*Econometric Theory*, **40**, 2024, 360–418.
doi:10.1017/S0266466622000469

# REGULARIZED ESTIMATION OF DYNAMIC PANEL MODELS

MARINE CARRASCO
*University of Montreal*

ADA NAYIHOUBA
*World Bank*

In a dynamic panel data model, the number of moment conditions increases rapidly with the time dimension, resulting in a large dimensional covariance matrix of the instruments. As a consequence, the generalized method of moments (GMM) estimator exhibits a large bias in small samples, especially when the autoregressive parameter is close to unity. To address this issue, we propose a regularized version of the one-step GMM estimator using three regularization schemes based on three different ways of inverting the covariance matrix of the instruments. Under double asymptotics, we show that our regularized estimators are consistent and asymptotically normal. These regularization schemes involve a tuning or regularization parameter which needs to be chosen. We derive a data-driven selection of this regularization parameter based on an approximation of the higher-order mean square error and show its optimality. As an empirical application, we estimate a model of income dynamics.

## 1. INTRODUCTION

In this paper, we propose a regularization approach to the estimation of a dynamic panel data model (DPM) with individual fixed effect. The presence of this last element creates a correlation between the error term of the model and one of the explanatory variable which is the lagged value of the dependent variable. Hence, the generalized method of moments (GMM) is widely used to estimate such models using lags of the dependent variable as instruments. A feature of the DPM is that, if a variable at a certain time period can be used as an instrument, then all the past realizations of that variable can also be used as instruments. Therefore, the number of moment conditions can be very large even if the time dimension is moderately large.

Increasing the number of moments has two opposite effects on the GMM estimator. On the one hand, it improves its efficiency; on the other hand, it increases

The authors thank the Editor (P.C.B. Phillips), the Co-Editor (Guido Kuersteiner), three anonymous referees, Ryo Okui, and the participants of the NBER-NSF conference (2016), NY Camp Econometrics (2017), Canadian Economic Association (2018), International Association for Applied Econometrics (2018), and the Econometric Study Group (2019) for their helpful comments. Carrasco thanks SSHRC for partial financial support. Address correspondence to Marine Carrasco, CIREQ, University of Montreal, Montreal, QC, Canada; e-mail: marine.carrasco@umontreal.ca

 

its bias. 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 this problem, we propose to use regularized inverses of the covariance matrix as in Carrasco (2012) and Carrasco and Tchuente (2015). The impact of regularization is twofold. First, it yields to more reliable estimation of the inverse covariance matrix when it is ill-conditioned. Second, it performs a dimension reduction on the space of instruments by putting more weight on the principal components (PCs) associated with the largest eigenvalues of the covariance matrix. The three regularizations considered are spectral cutoff (SC), Tikhonov (TK), and Landweber–Fridman (LF). The SC regularization scheme is based on the first PCs, whereas the TK's one can be considered as the dynamic panel version of the ridge regression (also called Bayesian shrinkage), and the last one is an iterative method. All these methods involve a regularization parameter similar to the smoothing parameter in nonparametric regression. This parameter needs to converge to zero at an appropriate rate to obtain an asymptotically efficient estimator.

In this paper, we focus on regularized versions of the one-step GMM estimator where the individual effect has been removed by forward filtering as in Alvarez and Arellano (2003; hereafter AA). The one-step GMM estimator corresponds to the efficient GMM estimator when the error is conditionally homoskedastic. We derive the first-order asymptotic properties of the regularized estimators under double asymptotics where $N$ and $T$ go to infinity and assuming homoskedastic error. Then, we derive the leading term of the mean square error (MSE) using a second-order expansion. We show that the leading terms consist of a squared bias and a variance term which are both functions of the regularization parameter. The main contribution of the paper is to develop a data-driven selection of the regularization parameter as minimum of the approximate MSE and to prove its optimality in the sense of Li (1986, 1987).

The literature related to the many instruments problem is very large. Working on cross-sectional data, Donald and Newey (2001) propose to select the number of instruments that minimizes the MSE of their estimators. Kuersteiner and Okui (2010) propose a model averaging two-stage least-squares (2SLS) estimator where the first-stage estimator is the average of estimators obtained by projecting on subsets of instruments. The weights are chosen to minimize the asymptotic MSE of the model averaging 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. Belloni et al. (2012) apply Lasso on the first-stage equation to select a subset of instruments. Doran and Schmidt (2006) use PCs in a DPM to reduce the bias. Our work complements their paper by proposing a data-driven method to choose the optimal number of PCs to use in order to improve the finite sample properties of the estimator.

Carrasco (2012) proposes regularization approaches to two-stage least-squares estimation, whereas Carrasco and Tchuente (2015) focus on a regularized version of the limited information maximum likelihood (LIML) estimator. While the same regularizations are considered here, the proof technique is very different because of the dynamic nature of the model. Indeed, because the instruments are not exogenous but only predetermined, the analysis of the MSE is much more challenging than in the cross-sectional framework of Carrasco (2012). In particular, one cannot rely on an expansion of the MSE conditional on the instruments, but on the unconditional MSE instead.

Several bias-corrected estimators have been proposed for DPM. Kiviet (1995) proposes a bias-corrected version of within-group estimator. Hahn and Kuersteiner (2002) propose a bias correction of OLS estimator when both $N$ and $T$ are large. Hahn, Hausman, and Kuersteiner (2004) develop a bias correction for 2SLS based on the Jackknife. Bun and Kiviet (2006) study the bias of the method of moment estimator of DPM with weakly exogenous regressors. Hayakawa (2009) proposes to use instrumental variables (IV) with an approximate optimal instrument, whereas Hayakawa, Qi, and Breitung (2019) construct instruments by backward filtering and show the equivalence with a bias-corrected estimator. Our methodology complements those methods as regularization provides a partial bias correction. In an identical framework as ours, Okui (2009) derives a higher-order expansion of the MSE and proposes to choose the optimal number of moment conditions to minimize an estimated version of this expansion. However, the finite sample bias problem is not completely addressed since his simulations present a large bias for the GMM estimator when the autoregressive parameter is close to unity.

The remainder of this paper is organized as follows. Section 2 presents the DPM and the one-step GMM estimator. Section 3 presents regularized estimators, whereas Sections 4 and 5, respectively, present asymptotic properties and higher-order properties of regularized GMM estimators. A data-driven selection of the regularization parameter is developed in Section 6. Section 7 presents the extension of the model to exogenous covariates, and Section 8 discusses the results of Monte Carlo simulations. An empirical application on income dynamics is discussed in Section 9. It appears that the regularization corrects the bias of the usual GMM estimator, which seems to underestimate the estimated autoregressive coefficient. Throughout the paper, we use the notations $I$ and $I_{\bar{q}}$, respectively, for the $N \times N$ and $\bar{q} \times \bar{q}$ identity matrix. The proofs are collected in Appendix B.

## 2. THE MODEL

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

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

where $\delta$ is the parameter of interest satisfying $|\delta| < 1$, $\eta_i$ is the unobserved individual effect, and $v_{it}$ is the idiosyncratic error with conditional mean zero and variance $\sigma^2$ conditionally on $\eta_i, y_{it-1}, \dots, y_{i0}$. For simplicity, we assume that $y_{i0}$ is observed. Moreover, we denote $y_{i,t-1}$ by $x_{i,t}$.

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

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 difference (FD) and the forward orthogonal deviation (FOD) operator. In this paper, we use the latter for theoretical and computational purposes. Indeed, this transformation preserves homoskedasticity and no serial correlation properties of the error term. Let the $(T-1)\times T$ matrix $A$ denote the FOD operator as used by Arellano and Bover (1995) and define $v_i^* = Av_i$, $x_i^* = Ax_i$, and $y_i^* = Ay_i$, where $v_i = (v_{i1},\dots,v_{iT})'$, $x_i = (x_{i1},\dots,x_{iT})'$, and $y_i = (y_{i1},\dots,y_{iT})'$. Similarly, we denote $x_i^* = (x_{i1}^*,\dots,x_{iT}^*)'$ and $y_i^* = (y_{i1}^*,\dots,y_{iT}^*)'$. In particular, the $t$th element of $y_i^*$ is given by

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

with $c_t^2 = (T-t)/(T-t+1)$.

By multiplying the model by $A$, equation (1) becomes

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

We have $E(x_{i,t}^* v_{it}^*) \neq 0$ so that the ordinary least square (OLS) estimator of the transformed model is not consistent for fixed $T$ as $N$ tends to infinity. However, $E(x_{i,t-s}v_{it}^*) = 0$, for $s = 0,\dots,t-1$ and $t = 1,\dots,T-1$. Then, we consider the one-step GMM estimator of $\delta$ based on these moment conditions. The properties of this estimator have been studied by AA. The number of moment conditions is $\bar{q} = T(T-1)/2$, which can be very large even if $T$ is moderately large. Let $z_{it} = (x_{i1},\dots,x_{it})'$, and let $Z_i$ be the $(T-1)\times\bar{q}$ block diagonal matrix whose $t$th block is $z_{it}'$. The moment conditions are then given by $E(Z_i'v_i^*) = 0$ with $v_i^* = (v_{i1}^*,\dots,v_{i,T-1}^*)'$. Under the assumption of conditional homoskedasticity of $v_{it}$, the covariance matrix of the moment conditions is $\sigma^2 E(Z_i'Z_i)$. The one-step GMM estimator of the parameter $\delta$ is given by

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

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

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

where $M = Z(Z'Z)^{-1}Z'$ is an $N(T-1)\times N(T-1)$ matrix and $Z = (Z_1',\dots,Z_N')'$ is an $N(T-1)\times\bar{q}$ matrix. Note that $Z'Z$ is a block-diagonal matrix with blocks $Z_t'Z_t$, $t = 1,2,\dots,T-1$, on the diagonal.

Even though it is widely used by empirical researchers, this GMM estimator suffers from small-sample bias when the number of moments is large. Moreover, using a simple $AR(1)$ model, Blundell and Bond (1998) showed that the lagged levels of the dependent variable become weak instruments when the autoregressive parameter gets close to unity or when the variance of the unobserved individual effect increases toward the variance of the idiosyncratic error $v_{it}$. Doran and Schmidt (2006) argue that in the presence of many instruments, the marginal contribution of some of them may be small. As a result, many simulations including those in Okui (2009) showed that the one-step GMM estimator of dynamic panel data performs poorly in these settings. A solution could be to remove the instruments corresponding to farther lagged variables as in Okui (2009) or to reduce the weight given to these instruments as in Kuersteiner and Okui (2010) and Kuersteiner (2012). The solution we propose is to apply a regularization technique which will reduce the weight given to the eigenvectors corresponding to the smallest eigenvalues of $Z'Z$. Eigenvectors associated with the largest eigenvalues explain most of the variation of the instruments, so removing the smallest eigenvalues is a dimension reduction device which permits to reduce the bias of the GMM estimator without inducing an important efficiency loss. Another advantage is that regularizations use all the instruments to compute the eigenvectors, and no selection is needed. This is useful when the model includes exogenous covariates for which there is no natural ranking of the instruments.

## 3. THE REGULARIZED ESTIMATOR

Note that $M = Z(Z'Z)^{-1}Z' = Z(Z'Z/(NT^{3/2}))^{-1}Z'/(NT^{3/2}) = ZK_N^{-1}Z'/\left(NT^{3/2}\right)$. In the next section, we will see that the rescaled matrix $K_N = Z'Z/\left(NT^{3/2}\right)$ has square summable eigenvalues where the smallest eigenvalue goes to zero as $N$ and $T$ go to infinity. We are going to apply some regularization to $K_N = Z'Z/\left(NT^{3/2}\right)$. Regularization techniques were initially introduced in the inverse problem literature to obtain reliable inverses of matrices which are ill-conditioned (for which the ratio of the largest eigenvalue over the smallest is large; see Kress, 1999). The way regularization works is by dampening the effect of the smallest eigenvalues. As a result, one puts more weight on the eigenvectors associated with the largest eigenvalues.

Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_q \geq 0$ be the eigenvalues of $K_N$ in decreasing order. By spectral decomposition, we have $K_N = P_N D_N P_N'$ with $P_N P_N' = I_{\bar{q}}$ where $P_N$ is the matrix of eigenvectors and $D_N$ is the diagonal matrix with eigenvalues $\hat{\lambda}_j$ on the diagonal. The naive inverse $K_N^{-1} = P_N D_N^{-1} P_N'$ involves $1/\hat{\lambda}_j$, which may explode for values of $\hat{\lambda}_j$ tending to zero. Let $K_N^{\alpha}$ denote the regularized inverse of $K_N$, which is defined as

$$K_N^{\alpha} = P_N D_N^{\alpha} P_N',$$

where $D_N^{\alpha}$ is the diagonal matrix with elements $q(\alpha, \hat{\lambda}_j^2)/\hat{\lambda}_j$ with the convention $0/0 = 0$ and $\alpha$ is some nonnegative regularization (tuning) parameter. In each regularization scheme, the real-valued function $q(\alpha, \lambda^2)$ satisfies $0 \leq q(\alpha, \lambda^2) \leq 1$ and $\lim_{\alpha \to 0} q(\alpha, \lambda^2) = 1$ so that $q(\alpha, \hat{\lambda}_j^2)/\hat{\lambda}_j \leq 1/\hat{\lambda}_j$, reducing the impact of individual eigenvalues. The usual GMM estimator corresponds to a regularized estimator with $\alpha = 0$.

As in Carrasco (2012), three regularization schemes will be considered: TK, SC, and LF regularization schemes. More details on these schemes can be found in Carrasco, Florens, and Renault (2007b). If we let $\lambda$ be an arbitrary eigenvalue of the matrix $K_N$, we can define:

1. **Tikhonov Regularization:**
   This regularization scheme is close to the well-known ridge regression used in the presence of multicollinearity to improve the properties of the OLS estimator. In the TK regularization scheme, the real function $q(\alpha,\lambda^2)$ is given by
   $$q(\alpha,\lambda^2) = \frac{\lambda^2}{\lambda^2+\alpha}.$$
2. **The Spectral Cutoff**
   It consists in selecting the eigenvectors associated with the eigenvalues greater than some threshold:
   $$q(\alpha,\lambda^2) = I\{\lambda^2 \geq \alpha\} = \begin{cases} 1, \text{ if } \lambda^2 \geq \alpha, \\ 0, \text{ otherwise.} \end{cases}$$
   Another version of this regularization scheme is PC, which consists in using a certain number of eigenvectors to compute the inverse of the operator. PC and SC are perfectly equivalent, but only the definition of the regularization term $\alpha$ differs. In PC, $\alpha$ 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**
   Let $c$ be a constant satisfying $0 < c < 1/\hat{\lambda}_1^2$ where $\hat{\lambda}_1^2$ is the largest eigenvalue of the matrix $K_N$. The LF regularization is based on
   $$q(\alpha,\lambda^2) = 1-(1-c\lambda^2)^{\frac{1}{\alpha}},$$
   where $\frac{1}{\alpha}$ is an integer. Alternatively, one can compute the LF regularized inverse by iterations where the number of iterations $\frac{1}{\alpha}$ is the regularization parameter (see Carrasco et al., 2007b).

Let $M^\alpha = ZK_N^\alpha Z'/NT^{3/2}$. The regularized one-step GMM estimator for a given regularization scheme is

$$\widehat{\delta}^\alpha = \frac{x^{*'}M^\alpha y^*}{x^{*'}M^\alpha x^*}. \tag{2}$$

The matrix $K_N$ is a block diagonal matrix with the $t\times t$ matrix $Z_t'Z_t/NT^{3/2}$ at the $t$th block. Exactly as $K_N^{-1}$, the regularized inverse $K_N^\alpha$ is also a block diagonal matrix.[1]

[1] So the regularizations preserve the structure of the matrix as a block-diagonal matrix; hence, in some sense, it preserves the sparsity of the initial matrix. This holds because the regularizations transform only the eigenvalues, not the eigenvectors.

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

If we define $M_t^\alpha = Z_t (K_{Nt})^\alpha Z_t'/(NT^{3/2})$ with $(K_{Nt})^\alpha$ being the $t$th diagonal block of the matrix $K_N^\alpha$, the regularized estimator can be rewritten as

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

The different regularization schemes depend on the regularization parameter $\alpha$, which is a kind of smoothing parameter and needs to be selected in practice. The choice of $\alpha$ is based on the minimization of the approximate MSE of $\widehat{\delta^\alpha}$. Estimation of the MSE requires some preliminary estimates of $\delta$ and $\sigma^2$, independent of $\alpha$. Let $\hat{\delta}$ be a preliminary estimate of $\delta$, obtained, for instance, by GMM using $y_{t-1}$ as instrument, and let $\hat{\sigma}^2 = \sum_{i,t}\left(y_{it}^* - \hat{\delta} x_{it}^*\right)^2/(NT)$. Then, the regularization parameter is obtained by

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

where

$$\widehat{S}(\alpha) = (1+\hat{\delta})^2 \widehat{\mathcal{A}}^2(\alpha) + \frac{(1-\hat{\delta}^2)^2}{\hat{\sigma}^2}\widehat{R}(\alpha), \tag{3}$$

where the first term on the right-hand side comes from the squared bias of $\widehat{\delta}^\alpha$, whereas the second term comes from its variance with

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

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

and $\widehat{\phi}_j = \frac{1-\widehat{\delta}^j}{1-\widehat{\delta}}$. The index set $\mathcal{E}_T$ corresponds to $\mathbb{R}^+$ for TK, $\mathcal{E}_T$ is such that $\frac{1}{\alpha} \in \{1,2,\ldots,\bar{q}\}$ for PCs, and $\mathcal{E}_T$ is such that $\frac{1}{\alpha}$ is a positive integer and the cardinal of $\mathcal{E}_T$ is no larger than some finite multiple of $T^2$ for LF.

## 4. FIRST-ORDER ASYMPTOTIC PROPERTIES

In this section, we derive the asymptotic properties of the regularized estimator. As in Okui (2009), we make the following assumptions.

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

**Assumption 2.** The initial observation satisfies

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

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

**Assumption 3.** $\eta_i$ are *i.i.d.* 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, asymptotic properties are derived under the assumption that both $N$ and $T$ go to infinity with $T < N$. Under this restriction, the matrix $K_N$ is nonsingular and so has nonzero eigenvalues.

The regularization methods used in this paper are drawn from the literature on inverse problems (Kress, 1999). They are designed to stabilize the inverse of Hilbert–Schmidt operators. A symmetric matrix is Hilbert–Schmidt if it is bounded in Frobenius norm, or equivalently if its eigenvalues are square-summable. The matrix $E\left(Z'Z\right)$ is not Hilbert–Schmidt; however, we will show in Lemma 1 that $K = E[Z'Z/\left(NT^{3/2}\right)]$ is Hilbert–Schmidt.[2] Let $K$ denote the $\bar{q} \times \bar{q}$ matrix $E[Z'Z/\left(NT^{3/2}\right)]$. In the inverse problem literature, this matrix is referred to as the operator. In Carrasco and Tchuente (2015), the operator is assumed to be a trace-class operator which is satisfied if and only if its trace is finite. Here, however, $K$ is not trace class, but it is Hilbert–Schmidt, which is a slightly weaker condition. In the following lemma, we prove that $K$ is a Hilbert–Schmidt matrix.

LEMMA 1. *If Assumptions 1–3 are satisfied, then:*

(i) *$tr(K)$ diverges at rate $T^{1/2}$ when $T$ goes to infinity,*
(ii) *$tr\left(K^2\right) = O(1)$, for all $T$.*

Lemma 1 shows that even though the eigenvalues of $K$ are not summable as $T$ goes to infinity, they are square-summable. The Hilbert–Schmidt property is sufficient to derive proofs in our framework. This property is especially useful to establish the order of magnitude of the bias of the regularized estimator. Lemma 1 also implies some interesting properties on the eigenvalues of $K$. It follows from (ii) that the largest eigenvalue of $K^2$ (and hence of $K$) is bounded and the smallest one goes to zero but not too fast. It implies in particular that $K$ is ill-conditioned (the ratio of its largest eigenvalue over its smallest eigenvalue diverges as $T$ goes to infinity). Given that this property is invariant to the scaling, it means that $E\left(Z'Z\right)$ is also ill-conditioned. Inverses of ill-conditioned matrices are not reliable, and regularization is helpful to stabilize the inverse. Finally, the proof of Lemma 1 reveals that $K$ depends on $T$ but not on $N$.

The condition number of a matrix is defined as the ratio of the largest eigenvalue to the smallest one and is independent of the scaling. Table A1 in Appendix A presents the distribution of the condition number of the matrix $Z'Z$ for simulated data for various values of $\delta$ and $T$. The higher the condition number is, the more ill-conditioned the matrix is, and so inverting its eigenvalues is more problematic;

[2]The Hilbert–Schmidt property is used in the proofs to derive the asymptotic properties of the regularized estimators.

therefore, the need of regularization is higher. We present the min, the first quartile, the mean, the median, the third quartile, and the max. The last column gives the dimension of the matrix $Z'Z$, which is the total number of instruments $\bar{q} = 0.5 \times T \times (T-1)$ for each value of $T$. From Table A1, we observe that the need of regularization increases with $T$ for a given $\delta$ and also increases with $\delta$ for a given $T$. As $\delta$ gets closer to 1, the instruments become weak, yielding an ill-conditioned matrix.

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

PROPOSITION 1. *If Assumptions 1–3 are satisfied, $\alpha$ the parameter of regularization goes to 0, $\alpha\sqrt{NT}$ goes to infinity, and both $N$ and $T$ tend to infinity with $T < N$, then:*

(i) *Consistency: $\widehat{\delta}^{\alpha} \to \delta$ in probability;*
(ii) *Asymptotic normality: $\sqrt{NT}\left(\widehat{\delta}^{\alpha} - \delta\right) \xrightarrow{d} N\left(0, 1-\delta^2\right)$.*

For these properties, we need that $\alpha$ goes to zero slower than $\sqrt{NT}$ goes to infinity. Under similar assumptions, AA proved that the bias expression of the one-step GMM estimator of DPM is given by the limit of

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

where $\phi_t = \left(1-\delta^t\right)/(1-\delta)$. Hence, if $T/N$ tends to a positive scalar, the one-step GMM estimator is asymptotically biased. In our regularization setting, the bias is given by the limit of

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

This bias is of order $\frac{1}{(\alpha\sqrt{NT})}$ so that the asymptotic bias of the regularized estimator vanishes under the assumption that $\alpha\sqrt{NT}$ goes to $\infty$ even in the case where $T/N \to c$. Note that we do not have an explicit expression for $E[tr\left(M_t^{\alpha}\right)]$, but we know that $E[tr\left(M_t^{\alpha}\right)] < tr[M_t] = t$, for all $\alpha > 0$. From the rates of convergence of the two biases, we see that $b_{NT}^{\alpha}$ is negligible with respect to $b_{NT}$ as soon as $\alpha T \to \infty$. Regularization has the advantage of introducing a tuning parameter $\alpha$ which can be chosen in order to control the bias by dampening the weight attributed to the eigenvectors of the matrix $Z'Z$ without discarding any instrument a priori. Of course, $\alpha$ also has an impact on the variance as will become apparent in Section 5. The expression of the bias could be used to devise a bias-corrected estimator, but this is left for future research. Since the asymptotic properties of the regularized estimator presented in this section do not depend on the regularization scheme, we need to investigate higher-order properties to establish the impact of regularization.

## 5. HIGHER-ORDER ASYMPTOTIC PROPERTIES

In this section, we derive the Nagar's (1959) decomposition of the MSE of our estimators $E[(\hat{\delta}^{\alpha}-\delta)^2]$. This type of expansion is used in many papers on instrumental variables, such as Donald and Newey (2001), Carrasco (2012), and particularly Okui (2009), who works on a DPM. Moreover, this expansion will guide us in our goal to provide a data-driven method for selecting the regularization parameter. Let $H=\frac{\sigma^2}{1-\delta^2}\left(\frac{1}{T}\sum_{t=1}^{T}\psi_t^2\right)$, where $\psi_t=c_t(1-\delta\phi_{T-t}/(T-t)), \phi_j=(1-\delta^j)/(1-\delta)$, $\psi_t=c_t(1-\delta\phi_{T-t}/(T-t))$, and $c_t=\sqrt{(T-t)/(T-t+1)}$.

The Nagar approximation of the MSE is $\sigma^2H^{-1}+S(\alpha)$ in the following decomposition:

$$NT(\hat{\delta}^{\alpha}-\delta)^2=Q+r,\ \ E(Q)=\sigma^2H^{-1}+S(\alpha)+R, \tag{4}$$

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

PROPOSITION 2. *Suppose Assumptions 1–3 are satisfied and* $E\left(v_{it}^3\right)=0$. *If* $N\to\infty$, $T\to\infty$, $\alpha\to 0$, $\alpha\sqrt{NT}\to\infty$, *and* $\alpha(\ln T)\sqrt{T}\to 0$, *then for the regularized GMM estimator, the decomposition given in (4) holds with*

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

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

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

to $\sum_j q\left(\alpha,\widehat{\lambda}_j^2\right)$, where $q$ is the dampening function defined in Section 3. There is no simple expression for the term $E[w_{t-1}'(I-M_t^\alpha)^2 w_{t-1}]$. However, we know that $w_{t-1}'(I-M_t^\alpha)^2 w_{t-1} \geq w_{t-1}'(I-M_t)^2 w_{t-1} = w_{t-1}'(I-M_t)w_{t-1}$, for all $\alpha \geq 0$, with equality when $\alpha = 0$. Therefore, the higher-order variance of the regularized estimator is larger than that of the usual GMM estimator. Hence, regularization reduces the bias of the one-step GMM estimator at the cost of a larger small-sample variance.

In our panel data setting, the bias expression of $S(\alpha)$ is the sum of the bias of each period $H_0^{-1}E\left[tr\left(M_t^\alpha\right)\right]E[\tilde{v}_{tT}'v_t^*]$, where $H_0$ is the asymptotic variance. As the formula (3.14) in the special case of Kuersteiner (2012), this period bias expression is the product of the inverse of $H_0$, $E\left[tr\left(M_t^\alpha\right)\right]$ 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's (2012) is that the contribution of $E\left[tr\left(M_t^\alpha\right)\right]$ depends on $\alpha$ and is not the number of instruments.

## 6. DATA-DRIVEN SELECTION OF THE REGULARIZATION PARAMETER

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

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

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

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

Regarding the estimation of $R(\alpha)$, it follows from Okui (2009, p. 3) that

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

where $E_{t-1}$ denotes the conditional expectation conditional on $(\eta_i, x_{it}, x_{it-1}, \ldots)$ so that $x_t^*$ is an unbiased estimator of $\psi_t w_{it-1}$. We propose to estimate $R(\alpha)$ by $\widehat{R}(\alpha) = \frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*\prime}(I-M_t^\alpha)^2 x_t^*$ and we show that $\left(\widehat{R}(\alpha) - R(\alpha)\right)/R(\alpha)$ converges uniformly to 0 in the proof of Proposition 3. Regarding the estimation of $\mathcal{A}(\alpha)$, the term $E[tr\left(M_t^\alpha\right)]$ is not known in closed form; however,

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} tr\left(M_t^{\alpha}\right)\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)$ is an unbiased estimate of $\mathcal{A}(\alpha)$. This suggests that replacing $E[tr\left(M_t^{\alpha}\right)]$ by $tr\left(M_t^{\alpha}\right)$ in the expression of $\widehat{\mathcal{A}}(\alpha)$ might be a good idea, and this is confirmed in the proof of Proposition 3 where we show the uniform convergence of $\left(\widehat{\mathcal{A}}(\alpha)-\mathcal{A}(\alpha)\right)/\mathcal{A}(\alpha)$ to zero.

The optimal regularization parameter is selected by minimizing the estimate $\widehat{S}(\alpha)$ of $S(\alpha)$

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

where $\widehat{S}(\alpha)$ was defined in (3). Next, we analyze the impact of using an estimated version of $S(\alpha)$ to select $\alpha$ instead of the true and unknown criterion.

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

$$\frac{S(\hat{\alpha})}{\inf_{\alpha\in\mathcal{E}_T} S(\alpha)} \xrightarrow{p} 1, \tag{5}$$

as $N\to\infty$ and $T\to\infty$. It should be noticed that the result (5) is not a convergence result of $\hat{\alpha}$. It simply establishes that using an estimated version of $S(\alpha)$ to choose the regularization parameter is asymptotically equivalent to using the true and unknown value of $S(\alpha)$.

PROPOSITION 3. *Suppose that Assumptions 1–3 are satisfied, $\hat{\delta}\to\delta$, and $\hat{\sigma}^2\to\sigma^2$. If $N\to\infty$, $T\to\infty$, $\alpha\sqrt{NT}\to\infty$, $\alpha(\ln T)\sqrt{T}\to 0$, and $T^3/N(\ln T)^2\to 0$, then the regularization parameter selection criterion is asymptotically optimal in the sense of (5) for SC and LF regularization schemes provided that $\#\mathcal{E}_T=O(T^2)$ where $\#\mathcal{E}_T$ refers to the number of elements in the set $\mathcal{E}_T$.*

**Remarks.**

1. Proposition 3 proves the optimality for PC and LF regularization schemes which have discrete index set $\mathcal{E}_T$. The condition $\#\mathcal{E}_T = O(T^2)$ is a sufficient condition in the LF regularization scheme and necessarily holds for the PC 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 $\mathcal{E}_T$. A rigorous proof for the TK's continuous index set requires a more complicated material, which is beyond the scope of this work. However, optimality could be established in this case using a discretization of the compact set $\mathcal{E}_T$ and the fact that the regularization function $q(\alpha,\lambda^2)$ of TK regularization scheme is a real continuous function as in Hansen (2007).

2. Proposition 3 is related to Donald and Newey (2001) optimality result for the selection of the number of instruments in a linear IV model and Carrasco and Tchuente (2015) for the selection of the regularization parameter for the

[3]Recall that $\#\mathcal{E}_T=\bar{q}$ for PC case.

regularized LIML estimator. However, their proofs rely on applying results by Li (1986, 1987) on cross-validation for the first-stage equation. From $x^*_{it} = \psi_t w_{it} - c_t\tilde{v}_{it}$, the term $-c_t\tilde{v}_{it}$ can be regarded as the error term of the first-stage equation since $\psi_t w_{it}$ is considered as the optimal instrument in Okui (2009). However, Li's (1986, 1987) results do not apply in our framework because of the autocorrelation of this error term. Instead, our proof combines the strategies of Okui (2009) and Kuersteiner (2012).

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

## 7. INTRODUCTION OF EXOGENOUS REGRESSORS

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

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

where $m_{it}$ is an $L_m$-dimensional vector of strictly exogenous variables in the sense that $E(m_{it}v_{is}) = 0$ for each $t$ and $s$.

Let us define $\theta = (\delta, \gamma')'$ and $x_{it} = (y_{i,t-1}, m'_{it})'$, and denote $y_i = (y_{i1}, \ldots, y_{iT})'$, $x_i = (x_{i1}, \ldots, x_{iT})'$, and $v_i = (v_{i1}, \ldots, v_{iT})'$. Let $A$ be the matrix of FOD operator and denote $y^*_i = Ay_i$, $x^*_i = Ax_i$, and $v^*_i = Av_i$. The model becomes

$$y^*_{it} = \theta x^*_{it} + v^*_{it}.$$

Following Okui (2009), we assume that time-invariant variables $f_i$ that satisfy $E(f_i v_{it}) = 0$ for all $t$ are available and we denote by $L_f$ the dimension of this vector. Even though they are often omitted in proofs, time-invariant variables are widely used in empirical work. The vector of potential instruments for the endogenous regressor $x^*_{it}$ is the $q_t = (L_f + (T+1)L_m + t)$-dimensional vector $z_{it} = (f'_i, m'_{i0}, \ldots, m'_{iT}, y_{i0}, \ldots, y_{i,t-1})'$. In this setting, the total number of instruments is $\bar{q} = \sum_t q_t$. Let us define the following matrixes: $Z_t = (z'_{1t}, \ldots, z'_{Nt})'$, $x^*_t = (x^*_{1t}, \ldots, x^*_{Nt})'$, and $y^*_t = (y^*_{1t}, \ldots, y^*_{Nt})'$. If we denote by $K_N = Z'Z/NT^{3/2}$ and $K^\alpha_N$ the regularized inverse of $K_N$ given a regularization parameter $\alpha$, then the regularized one-step GMM estimator of $\theta$ is

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

with $M^\alpha = ZK^\alpha_N Z'/NT^{3/2}$, $Z = (Z'_1, \ldots, Z'_N)'$, and $Z_i$ has the same definition as in the model without covariates.

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

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

**Assumption 2'.** (i) $(y_{it}, m_{it})$ is a strictly stationary finite-order vector autoregressive process conditional on $\eta_i$ such that the distribution of $\{(y_{it}, m_{it}')', \ldots, (y_{i,t+s}, m_{i,t+s}')'\}$ conditional on $\eta_i$ 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). Let $K = E[Z'Z/NT^{3/2}]$. The matrix $K$ is assumed to be a Hilbert–Schmidt matrix. Moreover, in the extended model, we make an assumption on the growth rate of the eigenvalues of $K$. If we define $\tilde{W} = (\tilde{w}_1', \ldots, \tilde{w}_{T-1}')'$ with $\tilde{w}_t = (\tilde{w}_{1t}, \ldots, \tilde{w}_{N,t})'$, then we impose the following condition.

**Assumption 3.** The matrix $K$ is Hilbert–Schmidt, and there is a $\beta > 0$ such that

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

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

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

Let us define $x_t^* = (u_t^*, m_t^*)$ with $u_t = y_{t-1}$. Okui (2009) states that $w_{i,t-1} = E_Z(u_{it}^*)$ is such that

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

where $c_t = \sqrt{\frac{T-t}{T-t+1}}$, $\mu_i = \frac{\eta_i}{1-\delta}$, and $\phi_j = \frac{1-\delta^j}{1-\delta}$. Hence, the optimal instrument is given by $E_Z(x_{it}^*) = \tilde{w}_{i,t-1} = (w_{i,t-1}, m_{it}^{*\prime})'$.

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

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

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

*where*

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

*and*

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

As in the model without covariates, the first part of $S(\alpha)$ is the squared bias and increases as the regularization parameter goes to zero. The second term of $S(\alpha)$ is the second-order variance of the regularized estimator and it decreases when $\alpha$ goes to zero.

We observe that the bias term depends only on the bias of the autoregressive coefficient $\hat{\delta}$ and not on $\hat{\gamma}$, whereas the variance term depends on both so that globally the MSE depends also on $\hat{\gamma}$. This is different from Okui (2009). In his Theorem 4, Okui imposes the extra assumption that the subset of instruments $z_{it}^K$ includes either $y_{it-1}$ and $m_{it}^*$ or linear combinations of these. He shows that, in this case, only the element (1,1) of the matrix $S(\alpha)$ is nonzero so that he can focus on this scalar to select the regularization parameter. Interestingly, it means that only the MSE of $\widehat{\delta}$ matters for selecting $\alpha$. In contrast, we do not impose this extra assumption and we show that the MSE of $\hat{\gamma}$ also plays a role.

Given $S(\alpha)$ is a matrix, $\alpha$ can be selected by minimizing $\ell'\widehat{S}(\alpha)\ell$ for an arbitrary $L_m+1$ vector $\ell$ and some estimator $\widehat{S}(\alpha)$ of $S(\alpha)$. For the estimation of $S(\alpha)$, similarly to the model without covariates, $\mathcal{A}(\alpha)$ can be estimated by

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

where the unknown parameters, $\sigma^2$ and $\phi_j$, are estimated using a preliminary estimate of $\theta$. Regarding the estimation of the variance term, $\frac{1}{NT}\sum_t E[\widetilde{w}_{t-1}'(I-M_t^\alpha)^2\widetilde{w}_{t-1}]$ can be estimated by $\frac{1}{NT}\sum_t x_t^{*'}(I-M_t^\alpha)^2 x_t^*$ because $x_t^*$ is an unbiased estimate of $\tilde{w}_{t-1}$.

$S(\alpha)$ depends also on the matrix $H$. In the simulations, we choose $\ell$ so that $\ell' H^{-1}$ is a vector of ones. This avoids estimating $H$. On the other hand, estimating $H$ is possible. Remark that $\frac{1}{T}\sum_{t=1}^T E\left(m_{it}^* m_{it}^{*'}\right)$ can be estimated by $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N m_{it}^* m_{it}^{*'}$ and the term $\frac{1}{T}\sum_{t=1}^T E\left(w_{it} m_{it}^{*'}\right)$ can be estimated by $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N u_{it}^* m_{it}^{*'}$ because $E\left(u_{it}^* m_{it}^{*'}\right) = E\left[E_Z\left(u_{it}^* m_{it}^{*'}\right)\right] = E\left[E_Z\left(u_{it}^*\right) m_{it}^{*'}\right] =$

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

$E\left(w_{it}m_{it}^{*'}\right)$. The estimation of the term involving $E\left(w_{it}^2\right)$ is not so straightforward and requires some calculations. Observe that $w_{it} = \psi_{t+1}(y_{i,t} - \mu_i) - \frac{c_{t+1}}{T-t-1}\gamma'(\phi_{T-t-1}m_{i,t+1} + \cdots + \phi_1 m_{i,T-1}) = \psi_{t+1}\widetilde{y}_{i,t} - \frac{c_{t+1}}{T-t-1}\gamma'(\phi_{T-t-1}m_{i,t+1} + \cdots + \phi_1 m_{i,T-1})$, where $\widetilde{y}_{i,t} = y_{i,t} - \mu_i$ is a stationary process which follows the dynamics.

$$\tilde{y}_{it} = \delta\widetilde{y}_{i,t-1} + \gamma' m_{it} + v_{it}.$$

To simplify assume that $m_{it}$ is not autocorrelated conditionally on $\eta_i$. Then, we have

$$E\left(w_{it}^2\right) = \psi_{t+1}^2 E\left(\widetilde{y}_{i,t}^2\right) + \left(\frac{c_{t+1}}{T-t-1}\right)^2 E\left[\left(\gamma'(\phi_{T-t-1}m_{i,t+1} + \cdots + \phi_1 m_{i,T-1})\right)^2\right]$$
$$- 2\psi_{t+1}\frac{c_{t+1}}{T-t-1}E\left(\widetilde{y}_{i,t}\right)E\left(\gamma'(\phi_{T-t-1}m_{i,t+1} + \cdots + \phi_1 m_{i,T-1})\right).$$

Using the stationarity of $\widetilde{y}_{i,t}$ and $m_{it}$, we have $V\left(\widetilde{y}_{i,t}\right) = \gamma' V(m_{it})\gamma / \left(1-\delta^2\right) + \sigma_v^2 / \left(1-\delta^2\right)$ and $E\left(\widetilde{y}_{i,t}\right) = \gamma' E(m_{it}) / (1-\delta), E\left(\gamma'(\phi_{T-t-1}m_{i,t+1} + \cdots + \phi_1 m_{i,T-1})\right) = (\phi_{T-t-1} + \cdots + \phi_1)\gamma' E(m_{it})$. The terms $E(m_{it})$, $V(m_{it})$, and $E\left[\left(\gamma'(\phi_{T-t-1}m_{i,t+1} + \cdots + \phi_1 m_{i,T-1})\right)^2\right]$ can be easily computed by replacing the expectation by an average over $i$. So an estimate of $E\left(w_{it}^2\right)$ follows using some preliminary estimates of $\theta$ and $\sigma_v^2$. The case where the $m_{it}$ are autocorrelated can be handled in a similar manner by replacing the autocorrelations of $m_{it}$ by their sample counterparts.

## 8. SIMULATION STUDY

In this section, we present Monte Carlo simulations to illustrate the finite sample properties of the regularized estimators and compare them to competing estimators.

### 8.1. Model with a Strictly Exogenous Regressor and Homoskedastic Error

We consider a model including one strictly exogenous covariate. The equation is given by

$$y_{it} = \delta y_{it-1} + \gamma m_{it} + \eta_i + v_{it}, |\delta| < 1, \tag{6}$$
$$m_{it} = \rho\eta_i + e_{it},$$

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

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

In this setting, for each period $t$, $m_{i0},\ldots,m_{iT}$ are potential instruments in addition to the lags of $y_{it}$. Hence, compared to a model without covariates, the total number

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

of instruments increases from $\bar{q} = 0.5 \times T \times (T-1)$ to $\bar{q} = 0.5 \times T \times (T-1) + (T-1) \times (T+1)$. We present the results with fixed $\gamma = 1, \rho = 0.5, \sigma^2 = 1, \sigma_\eta^2 = 1$, $\sigma_e^2 = 1$, and $N = 50$.

The optimal regularization parameter is selected following the procedure described in Section 7. The number of replications is 3,000 for all cases. Different estimators of the parameter of interest $\delta$ are presented. We denote by IV1 and IV2 the one-step GMM estimators using, respectively, one and two lags of $(y_{it}, m_{it+1})$ as instruments and GMM the usual one-step GMM estimator using all available lags of $y_{it}$ and $(m_{i0}, \ldots, m_{iT})$ as instruments. The estimator denoted by OKUI corresponds to the estimator proposed in Okui (2009) using as instruments $(y_{it-1}, m_{it}, y_{t-2}, m_{it-1}, \ldots, y_{it-k}, m_{it-k+1})$, where $k$ is selected to minimize the approximate MSE.

The regularization parameter is selected by minimizing the approximate MSE $\widehat{S}(\alpha)$ as described in Section 6. The IV1 estimator is used as plug-in in the minimization criterion. Hence, the estimate for the variance $\sigma^2$ is given by

$$\widehat{\sigma^2} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T-1} (y_{it}^* - \hat{\delta} x_{it}^*)^2,$$

where $\hat{\delta}$ denotes the IV1 estimator of the parameter of interest $\delta$. For each estimator, we compute the median bias, the median absolute deviation (mad) defined as $\text{Med}\left|\hat{\delta} - \delta\right|$, the empirical standard error, the length of the interquartile range, and the empirical coverage probabilities of the 95 % confidence intervals. The standard error is computed with the formula

$$\bar{V} = \sqrt{\widehat{\sigma^2} \frac{\left(x^{*\prime} M^{\alpha 2} x^*\right)}{\left(x^{*\prime} M^{\alpha} x^*\right)^2}}.$$

Tables A2–A5 in Appendix A contain simulations for two values of $T$ ($T = 10$ and $T = 25$) and two values of $\delta$ ($\delta = 0.5$ and $\delta = 0.95$). For each simulation setting, we provide a table of the properties of the optimal regularization parameter. Table A2 contains summary statistics for the value of the regularization parameter, which minimizes the approximate MSE for the different values of $\delta$ and $T$. The regularization parameter is the number of lags $k$ included (the corresponding number of instruments is $2k$) for OKUI, the optimal $\alpha$ for TK, the optimal number of iterations for LF, and the optimal number of PCs for PC. We report the mean, the standard deviation (std), the mode, and the three quartiles of the distribution of the regularization parameter. The regularization parameter is the optimal $\alpha$ for TK, the optimal number of iterations for LF, and the optimal number of PCs for PC. Noting that the standard GMM estimator corresponds to the TK estimator with $\alpha = 0$, to the LF estimator with an infinite number of iterations, and to the PC using all the PCs, a higher level of regularization will correspond to a larger value of $\alpha$, a smaller number of PCs, and a smaller number of iterations for the LF estimator.

When comparing the regularization parameters selected for $\delta = 0.5$ and $\delta = 0.95$, we observe that the regularization is stronger for $\delta = 0.95$ (fewer instruments for OKUI, larger $\alpha$ for TK, fewer PCs for PC, and fewer iterations for LF), this is due to the fact that the instruments become weak and the covariance matrix is more ill-conditioned for larger $\delta$. For $\delta = 0.95$, the automatic selection of $k$ for OKUI selects always $k = 1$, i.e., OKUI estimator coincides with IV1 which uses two instruments $(y_{it-1}, m_{it})$. We also observe that bias and standard errors increase for all the methods when $\delta$ goes from 0.5 to 0.95 and that the coverage deteriorates. This is again consistent with the fact that the instruments become weak when $\delta$ approaches 1.

The comparison between the one-step GMM using all instruments and the regularized methods depends on the estimated parameter. For $\hat{\delta}$, we found that the regularized estimators exhibit a smaller median bias and empirical standard error than GMM. For $\hat{\gamma}$, the median bias is still smaller for the regularized methods, but GMM exhibits smaller s.e. than regularized methods. So, the reduction of the MSE comes from reducing bias rather than variance. This is consistent with our theoretical results (see Section 5).

In terms of the mad, we see that regularized estimators have smaller mad than all other estimators for $\hat{\delta}$. For $\hat{\gamma}$, the smallest mad is obtained for GMM followed by the regularized methods. Among the three regularized estimators, PC and TK have, in general, the smallest mad for $\hat{\delta}$ and $\hat{\gamma}$, respectively. In terms of bias, OKUI has the smallest median bias for $\hat{\delta}$ and $\hat{\gamma}$ when $\delta = 0.5$. However, for $\delta = 0.95$, the median bias is smallest for PC and the bias of OKUI has substantially increased compared to the case where $\delta = 0.5$. Regarding coverage, IV1, IV2, and OKUI give the best coverage for $\hat{\delta}$ and $\hat{\gamma}$ when $\delta = 0.5$ and for $\hat{\delta}$ for $\delta = 0.95$, whereas the regularized methods give the best coverage for $\hat{\gamma}$ when $\delta = 0.95$. For $\delta = 0.95$ and $T = 10$, LF and PC tend to have large standard errors and interquartile range compared to TK, which has smaller dispersion in small samples. PC has almost always the smallest median bias and the best coverage among regularized estimators for $\delta = 0.5$ and $\delta = 0.95$.

## 8.2. Model with Heteroskedasticity

All the theory has been developed assuming homoskedastic errors. However, in practice, one rarely knows for sure whether the observations are homoskedastic or not. Therefore, we conduct simulations to investigate the impact of heteroskedasticity. The data are generated as in (6) except for the way $v_{it}$ is generated. The $v_{it}$ are conditionally independent and follow a mean-zero normal distribution with variance equal to $0.8m_{it}^2$. We implement the exact same estimators as before with the same weighting matrix as in the homoskedastic case and the same calculation of the standard error to construct the confidence intervals. It is expected that the resulting estimators will lack efficiency and that the coverage rates may be off.

From Tables A6–A9, we see that all methods exhibit a larger median bias, mad, and dispersion in the heteroskedastic case compared to the homoskedastic case.

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

As expected, the empirical coverage rates are further from the nominal coverage rates because the standard errors do not account for the heteroskedasticity. The regularized methods seem to be less sensitive to heteroskedasticity than other methods, and this robustness is more evident when $\delta = 0.95$. Indeed, the coverage of nonregularized methods and OKUI is far from the nominal coverage when $\delta = 0.95$, whereas that of the regularized estimators remains close to the nominal coverage. PC and TK estimators exhibit a much smaller mad than IV1, IV2, and OKUI, especially when $\delta$ is close to unity. PC has the smallest median absolute bias for $\hat{\delta}$ and $\hat{\gamma}$ and the smallest mad for $\hat{\delta}$. For $\hat{\gamma}$, GMM has the smallest mad. Overall, PC is the preferred method as it gives the best results in terms of median absolute bias, mad, and coverage most of the time.

## 9. EMPIRICAL APPLICATION

In this section, we apply the regularization approach to estimate a model of income dynamics. We consider the following equation:

$$y_{it} = \delta y_{i,t-1} + \eta_i + v_{it} \quad \text{for } i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \tag{7}$$

where $y_{it}$ is the earning residuals of person $i$ at time period $t$, controlled for his observable personal characteristics. This empirical application is similar to the one in Hirano (2002). The data come from the Panel Study of Income Dynamics[4] and include male household heads of 24 and 45 years old between 1967 and 2016. Similarly to Hirano (2002), the time dimension in equation (7) is age and can then be any calendar year. The estimation sample includes individuals with positive earnings at least 1 year prior to the age of 24, with complete data on education, race, and with earnings for all the successive ages. The final sample is a balanced panel of $N = 92$ and $T + 1 = 22$.

For each age $t$, we regress the logarithm of the real earnings (based on the consumer price index) on a constant, a dummy for non-white person, education dummies including high school, some college, college, and beyond college, and calendar year. The residual of this regression for each individual $i$ at each age $t$ is denoted by $y_{it}$ and used to estimate equation (7).

Table 1 presents the estimation results. We compute the following seven estimators: IV1 and IV2 which uses, respectively, one and two lags of $y_{it}$ to compute the one-step GMM estimator, GMM the usual one-step GMM estimator using all available lags of $y_{it}$, OKUI the estimator proposed in Okui (2009), TGMM the regularized estimator using TK, PGMM the regularized estimator using PC, and LGMM the regularized estimator using LF regularization scheme. We also report the standard errors for all these estimators (in parentheses) and the regularization parameter for regularized estimators. This parameter is the optimal $\alpha$ for TK, the number of PCs for PC, and the number of iterations for LF. For the estimator OKUI,

[4]We thank Keisuke Hirano for providing his dataset. We added extra years to the original data to obtain a larger sample.

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

**Table 1.** Application to income dynamics

| | IV1 | IV2 | GMM | OKUI | TGMM | PGMM | LGMM |
|---|---|---|---|---|---|---|---|
| $\delta$ | 0.678 | 0.654 | 0.652 | 0.667 | 0.676 | 0.665 | 0.678 |
| | (0.020) | (0.019) | (0.016) | (0.016) | (0.020) | (0.016) | (0.020) |
| $\alpha^*$ | | | | 8 | 6.6e-05 | 116 | 15,126 |

we report the number of lags that minimizes the approximated MSE proposed in Okui (2009).

The covariance matrix of the instruments is of dimension $\bar{q} = 210$, and its condition number is of order $10^6$, which is very high and motivates the use of regularization.

All the coefficients are larger than the GMM coefficient. This is consistent with the negative bias we found in Section 8. The methods giving the largest coefficient are IV1 and LGMM. Regularization provides a bias correction as the regularized coefficients are larger than GMM. The LF regularization scheme provides the largest bias correction, whereas PC provides the smallest. The bias reduction of regularization for the GMM estimator is obtained at the expense of a small increase of the standard errors. Note that the standard errors in Table 1 are not robust to heteroskedasticity. All the estimators are close to each other, suggesting that the autoregressive coefficient is around 0.67.

## 10. CONCLUSION AND FURTHER EXTENSIONS

In DPMs, the number of moment conditions increases with the sample size so that the one-step GMM estimator has poor finite sample properties. Instead of selecting a subset of moment conditions, we propose a regularization approach based on 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 these estimators outperform the classical one-step GMM estimator especially when the autoregressive coefficient is close to 1.

The regularization methods introduced in this paper can be extended to several other estimators. Under the i.i.d. assumption on the error term $v_{it}$, the weighting matrix of the GMM estimator is estimated by the inverse of $\sigma^2 Z'Z$. In the presence of heteroskedasticity or autocorrelation (or both), the covariance matrix of the instruments will be of the form $Z'A_N Z$ where $A_N$ is the estimated covariance matrix of the error terms. For example, when FDs are used to eliminate the unobserved fixed effect instead of the FOD used in this paper, the FD error term exhibits autocorrelation even under i.i.d. assumptions on the error term $v_{it}$. As a result, the weighting matrix depends on the FD error term. The Arellano and Bond (1991)

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

two-step FD GMM estimator is given by

$$\hat{\delta} = (\bar{y}_{-1}^{'} Z A_N Z^{'} \bar{y}_{-1})^{-1} (\bar{y}_{-1}^{'} Z A_N Z^{'} \bar{y}_{-1}),$$

where $\bar{y} = (\bar{y}_1^{'}, \ldots, \bar{y}_N^{'})$, $\bar{y}_{-1} = (\bar{y}_{1,-1}^{'}, \ldots, \bar{y}_{N,-1}^{'})$, $\bar{y}_i^{'} = (\Delta y_{i3}, \ldots, \Delta y_{iT})$, and $\bar{y}_{i,-1}^{'} = (\Delta y_{i2}, \ldots, \Delta y_{i,T-1})$, and, for a given $t$, we have $\Delta y_{it} = y_{it} - y_{i,t-1}$. The matrix $A_N$ is given by

$$A_N = \left( \frac{1}{N} \sum_{i=1}^{N} Z_i^{'} \hat{\hat{u}}_i \hat{\hat{u}}_i^{'} Z_i \right)^{-1},$$

where $\widehat{\hat{u}_i} = (\Delta \widehat{v}_{i3}, \ldots, \Delta \widehat{v}_{iT})$ is computed from an initial consistent estimator. A regularized version of the inverse in $A_N$ is likely to improve the properties of this estimator when the number of instruments is large.

More generally, regularization can be applied to heteroskedasticity and autocorrelation consistent-type weighting matrices as in Carrasco et al. (2007a). Establishing consistency and asymptotic normality of this estimator should not be difficult. However, deriving the MSE expansion to select the regularization parameter will be more complicated. As a solution to the poor finite sample properties of the GMM estimator of the DPM, Blundell and Bond (1998) proposed the system GMM estimator that combines moment conditions for the model in FDs with moment conditions for the model in levels. However, even though the system GMM estimator is widely used in the literature (Blundell and Bond, 2000; Levine, Loayza, and Beck, 2000 among others), it exhibits a bias when the instruments are weak or nearly weak. In fact, 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. Extending the regularization approach to the GMM system would be of great interest.

In this paper, we consider a single equation version of the DPM. The regularization approach can be extended to the multiple equations setting referred to as the panel dynamic simultaneous equations model. Several estimators including the GMM and the System GMM estimator of the panel dynamic simultaneous equation models are presented in Mitze (2012), whereas Hsiao and Zhou (2018) investigate the asymptotic properties of the GMM estimator in this model. In comparison to the single equation model, the covariance matrix of the instruments has a larger dimension in the multiple equations setting because of the use of instruments from multiple equations. Since our simulations show that the relative performance of regularization increases with the number of instruments (large $T$ or model with covariates), one should expect regularization to significantly improve the finite sample properties of the GMM estimator of the panel dynamic simultaneous equations model.

# APPENDIX A: Simulation Tables

**Table A1.** Properties of the condition number with $N = 50$, $\sigma^2 = 1$, and $\sigma_\eta^2 = 1$ for 1,000 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 |
| $T = 50$ | 17,418.9 | 68,429.4 | 734,508.6 | 137,592.7 | 310,565.6 | 92,594,681.5 | 1,225 |
| $\delta = 0.50$ | | | | | | | |
| $T = 5$ | 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 | 1,176.6 | 1,119.4 | 1,361.3 | 3,137.3 | 300 |
| $T = 50$ | 33,789.1 | 217,795.9 | 2,097,168.8 | 428,699.4 | 970,051.8 | 356,480,210.5 | 1,225 |
| $\delta = 0.90$ | | | | | | | |
| $T = 5$ | 626.2 | 1,183.7 | 1,509.0 | 1,440.7 | 1,751.9 | 3,901.3 | 10 |
| $T = 10$ | 2,168.4 | 4,245.2 | 5,316.8 | 5,067.7 | 6,214.7 | 13,130.8 | 45 |
| $T = 25$ | 16,422.2 | 32,971.4 | 42,706.8 | 40,619.7 | 49,839.7 | 108,698.3 | 300 |
| $T = 50$ | 1,124,835.1 | 7,618,036.7 | 242,677,244.8 | 14,558,610.1 | 36,715,676.3 | 178,194,140,502.4 | 1,225 |

**Table A2.** Properties of the distribution of the regularization parameters with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\delta = 0.50$, $\gamma = 1$, and $\rho = 0.50$ for 3,000 replications (homoskedastic case)

| | | Mean | std | Mode | q1 | Median | q3 |
|---|---|---|---|---|---|---|---|
| $T = 10$ | | | | | | | |
| | OKUI | 1.5 | 0.5 | 1.0 | 1.0 | 1.0 | 2.0 |
| | TK | 7.00e-04 | 3.79e-04 | 5.63e-04 | 4.69e-04 | 6.25e-04 | 8.13e-04 |
| | PC | 86.0 | 13.7 | 95.0 | 77.0 | 87.0 | 96.0 |
| | LF | Inf | NaN | 85,332.0 | 55,350.0 | 81,919.0 | 120,470.0 |
| $T = 25$ | | | | | | | |
| | OKUI | 3.4 | 0.5 | 3.0 | 3.0 | 3.0 | 4.0 |
| | TK | 1.48e-04 | 4.05e-05 | 1.56e-04 | 1.17e-04 | 1.41e-04 | 1.72e-04 |
| | PC | 333.2 | 51.9 | 332.0 | 298.0 | 332.0 | 370.0 |
| | LF | 161,802.0 | 80,383.1 | 127,999.0 | 102,399.0 | 146,285.0 | 204,799.0 |

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

**Table A3.** Simulations results for $\hat{\delta}$ and $\hat{\gamma}$ with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\delta = 0.50$, $\gamma = 1$, and $\rho = 0.50$ for 3,000 replications (homoskedastic case)

| | GMM | IV1 | IV2 | OKUI | TK | PC | LF |
|---|---|---|---|---|---|---|---|
| $\hat{\delta}$ | | | | | | | |
| $T = 10$ | | | | | | | |
| Median bias | −0.0376 | −0.0187 | −0.0217 | −0.0109 | −0.0243 | −0.0233 | −0.0238 |
| Median absolute deviation | 0.0391 | 0.0532 | 0.0416 | 0.0450 | 0.0313 | 0.0309 | 0.0311 |
| Empirical standard error | 0.0525 | 0.0832 | 0.0620 | 0.0683 | 0.0460 | 0.0456 | 0.0458 |
| Interquartile range | 0.0485 | 0.1053 | 0.0785 | 0.0896 | 0.0509 | 0.0518 | 0.0519 |
| Coverage rate | 0.8163 | 0.9450 | 0.9400 | 0.9577 | 0.8897 | 0.8943 | 0.8940 |
| $T = 25$ | | | | | | | |
| Median bias | −0.0306 | −0.0075 | −0.0073 | −0.0064 | −0.0149 | −0.0135 | −0.0143 |
| Median absolute deviation | 0.0306 | 0.0268 | 0.0213 | 0.0184 | 0.0179 | 0.0174 | 0.0176 |
| Empirical standard error | 0.0362 | 0.0409 | 0.0316 | 0.0274 | 0.0258 | 0.0255 | 0.0256 |
| Interquartile range | 0.0264 | 0.0532 | 0.0414 | 0.0367 | 0.0289 | 0.0292 | 0.0291 |
| Coverage rate | 0.6267 | 0.9550 | 0.9427 | 0.9390 | 0.8840 | 0.9010 | 0.8910 |
| $\hat{\gamma}$ | | | | | | | |
| $T = 10$ | | | | | | | |
| Median bias | −0.0081 | −0.0147 | −0.0137 | −0.0030 | −0.0059 | −0.0030 | −0.0053 |
| Median absolute deviation | 0.0330 | 0.0515 | 0.0436 | 0.0455 | 0.0333 | 0.0353 | 0.0339 |
| Empirical standard error | 0.0478 | 0.0791 | 0.0650 | 0.0706 | 0.0496 | 0.0532 | 0.0507 |
| Interquartile range | 0.0646 | 0.1005 | 0.0871 | 0.0921 | 0.0663 | 0.0702 | 0.0682 |
| Coverage rate | 0.9473 | 0.9473 | 0.9517 | 0.9573 | 0.9520 | 0.9520 | 0.9517 |
| $T = 25$ | | | | | | | |
| Median bias | −0.0019 | −0.0029 | −0.0020 | 0.0002 | −0.0006 | 0.0007 | −0.0008 |
| Median absolute deviation | 0.0194 | 0.0253 | 0.0232 | 0.0229 | 0.0219 | 0.0242 | 0.0228 |
| Empirical standard error | 0.0289 | 0.0382 | 0.0342 | 0.0338 | 0.0326 | 0.0362 | 0.0339 |
| Interquartile range | 0.0383 | 0.0518 | 0.0465 | 0.0460 | 0.0438 | 0.0485 | 0.0459 |
| Coverage rate | 0.9477 | 0.9467 | 0.9460 | 0.9470 | 0.9507 | 0.9513 | 0.9487 |

**Table A4.** Properties of the distribution of the regularization parameters with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\delta = 0.95$, $\gamma = 1$, and $\rho = 0.50$ for 3,000 replications (homoskedastic case)

| | | Mean | std | Mode | q1 | Median | q3 |
|---|---|---|---|---|---|---|---|
| $T = 10$ | | | | | | | |
| | OKUI | 1.0 | | 1.0 | 1.0 | 1.0 | 1.0 |
| | TK | 3.03e-03 | 1.64e-03 | 2.25e-03 | 2.06e-03 | 2.63e-03 | 3.50e-03 |
| | PC | 44.4 | 18.3 | 9.0 | 34.0 | 46.0 | 57.0 |
| | LF | 1,241,287.7 | 13,090,177.6 | 2,285.0 | 45.0 | 999.0 | 1,999.0 |
| $T = 25$ | | | | | | | |
| | OKUI | 1.0 | | 1.0 | 1.0 | 1.0 | 1.0 |
| | TK | 9.69e+00 | 4.62e+02 | 1.31e-03 | 1.31e-03 | 1.94e-03 | 3.44e-03 |
| | PC | 60.7 | 43.1 | 24.0 | 34.0 | 49.0 | 67.0 |
| | LF | 65,066,505.1 | 96,025,335.9 | 15,999.0 | 7,999.0 | 21,845,332.0 | 99,864,380.0 |

**Table A5.** Simulations results for $\hat{\delta}$ and $\hat{\gamma}$ with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\delta = 0.95$, $\gamma = 1$, and $\rho = 0.50$ for 3,000 replications (homoskedastic case)

| | GMM | IV1 | IV2 | OKUI | TK | PC | LF |
|---|---|---|---|---|---|---|---|
| $\hat{\delta}$ | | | | | | | |
| $T = 10$ | | | | | | | |
| Median bias | −0.0706 | −0.1267 | −0.1261 | −0.1267 | −0.0617 | −0.0543 | −0.1700 |
| Median absolute deviation | 0.0706 | 0.1331 | 0.1270 | 0.1331 | 0.0617 | 0.0581 | 0.1918 |
| Empirical standard error | 0.0785 | 0.1983 | 0.1623 | 0.1983 | 0.0767 | 1.0500 | 0.3621 |
| Interquartile range | 0.0444 | 0.1821 | 0.1271 | 0.1821 | 0.0544 | 0.0726 | 0.3314 |
| Coverage rate | 0.4147 | 0.8470 | 0.7077 | 0.8470 | 0.6277 | 0.7867 | 0.9667 |
| $T = 25$ | | | | | | | |
| Median bias | −0.0492 | −0.0469 | −0.0463 | −0.0469 | −0.0390 | −0.0368 | −0.0437 |
| Median absolute deviation | 0.0492 | 0.0503 | 0.0465 | 0.0503 | 0.0391 | 0.0410 | 0.0472 |
| Empirical standard error | 0.0515 | 0.0779 | 0.0613 | 0.0779 | 0.0482 | 0.1185 | 0.1068 |
| Interquartile range | 0.0179 | 0.0736 | 0.0523 | 0.0736 | 0.0335 | 0.0618 | 0.0592 |
| Coverage rate | 0.0133 | 0.8687 | 0.7727 | 0.8687 | 0.5757 | 0.8613 | 0.7663 |

(Continues)

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

**Table A5.** Continued

| | GMM | IV1 | IV2 | OKUI | TK | PC | LF |
|---|---|---|---|---|---|---|---|
| $\hat{\gamma}$ | | | | | | | |
| $T = 10$ | | | | | | | |
| Median bias | −0.0329 | −0.1184 | −0.1140 | −0.1184 | −0.0293 | −0.0195 | −0.0629 |
| Median absolute deviation | 0.0406 | 0.1291 | 0.1168 | 0.1291 | 0.0460 | 0.0623 | 0.2624 |
| Empirical standard error | 0.0580 | 0.1912 | 0.1566 | 0.1912 | 0.0653 | 0.7765 | 0.4366 |
| Interquartile range | 0.0669 | 0.1784 | 0.1377 | 0.1784 | 0.0813 | 0.1224 | 0.5072 |
| Coverage rate | 0.8973 | 0.8670 | 0.7760 | 0.8670 | 0.9180 | 0.9473 | 0.9943 |
| $T = 25$ | | | | | | | |
| Median bias | −0.0165 | −0.0402 | −0.0370 | −0.0402 | −0.0153 | −0.0075 | −0.0163 |
| Median absolute deviation | 0.0231 | 0.0460 | 0.0399 | 0.0460 | 0.0392 | 0.0924 | 0.0657 |
| Empirical standard error | 0.0342 | 0.0719 | 0.0579 | 0.0719 | 0.0625 | 0.1917 | 0.1550 |
| Interquartile range | 0.0402 | 0.0741 | 0.0598 | 0.0741 | 0.0737 | 0.1815 | 0.1285 |
| Coverage rate | 0.9023 | 0.8870 | 0.8607 | 0.8870 | 0.9423 | 0.9583 | 0.9597 |

**Table A6.** Properties of the distribution of the regularization parameters with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\delta = 0.50$, $\gamma = 1$, and $\rho = 0.50$ for 3,000 replications (heteroskedastic case)

| | Mean | std | Mode | q1 | Median | q3 |
|---|---|---|---|---|---|---|
| $T = 10$ | | | | | | |
| OKUI | 1.5 | 0.5 | 1.0 | 1.0 | 1.0 | 2.0 |
| TK | 6.41e-04 | 5.02e-04 | 5.63e-04 | 3.75e-04 | 5.31e-04 | 7.50e-04 |
| PC | 88.1 | 14.6 | 92.0 | 79.0 | 90.0 | 99.0 |
| LF | 109,021.9 | 71,178.5 | 127,999.0 | 62,060.0 | 93,090.0 | 136,532.0 |
| $T = 25$ | | | | | | |
| OKUI | 3.4 | 0.6 | 3.0 | 3.0 | 3.0 | 4.0 |
| TK | 1.38e-04 | 4.86e-05 | 1.25e-04 | 1.02e-04 | 1.29e-04 | 1.64e-04 |
| PC | 340.1 | 54.7 | 325.0 | 301.0 | 340.0 | 378.0 |
| LF | 177,463.4 | 88,421.5 | 170,666.0 | 113,777.0 | 157,537.0 | 215,578.0 |

**TABLE A7.** Simulations results for $\hat{\delta}$ and $\hat{\gamma}$ with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\delta = 0.50$, $\gamma = 1$, and $\rho = 0.50$ for 3,000 replications (heteroskedastic case)

| | GMM | IV1 | IV2 | OKUI | TK | PC | LF |
|---|---|---|---|---|---|---|---|
| $\hat{\delta}$ | | | | | | | |
| $T = 10$ | | | | | | | |
| Median bias | −0.0397 | −0.0450 | −0.0357 | −0.0300 | −0.0269 | −0.0263 | −0.0261 |
| Median absolute deviation | 0.0403 | 0.0686 | 0.0494 | 0.0548 | 0.0332 | 0.0335 | 0.0330 |
| Empirical standard error | 0.0550 | 0.1150 | 0.0750 | 0.0853 | 0.0486 | 0.0484 | 0.0483 |
| Interquartile range | 0.0489 | 0.1338 | 0.0890 | 0.1026 | 0.0536 | 0.0547 | 0.0544 |
| Coverage rate | 0.7847 | 0.8393 | 0.8757 | 0.8923 | 0.8733 | 0.8763 | 0.8763 |
| $T = 25$ | | | | | | | |
| Median bias | −0.0310 | −0.0193 | −0.0135 | −0.0111 | −0.0151 | −0.0140 | −0.0145 |
| Median absolute deviation | 0.0310 | 0.0349 | 0.0240 | 0.0197 | 0.0179 | 0.0178 | 0.0178 |
| Empirical standard error | 0.0359 | 0.0554 | 0.0359 | 0.0297 | 0.0260 | 0.0257 | 0.0258 |
| Interquartile range | 0.0253 | 0.0666 | 0.0442 | 0.0372 | 0.0283 | 0.0288 | 0.0290 |
| Coverage rate | 0.6263 | 0.8567 | 0.8980 | 0.9170 | 0.8773 | 0.8883 | 0.8823 |
| $\hat{\gamma}$ | | | | | | | |
| $T = 10$ | | | | | | | |
| Median bias | −0.0105 | −0.0334 | −0.0231 | −0.0116 | −0.0063 | −0.0044 | −0.0061 |
| Median absolute deviation | 0.0499 | 0.0770 | 0.0645 | 0.0664 | 0.0503 | 0.0514 | 0.0507 |
| Empirical standard error | 0.0732 | 0.1185 | 0.0938 | 0.1022 | 0.0756 | 0.0782 | 0.0766 |
| Interquartile range | 0.0987 | 0.1460 | 0.1236 | 0.1312 | 0.1005 | 0.1023 | 0.1022 |
| Coverage rate | 0.7903 | 0.7900 | 0.8157 | 0.8293 | 0.7960 | 0.8093 | 0.8000 |
| $T = 25$ | | | | | | | |
| Median bias | −0.0033 | −0.0095 | −0.0054 | −0.0001 | −0.0026 | −0.0013 | −0.0030 |
| Median absolute deviation | 0.0313 | 0.0393 | 0.0347 | 0.0342 | 0.0341 | 0.0354 | 0.0346 |
| Empirical standard error | 0.0458 | 0.0595 | 0.0516 | 0.0509 | 0.0494 | 0.0522 | 0.0505 |
| Interquartile range | 0.0614 | 0.0788 | 0.0681 | 0.0686 | 0.0681 | 0.0708 | 0.0691 |
| Coverage rate | 0.7797 | 0.7883 | 0.7940 | 0.8120 | 0.8013 | 0.8337 | 0.8107 |

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

**TABLE A8.** Properties of the distribution of the regularization parameters with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\delta = 0.95$, $\gamma = 1$, and $\rho = 0.50$ for 3,000 replications (heteroskedastic case)

| | Mean | std | Mode | q1 | Median | q3 |
|---|---|---|---|---|---|---|
| $T = 10$ | | | | | | |
| OKUI | 1.0 | | 1.0 | 1.0 | 1.0 | 1.0 |
| TK | 2.77e-03 | 1.49e-03 | 2.31e-03 | 1.88e-03 | 2.44e-03 | 3.25e-03 |
| PC | 46.5 | 17.9 | 9.0 | 37.0 | 49.0 | 59.0 |
| LF | 2,714,292.0 | 22,736,444.0 | 2,285.0 | 95.0 | 1,142.0 | 1,999.0 |
| $T = 25$ | | | | | | |
| OKUI | 1.0 | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| TK | 9.55e+00 | 4.47e+02 | 1.31e-03 | 1.22e-03 | 1.81e-03 | 3.13e-03 |
| PC | 63.6 | 45.4 | 24.0 | 35.0 | 50.0 | 70.5 |
| LF | 76,597,852.3 | 103,442,858.4 | 15,999.0 | 15,999.0 | 37,449,142.0 | 116,508,443.0 |

**TABLE A9.** Simulations results for $\hat{\delta}$ and $\hat{\gamma}$ with $N = 50$, $\sigma^2 = 1$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\delta = 0.95$, $\gamma = 1$, and $\rho = 0.50$ for 3,000 replications (heteroskedastic case)

| | GMM | IV1 | IV2 | OKUI | TK | PC | LF |
|---|---|---|---|---|---|---|---|
| $\hat{\delta}$ | | | | | | | |
| $T = 10$ | | | | | | | |
| Median bias | −0.0732 | −0.2404 | −0.1862 | −0.2404 | −0.0669 | −0.0589 | −0.1879 |
| Median absolute deviation | 0.0732 | 0.2409 | 0.1862 | 0.2409 | 0.0670 | 0.0615 | 0.2090 |
| Empirical standard error | 0.0833 | 0.3076 | 0.2218 | 0.3076 | 0.0825 | 0.2357 | 0.3794 |
| Interquartile range | 0.0466 | 0.2323 | 0.1475 | 0.2323 | 0.0571 | 0.0774 | 0.3582 |
| Coverage rate | 0.3687 | 0.4863 | 0.4260 | 0.4863 | 0.5617 | 0.7320 | 0.9240 |
| $T = 25$ | | | | | | | |
| Median bias | −0.0493 | −0.1087 | −0.0764 | −0.1087 | −0.0417 | −0.0413 | −0.0465 |
| Median absolute deviation | 0.0493 | 0.1087 | 0.0764 | 0.1087 | 0.0418 | 0.0447 | 0.0487 |
| Empirical standard error | 0.0518 | 0.1430 | 0.0924 | 0.1430 | 0.0528 | 0.0978 | 0.1082 |
| Interquartile range | 0.0184 | 0.1130 | 0.0626 | 0.1130 | 0.0346 | 0.0619 | 0.0567 |
| Coverage rate | 0.0133 | 0.4493 | 0.4420 | 0.4493 | 0.5150 | 0.8090 | 0.6957 |

(Continues)

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

**Table A9.** Continued

| | GMM | IV1 | IV2 | OKUI | TK | PC | LF |
|---|---|---|---|---|---|---|---|
| $\hat{\gamma}$ | | | | | | | |
| $T = 10$ | | | | | | | |
| Median bias | −0.0337 | −0.2257 | −0.1694 | −0.2257 | −0.0322 | −0.0222 | −0.0714 |
| Median absolute deviation | 0.0565 | 0.2271 | 0.1706 | 0.2271 | 0.0605 | 0.0791 | 0.3003 |
| Empirical standard error | 0.0824 | 0.2945 | 0.2155 | 0.2945 | 0.0902 | 0.3305 | 0.4973 |
| Interquartile range | 0.1007 | 0.2315 | 0.1730 | 0.2315 | 0.1119 | 0.1529 | 0.5905 |
| Coverage rate | 0.7403 | 0.5493 | 0.5427 | 0.5493 | 0.7727 | 0.8463 | 0.9797 |
| $T = 25$ | | | | | | | |
| Median bias | −0.0175 | −0.0935 | −0.0633 | −0.0935 | −0.0173 | −0.0051 | −0.0139 |
| Median absolute deviation | 0.0335 | 0.0944 | 0.0662 | 0.0944 | 0.0472 | 0.0992 | 0.0743 |
| Empirical standard error | 0.0500 | 0.1280 | 0.0889 | 0.1280 | 0.0771 | 0.1888 | 0.1759 |
| Interquartile range | 0.0620 | 0.1136 | 0.0825 | 0.1134 | 0.0922 | 0.1980 | 0.1443 |
| Coverage rate | 0.7407 | 0.5483 | 0.6120 | 0.5483 | 0.8537 | 0.9160 | 0.8940 |

# APPENDIX B: Proofs

## B.1. Proof of Lemma 1

(i) We have

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

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

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

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

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

Now, using the autoregressive equation defining the DPM model, we introduce the notation $w_{it} = y_{it} - \eta_i/(1-\delta) = y_{it} - \mu_i$. Note that $w_{it}$ is a stationary AR(1) process with mean 0

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

and variance $\sigma^2/\left(1-\delta^2\right)$. It follows that

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

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

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

Now,

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

It follows that

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

(ii) For any symmetric matrix $A = (a_{ij})$, we have $tr[A^2] = \sum_{i,j} a_{i,j}^2$. So the trace of $tr[K^2]$ is given by the sum of the squares of all the elements of the matrix $K$. By construction, we have

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

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

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

with $1 \le a,b \le t$.

We now calculate $tr[K_t^2]$ by summing the squares of the elements of $K_t$.

$$tr[K_t^2] = \sum_{a,b=1}^{t} K_{ab,t}^2 = \sum_{a=1}^{t} K_{aa,t}^2 + \sum_{a\neq b=1}^{t} K_{ab,t}^2.$$

We have

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

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

$$= \frac{1}{T^3}\frac{\sigma^4}{(1-\delta^2)^2}\sum_{a\neq b=1}^{t}\delta^{2(a-b)} + \frac{2}{T^3}\frac{\sigma^2}{(1-\delta^2)}\frac{\sigma_\eta^2}{(1-\delta)^2}\sum_{a\neq b=1}^{t}\delta^{|a-b|}$$

$$+\frac{1}{T^3}\sum_{a\neq b=1}^{t}\frac{\sigma_\eta^4}{(1-\delta)^4}.$$

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

However, $\sum_{a\neq b=1}^{t} \delta^{2(a-b)} = 2\sum_{a=2}^{t}\sum_{b=1}^{a-1}\delta^{2b} = 2\sum_{a=2}^{t}\left[\frac{1-\delta^{2a}}{(1-\delta^2)} - 1\right] = \frac{2}{(1-\delta^2)}\sum_{a=2}^{t}\left[\delta^2 - \delta^{2a}\right] = \frac{2}{(1-\delta^2)}\left[\sum_{a=2}^{t}\delta^2 - \sum_{a=2}^{t}\delta^{2a}\right] = \frac{2}{(1-\delta^2)}\left[\delta^2(t-1) - \sum_{a=2}^{t}\delta^{2a}\right] = \frac{2}{(1-\delta^2)}\left[\delta^2(t-1) - \left(\frac{1-\delta^{2t+2}}{(1-\delta^2)} - 1 - \delta^2\right)\right] = O(t)$. Similarly, we have $\sum_{a\neq b=1}^{t}\delta^{(a-b)} = O(t)$. Moreover, $\sum_{a\neq b=1}^{t}\frac{\sigma_\eta^4}{(1-\delta)^4} = O(t^2)$. From the last three results, we have $\sum_{a\neq b=1}^{t} K_{ab,t}^2 = O\left(\frac{t}{T^3}\right)$ and then $tr[K_t^2] = \sum_{a,b=1}^{t} K_{ab,t}^2 = O\left(\frac{t}{T^3}\right) + O\left(\frac{t^2}{T^3}\right) = O\left(\frac{t^2}{T^3}\right)$. And finally,

$$tr[K^2] = \sum_{t=1}^{T-1} tr[K_t^2] = \sum_{t=1}^{T-1} O\left(\frac{t^2}{T^3}\right) = O(1).$$

## B.2. Preliminary Results for the Proof of Proposition 1

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

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

(i) $||K_N - K|| = O_p(1/\sqrt{N})$;
(ii) $Etr[M^\alpha] = O(1/\alpha)$.

**Proof of Lemma 2.** (i) Let us define $K_{N,t}$ and $K_t$ as the $t$th blocks of the matrixes $K_N$ and $K$.

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

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

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

Hence,

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

By Cauchy–Schwarz, we have

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

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

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

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

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

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

(ii)

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

where $\widehat{\lambda}_j$ are the eigenvalues of the matrix $K_N$. From Kress (1999) and Carrasco et al. (2007b, Sect. 3.3), we have that for the three regularizations, $q(\alpha, \lambda^2) \leq C\lambda^2/\alpha$ for some positive constant $C$. Then we have

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

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

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

where $E||K_N - K||^2 = O(1/N)$ comes from $(i)$ and $E||K||^2 = O(1)$ comes from Lemma 1$(ii)$. This completes the proof of Lemma 2. □

LEMMA 3. *Let us denote by $d_t(\alpha)$ the $N \times 1$ vectors containing the diagonal elements of $M_t^\alpha$, and let $\kappa_3$ and $\kappa_4$ be the third- and fourth-order cumulants of $v_{it}$. Under Assumptions 1–3:*

(i) $tr(M_t^\alpha) \le t$,
(ii) $Var(v_t' M_t^\alpha v_t) \le (2\sigma^4 + \kappa_4) E\left[tr\left(M_t^{\alpha 2}\right)\right] \le (2\sigma^4 + \kappa_4)t$,
(iii) $Var(v_t' M_t^\alpha v_{t+j}) = \sigma^4 Etr(M_t^\alpha) \le \sigma^4 t$, for $j > 0$,
(iv) $Cov(v_t' M_t^\alpha v_{t+j}, v_{t+j}' M_{t+j}^\alpha v_{t+j}) \le \kappa_3 E\left(d_{t+j}(\alpha)' M_t^\alpha v_t\right) \le \kappa_3 \sigma \sqrt{t+j}\left(E\left(tr\left(M_t^{\alpha 2}\right)\right)\right)^{1/2}$, for $j > 0$.

**Proof of Lemma 3.** (i) The $t \times t$ symmetric matrix $Z_t' Z_t / NT^{3/2}$ can be decomposed as $P_t D_t P_t'$ with $P_t P_t' = I_t$ the $t$-dimensional identity matrix and $D_t = diag(\lambda_1^t, \lambda_2^t, \ldots, \lambda_t^t)$. The regularized inverse of $D_t$ is $D_t(\alpha)=diag(\frac{q(\alpha, \lambda_1^{t2})}{\lambda_1^t}, \ldots, \frac{q(\alpha, \lambda_t^{t2})}{\lambda_t^t})$. If we denote by $(Z_t' Z_t / NT^{3/2})^\alpha$ the regularized inverse of $(Z_t' Z_t / NT^{3/2})$, then

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

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

(ii)

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

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

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

so that

$$var_t(v_t' M_t^\alpha v_t) = E_t(v_t' M_t^\alpha v_t v_t' M_t^\alpha v_t) - \left(E_t(v_t' M_t^\alpha v_t)\right)^2 = \kappa_4 d_t'(\alpha) d_t(\alpha) + 2\sigma^4 tr(M_t^\alpha M_t^\alpha).$$

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

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

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

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

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

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

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

Moreover, by the Cauchy–Schwarz inequality,

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

Since $d_{t+k}'(\alpha) d_{t+k}(\alpha) \leq tr[M_{t+k}^{\alpha 2}] \leq t+k$ and $E(v_t' M_t^{\alpha 2} v_t) \leq \sigma^2 E\left[tr\left(M_t^{\alpha 2}\right)\right] \leq \sigma^2 t$, by taking expectation of the previous inequality, we have $E[(d_{t+k}'(\alpha) M_t^\alpha v_t)^2] \leq (t+k)\sigma^2 E\left[tr\left(M_t^{\alpha 2}\right)\right]$. The result $(iv)$ follows by noting that $[E(d_{t+k}'(\alpha) M_t^\alpha v_t)]^2 \leq E\left[(d_{t+k}'(\alpha) M_t^\alpha v_t)^2\right]$. This completes the proof of Lemma 3. □

LEMMA 4. *Let* $\tilde{v}_{tT} = \frac{1}{T-t}\left(\phi_{T-t} v_t + \cdots + \phi_1 v_{T-1}\right)$ *and* $\phi_j = \frac{1-\delta^j}{1-\delta}$. *If* $N \to \infty$, $T \to \infty$, *and* $\alpha \to 0$, *then:*
(i)

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

(ii)

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

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

$$\Upsilon_{21NT}^\alpha = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \tilde{v}_{tT}' M_t^\alpha v_t, \tag{8}$$

$$\Upsilon_{22NT}^\alpha = -\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \tilde{v}_{tT}' M_t^\alpha \bar{v}_{tT}. \tag{9}$$

*Moreover,*

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

*(iv)*

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

**Proof of Lemma 4.** (i) Let us define $W = (w_0', \ldots, w_{T-2}')'$, then

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

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

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

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

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

where the last inequality follows from AA. We have just proved that

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

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

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

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

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

(iii)

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

where $a_{0NT}^{\alpha}$ and $a_{1NT}^{\alpha}$ have the same form as $a_{0NT}$ and $a_{1NT}$ of AA (eqn. (A63)) but with $M_t^{\alpha}$ instead of $M_t$. First, consider $a_{0NT}^{\alpha}$ :

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

Using Lemma 3(*ii*) and (*iii*), we have that

$$Var(v_t' M_t^{\alpha} v_t) \leq (\kappa^4 + 2\sigma^4) E\left[tr\left(M_t^{\alpha 2}\right)\right],$$

and, for $j > 0$,

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

Hence, using $\phi_t^2 \leq 1/(1-\delta)^2$, we have

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

for some constant $C > 0$ and we can conclude that $a_{0NT}^{\alpha} = O\left(\frac{1}{\alpha NT}\right)$. Now, looking at $a_{1NT}^{\alpha}$, we have

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

Using Lemma 3($iv$), we have

$$\begin{aligned}|a^{\alpha}_{1NT}| &= \left|\frac{2}{NT}\sum_{t=1}^{T-2}\left[\sum_{j=1}^{T-t-1}\frac{\phi^2_{T-t-j}cov(v_t'M^{\alpha}_t v_{t+j}, v_{t+j}'M^{\alpha}_{t+j}v_{t+j})}{(T-t-j)(T-t)}\right]\right| \\ &\le \frac{1}{(1-\delta)^2}\frac{2}{NT}\left|\sum_{t=1}^{T-2}\frac{1}{(T-t)}\left[\sum_{j=1}^{T-t-1}\frac{\kappa_3 E(d_{t+j}(\alpha)M^{\alpha}_t v_t)}{T-t-j}\right]\right| \\ &\le \frac{\sigma}{(1-\delta)^2}\frac{2|\kappa_3|}{NT}\sum_{t=1}^{T-2}\frac{\sqrt{EtrM^{\alpha 2}_t}}{(T-t)}\left[\sum_{j=1}^{T-t-1}\frac{\sqrt{t+j}}{T-t-j}\right] \\ &\le \frac{1}{NT}\sum_{t=1}^{T-2}E\left[tr\left(M^{\alpha}_t\right)\right]O(\ln T)\end{aligned}$$

so that, by Lemma 2(ii), we have $a^{\alpha}_{1NT} = O\left(\ln T/(\alpha NT)\right)$. This allows us to conclude that $Var(\Upsilon^{\alpha}_{21NT}) = O\left(\ln T/(\alpha NT)\right)$.

We now look at the term $\Upsilon^{\alpha}_{22NT}$:

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

where using arguments similar to those of AA (eqns. (A72) and (A73)) and Okui (2011),

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

and

$$\begin{aligned}|b^{\alpha}_{1NT}| &\le \frac{2}{NT}\sum_s\sum_{t>s}|cov(\tilde{v}_{tT}'M^{\alpha}_t\bar{v}_{tT}, \tilde{v}_{sT}'M^{\alpha}_s\bar{v}_{sT})| \\ &= \frac{1}{NT}O\left(\sum_s\sum_{t>s}\frac{(E\left[tr\left(M^{\alpha 2}_t\right)\right])^{1/2}}{T-t}\frac{(Etr[M^{\alpha 2}_s])^{1/2}}{T-s}\right).\end{aligned}$$

But, for $s < t$,

$$E\left[tr\left(M^{\alpha 2}_t\right)\right] \le Etr[M^{\alpha 2}_s]$$

so that

$$|b^{\alpha}_{1NT}| \le \frac{C}{NT}\sum_{t=1}^{T-1}\frac{E\left[tr\left(M^{\alpha 2}_t\right)\right]}{T-t}\sum_{s=1}^{T-1}\frac{1}{T-s} = O\left(\frac{\ln(T)}{\alpha NT}\right)$$

and finally

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

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

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

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

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

Using the Cauchy–Schwarz inequality, we have

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

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

(iv) By the law of iterated expectations and equation (A47) of AA, we have

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

Hence,

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

Moreover, note that $\left|\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right| \le |\phi_{T-t}| + |\phi_{T-t+1}| \le 2/(1-\delta)$ so that

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

Therefore, Result (*iv*) follows from Lemma 2(*ii*). This completes the proof of Lemma 4. □

## B.3. Proof of Proposition 1

B.3.1. *Proof of Consistency.*

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

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

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} v_t^* = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} \psi_t w_{t-1}' M_t^{\alpha} v_t^* - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t \tilde{v}_{tT}' M_t^{\alpha} v_t^* \tag{10}$$

using $w_{t-1} = y_{t-1} - \mu$ with $\mu = \eta/(1-\delta)$, $c_t = \sqrt{(T-t)/(T-t+1)}$, $\phi_j = \frac{1-\delta^j}{1-\delta}, x_t^* = \psi_t w_{t-1} - c_t \tilde{v}_{tT}, \psi_t = c_t(1 - \frac{\delta\phi_{T-t}}{T-t})$, and $\tilde{v}_{tT} = \frac{(\phi_{T-t}v_t + \cdots + \phi_1 v_{T-1})}{T-t}$. The expectation of the first term of the right side of ( 10) is null, and by Lemma 4(iv), we have

$$E\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} v_t^*\right) = -E\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} c_t \tilde{v}'_{tT} M_t^{\alpha} v_t^*\right) = O\left(\frac{1}{\alpha\sqrt{NT}}\right), \tag{11}$$

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

We now look at the variance of $(x^{*'} M^{\alpha} v^*)/\sqrt{NT}$. Following the decomposition (A49) of AA, we can write

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

where $\Upsilon_{21NT}^{\alpha}$ and $\Upsilon_{22NT}^{\alpha}$ are defined in equations (8) and (9), respectively, and $\Upsilon_{11NT}^{\alpha} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w_{t-1}' M_t^{\alpha} \bar{v}_{tT}$, and $\Upsilon_{12NT}^{\alpha} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} \frac{c_t\delta\phi_{T-t}}{T-t} w_{t-1}' M_t^{\alpha} v_t^*$.

We have

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

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

From AA, $\frac{\sigma^2}{NT}\sum_{t=1}^{T-1} E(w_{t-1}' M_t w_{t-1}) \xrightarrow{m.s.} \frac{\sigma^4}{(1-\delta^2)}$.

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

Now, we give the order of magnitude of $\Upsilon_{11NT}^{\alpha}$, $\Upsilon_{12NT}^{\alpha}$, $\Upsilon_{21NT}^{\alpha}$, and $\Upsilon_{22NT}^{\alpha}$.

$$Var(\Upsilon_{11NT}^{\alpha}) = \frac{1}{NT}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} E(w_{t-1}' M_t^{\alpha} \bar{v}_{tT} \bar{v}_{sT} M_s^{\alpha} w_{s-1}).$$

For $t \geq s$,

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

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

$$\leq [E(w_0' M_1 w_0)]^{\frac{1}{2}} [E(w_0' M_1 w_0)]^{\frac{1}{2}}$$
$$\leq E(w_0' M_1 w_0).$$

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

Next, following (A60) of AA, we have

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

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

Turning to the denominator, we have

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

We can write the first term in the following way:

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

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

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

By Lemma 4(i), $\frac{1}{NT}\sum_{t=1}^{T-1} w'_{t-1}[M_t - M_t^\alpha]w_{t-1} = o_p(1)$. As a result, similarly to AA, 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 $\frac{\sigma^2}{(1-\delta^2)}$. The term $\frac{2}{NT}\sum_{t=1}^{T-1} c_t\psi_t w'_{t-1}$ $M_t^\alpha \tilde{v}_{tT}$ is identical to $\Upsilon_{11NT}^\alpha$ and is $o_p(1)$.

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

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

$$= \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 E\left[tr\left(M_t^\alpha\right)E_t(\tilde{v}_{itT}^2)\right]$$

$$\leq \frac{C}{NT}E\left(\sum_{t=1}^{T-1}[tr(M_t^\alpha)]\right) = O\left(\frac{1}{\alpha NT}\right),$$

where the last equality comes from Lemma 2(ii). By Markov's inequality,

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

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

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

B.3.2. *Proof of the Asymptotic Normality.* From (11) and Lemma 4(iv), we have

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

From (12) and since the variances of $\Upsilon_{11NT}^\alpha$, $\Upsilon_{12NT}^\alpha$, $\Upsilon_{22NT}^\alpha$, and $\Upsilon_{21NT}^\alpha$ go to zero, we obtain

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

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

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

Let us denote $h = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w'_{t-1}[M_t - M_t^\alpha]v_t$. By the law of iterated expectations, $E(h) = 0$ and $Var(h) = \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}E(w'_{t-1}[M_t - M_t^\alpha]^2 w_{t-1})$. By Lemma 4($i$), we have $Var(h) = o(1)$ so that $h = o_p(1)$.

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

From AA, $\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1} w'_{t-1}M_t v_t \xrightarrow{d} N(0,\frac{\sigma^2}{1-\delta^2})$ and we proved that $(x^{*'}M^{\alpha}x^*)/NT$ tends to $\frac{\sigma^2}{(1-\delta^2)}$ in probability so that, by Slutsky's theorem,

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

or equivalently

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

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

## B.4. Preliminary Results for the Proof of Proposition 2

Let

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

LEMMA 5. *If Assumptions 1–3 are satisfied, then:*

(i) $\Delta_{\alpha} = O\left(\frac{\ln(T)}{T}\right) = o(1).$

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

(iii)

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

(iv)

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

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

**Proof of Lemma 5.**

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

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

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

From equation (A86) in AA, we have

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

The last two terms are also $o(1)$ using results from Lemma 4(i) and (ii).

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

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

and the result follows from Markov's inequality.

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

(iv)

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

Now, for the variance, note that

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

For $t\ge s$,

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

However,

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

so that

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

Now, by the Cauchy–Schwarz inequality,

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

Then,

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

Hence,

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

so that $(iv)$ holds by Markov's inequality.

(v)

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

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

For completeness, we reproduce here a lemma from Okui (2009), which is essential to derive the higher-order expansion of the MSE. This lemma is similar to Lemma A1 of Donald and Newey (2001), but the expectation is unconditional. □

LEMMA 6 (Lemma 2 of Okui (2009)). *Let* $\rho_\alpha = trS(\alpha)$. *Suppose that an estimator* $\hat{\delta}$ *has a decomposition* $\sqrt{NT}\left(\hat{\delta}-\delta\right) = \widehat{H}^{-1}\hat{h}$, $\hat{h} = h + T^h + Z^h$, $\widehat{H} = H + T^H + Z^H$, $\left(h+T^h\right)\left(h+T^h\right)' - hh'H^{-1}T^{H\prime} - T^H H^{-1}hh' = \hat{A} + Z^A$, *such that* $T^h = o_p(1)$, $h = O_p(1)$, $H = O_p(1)$, *the determinant of* $H$ *is bounded away from zero with probability approaching 1,* $\rho_\alpha = o_p(1)$, $\left\|T^H\right\|^2 = o_p(\rho_\alpha)$, $\left\|T^h\right\|\left\|T^H\right\| = o_p(\rho_\alpha)$, $\left\|Z^h\right\| = o_p(\rho_\alpha)$, $\left\|Z^H\right\| = o_p(\rho_\alpha)$, $\left\|Z^A\right\| = o_p(\rho_\alpha)$, $E\left(\hat{A}\right) = \sigma^2 H + HS(\alpha)H + o_p(\rho_\alpha)$. *Then, the decomposition (4) holds for* $\hat{\delta}$.

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

## B.5. Proof of Proposition 2

Here, $S(\alpha)$ is a scalar so that $\rho_\alpha = S(\alpha)$. Notice that

$$\rho_\alpha \geq \frac{(1+\delta)^2}{NT}\left\{\sum_{t=1}^{T-1} E\left[tr\left(M_t^\alpha\right)\right]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2 \tag{13}$$

and

$$\rho_\alpha \geq \frac{(1-\delta^2)^2}{\sigma^2}\frac{1}{NT}\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t^\alpha)^2 w_{t-1}]. \tag{14}$$

First, we establish the rate of the RHS of (13). Because

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

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

$$\begin{aligned}&\frac{(1+\delta)^2}{NT}\left\{\sum_{t=1}^{T-1} E\left[tr\left(M_t^\alpha\right)\right]\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right)\right\}^2\\ &\leq \frac{4(1+\delta)^2}{NT(1-\delta)^2}\left(\sum_{t=1}^{T-1} E\left[tr\left(M_t^\alpha\right)\right]\right)^2 = \frac{4(1+\delta)^2}{NT(1-\delta)^2}\left(Etr[M^\alpha]\right)^2 = O(1/(\alpha^2 NT))\end{aligned}$$

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

Moreover, since

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

and by equation (A86) of AA, $\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]/NT = O(\ln(T)/T)$ so that $o(lnT/T) = o(\rho_\alpha)$ by inequality (14). To prove Proposition 2, we use Lemma 6 and

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

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

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

where

$$h = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \psi_t w'_{t-1} v^*_t,$$

$$T_1^h = -\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \psi_t w'_{t-1} (I - M_t^\alpha) v_t^* = O_p(\Delta_\alpha^{1/2}),$$

$$T_2^h = -\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} c_t \tilde{v}'_{tT} M_t^\alpha v_t^* = O_p\left(\frac{1}{\alpha\sqrt{NT}}\right),$$

where the rate for $T_2^h$ follows from Lemma 4(iii) and (iv). Moreover,

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

with

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

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

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

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

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

By $1/\sqrt{NT} = o(logT/T)$ and $1/(\alpha NT) = o(1/(\alpha^2 NT)$, we have that $Z_j^H$ are $o_p(\rho_\alpha)$ for $j = 1,2,3$ so that $\| \sum_{j=1}^3 Z_j^H \| = o_p(\rho_\alpha)$ by the triangular inequality.

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

We now apply Lemma 6 with $Z^A = 0$ and

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

Lemma 6 states that $S(\alpha)$ satisfies $E\left(\widehat{A}\right) = \sigma^2 H + HS(\alpha) H + o_p(\rho_\alpha)$. To calculate the expectation of $\widehat{A}$, we need to compute the expectation of each term. By the third moment condition and the independence assumption both on the error term $v_{it}$, we can show that $E(hT_2^h) = E(T_1^h T_2^h) = 0$. It can easily be proved that

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

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

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

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

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

And therefore,

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

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

## B.6. Preliminary Results for the Proof of Proposition 3

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

LEMMA 7.

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

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

**Proof of Lemma 7.**

(i)

$$tr[K_N^2] - tr[K^2] = ||K_N||^2 - ||K||^2 = (||K_N|| + ||K||)(||K_N|| - ||K||).$$

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

From Lemmas 1(ii) and 2(i), we have $||K_N|| + ||K|| = O_p(1)$. Moreover,

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

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

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

$$E\left[\left[\sum_{t=1}^{T-1}\left(tr(M_t^\alpha) - E[tr(M_t^\alpha)]\right)\right]^2\right]$$
$$= E\left[\left[tr(M^\alpha) - E[tr(M^\alpha)]\right]^2\right] = E\left[\left[tr(\left(Z'Z\right)^\alpha Z'Z) - E[tr(\left(Z'Z\right)^\alpha Z'Z)]\right]^2\right]$$
$$= E\left[\left[tr(K_N^\alpha K_N) - trE(K_N^\alpha K_N)\right]^2\right] = E\left[\left\|K_N^\alpha K_N - E(K_N^\alpha K_N)\right\|^2\right].$$

Moreover,

$$E\left[\left\|K_N^\alpha K_N - E(K_N^\alpha K_N)\right\|^2\right]$$
$$= E\left[\left\|K_N^\alpha (K_N - K) + \left(K_N^\alpha - E\left(K_N^\alpha\right)\right)K + E\left(K_N^\alpha (K - K_N)\right)\right\|^2\right]$$
$$\leq 3E\left\|K_N^\alpha (K_N - K)\right\|^2 + 3E\left\|\left(K_N^\alpha - E\left(K_N^\alpha\right)\right)K\right\|^2 + 3\left\|E\left(K_N^\alpha (K - K_N)\right)\right\|^2.$$

We have

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

The result follows. □

### B.7. Proof of Proposition 3

We want to prove that

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

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

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

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

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

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

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

For the first equality, we have

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

Moreover,

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

So it is sufficient to prove that

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

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

where $\nu_t = \frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}$ and $\widehat{\nu}_t = \frac{\widehat{\phi}_{T-t}}{T-t}-\frac{\widehat{\phi}_{T-t+1}}{T-t+1}$. We will use the following result (see Okui, 2009, p. 13): for a random sequence $\{a_k\}_k$, $\sum_k E(a_k^2) = o(1)$ implies that $\sup_k a_k = o_p(1)$.

$$E\left\{\left[\sum_{t=1}^{T-1} tr(M_t^\alpha)\left(\widehat{\nu}_t-\nu_t\right)\right]^2\right\} = E\sum_{t=1}^{T-1}\left(tr(M_t^\alpha)\right)^2\left(\widehat{\nu}_t-\nu_t\right)^2 + \sum_{t\neq s} E\left[tr\left(M_t^\alpha\right)\left(\widehat{\nu}_t-\nu_t\right)tr(M_s^\alpha)\left(\widehat{\nu}_s-\nu_s\right)\right]. \qquad \textbf{(15)}$$

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

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

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

By the Cauchy–Schwarz inequality, the second term of the RHS of (15) is also $O(1/N\alpha^2)$. By Lemma 7(ii),

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

We previously established that

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

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

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

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

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

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

Now, we want to prove that

$$\sup_{\mathcal{E}_T}\left|\frac{\widehat{R}(\alpha) - R(\alpha)}{R(\alpha)}\right| = o_p(1). \quad \textbf{(16)}$$

*We first consider the SC regularization scheme.*

In this case, $(I - M_t^{\alpha})^2 = (I - M_t^{\alpha})$ because $M_t^{\alpha}$ is a projection matric for this regularization so that

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

and

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

Let us define

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

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

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

Hence, to prove (16), we have to prove that

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

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

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

Noting that $w_{t-1}'(I - M_t^\alpha) w_{t-1} \geq w_{t-1}'(I - M_t) w_{t-1}$, we obtain

$$\begin{aligned}
&\sup_{\mathcal{E}_T} \left| \frac{\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}'(I - M_t^\alpha) w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t^\alpha) w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t^\alpha) w_{t-1}]} \right| \\
&\leq \sup_{\mathcal{E}_T} \left| \frac{\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}'(I - M_t^\alpha) w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t^\alpha) w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t) w_{t-1}]} \right| \\
&\leq \sup_{\mathcal{E}_T} \left| \frac{\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}' w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}' w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t) w_{t-1}]} \right| \\
&\quad + \sup_{\mathcal{E}_T} \left| \frac{\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 w_{t-1}' M_t^\alpha w_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}' M_t^\alpha w_{t-1}]}{\frac{1}{NT}\sum_{t=1}^{T-1} \psi_t^2 E[w_{t-1}'(I - M_t) w_{t-1}]} \right|. \qquad \textbf{(17)}
\end{aligned}$$

Now, we want to prove that

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

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

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

Moreover, from AA,

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

so that we can conclude that

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

We now turn our attention to (17). This term depends on $\alpha$. From Lemma 5($iii$), we have

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

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

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

Then, we conclude that

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

We now consider the proof of

$$\sup_{\mathcal{E}_T}\left|\frac{\sum_{t=1}^{T-1}\psi_t c_t w_{t-1}'M_t^{\alpha}\tilde{v}_{tT}/(NT)}{\sum_{t=1}^{T-1}\psi_t^2 E[w_{t-1}'(I-M_t)w_{t-1}]/(NT)}\right| = o_p(1). \quad \textbf{(18)}$$

We have

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

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

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

Then, we can conclude that (18) holds.

Now, we want to prove that

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

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

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

which is $o_p(1)$ under the assumption that $T/N \to 0$. This ends the proof of (16) for the SC regularization scheme.

*We now consider the LF regularization scheme.*

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

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

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

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

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

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

Noting that

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

we have to prove that

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

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

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

Since $\widetilde{M}_t^{\alpha}\leq 2M_t^{\alpha}\leq 2M_t$, we can apply the same strategy as in the case of SC regularization scheme provided that $\#\mathcal{E}_T=O(T^2)$ with $\#\mathcal{E}_T$ being the number of elements in the parameter set $\mathcal{E}_T$. Imposing that $\#\mathcal{E}_T=O(T^2)$ is a sufficient condition to have optimality in the LF 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 $\alpha$ is optimal under the assumption $\#\mathcal{E}_T=O(T^2)$ for the LF regularization scheme.

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

LEMMA 8. *If Assumptions 1', 2', and 3 are satisfied, then*
(i)

$$E\left[\left[\sum_{t=1}^{T-1}c_t\tilde{v}'_{tT}M_t^{\alpha}v_t^*\right]^2\right]$$

$$=\frac{\sigma^4}{(1-\delta)^2}E\left[\sum_{t=1}^{T-1}tr\left(M_t^{\alpha}\right)\left(\frac{\phi_{T-t}}{T-t}-\frac{\phi_{T-t+1}}{T-t+1}\right)\right]^2+o\left(\left(\sum_{t=1}^{T-1}tr\left(M_t^{\alpha}\right)\right)^2\right)$$

$$=O\left(\ln(T)/(\alpha NT)\right).$$

(ii)

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

(iii) *Let* $\Delta_\alpha$ *be defined as*

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

*Then,*

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

**Proof of Lemma 8.** (i) and (ii) These results can be established using proofs similar to those of Lemma 4(iii) and (iv).

(iii)

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

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

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

is finite by Assumption 3. Hence, the rate of $\Delta_\alpha$ follows. □

## B.8. Proof of Proposition 4

Let $\rho_\alpha = trace(S(\alpha))$. It follows from Lemma 8 that a term is $o_p(\rho_\alpha)$ if it is either $o_p\left(1/\left(\alpha^2 NT\right)\right)$ or $o_p(\alpha^\beta)$. Recall that $x_t$ is an $N\times(L_m+1)$ matrix with $x_t = \left(y_{t-1}, m_t\right) \equiv (u_t, m_t)$. $x_t^* = \left(u_t^*, m_t^*\right)$ with $u_t^* = w_{t-1} - c_t\widetilde{v}_{tT}$, $E_Z\left(u_t^*\right) = w_{t-1}$, $E_Z\left(x_t^*\right) = \left(w_{t-1}, m_t^*\right) \equiv \widetilde{w}_{t-1}$, and $x_t^* - E_Z\left(x_t^*\right) = (-c_t\widetilde{v}_{tT}, 0)$.

First, we note that

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

Adapting Okui (2009) to our setting, 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

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

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

Now, consider the denominator:

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} x_t^*$$

$$= \frac{1}{NT}\sum_{t=1}^{T-1} \left(x_t^* - E_Z\left(x_t^*\right)\right)' M_t^{\alpha} \left(x_t^* - E_Z\left(x_t^*\right)\right) \text{ (term } Z_4^H\text{)}$$

$$+ \frac{1}{NT}\sum_{t=1}^{T-1} E_Z\left(x_t^*\right)' M_t^{\alpha} E_Z\left(x_t^*\right)$$

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

$$= Z_3^H + Z_4^H + H$$

$$+ \frac{1}{NT}\sum_{t=1}^{T-1} E_Z\left(x_t^*\right)' E_Z\left(x_t^*\right) - H \text{ (term } Z_1^H\text{)}$$

$$- \left[\frac{1}{NT}\sum_{t=1}^{T-1}\left\{E_Z\left(x_t^*\right)'\left(I - M_t^{\alpha}\right) E_Z\left(x_t^*\right) - E\left[E_Z\left(x_t^*\right)'\left(I - M_t^{\alpha}\right) E_Z\left(x_t^*\right)\right]\right\}\right] \text{ (term } Z_2^H\text{)}$$

$$+ \frac{1}{NT}\sum_{t=1}^{T-1} E\left\{E_Z\left(x_t^*\right)'\left(I - M_t^{\alpha}\right) E_Z\left(x_t^*\right)\right\} \text{ (term } T^H\text{)}$$

so that

$$\frac{1}{NT}\sum_{t=1}^{T-1} x_t^{*'} M_t^{\alpha} x_t^* = H + T^H + \sum_{j=1}^{4} Z_j^H,$$

where

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

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

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

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

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

$$Z^H_{2,21} = \frac{1}{NT}\sum_{t=1}^{T-1}[m^{*'}_t(I-M^\alpha_t)w_{t-1} - E\{m^{*'}_t(I-M^\alpha_t)w_{t-1}\}],$$

$$Z^H_{2,12} = \frac{1}{NT}\sum_{t=1}^{T-1}[w'_{t-1}(I-M^\alpha_t)m^*_t - E\{w'_{t-1}(I-M^\alpha_t)m^*_t\}],$$

$$Z^H_{2,22} = \frac{1}{NT}\sum_{t=1}^{T-1}[m^{*'}_t(I-M^\alpha_t)m^*_t - E\{m^{*'}_t(I-M^\alpha_t)m^*_t\}],$$

$$Z^H_3 = -2\begin{bmatrix} \frac{1}{NT}\sum_{t=1}^{T-1} c_t w'_{t-1} M^\alpha_t \tilde{v}_{tT} & 0 \\ \frac{1}{NT}\sum_{t=1}^{T-1} c_t m^{*'}_t M^\alpha_t \tilde{v}_{tT} & 0 \end{bmatrix},$$
$$Z^H_4 = \begin{bmatrix} \frac{1}{NT}\sum_{t=1}^{T-1} c^2_t \tilde{v}'_{tT} M^\alpha_t \tilde{v}_{tT} & 0 \\ 0 & 0 \end{bmatrix},$$

and

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

The terms $h$, $H$, and $Z^H_1$ do not depend on the matrix $M^\alpha_t$ so that we can use their order given in Okui (2009). We then have that $H = O(1)$, $h = o_p(1)$, and $Z^H_1 = O_p(1/\sqrt{NT}) = o(\rho_\alpha)$ provided that either $\alpha^\beta\sqrt{NT} \to \infty$ or $\alpha^2\sqrt{NT} \to 0$.

Using $\tilde{w}_{t-1} = (w_{t-1}, m^*_t)$, the term $T^h_1$ can be rewritten as

$$T^h_1 = \frac{1}{NT}\sum_{t=1}^{T-1}\tilde{w}'_{t-1}(I-M^\alpha_t)v^*_t.$$

We have then $E(T^h_1) = 0$ and $V(T^h_1) = \sigma^2\Delta_\alpha$ so that $T^h_1 = O_p(\Delta^{1/2}_\alpha)$ by Markov's inequality. Since from Lemma 8, we have $\Delta_\alpha = O_p(\alpha^\beta)$, we can conclude that $T^h_1 = o_p(1)$.

Regarding the term $T^h_2$, we have from Lemma 8 that $E\left(T^h_2\right) = O\left(1/\left(\alpha\sqrt{NT}\right)\right)$ and $V\left(T^h_2\right) = O\left((\ln T)^2/N\right)$ so that $T^h_2 = o_p(1)$ provided that $\alpha\sqrt{NT} \to \infty$.

Next, we consider $T^H$:

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

We now look at the term $Z^H_2$. In the same way as in the model without covariates, we can prove that diagonal elements of $Z^H_2$ are $O_p(1/\sqrt{NT})$. For the other terms, we have

$$Z^H_{2,21} = \frac{1}{NT}\sum_{t=1}^{T-1}[m^{*'}_t(I-M^\alpha_t)\tilde{w}_{t-1} - E\{m^{*'}_t(I-M^\alpha_t)\tilde{w}_{t-1}\}].$$

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

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

$$= \frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}\tilde{w}_{t-1} - E\{m_t^{*',k}\tilde{w}_{t-1}\}] \quad \textbf{(19)}$$

$$-\frac{1}{NT}\sum_{t=1}^{T-1}[m_t^{*',k}M_t^{\alpha}\tilde{w}_{t-1} - E\{m_t^{*',k}M_t^{\alpha}\tilde{w}_{t-1}\}]. \quad \textbf{(20)}$$

First, consider the term (19), its expectation is equal to zero, and its variance is

$$V\left(\frac{1}{NT}\sum_{t=1}^{T-1}m_t^{*',k}\tilde{w}_{t-1}\right) = \frac{1}{N^2}V\left(\frac{1}{T}\sum_{t=1}^{T-1}\sum_{i=1}^{N}m_{it}^{*',k}\tilde{w}_{it-1}\right)$$

$$= \frac{1}{N}V\left(\frac{1}{T}\sum_{t=1}^{T-1}m_{it}^{*',k}\tilde{w}_{it-1}\right) = \frac{1}{N}O\left(\frac{1}{T}\right).$$

Hence, by Markov's theorem,

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

Concerning the term (20), its variance is bounded for all $\alpha$, and thus it is enough to prove the result for $\alpha = 0$. Using proofs similar to before, we can show that it is $O_p(1/\sqrt{NT})$. We have just established that elements $Z_{21}^{H,k}$ are $O_p(1/\sqrt{NT})$. The same strategy can be applied to the nondiagonal elements of the $l_m$-dimensional matrix $Z_{22}^H$, allowing us to conclude that $Z_2^H = O_p(1/\sqrt{NT})$ so that the $Z_2^H = o_p(\rho_\alpha)$ provided that $\alpha^\beta\sqrt{NT} \to \infty$.

For the term $Z_3^H$, we note that

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

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

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

so that

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

Now, note that the other terms of $Z_3^H$ take the form $\frac{1}{NT}\sum_{t=1}^{T-1}c_t m_t^{l*'}M_t^{\alpha}\tilde{v}_{tT}$, $l = 1,\ldots,l_m$. Hence, by conditioning on the instruments, we can prove in the same way that those terms are $o_p(\rho_\alpha)$.

https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

For $Z_4^H$, following the same strategy as in the model without covariates, we have $\frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 \tilde{v}'_{tT} M_t^\alpha \tilde{v}_{tT} = O_p(\frac{1}{\alpha NT})$. Hence, $Z_4^H = o_p(\rho_\alpha)$.

We now apply Lemma 6. Let us define $Z^A = 0$ and

$$\hat{A} = (h + T_1^h + T_2^h)(h + T_1^h + T_2^h)' - hh'H^{-1}T^H - T^H H^{-1}hh'.$$

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 $E(hT_2^{h'}) = E(T_2^h h') = E(T_1^h T_2^{h'}) = E(T_2^h T_1^{h'})$.

It can easily be proved that $E(hh') = \sigma^2 H$ and $E\{hT_1^{h'}\} = E\{T_1^h h'\} = E(hh'H^{-1}T^H) = E(T^H H^{-1}hh') = \sigma^2 T^H$. Given these equalities,

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

and

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

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

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

with

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

## *REFERENCES*


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

Arellano, M. & S. Bond (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* 58, 277–297.

Arellano, M. & O. Bover (1995) Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68(1), 29–51.

Belloni, A., D. Chen, V. Chernozhukov, & C. Hansen (2012) Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80, 2369–2429.

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

Blundell, R. & S. Bond (2000) GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews* 19, 321–340.

Bun, M.J. & J.F. Kiviet (2006) The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132(2), 409–444.

Bun, M.J.G. & F. Windmeijer (2010) The weak instrument problem of the system GMM estimator in dynamic panel data models. *The Econometrics Journal* 13(1), 95–126.

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


https://doi.org/10.1017/S0266466622000469 Published online by Cambridge University Press

Carrasco, M., M. Chernov, J.-P. Florens, & E. Ghysels (2007a) Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics* 140(2), 529–573.

Carrasco, M., J.-P. Florens, & E. Renault (2007b) Chapter 17: Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In *Handbook of Econometrics*, pp. 5633–5751. Elsevier.

Carrasco, M. & G. Tchuente (2015) Regularized LIML for many instruments. *Journal of Econometrics* 186(2), 427–442.

Donald, S.G. & W.K. Newey (2001) Choosing the number of instruments. *Econometrica* 69(5), 1161–1191.

Doran, H.E. & P. Schmidt (2006) 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.

Groetsch, C. (1993) *Inverse Problems in the Mathematical Sciences*. Theory and Practice of Applied Geophysics Series. Vieweg.

Hahn, J., & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effect when both $n$ and $T$ are large. *Econometrica* 70(4), 1639–1657.

Hahn, J., J. Hausman, & G. Kuersteiner (2004) Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *The Econometrics Journal* 7, 272–306.

Hansen, B.E. (2007) Least squares model averaging. *Econometrica* 75(4), 1175–1189.

Hayakawa, K. (2009) A simple efficient instrumental variable estimator for panel AR($p$) models when both $N$ and $T$ are large. *Econometric Theory* 25, 873–890.

Hayakawa, K., M. Qi, & J. Breitung (2019) Double filter instrumental variable estimation of panel data models with weakly exogenous variables. *Econometric Reviews* 38(9), 1055–1088.

Hirano, K. (2002) Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica* 70(2), 781–799.

Hsiao, C. & Q. Zhou (2018) JIVE for panel dynamic simultaneous equations models. *Econometric Theory* 34(6), 1325–1369.

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

Kress, R. (1999) *Linear Integral Equations*. Springer.

Kuersteiner, G. (2012) Kernel-weighted GMM estimators for linear time series models. *Journal of Econometrics* 170(2), 399–421.

Kuersteiner, G. & R. Okui (2010) Constructing optimal instruments by first-stage prediction averaging. *Econometrica* 78, 697–718.

Levine, R., N. Loayza, & T. Beck (2000) Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics* 46(1), 31–77.

Li, K.-C. (1986) Asymptotic optimality of CL and generalized cross-validation in ridge regression with application to spline smoothing. *Annals of Statistics* 14(3), 1101 – 1112.

Li, K.-C. (1987) Asymptotic optimality for CP, CL, cross-validation and generalized cross-validation: Discrete index set. *Annals of Statistics* 15(3), 958–975.

Mitze, T. (2012) *Empirical Modelling in Regional Science*. Lecture Notes in Economics and Mathematical Systems. Springer.

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

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

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