A regularization approach to the dynamic panel data model estimation

Marine Carrasco [*] Ada Nayihouba [†]

University of Montreal, CIREQ [‡]

November 2017

[*]Email:marine.carrasco@umontreal.ca

[†]Email:ada.nayihouba@umontreal.ca

[‡]Carrasco thanks SSHRC for partial support.

Abstract

In a dynamic panel data model, the number of moment conditions may be very large even if the time dimension is moderately large. Even though the use of many moment conditions improves the asymptotic efficiency, the inclusion of an excessive number of moment conditions increases the bias in finite samples. An immediate consequence of a large number of instruments is a large dimensional covariance matrix of the instruments. As a consequence, the condition number (the largest eingenvalue divided by the smallest one) is very high especially when the autoregressive parameter is close to unity. Inverting covariance matrix of instruments with high condition number can badly impact the properties of the estimators. This paper proposes a regularization approach to the estimation of such models using three regularization schemes based on three different ways of inverting the covariance matrix of the instruments. Under double asymptotic, we show that our regularized estimators are consistent and asymptotically normal. These regularization schemes involve a regularization or smoothing parameter so that we derive a data driven selection of this regularization parameter based on an approximation of the 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. Regularization corrects the bias of the usual GMM estimator which seems to underestimate the financial development - economic growth effect.

1 Introduction

In this paper, we propose a regularization approach to the estimation of a dynamic panel data model 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 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. 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 regularization 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 dynamic panel model framework.

As in Carrasco (2012) and Carrasco and Tchuente (2015) on cross -sectional data, we compute three regularized estimators : 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 in this work 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 efficient estimator.

We derive the first order asymptotic properties of the modified estimator under double asymptotics following Alvarez and Arellano (2003). As mentionned in Okui (2009), first order analysis even under double asymptotic

does not provide information about the variance of the estimator. 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.
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 Square (2sls) 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 dynamic panel data 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 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. 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 presents the dynamic panel data model and the classical GMM estimator. Section 2 presents regularized estimator whereas section 3 and 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 5. In section 6 is presented the extension of the model to exogenous covariates and the section 7 presents the results of Monte Carlo simulations. An empirical application is presented in section 8 whereas section 9 concludes the paper.
Throughout the paper, we use the notations I and I_q respectively for the $N \times N$ and $\bar{q} \times \bar{q}$ identity matrix.

2 The model

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

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

δ is the parameter of interest and the following assumptions are made on the model such that $|\delta| < 1$, η_i is the unobserved individual effect, v_{it} the idiosyncratic error with mean zero and variance σ^2 conditional on $\eta_i, y_{it-1}...y_{i0}$. For simplicity, we assume that y_{i0} is observed. Morever, 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 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

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

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

$$y_{it}^* = \delta x_{it}^* + v_{it}^* \tag{2}$$

We have $E(x_{i,t}^* v_{it}^*) \neq 0$ so that 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 are interested in 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 the $(T-1) \times \bar{q}$ block diagonal matrix whose t-th block is z_{it}'. The moment condition is then given by
$E(Z_i' v_i^*) = 0$ with $v_i^* = (v_{i1}^*, ..., v_{i,T-1}^*)$. Under 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

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

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

$$\hat{\delta} = \frac{x^{*'}My^*}{x^{*'}Mx^*} \tag{3}$$

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) proved that lagged levels 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, following Doran et Schmidt (2006) in presence of models with many instruments, the marginal contribution of some of them can be small. As a result, many simulations including those of Okui (2009) show that the GMM estimator of dynamic panel data performs poorly in these settings.
This paper proposes a regularization approach to improve the finite sample properties of the estimator. The intuition is that when T is very large, the dimension of the $\bar{q} \times \bar{q}$ matrix $K_n = Z'Z$ is also large and some of the eigenvalues can be too small. The poor finite sample properties of the estimator in such models arise because inverting these small eigenvalues amplifies the potential sampling errors. So regularization can be seen as a way to smooth the procedure of inverting the eigenvalues to reduce the variability of estimated weighting matrix and improve the finite sample properties of the estimator. We propose to use a regularized inverse of K_n instead of K_n^{-1}, the usual inverse of a matrix, to compute the GMM estimator. Computing the regularized inverse of this matrix introduces a regularization parameter which can be chosen to minimize the approximate MSE. The regularized estimator is presented in details in the next section.

3 The regularized estimator

Let $K = E(Z_i'Z_i)/\bar{q}$ and $(\lambda_l$,ϕ_l,l=1,2,...) be the eigenvalues and orthonormal eigenvectors of K. The matrix K is the operator using the genelarized inverse vocabulary and is assumed to be a nuclear (also called trace-class) operator which is satisfied if and only if its trace is finite. This assumption is discussed in detail in Carrasco and Florens (2014). Moreover, K is also assumed to be nonsingular so that it has only nonzero eigenvalues. In regularization approach, a generalized inverse of K is used instead of its usual inverse K^{-1} to compute the regularized GMM estimator. By spectral (eigenvalue–eigenvector) decomposition, we have $K= P'DP$ with $P'P =I_{\bar{q}}$ where P is the matrix of eigenvectors and D the diagonal matrix with eigenvalues λ_j on the diagonal. Let K^α denote the regularized inverse of K which is defined as

$$K^\alpha = P'D^\alpha P$$

where D^α is the diagonal matrix with elements $q(\alpha, \lambda_j^2)/\lambda_j$.

The real parameter α is the parameter of regularization, a kind of smoothing parameter, and the real function $q(\alpha, \lambda)$ depends on the regularization scheme used. As in Carrasco (2012), three regularization schemes will be used: 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 α be the regularization parameter and λ a given eigenvalue of the matrix K:

1. **Tikhonov regularization (TH):**
   This regularization scheme is close to the well known ridge regression used in presence of multicolinearity to improve properties of Ordinary Least Squares (OLS) estimators. 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 (SC)**
   It consists in selecting the eigenvectors associated with the eigenvalues greater than some threshold.

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

   Another version of this regularization scheme is Principal Components (PC) which consists in using a certain number of eigenvalues 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^α is computed by an iterative procedure with the formula

$$\begin{cases} K_l^\alpha = (I - cK)K_{l-1}^\alpha + cK, l = 1, 2, 3.........., 1/\alpha - 1, \\ \hfill K_0^\alpha = cK \end{cases}$$

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

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

In each regularization scheme, the real valued function $q(\alpha, \lambda)$ satisfies $0 \leq q(\alpha, \lambda) \leq 1$ and $\lim_{\alpha \to 0} q(\alpha, \lambda) = 1$ so that the usual GMM estimator corresponds to a regularized estimator with $\alpha = 0$.
Let us denote by K_n the sample counterpart of K, K_n^α the sample counterpart of K^α the regularized inverse of K, and the matrix $M^\alpha = ZK_n^\alpha Z'$. The regularized GMM estimator for a given regularization scheme is :

$$\widehat{\delta}^\alpha = \frac{x^{*'} M^\alpha y^*}{x^{*'} M^\alpha x^*}. \tag{4}$$

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

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

4 First order asymptotic properties

In this section, we derive the asymptotic properties of the regularized estimator. Following Okui (2009), we make the following assumptions:
Assumption 1: $\{v_{it}\}$ $(t = 1...T; i = 1...N)$ are iid across time and individuals and independent of η_i and y_{i0} with $E(v_{it}) = 0$, $\text{var}(v_{it}) = \sigma^2$, $E(v_{it}^3) = 0$,

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

$E(v_{it}^4) < \infty$.
Assumption 2: The initial observations satisfy

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

where w_{i0} is independent of η_i and iid 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 iid across individuals with $E(\eta_i) = 0$, $\text{var}(\eta_i)=\sigma_\eta^2$ with $0 < \sigma_\eta^2 < \infty$, and finite fourth order moment.
Moreover, asymptotics properties are derived under the assumption that both N and T tend to infinity but with $T < N$. Under this restriction the matrix K_n is non singular and so has nonzero eigenvalues.

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

(1) Consistency: $\hat{\delta}^\alpha \longrightarrow \delta$ in probability.

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

For these properties, we need that α goes to zero slower than $\sqrt{NT}$ goes to infinity. Alvarez and Arellano (2003) show that if T/N tends to a positive scalar, the GMM estimator is asymptotically biased. In our setting, the assymptotic bias vanishes under the assumption $\alpha\sqrt{NT}$ goes to ∞. Since the previous asymptotic properties of the estimator do not depend on the regularization scheme we need to investigate higher order properties to distinguish among them.

5 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 IV literature such as Carrasco (2012), Donald and Newey (2001) and particularly Okui (2009) who works on a dynamic panel 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 $\sigma^2 H^{-1} + S(\alpha)$ in the following

decomposition:

$$NT(\hat{\delta}^{\alpha}-\delta)^2 = Q + r, \quad E(Q) = \sigma^2 H^{-1} + S(\alpha) + R \tag{5}$$

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

Proposition 2
Suppose assumptions 1-3 are satisfied. If $N \to \infty, T \to \infty$, $\alpha \to 0$, $\alpha\sqrt{NT} \to \infty$, *and* $\alpha logT \to 0$, *then for the regularized GMM estimator, the decomposition given in* (5) *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 $\psi_t = c_t(1-\delta\phi_{T-t}/(T-t))$.

With this decomposition we have $S(\alpha) = O(1/(\alpha^2 NT) + \alpha^2)$. The first term of $S(\alpha)$ comes from the square of the bias whereas the second term is from the second order expansion of the variance. As in Carrasco (2012), $S(\alpha)$ is composed of a bias term that increases when α goes to zero and a variance term that decreases when α goes to zero. Unlike in Carrasco (2012) and in Carrasco and Tchuente (2015), our expression of MSE is unconditional as the one 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 our panel 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 assymptotic 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 since it is not an integer except in the case of spectral cut-off regularization scheme. In this latter case, our regularized estimator 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.

6 Data driven selection of the regularization parameter

6.1 Estimation of the approximate MSE

In propositon 2, we derive the leading term 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 notation:

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

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,

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

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

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

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

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

The optimal parameter of regularization is selected by minimizing this estimated $S(\alpha)$

$$\hat{\alpha} = \arg\min_{\alpha \in E_n} \widehat{S}(\alpha) \tag{6}$$

where E_n is the index set of α. E_n is a real compact subset for TH, E_n is such that $\frac{1}{\alpha} \in \{1, 2, ..., \bar{q}\}$ for PC, and E_n is such that $\frac{1}{\alpha}$ is a positive integer no larger than some finite multiple of NT. Next, we analyse the impact of using an estimated version of $S(\alpha)$ to select α instead of the true and and unknow criterion.

6.2 Optimality

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

$$\frac{S(\hat{\alpha})}{\inf_{\alpha \in E_n} S(\alpha)} \xrightarrow{p} 1 \tag{7}$$

as $N \to \infty$, $T \to \infty$. It should be noticied that the result (7) 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 use 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 $TlogT/N \to 0$, then the regularization parameter selection criterion is asymptotically optimal in the sense of (7) *for Principal Components and Landweber Fridman regularization schemes provided that $\#E_n{}^2 = O(T^2)$.*

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 regarded as the error term of the second 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 Kuersteiner (2012) and Okui (2009).
Proposition 3 proves optimality for Principal Components and Landweber

2. $\#E_n$ refers to the number of elements in the set E_n.

Fridman regularization schemes which have discrete index set E_n. The condition $\#E_n = O(T^2)$ is a sufficient condition in the Landweber Fridman regularization scheme since it holds for the principal components case. [3] Rather than imposing a maximum number of iterations, this condition restricts the order of magnitude of the number of elements of the index set E_n. 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 E_n and the fact that the regularization function $q(\alpha, \lambda)$ for in Tikhonov regularization scheme is a real continous function;See Hansen (2007) .

7 Introduction of exogenous covariates

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

$$y_{it} = \delta y_{i,t-1} + \gamma' m_{it} + \eta_i + v_{it} \tag{8}$$

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 omitted in proofs, time-invariant variables are widely used in applied in empirical works.
Let us define $\theta = (\delta, \gamma')'$, $x_{it} = (y_{i,t-1}, m'_{it})'$ et 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 foward orthogonal deviation operator and denote $y_i^* = Ay_i$, $x_i^* = Ax_i$, $v_i^* = Av_i$. The model becomes:

$$y_{it}^* = \theta x_{it}^* + v_{it}^* \tag{9}$$

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 $K_n = Z'Z$ and K_n^α a regularized inverse of K_n given the parameter of regularization α, then the regularized GMM estimator of θ is :

$$\widehat{\theta^\alpha} = \left(x^{*'} M^\alpha y^*\right)\left(x^{*'} M^\alpha x^*\right)^{-1} \tag{10}$$

3. Recall that $\#E_n = \bar{q}$ for principal components case.

with $M^\alpha = ZK_n^\alpha Z'$, $Z = (Z_1', ..., Z_N')'$ 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 this general model. Let $E_Z(a) = E(a|z_{it}, z_{i,t-1}, ...)$ 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 iid across individuals with $E(\eta_i) = 0$, var(η_i)=σ_η^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 in the model with covariates, the optimal instrument for the endogenous variable x_{it}^* is $E_Z(x_{it}^*) = w_{i,t-1} = (\widetilde{w}_{i,t-1}, m_{it}^*)$ with

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

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 $K = E(Z_i'Z_i)/\bar{q}$ and (λ_l,ϕ_l,l=1,2,...) be the eigenvalues and orthonormal eigenvectors of K. The matrix K is assumed to have a finite trace as in the simple model. However, in the extended model we made an assumption on the growth rate of its eigenvalues. If we define $W = (w_1', ..., w_N')'$ with $w_i = (w_{i1}, ..., w_{i,T-1})'$, then we make the following assumption:
Assumption 3: There is a $\beta \geq 1/2$ such that:

$$\sum_{l=1}^{\infty} \frac{E(< \phi_l, ZW^j >^2)}{\lambda_l^{2\beta+1}} < \infty$$

for $j = 1, 2, ...L_m + 1$. $<,>$ denotes inner product in $R^{\bar{q}}$.
This last assumption is similar to Assumption 2(ii) in Carrasco (2012). It allows us used to derive the rate of convergence of the MSE. More precisely, under this assumption we have that $E[||W - M^\alpha W||^2] = O(\alpha^\beta)$ for PC and LF and $E[||W - M^\alpha W||^2] = O(\alpha^{min(\beta,2)})$ for TH. We now prove that under these assumptions, we can isolate the leading term of a second order expansion of the MSE.

Proposition 4:
Assume that assumption 1',2', 3 are satisfied. If α *the parameter of regularization goes to 0,* N *and* T *tend to infinity and* $\alpha^{min\{\beta,1\}}\sqrt{NT} \to \infty$*, then* $S(\alpha)$ *the leading term in a high order expansion of the estimator has the following form:*

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

where

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

As in the model without covariates, the first part of $S(\alpha)$ is the square of the bias and it can be proved to decrease with the parameter of regularization. The second term of $S(\alpha)$ is the second order variance of the regularized estimator and it increases with the regularized parameter.
Moreover, with spectral of cut-off regularization scheme, the form of $S(\alpha)$ is simplified by the projection properties of the matrix M^α. In this case, only the element (1,1) of the matrix $S(\alpha)$ is nonzero so that we can focus on this scalar to select the parameter of regularization as in Okui (2009). In contrast, for Tikhonov and Landweber regularization scheme, the criteria to minimize in order to select the parameter of regularization is a matrix. Hence, α can be selected to minimized $\ell' S(\alpha)\ell$ with ℓ a $l_m + 1$ vector.
For the estimation of $S(\alpha)$, similarly to the model without covariates, $\mathcal{A}(\alpha)$ can be estimated by

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

where $K_{n,t}$ and $K_{n,t}^\alpha$ are respectivly the $th-block$ of K_n and its regularized inverse of K_n^α. Unknown parameters as σ^2 and ϕ_j are estimated using estimated using a preleminairy estimation of θ. Moreover $E[w'_{t-1}(I - M_t^\alpha)^2 w_{t-1}]$ can be estimated by $x_t^{*'}(I - M_t^\alpha)^2 x_t^*$ where $x_t^* = (u_t, m_t')'$ with $u_t = y_{t-1}$.

8 Simulation study

In this section, we present Monte Carlo simulations to illustrate the quality of our regularized estimators and compare them to others estimators such as the usual GMM estimator and the one presented in Okui (2009). In our simulations, we consider the following autoregressive model :

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

with $\eta_i \sim iidN(0, \sigma_\eta^2)$, $y_{i0} \sim iidN(\eta_i/(1-\delta), \sigma^2/(1-\delta^2))$ and $v_{it} \sim iidN(0, \sigma^2)$. Each simulation corresponds to a choice of vector $(N, T, \sigma^2, \sigma_\eta^2, \delta)$. We consider $N = 50$ and $N = 100$. For each values of N, we simulated for $T = 5$, $T = 10$, $T = 15$, $T = 25$ and for three values of δ=(0.5,0.75,0.9). The number of replications is 1000 for all cases. Five 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 estimators are denoted as TH for the Tikhonov one, PC for principal components and LF for Landweber-Fridman.
In order to choose the parameter of regularization α, we minimized the estimated version of $S(\alpha)$ given in the previous section. As a convergent estimator of δ, we used the IVK estimator, and 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 biais (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 :

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

Table 1 presents the distribution of the condition number of the matrix $K_n = Z'Z$. The condition number is defined as the ratio of the highest eigenvalue on the smallest one. The higher the condition number is, the more ill-conditioned the matrix is from non singularity and so inverting its eigeinvalues can negatively affects the estimator, therefors 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 K_n which is the total number of instruments $\bar{q} = 0.5 \times T \times (T-1)$ for each value of T. From Table 1, the need of regularization increases with

Table 1 – Properties of of the condition number with N=50, $\sigma^2 = 1$, $\sigma^2_\eta = 1$, for 1000 replications

	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

T for a given δ and also increases with δ for a given T.

Table 2 contains summary statistics for the value of the regularization parameter which minimizes the approximate MSE. This regularization parameter is the optimal α for TH, the number of iterations for LF, and the number of principal components for PC. We report the mean, the standard error (std), the mode, the first, the second and the third quartile of the distribution of the regularization parameter. As one can expect, the % of optimal principal components selected in PC decreases where the need of the regularization increases. Hence, when $\delta = 0.5$, on average 80%(a mean of 8 out of a total of $\bar{q} = 10$) of the principal components are selected for $T = 5$ whereas around 62% are selected when T increases to 10. For $\delta = 0.9$, the percentage of selected principal components are 50% and 28% respectivly for $T = 5$ and $T = 10$. Those percentage are even smaller for high values of T especially when δ is close to unity.
For TH, for a given value of T, the optimal α increases with δ as 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 number are associated with small α. This is because the bias of GMM estimator decreases as T increases so that regularized estimator tends to be close to the GMM one by selected an α close to 0. Note that Tikhonov

Table 2 – Properties of the distribution of the regularization parameters with N=50 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 1$, for 1000 replications

		Mean	std	q1	Median	q3
$\delta = 0.50$						
	TH	0.00647	0.00056	0.00056	0.00094	0.00163
T= 5	PC	7.851	8.000	7.000	8.000	9.000
	LF	885.982	0.651	423.873	709.142	1130.685
	TH	0.00050	0.00038	0.00031	0.00044	0.00063
T= 10	PC	28.962	30.000	26.000	29.000	32.000
	LF	442.521	56.303	270.036	396.889	558.524
$\delta = 0.75$						
	TH	1.73093	0.00013	0.00009	0.00025	0.00169
T= 5	PC	5.889	8.000	4.000	6.000	8.000
	LF	5276.562	0.861	7.625	1179.171	6600.840
	TH	0.01709	0.00019	0.00006	0.00013	0.00025
T= 10	PC	19.439	21.000	17.000	20.000	22.000
	LF	2341.274	0.045	695.787	1649.360	3177.710
$\delta = 0.90$						
	TH	2.80456	0.00013	0.00001	0.00013	0.02975
T= 5	PC	4.977	4.000	4.000	4.000	7.000
	LF	32147.39486	0.67158	1.50986	24.39590	51.32681
	TH	0.11643	0.00013	0.00001	0.00003	0.00216
T= 10	PC	13.544	9.000	9.000	13.000	18.000
	LF	36378.358	0.019	2.847	53.777	21089.670

estimator corresponds to the GMM one when α is egal to zero.
Similarly, GMM corresponds to the LF estimator with infinite number of iterations. So as the optimal α in for TH, the median optimal number of iterations for LF increases as T becomes large especially for high values of δ.

Results comparing the five estimators are presented in Tables 1-8 in Appendix[4]. In the LF regularization scheme, the parameter c is set to 0.1 since it gives the best results. IVK and regularized estimators improve the properties of the GMM estimators since this one is dominated by either the IVK or a regularized estimator whatever the criteria we consider. Comparing IVK and regularized estimators, we note that in terms of bias, IVK dominated regularized estimators when δ is not close to unity. When δ is close to unity, regularized estimators outperform IVK especially PC. The domination of regularized estimator is even larger when σ_η^2/σ^2 is large. Regularized estimators have more concentrated distribution since they have the smallest Iqr. Regularized esitmators have also generally the smallest MSE especially when σ_η^2/σ^2 is large. Finally, regularized estimators have much better coverage rates especially when δ is close to unity. Now, comparisons between regularized estimators show that PC dominates TH and LF for all criteria except for coverage rate. Moreover, we note that the performances of regularized estimators toward IVK are better for small T.

4. It should be noticed that we normalized the instruments matrix so that it becomes a correlation matrix (diagonal elements equal one).This normalization only affects the regularization parameter and not the regularized estimator.

To summarize these results, we note IVK has generally the smallest bias whereas regularized estimators especially PC have the smallest MSE. Moreover, regularized estimators have the best properties when δ is close to unity or σ_η^2/σ^2 is large, settings where GMM and IVK estimators perform poorly.

9 Empirical application

In this section, we estimate the effect of financial development on economic growth using dynamic panel models (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 mesureament errors, DPM are widely used to evaluate the effect of the exogenous components of financial development on economic growth (Levine et al. (2000) ,Beck et al. (2000) among others). This model takes advantage over the usual purely cross-sectional model by adding more variability to the model 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 litterature to adress the financial development-economic growth question is the following :

$$y_{it} = \delta y_{i,t-1} + \beta_1 FinancialDevelopment_{it} + \beta' X_{it} + \eta_i + \epsilon_{it} \tag{11}$$

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 rigorously removes the autoregressive part of the model as g_{it-1} is not inlcuded as a regressor. So we will estimate exactly the regression equation (11) following our theorical model.
Many indicators are used in the litterature 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

indicator 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) considere 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 don't 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 équation (11), 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 roots 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 [5]. We also included time dummies to account for time-specific effects [6].
The GMM dynamic panel estimator with lagged levels as instruments is widely used in litterature to estimate this model. However, this estimator suffers from finite sample bias since the lagged levels are potentially weak instruments. Several papers including Blundell and Bond (2008) show that when the explanatory variables are persistent over time, lagged levels of these variables are weak instruments for the regression equation in differences. In this paper, we derive the regularized GMM estimator of the effect of financial development on economic growth since we have seen through our simulations that regularized estimators outperform the classical GMM estimators in weak

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

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

Table 3 – Impact of financial development on economic growth

	GMM	**TH**	**PC**	**LF**
δ	0.849	0.615	0.760	0.634
	(0.021)	(0.058)	(0.040)	(0.051)
β_1	0.022	0.066	0.035	0.055
	(0.004)	(0.016)	(0.007)	(0.008)
α^*		0.00010	26	5000

instruments settings. We have a total of $\bar{q} = 141$ instruments so that the covariance matrix $K_n = Z'Z$ has a large condition number (of the order of 10^7) and many very small eigenvalues which are a motivation for the use of regularization. The Fisher statistic of the first stage regression is 4 which is a sign of weak instruments framework. Regularization will improve the classical nonregularized GMM estimator.
In Table 1, we report various estimates of the financial development indicator and their standard errors (in brackets). We also report for regularized estimators, the optimal parameter of regularization 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 6. 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 9 points between 0.1 and 0.09 for TH whereas the optimal number of iterations for LF is searched from 1 to 10000. For PC, the number of principal components is selected between 1 and $\bar{q}$.
The GMM estimation leads to a positive and significant estimate for the indicator of financial development: an increase of 1 % in the financial development indicator would be associated with a 0.02 % increase of the real GDP per capita.

The estimate of the effect of financial development on economic growth, regularized estimates 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 20 % of the principal components (26 out of 141) are selected whereas a relatively low number of iterations is selected in the case of LF.

Analysis in terms of confidence interval shows that the coefficient using GMM without regularization is not in the confidence interval of TH([0.0346,0.0974]) and LF([0.0393,0.0707]). At the same time, none the regularized estimators is included in the confidence intervall of the GMM coefficient ([0.0142,0.0298]). 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 dynamic panel data model. It is worth noting that our results may not be comparable to those in the litterature 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.

10 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 the 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.

In this paper, we extended the regularization approach of Carrasco (2012) to the GMM estimator of the dynamic panel data model. 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 addressed 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 interresting extension of this paper will 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.

Appendix. Proofs

We begin by two lemmas which establish some preliminary useful results. We essentially show how we 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 $v_{i(t-j)}{}_{j=1}^{\infty}$.

Lemma 1: 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) $Var(v_t' M_t^\alpha v_t) \leq (2\sigma^4 + \kappa_4) tr[M_t^\alpha M_t^\alpha] \leq (2\sigma^4 + \kappa_4)t$,

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

(iv) $Cov(v_t' M_t^\alpha v_{t+j}, v_{t+j}' M_{t+j}^\alpha v_{t+j}) \leq \kappa_3 \sigma tr(M_{t+j}^\alpha) \leq \kappa_3 (t+j)\sigma$, for j $>$ 0.

Proof of Lemma 1

(i) The $t \times t$ symmetric matrix $(Z_t' Z_t)$ 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$(Z_t' Z_t)^\alpha$ the regularized inverse of $(Z_t' Z_t)$, then

$$tr(M_t^\alpha) = tr[Z_t (Z_t' Z_t)^\alpha Z_t'] = tr[P_t D_t P_t' P_t D(\alpha) P_t'] = tr[D_t D_t(\alpha)] = \sum_{l=1}^{t} q(\alpha, \lambda_l^t).$$

The result follows from $0 \leq q(\alpha, \lambda) \leq 1$.

(ii) As in AA(2003) we denote by $E_t()$ the expectation conditional on η_i and $\{v_{i(t-j)}\}_{j=1}^{\infty}$.

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

where $m(\alpha)^t_{ij}$ is the (i,j) element of the matrix M^α_t. Moreover,

$$E_t(v'_t M^\alpha_t v_t) = tr[M^\alpha_t E_t(v_t v'_t)] = \sigma^2 tr(M^\alpha_t).$$

So that,

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

By definition $d'_t(\alpha)d_t(\alpha)= \sum_i m(\alpha)^t_{ii} m(\alpha)^t_{ii} = tr(M^\alpha_t M^\alpha_t) \leq t$ so that $var_t(v'_t M^\alpha_t v_t) \leq (\kappa_4 + 2\sigma^4)t$ and the result follows by decomposition of conditional variance.

(iii) By the iterative law of expectations, the expectation of $v'_t M^\alpha_t v_{t+j}$ is null for $j > 0$, so that $Var(v'_t M^\alpha_t v_{t+j}) = E(v'_t M^\alpha_t v_{t+j} v'_{t+j} M^\alpha_t v_t)$. Conditioning on t, it follows that

$$\begin{aligned} E_t(v'_t M^\alpha_t v_{t+j} v'_{t+j} M^\alpha_t v_t) &= E_t[tr(M^\alpha_t v_{t+j} v'_{t+j} M^\alpha_t v_t v'_t)] \\ &= tr[M^\alpha_t E_t(v_{t+j} v'_{t+j}) M^\alpha_t E_t(v_t v'_t)] \\ &= \sigma^4 tr[M^\alpha_t M^\alpha_t] \\ &\leq \sigma^4 t. \end{aligned}$$

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

(iv) $Cov(v'_t M^\alpha_t v_{t+k}, v'_{t+k} M^\alpha_{t+k} v_{t+k}) = E(v'_{t+k} M^\alpha_{t+k} v_{t+k} v'_{t+k} M^\alpha_t v_t)$

$$\begin{aligned} E_{t+k}(v'_{t+k} M^\alpha_{t+k} v_{t+k} v'_{t+k} M^\alpha_t v_t) &= E_{t+k}(v'_{t+k} M^\alpha_{t+k} v_{t+k} v'_{t+k}) M^\alpha_t v_t \\ &= \sum_l \sum_i \sum_j m(\alpha)^{t+k}_{ij} E_{t+k}(v_{it+k} v_{jt+k} v_{lt+k}) M^\alpha_t v_t \\ &= \kappa_3 d'_{t+k}(\alpha) M^\alpha_t v_t \end{aligned}$$

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 $E(v'_{t+k} M^\alpha_{t+k} v_{t+k} v'_{t+k} M^\alpha_t v_t) = E(\kappa_3 d'_{t+k}(\alpha) M^\alpha_t v_t)$. Morever, by Cauchy–Schwarz inequality,

$$(d'_{t+k}(\alpha) M^\alpha_t v_t)^2 \leq (d'_{t+k}(\alpha) d_{t+k}(\alpha))(v'_t M^\alpha_t M^\alpha_t v_t)$$

Since $d'_{t+k}(\alpha)d_{t+k}(\alpha) \leq tr[M^{\alpha}_{t+k}M^{\alpha}_{t+k}] \leq t+k$ and $E(v'_t M^{\alpha}_t M^{\alpha}_t v_t) \leq \sigma^2 tr[M^{\alpha}_t M^{\alpha}_t] \leq \sigma^2 t$, by taking expectation of the previous inequality, we have
$E[(d'_{t+k}(\alpha)M^{\alpha}_t v_t)^2] \leq [(t+k)\sigma]^2$. The results (iv) follows by noting that $[E(d'_{t+k}(\alpha)M^{\alpha}_t v_t)]^2 \leq E[(d'_{t+k}(\alpha)M^{\alpha}_t v_t)^2$.

Lemma 2: Let α be the parameter of regularization.

(i) If $N \to \infty$,$T \to \infty$ and $\alpha \to 0$, then

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

(ii) If $N \to \infty$,$T \to \infty$ and $\alpha \to 0$, then

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

(iii)

$$\frac{1}{\sqrt{NT}} E\left[\sum_{t=1}^{T-1} c_t \tilde{v}'_{tT} M^{\alpha}_t v^*_t\right] = O\left(\frac{1}{\alpha\sqrt{NT}}\right).$$

(iv) If $LnT/\alpha NT \to 0$, then

$$Var\left[\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t \tilde{v}'_{tT} M^{\alpha}_t v^*_t\right] = o(1).$$

Proof of Lemma 2

(i) To prove this result, we will prove that

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

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

$$\frac{1}{NT}\sum_{t=1}^{T-1} w'_{t-1}[M_t - M^{\alpha}_t]w_{t-1} = \frac{1}{NT}W'Z[K^{-1}_n - K^{\alpha}_n]Z'W.$$

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 $D_n^\alpha = diag[\frac{q_1}{\lambda_1}......\frac{q_{\bar{q}}}{\lambda_{\bar{q}}}]$ where $\frac{q_l}{\lambda_l}$ is a notation for $q(\alpha, \hat{\lambda}_l)/\hat{\lambda}_l$. Let $U_n = P_n Z' W$ a $\bar{q} \times 1$ vector, then

$$\begin{aligned}
\frac{1}{NT} W' Z [K_n^{-1} - K_n^\alpha] Z' W &= \frac{1}{NT} W' Z P_n' [D_n^{-1} - D_n^\alpha] P_n Z' W \\
&= \frac{1}{NT} U_n' [D_n^{-1} - D_n^\alpha] U_n \\
&= \frac{1}{NT} \sum_{l=1}^{l=\bar{q}} (1 - q_l) \frac{U_{n,l}^2}{\lambda_l} \\
&\leq sup_{\lambda_l} (1 - q_l) \frac{1}{NT} \sum_{l=1}^{l=\bar{q}} \frac{U_{n,l}^2}{\lambda_l} \\
&\leq sup_{\lambda_l} (1 - q_l) \frac{1}{NT} W' Z K_n^{-1} Z' W.
\end{aligned}$$

From Lemma 2 of AA(2003),

$$\frac{1}{NT} W' Z K_n^{-1} Z' W = \frac{1}{NT} \sum_{t=1}^{T-1} w_{t-1}' M_t w_{t-1} < \infty.$$

Moreover, q_l is between 0 and 1 so that $sup_{\lambda_l}(1 - q_l)$ is bounded.
We have just proved that

$$\frac{1}{NT} W' Z [K_n^{-1} - K_n^\alpha] Z' W < \infty.$$

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

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

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

$$\frac{1}{NT} \sum_{t=1}^{T-1} w_{t-1}' [M_t - M_t^\alpha M_t^\alpha] w_{t-1} \leq sup_{\lambda_l} (1 - q_l^2) \frac{1}{NT} \sum_{t=1}^{T-1} w_{t-1}' M_t w_{t-1} < \infty.$$

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

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

Hence

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

But

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

where the last equality comes from Lemma 4 in Carrasco (2012). So that the expectation of the term is $O(1/\alpha\sqrt{NT})$.

(iv) Let's define $\bar{v}_{tT} = (v_t + ... + v_T)/(T-t+1)$. Then we have

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t\tilde{v}_{tT}'M_t^\alpha v_t^* = \Upsilon_{21NT}^\alpha + \Upsilon_{22NT}^\alpha$$

where

1.
$$\Upsilon_{21NT}^\alpha = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} \tilde{v}_{tT}'M_t^\alpha v_t,$$

2.

$$\Upsilon^{\alpha}_{22NT} = -\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\tilde{v}'_{tT}M^{\alpha}_{t}\bar{v}_{tT}.$$

$$Var(\Upsilon^{\alpha}_{21NT}) = \frac{1}{NT}Var\left[\sum_{t=1}^{T-1}\frac{1}{T-t}v'_{t}M^{\alpha}_{t}(\phi_{T-t}v_{t} + ... + \phi_{1}v_{T-1})\right] \;\; = a^{\alpha}_{0NT} + a^{\alpha}_{1NT}.$$

Note that a^{α}_{0NT} and a^{α}_{1NT} have the same form as a_{0NT} and a_{1NT} of AA(2003) but with M^{α}_{t} instead of M_{t}.

$$a^{\alpha}_{0NT} = \frac{1}{NT}\sum_{t=1}^{T-1}\frac{1}{(T-t)^{2}}[\phi^{2}_{T-t}Var(v'_{t}M^{\alpha}_{t}v_{t}) + ... + \phi^{2}_{1}Var(v'_{t}M^{\alpha}_{t}v_{T-1})].$$

Using Lemma 1 (i)-(iii), we can note that

$$Var_{t}(v'_{t}M^{\alpha}_{t}v_{t}) \leq (\kappa^{4} + 2\sigma^{4})tr[M^{\alpha}_{t}M^{\alpha}_{t}]$$

so that

$$Var(v'_{t}M^{\alpha}_{t}v_{t}) = EVar_{t}(v'_{t}M^{\alpha}_{t}v_{t}) + VE_{t}(v'_{t}M^{\alpha}_{t}v_{t}) \;\; \leq (\kappa^{4} + 2\sigma^{4})Etr[M^{\alpha}_{t}M^{\alpha}_{t}].$$

Moreover, for $j > 0$, we have

$$Var(v'_{t}M^{\alpha}_{t}v_{t+j}) = \sigma^{4}Etr[M^{\alpha}_{t}M^{\alpha}_{t}].$$

Hence,

$$\begin{aligned}
a^{\alpha}_{0NT} &= \frac{1}{NT}\sum_{t=1}^{T-1}\frac{1}{1-\delta^{2}}\frac{Etr[M^{\alpha}_{t}M^{\alpha}_{t}]}{(T-t)^{2}}[\kappa_{4} + 2\sigma^{4} + (T-t-1)\sigma^{4}] \\
&\leq \frac{(\kappa_{4} + 2\sigma^{4})}{NT}\sum_{t=1}^{T-1}\frac{Etr[M^{\alpha}_{t}M^{\alpha}_{t}]}{(T-t)} \\
&\leq \frac{(\kappa_{4} + 2\sigma^{4})}{NT}\sum_{t=1}^{T-1}\frac{Etr[M^{\alpha}_{t}]}{(T-t)} \\
&\leq \frac{(\kappa_{4} + 2\sigma^{4})}{NT}\sum_{t=1}^{T-1}Etr[M^{\alpha}_{t}].
\end{aligned}$$

and we can conlcude that

$$a^{\alpha}_{0NT} = O\Big(1/\alpha NT\Big)$$

Now looking to a^{α}_{1NT}, we have

$$a^{\alpha}_{1NT} = \frac{2}{NT}\sum_{t=1}^{T-2}\Big[\sum_{j=1}^{T-t-1}\frac{\phi^2_{T-t-j}cov(v'_t M^{\alpha}_t v_{t+j}, v'_{t+j}M^{\alpha}_{t+j}v_{t+j})}{(T-t-j)(T-t)}\Big]$$

Using (iv) of Lemma 1, we have

$$\begin{aligned}
a^{\alpha}_{1NT} &= \frac{2}{NT}\sum_{t=1}^{T-2}\Big[\sum_{k=1}^{T-t-1}\frac{\phi^2_{T-t-j}cov(v'_t M^{\alpha}_t v_{t+j}, v'_{t+j}M^{\alpha}_{t+j}v_{t+j})}{(T-t-j)(T-t)}\Big] \\
&\leq \frac{2}{NT}\sum_{t=1}^{T-2}\frac{1}{(T-t)}\Big[\sum_{j=1}^{T-t-1}\frac{\kappa_3\sigma Etr(M^{\alpha}_{t+j})}{T-t-j}\Big] \\
&\leq \frac{\kappa_3\sigma}{NT}\sum_{t=1}^{T-2}Etr[M^{\alpha}_t]
\end{aligned}$$

so that

$$a^{\alpha}_{1NT} = O\Big(1/\alpha NT\Big)$$

This allows us to conclude that

$$Var(\Upsilon^{\alpha}_{21NT}) = O\Big(1/\alpha NT\Big)$$

We now look to the term Υ^{α}_{21NT}.

$$Var(\Upsilon^{\alpha}_{22NT}) = b^{\alpha}_{0NT} + b^{\alpha}_{1NT}.$$

where

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

and

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

But, for $s<t$

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

so that

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

and finally

$$
Var(\Upsilon_{22NT}^{\alpha}) = O\left(\frac{1}{\alpha NT}\right) + O\left(\frac{LnT}{\alpha NT}\right) = O\left(\frac{LnT}{\alpha NT}\right).
$$

To end the proof of iv, we note that

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

Using Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
Var\left[\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}c_t\tilde{v}_{tT}'M_t^{\alpha}v_t^*\right] &\leq Var(\Upsilon_{21NT}^{\alpha}) + Var(\Upsilon_{22NT}^{\alpha}) \\
&\quad + 2\left(Var(\Upsilon_{21NT}^{\alpha})\right)^{1/2}\left(Var(\Upsilon_{22NT}^{\alpha})\right)^{1/2} \\
&\leq O\left(\frac{1}{\alpha NT}\right) + O\left(\frac{LnT}{\alpha NT}\right) + O\left(\frac{1}{\sqrt{\alpha NT}}\frac{\sqrt{LnT}}{\sqrt{\alpha NT}}\right) \\
&= O\left(\frac{LnT}{\alpha NT}\right).
\end{aligned}
$$

and provided that $LnT/\alpha NT \to 0$, iv holds.

Proof of Proposition 1

Proof of consistency

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

As in the proof of lemma 2 of AA(2003), we can decompose the numerator as :

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}x_t^{*'}M_t^{\alpha}v_t^*=\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\psi_t w_{t-1}'M_t^{\alpha}v_t^*-\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}c_t\tilde{v}_{tT}'M_t^{\alpha}v_t^* \quad (12)$$

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 (12) is null so that

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

It follows from *iii* of lemma 2 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 to the variance of $(x^{*'}M^{\alpha}v^*)/\sqrt{NT}$. Following the decomposition (A49) in AA(2003)we can write:

$$\frac{1}{\sqrt{NT}}x^{*'}M^{\alpha}v^*=\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{t=T-1}w_{t-1}'M_t^{\alpha}v_t-\Upsilon_{11NT}^{\alpha}-\Upsilon_{12NT}^{\alpha}\right)-\left(\Upsilon_{21NT}^{\alpha}-\Upsilon_{22NT}^{\alpha}\right) \quad (13)$$

where

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

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

We have

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

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

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 2 *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 $Var(\frac{1}{\sqrt{NT}} \sum_{t=1}^{t=T-1} w'_{t-1} M^{\alpha}_t v^*_t)$ goes to $\frac{\sigma^4}{(1-\delta^2)}$.
Now we give the order of magnitude of $\Upsilon^{\alpha}_{11NT}, \Upsilon^{\alpha}_{12NT}, \Upsilon^{\alpha}_{21NT}, \Upsilon^{\alpha}_{22NT}$.

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

For $t \geq s$,

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

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

By similar calculation as AA(2003) we have that Var($\Upsilon^{\alpha}_{11NT}) \longrightarrow 0$.

Next, following calculation (A60) from AA(2003)

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

The last inequality comes from the fact that $M_t - M^{\alpha}_t M^{\alpha}_t$ is non-negative definite so that $E(w'_{t-1} M^{\alpha}_t M^{\alpha}_t w_{t-1}) \leq E(w'_{t-1} M_t w_{t-1})$. From *iv* of lemma 2, the variance of $\Upsilon^{\alpha}_{21NT} - \Upsilon^{\alpha}_{22NT}$ goes to 0 if $lnT/\alpha NT \to 0$.
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 $\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} x^{*'}_t M^{\alpha}_t v^*_t$ tends to $\frac{\sigma^4}{(1-\delta^2)}$ as N and T go to infinity, α goes to zero and $lnT/\alpha NT \to 0$. The expectation of $(x^{*'} M^{\alpha} v^*)/\sqrt{NT}$ tends 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:

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

From Lemma C2 of AA(2003), when T goes to infinity and regardless of whether N goes to infinity or not, we have

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

By (i) of Lemma 2, $\frac{1}{NT}\sum_{t=1}^{T-1} w'_{t-1}[M_t - M_t^\alpha]w_{t-1} = o_p(1)$. 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 we have that

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

where the last equality comes from Lemma 4 in Carrasco (2012). By the Markov inequality,

$$\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 \tilde{v'}_{tT} M_t^\alpha \tilde{v}_{tT} = O\left(\frac{1}{\alpha NT}\right) \tag{14}$$

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 $(x^{*'} M^\alpha v^*)/NT$ tends 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 the decomposition (13) we have

$$\mu_{NT}^\alpha = E((x^{*'} M^\alpha v^*)/\sqrt{NT}) = E[-(\Upsilon_{21NT}^\alpha - \Upsilon_{22NT}^\alpha)/\sqrt{NT}].$$

since we proved that Υ_{11NT}^α and Υ_{12NT}^α have mean zero. It can be proved that $\mu_{NT}^\alpha = \frac{\sigma^2}{1-\delta^2}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} E[tr(M_t^\alpha)]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)$.

Since the variance of Υ_{11NT}^α, Υ_{12NT}^α, Υ_{22NT}^α and Υ_{21NT}^α tend 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 term can be written 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 1 (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

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

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 (iii) of Lemma2, $\mu_{NT}^{\alpha} = o(1)$ 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 3:*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)

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

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

Proof of the Lemma 3

(i) Noting that $\psi_t^2 = O(1)$, this term can ommitted in the proof.

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

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

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

The last two terms are also $o_p(1)$ using results from Lemma 2.

(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)

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

But

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w_{t-1}'M_t^\alpha w_{t-1} - E\{w_{t-1}'w_{t-1}\}] &\leq \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w_{t-1}'M_t w_{t-1} - E\{w_{t-1}'w_{t-1}\}] \\ &\leq \frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2[w_{t-1}'w_{t-1} - E\{w_{t-1}'w_{t-1}\}] \\ &= O_p(1/\sqrt{NT}).\end{aligned}$$

where the last equality is from Okui (2009).

(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

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

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

$$E_t(\tilde{v}_{tT}\tilde{v}'_{sT}) = \frac{\sigma^2}{(T-t)(T-s)}[\phi^2_{T-s} + ... + \phi^2_1] \leq \frac{\sigma^2}{(T-t)}$$

so that

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

Now by Cauchy-Schwarz inequality's inequality

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

Then,

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

Hence

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

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

Proof of Proposition 2

Let $\rho_\alpha = trace(S(\alpha))$. Since

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

and

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

From

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

a term is $o(\rho_\alpha)$ if it is $o(1/(\alpha^2 NT))$. 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 this proposition, we use lemma 2 of Okui (2009)

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

As in Okui (2009) the numerator can be write 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

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

Moreover

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*'} M_t^\alpha x_t^* = H + T_1^H + \sum_{j=1}^{3} Z_j^H$$

with

$$\begin{aligned} T_1^H &= -\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t^\alpha) w_{t-1}] = O_p(\Delta_\alpha), \\ Z_1^H &= \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 [w_{t-1}' M_t^\alpha w_{t-1} - E\{w_{t-1}' M_t^\alpha w_{t-1}\}] = O_p(1/\sqrt{NT}), \\ Z_2^H &= -2\frac{1}{NT}\sum_{t=1}^{T-1} c_t \psi_t w_{t-1}' M_t^\alpha \tilde{v}_{tT} = O_p(1/\sqrt{NT}), \\ Z_3^H &= \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 \tilde{v}_{tT}' M_t^\alpha \tilde{v}_{tT} = O_p(1/\alpha NT). \end{aligned}$$

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 $||T_1^H||||T_1^h|| = O(\Delta_\alpha/\alpha\sqrt{NT}) = o_p(\rho_\alpha)$ and $||T_1^H||||T_2^h|| = O(\Delta_\alpha^{3/2}) = o_p(\rho_\alpha)$ so that we can conclude that $||T_1^H||||T_1^h + T_2^h|| = o_p(\rho_\alpha)$.

We now apply lemma2 of Okui (2009) with Z^A =0 and

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

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 $E(hT_2^h)$=$E(T_1^h T_2^h)$=0. It can easily proved that

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

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

Since $logT/\alpha NT = o_p(\rho_\alpha)$ provided that $\alpha lnT \to 0$. Finally

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

And therefore

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

This ends the proof of the proposition.

Proof of the proposition 3

Lemma 4

By construction the matrix K is a block diagonal matrix and we define K_t its $t-th$ block. Moreover, let us define K_t^α the $t-th$ block of K^α. Then, if $TlogT/N \to 0$

$$\frac{1}{NT}\mathcal{A}^2 = \frac{1}{NT}\left[\sum_{t=1}^{T-1} tr(K_t^\alpha K_t)\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]^2 + o_p(\rho_\alpha).$$

Proof of the Lemma 4

Let us denote by K_n^α the regularized inverse of K_n. If K_n^α is the $t-th$ block of K_n and $K_{n,t}^\alpha$ the $t-th$ block of K_n^α.

$$\begin{aligned}
\mathcal{A} =& \mathcal{A} - \sum_{t=1}^{T-1} tr[K_t^\alpha K_t]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) + \sum_{t=1}^{T-1} tr[K_t^\alpha K_t]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \\
=& \sum_{t=1}^{T-1} E[tr[M_t^\alpha]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) - \sum_{t=1}^{T-1} tr[K_t^\alpha K_t]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \\
&+ \sum_{t=1}^{T-1} tr[K_t^\alpha K_t]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \\
=& E\left[\sum_{t=1}^{T-1}[tr[M_t^\alpha] - tr[K_t^\alpha K_t]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right] \\
&+ \sum_{t=1}^{T-1} tr[K_t^\alpha K_t]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \\
=& E\left[\sum_{t=1}^{T-1}[tr[K_{n,t}^\alpha K_{n,t} - tr[K_t^\alpha K_t]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right] \\
&+ \sum_{t=1}^{T-1} tr[K_t^\alpha K_t]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right).
\end{aligned}$$

so that

$$\frac{1}{NT}\mathcal{A}^2 = \frac{1}{NT}\left[\sum_{t=1}^{T-1} tr(K_t^\alpha K_t)\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]^2$$
$$+ 2\frac{1}{NT}\left[\sum_{t=1}^{T-1} tr(K_t^\alpha K_t)\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]$$
$$\left\{E\left[\sum_{t=1}^{T-1}[tr[K_{n,t}^\alpha K_{n,t}] - tr[K_t^\alpha K_t]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]\right\}$$
$$+ \frac{1}{NT}\left\{E\left[\sum_{t=1}^{T-1}[tr[K_{n,t}^\alpha K_{n,t}] - tr[K_t^\alpha K_t]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]\right\}^2.$$

But for a given t

$$tr[K_{n,t}^\alpha K_{n,t}] - tr[K_t^\alpha K_t] = tr[K_{n,t}^\alpha K_{n,t} - K_t^\alpha K_t] = \sum_{l=1}^{t}[\widehat{q}_l - q_l]$$
$$\leq T sup(\widehat{q}_l - q_l) = O_p(T/\sqrt{N})$$

under our assumtions. Moreover we have

$$\sum_{t=1}^{T-1}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) = \sum_{j=1}^{T-1}\frac{\phi_j}{j} - \sum_{j=2}^{T}\frac{\phi_j}{j} = \phi_1 - \frac{\phi_T}{T} \to 1.$$

Hence

$$\left|\sum_{t=1}^{T-1}[tr[K_{n,t}^\alpha K_{n,t}] - tr[K_t^\alpha K_t]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right| = O_p(T/\sqrt{N})\left|\sum_{t=1}^{T-1}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right|$$
$$= O_p(T/\sqrt{N}).$$

In the proof of proposition 2, we show that for a term to be $o_p(\rho_\alpha)$ it enough to be $o_p(logT/T)$. Hence

$$\frac{1}{NT}\left\{E\left[\sum_{t=1}^{T-1}[tr[K_{n,t}^\alpha K_{n,t}] - tr[K_t^\alpha K_t]]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right]\right\}^2 = O(T/N^2) = o_p(\rho_\alpha).$$

provided that $T/N \to 0$. Moreover since $tr[K_t^\alpha K_t] \leq t$

$$\sum_{t=1}^{T-1} tr[K_t^\alpha K_t]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \quad \leq T\sum_{t=1}^{T-1}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right) \leq T$$

so that

$$\frac{1}{NT}\left[\sum_{t=1}^{T-1} tr(K_t^{\alpha}K_t)\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)\right]$$
$$\left\{E\left[\sum_{t=1}^{T-1}[tr[K_{n,t}^{\alpha}K_{n,t}]-tr[K_t^{\alpha}K_t]]\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)\right]\right\} = O\left(\frac{T}{N^{3/2}}\right) = o_p(\rho_\alpha).$$

For $T/N^{3/2}$ to be $o_p(\rho_\alpha)$, it is enough that $T^2/N^{3/2}logT \to 0$. But this condition is satisfied even for moderatly large T under the assumption that $TLogT/N \to 0$ provided $T < N$.
We can then conclude that

$$\frac{1}{NT}\mathcal{A}^2 = \frac{1}{NT}\left[\sum_{t=1}^{T-1} tr[K_t^{\alpha}K_t]\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)\right]^2 + o_p(\rho_\alpha).$$

Given this Lemma, we now prove the proposition 3 by using the leading term of $\mathcal{A}$

$$\sum_{t=1}^{T-1} tr[K_t^{\alpha}K_t]\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right).$$

So throught out the proof of Proposition 3, $\mathcal{A}$ is actually this leading term.
Proof of proposition 3
To prove that our estimation $\widehat{S(\alpha)}$ is optimal we have to prove that

$$\frac{S(\hat{\alpha})}{\inf_{\alpha\in M_n} S(\alpha)} \underset{\mathrm{P}}{\rightarrow} 1$$

where E_n is the parameter set for a given regularization scheme. By Lemma A9 of Donald and Newey (2001), it is sufficient to prove that

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

But,

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

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

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

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

For the fisrt equality, we have

$$\sup_{E_n}\left|\frac{\widehat{\mathcal{A}}^2-\mathcal{A}^2}{\mathcal{A}^2}\right| = \sup_{M_n}\left|\frac{\widehat{\mathcal{A}}-\mathcal{A}}{\mathcal{A}}\right|\left|\frac{\widehat{\mathcal{A}}+\mathcal{A}}{\mathcal{A}}\right|.$$

Note that, using $K_{n,t}^{\alpha}K_{n,t} = K_t^{\alpha}K_t + O(1/\sqrt{N})$ and $\frac{\widehat{\phi}_{T-t}}{T-t} - \frac{\widehat{\phi}_{T-t+1}}{T-t+1} = \frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} + O(1/\sqrt{NT})$

$$\begin{aligned}
\widehat{\mathcal{A}} &= \sum_{t=1}^{T-1} tr(K_{n,t}^{\alpha}K_{n,t})\left(\frac{\widehat{\phi}_{T-t}}{T-t} - \frac{\widehat{\phi}_{T-t+1}}{T-t+1}\right)\\
&= \sum_{t=1}^{T-1}\left(tr(K_t^{\alpha}K_t) + O_p(1/\sqrt{N})\right)\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1} + O_p(1/\sqrt{NT})\right)\\
&= \mathcal{A} + O\left(\frac{1}{\sqrt{NT}}\right)\sum_{t=1}^{T-1} tr(K_t^{\alpha}K_t) + O\left(\frac{1}{\sqrt{N}}\right)\sum_{t=1}^{T-1}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\\
&\leq \mathcal{A} + o(1)O(\mathcal{A}) + o(1)O(\mathcal{A})
\end{aligned}$$

so that for any α,

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

Moreover

$$\left|\frac{\widehat{\mathcal{A}}+\mathcal{A}}{\mathcal{A}}\right| \leq 2 + \left|\frac{\widehat{\mathcal{A}}-\mathcal{A}}{\mathcal{A}}\right| = O_p(1).$$

We have just proved that

$$\sup_{E_n}\left|\frac{\widehat{\mathcal{A}}^2-\mathcal{A}^2}{\mathcal{A}^2}\right| = o_p(1).$$

To end the proof of optimality we have to prove that

$$\sup_{E_n}\left|\frac{\widehat{R(\alpha)}-R(\alpha)}{R(\alpha)}\right| = o_p(1). \qquad (15)$$

We fisrt consider the spectral cut-off/principal component 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

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

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 we can prove optimality using $\widetilde{R}(\alpha)$ instead of $\widehat{R(\alpha)}$. Given that :

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

Hence to prove (15), we have to prove that:

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

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

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

To prove these three equalities, we will use the following result : For a random sequence, $\{a_k\}_{k=1}^{T-1}$,$sup_k a_k = o_p(1)$, because $Pr(sup_k a_k > \epsilon) \leq \sum_{k=1}^{T-1} E(a_k^2)/\epsilon^2$ Noting that $w'_{t-1}(I - M_t^{\alpha})w_{t-1} \geq w'_{t-1}(I - M_t)w_{t-1}$

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

To prove that

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

We note from Okui (2009), that

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

But since E_n is such that

$$\sum_{E_n} O_p(\frac{1}{NT}) = O_p(\frac{T}{N}).$$

Morever from AA(2003)

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

so that if $T^2/NlogT \to 0$, we have $T/N = o(\frac{logT}{T})$ and we can conclude that

$$\sup_{E_n}\left|\frac{\sum_{t=1}^{T-1}\psi_t^2 w'_{t-1}w_{t-1}/(NT) - \sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}w_{t-1}]/(NT)}{\sum_{t=1}^{T-1}\psi_t^2 E[w'_{t-1}(I - M_t)w_{t-1}]/(NT)}\right| = o_p(1). \quad (16)$$

Having previously proved that

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

we can use the same stategy as in (16) to prove that

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

and then conclude that

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

We now consider the proof of

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

For a given α, we have

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

where the last equality is from Okui (2009). Since if $T/N\to 0$

$$\sum_{E_n}O_p(logT/NT^2)=O_(logT/N)=o(logT/T).$$

then we have

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

For the proof of

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

Following Okui (2009), we can note that

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

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

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

for spectral cut off regularisation scheme.

We now consider the Landweber 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 we $\widetilde{M}_t^\alpha = 2M_t^\alpha - M_t^\alpha M_t^\alpha$. As in the case of spectral cut-off regularization scheme, let's 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 depends on α, we can prove optimaly using $\widetilde{R}(\alpha)$ instead of $\widehat{R}(\alpha)$. Hence we have to prove that

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

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

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

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

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

Since $\widetilde{M_t^\alpha}\le 2M_t^\alpha\le 2M_t$, we can apply the same strategy as in the case of spectral cut off regularization scheme provided that $\#E_n=O(T^2)$ with $\#E_n$ being the number of elements in the parameter set E_n. Impose that $\#E_n=O(T^2)$ is a sufficeint condition to have optimality in the Landweber 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 $\#E_n=O(T^2)$ in the Landweber regularization scheme.

Lemma 5 :*If assumptions 1',2',3 are satisfied, then*

i.

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

ii. Let's define Δ_α by

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

Then

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

Proof of the Lemma 5

In the following proof m_{ij}^t denote the elements of the matrix M_t^α and $\sum_{i\neq k}$ denote the double sum $\sum_i\sum_{i\neq k}$. For manipulation propurse we denote $u_t=-c_t\tilde{v}'_{tT}$.

i

$$
\begin{aligned}
E\left[\sum_{t=1}^{T-1} u_t' M_t^\alpha v_t^*\right]^2 &= E\left[\sum_{t,s} u_t' M_t^\alpha v_t^* v_s^{*'} M_s^\alpha u_s\right] \\
&= E\left[\sum_{t,s}\sum_{i,j,k,l} m_{ij}^t m_{kl}^s u_{it} u_{ks} v_{jt}^* v_{ls}^*\right] \\
&= E\sum_{t,s}\left[\sum_{i,j,k,l} m_{ij}^t m_{kl}^s E[u_{it} u_{ks} v_{jt}^* v_{ls}^*]\right] \\
&= E\sum_{t,s}\left[\sum_{i} m_{ii}^t m_{ii}^s E[u_{it} u_{is} v_{it}^* v_{is}^*] + \sum_{i\neq k} m_{ii}^t m_{kk}^s E[u_{it} v_{it}^*] E[u_{ks} v_{ks}^*]\right. \\
&\left. + \sum_{i\neq j} m_{ij}^t m_{ij}^s E[u_{it} v_{is}^*] E[u_{js} v_{jt}^*] + \sum_{i\neq j} m_{ij}^t m_{ij}^s E[u_{it} u_{is}] E[v_{jt}^* v_{js}^*]\right] \\
&= E\left[\sum_{t,s}[E[u_{it} u_{is} v_{it}^* v_{is}^*]\sum_{i} m_{ii}^t m_{ii}^s\right] \\
&+ E\left[\sum_{t,s} E[u_{it} v_{it}^*] E[u_{ks} v_{ks}^*]\sum_{i\neq k} m_{ii}^t m_{kk}^s\right] \\
&+ E\left[\sum_{t,s}[u_{it} v_{is}^*] E[u_{js} v_{jt}^*]\sum_{i\neq j} m_{ij}^t m_{ij}^s\right] \\
&+ E\left[\sum_{t,s} E[u_{it} u_{is}] E[v_{jt}^* v_{js}^*]\sum_{i\neq j} m_{ij}^t m_{ij}^s\right]
\end{aligned}
$$

But noting that

$$
\begin{aligned}
&\sum_{i\neq j} m_{ij}^t m_{ij}^t = \sum_{ij} m_{ij}^t m_{ij}^t - \sum_{i} m_{ii}^t m_{ii}^t = tr[M_t^\alpha M_t^\alpha] - \sum_{i} m_{ii}^t m_{ii}^t. \\
&\sum_{i\neq j} m_{ij}^t m_{ij}^s = \sum_{ij} m_{ij}^t m_{ij}^s - \sum_{i} m_{ii}^t m_{ii}^s = tr[M_t^\alpha M_s^\alpha] - \sum_{i} m_{ii}^t m_{ii}^s. \\
&\sum_{i\neq k} m_{ii}^t m_{kk}^s = tr[M_t^\alpha] tr[M_s^\alpha].
\end{aligned}
$$

we have

$$E\left[\sum_{t,s} E[u_{it}v_{it}^*]E[u_{ks}v_{ks}^*]\sum_{i\neq k} m_{ii}^t m_{kk}^s\right] = E\left[\sum_{t,s} E[u_{it}v_{it}^*]E[u_{ks}v_{ks}^*]tr[M_t^\alpha]tr[M_s^\alpha]\right]$$
$$= E\left[\sum_{t=1}^{T-1} E[u_{it}v_{it}^*]tr[M_t^\alpha]\right]^2.$$

$$E\left[\sum_{t,s}[u_{it}v_{is}^*]E[u_{js}v_{jt}^*]\sum_{i\neq j} m_{ij}^t m_{ij}^s\right] = E\left[\sum_{t,s} E[u_{it}v_{is}^*]E[u_{js}v_{jt}^*]tr[M_t^\alpha M_s^\alpha]\right]$$
$$- E\left[\sum_{t,s} E[u_{it}v_{is}^*]E[u_{js}v_{jt}^*]\sum_i m_{ii}^t m_{ii}^s\right].$$

$$E\left[\sum_{t,s} E[u_{it}u_{is}]E[v_{jt}^*v_{js}^*]\sum_{i\neq j} m_{ij}^t m_{ij}^s\right] = E\left[\sum_{t,s} E[u_{it}u_{is}]E[v_{jt}^*v_{js}^*]]tr[M_t^\alpha M_s^\alpha]\right]$$
$$- E\left[\sum_{t,s} E[u_{it}u_{is}]E[v_{jt}^*v_{js}^*]\sum_i m_{ii}^t m_{ii}^s\right].$$

Now,

$$E\left[\sum_{t=1}^{T-1} u_t^{'} M_t^\alpha v_t^*\right]^2$$
$$= E\left[\sum_{t,s}\left[E[u_{it}u_{is}v_{it}^*v_{is}^*] - E[u_{it}v_{is}^*]E[u_{js}v_{jt}^*] - E[u_{it}u_{is}\right]E[v_{jt}^*v_{js}^*]]\sum_i m_{ii}^t m_{ii}^s\right]$$
$$+ E\left[\sum_{t,s}\left[E[u_{it}v_{is}^*]E[u_{js}v_{jt}^*] + E[u_{it}u_{is}]E[v_{jt}^*v_{js}^*\right]tr[M_t^\alpha M_s^\alpha]\right]$$
$$+ E\left[\sum_{t,s} E[u_{it}v_{it}^*]E[u_{ks}v_{ks}^*]tr[M_t^\alpha]tr[M_s^\alpha]\right].$$

The final result is obtained by noting/proving that

$$\sum_{t,s}\sum_i m_{ii}^t m_{ii}^s = o\left(\left(\sum_{t=1}^{T-1} tr[M_t^\alpha]\right)^2\right).$$

$$\sum_{t,s} tr[M_t^\alpha M_s^\alpha] = o\left(\left(\sum_{t=1}^{T-1} tr[M_t^\alpha]\right)^2\right).$$

ii . Let ($\hat{\lambda}_l$,$\hat{\phi}_l$, l=1,2,...) be the eigenvalues and orthonormal eigenvectors of K_n, the sample version of K. The element (a,b) of the matrice

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

is

$$\frac{1}{NT}E\sum_l \frac{(1-q_l)^2}{\lambda_l} < ZW_a, \hat{\phi}_l >< ZW_b, \hat{\phi}_l >$$

where W_a is the ath column of the matrix W. Hence,

$$\begin{aligned}\Delta_\alpha &= \frac{1}{NT}tr\left[\sum_{t=1}^{T-1} E[w_{t-1}'(I - M_t^\alpha)^2 w_{t-1}\right] \\ &= \frac{1}{NT}E\sum_a\sum_l \frac{(1-\hat{q}_l)^2}{\hat{\lambda}_l} < ZW_a, \hat{\phi}_l >^2 \\ &\le sup_{\lambda_l}\lambda_l^{2\beta}(1-q_l)^2 \frac{1}{NT}E\sum_a\sum_l \frac{1}{\hat{\lambda}_l^{2\beta+1}} < ZW_a, \hat{\phi}_l >^2 .\end{aligned}$$

At the limit, the sum

$$\frac{1}{NT}E\sum_l \frac{1}{\hat{\lambda}_l^{2\beta+1}} < ZW_a, \hat{\phi}_l >^2$$

is finite by Assumption 3. Then, it is also true for N and T sufficiently large. Hence, the rate of Δ_α is given by $sup_{\lambda_l}\lambda_l^{2\beta}(1-q_l)^2$ which is given in Carrasco et al. (2007, Proposition 3.11).

Proof of the proposition 4:
Let $\rho_\alpha = trace(S(\alpha))$. In the following proof we will use the fact that a term is $o_p(\rho_\alpha)$ if it is $o_p(1/\alpha^2 NT)$ or $o_p(\alpha^\beta)$.
First, we note that

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

Following Okui (2009), we have the following decomposition

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

where

$$\begin{aligned} h &= \begin{bmatrix} \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} \tilde{w}_{t-1}' v_t^* \\ \\ \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} m_{t-1}^* v_t^* \end{bmatrix}, \\ T_1^h &= -\begin{bmatrix} \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} \tilde{w}_{t-1}'(I - M_t^\alpha) v_t^* \\ \\ \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} m_t^*(I - M_t^\alpha) v_t^* \end{bmatrix}, \\ T_2^h &= -\begin{bmatrix} \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} \tilde{v}_{tT}' M_t^\alpha v_t^* \\ \\ 0 \end{bmatrix}. \end{aligned}$$

and

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*'} M_t^\alpha x_t^* = H + T_1^H + \sum_{j=1}^{j=5} Z_j^H$$

where

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

$$Z_1^H = \begin{bmatrix} \frac{1}{NT}\sum_{t=1}^{T-1} \tilde{w}_{t-1}'\tilde{w}_{t-1} & \frac{1}{NT}\sum_{t=1}^{T-1} \tilde{w}_{t-1}' m_t^{*} \\ \frac{1}{NT}\sum_{t=1}^{T-1} m_t^{*'}\tilde{w}_t & \frac{1}{NT}\sum_{t=1}^{T-1} m_t^{*'} m_t^{*} \end{bmatrix} - H,$$

$$T_1^H = -\begin{bmatrix} \frac{1}{NT}\sum_{t=1}^{T-1} E[\tilde{w}_{t-1}'(I - M_t^{\alpha})\tilde{w}_{t-1}] & \frac{1}{NT}\sum_{t=1}^{T-1} E[\tilde{w}_{t-1}'(I - M_t^{\alpha}) m_t^{*}] \\ \frac{1}{NT}\sum_{t=1}^{T-1} E[m_t^{*'}(I - M_t^{\alpha})\tilde{w}_{t-1}] & \frac{1}{NT}\sum_{t=1}^{T-1} E[m_t^{*'}(I - M_t^{\alpha}) m_t^{*}] \end{bmatrix},$$

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

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

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

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

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

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

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

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 $\alpha^{\beta}\sqrt{NT} \to \infty$.

For the term T_1^h, note that using $w_{t-1} = (\tilde{w}_{t-1}, m_t^*)$ it be writen as

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

We have then $E(T_1^h) = 0$.
We have $var(T_1^h) = \sigma^2\Delta_\alpha$ so that $T_1^h = O_p(\Delta_\alpha^{1/2})$ by Markov's inequality. Since from assumption 3, we have $\Delta_\alpha = O_p(\alpha^\beta)$ we can conclude that $T_1^h = o_p(1)$.
Looking to the term T_2^h, using the same strategy as in the model without covariates

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

so by the ii of the Lemma, we have

$$T_2^h = O_p\left(\frac{1}{\alpha\sqrt{NT}}\right).$$

So under the assumption that $\alpha\sqrt{NT} \to \infty$, $T_2^h = o_p(1)$.
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

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

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 lemma2 of Okui (2009). Let define Z^A =0 and

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

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 $E(hh') = \sigma^2 H$,$E\{hT_1^{h'}\} = E\{T_1^h h'\}$,$E(hh'H^{-}T_1^{H'}) = E(T_1^H H^{-1}hh')$. Moreover we have $E\{hT_1^{h'}\} = E(T_1^H H^{-1}hh')$
Given these equalities,

$$E(\widehat{A}) = \sigma^2 H + E(T_1^h T_1^{h'}) + E(T_2^h T_2^{h'}),$$
$$E(T_1^h T_1^{h'}) = \frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E[w_{t-1}'(I - M_t^{\alpha})^2 w_{t-1}]$$

and

$$
\begin{aligned}
E(T_2^h T_2^{h'}) &= E(T_2^h)E(T_2^h)' + var(T_2^h) \\
&= \frac{\sigma^4}{(1-\delta)^2} \begin{bmatrix} \mathcal{A}(\alpha) & 0 \\ 0 & 0 \end{bmatrix} \\
&+ o_p(\rho_\alpha)
\end{aligned}
$$

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

with

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

where

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

Table 1 – Simulations results with N=50 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 1$, for 1000 replications

		GMM	IVK	TH	PC	LF
$\delta = 0.50$						
T= 5	Med.bias	-0.0925	-0.0856	-0.0918	-0.0839	-0.0895
	Med.abs	0.1254	0.1372	0.1312	0.1272	0.1306
	Iqr	0.2115	0.2449	0.2244	0.2238	0.2293
	Med.mse	0.0157	0.0188	0.0172	0.0162	0.0171
	Cov	0.8900	0.9120	0.9150	0.9050	0.9120
T= 10	Med.bias	-0.0618	-0.0387	-0.0518	-0.0444	-0.0512
	Med.abs	0.0667	0.0630	0.0658	0.0623	0.0657
	Iqr	0.1020	0.1146	0.1063	0.1053	0.1084
	Med.mse	0.0045	0.0040	0.0043	0.0039	0.0043
	Cov	0.8830	0.9270	0.9310	0.8970	0.9180
$\delta = 0.75$						
T= 5	Med.bias	-0.2644	-0.2181	-0.2246	-0.2294	-0.2330
	Med.abs	0.2706	0.2780	0.2646	0.2633	0.2772
	Iqr	0.3085	0.4548	0.3635	0.3550	0.3974
	Med.mse	0.0732	0.0773	0.0700	0.0693	0.0769
	Cov	0.7990	0.8570	0.8660	0.8430	0.8700
T= 10	Med.bias	-0.1328	-0.0925	-0.1027	-0.0916	-0.1037
	Med.abs	0.1328	0.1111	0.1142	0.1067	0.1187
	Iqr	0.1253	0.1670	0.1563	0.1600	0.1671
	Med.mse	0.0176	0.0124	0.0130	0.0114	0.0141
	Cov	0.6790	0.8720	0.8710	0.8340	0.8680
$\delta = 0.90$						
T= 5	Med.bias	-0.5226	-0.5375	-0.5023	-0.4874	-0.5168
	Med.abs	0.5226	0.5579	0.5238	0.5170	0.5609
	Iqr	0.4491	0.6267	0.5516	0.5301	0.6102
	Med.mse	0.2731	0.3112	0.2744	0.2673	0.3146
	Cov	0.6160	0.8170	0.8200	0.7670	0.8950
T= 10	Med.bias	-0.2674	-0.2393	-0.2049	-0.1847	-0.2128
	Med.abs	0.2674	0.2481	0.2161	0.1970	0.2236
	Iqr	0.1921	0.3021	0.2464	0.2608	0.2926
	Med.mse	0.0715	0.0616	0.0467	0.0388	0.0500
	Cov	0.3360	0.8020	0.8250	0.8180	0.8560

Table 2 – Simulations results with N=50 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 10$, for 1000 replications

		GMM	IVK	TH	PC	LF
$\delta = 0.50$						
T= 5	Med.bias	-0.1396	-0.1729	-0.1447	-0.1287	-0.1911
	Med.abs	0.1654	0.2036	0.1738	0.1660	0.2426
	Iqr	0.2509	0.2691	0.2630	0.2710	0.3746
	Med.mse	0.0274	0.0414	0.0302	0.0276	0.0589
	Cov	0.8920	0.9070	0.9100	0.8930	0.9200
T= 10	Med.bias	-0.0719	-0.0505	-0.0625	-0.0543	-0.0683
	Med.abs	0.0766	0.0704	0.0725	0.0698	0.0782
	Iqr	0.1023	0.1163	0.1082	0.1139	0.1208
	Med.mse	0.0059	0.0050	0.0053	0.0049	0.0061
	Cov	0.8480	0.9090	0.9110	0.8950	0.9100
$\delta = 0.75$						
T= 5	Med.bias	-0.3275	-0.4233	-0.3286	-0.3180	-0.4161
	Med.abs	0.3402	0.4485	0.3671	0.3609	0.4682
	Iqr	0.3720	0.5054	0.4295	0.4149	0.6004
	Med.mse	0.1158	0.2012	0.1348	0.1302	0.2192
	Cov	0.7570	0.8540	0.8540	0.8130	0.9450
T= 10	Med.bias	-0.1545	-0.1430	-0.1244	-0.1057	-0.1537
	Med.abs	0.1545	0.1540	0.1412	0.1174	0.1749
	Iqr	0.1396	0.1980	0.1708	0.1742	0.2722
	Med.mse	0.0239	0.0237	0.0199	0.0138	0.0306
	Cov	0.6470	0.8410	0.8750	0.8450	0.9010
$\delta = 0.90$						
T= 5	Med.bias	-0.5492	-0.5848	-0.5401	-0.5026	-0.5904
	Med.abs	0.5492	0.6026	0.5767	0.5674	0.6250
	Iqr	0.4429	0.6308	0.6038	0.5463	0.6938
	Med.mse	0.3016	0.3631	0.3325	0.3220	0.3906
	Cov	0.6450	0.8360	0.8520	0.7710	0.9310
T= 10	Med.bias	-0.2845	-0.2964	-0.3225	-0.2261	-0.3041
	Med.abs	0.2845	0.3020	0.3315	0.2333	0.3156
	Iqr	0.1859	0.3212	0.3831	0.2644	0.3476
	Med.mse	0.0809	0.0912	0.1099	0.0544	0.0996
	Cov	0.3150	0.7720	0.9180	0.7610	0.8980

Table 3 – Simulations results with N=50 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 1$, for 1000 replications

		GMM	IVK	TH	PC	LF
$\delta = 0.50$						
T= 15	Med.bias	-0.0509	-0.0262	-0.0422	-0.0345	-0.0407
	Med.abs	0.0525	0.0408	0.0496	0.0455	0.0488
	Iqr	0.0665	0.0715	0.0714	0.0759	0.0738
	Med.mse	0.0028	0.0017	0.0025	0.0021	0.0024
	Cov	0.8360	0.9200	0.9240	0.9110	0.9160
T= 25	Med.bias	-0.0404	-0.0161	-0.0274	-0.0211	-0.0273
	Med.abs	0.0405	0.0253	0.0313	0.0292	0.0310
	Iqr	0.0399	0.0479	0.0470	0.0497	0.0483
	Med.mse	0.0016	0.0006	0.0010	0.0009	0.0010
	Cov	0.7820	0.9340	0.9300	0.9110	0.9240
$\delta = 0.75$						
T= 15	Med.bias	-0.0917	-0.0540	-0.0630	-0.0476	-0.0611
	Med.abs	0.0917	0.0650	0.0677	0.0623	0.0672
	Iqr	0.0774	0.1013	0.0923	0.0936	0.0943
	Med.mse	0.0084	0.0042	0.0046	0.0039	0.0045
	Cov	0.6180	0.8800	0.8930	0.8670	0.8830
T= 25	Med.bias	-0.0618	-0.0285	-0.0364	-0.0283	-0.0355
	Med.abs	0.0618	0.0351	0.0398	0.0347	0.0396
	Iqr	0.0444	0.0577	0.0536	0.0558	0.0548
	Med.mse	0.0038	0.0012	0.0016	0.0012	0.0016
	Cov	0.4950	0.8980	0.9210	0.8910	0.9060
$\delta = 0.90$						
T= 15	Med.bias	-0.1761	-0.1365	-0.1241	-0.1119	-0.1298
	Med.abs	0.1761	0.1429	0.1268	0.1176	0.1378
	Iqr	0.0976	0.1793	0.1532	0.1578	0.1699
	Med.mse	0.0310	0.0204	0.0161	0.0138	0.0190
	Cov	0.2010	0.8110	0.8310	0.8330	0.8570
T= 25	Med.bias	-0.1052	-0.0631	-0.0625	-0.0511	-0.0608
	Med.abs	0.1052	0.0663	0.0661	0.0581	0.0657
	Iqr	0.0547	0.0912	0.0797	0.0820	0.0856
	Med.mse	0.0111	0.0044	0.0044	0.0034	0.0043
	Cov	0.0920	0.8200	0.8690	0.8630	0.8790

Table 4 – Simulations results with N=50 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 10$, for 1000 replications

		GMM	IVK	TH	PC	LF
$\delta = 0.50$						
	Med.bias	-0.0564	-0.0337	-0.0466	-0.0373	-0.0451
T= 15	Med.abs	0.0576	0.0458	0.0515	0.0480	0.0512
	Iqr	0.0710	0.0760	0.0764	0.0768	0.0771
	Med.mse	0.0033	0.0021	0.0027	0.0023	0.0026
	Cov	0.8040	0.9120	0.9250	0.9050	0.9180
	Med.bias	-0.0422	-0.0170	-0.0294	-0.0228	-0.0283
T= 25	Med.abs	0.0423	0.0279	0.0331	0.0318	0.0326
	Iqr	0.0473	0.0500	0.0522	0.0540	0.0538
	Med.mse	0.0018	0.0008	0.0011	0.0010	0.0011
	Cov	0.7310	0.9190	0.9200	0.8910	0.9030
$\delta = 0.75$						
	Med.bias	-0.1000	-0.0646	-0.0730	-0.0587	-0.0881
T= 15	Med.abs	0.1000	0.0757	0.0770	0.0668	0.0955
	Iqr	0.0791	0.1158	0.0938	0.0972	0.1424
	Med.mse	0.0100	0.0057	0.0059	0.0045	0.0091
	Cov	0.6010	0.8670	0.8880	0.8680	0.8880
	Med.bias	-0.0657	-0.0323	-0.0417	-0.0326	-0.0412
T= 25	Med.abs	0.0657	0.0377	0.0446	0.0380	0.0449
	Iqr	0.0480	0.0583	0.0588	0.0628	0.0623
	Med.mse	0.0043	0.0014	0.0020	0.0014	0.0020
	Cov	0.4700	0.8870	0.8940	0.8820	0.8820
$\delta = 0.90$						
	Med.bias	-0.1875	-0.1672	-0.1856	-0.1160	-0.1570
T= 15	Med.abs	0.1875	0.1681	0.1877	0.1217	0.1585
	Iqr	0.0989	0.1767	0.2705	0.1646	0.2130
	Med.mse	0.0352	0.0283	0.0352	0.0148	0.0251
	Cov	0.1600	0.7770	0.9290	0.8240	0.9010
	Med.bias	-0.1130	-0.0786	-0.0924	-0.0548	-0.0757
T= 25	Med.abs	0.1130	0.0799	0.0963	0.0594	0.0811
	Iqr	0.0539	0.0903	0.1566	0.0847	0.1177
	Med.mse	0.0128	0.0064	0.0093	0.0035	0.0066
	Cov	0.0650	0.7830	0.9360	0.8600	0.8940

Table 5 – Simulations results with N=100 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 1$, for 1000 replications

		GMM	IVK	TH	PC	LF
$\delta = 0.50$						
	Med.bias	-0.0580	-0.0487	-0.0555	-0.0491	-0.0541
T= 5	Med.abs	0.0906	0.0958	0.0928	0.0911	0.0929
	Iqr	0.1660	0.1690	0.1729	0.1655	0.1707
	Med.mse	0.0082	0.0092	0.0086	0.0083	0.0086
	Cov	0.9160	0.9260	0.9340	0.9180	0.9310
	Med.bias	-0.0364	-0.0271	-0.0333	-0.0305	-0.0325
T= 10	Med.abs	0.0441	0.0415	0.0422	0.0420	0.0421
	Iqr	0.0689	0.0726	0.0718	0.0723	0.0728
	Med.mse	0.0019	0.0017	0.0018	0.0018	0.0018
	Cov	0.8990	0.9190	0.9260	0.9150	0.9200
$\delta = 0.75$						
	Med.bias	-0.1546	-0.1300	-0.1406	-0.1287	-0.1281
T= 5	Med.abs	0.1792	0.2077	0.1865	0.1844	0.2085
	Iqr	0.2525	0.3640	0.2809	0.2759	0.3074
	Med.mse	0.0321	0.0432	0.0348	0.0340	0.0435
	Cov	0.8520	0.8880	0.8900	0.8720	0.8910
	Med.bias	-0.0810	-0.0559	-0.0649	-0.0561	-0.0621
T= 10	Med.abs	0.0830	0.0779	0.0762	0.0710	0.0774
	Iqr	0.0960	0.1272	0.1107	0.1092	0.1138
	Med.mse	0.0069	0.0061	0.0058	0.0050	0.0060
	Cov	0.7860	0.8920	0.8940	0.8870	0.8830
$\delta = 0.90$						
	Med.bias	-0.4230	-0.4096	-0.3854	-0.3807	-0.4073
T= 5	Med.abs	0.4258	0.4506	0.4102	0.4164	0.4430
	Iqr	0.4154	0.6108	0.4842	0.4778	0.6090
	Med.mse	0.1813	0.2030	0.1683	0.1734	0.1963
	Cov	0.6860	0.8250	0.8280	0.7930	0.9000
	Med.bias	-0.2116	-0.1731	-0.1502	-0.1435	-0.1599
T= 10	Med.abs	0.2116	0.1876	0.1637	0.1543	0.1780
	Iqr	0.1667	0.2608	0.2179	0.2153	0.2370
	Med.mse	0.0448	0.0352	0.0268	0.0238	0.0317
	Cov	0.4290	0.8090	0.8410	0.8170	0.8720

Table 6 – Simulations results with N=100 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 10$, for 1000 replications

		GMM	IVK	TH	PC	LF
$\delta = 0.50$						
	Med.bias	-0.0730	-0.0871	-0.0770	-0.0708	-0.1065
T= 5	Med.abs	0.1068	0.1264	0.1100	0.1115	0.1516
	Iqr	0.1934	0.2042	0.1933	0.1908	0.2529
	Med.mse	0.0114	0.0160	0.0121	0.0124	0.0230
	Cov	0.9260	0.9300	0.9360	0.9310	0.9460
	Med.bias	-0.0423	-0.0269	-0.0390	-0.0312	-0.0415
T= 10	Med.abs	0.0516	0.0484	0.0493	0.0492	0.0531
	Iqr	0.0753	0.0839	0.0772	0.0797	0.0847
	Med.mse	0.0027	0.0023	0.0024	0.0024	0.0028
	Cov	0.9090	0.9300	0.9380	0.9200	0.9350
$\delta = 0.75$						
	Med.bias	-0.2376	-0.3532	-0.2411	-0.2167	-0.3448
T= 5	Med.abs	0.2518	0.3727	0.2688	0.2408	0.3973
	Iqr	0.3024	0.4135	0.3300	0.3075	0.5617
	Med.mse	0.0634	0.1389	0.0723	0.0580	0.1579
	Cov	0.8110	0.8560	0.8650	0.8370	0.9500
	Med.bias	-0.0940	-0.0886	-0.0762	-0.0662	-0.1061
T= 10	Med.abs	0.0940	0.1018	0.0854	0.0804	0.1324
	Iqr	0.1102	0.1457	0.1242	0.1226	0.2283
	Med.mse	0.0088	0.0104	0.0073	0.0065	0.0175
	Cov	0.7550	0.8690	0.8960	0.8690	0.9160
$\delta = 0.90$						
	Med.bias	-0.4979	-0.5664	-0.5164	-0.4869	-0.5953
T= 5	Med.abs	0.4986	0.5795	0.5361	0.5046	0.6147
	Iqr	0.4202	0.5815	0.5507	0.4722	0.6757
	Med.mse	0.2486	0.3358	0.2874	0.2546	0.3779
	Cov	0.6260	0.8200	0.8470	0.7590	0.9140
	Med.bias	-0.2479	-0.2715	-0.3079	-0.1716	-0.2715
T= 10	Med.abs	0.2479	0.2718	0.3097	0.1815	0.2766
	Iqr	0.1731	0.2580	0.3623	0.2230	0.3364
	Med.mse	0.0615	0.0739	0.0959	0.0330	0.0765
	Cov	0.3790	0.8030	0.9300	0.7680	0.9230

Table 7 – Simulations results with N=100 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 1$, for 1000 replications

		GMM	IVK	TH	PC	LF
$\delta = 0.50$						
	Med.bias	-0.0265	-0.0133	-0.0204	-0.0165	-0.0194
T= 15	Med.abs	0.0304	0.0271	0.0291	0.0282	0.0290
	Iqr	0.0488	0.0526	0.0511	0.0505	0.0527
	Med.mse	0.0009	0.0007	0.0008	0.0008	0.0008
	Cov	0.9090	0.9350	0.9520	0.9320	0.9510
	Med.bias	-0.0231	-0.0108	-0.0168	-0.0123	-0.0159
T= 25	Med.abs	0.0242	0.0172	0.0205	0.0195	0.0198
	Iqr	0.0301	0.0325	0.0308	0.0320	0.0320
	Med.mse	0.0006	0.0003	0.0004	0.0004	0.0004
	Cov	0.8640	0.9490	0.9440	0.9270	0.9400
$\delta = 0.75$						
	Med.bias	-0.0558	-0.0318	-0.0384	-0.0326	-0.0366
T= 15	Med.abs	0.0563	0.0427	0.0438	0.0410	0.0442
	Iqr	0.0610	0.0740	0.0718	0.0694	0.0723
	Med.mse	0.0032	0.0018	0.0019	0.0017	0.0019
	Cov	0.7500	0.9040	0.9130	0.8940	0.9030
	Med.bias	-0.0373	-0.0155	-0.0213	-0.0162	-0.0206
T= 25	Med.abs	0.0373	0.0233	0.0257	0.0241	0.0255
	Iqr	0.0314	0.0423	0.0396	0.0406	0.0403
	Med.mse	0.0014	0.0005	0.0007	0.0006	0.0007
	Cov	0.6730	0.9110	0.9250	0.9100	0.9190
$\delta = 0.90$						
	Med.bias	-0.1288	-0.0935	-0.0781	-0.0672	-0.0789
T= 15	Med.abs	0.1288	0.1018	0.0841	0.0769	0.0878
	Iqr	0.0851	0.1499	0.1184	0.1138	0.1227
	Med.mse	0.0166	0.0104	0.0071	0.0059	0.0077
	Cov	0.3450	0.8340	0.8630	0.8520	0.8670
	Med.bias	-0.0733	-0.0397	-0.0340	-0.0261	-0.0327
T= 25	Med.abs	0.0733	0.0475	0.0394	0.0352	0.0400
	Iqr	0.0440	0.0710	0.0589	0.0606	0.0625
	Med.mse	0.0054	0.0023	0.0015	0.0012	0.0016
	Cov	0.2380	0.8510	0.9010	0.8860	0.9030

Table 8 – Simulations results with N=100 ,$\sigma^2 = 1$, $\sigma_\eta^2 = 10$, for 1000 replications

		GMM	IVK	TH	PC	LF
$\delta = 0.50$						
	Med.bias	-0.0279	-0.0166	-0.0223	-0.0176	-0.0208
T= 15	Med.abs	0.0325	0.0296	0.0309	0.0314	0.0313
	Iqr	0.0559	0.0591	0.0598	0.0609	0.0601
	Med.mse	0.0011	0.0009	0.0010	0.0010	0.0010
	Cov	0.8680	0.9240	0.9330	0.9080	0.9270
	Med.bias	-0.0249	-0.0121	-0.0204	-0.0153	-0.0195
T= 25	Med.abs	0.0261	0.0195	0.0232	0.0209	0.0227
	Iqr	0.0328	0.0358	0.0358	0.0360	0.0367
	Med.mse	0.0007	0.0004	0.0005	0.0004	0.0005
	Cov	0.8250	0.9260	0.9330	0.9080	0.9230
$\delta = 0.75$						
	Med.bias	-0.0612	-0.0395	-0.0455	-0.0379	-0.0550
T= 15	Med.abs	0.0613	0.0503	0.0501	0.0450	0.0640
	Iqr	0.0649	0.0770	0.0734	0.0780	0.0969
	Med.mse	0.0038	0.0025	0.0025	0.0020	0.0041
	Cov	0.7140	0.8920	0.9190	0.8960	0.9250
	Med.bias	-0.0402	-0.0202	-0.0251	-0.0191	-0.0243
T= 25	Med.abs	0.0402	0.0266	0.0278	0.0257	0.0279
	Iqr	0.0359	0.0417	0.0419	0.0439	0.0422
	Med.mse	0.0016	0.0007	0.0008	0.0007	0.0008
	Cov	0.6350	0.9030	0.9130	0.8910	0.9080
$\delta = 0.90$						
	Med.bias	-0.1447	-0.1490	-0.1827	-0.0851	-0.1236
T= 15	Med.abs	0.1447	0.1504	0.1883	0.0877	0.1320
	Iqr	0.0983	0.1676	0.2883	0.1342	0.2108
	Med.mse	0.0209	0.0226	0.0355	0.0077	0.0174
	Cov	0.2960	0.7760	0.9450	0.8140	0.9180
	Med.bias	-0.0811	-0.0571	-0.1121	-0.0366	-0.0595
T= 25	Med.abs	0.0811	0.0602	0.1135	0.0416	0.0641
	Iqr	0.0432	0.0755	0.1956	0.0618	0.1110
	Med.mse	0.0066	0.0036	0.0129	0.0017	0.0041
	Cov	0.1780	0.8360	0.9360	0.8610	0.9110

References

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

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

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.

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

Richard Blundell and Steve Bond. GMM estimation with persistent panel data: an application to production functions. IFS Working Papers W99/04, Institute for Fiscal Studies, 1999.

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.

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

Marine Carrasco and Jean-Pierre Florens. On The Asymptotic Efficiency Of Gmm. *Econometric Theory*, 30(02):372–406, 2014.

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.

Marine Carrasco and Guy Tchuente. Regularized liml for many instruments. *Journal of Econometrics*, 186(2):427 – 442, 2015. High Dimensional Problems in Econometrics.

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

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.

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

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

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.

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

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

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.

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

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.

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

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.

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

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