_p(1).$$

Using the limit of $\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} x_t^*$, we can conclude that

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} y_t^* \xrightarrow{p} \frac{\sigma^2\delta}{1-\delta^2}.$$

We now look at the first element of the matrix $\hat{A}$

$$\frac{1}{NT}\sum_{t=1}^{T-1} y_t^{*\prime} M_t^{\alpha} y_t^* = \frac{\delta^2}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} x_t^* + \frac{2\delta}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{\alpha} v_t^* + \frac{1}{NT}\sum_{t=1}^{T-1} v_t^{*\prime} M_t^{\alpha} v_t^*.$$

By Lemma 1 (i), we have

$$\frac{1}{NT}\sum_{t=1}^{T-1} v_t^{*\prime} M_t^{\alpha} v_t^* = O_p\bigg(\frac{1}{\alpha NT}\bigg).$$

which is $o_p(1)$ provided that $\alpha\sqrt{NT} \to \infty$.
Hence the matrix $\hat{A}$ converges to the following positive semi definite matrix A

$$A = \frac{\sigma^2}{(1-\delta^2)}\begin{bmatrix}\delta^2 & \delta \\ & \\ \delta & 1\end{bmatrix}$$

from the result (A91) of Alvarez and Arellano (2003),

$$\hat{B} \xrightarrow{p} B = \frac{\sigma^2}{(1-\delta^2)}\begin{bmatrix}1 & \delta \\ & \\ \delta & 1\end{bmatrix}.$$

The determinant of B is $\sigma^4/(1-\delta^2)$ and the matrix B is positive definite as $|\delta| < 1$. To end the proof of consistency, we need to prove that δ is the unique minimun of $\frac{\tau' A\tau}{\tau' B\tau}$. For that, let's consider $\tau = (1, -\tau_1)$ for any τ_1, we have

$$\begin{aligned}\tau' A\tau &= (1, -\tau_1)' A(1, -\tau_1) \\ &= \frac{\sigma^2}{(1-\delta^2)}(\tau_1 - \delta)^2 \\ &\geq 0\end{aligned}$$

with equality if and only $\tau_1 = \delta$. Since the matrix B is positive definite, we can conclude that δ is the unique minimun of $\frac{\tau' A\tau}{\tau' B\tau}$ and this ends the consistency proof.

Proof of Asymptotic Normality

In the sequel, by abuse of notation, δ sometimes denotes the true value of

the parameter and sometimes denotes a generic value of the parameter.
Let

$$A(\delta) = (y^* - x^*\delta)' M^\alpha (y^* - x^*\delta)$$
$$B(\delta) = (y^* - x^*\delta)'(y^* - x^*\delta)$$

and

$$\wedge(\delta) = \frac{A(\delta)}{B(\delta)}.$$

By definition of the LIML estimator, we have

$$\widehat{\delta^\alpha} = \operatorname{argmin} \wedge(\delta).$$

The gradient and the Hessian are given by

$$\wedge_\delta(\delta) = B(\delta)^{-1}[A_\delta(\delta) - \wedge(\delta) B_\delta(\delta)]$$

$$\begin{aligned}\wedge_{\delta\delta}(\delta) &= B(\delta)^{-1}[A_{\delta\delta}(\delta) - \wedge(\delta) B_{\delta\delta}(\delta)] \\ &\quad - B(\delta)^{-1}[B_\delta(\delta) \wedge'_\delta (\delta) - A_\delta(\delta) \wedge'_\delta (\delta)].\end{aligned}$$

Then by a standard mean-value expansion of the first-order conditions $\wedge_\delta(\hat{\delta}) = 0$, we have

$$\sqrt{NT}(\widehat{\delta^\alpha} - \delta) = - \wedge_{\delta\delta}^{-1} (\tilde{\delta}) \sqrt{NT} \wedge_\delta (\delta)$$

where $\tilde{\delta}$ is the mean-value. Because $\widehat{\delta^\alpha}$ is consistent, $\tilde{\delta} \xrightarrow{p} \delta$. Moreover, note that from Alvarez and Arellano (2003, page 1153)

$$\frac{x^{*\prime}x^*}{NT} \xrightarrow{p} \frac{\sigma^2}{(1-\delta^2)},$$
$$\frac{x^{*\prime}v^*}{NT} \xrightarrow{p} 0,$$
$$\frac{v^{*\prime}v^*}{NT} \xrightarrow{p} \sigma^2.$$

Hence,

$$\begin{aligned}B(\tilde{\delta}) &= \frac{(y^* - x^*\tilde{\delta})'(y^* - x^*\tilde{\delta})}{NT} \\ &= (\delta - \tilde{\delta})' \frac{x^{*\prime}x^*}{NT}(\delta - \tilde{\delta}) + (\delta - \tilde{\delta})\frac{x^{*\prime}v^*}{NT} + \frac{v^{*\prime}v^*}{NT} \\ &\xrightarrow{p} \sigma^2\end{aligned}$$

using the consistency of $\tilde{\delta}$.

$$B_\delta(\tilde{\delta}) = \frac{-2x^{*\prime}(y^* - x^*\tilde{\delta})}{NT} = -2(\delta - \tilde{\delta})\frac{x^{*\prime}x^*}{NT} - \frac{2x^{*\prime}v^*}{NT} \xrightarrow{p} -2\frac{\sigma^2}{1-\delta^2}.$$

$$A(\tilde{\delta}) = (\delta - \tilde{\delta})'\frac{x^{*\prime}M^\alpha x^*}{NT}(\delta_0 - \tilde{\delta}) + 2(\delta - \tilde{\delta})\frac{x^{*\prime}M^\alpha v^*}{NT} + \frac{v^{*\prime}M^\alpha v^*}{NT}.$$

We have by Lemma 1 (i) :

$$\begin{aligned} \frac{x^{*\prime}M^\alpha x^*}{NT} &\leq \frac{x^{*\prime}x^*}{NT} = O_p(1) \\ \frac{v^{*\prime}M^\alpha v^*}{NT} &= O_p\left(\frac{1}{\alpha NT}\right). \end{aligned}$$

Moreover in the proof of the consistency of the regularized GMM estimator, CN (2019) proved that

$$\frac{x^{*\prime}M^\alpha v^*}{\sqrt{NT}} = o_p(1)$$

so that

$$\frac{x^{*\prime}M^\alpha v^*}{NT} = o_p(1).$$

We then have $A(\tilde{\delta}) = o_p(1)$ provided $\alpha NT \to \infty$.
From $B(\tilde{\delta}) = O_p(1)$, we have $\wedge(\tilde{\delta}) = A(\tilde{\delta})/B(\tilde{\delta}) = o_p(1)$.

$$\begin{aligned} A_\delta(\tilde{\delta}) &= \frac{-2x^{*\prime}M^\alpha(y^* - x^*\tilde{\delta})}{NT} \\ &= -2(\delta - \tilde{\delta})\frac{x^{*\prime}M^\alpha x^*}{NT} - 2(\delta - \tilde{\delta})\frac{x^{*\prime}M^\alpha v^*}{NT}. \end{aligned}$$

From $x^{*\prime}M^\alpha x^*/NT = O_p(1)$ and $v^{*\prime}M^\alpha v^*/NT = o_p(1)$ and by the consistency of $\tilde{\delta}$, we obtain $\wedge_\delta(\tilde{\delta}) \to 0$.

$$B_{\delta\delta}(\tilde{\delta}) = \frac{2x^{*\prime}x^*}{NT} \xrightarrow{p} \frac{2\sigma^2}{1-\delta^2},$$

$$A_{\delta\delta}(\tilde{\delta}) = \frac{2x^{*\prime}M^\alpha x^*}{NT} \xrightarrow{p} \frac{2\sigma^2}{1-\delta^2}.$$

Let $\tilde{\sigma}^2 = v^{*\prime}v^*/NT$. We have by continuity

$$\wedge_{\delta\delta}(\tilde{\delta}) = \wedge_{\delta\delta}(\delta) + O_p\left(\frac{1}{\sqrt{NT}}\right)$$

and hence $\tilde{\sigma}^2 \wedge_{\delta\delta} (\tilde{\delta})/2 \xrightarrow{p} \frac{\sigma^2}{1-\delta^2}$.
Moreover noting that for the true value of the parameter δ

$$\begin{aligned}
&B(\delta) = v^{*\prime}v^*,\\
&A_\delta(\delta) = -2x^{*\prime}M^\alpha v^*,\\
&B_\delta(\delta) = -2x^{*\prime}v^*,\\
&\wedge(\delta) = \frac{A(\delta))}{B(\delta)} = \frac{v^{*\prime}M^\alpha v^*}{v^{*\prime}v^*},
\end{aligned}$$

we have

$$\begin{aligned}
-\sqrt{NT}\tilde{\sigma}^2 \wedge_\delta (\delta)/2 &= -\frac{1}{2}\sqrt{NT}\frac{v^{*\prime}v^*}{NT}\left[\frac{-2x^{*\prime}M^\alpha v^*}{v^{*\prime}v^*} - \frac{v^{*\prime}M^\alpha v^*}{v^{*\prime}v^*}\left(\frac{-2x^{*\prime}v^*}{v^{*\prime}v^*}\right)\right]\\
&= \frac{1}{\sqrt{NT}}\left[x^{*\prime}M^\alpha v^* - v^{*\prime}M^\alpha v^*\frac{x^{*\prime}v^*}{v^{*\prime}v^*}\right]\\
&= \frac{1}{\sqrt{NT}}\Bigg[\sum_{t=1}^{T-1}\psi_t w_{t-1}' v_t^* - \sum_{t=1}^{T-1}\psi_t w_{t-1}'(I - M_t^\alpha)v_t^* + \sum_{t=1}^{T-1}\epsilon_t' M_t^\alpha v_t^*\\
&\quad - v^{*\prime}M^\alpha v^*\left(\frac{x^{*\prime}v^*}{v^{*\prime}v^*} - \frac{E[x^{*\prime}v^*]}{\sigma^2 NT}\right) - \sum_{t=1}^{T-1}\left(\frac{E[x^{*\prime}v^*]}{\sigma^2 NT} - \rho_t\right)v_t^{*\prime}M_t^\alpha v_t^*\Bigg].
\end{aligned}$$

Let

$$\Delta_\alpha = \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I - M_t^\alpha)^2 w_{t-1}].$$

From Lemma 5 (i) and (ii) of CN (2019), we have

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w_{t-1}'(I-M_t^{\alpha})v_t^* = O_p\Big(\Delta_{\alpha}^{1/2}\Big) = o_p(1).$$

From Lemma 2,

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\epsilon_t' M_t^{\alpha} v_t^* = O_p\left(\frac{1}{\sqrt{\alpha NT}}\right).$$

which tends to zero if $\sqrt{\alpha NT}\to\infty$.
From Alvarez and Arellano (2003),

$$\frac{x^{*\prime}v^*}{NT} - \frac{E[x^{*\prime}v^*]}{NT} = O_p(1/\sqrt{NT})$$
$$\frac{[v^{*\prime}v^*]}{NT} \xrightarrow{p} \sigma^2.$$

Hence

$$\frac{x^{*\prime}v^*}{v^{*\prime}v^*} - \frac{E[x^{*\prime}v^*]}{\sigma^2 NT} = O_p\left(\frac{1}{\sqrt{NT}}\right)$$

and using Lemma 1 (i)

$$\frac{v^{*\prime}M^{\alpha}v^*}{\sqrt{NT}}\left(\frac{x^{*\prime}v^*}{v^{*\prime}v^*} - \frac{E[x^{*\prime}v^*]}{\sigma^2 NT}\right) = O_p\left(\frac{1}{\alpha\sqrt{NT}}\right)O_p\left(\frac{1}{\sqrt{NT}}\right)$$

which tends to zero provided that $\alpha\sqrt{NT}\to\infty$.
By Lemma 1 (iv)

$$\sum_{t=1}^{T-1}\left(\frac{E[x^{*\prime}v^*]}{\sigma^2 NT} - \rho_t\right)v_t^{*\prime}M_t^{\alpha}v_t^* = O_p\left(\frac{1}{\alpha T\sqrt{NT}}\right)$$

which tends to zero under the assumption that $\alpha\sqrt{NT}\to\infty$.

Finally we have

$$
\begin{aligned}
-\sqrt{NT}\tilde{\sigma}^2 \wedge_\delta (\delta)/2 &= \frac{1}{\sqrt{NT}}\Bigg[\sum_{t=1}^{T-1}\psi_t w_{t-1}' v_t^* - \sum_{t=1}^{T-1}\psi_t w_{t-1}'(I - M_t^\alpha)v_t^* + \sum_{t=1}^{T-1}\epsilon_t' M_t^\alpha v_t^* \\
&\quad - v^{*\prime} M^\alpha v^* \left(\frac{x^{*\prime}v^*}{v^{*\prime}v^*} - \frac{E[x^{*\prime}v^*]}{\sigma^2 NT}\right) - \sum_{t=1}^{T-1}\left(\frac{E[x^{*\prime}v^*]}{\sigma^2 NT} - \rho_t\right) v_t^{*\prime} M_t^\alpha v_t^*\Bigg] \\
&= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w_{t-1}' v_t^* + o_p(1) \\
&\xrightarrow{d} N\left(0, \frac{\sigma^4}{1-\delta^2}\right).
\end{aligned}
$$

The conclusion follows from Slutzky theorem.

Preliminary results for the proof of Proposition 2.

Lemma 3. Let $\tilde{\wedge} = v^{*\prime}M^\alpha v^*/(NT\sigma^2)$, $\wedge_0 = \wedge(\delta_0)$ and $\hat{\wedge} = \wedge(\hat{\delta})$ with $\wedge(\delta) = \frac{(y^*-x^*\delta)' M^\alpha (y^*-x^*\delta)}{(y^*-x^*\delta)'(y^*-x^*\delta)}$. If the assumptions of Proposition 2 are satisfied, then

$$
\begin{aligned}
\hat{\wedge} &= \tilde{\wedge} - (\frac{\sigma^2}{\tilde{\sigma}^2} - 1)\tilde{\wedge} - hH^{-1}h/(NT\sigma^2) + \hat{R}_\wedge \\
&= \tilde{\wedge} + o(1/NT\alpha), \\
\sqrt{NT}\hat{R}_\wedge &= o(\rho_\alpha)
\end{aligned}
$$

where $\rho_\alpha = S(\alpha)$.

Proof of Lemma 3.

Similarly to the calculations in the previous proposition it can be shown that $\wedge(\delta)$ is three times continuously differentiable with derivative that is bounded in probability uniformly in a neighborhood of δ_0, so that for any $\tilde{\hat{\delta}}$ between $\hat{\delta}$ and δ_0, $\wedge_{\delta\delta}(\tilde{\hat{\delta}}) = \wedge_{\delta\delta}(\delta_0) + O_p(1/\sqrt{NT})$. Hence, $\hat{\delta} = \delta_0 + [\wedge_{\delta\delta}(\delta_0)]^{-1}\wedge_\delta(\delta_0) + O_p(1/NT)$. By an expansion of $\wedge(\delta_0)$ around $\hat{\delta}$ we have

$$
\begin{aligned}
\hat{\wedge} &= \wedge(\delta_0) - \frac{1}{2}(\hat{\delta} - \delta_0)^2[\wedge_{\delta\delta}(\delta_0)]^{-1} + O(1/NT\sqrt{NT}) \\
&= \wedge(\delta_0) - \frac{1}{2}(\wedge_\delta(\delta_0))^2[\wedge_{\delta\delta}(\delta_0)]^{-1} + O(1/NT\sqrt{NT}).
\end{aligned}
$$

As in the proof of proposition 1

$$
\begin{aligned}
-\sqrt{NT}\tilde{\sigma}_v^2 \wedge_\delta (\delta_0)/2 &= \frac{1}{\sqrt{NT}}\left(x^{*\prime}M^\alpha v^* - \frac{v^{*\prime}M^\alpha v^*}{v^{*\prime}v^*}x^{*\prime}v^*\right) \\
&= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w_{t-1}' v_t^* - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w_{t-1}'(I - M_t^\alpha)v_t^* \\
&+ \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\epsilon_t' M_t^\alpha v_t^* - \frac{1}{\sqrt{NT}}v^{*\prime}M^\alpha v^*\left[\frac{x^{*\prime}v^*}{v^{*\prime}v^*} - \frac{E[x^{*\prime}v^*]}{NT\sigma^2}\right] \\
&- \frac{1}{\sqrt{NT}}v^{*\prime}M^\alpha v^*\frac{E[x^{*\prime}v^*]}{NT\sigma^2} + \frac{1}{\sqrt{NT}}\sum_t^{T-1}\rho_t v_t^{*\prime}M_t^\alpha v_t^* \\
&= h + O_p(\Delta_\alpha^{1/2}) + O_p\left(\frac{1}{\sqrt{\alpha NT}}\right) + O_p\left(\frac{1}{\alpha\sqrt{NT}}\right)O_p\left(\frac{1}{\sqrt{NT}}\right) \\
&+ O_p\left(\frac{1}{\alpha T\sqrt{NT}}\right)
\end{aligned}
$$

where $h = \sum_{t=1}^{T-1}\psi_t w_{t-1}' v_t^*/\sqrt{NT}$. If we define

$$
H = \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}' w_{t-1}] = \frac{\sigma^2}{1-\delta^2}\frac{1}{T}\sum_{t=1}^{T-1}\psi_t^2
$$

we have

$$
\begin{aligned}
\tilde{\sigma}_v^2 \wedge_{\delta\delta} (\delta_0)/2 &= \frac{1}{NT}\left(x^{*\prime}M^\alpha x^* - \frac{v^{*\prime}M^\alpha v^*}{v^{*\prime}v^*}x^{*\prime}x^*\right) \\
&= \frac{\sigma^2}{1-\delta^2}\frac{1}{T}\sum_{t=1}^{T-1}\psi_t^2 + \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w_{t-1}'w_{t-1}] - H \\
&+ \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t^\alpha)w_{t-1}] \\
&+ \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w_{t-1}'(I-M_t^\alpha)w_{t-1} - E\{w_{t-1}'(I-M_t^\alpha)w_{t-1}\}] \\
&+ \frac{1}{NT}\sum_{t=1}^{T-1}c_t^2\tilde{v}_{tT}'M_t^\alpha\tilde{v}_{tT} \\
&- \frac{2}{NT}\sum_{t=1}^{T-1}c_t\psi_t w_{t-1}'M_t^\alpha\tilde{v}_{tT} \\
&- \wedge_0\left[\frac{1}{NT}\sum_{t=1}^{T-1}x_t^{*\prime}x_t^* - H\right] - H\wedge_0 \\
&= H + O_p(\Delta_\alpha) + O_p(1/\sqrt{NT}) + O_p(1/\alpha NT) \\
&+ O_p(1/\sqrt{NT}) + O_p(1/\alpha NT) \\
&= H + O_p(\Delta_\alpha^{1/2} + 1/\sqrt{\alpha NT})
\end{aligned}
$$

where we used results from Lemma 5 of CN (2019).
Combining the two equations we have

$$\wedge_\delta(\delta_0)[\wedge_{\delta\delta}(\delta_0)]^{-1}\wedge_\delta(\delta_0) = hH^{-1}h/(NT\sigma^2) + O_p(\Delta_\alpha^{1/2}/NT) + 1/\sqrt{\alpha NTN^2T^2}).$$

Using the same decomposition as in Donald and Newey (2001)

$$\begin{aligned}\wedge(\delta_0) &= \frac{v^{*\prime} M^{\alpha} v^*}{v^{*\prime} v^*} \\ &= \frac{\sigma^2}{\tilde{\sigma}^2}\tilde{\wedge} \\ &= \tilde{\wedge} - (\frac{\sigma^2}{\tilde{\sigma}^2} - 1)\tilde{\wedge} + \frac{(\tilde{\sigma}^2 - \sigma^2)^2}{\tilde{\sigma}^2\sigma^2}\tilde{\wedge} \\ &= \tilde{\wedge} - (\frac{\sigma^2}{\tilde{\sigma}^2} - 1)\tilde{\wedge} + O_p\left(\frac{1}{\alpha N^2 T^2}\right) \\ &= \tilde{\wedge} - (\frac{\sigma^2}{\tilde{\sigma}^2} - 1)\tilde{\wedge} + O_p(1/\sqrt{\alpha NTN^2T^2})\end{aligned}$$

where we used Lemma 1 (ii) and delta method to obtain $\tilde{\sigma}^2/\sigma^2 = 1 + O_p(1/\sqrt{NT})$. We have

$$\begin{aligned}\rho_\alpha = &\frac{(1-\delta^2)^2}{\sigma^2}\Bigg\{\frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right) \\ &+ \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I - M_t^\alpha)^2 w_{t-1}]\Bigg\} \\ = &\frac{(1-\delta^2)^2}{\sigma^2}\Bigg\{\frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right) \\ &+ \Delta_\alpha\Bigg\}\end{aligned}$$

Notice that

$$\rho_\alpha \geq \frac{(1-\delta^2)^2}{NT}\left[\sum_{t=1}^{T-1} E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right)\right] \tag{B.2}$$

We will establish the rate of the RHS of (B.4). Because $\phi_j \leq 1/(1-\delta)$

$$
\begin{aligned}
&\left| \frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)^2 \right| \\
&\leq \frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} + \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)^2 \\
&\leq C
\end{aligned}
$$

we have

$$
\begin{aligned}
&\frac{(1-\delta^2)^2}{NT}\left[\sum_{t=1}^{T-1} E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right)\right] \\
&= O\left(\frac{1}{NT}\sum_{t=1}^{T-1} E[tr(M_t^\alpha)^2]\right) \\
&= O\left(\frac{E[tr(M^\alpha)]}{NT}\right) \\
&= O\left(\frac{1}{\alpha NT}\right)
\end{aligned}
$$

so that

$$
\frac{(1-\delta^2)^2}{NT}\left[\sum_{t=1}^{T-1} E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right)\right] = O\left(\frac{1}{\alpha NT}\right) \tag{B.3}
$$

and

$$
\begin{aligned}
\rho_\alpha = &\frac{(1-\delta^2)^2}{\sigma^2}\left\{\frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right)\right. \\
&\left. + \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t^\alpha)^2 w_{t-1}]\right\} \\
&= O_p(1/\alpha NT + \Delta_\alpha).
\end{aligned}
$$

We have

$$\sqrt{NT}\times 1/\sqrt{\alpha NTN^2T^2} = 1/\sqrt{\alpha N^2T^2} = o(1/\alpha NT) = o(\rho_\alpha)$$
$$\sqrt{NT}\Delta_\alpha^{1/2}/NT = o(\rho_\alpha)$$

and

$$\hat{\wedge} = \tilde{\wedge} - (\frac{\sigma^2}{\tilde{\sigma}^2} - 1)\tilde{\wedge} - hH^{-1}h/(2NT\sigma^2) + \hat{R}_\wedge$$

and $\sqrt{NT}\hat{R}_\wedge = o(\rho_\alpha)$.
To end the proof of this Lemma we have to prove that $\hat{\wedge} = \tilde{\wedge} + o(1/NT\alpha)$.
By $\tilde{\sigma}^2/\sigma^2 = 1 + O_p(1/\sqrt{NT})$, $\tilde{\wedge} = O_p(1/\alpha NT)$, $h = O_p(1)$ and $\sqrt{NT}\hat{R}_\wedge = o(\rho_\alpha)$ we have

$$\hat{\wedge} - \tilde{\wedge} = -(\frac{\sigma^2}{\tilde{\sigma}^2} - 1)\tilde{\wedge} - \frac{hH^{-1}h}{2NT\sigma^2} + \sqrt{NT}\hat{R}_\wedge\frac{1}{\sqrt{NT}} = o(\rho_\alpha).$$

Lemma 4. (Lemma 2 of Okui (2009))
Let $\rho_\alpha = trS(\alpha)$. Suppose that an estimator $\hat{\delta}$ has a decomposition $\sqrt{NT}\left(\hat{\delta} - \delta\right) = \widehat{H}^{-1}\hat{h}$, $\hat{h} = h + T^h + Z^h$, $\widehat{H} = H + T^H + Z^H$,
$\left(h + T^h\right)\left(h + T^h\right)' - hh'H^{-1}T^{H\prime} - T^H H^{-1}hh' = \hat{A} + Z^A$,
such that $T^h = o_p(1)$, $h = O_p(1)$, $H = O_p(1)$, the determinant of H is bounded away from zero with probability approaching 1, $\rho_\alpha = o_p(1)$,
$\left\|T^H\right\|^2 = o_p(\rho_\alpha)$, $\left\|T^h\right\|\left\|T^H\right\| = o_p(\rho_\alpha)$, $\left\|Z^h\right\| = o_p(\rho_\alpha)$, $\left\|Z^H\right\| = o_p(\rho_\alpha)$, $\left\|Z^A\right\| = o_p(\rho_\alpha)$,
$E\left(\hat{A}\right) = \sigma^2 H + HS(\alpha)H + o_p(\rho_\alpha)$.
Then, the decomposition (2.10) holds for $\hat{\delta}$.

Proof of Proposition 2.
Let $\rho_\alpha = S(\alpha)$. Notice that

$$\rho_\alpha \geq \frac{(1-\delta^2)^2}{NT}\left[\sum_{t=1}^{T-1} E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right)\right] \tag{B.4}$$

and

$$\rho_\alpha \geq \frac{(1-\delta^2)^2}{\sigma^2 NT} \sum_{t=1}^{T-1} \psi_t^2 E[w'_{t-1}(I - M_t^\alpha)^2 w_{t-1}].$$

From (B.3) a term that is $o(1/(\alpha NT))$ is necessarily $o(\rho_\alpha)$.
Since

$$\Delta_\alpha \geq \frac{1}{NT} \sum_{t=1}^{T-1} \psi_t^2 E[w'_{t-1}(I - M_t) w_{t-1}]$$

and from Okui (2009), $logT/T = o(\sum_{t=1}^{T-1} \psi_t^2 E[w'_{t-1}(I - M_t)w_{t-1}]/NT)$ so that $o(logT/T) = o(\rho_\alpha)$.

Our proof of proposition 2 will be very close to those of Carrasco and Tchuente (2015) and Donald and Newey (2001). The LIML estimator is obtained by solving the following first order condition

$$x^{*\prime} M^\alpha (y^* - x^* \hat{\delta}) - \hat{\wedge} x^{*\prime}(y^* - x^* \hat{\delta}) = 0$$

with $\hat{\wedge} = \wedge(\hat{\delta})$. Let us consider

$$\sqrt{NT}(\delta - \hat{\delta}^\alpha) = \widehat{H}^{-1}\widehat{h}$$

with

$$\widehat{h} = \frac{x^{*\prime} M^\alpha v^*}{\sqrt{NT}} - \hat{\wedge} \frac{x^{*\prime} v^*}{\sqrt{NT}}$$

$$\widehat{H} = \frac{x^{*\prime} M^\alpha x^*}{NT} - \hat{\wedge} \frac{x^{*\prime} x^*}{NT}$$

The term $\widehat{h}$ can be decomposed as follows $\hat{h} = h + \sum_{j=1}^{5} T_j^h + Z^h$ with

$$
\begin{aligned}
h &= \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \psi_t w'_{t-1} v_t^*, \\
T_1^h &= -\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \psi_t w'_{t-1} (I - M_t^\alpha) v_t^* = O_p(\Delta_\alpha^{1/2}), \\
T_2^h &= \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \epsilon'_t M_t^\alpha v_t^* = O_p(1/\sqrt{\alpha NT}), \\
T_3^h &= -\tilde{\wedge} h = O_p(1/\alpha NT), \\
T_4^h &= -\tilde{\wedge} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \epsilon'_t v_t^* = O_p(1/\alpha NT), \\
T_5^h &= \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \psi_t w'_{t-1} v_t^* \right)^2 H^{-1} \frac{E[x^{*\prime} v^*]}{\sigma^2 NT} \frac{1}{2\sqrt{NT}} = O_p(1/\sqrt{NT}), \\
Z^h &= -\hat{R}_\wedge \frac{x^{*\prime} v^*}{\sqrt{NT}} - (\hat{\wedge} - \tilde{\wedge} - \hat{R}_\wedge) \sqrt{NT} \left(\frac{x^{*\prime} v^*}{NT} - \frac{E[x^{*\prime} v^*]}{NT} \right)
\end{aligned}
$$

where $\hat{R}_\wedge$ defined in Lemma 3.
From Alvarez and Arellano (2003), $\sqrt{NT}(x^{*\prime} v^*/NT - E[x^{*\prime} v^*]/NT) = O_p(1)$ and by Lemma 3, $Z^h = O(\rho_\alpha)$.

We also have $\hat{H} = H + \sum_{j=1}^{3} T_j^H + Z^H$ with

$$H = \frac{\sigma^2}{1-\delta^2}\frac{1}{T}\sum_{t=1}^{T-1}\psi_t^2,$$

$$T_1^H = -\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I - M_t^\alpha)w_{t-1}] = O_p(\Delta_\alpha),$$

$$T_2^H = -\frac{2}{NT}\sum_{t=1}^{T-1} c_t\psi_t w_{t-1}'\tilde{v}_{tT} = O_p\left(\frac{1}{\sqrt{NT}}\right),$$

$$T_3^H = -\tilde{\wedge}H = O_p\left(\frac{1}{\alpha NT}\right),$$

$$Z^H = \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2\tilde{v}_{tT}'M_t^\alpha\tilde{v}_{tT} - \frac{2}{NT}\sum_{t=1}^{T-1} c_t\psi_t w_{t-1}'(I - M_t^\alpha)\tilde{v}_{tT} - \hat{\wedge}\left[\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime}x_t^*\right] + \tilde{\wedge}H$$
$$+\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w_{t-1}'w_{t-1} - E\{w_{t-1}'w_{t-1}\}] + \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w_{t-1}'M_t^\alpha w_{t-1} - E\{w_{t-1}'M_t^\alpha w_{t-1}\}].$$

From Okui (2009, page 11), $H = O(1)$. We derive the rate of convergence of the Z^H term.
From equation (14) of CN (2019) we have

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2\tilde{v}_{tT}'M_t^\alpha\tilde{v}_{tT} = O_p\left(\frac{1}{\alpha NT}\right).$$

Moreover

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t\psi_t w_{t-1}'M_t^\alpha\tilde{v}_{tT} = O_p\left(\frac{1}{\sqrt{NT}}\right)$$

from Lemma 5 (iv) of CN (2019).
By the result (A14) (page 1139) of Alvarez and Arellano (2003) we have

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime}x_t^* - H = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Moreover by Lemma 3, $\hat{\wedge} - \tilde{\wedge} = o_p(1/\alpha NT)$. Hence, we have

$$\begin{aligned}\hat{\wedge}\left[\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} x_t^*\right] - \tilde{\wedge} H &= (\hat{\wedge} - \tilde{\wedge})\left[\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} x_t^*\right] + \tilde{\wedge}\left[\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime} x_t^* - H\right] \\ &= o_p(1/\alpha NT) + O_p(1/\alpha NT) O_p(1/\sqrt{NT}) \\ &= o(\rho_\alpha).\end{aligned}$$

From Okui (2009),

$$\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 [w_{t-1}' w_{t-1} - E\{w_{t-1}' w_{t-1}\}] = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Hence, we can conclude that

$$\begin{aligned}Z^H &= O_p(1/\alpha NT) + O_p(1/\sqrt{NT}) + O_p(1/\alpha NT) O_p(1/\sqrt{NT}) + O_p(1/\sqrt{NT}) + O_p(1/\sqrt{NT}) \\ &= O_p(1/\sqrt{NT}).\end{aligned}$$

We now apply Lemma 4 with $T^h = \sum_{j=1}^{5} T_j^h$, $T^H = \sum_{j=1}^{3} T_j^H$,

$$\begin{aligned}Z^A &= \left(\sum_{j=3}^{5} T_j^h\right)^2 + 2\left(\sum_{j=3}^{5} T_j^h\right)\left(T_1^h + T_2^h\right), \\ \hat{A} &= h^2 + 2\sum_{j=1}^{5} T_j^h h + \left(T_1^h + T_2^h\right)^2 - 2h^2 H^{-1} \sum_{j=1}^{3} T_j^H\end{aligned}$$

By $\sqrt{\alpha NT} \to \infty$ and $\Delta_\alpha = o(1)$, we have $||T_1^h||||T_j^H|| = o(\rho_\alpha)$, $||T_2^h||||T_j^H|| = o(\rho_\alpha)$ for each j, $||T_k^h||||T_j^H|| = o(\rho_\alpha)$ for each j and $k > 2$, $||T_j^H||^2 = o(\rho_\alpha)$ for each j and $Z^A = o(\rho_\alpha)$.
We now derive $E(\hat{A})$.

$$E(\hat{A}) = E(h^2) + 2\sum_{j=1}^{5} E(T_j^h h) + E\left(T_1^h + T_2^h\right)^2 - 2\sum_{j=1}^{3} E(h^2 H^{-1} T_j^H)$$

We can note that $hT_3^h - h^2 H^{-1} T_3^H = 0$, $E(h^2 H^{-1}(T_1^H + T_2^H)) = -\sigma^2 T_1^H + H^{-1} E(h^2 T_2^H)$, $E(T_1^h h) = \sigma^2 \Delta_\alpha$, $E(T_1^h T_1^h) = \sigma^2 \Delta_\alpha$.

Moreover by Lemma 2, we have

$$\begin{aligned} E[(T_2^h)^2] &= \frac{1}{NT} E\bigg[\sum_t \sum_i q_{it}^2 E[\epsilon_{it}\epsilon_{it}] E[v_{it}^* v_{it}^*] \bigg] + o(1/\alpha NT) \\ &= \frac{1}{NT} E\bigg[\sum_t \sum_i q_{it}^2 E[\epsilon_{it}\epsilon_{it}] E[v_{it}^* v_{it}^*] \bigg] + o(\rho_\alpha) \end{aligned}$$

We have $\sum_i q_{it}^2 = tr(M_t^\alpha)^2$ and $E[v_{it}^* v_{it}^*] = \sigma^2$. Moreover,

$$\begin{aligned} E[\epsilon_{it}\epsilon_{it}] &= Var(\epsilon_{it}) \\ &= Var(u_{it} - \rho_t v_{it}^*) \\ &= Var(u_{it}) - 2Cov(u_{it}, \rho_t v_{it}^*) + \rho_t^2 Var(v_{it}^*) \\ &= c_t^2 Var\bigg(\frac{\phi_{T-t} v_{it} + ... + \phi_1 v_{i,T-1}}{T-t} \bigg) - 2\sigma^2 \rho_t^2 + \rho_t^2 \sigma^2 \\ &= \frac{1}{(T-t)(T-t+1)} Var\bigg(\phi_{T-t} v_{it} + ... + \phi_1 v_{i,T-1} \bigg) - \sigma^2 \rho_t^2 \\ &= \frac{\sigma^2 (\phi_{T-t}^2 + ... + \phi_1^2)}{(T-t)(T-t+1)} - \sigma^2 \rho_t^2. \end{aligned}$$

But from the proof of Lemma 1 (iv), we have

$$\begin{aligned} \rho_t &= \frac{E[-c_t \tilde{v}_{itT} v_{it}^*]}{\sigma^2} \\ &= \frac{1}{\sigma^2} \frac{-\sigma^2}{1-\delta} \bigg(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} \bigg) \\ &= \frac{-1}{1-\delta} \bigg(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} \bigg) \end{aligned}$$

and finally

$$E[\epsilon_{it}\epsilon_{it}] = \frac{\sigma^2 (\phi_{T-t}^2 + ... + \phi_1^2)}{(T-t)(T-t+1)} - \frac{\sigma^2}{(1-\delta)^2} \bigg(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} \bigg)^2.$$

So we have

$$E[(T_2^h)^2] = \frac{\sigma^4}{NT}\sum_{t=1}^{T-1}\left[\frac{[\phi_{T-t}^2+...+\phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)^2\right]E[tr(M_t^\alpha)^2] + o(\rho_\alpha).$$

By the third moment condition and the independence assumption both on the error term v_{it}, we have $E(hT_2^h) = E(hT_5^h) = E(h^2T_2^H) = 0$. Moreover, because $E(v_{it}^5) = 0$, we have $E(hT_4^h) = 0$.
Finally

$$E(\hat{A}) = \frac{\sigma^4}{NT}\sum_{t=1}^{T-1}E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2+...+\phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right)$$
$$+\sigma^2 H + \sigma^2\Delta_\alpha + o(\rho_\alpha).$$

And therefore

$$S(\alpha) = \left(\lim_{T\to\infty} H\right)^{-2}\times$$
$$\left\{\frac{\sigma^4}{NT}\sum_{t=1}^{T-1}E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2+...+\phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right) + \sigma^2\Delta_\alpha\right\}$$
$$= \frac{(1-\delta^2)^2}{\sigma^2}\left\{\frac{\sigma^2}{NT}\sum_{t=1}^{T-1}E[tr(M_t^\alpha)^2]\left(\frac{[\phi_{T-t}^2+...+\phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left[\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right]^2\right)\right.$$
$$\left. + \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t^\alpha)^2 w_{t-1}]\right\}.$$

using the fact that $\lim_{T\to\infty} H = \sigma^2/(1-\delta^2)$. This ends the proof of Proposition 2.

Proof of proposition 3

We want to prove that

$$\frac{S(\hat{\alpha})}{\inf_{\alpha\in\mathcal{E}_T} S(\alpha)} \xrightarrow{P} 1$$

where $\mathcal{E}_T$ is the parameter set for a given regularization scheme. By Lemma A9 of Donald and Newey (2001), it is sufficient to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{S}(\alpha) - S(\alpha)}{S(\alpha)}\right| = o_p(1).$$

Using the fact that, $\frac{(1-\delta^2)^2}{\sigma^2}R(\alpha) \leq S(\alpha)$ and $\frac{(1-\delta^2)^2}{NT}\mathcal{A}(\alpha) \leq S(\alpha)$, we have for some constant C

$$\frac{1}{C}\left|\frac{\widehat{S}(\alpha)-S(\alpha)}{S(\alpha)}\right| \leq \frac{(1-\hat{\delta}^2)^2}{(1-\delta^2)^2}\left|\frac{\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha)}{\mathcal{A}(\alpha)}\right| + \left|\frac{(1-\hat{\delta}^2)^2-(1-\delta^2)^2}{(1-\delta^2)^2}\right|$$
$$+\frac{(1-\hat{\delta}^2)^2/\hat{\sigma}^2}{(1-\delta^2)^2/\sigma^2}\left|\frac{\widehat{R}(\alpha)-R(\alpha)}{R(\alpha)}\right| + \left|\frac{(1-\hat{\delta}^2)^2/\hat{\sigma}^2-(1-\delta^2)^2/\sigma^2}{(1-\delta^2)^2/\sigma^2}\right|.$$

By the consistency of $\hat{\delta}$ and $\hat{\sigma}^2$, we just need to prove that :

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha)}{\mathcal{A}(\alpha)}\right| = o_p(1),$$

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{R}(\alpha)-R(\alpha)}{R(\alpha)}\right| = o_p(1).$$

For the first equality, we have

$$\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha) = \sum_{t=1}^{T-1} tr(M_t^\alpha)^2\left(\widehat{\nu}_t-\nu_t\right) + \sum_{t=1}^{T-1}\left(tr(M_t^\alpha)^2 - E[tr(M_t^\alpha)^2]\right)\nu_t$$

where

$$\nu_t = \frac{[\phi_{T-t}^2+...+\phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)^2$$
$$\hat{\nu_t} = \frac{[\widehat{\phi}_{T-t}^2+...+\widehat{\phi}_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\hat{\delta})^2}\left(\frac{\widehat{\phi}_{T-t}}{T-t}-\frac{\widehat{\phi}_{T-t+1}}{T-t+1}\right)^2.$$

We will use the following result (see Okui (2009, p.13) : For a random sequence, $\{a_k\}_k$, $\sum_k E(a_k^2) = o(1)$ implies that $\sup_k a_k = o_p(1)$.

$$E\left\{\left[\sum_{t=1}^{T-1} tr(M_t^\alpha)^2\left(\widehat{\nu}_t-\nu_t\right)\right]^2\right\} = E\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)^2\right)^2\left(\widehat{\nu}_t-\nu_t\right)^2 \quad \text{(B.5)}$$
$$+\sum_{t\neq s} E\left[tr\left(M_t^\alpha\right)^2\left(\widehat{\nu}_t-\nu_t\right)tr(M_s^\alpha)^2\left(\widehat{\nu}_s-\nu_s\right)\right] \quad \text{(B.6)}$$

Using $tr(M_t^\alpha)^2 \leq tr(M_t^\alpha) \leq C/\alpha$ and, by the consistency of $\hat{\delta}$, $E\left(\widehat{\nu}_t - \nu_t\right)^2 = O\left(1/NT\right)$, we have

$$E\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)\right)^2\left(\widehat{\nu}_t - \nu_t\right)^2 = O\left(\frac{1}{\alpha^2 N}\right).$$

By Cauchy-Schwarz's inequality, the second term on the RHS of (B.5) is also $O(1/N\alpha^2)$.
Moreover given that $\phi_j \leq 1/(1-\delta)$ we have

$$\begin{aligned}&\left|\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)^2\right| \\ &\leq \frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} + \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)^2 \\ &\leq C\end{aligned}$$

so that the result of Lemma 7 (ii) of CN (2019) applies and we can write

$$E\left[\left[\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)^2 - E[tr(M_t^\alpha)^2]\right)\nu_t\right]^2\right] = O\left(\frac{1}{\alpha^2 N}\right).$$

From Lemma 1 (iv) we have $\sigma^2\mathcal{A}\left(\alpha\right)/NT = O(1/\alpha NT)$ so that

$$\begin{aligned}\mathcal{A}(\alpha) &= \sum_{t=1}^{T-1}\left[\frac{[\phi_{T-t}^2 + ... + \phi_1^2]}{(T-t)(T-t+1)} - \frac{1}{(1-\delta)^2}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)^2\right]E[tr(M_t^\alpha)^2] \\ &= O(1/\alpha).\end{aligned}$$

Hence for any $\alpha \in \mathcal{E}_T$ (which is discrete and finite for SC and LF and is compact for TH) we have

$$E\left|\frac{\widehat{\mathcal{A}}\left(\alpha\right) - \mathcal{A}\left(\alpha\right)}{\mathcal{A}\left(\alpha\right)}\right|^2 = O\left(\frac{1}{N}\right).$$

Now summing up over the elements of $\mathcal{E}_T$, we obtain

$$\sum_{\alpha\in\mathcal{E}_T} O\left(\frac{1}{N}\right) = O\left(\frac{T^2}{N}\right)$$

because the cardinal of $\mathcal{E}_T$ is equal to T^2. Hence, $\sup_\alpha \frac{\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha)}{\mathcal{A}(\alpha)} = o_p(1)$ provided $T^2/N \to 0$ (which is true under the condition $T^3/(N\ln(T)^2) \to 0$). To end the proof of proposition 3 we have to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{R}(\alpha)-R(\alpha)}{R(\alpha)}\right| = o_p(1) \tag{B.7}$$

which is exactly the equality (18) of CN (2019) who proved that equation (B.7) holds provided that $T/N \to 0$, $\#\mathcal{E}_T = O(T^2)$ and $T^3/(N\ln(T)^2) \to 0$.

C Chapter 3

	1981	1991	2001	2011	All
Wage	18.50	22.56	26.40	28.85	25.00
	[8.776]	[11.50]	[13.88]	[16.33]	[13.88]
Annual hours worked ($1000)	2.202	2.220	2.279	2.166	2.221
	[0.435]	[0.443]	[0.469]	[0.400]	[0.441]
Earned income ($1000)	39.76	49.05	59.14	61.64	54.53
	[18.10]	[24.61]	[31.28]	[34.65]	[30.09]
Non labor income ($1000)	1.607	2.787	2.144	3.643	2.689
	[4.316]	[5.809]	[7.849]	[12.98]	[8.825]
Newfoundland and Labrador	0.0205	0.0160	0.0142	0.0156	0.0158
	[0.142]	[0.125]	[0.118]	[0.124]	[0.125]
Prince Edward Island	0.00746	0.00740	0.00725	0.00726	0.00732
	[0.0861]	[0.0857]	[0.0848]	[0.0849]	[0.0852]
Nova Scotia	0.0295	0.0315	0.0294	0.0288	0.0299
	[0.169]	[0.175]	[0.169]	[0.167]	[0.170]
New Brunswick	0.0265	0.0238	0.0232	0.0235	0.0239
	[0.161]	[0.153]	[0.151]	[0.152]	[0.153]
Quebec	0.262	0.249	0.251	0.247	0.250
	[0.440]	[0.433]	[0.433]	[0.431]	[0.433]
Ontario	0.352	0.384	0.390	0.384	0.382
	[0.478]	[0.486]	[0.488]	[0.486]	[0.486]
Manitoba	0.0367	0.0383	0.0370	0.0352	0.0368
	[0.188]	[0.192]	[0.189]	[0.184]	[0.188]
Saskatchewan	0.0350	0.0305	0.0271	0.0296	0.0297
	[0.184]	[0.172]	[0.162]	[0.170]	[0.170]
Alberta	0.115	0.104	0.103	0.105	0.105
	[0.319]	[0.305]	[0.304]	[0.307]	[0.307]
British Columbia	0.116	0.116	0.119	0.123	0.119
	[0.320]	[0.320]	[0.324]	[0.329]	[0.324]
Observations	18228	72991	66311	61786	219316

mean coefficients ; sd in brackets.

TABLE C.1 – Descriptive statistics.