cemmap
centre for microdata methods and practice

Dynamic linear panel regression models with interactive fixed effects

Hyungsik Roger Moon
Martin Weidner

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP63/13

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Dynamic Linear Panel Regression Models with Interactive Fixed Effects*

Hyungsik Roger Moon‡ Martin Weidner§

December 25, 2013

Abstract

We analyze linear panel regression models with interactive fixed effects and predetermined regressors, e.g. lagged-dependent variables. The first order asymptotic theory of the least squares (LS) estimator of the regression coefficients is worked out in the limit where both the cross sectional dimension and the number of time periods become large. We find that there are two sources of asymptotic bias of the LS estimator: bias due to correlation or heteroscedasticity of the idiosyncratic error term, and bias due to predetermined (as opposed to strictly exogenous) regressors. We provide an estimator for the bias and a bias corrected LS estimator for the case where idiosyncratic errors are independent across both panel dimensions. Furthermore, we provide bias corrected versions of the three classical test statistics (Wald, LR and LM test) and show that their asymptotic distribution is a χ^2-distribution. Monte Carlo simulations show that the bias correction of the LS estimator and of the test statistics also work well for finite sample sizes.

1 Introduction

In this paper we study a linear panel regression model where the individual fixed effects λ_i, called factor loadings, interact with common time specific effects f_t, called factors. This interactive fixed effect specification contains the conventional fixed effects and time-specific effects as special cases, but is significantly more flexible since it allows the factors f_t to affect each individual with a different loading λ_i.

*This paper is based on an unpublished manuscript of the authors which was circulated under the title "Likelihood Expansion for Panel Regression Models with Factors", but is now completely assimilated by the current paper and Moon and Weidner (2013). We greatly appreciate comments from the participants in the Far Eastern Meeting of the Econometric Society 2008, the SITE 2008 Conference, the All-UC-Econometrics Conference 2008, the July 2008 Conference in Honour of Peter Phillips in Singapore, the International Panel Data Conference 2009, the North American Summer Meeting of the Econometric Society 2009, and from seminar participants at Penn State, UCLA, and USC. We are also grateful for the comments and suggestions of Guido Kuersteiner and three anonymous referees. Moon is grateful for the financial support from the NSF via grant SES 0920903 and the faculty development award from USC. Weidner acknowledges support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0002.

‡Department of Economics, University of Southern California, Los Angeles, CA 90089-0253. Email: moonr@usc.edu. Department of Economics, Yonsei University, Seoul, Korea.

§Department of Economics, University College London, Gower Street, London WC1E 6BT, U.K., and CeMMaP. Email: m.weidner@ucl.ac.uk.

Factor models have been widely studied in various economics disciplines, for example in asset pricing, empirical macro, forecasting, and empirical labor economics.[1] In the panel literature, factor models are often used to represent time varying individual effects (or heterogenous time effects), so called interactive fixed effects. For panels with a large cross sectional dimension (N) but a short time dimension (T), Holtz-Eakin, Newey, and Rosen (1988) (hereafter HNR) study a linear panel regression model with interactive fixed effects and lagged dependent variables. To solve the incidental parameter problem caused by the λ_i's, they estimate a quasi-differenced version of the model using appropriate lagged variables as instruments, and treating f_t's as a fixed number of parameters to estimate. Ahn, Lee and Schmidt (2001) also consider large N but short T panels. Instead of eliminating the individual effects λ_i by transforming the panel data, they impose various second moment restrictions including the correlated random effects λ_i, and derive moment conditions to estimate the regression coefficients. More recent literature considers panels with comparable size of N and T. The interactive fixed effect panel regression model of Pesaran (2006) allows heterogenous regression coefficients. Pesaran's estimator is the common correlated effect (CCE) estimator that uses the cross sectional averages of the dependant variable and the independent variables as control functions for the interactive fixed effects.[2]

Among the interactive fixed effect panel literature, most closely related to our paper is Bai (2009). Bai assumes that the regressors are *strictly* exogenous and the number of factors is known. The estimator that he investigates is the least squares (LS) estimator, which minimizes the sum of squared residuals of the model jointly over the regression coefficients and the fixed effect parameters λ_i and f_t.[3] Using the alternative asymptotics where $N, T \to \infty$ at the same rate,[4] Bai shows that the LS estimator is $\sqrt{NT}$-consistent and asymptotically normal, but may have an asymptotic bias. The bias in the normal limiting distribution occurs when the regression errors are correlated or heteroscedastic. Bai also shows how to estimate the bias and proposes a bias corrected estimator.

Following the methodology in Bai (2009), we investigate the LS estimator for a linear panel regression with a known number of interactive fixed effects. The main difference from Bai is that we consider *predetermined* regressors, thus allowing feedback of past outcomes to future regressors. One of the main findings of the present paper is that under the alternative asymptotics, the limit distribution of the LS estimator has two types of biases, one type of bias due to correlated or heteroscedastic errors (the same bias as in Bai) and the other type of bias due to the predetermined regressors. This additional bias term is analogous to the incidental parameter bias of Nickell (1981) in finite T and the bias in Hahn and Kuersteiner (2002) in large T.

In addition to allowing for predetermined regressors, we also extend Bai's results to models where both "low-rank regressor" (e.g. time-invariant and common regressors, or interactions of those two) and "high-rank-regressor" (almost all other regressors that vary across individuals and

[1] See, e.g., Chamberlain and Rothschild (1983), Ross (Ross, 1976), and Fama and French (1993) for asset pricing, Bernanke, Boivin and Eliasz (2005) for empirical macro, Stock and Watson (2002) and Bai and Ng (2006) for forecasting, and Holtz-Eakin, Newey, and Rosen (1988) for empirical labor economics.

[2] The theory of the CCE estimator was further developed in e.g. Harding and Lamarche (2009; 2011), Kapetanios, Pesaran and Yamagata (2011), Pesaran and Tosetti (2011), Chudik, Pesaran and Tosetti (2011), and Chudik and Pesaran (2013).

[3] The LS estimator is sometimes called "concentrated" least squares estimator in the literature, and in an earlier version of the paper we referred to it as the "Gaussian Quasi Maximum Likelihood Estimator", since LS estimation is equivalent to maximizing a conditional Gaussian likelihood function.

[4] This alternative asymptotics is known to be a convenient tool in the fixed effect panel literature to characterize the asymptotic bias due to incidental parameter problems. See, e.g., Hahn and Kuersteiner (2002; 2011), Alvarez and Arellano (2003), Hahn and Newey (2004), and Hahn and Moon (2006).

over time) are present simultaneously, while Bai (2009) only considers the "low-rank regressor" separately and in a restrictive setting (in particular not allowing for regressors that are obtained by interacting time-invariant and common variables). A general treatment of "low-rank regressors" is desirable since they often occur in applied work, e.g., Gobillon and Magnac (2013). The analysis of those regressors is challenging, however, since the unobserved interactive fixed effects also represent a low-rank $N \times T$ matrix, thus posing a non-trivial identification problem for low-rank regressors, which needs to be addressed. We provide conditions under which the different type of regressors are identified jointly and under which they can be estimated consistently as N and T grow large.

Another contribution of this paper is to establish the asymptotic theory of the three classical test statistics (Wald test, LR test, and LM (or score) test) for testing restrictions on the regression coefficients in a large N, T panel framework.[5] Regarding testing for coefficient restrictions, Bai (2009) investigates the Wald test based on the bias corrected LS estimator, and HNR consider the LR test in their 2SLS estimation framework with fixed T.[6] What we show is that the conventional LR and LM test statistics based on the LS profile objective function have non-central chi-square limit due to incidental parameters in the interactive fixed effects. We therefore propose modified LR and LM tests whose asymptotic distributions are conventional chi-square distributions.

In order to establish the asymptotic theories of the LS estimator and the three classical tests, we use the quadratic approximation of the profile LS objective function that was derived in Moon and Weidner (2013). This method is different from Bai (2009), who uses the first order condition of the LS optimization problem as the starting point of his analysis. One advantage of our methodology is that it can also directly be applied to derive the asymptotic properties of the LR and LM test statistics.

In this paper, we assume that the regressors are not endogenous and the number of factors is known, which might be restrictive in some applications. In other papers we study how to relax these restrictions. Moon and Weidner (2013) investigates the asymptotic properties of the LS estimator of the linear panel regression model with factors when the number of factors is unknown and extra factors are included unnecessarily in the estimation. It turns out that under suitable conditions the limit distribution of the LS estimator is unchanged when the number of factors is overestimated. Moon and Weidner (2013) is complementary to the current paper, since there we do not derive the limiting distribution of the estimator, do not correct for the bias, and also do not consider low-rank regressors or testing problems. The extension to allow endogenous regressors is closely related with the result in Moon, Shum and Weidner (2012) (hereafter MSW). MSW's main purpose is to extended the random coefficient multinomial logit demand model (known as the BLP demand model from Berry, Levinsohn and Pakes (1995)) by allowing for interactive product and market specific fixed effects. Although the main model of interest is quite different from the linear panel regression model of the current paper, MSW's econometrics framework is directly applicable to the model of the current paper with endogenous regressors. In Section 6, we briefly discuss how to apply the estimation method of MSW in the current framework with endogenous regressors.[7]

[5] The "likelihood ratio" and the score used in the tests are based on the LS objective function, which can be interpreted as the (misspecified) conditional Gaussian likelihood function.

[6] Another type of widely studied tests in the interactive fixed effect panel literature are panel unit root test, e.g., Bai and Ng (2004), Moon and Perron (2004), and Phillips and Sul (2003).

[7] Lee, Moon, and Weidner (2012) also apply the MSW estimation method to estimate a simple dynamic panel regression with interactive fixed effect and classical measurement errors.

Comparing the different estimation approaches for interactive fixed effect panel regressions proposed in the literature, it seems fair to say that the LS estimator in Bai (2009) and our paper, the CCE estimator of Pesaran (2006), and the IV estimator based on quasi-differencing in HNR, all have their own relative advantages and disadvantages. These three estimation methods handle the interactive fixed effects quite differently. The LS method concentrates out the interactive fixed effects by taking out the principal components. The CCE method controls the factor (or time effects) using the cross sectional averages of the dependent and independent variables. The NHR's approach quasi-differences out the individual effects, treating the remaining time effects as parameters to estimate. The IV estimator of HNR should work well when T is short, but should be expected to also suffer from an incidental parameter problem due to estimation of the factors when T becomes large. Pesaran's CCE estimation method does not require the number of factors to be known and does not require the strong factor assumption that we will impose below, but in order for the CCE to work, not only the DGPs for the dependent variable (e.g., the regression model) but also the DGP of the explanatory variables should be restricted in a way that their cross sectional average can control the unobserved factors. The LS estimator and its bias corrected version perform well under relatively weak restrictions on the regressors, but it requires that T should not be too small and the factors should be sufficiently strong to be correctly picked up as the leading principal components.

The paper is organized as follows. In Section 2 we introduce the interactive fixed effect model and provide conditions for identifying the regression coefficients in the presence of the interactive fixed effects. In Section 3 we define the LS estimator of the regression parameters and provide a set of assumptions that are sufficient to show consistency of the LS estimator. In Section 4 we work out the asymptotic distribution of the LS estimator under the alternative asymptotic. We also provide a consistent estimator for the asymptotic bias and a bias corrected LS estimator. In Section 5 we consider the Wald, LR and LM tests for testing restrictions on the regression coefficients of the model. We present bias corrected versions of these tests and show that they have chi-square limiting distribution. In Section 6 we briefly discuss how to estimate the interactive fixed effect linear panel regression when the regressors are endogenous. In Section 7 we present Monte Carlo simulation results for an AR(1) model with interactive fixed effect. The simulations show that the LS estimator for the AR(1) coefficient is biased, and that the tests based on it can have severe size distortions and power asymmetries, while the bias corrected LS estimator and test statistics have better properties. We conclude in Section 8. All proofs of theorems and some technical details are presented in the appendix.

A few words on notation. For a column vector v the Euclidean norm is defined by $\|v\| = \sqrt{v'v}$. For the n-th largest eigenvalues (counting multiple eigenvalues multiple times) of a symmetric matrix B we write $\mu_n(B)$. For an $m \times n$ matrix A the Frobenius norm is $\|A\|_F = \sqrt{\text{Tr}(AA')}$, and the spectral norm is $\|A\| = \max_{0 \neq v \in \mathbb{R}^n} \frac{\|Av\|}{\|v\|}$, or equivalently $\|A\| = \sqrt{\mu_1(A'A)}$. Furthermore, we define $P_A = A(A'A)^{-1}A'$ and $M_A = \mathbb{I} - A(A'A)^{-1}A'$, where $\mathbb{I}$ is the $m \times m$ identity matrix, and $(A'A)^{-1}$ may be a pseudo-inverse in case A is not of full column rank. For square matrices B, C, we write $B > C$ (or $B \geq C$) to indicate that $B - C$ is positive (semi) definite. For a positive definite symmetric matrix A we write $A^{1/2}$ and $A^{-1/2}$ for the unique symmetric matrices that satisfy $A^{1/2}A^{1/2} = A$ and $A^{-1/2}A^{-1/2} = A^{-1}$. We use ∇ for the gradient of a function, i.e. $\nabla f(x)$ is the row vector of partial derivatives of f with respect to each component of x. We use "wpa1" for "with probability approaching one".

2 Model and Identification

We study the following panel regression model with cross-sectional size N, and T time periods,

$$Y_{it} = \beta^{0\prime} X_{it} + \lambda_i^{0\prime} f_t^0 + e_{it}, \qquad i = 1 \dots N, \quad t = 1 \dots T, \tag{2.1}$$

where X_{it} is a $K \times 1$ vector of observable regressors, β^0 is a $K \times 1$ vector of regression coefficients, λ_i^0 is an $R \times 1$ vector of unobserved factor loadings, f_t^0 is an $R \times 1$ vector of unobserved common factors, and e_{it} are unobserved errors. The superscript zero indicates the true parameters. We write f_{tr}^0 and λ_{ir}^0, where $r = 1, \dots, R$, for the components of λ_i^0 and f_t^0, respectively. R is the number of factors. Note that we can have $f_{tr}^0 = 1$ for all t and a particular r, in which case the corresponding λ_{ir}^0 become standard individual specific effects. Analogously we can have $\lambda_{ir}^0 = 1$ for all i and a particular r, so that the corresponding f_{tr}^0 become standard time specific effects.

Throughout this paper we assume that the true number of factors R is known.[8] We introduce the notation $\beta^0 \cdot X \equiv \sum_{k=1}^K \beta_k^0 X_k$. In matrix notation the model can then be written as

$$Y = \beta^0 \cdot X \; + \; \lambda^0 f^{0\prime} + e \; ,$$

where Y, X_k and e are $N \times T$ matrices, λ^0 is an $N \times R$ matrix, and f^0 is a $T \times R$ matrix. The elements of X_k are denoted by $X_{k,it}$.

We separate the K regressors into K_1 "low-rank regressors" X_l, $l = 1, \dots, K_1$, and $K_2 = K - K_1$ "high-rank regressors" X_m, $m = K_1 + 1, \dots, K$. Each low-rank regressor $l = 1, \dots, L$ is assumed to satisfy $\text{rank}(X_l) = 1$. This implies that we can write $X_l = w_l v_l'$, where w_l is an N-vector and v_l is a T-vector, and we also define the $N \times K_1$ matrix $w = (w_1, \dots, w_{K_1})$ and the $T \times K_1$ matrix $v = (v_1, \dots, v_{K_1})$.

Let $l = 1, \dots, K_1$. The two most prominent types of low-rank regressors are time-invariant regressors, which satisfy $X_{l,it} = Z_i$ for all i, t, and common (or cross-sectionally invariant) regressors, in which case $X_{l,it} = W_t$ for all i, t. Here, Z_i and W_t are some observed variables, which only vary over i or t, respectively. A more general low-rank regressor can be obtained by interacting Z_i and W_t multiplicatively, i.e. $X_{l,it} = Z_i W_t$, an empirical example of which is given in Gobillon and Magnac (2013). In these examples, and probably for the vast majority of applications, the low-rank regressors all satisfy $\text{rank}(X_l) = 1$, but our results can easily be extended to more general low-rank regressors.[9]

High-rank regressor are those whose distribution guarantees that they have high rank (usually full rank) when considered as an $N \times T$ matrix. For example, a regressor whose entries satisfy $X_{m,it} \sim iid\mathcal{N}(\mu, \sigma)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$, satisfies $\text{rank}(X_m) = \min(N, T)$ with probability one.

This separation of the regressors into low- and high-rank regressors is important to formulate our assumptions for identification and consistency, but actually plays no role in the estimation and inference procedures for $\widehat{\beta}$ discussed below.

[8] To remove this restriction, one could estimate R consistently in the presence of the regressors. In the literature so far, however, consistent estimation procedures for R are established mostly in pure factor models (e.g., Bai and Ng (2002), Onatski (2005) and Harding (2007)). Alternatively, one could rely on Moon and Weidner (2013) who consider a regression model with interactive fixed effects when only an upper bound on the number of factors is known — but it is mathematically very challenging to extend those results to the more general setup considered here.

[9] If we have low-rank regressors with rank larger than one, then we write $X_l = w_l v_l'$, where w_l is an $N \times \text{rank}(X_l)$ matrix and v_l is a $T \times \text{rank}(X_l)$ matrix, and we define $w = (w_1, \dots, w_{K_1})$ as a $N \times \sum_{l=1}^L \text{rank}(X_l)$ matrix, and $v = (v_1, \dots, v_{K_1})$ ae a $T \times \sum_{l=1}^L \text{rank}(X_l)$ matrix. All our results would then be unchanged, as long as $\text{rank}(X_l)$ is a finite constant for all $l = 1, \dots, K_1$, and we replace $2R + K_1$ by $2R + \text{rank}\,(w)$ in Assumption ID(v) and Assumption 4(ii)(a).

Assumption ID (Assumptions for Identification).

(i) **Existence of Second Moments:**
The second moments of $X_{k,it}$ and e_{it} conditional on λ^0, f^0, w exist for all i, t, k.

(ii) **Mean Zero Errors and Exogeneity:**
$\mathbb{E}\left(e_{it}|\lambda^0, f^0, w\right) = 0$, $\mathbb{E}(X_{k,it}e_{it}|\lambda^0, f^0, w) = 0$, *for all i, t, k.*

The following two assumptions only need to be imposed if $K_1 > 0$, i.e. if low-rank regressors are present:

(iii) **Non-Collinearity of Low-Rank Regressors:**
Consider linear combinations $\alpha \cdot X_{\text{low}} \equiv \sum_{l=1}^{K_1} \alpha_l X_l$ of the low-rank regressors X_l with $\alpha \in \mathbb{R}^{K_1}$. For all $\alpha \neq 0$ we assume that

$$\mathbb{E}\left[(\alpha \cdot X_{\text{low}}) M_{f^0} (\alpha \cdot X_{\text{low}})' \big| \lambda^0, f^0, w\right] \neq 0.$$

(iv) **No Collinearity between Factor Loadings and Low-Rank Regressors:**
$\text{rank}(M_w \lambda^0) = \text{rank}(\lambda^0)$.[10]

The following assumption only needs to be imposed if $K_2 > 0$, i.e. if high-rank regressors are present:

(v) **Non-Collinearity of High-Rank Regressors:**
Consider linear combinations $\alpha \cdot X_{\text{high}} \equiv \sum_{m=K_1+1}^{K} \alpha_m X_m$ of the high-rank regressors X_m for $\alpha \in \mathbb{R}^{K_2}$.[11] *For all $\alpha \neq 0$ we assume that*

$$\text{rank}\left\{\mathbb{E}\left[(\alpha \cdot X_{\text{high}})(\alpha \cdot X_{\text{high}})' \big| \lambda^0, f^0, w\right]\right\} > 2R + K_1.$$

All expectations in the assumptions are conditional on λ^0, f^0, and w, in particular e_{it} is not allowed to be correlated with λ^0, f^0, and w. However, e_{it} is allowed to be correlated with v (i.e. predetermined low-rank regressors are allowed). If desired, one can interchange the role of N and T in the assumptions, by using the formal symmetry of the model under exchange of the panel dimensions ($N \leftrightarrow T$, $\lambda^0 \leftrightarrow f^0$, $Y \leftrightarrow Y'$, $X_k \leftrightarrow X_k'$, $w \leftrightarrow v$).

Assumptions ID(i) and (ii) have standard interpretations, but the other assumptions require some further discussion.

Assumption ID(iii) states that the low-rank regressors are non-collinear even after projecting out all variation that is explained by the true factors f^0. This would, for example, be violated if $v_l = f_r^0$ for some $l = 1, \ldots, K_1$ and $r = 1, \ldots, R$, since then $X_l M_{f^0} = 0$ and we can choose α such that $X_{\text{low}} = X_l$. Similarly, Assumption ID(iv) rules out, for example, that $w_l = \lambda_r^0$ for some $l = 1, \ldots, K_1$ and $r = 1, \ldots, R$, since then $\text{rank}(M_w \lambda^0) < \text{rank}(\lambda^0)$, in general. It ought to be expected that λ^0 and f^0 have to feature in the identification conditions for the low-rank regressors, since the interactive fixed effects structure and the low-rank regressors represent similar types of low-rank $N \times T$ structures.

[10] Note that $\text{rank}(\lambda^0) = R$ if R factors are present. Our identification results are consistent with the possibility that $\text{rank}(\lambda^0) < R$, i.e. that R only represents an upper bound on the number of factors, but later we assume $\text{rank}(\lambda^0) = R$ to show consistency.

[11] The components of the K_2-vector α are denoted by α_{K_1+1} to α_K.

Assumption ID(v) would be a standard non-collinearity assumption if it would impose $\text{rank}\left\{\mathbb{E}\left[(\alpha \cdot X_{\text{high}})(\alpha \cdot X_{\text{high}})'\middle|\lambda^0, f^0, w\right]\right\} > 0$, which is equivalent to demanding that the $N \times N$ matrix $\mathbb{E}\left[(\alpha \cdot X_{\text{high}})(\alpha \cdot X_{\text{high}})'\middle|\lambda^0, f^0, w\right]$ is non-zero for all $\alpha \in \mathbb{R}^{K_2}$. The assumption strengthens this standard non-collinearity assumption by imposing the rank of this $N \times N$ matrix to be larger than $2R + K_1$, thus guaranteeing that any linear combination $\alpha \cdot X_{\text{high}}$ is sufficiently different from the low-rank regressors and from the interactive fixed effects. This also explain the name "high-rank regressors" since their rank has to be sufficiently large in order to satisfy Assumption ID(v). Note also that only the number of factors R, but not λ^0 and f^0 itself feature in Assumption ID(v).

Theorem 2.1 (**Identification**). *Suppose that the Assumptions ID are satisfied. Then, the minima of the expected objective function* $\mathbb{E}\left(\|Y - \beta \cdot X - \lambda f'\|_F^2 \middle| \lambda^0, f^0, w\right)$ *over* $(\beta, \lambda, f) \in \mathbb{R}^{K+N\times R+T\times R}$ *satisfy* $\beta = \beta^0$ *and* $\lambda f' = \lambda^0 f^{0\prime}$. *This shows that* β^0 *and* $\lambda^0 f^{0\prime}$ *are identified.*

The theorem shows that the true parameters are identified as minima of the expected value of $\|Y - \beta \cdot X - \lambda f'\|_F^2 = \sum_{i,t}(Y_{it}\beta' - X_{it} - \lambda_i' f_t)^2$, which is the sum of squared residuals. The same objective function is used to define the estimators $\widehat{\beta}$, $\widehat{\lambda}$ and $\widehat{f}$ below. Without further normalization conditions the parameters λ^0 and f^0 are not separately identified, because the outcome variable Y is invariant under transformations $\lambda^0 \to^0 \lambda A'$ and $f^0 \to f^0 A^{-1}$, where A is a non-singular $R \times R$ matrix. However, the product $\lambda^0 f^{0\prime}$ is uniquely identified according to the theorem. Since our focus is on identification and estimation of β^0, there is no need to discuss those additional normalization conditions for λ^0 and f^0 in this paper.

3 Estimator and Consistency

The objective function of the model is simply the sum of squared residuals, which in matrix notation can be expressed as

$$\begin{aligned}\mathcal{L}_{NT}(\beta, \lambda, f) &= \frac{1}{NT}\left\|Y - \beta \cdot X - \lambda f'\right\|_F^2 \\ &= \frac{1}{NT}\,\text{Tr}\left[\left(Y - \beta \cdot X - \lambda f'\right)'\left(Y - \beta \cdot X - \lambda f'\right)\right] . \end{aligned} \tag{3.1}$$

The estimator we consider is the LS estimator that jointly minimizes $\mathcal{L}_{NT}(\beta, \lambda, f)$ over β, λ and f. Our main object of interest are the regression parameters $\beta = (\beta_1, ..., \beta_K)'$, whose estimator is given by

$$\widehat{\beta} = \underset{\beta \in \mathbb{B}}{\text{argmin}}\; L_{NT}(\beta) \;, \tag{3.2}$$

where $\mathbb{B} \subset \mathbb{R}^K$ is a compact parameter set that contains the true parameter, i.e. $\beta^0 \in \mathbb{B}$, and the objective function is the profile objective function

$$\begin{aligned} L_{NT}(\beta) &= \min_{\lambda, f}\; \mathcal{L}_{NT}(\beta, \lambda, f) \\ &= \min_{f}\; \frac{1}{NT}\,\text{Tr}\left[(Y - \beta \cdot X)\, M_f\, (Y - \beta \cdot X)'\right] \\ &= \frac{1}{NT}\sum_{r=R+1}^{T} \mu_r\left[(Y - \beta \cdot X)'(Y - \beta \cdot X)\right] . \end{aligned} \tag{3.3}$$

Here, the first expression for $L_{NT}(\beta)$ is its definition as the the minimum value of $\mathcal{L}_{NT}(\beta,\lambda,f)$ over λ and f. We denote the minimizing incidental parameters by $\widehat{\lambda}(\beta)$ and $\widehat{f}(\beta)$, and we define the estimators $\widehat{\lambda}=\widehat{\lambda}(\widehat{\beta})$ and $\widehat{f}=\widehat{f}(\widehat{\beta})$. Those minimizing incidental parameters are not uniquely determined – for the same reason that λ^0 and f^0 are non uniquely identified –, but the product $\widehat{\lambda}(\beta)\widehat{f}'(\beta)$ is unique.

The second expression for $L_{NT}(\beta)$ in equation (3.3) is obtained by concentrating out λ (analogously, one can concentrate out f to obtain a formulation where only the parameter λ remains). The optimal f in the second expression is given by the R eigenvectors that correspond to the R largest eigenvalues of the $T\times T$ matrix $(Y-\beta\cdot X)'(Y-\beta\cdot X)$. This leads to the third line that presents the profile objective function as the sum over the $T-R$ smallest eigenvalues of this $T\times T$ matrix. Theorem C.1 in the appendix shows equivalence of the three expressions for $L_{NT}(\beta)$ given above.

Multiple local minima of $L_{NT}(\beta)$ may exist, and one should use multiple starting values for the numerical optimization of β to guarantee that the true global minimum $\widehat{\beta}$ is found.

To show consistency of the LS estimator $\widehat{\beta}$ of the interactive fixed effect model, and also later for our first order asymptotic theory, we consider the limit $N,T\to\infty$. In the following we present assumptions on X_k, e, λ, and f that guarantee consistency.[12]

Assumption 1. *(i)* $\operatorname{plim}_{N,T\to\infty}\left(\lambda^{0\prime}\lambda^0/N\right)>0$, *(ii)* $\operatorname{plim}_{N,T\to\infty}\left(f^{0\prime}f^0/T\right)>0$.

Assumption 2. $\operatorname{plim}_{N,T\to\infty}\left[(NT)^{-1}\mathrm{Tr}(X_k\,e')\right]=0$, *for all* $k=1,\ldots,K$.

Assumption 3. $\operatorname{plim}_{N,T\to\infty}\left(\|e\|/\sqrt{NT}\right)=0$.

Assumption 1 guarantees that the matrices f^0 and λ^0 have full rank, i.e. that there are R distinct factors and factor loadings asymptotically, and that the norm of each factor and factor loading grows at a rate of $\sqrt{T}$ and $\sqrt{N}$, respectively. Assumption 2 demands that the regressors are weakly exogenous. Assumption 3 restricts the spectral norm of the $N\times T$ error matrix e. We discuss this assumption in more detail in the next section, and we give examples of error distributions that satisfy this condition in Appendix A. The final assumption needed for consistency is an assumption on the regressors X_k. We already introduced the distinction between the K_1 "low-rank regressors" X_l, $l=1,\ldots,K_1$, and the $K_2=K-K_1$ "high-rank regressors" X_m, $m=K_1+1,\ldots,K$ above.

Assumption 4.

(i) $\operatorname{plim}_{N,T\to\infty}\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\right]>0$.

(ii) The two types of regressors satisfy:

(a) Consider linear combinations $\alpha\cdot X_{\text{high}}=\sum_{m=K_1+1}^{K}\alpha_m X_m$ *of the high-rank regressors* X_m *for* K_2*-vectors*[13] α *with* $\|\alpha\|=1$. *We assume that there exists a constant* $b>0$ *such that*

$$\min_{\{\alpha\in\mathbb{R}^{K_2},\|\alpha\|=1\}}\sum_{r=2R+K_1+1}^{N}\mu_r\left[\frac{(\alpha\cdot X_{\text{high}})(\alpha\cdot X_{\text{high}})'}{NT}\right]\geq b\qquad wpa1.$$

[12] We could write $X_k^{(N,T)}$, $e^{(N,T)}$, $\lambda^{(N,T)}$ and $f^{(N,T)}$, because all these matrices, and even their dimensions, are functions on N and T, but we suppress this dependence throughout the paper.

[13] The components of the K_2-vector α are denoted by α_{K_1+1} to α_K.

(b) For the low-rank regressors we assume $\text{rank}(X_l) = 1$, $l = 1, \ldots, K_1$, *i.e. they can be written as* $X_l = w_l v_l'$ *for* N*-vectors* w_l *and* T*-vectors* v_l*, and we define the* $N \times K_1$ *matrix* $w = (w_1, \ldots, w_{K_1})$ *and the* $T \times K_1$ *matrix* $v = (v_1, \ldots, v_{K_1})$*. We assume that there exists a constant* $B > 0$ *such that* $N^{-1} \lambda^{0\prime} M_w \lambda^0 > B \mathbb{I}_R$ *and* $T^{-1} f^{0\prime} M_v f^0 > B \mathbb{I}_R$*, wpa1.*

Assumption $4(i)$ is a standard non-collinearity condition for all the regressors. Assumption $4(ii)(a)$ is an appropriate sample analog of the identification Assumption ID(v). If the sum in Assumption $4(ii)(a)$ would start from $r = 1$, then we would have $\sum_{r=1}^{N} \mu_r \left[\frac{(\alpha \cdot X_{\text{high}})(\alpha \cdot X_{\text{high}})'}{NT} \right] = \frac{1}{NT} \text{Tr}[(\alpha \cdot X_{\text{high}})(\alpha \cdot X_{\text{high}})']$, so that the assumption would become a standard non-collinearity condition. Not including the first $2R + K_1$ eigenvalues in the sum implies that the $N \times N$ matrix $(\alpha \cdot X_{\text{high}})(\alpha \cdot X_{\text{high}})'$ needs to have rank larger than $2R + K_1$.

Assumption $4(ii)(b)$ is closely related to the identification Assumptions ID(iii) and (iv). The appearance of the factors and factor loadings in this assumption on the low-rank regressors is inevitable in order to guarantee consistency. For example, consider a low-rank regressor that is cross-sectionally independent and proportional to the r'th unobserved factor, e.g. $X_{l,it} = f_{tr}$. The corresponding regression coefficient β_l is then not identified, because the model is invariant under a shift $\beta_l \mapsto \beta_l + a$, $\lambda_{ir} \mapsto \lambda_{ir} - a$, for an arbitrary $a \in \mathbb{R}$. This phenomenon is well known from ordinary fixed effect models, where the coefficients of time-invariant regressors are not identified. Assumption 4(ii)(b) therefore guarantees for $X_l = w_l v_l'$ that w_l is sufficiently different from λ^0, and v_l is sufficiently different from f^0.

We can now state our consistency result for the LS estimator.

Theorem 3.1. *Let Assumption 1, 2, 3, 4 be satisfied, let the parameter set* $\mathbb{B}$ *be compact, and let* $\beta^0 \in \mathbb{B}$*. In the limit* $N, T \to \infty$ *we then have*

$$\widehat{\beta} \underset{p}{\longrightarrow} \beta^0 .$$

The proof of the theorem and of all theorems below can be found in the appendix. We assume compactness of $\mathbb{B}$ to guarantee existence of the minimizing $\widehat{\beta}$. We also use boundedness of $\mathbb{B}$ in the consistency proof, but only for those parameters β_l, $l = 1 \ldots K_1$, that correspond to low-rank regressors, i.e. if there are only high-rank regressors ($K_1 = 0$) the compactness assumption can be omitted, as long as existence of $\widehat{\beta}$ is guaranteed (e.g. for $\mathbb{B} = \mathbb{R}^K$).

Bai (2009) also proves consistency of the LS estimator of the interactive fixed effect model, but under somewhat different assumptions. He also employs, what we call Assumptions 1 and 2, and he uses a low-level version of Assumption 3. He demands the regressors to be strictly exogenous. Regarding consistency, the real difference between our assumptions and his is the treatment of high- and low-rank regressors. He first gives a condition on the regressors (his assumption A) that rules out low-rank regressors, and later discussed the case where all regressors are either time-invariant or common regressors (*i.e.* are all low-rank). In contrast, our Assumption 4 allows for a combination of high- and low-rank regressors, and for low-rank regressors that are more general than time-invariant and common regressors.

4 Asymptotic Distribution and Bias Correction

Since we have already shown consistency of the LS estimator $\widehat{\beta}$, it is sufficient to study the local properties of the objective function $L_{NT}(\beta)$ around β^0 in order to derive the first order

asymptotic theory of $\widehat{\beta}$. A useful approximation of $L_{NT}(\beta)$ around β^0 was derived in Moon and Weidner (2013), and we briefly summarize the ideas and results of this approximation in the following subsection. We then apply those results to derive the asymptotic distribution of the LS estimator, including working out the asymptotic bias, which was not done previously. Afterwards we discuss bias correction and inference.

4.1 Expansion of the Profile Objective Function

The last expression in equation (3.3) for the profile objective function is convenient because it does not involve any minimization over the parameters λ or f. On the other hand, this is not an expression that can be easily discussed by analytic means, because in general there is no explicit formula for the eigenvalues of a matrix. The conventional method that involves a Taylor series expansion in the regression parameters β *alone* seems infeasible here. In Moon and Weidner (2013) we showed how to overcome this problem by expanding the profile objective function *jointly* in β and $\|e\|$. The key idea is the following decomposition

$$Y - \beta \cdot X = \underbrace{\lambda^0 f^{0\prime}}_{\substack{\text{leading} \\ \text{term}}} \underbrace{- \left(\beta - \beta^0\right) \cdot X + e}_{\text{perturbation term}}.$$

If the perturbation term is zero, then the profile objective $L_{NT}(\beta)$ is also zero, since the leading term $\lambda^0 f^{0\prime}$ has rank R, so that the $T - R$ smallest eigenvalues of $f^0 \lambda^{0\prime} \lambda^0 f^{0\prime}$ all vanish. One may thus expect that small values of the perturbation term should correspond to small values of $L_{NT}(\beta)$. This idea can indeed be made mathematically precise. By using the perturbation theory of linear operators (see e.g. Kato (1980)) one can work out an expansion of $L_{NT}(\beta)$ in the perturbation term, and one can show that this expansion is convergent as long as the spectral norm of the perturbation term is sufficiently small.

The assumptions on the model made so far are in principle already sufficient to apply this expansion of the profile objective function, but in order to truncate the expansion at an appropriate order and to provide a bound on the remainder term which is sufficient to derive the first order asymptotic theory of the LS estimator, we need to strengthen Assumption 3 as follows.

Assumption 3*. $\|e\| = o_p(N^{2/3})$.

In the rest of the paper we only consider asymptotics where N and T grow at the same rate, i.e. we could equivalently write $o_p(T^{2/3})$ instead of $o_p(N^{2/3})$ in Assumption 3*. In Appendix A we provide examples of error distributions that satisfy Assumption 3*. In fact, for these examples, we have $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)})$. There is large literature that studies the asymptotic behavior of the spectral norm of random matrices, see e.g. Geman (1980), Silverstein (1989), Bai, Silverstein, Yin (1988), Yin, Bai, and Krishnaiah (1988), and Latala (2005). Loosely speaking, we expect the result $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)})$ to hold as long as the errors e_{it} have mean zero, uniformly bounded fourth moment, and weak time-serial and cross-sectional correlation (in some well-defined sense, see the examples).

We can now present the quadratic approximation of the profile objective function $L_{NT}(\beta)$ that was derived in Moon and Weidner (2013).

Theorem 4.1. *Let Assumption 1, 3*, and 4(i) be satisfied, and consider the limit $N, T \to \infty$ with $N/T \to \kappa^2$, $0 < \kappa < \infty$. Then, the profile objective function satisfies $L_{NT}(\beta) = L_{q,NT}(\beta) +$*

$(NT)^{-1} R_{NT}(\beta)$, *where the remainder* $R_{NT}(\beta)$ *is such that for any sequence* $\eta_{NT} \to 0$ *we have*

$$\sup_{\{\beta:\|\beta-\beta^0\|\leq\eta_{NT}\}} \frac{|R_{NT}(\beta)|}{\left(1+\sqrt{NT}\,\left\|\beta-\beta^0\right\|\right)^2} = o_p\left(1\right),$$

and $L_{q,NT}(\beta)$ *is a second order polynomial in* β*, namely*

$$L_{q,NT}(\beta) = L_{NT}(\beta^0) - \frac{2}{\sqrt{NT}}\left(\beta-\beta^0\right)' C_{NT} + \left(\beta-\beta^0\right)' W_{NT}\left(\beta-\beta^0\right),$$

with $K \times K$ *matrix* W_{NT} *defined by* $W_{NT,k_1k_2} = (NT)^{-1}\operatorname{Tr}(M_{f^0} X_{k_1}' M_{\lambda^0} X_{k_2})$*, and* K*-vector* C_{NT} *with entries* $C_{NT,k} = C^{(1)}\left(\lambda^0, f^0, X_k\, e\right) + C^{(2)}\left(\lambda^0, f^0, X_k\, e\right)$*, where*

$$\begin{aligned}
C^{(1)}\left(\lambda^0,\, f^0,\, X_k,\, e\right) &= \frac{1}{\sqrt{NT}}\operatorname{Tr}(M_{f^0}\, e'\, M_{\lambda^0}\, X_k)\,, \\
C^{(2)}\left(\lambda^0,\, f^0,\, X_k,\, e\right) &= -\frac{1}{\sqrt{NT}}\Big[\operatorname{Tr}\left(eM_{f^0}\, e'\, M_{\lambda^0}\, X_k\, f^0\, (f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right) \\
&\qquad + \operatorname{Tr}\left(e' M_{\lambda^0}\, e\, M_{f^0}\, X_k'\, \lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\, f^{0\prime}\right) \\
&\qquad + \operatorname{Tr}\left(e' M_{\lambda^0}\, X_k\, M_{f^0}\, e'\, \lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\, f^{0\prime}\right)\Big]\,.
\end{aligned}$$

We refer to W_{NT} and C_{NT} as the approximated Hessian and the approximated score (at the true parameter β^0). The exact Hessian and the exact score (at the true parameter β^0) contain higher order expansion terms in e, but the expansion up to the particular order above is sufficient to work out the first order asymptotic theory of the LS estimator, as the following corollary shows.

Corollary 4.2. *Let the assumptions of Theorem 3.1 and 4.1 hold, let* β^0 *be an interior point of the parameter set* $\mathbb{B}$*, and assume that* $C_{NT} = \mathcal{O}_p(1)$*. We then have* $\sqrt{NT}\big(\widehat{\beta}-\beta^0\big) = W_{NT}^{-1}C_{NT} + o_p(1) = \mathcal{O}_p(1)$*.*

Combining consistency of the LS estimator and the expansion of the profile objective function in Theorem 4.1, one obtains $\sqrt{NT}\, W_{NT}\big(\widehat{\beta}-\beta^0\big) = C_{NT} + o_p(1)$ (see e.g. Andrews (1999)). To obtain the corollary one needs in addition that W_{NT} does not become degenerate as $N, T \to \infty$, i.e. the smallest eigenvalue of W_{NT} should by bounded by a positive constant. Our assumptions already guarantee this, as is shown in the supplementary material.

4.2 Asymptotic Distribution

We now apply Corollary 4.2 to work out the asymptotic distribution of the LS estimator $\widehat{\beta}$. For this purpose we need more specific assumptions on λ^0, f^0, X_k, and e.

Assumption 5. *There exists a conditioning set* $\mathcal{C} = \mathcal{C}_{NT}$*, which contains the sigma-algebra generated by* λ^0 *and* f^0*, such that*

(i) $\mathbb{E}e_{it} = 0$ *for all* i, t*.*

(ii) $\{(X_{it}, e_{it}), t = 1, \ldots, T\}$ *is independent across* i*, conditional on* $\mathcal{C}$*.*

(iii) $e_{it} \perp \mathcal{C}$*, and* $e_{it} \perp \{(X_{is}, e_{i,s-1}), s \leq t\} \,\big|\, \mathcal{C}$*, for all* i, t*.*

(iv) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\left|\text{Cov}\left(X_{k,it},X_{k,is}\middle|\mathcal{C}\right)\right| = \mathcal{O}_p(1)$, *for all* $k=1,\ldots,K$.

(v) $\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\text{Cov}\left(e_{it}\widetilde{X}_{k,is},\, e_{iu}\widetilde{X}_{k,iv}\middle|\mathcal{C}\right)\right| = \mathcal{O}_p(1)$, *where* $\widetilde{X}_{k,it} \equiv X_{k,it} - \mathbb{E}\left[X_{k,it}\middle|\mathcal{C}\right]$, *for all* $k=1,\ldots,K$.

(vi) $\mathbb{E}e_{it}^8$ *and* $\mathbb{E}\left(\|X_{it}\|^{8+\epsilon}\middle|\mathcal{C}\right)$ *and* $\mathbb{E}\|\lambda_i^0\|^4$ *and* $\mathbb{E}\|f_t^0\|^{4+\epsilon}$ *are bounded uniformly over* i,t *and* N,T, *and over all realizations of* $\mathcal{C}$, *for some* $\epsilon>0$.

(vii) β^0 *is an interior point of the compact parameter set* $\mathbb{B}$.

Remarks on Assumption 5

(1) Assumption 5 imposes (i) mean zero errors, (ii) cross-sectional independence, conditional on $\mathcal{C}$, (iii) strict exogeneity of $\mathcal{C}$, sequential exogeneity of X_{it}, time-serial independence of errors, (iv) weak time-serial correlation of X_{it}, (v) weak time-serial correlation of $\widetilde{X}_{k,it} = X_{k,it} - \mathbb{E}\left[X_{k,it}\middle|\mathcal{C}\right]$ and e_{it}, (vi) bounded moments, and $(viii)$ a compact parameter set with interior true parameter.

(2) Assumption 5(i) and (iii) imply that $\mathbb{E}\left(X_{it}e_{it}|\mathcal{C}\right) = \mathbb{E}\left(e_{it}|\mathcal{C}\right)\mathbb{E}\left(X_{it}|\mathcal{C}\right) = \mathbb{E}\left(e_{it}\right)\mathbb{E}\left(X_{it}|\mathcal{C}\right) = 0$. Analogously we obtain $\mathbb{E}\left(X_{it}e_{it}X_{is}e_{is}|\mathcal{C}\right) = 0$ for $t\neq s$. Thus, the assumption guarantees that $X_{it}e_{it}$ is mean zero, uncorrelated over t, and independent across i, conditional on $\mathcal{C}$.

(3) Assumption 5 is sufficient for Assumption 2. To see this, notice that $\text{Tr}(X_k\, e') = \sum_{i,t} X_{k,it}e_{it}$, and that the sequential exogeneity and the cross-sectional independence assumption imply that $\mathbb{E}\left[\left((NT)^{-1}\sum_{i,t}X_{k,it}e_{it}\right)^2\middle|\mathcal{C}\right] = (NT)^{-2}\sum_{i,t}\mathbb{E}\left[\left(X_{k,it}e_{it}\right)^2\middle|\mathcal{C}\right]$. Together with the assumption of bounded moments this gives $(NT)^{-1}\sum_{i,t}X_{k,it}e_{it} = o_p(1)$.

(4) Assumption 5 is also sufficient for Assumption 3* (and thus for Assumption 3). This is because e_{it} is assumed independent over t and across i and has bounded 4'th moment, which according to Latala (2005) implies that the spectral norm satisfies $\|e\| = \sqrt{\max(N,T)}$ as N and T become large.

(5) Examples of regressor processes, which satisfy assumption Assumption 5(iv) and (v) are discussed in the following. This will also illuminate the role of the conditioning set $\mathcal{C}$.

Examples of DGPs for X_{it}

Here we provide examples of the DGPs of the regressors X_{it} that satisfy the conditions in Assumption 5. Proofs for these examples are provided in the supplementary material.

Example 1. *The first example is a simple AR(1) interactive fixed effect regression*

$$Y_{it} = \beta^0 Y_{i,t-1} + \lambda_i^{0\prime} f_t^0 + e_{it},$$

where e_{it} *is mean zero, independent across* i *and* t, *and independent of* λ^0 *and* f^0. *Assume that* $|\beta^0|<1$ *and that* e_{it}, λ_i^0 *and* f_t^0 *all possess uniformly bounded moments of order* $8+\epsilon$. *In this case, the regressor is* $X_{it} = Y_{it-1} = \lambda_i^{0\prime}F_t^0 + U_{it}$, *where* $F_t^0 = \sum_{s=0}^{\infty}(\beta^0)^s f_{t-1-s}^0$ *and* $U_{it} = \sum_{s=0}^{\infty}(\beta^0)^s e_{i,t-1-s}$. *For the conditioning sigma field* $\mathcal{C}$ *in Assumption 5, we choose* $\mathcal{C} = \sigma\left(\{\lambda_i^0 : 1\leq i\leq N\}, \{f_t^0 : 1\leq t\leq T\}\right)$. *Conditional on* $\mathcal{C}$ *the only variation in* X_{it} *stems from*

U_{it}, which is independent across i and weakly correlated over t, so that Assumption 5(iv) holds. Furthermore, we have $\mathbb{E}(X_{it}|\mathcal{C}) = \lambda_i^{0\prime} F_t^0$ and $\widetilde{X}_{it} = U_{it}$, which allows to verify Assumption 5(v).

This example can be generalized to a VAR(1) *model as follows:*

$$\begin{pmatrix} Y_{it} \\ Z_{it} \end{pmatrix} = \mathcal{B} \underbrace{\begin{pmatrix} Y_{i,t-1} \\ Z_{i,t-1} \end{pmatrix}}_{=X_{it}} + \begin{pmatrix} \lambda_i^{0\prime} f_t^0 \\ d_{it} \end{pmatrix} + \underbrace{\begin{pmatrix} e_{it} \\ u_{it} \end{pmatrix}}_{=E_{it}}, \tag{4.1}$$

where Z_{it} is an $m \times 1$ vector of additional variables and $\mathcal{B}$ is an $(m+1) \times (m+1)$ matrix of VAR parameters. The $m \times 1$ vector d_{it} and the factors f_t^0 and factor loadings λ_i^0 are assumed to be independent of the $(m+1) \times 1$ vector of innovations E_{it}. Suppose that our interest is to estimate the first row in equation (4.1)*, which corresponds exactly to our interactive fixed effects model with regressors $Y_{i,t-1}$ and $Z_{i,t-1}$. Choosing $\mathcal{C}$ to be the sigma field generated by all f_t^0, λ_i^0, d_{it} we obtain $\widetilde{X}_{it} = \sum_{s=0}^{\infty} \mathcal{B}^s E_{i,t-1-s}$. Analogous to the AR(1) case we then find that Assumption 5(iv) and (v) are satisfied in this example if the innovations E_{it} are independent across i and over t, have appropriate bounded moments (higher than four), and the absolute values of the eigenvalues of $\mathcal{B}$ are all smaller than one.*

Example 2. *Consider a scalar X_{it} for simplicity, and let $X_{it} = g(v_{it}, \delta_i, h_t)$. We assume that (i) $\left\{(e_{it}, v_{it})_{i=1,\ldots,N;t=1,\ldots,T}\right\} \perp \left\{\left(\lambda_i^0, \delta_i\right)_{i=1,\ldots,N}, \left(f_t^0, h_t\right)_{t=1,\ldots,T}\right\}$, (ii) $(e_{it}, v_{it}, \delta_i)$ are independent across i for all t, and (iii) $v_{is} \perp e_{it}$ for $s \leq t$ and all i. Furthermore assume that $\sup_{it} \mathbb{E}|X_{it}|^{8+\epsilon} < \infty$ for some positive ϵ. For the conditioning sigma field $\mathcal{C}$ in Assumption 5 we choose $\mathcal{C} = \sigma\left(\left\{\lambda_i^0 : 1 \leq i \leq N\right\}, \left\{\delta_i : 1 \leq i \leq N\right\}, \left\{f_t^0 : 1 \leq t \leq T\right\}, \left\{h_t : 1 \leq t \leq T\right\}\right)$. Furthermore, let $\mathcal{F}_\tau^t(i) = \sigma\left(\{(e_{is}, v_{is}) : \tau \leq s \leq t\}, \mathcal{C}\right)$, and define the conditional α-mixing coefficient on $\mathcal{C}$,*

$$\alpha_m(i) = \sup_{A \in \mathcal{F}_{-\infty}^t(i), B \in \mathcal{F}_{t+m}^{\infty}(i)} \left[\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\mathcal{C}\right].$$

Let $\alpha_m = \sup_i \alpha_m(i)$, and assume that $\alpha_m = O\left(m^{-\zeta}\right)$, where $\zeta > \frac{12p}{4p-1}$ for $p > 4$. Then, Assumption 5(iv) and (v) are satisfied.

In this example, the shocks h_t (which may contain the factors f_t^0), δ_i (which may contain the factor loadings λ_i^0), and v_{it} (which may contain past values of e_{it}) can enter in a general non-linear way into the regressor X_{it}.

The following assumption guarantees that the limiting variance and the asymptotic bias converge to constant values.

Assumption 6. *Let $\mathcal{X}_k = M_{\lambda^0} X_k M_{f^0}$, which is an $N \times T$ matrix with entries $\mathcal{X}_{k,it}$. For each i and t, define the K-vector $\mathcal{X}_{it} = (\mathcal{X}_{1,it}, \ldots, \mathcal{X}_{K,it})'$. We assume existence of the following*

probability limits for all $k = 1, \ldots, K$,

$$
\begin{aligned}
W &= \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathcal{X}_{it}\, \mathcal{X}_{it}' \,, \\
\Omega &= \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}\left(e_{it}^2\right)\, \mathcal{X}_{it}\, \mathcal{X}_{it}' \,, \\
B_{1,k} &= \operatorname*{plim}_{N,T\to\infty} \frac{1}{N} \operatorname{Tr}\left[P_{f^0} \mathbb{E}\left(e' X_k \,\middle|\, \mathcal{C}\right)\right] \,, \\
B_{2,k} &= \operatorname*{plim}_{N,T\to\infty} \frac{1}{T} \operatorname{Tr}\left[\mathbb{E}\left(ee'\right)\, M_{\lambda^0}\, X_k\, f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime}\right] \,, \\
B_{3,k} &= \operatorname*{plim}_{N,T\to\infty} \frac{1}{N} \operatorname{Tr}\left[\mathbb{E}\left(e'e\right)\, M_{f^0}\, X_k'\, \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \left(f^{0\prime} f^0\right)^{-1} f^{0\prime}\right] \,,
\end{aligned}
$$

where $\mathcal{C}$ *is the same conditioning set that appears in Assumption 5.*

Here, W and Ω are $K \times K$ matrices, and we define the K-vectors B_1, B_2 and B_3 with components $B_{1,k}$, $B_{2,k}$ and $B_{3,k}$, $k = 1, \ldots, K$.

Theorem 4.3. *Let Assumptions 1, 4, 5 and 6 be satisfied,*[14] *and consider the limit* $N, T \to \infty$ *with* $N/T \to \kappa^2$, *where* $0 < \kappa < \infty$. *Then we have*

$$
\sqrt{NT}\left(\widehat{\beta} - \beta^0\right) \underset{d}{\to} \mathcal{N}\left(W^{-1} B,\; W^{-1}\, \Omega\, W^{-1}\right) \,,
$$

where $B = -\kappa B_1 - \kappa^{-1} B_2 - \kappa B_3$.

From Corollary 4.2 we already know that the limiting distribution of $\widehat{\beta}$ is given by the limiting distribution of $W_{NT}^{-1} C_{NT}$. Note that $W_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathcal{X}_{it}\, \mathcal{X}_{it}'$, i.e. W is simply defined as the probability limit of W_{NT}. Assumption 4 guarantees that W is positive definite.

Thus, the main task in showing Theorem 4.3 is to show that the approximated score at the true parameter satisfies $C_{NT} \to_d \mathcal{N}(B, \Omega)$. It turns out that the asymptotic variance Ω and the asymptotic bias B_1 originate from the $C^{(1)}$ term, while the two further bias terms B_2 and B_3 originate from the $C^{(2)}$ term of C_{NT}.

The bias B_1 is due to correlation of the errors e_{it} and the regressors $X_{k,i\tau}$ in the time direction (for $\tau > t$). This bias term generalizes the Nickell (1981) bias that occurs in dynamic models with standard fixed effects, and it is not present in Bai (2009), where only strictly exogenous regressors are considered.

The other two bias terms B_2 and B_3 are already described in Bai (2009). If e_{it} is homoscedastic, i.e. if $\mathbb{E}(e_{it}) = \sigma^2$, then $\mathbb{E}\left(ee'\right) = \sigma^2 \mathbb{I}_N$ and $\mathbb{E}\left(e'e\right) = \sigma^2 \mathbb{I}_T$, so that $B_2 = 0$ and $B_3 = 0$ (because the trace is cyclical and $f^{0\prime} M_{f^0} = 0$ and $\lambda^{0\prime} M_{\lambda^0} = 0$). Thus, B_2 is only non-zero if e_{it} is heteroscedastic across i, and B_3 is only non-zero if e_{it} is heteroscedastic over t. Correlation in e_{it} across i or over t would also generate non-zero bias terms of exactly the form B_2 and B_3, but is ruled out by our assumptions.

[14] Assumption 2 and 3* are implied by Assumption 5 and therefore need not be explicitly assumed here.

4.3 Bias Correction

In order to express our estimators for the asymptotic bias and the asymptotic variance of $\widehat{\beta}$ we first have to introduce some notation.

Definition 1. *Let $\Gamma : \mathbb{R} \to \mathbb{R}$ be the truncation kernel defined by $\Gamma(x) = 1$ for $|x| \leq 1$, and $\Gamma(x) = 0$ otherwise. Let M be a bandwidth parameter that depends on N and T. For an $N \times N$ matrix A with elements A_{ij} and a $T \times T$ matrix B with elements B_{ts} we define*

(i) the diagonal truncations $A^{\text{truncD}} = \text{diag}[(A_{ii})_{i=1,\ldots,N}]$ and $B^{\text{truncD}} = \text{diag}[(B_{tt})_{t=1,\ldots,T}]$.

(ii) the right-sided Kernel truncation of B, which is a $T \times T$ matrics B^{truncR} with elements $B^{\text{truncR}}_{ts} = \Gamma\left(\frac{s-t}{M}\right) B_{ts}$ for $t < s$, and $B^{\text{truncR}}_{ts} = 0$ otherwise.

Here, we suppress the dependence of B^{truncR} on the bandwidth parameter M. Estimators for W, Ω, B_1, B_2, and B_3 are obtained by forming suitable sample analogs and replacing the unobserved λ^0, f^0 and e by the estimates $\widehat{\lambda}$, $\widehat{f}$ and the residuals $\widehat{e}$.

Definition 2. *Let $\widehat{\mathcal{X}}_k = M_{\widehat{\lambda}} X_k M_{\widehat{f}}$. For each i and t, define the K-vector $\widehat{\mathcal{X}}_{it} = (\widehat{\mathcal{X}}_{1,it}, \ldots, \widehat{\mathcal{X}}_{K,it})'$. We define the $K \times K$ matrices $\widehat{W}$ and $\widehat{\Omega}$, and the K-vectors $\widehat{B}_1$, $\widehat{B}_2$ and $\widehat{B}_3$ as follows*

$$
\begin{aligned}
\widehat{W} &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\mathcal{X}}_{it} \, \widehat{\mathcal{X}}_{it}' \,, \\
\widehat{\Omega} &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\widehat{e}_{it})^2 \, \widehat{\mathcal{X}}_{it} \, \widehat{\mathcal{X}}_{it}' \,, \\
\widehat{B}_{1,k} &= \frac{1}{N} \operatorname{Tr}\left[P_{\widehat{f}} \left(\widehat{e}' X_k \right)^{\text{truncR}} \right] \,, \\
\widehat{B}_{2,k} &= \frac{1}{T} \operatorname{Tr}\left[\left(\widehat{e}\, \widehat{e}' \right)^{\text{truncD}} \, M_{\widehat{\lambda}} X_k \, \widehat{f} \, (\widehat{f}' \widehat{f})^{-1} \, (\widehat{\lambda}' \widehat{\lambda})^{-1} \, \widehat{\lambda}' \right] \,, \\
\widehat{B}_{3,k} &= \frac{1}{N} \operatorname{Tr}\left[\left(\widehat{e}' \, \widehat{e} \right)^{\text{truncD}} \, M_{\widehat{f}} X_k' \, \widehat{\lambda} \, (\widehat{\lambda}' \widehat{\lambda})^{-1} \, (\widehat{f}' \widehat{f})^{-1} \, \widehat{f}' \right] \,,
\end{aligned}
$$

where $\widehat{e} = Y - \widehat{\beta} \cdot X - \widehat{\lambda} \, \widehat{f}'$.

Notice that the estimators $\widehat{\Omega}$, $\widehat{B}_2$, and $\widehat{B}_3$ are similar to White's standard error estimator under heteroskedasticity and the estimator $\widehat{B}_1$ is similar to the HAC estimator with a kernel. To show consistency of these estimators we impose some additional assumptions.

Assumption 7.

(i) $\|\lambda_i^0\|$ and $\|f_t^0\|$ are uniformly bounded over i, t and N, T.

(ii) There exists $c > 0$ and $\epsilon > 0$ such that for all i, t, m, N, T we have $\left| \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}(e_{it} X_{k,it+m} \,|\, \mathcal{C}) \right| \leq c \, m^{-(1+\epsilon)}$.

Assumption 7(i) is made for convenience in order to simplify the consistency proof for the estimators in Definition 2. It is possible to weaken this assumption by only assuming suitable bounded moments of $\|\lambda_i^0\|$ and $\|f_t^0\|$. In order to show consistency of $\widehat{B}_1$ we need to control how strongly e_{it} and $X_{k,i\tau}$, $t < \tau$, are allowed to be correlated, which is done by Assumption 7(ii). It is straightforward to verify that Assumption 7(ii) is satisfied in the two examples of regressors processes presented below Assumption 5.

Theorem 4.4. *Let Assumptions 1, 4, 5, 6 and 7 hold, and consider a limit $N,T \to \infty$ with $N/T \to \kappa^2$, $0 < \kappa < \infty$, such that the bandwidth $M = M_{NT}$ satisifies $M \to \infty$ and $M^5/T \to 0$. We then have $\widehat{W} = W + o_p(1)$, $\widehat{\Omega} = \Omega + o_p(1)$, $\widehat{B}_1 = B_1 + o_p(1)$, $\widehat{B}_2 = B_2 + o_p(1)$, and $\widehat{B}_3 = B_3 + o_p(1)$.*

The assumption $M^5/T \to 0$ can be relaxed if additional higher moment restrictions on e_{it} and $X_{k,it}$ are imposed. Note also that for the construction of the estimators $\widehat{W}$, $\widehat{\Omega}$, and $\widehat{B}_i$, $i = 1, 2, 3$, it is not necessary to know whether the regressors are strictly exogenous or predetermined; in both cases the estimators for W, Ω, and B_i, $i = 1, 2, 3$, are consistent. We can now present our bias corrected estimator and its limiting distribution.

Corollary 4.5. *Under the assumptions of Theorem 4.4 the bias corrected estimator*

$$\widehat{\beta}^* = \widehat{\beta} + \widehat{W}^{-1}\left(T^{-1}\widehat{B}_1 + N^{-1}\widehat{B}_2 + T^{-1}\widehat{B}_3\right)$$

satisfies $\sqrt{NT}\left(\widehat{\beta}^* - \beta^0\right) \to_d \mathcal{N}\left(0,\ W^{-1}\,\Omega\,W^{-1}\right)$.

According to Theorem 4.4, a consistent estimator of the asymptotic variance of $\widehat{\beta}^*$ is given by $\widehat{W}^{-1}\,\widehat{\Omega}\,\widehat{W}^{-1}$.

An alternative to the analytical bias correction result given by Corollary 4.5 is to use Jackknife bias correction in order to eliminate the asymptotic bias. For panel models with incidental parameters only in the cross-sectional dimensions one typical finds a large N,T leading incidental parameter bias of order $1/T$ for the parameters of interest. To correct for this $1/T$ bias one can use the delete-one Jackknife bias correction if observations are iid over t (Hahn and Newey, 2004) and the split-panel Jackknife bias correction if observations are correlated over t (Dhaene and Jochmans, 2010). In our current model we have incidental parameters in both panel dimensions (λ_i^0 and f_t^0), resulting in leading bias terms of order $1/T$ (bias term B_1 and B_3) and of order $1/N$ (bias term B_2). The generalizations of the split-panel Jackknife bias correction to that case was discussed in Fernández-Val and Weidner (2013).

The corresponding bias corrected split-panel Jackknife estimator reads $\widehat{\beta}^J = 3\widehat{\beta}_{NT} - \overline{\beta}_{N,T/2} - \overline{\beta}_{N/2,T}$, where $\widehat{\beta}_{NT} = \widehat{\beta}$ is the LS estimator obtained from the full sample, $\overline{\beta}_{N,T/2}$ is average of the two LS estimators that leave out the first and second halves of the time periods, and $\overline{\beta}_{N/2,T}$ is the average of the two LS estimators that leave out half of the individuals. Jackknife bias correction is convenient since only the order of the bias, but not the structure of the terms B_1, B_2, B_3 needs not be known in detail. However, one requires additional stationarity assumptions over t and homogeneity assumptions across i in order to justify the Jackknife correction and to show that $\widehat{\beta}^J$ has the same limiting distribution as $\widehat{\beta}^*$ in Corollary 4.5, see Fernández-Val and Weidner (2013) for more details. Jackknife bias correction is not explored further in this paper.

5 Testing Restrictions on β^0

In this section we discuss the three classical test statistics for testing linear restrictions on β^0. The null-hypothesis is $H_0: \ H\beta^0 = h$, and the alternative is $H_a: \ H\beta^0 \neq h$, where H is an $r \times K$ matrix of rank $r \leq K$, and h is an $r \times 1$ vector. We restrict the presentation to testing a linear hypothesis for ease of exposition. One can easily generalize the discussion to the testing of non-linear hypotheses. Throughout this subsection we assume that β^0 is an interior point of $\mathbb{B}$,

i.e. there are no local restrictions on β as long as the null-hypothesis is not imposed. Using the expansion of $L_{NT}(\beta)$ one could also discuss testing when the true parameter is on the boundary, as shown in Andrews (2001).

The restricted estimator is defined by

$$\widetilde{\beta} = \underset{\beta \in \widetilde{\mathbb{B}}}{\operatorname{argmin}} \; L_{NT}(\beta) \;,$$

where $\widetilde{\mathbb{B}} = \{\beta \in \mathbb{B} | \, H\beta = h\}$ is the restricted parameter set. Analogous to Theorem 4.3 for the unrestricted estimator $\widehat{\beta}$, we can use the expansion of the profile objective function to derive the limiting distribution of the restricted estimator. Under the assumptions of Theorem 4.3 we have

$$\sqrt{NT}(\widetilde{\beta} - \beta^0) \;\underset{d}{\longrightarrow}\; \mathcal{N}\left(\mathfrak{W}^{-1} B,\; \mathfrak{W}^{-1}\,\Omega\,\mathfrak{W}^{-1}\right) \;,$$

where $\mathfrak{W}^{-1} = W^{-1} - W^{-1}H'(HW^{-1}H')^{-1}HW^{-1}$. The $K \times K$ covariance matrix in the limiting distribution of $\widetilde{\beta}$ is not full rank, but satisfies $\operatorname{rank}(\mathfrak{W}^{-1}\,\Omega\,\mathfrak{W}^{-1}) = K - r$, because $H\mathfrak{W}^{-1} = 0$ and thus $\operatorname{rank}(\mathfrak{W}^{-1}) = K - r$. The asymptotic distribution of $\sqrt{NT}(\widetilde{\beta} - \beta^0)$ is therefore $K - r$ dimensional, as it should be for the restricted estimator.

Wald Test

Using the result of Theorem 4.3 we find that under the null-hypothesis $\sqrt{NT}\left(H\widehat{\beta} - h\right)$ is asymptotically distributed as $\mathcal{N}\left(HW^{-1}B,\; HW^{-1}\,\Omega\,W^{-1}H'\right)$. Thus, due to the presence of the bias B, the standard Wald test statistics $WD_{NT} = NT\left(H\widehat{\beta} - h\right)'\left(H\widehat{W}^{-1}\,\widehat{\Omega}\,\widehat{W}^{-1}H'\right)^{-1}\left(H\widehat{\beta} - h\right)$ is not asymptotically χ^2_r distributed. Using the estimator $\widehat{B} \equiv -\sqrt{\frac{N}{T}}\,\widehat{B}_1 - \sqrt{\frac{T}{N}}\,\widehat{B}_2 - \sqrt{\frac{N}{T}}\,\widehat{B}_3$ for the bias we can define the bias corrected Wald test statistics as

$$WD^*_{NT} = \left[\sqrt{NT}\left(H\widehat{\beta} - h\right) - H\widehat{W}^{-1}\widehat{B}\right]'\left(H\widehat{W}^{-1}\,\widehat{\Omega}\,\widehat{W}^{-1}H'\right)^{-1}\left[\sqrt{NT}\left(H\widehat{\beta} - h\right) - H\widehat{W}^{-1}\widehat{B}\right]. \tag{5.1}$$

Under the null hypothesis and the Assumptions of Theorem 4.4 we find $WD^*_{NT} \to_d \chi^2_r$.

Likelihood Ratio Test

To implement the LR test we need the relationship between the asymptotic Hessian W and the asymptotic score variance Ω of the profile objective function to be of the form $\Omega = cW$, where $c > 0$ is a scalar constant. This is satisfied in our interactive fixed effect model if $\mathbb{E}e^2_{it} = c$, i.e. if the error is homoskedastic. A consistent estimator for c is then given by $\widehat{c} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}^2_{it}$, where $\widehat{e} = Y - \widehat{\beta}\cdot X - \widehat{\lambda}\,\widehat{f}'$. Since the likelihood function for the interactive fixed effect model is just the sum of squared residuals, we have $\widehat{c} = L_{NT}(\widehat{\beta})$. The likelihood ratio test statistics is defined by

$$LR_{NT} = \widehat{c}^{-1}\,NT\left[L_{NT}\left(\widetilde{\beta}\right) - L_{NT}\left(\widehat{\beta}\right)\right] \;.$$

Under the assumption of Theorem 4.3 we then have

$$LR_{NT} \;\underset{d}{\longrightarrow}\; c^{-1}C'W^{-1}H'(HW^{-1}H')^{-1}HW^{-1}C \;,$$

where $C \sim \mathcal{N}(B, \Omega)$, i.e. $C_{NT} \to_d C$. This is the same limiting distribution that one finds for the Wald test if $\Omega = cW$ (in fact, one can show $WD_{NT} = LR_{NT} + o_p(1)$). Therefore, we need to do a bias correction for the LR test in order to achieve a χ^2 limiting distribution. We define

$$LR^*_{NT} = \widehat{c}^{-1} NT \left[\min_{\{\beta \in \mathbb{B} |\, H\beta = h\}} L_{NT}\left(\beta + (NT)^{-1/2}\widehat{W}^{-1}\widehat{B}\right) - \min_{\beta \in \mathbb{B}} L_{NT}\left(\beta + (NT)^{-1/2}\widehat{W}^{-1}\widehat{B}\right)\right], \tag{5.2}$$

where $\widehat{B}$ and $\widehat{W}$ do not depend on the parameter β in the minimization problem.[15] Asymptotically we have $\min_{\beta\in\mathbb{B}} L_{NT}\left(\beta + (NT)^{-1/2}\widehat{W}^{-1}\widehat{B}\right) = L_{NT}(\widehat{\beta})$, because $\beta \in \mathbb{B}$ does not impose local constraints, i.e. close to β^0 it does not matter for the value of the minimum whether one minimizes over β or over $\beta + (NT)^{-1/2}\widehat{W}^{-1}\widehat{B}$. The correction to the LR test therefore originates from the first term in LR^*_{NT}. For the minimization over the restricted parameter set it matters whether the argument of L_{NT} is β or $\beta + (NT)^{-1/2}\widehat{W}^{-1}\widehat{B}$, because generically we have $HW^{-1}B \neq 0$ (otherwise no correction would be necessary for the LR statistics). One can show that

$$LR^*_{NT} \underset{d}{\longrightarrow} c^{-1}(C-B)'W^{-1}H'(HW^{-1}H')^{-1}HW^{-1}(C-B)\,,$$

i.e. we obtain the same formula as for LR_{NT}, but the limit of the score C is replaced by the bias corrected term $C - B$. Under the Assumptions of Theorem 4.4, if H_0 is satisfied, and for homoscedastic errors e_{it}, we then have $LR^*_{NT} \to_d \chi^2_r$. In fact, one can show that $LR^*_{NT} = WD^*_{NT} + o_p(1)$.

Lagrange Multiplier Test

Let $\widetilde{\nabla}\mathcal{L}_{NT}$ be the gradient of the LS objective function (3.1) with respect to β, evaluated at the restricted parameter estimates, i.e.

$$\begin{aligned}\widetilde{\nabla}\mathcal{L}_{NT} \;\equiv\; \nabla\mathcal{L}_{NT}(\widetilde{\beta},\widetilde{\lambda},\widetilde{f}) &= \left(\left.\frac{\partial\mathcal{L}_{NT}(\beta,\widetilde{\lambda},\widetilde{f})}{\partial\beta_1}\right|_{\beta=\widetilde{\beta}}, \ldots, \left.\frac{\partial\mathcal{L}_{NT}(\beta,\widetilde{\lambda},\widetilde{f})}{\partial\beta_K}\right|_{\beta=\widetilde{\beta}}\right)' \\ &= -\frac{2}{NT}\Big(\mathrm{Tr}\left(X_1'\widetilde{e}\right), \ldots, \mathrm{Tr}\left(X_K'\widetilde{e}\right)\Big)',\end{aligned}$$

where $\widetilde{\lambda} = \widehat{\lambda}(\widetilde{\beta})$, $\widetilde{f} = \widehat{f}(\widetilde{\beta})$, and $\widetilde{e} = Y - \widetilde{\beta}\cdot X - \widetilde{\lambda}\,\widetilde{f}'$. Under the Assumptions of Theorem 4.3, and if the null hypothesis $H_0:\ H\beta^0 = h$ is satisfied, one finds that[16]

$$\sqrt{NT}\,\widetilde{\nabla}\mathcal{L}_{NT} = \sqrt{NT}\,\nabla L_{NT}(\widetilde{\beta}) + o_p(1). \tag{5.3}$$

Due to this equation, one can base the Lagrange multiplier test on the gradient of $\mathcal{L}_{NT}(\widetilde{\beta},\widetilde{\lambda},\widetilde{f})$, or on the gradient of the profile quasi-likelihood function $L_{NT}(\widetilde{\beta})$ and obtains the same limiting distribution.

[15] Alternatively, one could use $\widehat{B}(\widetilde{\beta})$ and $\widehat{W}(\widetilde{\beta})$ as estimates for B and W, and would obtain the same limiting distribution of LR^*_{NT} under the null hypothesis H_0. These alternative estimators are not consistent if H_0 is false, i.e. the power-properties of the test would be different. The question which specification should be preferred is left for future research.

[16] The proof of the statement is given in the appendix as part of the proof of Theorem 5.2.

Using the bound on the remainder $R_{NT}(\beta)$ given in Theorem 4.1, one cannot infer any properties of the score function, i.e. of the gradient $\nabla L_{NT}(\beta)$, because nothing is said about $\nabla R_{NT}(\beta)$. The following theorem gives the bound on $\nabla R_{NT}(\beta)$ that is sufficient to derive the limiting distribution of the Lagrange multiplier.

Theorem 5.1. *Under the assumptions of Theorem 4.1, and with W_{NT} and C_{NT} as defined there, the score function satisfies*

$$\nabla L_{NT}(\beta) \;=\; 2\,W_{NT}\,(\beta-\beta^0) \;-\; \frac{2}{\sqrt{NT}}\,C_{NT} \;+\; \frac{1}{NT}\,\nabla R_{NT}(\beta)\;,$$

where the remainder $\nabla R_{NT}(\beta)$ satisfies for any sequence $\eta_{NT}\to 0$

$$\sup_{\{\beta:\|\beta-\beta^0\|\leq\eta_{NT}\}} \frac{\|\nabla R_{NT}(\beta)\|}{\sqrt{NT}\left(1+\sqrt{NT}\;\|\beta-\beta^0\|\right)} = o_p\,(1)\,.$$

From this theorem, and the fact that $\widetilde{\beta}$ is $\sqrt{NT}$-consistent under H_0, we obtain

$$\begin{aligned}\sqrt{NT}\,\widetilde{\nabla}\mathcal{L}_{NT} &= \sqrt{NT}\,\nabla L_{q,NT}(\widetilde{\beta}) + o_p(1)\\ &= 2\sqrt{NT}\,W_{NT}\,(\widetilde{\beta}-\beta^0) - 2\,C_{NT} + o_p(1)\;.\end{aligned}$$

Using this result and the known limiting distribution of $\widetilde{\beta}$ we now find

$$\sqrt{NT}\,\widetilde{\nabla}\mathcal{L}_{NT} \;\underset{d}{\longrightarrow}\; -2H'(HW^{-1}H')^{-1}HW^{-1}C\;. \tag{5.4}$$

The LM test statistics is therefore given by[17]

$$LM_{NT} = \frac{NT}{4}\,(\widetilde{\nabla}\mathcal{L}_{NT})'\widetilde{W}^{-1}H'(H\widetilde{W}^{-1}\widetilde{\Omega}\widetilde{W}^{-1}H')^{-1}H\widetilde{W}^{-1}\widetilde{\nabla}\mathcal{L}_{NT}\;,$$

where $\widetilde{B}$, $\widetilde{W}$ and $\widetilde{\Omega}$ are defined like $\widehat{B}$, $\widehat{W}$ and $\widehat{\Omega}$, but with unrestricted parameter estimates replaced by restricted parameter estimates. One can show that the LM test is asymptotically equivalent to the Wald test: $LM_{NT} = WD_{NT} + o_p(1)$, i.e. again bias correction is necessary. We define the bias corrected LM test statistics as

$$LM^*_{NT} = \frac{1}{4}\,\left(\sqrt{NT}\,\widetilde{\nabla}\mathcal{L}_{NT}+2\widetilde{B}\right)'\widetilde{W}^{-1}H'(H\widetilde{W}^{-1}\widetilde{\Omega}\widetilde{W}^{-1}H')^{-1}H\widetilde{W}^{-1}\left(\sqrt{NT}\,\widetilde{\nabla}\mathcal{L}_{NT}+2\widetilde{B}\right)\;. \tag{5.5}$$

The following theorem summarizes the main results of the present subsection.

Theorem 5.2. *Let the assumptions of Theorem 4.4 and the null hypothesis $H_0:\; H\beta^0 = h$ be satisfied. For the bias corrected Wald and LM test statistics introduced in equation* (5.1) *and* (5.5) *we then have*

$$WD^*_{NT} \;\underset{d}{\longrightarrow}\; \chi^2_r\;, \qquad\qquad LM^*_{NT} \;\underset{d}{\longrightarrow}\; \chi^2_r\;.$$

If in addition we assume $\mathbb{E}e_{it}^2 = c$, i.e. the idiosyncratic errors are homoscedastic, and we use $\widehat{c} = L_{NT}(\widehat{\beta})$ as an estimator for c, then the LR test statistics defined in equation (5.2) *satisfies*

$$LR^*_{NT} \;\underset{d}{\longrightarrow}\; \chi^2_r\;.$$

[17] Note also that $\sqrt{NT}HW^{-1}\nabla L_{NT}(\widetilde{\beta}) \to_d -2HW^{-1}C$.

6 Extension to Endogenous Regressors

In this section we briefly discuss how to estimate the regression coefficient β^0 of Model (2.1) when some of the regressors in X_{it} are endogenous with respect to the regression error e_{it}. The question is how instrumental variables can be used to estimate the regression coefficients of the endogenous regressor in the presence of the interactive fixed effects $\lambda_i^{0\prime} f_t^0$.

In the existing literature similar questions were already investigated under various setups. Harding and Lamarche (2009; 2011) investigate the problem of estimating an endogenous panel (quantile) regression with interactive fixed effects and show how to use IVs in the CCE estimation framework. Moon, Shum and Weidner (2012) (hereafter MSW) estimate a random coefficient multinomial demand model (as in Berry, Levinsohn and Pakes (1995))) when the unobserved product-market characteristics have interactive fixed effects. The IVs are required to identify the parameters of the random coefficient distribution and to control for price endogeneity. They suggested a multi-step "least squares-minimum distance" (LS-MD) estimator.[18] The LS-MD approach is also applicable to linear panel regression models with endogenous regressors and interactive fixed effects, as demonstrated in Lee, Moon, and Weidner (2012) for the case of a dynamic linear panel regression model with interactive fixed effects and measurement error.

We now discuss how to implement the LS-MD estimation in our setup. Let X_{it}^{end} be the vectors of endogenous regressors, and let X_{it}^{exo} be the vector of exogenous regressors, with respect to e_{it}, such that $X_{it} = (X_{it}^{\text{end}\prime}, X_{it}^{\text{exo}\prime})'$. The model then reads

$$Y_{it} = \beta_{\text{end}}^{0\prime} X_{it}^{\text{end}} + \beta_{\text{exo}}^{0\prime} X_{it}^{\text{exo}} + \lambda_i^{0\prime} f_t^0 + e_{it},$$

where $\mathbb{E}\left(e_{it} X_{it}^{\text{exo}} | \lambda^0, f^0\right) = 0$, but $\mathbb{E}\left(e_{it} X_{it}^{\text{end}} | \lambda^0, f^0\right) \neq 0$. Suppose that Z_{it} is an additional L-vector of instrumental variables (IVs) such that $\mathbb{E}\left(e_{it} Z_{it} | \lambda^0, f^0\right) = 0$, but Z_{it} may be correlated with λ_i^0 and f_t^0. The LS-MD estimator of $\beta^0 = \left(\beta_{\text{end}}^{0\prime}, \beta_{\text{exo}}^{0\prime}\right)'$ can then be calculated by the following three steps:

(1) For given β_{end} we run the least squares regression of $Y_{it} - \beta_{\text{end}}' X_{it}^{\text{end}}$ on the included exogeneous regressors X_{it}^{exo}, the interactive fixed effects $\lambda_i' f_t$, and the IVs Z_{it} :

$$\begin{aligned}&\left(\widetilde{\beta}_{\text{exo}}\left(\beta_{\text{end}}\right), \widetilde{\gamma}\left(\beta_{\text{end}}\right), \widetilde{\lambda}\left(\beta_{\text{end}}\right), \widetilde{f}\left(\beta_{\text{end}}\right)\right) \\ &\quad = \underset{\{\beta_{\text{exo}}, \gamma, \lambda, f\}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \beta_{\text{end}}' X_{it}^{\text{end}} - \beta_{\text{exo}}' X_{it}^{\text{exo}} - \gamma' Z_{it} - \lambda_i' f_t\right)^2.\end{aligned}$$

(2) We estimate β_{end} by finding $\widetilde{\gamma}\left(\beta_{\text{end}}\right)$, obtained by step (1), that is closest to zero. For this, we choose a symmetric positive definite $L \times L$ weight matrix W_{NT}^{γ} and compute

$$\widehat{\beta}_{\text{end}} = \underset{\beta_{\text{end}}}{\operatorname{argmin}}\, \widetilde{\gamma}\left(\beta_{\text{end}}\right)' W_{NT}^{\gamma} \widetilde{\gamma}\left(\beta_{\text{end}}\right).$$

(3) We estimate β_{exo} (and λ, f) by running the least squares regression of $Y_{it} - \widehat{\beta}_{\text{end}}' X_{it}^{\text{end}}$ on the included exogeneous regressors X_{it}^{exo} and the interactive fixed effects $\lambda_i' f_t$:

$$\left(\widehat{\beta}_{\text{exo}}, \widehat{\lambda}, \widehat{f}\right) = \underset{\{\beta_{\text{exo}}, \gamma, \lambda, f\}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \widehat{\beta}_{\text{end}}' X_{it}^{\text{end}} - \beta_{\text{exo}}' X_{it}^{\text{exo}} - \lambda_i' f_t\right)^2.$$

[18] Chernazhukov and Hansen (2005) also used a similar method for estimating endogenous quantile regression models.

The idea behind this estimation procedure is that valid instruments are excluded from the model for Y_{it}, so that their first step regression coefficients $\widetilde{\gamma}(\beta_{\text{end}})$ should be close to zero if β_{end} is close to its true value β^0_{end}. Thus, as long as X^{exo}_{it} and Z_{it} jointly satisfy the assumptions of the current paper we obtain $\widetilde{\gamma}\left(\beta^0_{\text{end}}\right) = o_p(1)$ for the first step LS estimator, and we also obtain the asymptotic distribution of $\widetilde{\gamma}\left(\beta^0_{\text{end}}\right)$ from the results derived in Section 4.

However, to justify the second step minimization formally one needs to study the properties of $\widetilde{\gamma}(\beta_{\text{end}})$ also for $\beta_{\text{end}} \neq \beta^0_{\text{end}}$. For this we refer to MSW. Our $\beta_{\text{end}}, \beta_{\text{exo}}$, and $Y_{it} - \beta'_{\text{end}} X^{\text{end}}_{it}$ correspond to their α, β and $\delta_{jt}(\alpha)$, respectively. The Assumptions 1 to 5 in MSW can be translated accordingly, and the results in MSW show large N, T consistency and asymptotic normality of the LS-MD estimator.

The final step of the LS-MD estimation procedure is essentially a repetition of the first step, but without including Z_{it} in the set of regressors, which results in some efficiency gains for $\widehat{\beta}_{\text{exo}}$ compared to the first step.

7 Monte Carlo Simulations

We consider an AR(1) model with $R = 1$ factors:

$$Y_{it} \;=\; \rho^0 \, Y_{i,t-1} \;+\; \lambda^0_i \, f^0_t \;+\; e_{it} \; .$$

We estimate the model as an interactive fixed effect model, i.e. no distributional assumption on λ^0_i and f^0_t are made in the estimation. The parameter of interest is ρ^0. The estimators we consider are the OLS estimator (which completely ignores the presence of the factors), the least squares estimator with interactive fixed effects (denoted FLS in this section to differentiate from OLS) defined in equation (3.2),[19] and its bias corrected version (denoted BC-FLS), defined in Theorem 4.5.

For the simulation we draw the e_{it} independently and identically distributed from a t-distribution with five degrees of freedom, the λ^0_i independently distributed from $\mathcal{N}(1,1)$, and we generate the factors from an AR(1) specification, namely $f^0_t = \rho_f \, f^0_{t-1} + u_t$, where $u_t \sim \text{iid}\mathcal{N}(0, (1-\rho^2_f)\sigma^2_f)$, and σ_f is the standard deviation of f^0_t. For all simulations we generate 1000 initial time periods for f^0_t and Y_{it} that are not used for estimation. This guarantees that the simulated data used for estimation is distributed according to the stationary distribution of the model.

In this setup there is no correlation and heteroscedasticity in e_{it}, i.e. only the bias term B_1 of the LS estimator is non-zero, but we ignore this information in the estimation, i.e. we correct for all three bias terms (B_1, B_2, and B_3, as introduced in Assumption 6) in the bias corrected LS estimator.

Table 1 shows the simulation results for the bias, standard error and root mean square error of the three different estimators for the case $N = 100$, $\rho_f = 0.5$, $\sigma_f = 0.5$, and different values of ρ^0 and T. As expected, the OLS estimator is biased due to the factor structure and its bias does not vanish (it actually increases) as T increases. The FLS estimator is also biased, but as predicted by the theory its bias vanishes as T increases. The bias corrected FLS estimator performs even better than the non-corrected LS estimator, in particular its bias vanishes even faster. Since we only correct for the first order bias of the FLS estimator, we could not expect the bias corrected FLS estimator to be unbiased. However, as T gets larger more and more of

[19] Here we can either use $\mathbb{B} = (-1, 1)$, or $\mathbb{B} = \mathbb{R}$. In the present model we only have high-rank regressors, i.e. the parameter space need not be bounded to show consistency.

the LS estimator bias is corrected for, e.g. for $\rho^0 = 0.3$ we find that at $T = 5$ the bias correction only corrects for about half of the bias, while at $T = 80$ it already corrects for about 90% of it.

Table 2 is very similar to Table 1, with the only difference that we allow for misspecification in the number of factors R, namely the true number of factors is assumed to be $R = 1$ (i.e. same DGP as for Table 1), but we incorrectly use $R = 2$ factors when calculating the FLS and BC-FLS estimator. By comparing Table 2 with Table 1 we find that this type of misspecification of the number of factors increases the bias and the standard deviation of both the FLS and the BC-FLS estimator at finite sample. That increase, however, is comparatively small once both N and T are large. According to the results in Moon and Weidner (2013) we expect the limiting distribution of the correctly specified ($R = 1$) and incorrectly specified ($R = 2$) FLS estimator to be identical when N and T grow at the same rate. Our simulations suggest that the same is true for the BC-FLS estimator, which was not explored in Moon and Weidner (2013). The remaining simulation all assume correctly specified $R = 1$.

An import issue is the choice of bandwidth M for the bias correction. Table 3 gives the fraction of the FLS estimator bias that is captured by the estimator for the bias in a model with $N = 100$, $T = 20$, $\rho_f = 0.5$, $\sigma_f = 0.5$ and different values for ρ and M. The table shows that the optimal bandwidth (in the sense that most of the bias is corrected for) depends on ρ^0: it is $M = 1$ for $\rho = 0$, $M = 2$ for $\rho = 0.3$, $M = 3$ and $\rho = 0.6$, and $M = 5$ for $\rho = 0.9$. Choosing the bandwidth too large or too small results in a smaller fraction of the bias to be corrected. Table 4 also reports the properties of the BC-FLS estimator for different values of ρ^0, T and M. It shows that the effect of the bandwidth choice on the standard deviation of the BC-FLS estimator is relatively small at $T = 40$, but is more pronounced at $T = 20$. The issue of optimal bandwidth choice is therefore an important topic for future research. In the simulation results presented here we tried to choose reasonable values for M, but made no attempt of optimizing the bandwidth.

In our setup we have $\|\lambda^0 f^{0\prime}\| \approx \sqrt{2NT}\sigma_f$ and $\|e\| \approx \sqrt{N} + \sqrt{T}$.[20] Assumption 1 and 3 imply that asymptotically $\|\lambda^0 f^{0\prime}\| \gg \|e\|$. We can therefore only be sure that our asymptotic results for the FLS estimator distribution are a good approximation of the finite sample properties if $\|\lambda^0 f^{0\prime}\| \gtrsim \|e\|$, i.e. if $\sqrt{2NT}\sigma_f \gtrsim \sqrt{N} + \sqrt{T}$. To explore this we present in Table 5 simulation results for $N = 100$, $T = 20$, $\rho^0 = 0.6$, and different values of ρ_f and σ_f. In the case $\sigma_f = 0$ we have $0 = \|\lambda^0 f^{0\prime}\| \ll \|e\|$, and this case is equivalent to $R = 0$ (no factor at all). In this case the OLS estimator estimates the true model and is almost unbiased, and correspondingly the FLS estimator and the bias corrected FLS estimator perform worse than OLS at finite sample (though we expect that all three estimators are asymptotically equivalent), but the bias corrected FLS estimator has a lower bias and a lower variance than the non-corrected FLS estimator. The case $\sigma_f = 0.2$ corresponds to $\|\lambda^0 f^{0\prime}\| \approx \|e\|$, and one finds that the bias and the variance of the OLS estimator and of the LS estimator are of comparable size. However, the bias corrected FLS estimator already has much smaller bias and a bit smaller variance in this case. Finally, in the case $\sigma_f = 0.5$ we have $\|\lambda^0 f^{0\prime}\| > \|e\|$, and we expect our asymptotic results to be a good approximation of this situation. Indeed, one finds that for $\sigma_f = 0.5$ the OLS estimator is heavily biased and very inefficient compared to the FLS estimator, while the bias corrected FLS estimator performs even better in terms of bias and variance.

In Table 6 we present simulation results for the size of the various tests discussed in the last section when testing the Null hypothesis $H_0 : \rho = \rho^0$. We choose a nominal size of 5%, $\rho_f = 0.5$, $\sigma_f = 0.5$, and different values for ρ^0, N and T. In all cases, the size distortions of

[20] To be precise, we have $\|\lambda^0 f^{0\prime}\|/(\sqrt{2NT}\sigma_f) \to_p 1$, and $\|e\|/(\sqrt{N} + \sqrt{T}) \to_p 1$.

the uncorrected Wald, LR and LM test are rather large, and the size distortions of these test do not vanish as N and T increase: the size for $N = 100$ and $T = 20$ is about the same as for $N = 400$ and $T = 80$, and the size for $N = 400$ and $T = 20$ is about the same as for $N = 1600$ and $T = 80$. In contrast, the size distortions for the bias corrected Wald, LR, and LM test are much smaller, and tend to zero (i.e. the size becomes closer to 5%) as N, T increase, holding the ratio N/T constant. For fixed T an increase in N results in a larger size distortion, while for fixed N an increase in T results in a smaller size distortion (both for the non-corrected and for the bias corrected tests).

In Table 7 and 8 we present the power and the size corrected power when testing the left sided alternative $H_a^{\text{left}} : \rho = \rho^0 - (NT)^{-1/2}$ and the right-sided alternative $H_a^{\text{right}} : \rho = \rho^0 + (NT)^{-1/2}$. The model specifications are the same as for the size results in table 4. Since both the FLS estimator and the bias corrected FLS estimator for ρ have a negative bias one finds the power for the left-sided alternative to be much smaller than the power for the right-sided alternative. For the uncorrected tests this effect can be extreme and the size-corrected power of these tests for the left sided alternative is below 2% in all cases, and does not improve as N and T become large, holding N/T fixed. In contrast, the power for the bias corrected tests becomes more symmetric as N and T become large, and the size-corrected power for the left sided alternative is much larger than for the uncorrected tests, while the size corrected power for the right sided alternative is about the same.

8 Conclusions

This paper studies the least squares estimator for dynamic linear panel regression models with interactive fixed effects. We provide conditions under which the estimator is consistent, allowing for predetermined regressors, and for a general combination of "low-rank" and "high-rank" regressors. We then show how a quadratic approximation of the profile objective function $L_{NT}(\beta)$ can be used to derive the first order asymptotic theory of the LS estimator of β under the alternative asymptotic $N, T \to \infty$. We find that the asymptotic distribution of the LS estimator can be asymptotically biased (i) due to weak exogeneity of the regressors and (ii) due and heteroscedasticity (and correlation) of the idiosyncratic errors e_{it}. Consistent estimators for the asymptotic covariance matrix and for the asymptotic bias of the LS estimator are provided, and thus a bias corrected LS estimator is given. We furthermore study the asymptotic distributions of the Wald, LR and LM test statistics for testing a general linear hypothesis on β. The uncorrected test statistics are not asymptotically chi-square due to the asymptotic bias of the score and of the LS estimator, but bias corrected test statistics that are asymptotically chi-square distributed can be constructed. A possible extensions of the estimation procedure to the case of endogeneous regressors is also discussed. The findings of our Monte Carlo simulations show that our asymptotic results on the distribution of the (bias corrected) LS estimator and of the (bias corrected) test statistics provide a good approximation of their finite sample properties. Although the bias corrected LS estimator has a non-zero bias at finite sample, this bias is much smaller than the one of the LS estimator. Analogously, the size distortions and power asymmetries of the bias corrected Wald, LR and LM test are much smaller than for the non-bias corrected versions.

References

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2):219–255.

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

Andrews, D. W. K. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1384.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, Z. D., Silverstein, J. W., and Yin, Y. Q. (1988). A note on the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivar. Anal.*, 26(2):166–168.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, pages 841–890.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261.

Chudik, A. and Pesaran, H. (2013). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *CESIFO WORKING PAPER NO. 4232*.

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, 14(1):C45–C90.

Dhaene, G. and Jochmans, K. (2010). Split-panel jackknife estimation of fixed-effect models. *Unpublished manuscript*.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fernández-Val, I. and Weidner, M. (2013). Individual and time effects in nonlinear panel data models with large N, T. *CeMMAP working paper series*.

Geman, S. (1980). A limit theorem for the norm of random matrices. *Annals of Probability*, 8(2):252–261.

Gobillon, L. and Magnac, T. (2013). Regional policy evaluation: Interactive fixed effects and synthetic controls. *IZA Discussion Paper No. 7493*.

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

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27(6):1152.

Hahn, J. and Moon, H. R. (2006). Reducing bias of MLE in a dynamic panel model. *Econometric Theory*, 22(03):499–512.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

Harding, M. (2007). Structural estimation of high-dimensional factor models. *unpublished manuscript*.

Harding, M. and Lamarche, C. (2009). A quantile regression approach for estimating panel data models using instrumental variables. *Economics Letters*, 104(3):133–135.

Harding, M. and Lamarche, C. (2011). Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. *Economics Letters*, 111(3):197–199.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56(6):1371–95.

Kapetanios, G., Pesaran, M. H., and Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics*, 160(2):326–348.

Kato, T. (1980). *Perturbation Theory for Linear Operators*. Springer-Verlag.

Latala, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.*, 133:1273–1282.

Lee, N., Moon, H. R., and Weidner, M. (2012). Analysis of interactive fixed effects dynamic linear panel regression with measurement error. *Economics Letters*, 117(1):239–242.

Moon, H., Shum, M., and Weidner, M. (2012). Interactive fixed effects in the blp random coefficients demand model. *CeMMAP working paper series*.

Moon, H. and Weidner, M. (2013). Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects. *CeMMAP working paper series*.

Moon, H. R. and Perron, B. (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics*, 122(1):81–126.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Onatski, A. (2005). Determining the number of factors from empirical distribution of eigenvalues. Discussion Papers 0405-19, Columbia University, Department of Economics.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Phillips, P. C. B. and Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal*, 6(1):217–259.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.

Silverstein, J. W. (1989). On the eigenvectors of large dimensional sample covariance matrices. *J. Multivar. Anal.*, 30(1):1–16.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.

Yin, Y. Q., Bai, Z. D., and Krishnaiah, P. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probability Theory Related Fields*, 78:509–521.

Appendix

A Examples of Error Distributions

Under each of the following distributional assumptions on the errors e_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$, we have $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)})$. The proofs are given in the supplementary material.

(i) The e_{it} are independent across i and t, they satisfy $\mathbb{E}e_{it} = 0$, and $\mathbb{E}e_{it}^4$ is bounded uniformly over i, t and N, T.

(ii) The e_{it} follow different MA(∞) process for each i, namely

$$e_{it} = \sum_{\tau=0}^{\infty} \psi_{i\tau}\, u_{i,t-\tau}\ , \qquad \text{for } i = 1 \ldots N,\ t = 1 \ldots T\ , \tag{A.1}$$

where the u_{it}, $i = 1 \ldots N$, $t = -\infty \ldots T$ are independent random variables with $\mathbb{E}u_{it} = 0$ and $\mathbb{E}u_{it}^4$ uniformly bounded across i, t and N, T. The coefficients $\psi_{i\tau}$ satisfy

$$\sum_{\tau=0}^{\infty} \tau \max_{i=1\ldots N} \psi_{i\tau}^2 < B\ , \qquad\qquad \sum_{\tau=0}^{\infty} \max_{i=1\ldots N} |\psi_{i\tau}| < B\ , \tag{A.2}$$

for a finite constant B which is independent of N and T.

(iii) The error matrix e is generated as $e = \sigma^{1/2}\, u\, \Sigma^{1/2}$, where u is an $N \times T$ matrix with independently distributed entries u_{it} and $\mathbb{E}u_{it} = 0$, $\mathbb{E}u_{it}^2 = 1$, and $\mathbb{E}u_{it}^4$ is bounded uniformly across i, t and N, T. Here σ is the $N \times N$ cross-sectional covariance matrix, and Σ is $T \times T$ time-serial covariance matrix, and they satisfy

$$\max_{j=1\ldots N} \sum_{i=1}^{N} |\sigma_{ij}| < B\ , \qquad\qquad \max_{\tau=1\ldots T} \sum_{t=1}^{T} |\Sigma_{t\tau}| < B\ , \tag{A.3}$$

for some finite constant B which is independent of N and T. In this example we have $\mathbb{E}e_{it}e_{j\tau} = \sigma_{ij}\Sigma_{t\tau}$.

B Proof of Identification (Theorem 2.1)

Proof of Theorem 2.1. Let $Q(\beta, \lambda, f) \equiv \mathbb{E}\left(\|Y - \beta \cdot X - \lambda\, f'\|_F^2\right)$, where $\beta \in \mathbb{R}^K$, $\lambda \in \mathbb{R}^{N\times R}$ and $f \in \mathbb{R}^{T\times R}$. We have

$$\begin{aligned}
&Q(\beta, \lambda, f) \\
&= \mathbb{E}\left\{\mathrm{Tr}\left[\left(Y - \beta \cdot X - \lambda\, f'\right)'\left(Y - \beta \cdot X - \lambda\, f'\right)\right] \middle| \lambda^0, f^0, w\right\} \\
&= \mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta - \beta^0) \cdot X + e\right)'\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta - \beta^0) \cdot X + e\right)\right] \middle| \lambda^0, f^0, w\right\} \\
&= \mathbb{E}\left[\mathrm{Tr}\left(e'e\right) \middle| \lambda^0, f^0, w\right] \\
&\quad + \underbrace{\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta - \beta^0) \cdot X\right)'\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta - \beta^0) \cdot X\right)\right] \middle| \lambda^0, f^0, w\right\}}_{\equiv Q^*(\beta,\lambda,f)}.
\end{aligned}$$

In the last step we used Assumption ID(ii). Since $\mathbb{E}\left[\mathrm{Tr}\left(e'e\right)\middle|\lambda^0, f^0, w\right]$ is independent of β, λ, f, we find that minimizing $Q(\beta,\lambda,f)$ is equivalent to minimizing $Q^*(\beta,\lambda,f)$. We decompose $Q^*(\beta,\lambda,f)$ as follows

$$
\begin{aligned}
&Q^*(\beta,\lambda,f) \\
&= \mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta-\beta^0)\cdot X\right)'\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta-\beta^0)\cdot X\right)\right]\middle|\lambda^0, f^0, w\right\} \\
&= \mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta-\beta^0)\cdot X\right)' M_{(\lambda,\lambda^0,w)}\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta-\beta^0)\cdot X\right)\right]\middle|\lambda^0, f^0, w\right\} \\
&\quad + \mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta-\beta^0)\cdot X\right)' P_{(\lambda,\lambda^0,w)}\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta-\beta^0)\cdot X\right)\right]\middle|\lambda^0, f^0, w\right\} \\
&= \underbrace{\mathbb{E}\left\{\mathrm{Tr}\left[\left((\beta_{\text{high}}-\beta^0_{\text{high}})\cdot X_{\text{high}}\right)' M_{(\lambda,\lambda^0,w)}\left((\beta_{\text{high}}-\beta^0_{\text{high}})\cdot X_{\text{high}}\right)\right]\middle|\lambda^0, f^0, w\right\}}_{\equiv Q^{\text{high}}(\beta_{\text{high}},\lambda)} \\
&\quad + \underbrace{\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta-\beta^0)\cdot X\right)' P_{(\lambda,\lambda^0,w)}\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta-\beta^0)\cdot X\right)\right]\middle|\lambda^0, f^0, w\right\}}_{\equiv Q^{\text{low}}(\beta,\lambda,f)},
\end{aligned}
$$

where $(\beta_{\text{high}}-\beta^0_{\text{high}})\cdot X_{\text{high}} = \sum_{m=K_1+1}^{K}(\beta_m-\beta^0_m)X_m$. A lower bound on $Q^{\text{high}}(\beta_{\text{high}},\lambda)$ is given by

$$
\begin{aligned}
&Q^{\text{high}}(\beta_{\text{high}},\lambda) \\
&\geq \min_{\widetilde{\lambda}\in\mathbb{R}^{N\times(R+R+\text{rank}(w))}} \mathbb{E}\left\{\mathrm{Tr}\left[\left((\beta_{\text{high}}-\beta^0_{\text{high}})\cdot X_{\text{high}}\right)' M_{(\widetilde{\lambda},\lambda,w)}\left((\beta_{\text{high}}-\beta^0_{\text{high}})\cdot X_{\text{high}}\right)\right]\middle|\lambda^0, f^0, w\right\} \\
&= \sum_{r=R+R+\text{rank}(w)}^{\min(N,T)} \mathbb{E}\left[\left((\beta_{\text{high}}-\beta^0_{\text{high}})\cdot X_{\text{high}}\right)\left((\beta_{\text{high}}-\beta^0_{\text{high}})\cdot X_{\text{high}}\right)'\middle|\lambda^0, f^0, w\right]. \qquad \text{(B.1)}
\end{aligned}
$$

Since $Q^*(\beta,\lambda,f)$, $Q^{\text{high}}(\beta_{\text{high}},\lambda)$, and $Q^{\text{low}}(\beta,\lambda,f)$, are expectations of traces of positive semi-definite matrices we have $Q^*(\beta,\lambda,f)\geq 0$, $Q^{\text{high}}(\beta_{\text{high}},\lambda)\geq 0$, and $Q^{\text{low}}(\beta,\lambda,f)\geq 0$ for all β, λ, f. Let $\bar{\beta}$, $\bar{\lambda}$ and $\bar{f}$ be the parameter values that minimize $Q(\beta,\lambda,f)$, and thus also $Q^*(\beta,\lambda,f)$. Since $Q^*(\beta^0,\lambda^0,f^0)=0$ we have $Q^*(\bar{\beta},\bar{\lambda},\bar{f}) = \min_{\beta,\lambda,f} Q^*(\beta,\lambda,f) = 0$. This implies $Q^*(\bar{\beta},\bar{\lambda},\bar{f})$, and thus also $Q^{\text{high}}(\bar{\beta}_{\text{high}},\bar{\lambda})=0$, and $Q^{\text{low}}(\bar{\beta},\bar{\lambda},\bar{f})=0$. Assumption ID($v$), the lower bound (B.1), and $Q^{\text{high}}(\bar{\beta}_{\text{high}},\bar{\lambda})=0$ imply that $\bar{\beta}_{\text{high}}=\beta^0_{\text{high}}$. Using this we find that

$$
\begin{aligned}
&Q^{\text{low}}(\bar{\beta},\bar{\lambda},\bar{f}) \\
&= \mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \bar{\lambda}\bar{f}' - (\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}}\right)'\left(\lambda^0 f^{0\prime} - \bar{\lambda}\bar{f}' - (\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}}\right)\right]\middle|\lambda^0, f^0, w\right\}, \\
&\geq \min_f \mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \bar{\lambda} f' - (\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}}\right)'\left(\lambda^0 f^{0\prime} - \bar{\lambda} f' - (\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}}\right)\right]\middle|\lambda^0, f^0, w\right\} \\
&= \mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}}\right)' M_{\bar{\lambda}}\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}}\right)\right]\middle|\lambda^0, f^0, w\right\}, \qquad \text{(B.2)}
\end{aligned}
$$

where $(\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}} = \sum_{l=1}^{K_1}(\bar{\beta}_l-\beta^0_l)X_l$. Since $Q^{\text{low}}(\bar{\beta},\bar{\lambda},\bar{f})=0$ and the last expression in (B.2) is non-negative we must have

$$
\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}}\right)' M_{\bar{\lambda}}\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}}-\beta^0_{\text{low}})\cdot X_{\text{low}}\right)\right]\middle|\lambda^0, f^0, w\right\} = 0.
$$

Since $\text{rank}(\bar{\lambda}) \leq R$ this implies that

$$\text{rank}\left\{\mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right)\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right)' \middle| \lambda^0, f^0, w\right]\right\} \leq R.$$

We furthermore find

$$\begin{aligned}
R &\geq \text{rank}\left\{\mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right)\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right)' \middle| \lambda^0, f^0, w\right]\right\} \\
&\geq \text{rank}\left\{M_w \mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right) P_{f^0}\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right)' M_w \middle| \lambda^0, f^0, w\right]\right\} \\
&\quad + \text{rank}\left\{P_w \mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right) M_{f^0}\left(\lambda^0 f^{0\prime} - (\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right)' P_w \middle| \lambda^0, f^0, w\right]\right\} \\
&\geq \text{rank}\left[M_w \lambda^0 f^{0\prime} f^0 \lambda^{0\prime} M_w\right] \\
&\quad + \text{rank}\left\{\mathbb{E}\left[\left((\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right) M_{f^0}\left((\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right)' \middle| \lambda^0, f^0, w\right]\right\}.
\end{aligned}$$

Assumption ID(iv) guarantees that $\text{rank}\left(M_w \lambda^0 f^{0\prime} f^0 \lambda^{0\prime} M_w\right) = \text{rank}\left(\lambda^0 f^{0\prime} f^0 \lambda^{0\prime}\right) = R$, i.e. we must have

$$\mathbb{E}\left[\left((\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right) M_{f^0}\left((\bar{\beta}_{\text{low}} - \beta^0_{\text{low}}) \cdot X_{\text{low}}\right)' \middle| \lambda^0, f^0, w\right] = 0.$$

According to Assumption ID(iii) this implies $\bar{\beta}_{\text{low}} = \beta^0_{\text{low}}$, i.e. we have $\bar{\beta} = \beta^0$. This also implies $Q^*(\bar{\beta}, \bar{\lambda}, \bar{f}) = \|\lambda^0 f^{0\prime} - \bar{\lambda}\bar{f}'\|_F^2 = 0$, and thereofere $\bar{\lambda}\bar{f}' = \lambda^0 f^{0\prime}$. ∎

C Proof of Consistency (Theorem 3.1)

The following theorem is useful for the consistency proof and beyond.

Theorem C.1. *Let N, T, R, R_1 and R_2 be positive integers such that $R \leq N$, $R \leq T$, and $R = R_1 + R_2$. Let Z be an $N \times T$ matrix, λ be an $N \times R$, f be a $T \times R$ matrix, $\widetilde{\lambda}$ be an $N \times R_1$ matrix, and $\widetilde{f}$ be a $T \times R_2$ matrix. Then the following six expressions (that are functions of Z only) are equivalent:*

$$\begin{aligned}
&\min_{f,\lambda} \text{Tr}\left[\left(Z - \lambda f'\right)\left(Z' - f\lambda'\right)\right] = \min_{f} \text{Tr}(Z\, M_f\, Z') = \min_{\lambda} \text{Tr}(Z'\, M_\lambda\, Z) \\
&= \min_{\widetilde{\lambda}, \widetilde{f}} \text{Tr}(M_{\widetilde{\lambda}}\, Z\, M_{\widetilde{f}}\, Z') = \sum_{i=R+1}^{T} \mu_i(Z'Z) = \sum_{i=R+1}^{N} \mu_i(ZZ')
\end{aligned}$$

In the above minimization problems we do not have to restrict the matrices λ, f, $\widetilde{\lambda}$ and $\widetilde{f}$ to be of full rank. If for example λ is not of full rank we can still define $(\lambda'\lambda)^{-1}$ as a generalized inverse. The projector M_λ is therefore still defined in this case and satisfied $M_\lambda \lambda = 0$ and $\text{rank}(M_\lambda) = N - \text{rank}(\lambda)$. If $\text{rank}(Z) \geq R$ then the optimal λ, f, $\widetilde{\lambda}$ and $\widetilde{f}$ always have full rank.

Theorem C.1 shows the equivalence of the three different versions of the profile objective function in equation (3.3). It goes beyond this by also considering minimization of $\text{Tr}(M_{\widetilde{\lambda}}\, Z\, M_{\widetilde{f}}\, Z')$ over $\widetilde{\lambda}$ and $\widetilde{f}$, which will be used in the consistency proof below. The proof of the theorem is given in the supplementary material. The following lemma is due to Bai (2009).

Lemma C.2. *Under the assumptions of Theorem 3.1 we have*

$$\sup_f \left|\frac{\text{Tr}(X_k\, M_f\, e')}{NT}\right| = o_p(1)\,, \quad \sup_f \left|\frac{\text{Tr}(\lambda^0\, f^{0\prime}\, M_f\, e')}{NT}\right| = o_p(1)\,, \quad \sup_f \left|\frac{\text{Tr}(e\, P_f\, e')}{NT}\right| = o_p(1)\,,$$

where the parameters f are $T \times R$ matrices with $\text{rank}(f) = R$.

Proof. By Assumption 2 we know that the first equation in Lemma C.2 is satisfied when replacing M_f by the identity matrix. So we are left to show $\max_f \left|\frac{1}{NT}\text{Tr}(\Xi\, e')\right| = o_p(1)$, where Ξ is either $X_k P_f$, $\lambda^0 f^{0\prime} M_f$, or eP_f. In all three cases we have $\|\Xi\|/\sqrt{NT} = \mathcal{O}_p(1)$ by Assumption 1, 3, and 4, respectively, and we have $\text{rank}(\Xi) \leq R$. We therefore find[21]

$$\sup_f \left|\frac{1}{NT}\text{Tr}(\Xi\, P_f\, e')\right| \leq R\, \frac{\|e\|}{\sqrt{NT}}\, \frac{\|\Xi\|}{\sqrt{NT}} = o_p(1)\ .$$

∎

Proof of Theorem 3.1. For the second version of the profile objective function in equation (3.3) we write $L_{NT}(\beta) = \min_f\ S_{NT}(\beta, f)$, where

$$S_{NT}(\beta, f) = \frac{1}{NT}\ \text{Tr}\left[\left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k + e\right)\ M_f\ \left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k + e\right)'\right],$$

We have $S_{NT}(\beta^0, f^0) = \frac{1}{NT}\ \text{Tr}\left(e\, M_{f^0}\, e'\right)$. Using Lemma (C.2) we find that

$$\begin{aligned} S_{NT}(\beta, f) &= S_{NT}(\beta^0, f^0) + \widetilde{S}_{NT}(\beta, f) \\ &\quad + \frac{2}{NT}\ \text{Tr}\left[\left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k\right)\ M_f\, e'\right] + \frac{1}{NT}\ \text{Tr}\left(e\,(P_{f^0} - P_f)\, e'\right) \\ &= S_{NT}(\beta^0, f^0) + \widetilde{S}_{NT}(\beta, f) + o_p(\|\beta - \beta^0\|) + o_p(1)\ , \end{aligned} \tag{C.1}$$

where we defined

$$\widetilde{S}_{NT}(\beta, f) = \frac{1}{NT}\ \text{Tr}\left[\left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k\right)\ M_f\ \left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k\right)'\right].$$

Up to this point the consistency proof is almost equivalent to the one given in Bai (2009), but the remainder of the proof differs from Bai, since we allow for more general low-rank regressors, and since we allow for high-rank and low-rank regressors simultaneously. We split $\widetilde{S}_{NT}(\beta, f) = \widetilde{S}^{(1)}_{NT}(\beta, f) + \widetilde{S}^{(2)}_{NT}(\beta, f)$, where

$$\begin{aligned} \widetilde{S}^{(1)}_{NT}(\beta, f) &= \frac{1}{NT}\ \text{Tr}\left[\left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k\right)\ M_f\ \left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k\right)'\ M_{(\lambda_0, w)}\right] \\ &= \frac{1}{NT}\ \text{Tr}\left[\left(\sum_{m=K_1+1}^{K}(\beta_m^0 - \beta_m)X_m\right)\ M_f\ \left(\sum_{m=K_1+1}^{K}(\beta_m^0 - \beta_m)X_m\right)'\ M_{(\lambda_0, w)}\right], \\ \widetilde{S}^{(2)}_{NT}(\beta, f) &= \frac{1}{NT}\ \text{Tr}\left[\left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k\right)\ M_f\ \left(\lambda^0\, f^{0\prime} + \sum_{k=1}^{K}(\beta_k^0 - \beta_k)X_k\right)'\ P_{(\lambda_0, w)}\right], \end{aligned}$$

[21] Here we use $|\text{Tr}\,(C)| \leq \|C\|\,\text{rank}\,(C)$, which holds for all square matrices C, see the supplementary material.

and (λ_0, w) is the $N \times (R + K_1)$ matrix that is composed out of λ_0 and the $N \times K_1$ matrix w defined in Assumption 4. For $\widetilde{S}^{(1)}_{NT}(\beta, f)$ we can apply Theorem C.1 with $\widetilde{f} = f$ and $\widetilde{\lambda} = (\lambda^0, w)$ (the R in the theorem is now $2R + K_1$) to find

$$\begin{aligned}\widetilde{S}^{(1)}_{NT}(\beta, f) &\geq \frac{1}{NT} \sum_{i=2R+K_1+1}^{N} \mu_i \left[\left(\sum_{m=K_1+1}^{K} (\beta^0_m - \beta_m) X_m \right) \left(\sum_{m=K_1+1}^{K} (\beta^0_m - \beta_m) X_m \right)' \right] \\ &\geq b \left\| \beta^{\text{high}} - \beta^{\text{high}}_0 \right\|^2 , \qquad \text{wpa1,} \end{aligned} \tag{C.2}$$

where in the last step we used the existence of a constant $b > 0$ guaranteed by Assumption 4(ii)(a), and we introduced $\beta^{\text{high}} = (\beta_{K_1+1}, \ldots, \beta_K)'$, which refers to the $K_2 \times 1$ parameter vector corresponding to the high-rank regressors. Similarly we define $\beta^{\text{low}} = (\beta_1, \ldots, \beta_{K_1})'$ for the $K_1 \times 1$ parameter vector of low-rank regressors.

Using $P_{(\lambda_0,w)} = P_{(\lambda_0,w)} P_{(\lambda_0,w)}$ and the cyclicality of the trace we see that $\widetilde{S}^{(2)}_{NT}(\beta, f)$ can be written as the trace of a positive definite matrix, and therefore $\widetilde{S}^{(2)}_{NT}(\beta, f) \geq 0$. Note also that we can choose $\beta = \beta^0$ and $f = f^0$ in the minimization problem over $S_{NT}(\beta, f)$, i.e. the optimal $\beta = \widehat{\beta}$ and $f = \widehat{f}$ must satisfy $S_{NT}(\widehat{\beta}, \widehat{f}) \leq S_{NT}(\beta^0, f^0)$. Using this, equation (C.1), $\widetilde{S}^{(2)}_{NT}(\beta, f) \geq 0$, and the bound in (C.2) we find

$$0 \;\geq\; b \left\| \widehat{\beta}^{\text{high}} - \beta^{\text{high}}_0 \right\|^2 + o_p\left(\left\| \widehat{\beta}^{\text{high}} - \beta^{\text{high}}_0 \right\| \right) + o_p\left(\left\| \widehat{\beta}^{\text{low}} - \beta^{\text{low}}_0 \right\| \right) + o_p(1) \, .$$

Since we assume that $\widehat{\beta}^{\text{low}}$ is bounded, the last equation implies that $\left\| \widehat{\beta}^{\text{high}} - \beta^{\text{high}}_0 \right\| = o_p(1)$, i.e. $\widehat{\beta}^{\text{high}}$ is consistent. What is left to show is that $\widehat{\beta}^{\text{low}}$ is consistent, too. In the supplementary material we show that Assumption 4(ii)(b) guarantees that there exist finite positive constants a_0, a_1, a_2, a_3 and a_4 such that

$$\begin{aligned}\widetilde{S}^{(2)}_{NT}(\beta, f) \geq \; & \frac{a_0 \left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\|^2}{\left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\|^2 + a_1 \left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\| + a_2} \\ & - a_3 \left\| \beta^{\text{high}} - \beta^{\text{high}}_0 \right\| - a_4 \left\| \beta^{\text{high}} - \beta^{\text{high}}_0 \right\| \left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\| , \qquad \text{wpa1.} \end{aligned}$$

Using consistency of $\widehat{\beta}^{\text{high}}$ and again boundedness of β^{low} this implies that there exists $a > 0$ such that $\widetilde{S}^{(2)}_{NT}(\widehat{\beta}, f) \geq a \left\| \widehat{\beta}^{\text{low}} - \beta^{\text{low}}_0 \right\|^2 + o_p(1)$. With the same argument as for $\widehat{\beta}^{\text{high}}$ we therefore find $\left\| \widehat{\beta}^{\text{low}} - \beta^{\text{low}}_0 \right\| = o_p(1)$, i.e. $\widehat{\beta}^{\text{low}}$ is consistent. This is what we wanted to show. ∎

D Proof of Limiting Distribution (Theorem 4.3)

Theorem 4.1 and Corollary 4.2 are from Moon and Weidner (2013), and the proof can be found there. Note that Assumption 4(i) implies $\|X_k\| = \mathcal{O}_p(\sqrt{NT})$, which is assumed in Moon and Weidner (2013). There it is also assumed that $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)}) = \mathcal{O}_p(\sqrt{N})$, while we assume $\|e\| = o_p(\|N^{2/3}\|)$. It is, however, straightforward to verify that the proof of Theorem 4.1 is also valid under this weaker assumption. Moon and Weidner (2013) also employs different consistency assumptions than are demanded in Corollary 4.2, which is not important for the proof of the corollary, since only consistency result itself enters into the proof. In the supplementary

material we show that the assumptions of Corollary 4.2 already guarantee that W_{NT} does not become singular as $N, T \to \infty$.

For each $k = 1, \ldots, K$ we define the $N \times T$ matrices $\overline{X}_k$, $\widetilde{X}_k$ and $\mathfrak{X}_k$ as follows

$$\overline{X}_k \equiv \mathbb{E}\left(X_k \middle| \mathcal{C}\right), \qquad \widetilde{X}_k \equiv X_k - \mathbb{E}\left(X_k \middle| \mathcal{C}\right), \qquad \mathfrak{X}_k \equiv M_{\lambda^0}\, \overline{X}_k\, M_{f^0} + \widetilde{X}_k.$$

Note the difference between $\mathfrak{X}_k$ and $\mathcal{X}_k = M_{\lambda^0}\, X_k\, M_{f^0}$, which was defined in Assumption 6. In particular, conditional on $\mathcal{C}$, the elements $\mathfrak{X}_{k,it}$ of $\mathfrak{X}_k$ are contemporaneously independent of the error term e_{it}, while the same is not true for $\mathcal{X}_k$.

To present the proof of Theorem 4.3 it is convenient to first establish two technical lemmas.

Lemma D.1. *Under the assumptions of Theorem 4.3 we have*

$$(a) \qquad \frac{1}{\sqrt{NT}}\mathrm{Tr}\left(P_{f^0}\, e'\, P_{\lambda^0}\, \widetilde{X}_k\right) = o_p(1)\,,$$

$$(b) \qquad \frac{1}{\sqrt{NT}}\mathrm{Tr}\left(P_{\lambda^0}\, e\, \widetilde{X}_k'\right) = o_p(1)\,,$$

$$(c) \qquad \frac{1}{\sqrt{NT}}\mathrm{Tr}\left\{P_{f^0}\, \left[e'\, \widetilde{X}_k - \mathbb{E}\left(e'\, \widetilde{X}_k \,|\, \mathcal{C}\right)\right]\right\} = o_p(1)\,,$$

$$(d) \qquad \frac{1}{\sqrt{NT}}\mathrm{Tr}\left(eP_{f^0}\, e'\, M_{\lambda^0}\, X_k\, f^0\, (f^{0\prime}f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right) = o_p(1)\,,$$

$$(e) \qquad \frac{1}{\sqrt{NT}}\mathrm{Tr}\left(e'\, P_{\lambda^0}\, e\, M_{f^0}\, X_k'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime}f^0)^{-1}\, f^{0\prime}\right) = o_p(1)\,,$$

$$(f) \qquad \frac{1}{\sqrt{NT}}\mathrm{Tr}\left(e'M_{\lambda^0}\, X_k\, M_{f^0}\, e'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime}f^0)^{-1}\, f^{0\prime}\right) = o_p(1)\,,$$

$$(g) \qquad \frac{1}{\sqrt{NT}}\mathrm{Tr}\left\{\left[ee' - \mathbb{E}\left(ee'\right)\right]\, M_{\lambda^0}\, X_k\, f^0\, (f^{0\prime}f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\} = o_p(1)\,,$$

$$(h) \qquad \frac{1}{\sqrt{NT}}\mathrm{Tr}\left\{\left[e'e - \mathbb{E}\left(e'e\right)\right]\, M_{f^0}\, X_k'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime}f^0)^{-1}\, f^{0\prime}\right\} = o_p(1)\,,$$

$$(i) \qquad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbb{E}\left(e_{it}^2\right)\, \mathfrak{X}_{it}\, \mathfrak{X}_{it}' - \mathbb{E}\left(e_{it}^2\, \mathfrak{X}_{it}\, \mathfrak{X}_{it}' \,|\, \mathcal{C}\right)\right] = o_p(1),$$

$$(j) \qquad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(e_{it}^2\right)\left(\mathfrak{X}_{it}\, \mathfrak{X}_{it}' - \mathcal{X}_{it}\, \mathcal{X}_{it}'\right) = o_p(1).$$

Lemma D.2. *Under the assumptions of Theorem 4.3 we have*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\mathfrak{X}_{it} \underset{d}{\to} \mathcal{N}\left(0, \Omega\right).$$

The proofs of Lemma D.1 and Lemma D.2 are provided in the supplementary material. Regarding Lemma D.2, note that since $e_{it}\mathfrak{X}_{it}$ is mean zero and uncorrelated across both i and t,

conditional on $\mathcal{C}$, we have

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\mathfrak{X}_{it}\,\middle|\,\mathcal{C}\right) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(e_{it}^2\,\mathfrak{X}_{it}\,\mathfrak{X}_{it}'\,|\,\mathcal{C}\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(e_{it}^2\right)\,\mathfrak{X}_{it}\,\mathfrak{X}_{it}' + o_p(1) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(e_{it}^2\right)\,\mathcal{X}_{it}\,\mathcal{X}_{it}' + o_p(1) \\
&= \Omega + o_p(1), \qquad \text{(D.1)}
\end{aligned}$$

where we also used part (i) and (j) of Lemma D.1, and the definition of Ω in Assumptions 5. Note that Ω is a constant, which implies that the probability limit of $\operatorname{Var}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\mathfrak{X}_{it}\,|\,\mathcal{C}\right]$ is independent of $\mathcal{C}$. This explains why the asymptotic variance-covariance matrix in Lemma D.2 turns out to be Ω.

Using those lemmas we can now prove the theorem on the limiting distribution of $\widehat{\beta}$ in the main text.

Proof of Theorem 4.3. We have $\|e\| = \mathcal{O}_p(N^{1/2})$, i.e. Assumption 3* is satisfied. We can therefore apply Corollary 4.2 to calculate the limiting distribution of $\widehat{\beta}$. Note that $\mathcal{X}_k = \mathfrak{X}_k - \widetilde{X}_k\,P_{f^0} - P_{\lambda^0}\,\widetilde{X}_k + P_{\lambda^0}\,\widetilde{X}_k\,P_{f^0}$. Using Lemmas D.1 and D.2 and Assumption 6 we find

$$\begin{aligned}
\frac{1}{\sqrt{NT}}C^{(1)}\left(\lambda^0,\,f^0,\,X_k,\,e\right) =& \frac{1}{\sqrt{NT}}\operatorname{Tr}\left(M_{f^0}\,e'\,M_{\lambda^0}\,X_k\right) \\
=& \frac{1}{\sqrt{NT}}\operatorname{Tr}\left(e'\mathfrak{X}_k\right) - \frac{1}{\sqrt{NT}}\operatorname{Tr}\left[P_{f^0}\,\mathbb{E}\left(e'\,\widetilde{X}_k\,|\,\mathcal{C}\right)\right] \\
& - \frac{1}{\sqrt{NT}}\operatorname{Tr}\left(e'\,P_{\lambda^0}\,\widetilde{X}_k\right) + \frac{1}{\sqrt{NT}}\operatorname{Tr}\left(P_{f^0}\,e'\,P_{\lambda^0}\,\widetilde{X}_k\right) \\
& - \frac{1}{\sqrt{NT}}\operatorname{Tr}\left\{P_{f^0}\,\left[e'\,\widetilde{X}_k - \mathbb{E}\left(e'\,\widetilde{X}_k\,|\,\mathcal{C}\right)\right]\right\} \\
=& \frac{1}{\sqrt{NT}}\operatorname{Tr}\left(e'\mathfrak{X}_k\right) - \frac{1}{\sqrt{NT}}\operatorname{Tr}\left[P_{f^0}\,\mathbb{E}\left(e'\,X_k\,|\,\mathcal{C}\right)\right] + o_p(1)\,. \\
\underset{d}{\to}& \mathcal{N}\left(-\kappa B_1,\,\Omega\right)\,,
\end{aligned}$$

where we also used that $\mathbb{E}\left(e'\,\widetilde{X}_k \,\middle|\, \mathcal{C}\right) = \mathbb{E}\left(e'\,X_k \,\middle|\, \mathcal{C}\right)$. Using Lemmas D.1 we also find

$$\begin{aligned}
\frac{1}{\sqrt{NT}}C^{(2)}\left(\lambda^0,\, f^0,\, X_k,\, e\right) =& -\frac{1}{\sqrt{NT}}\Big[\mathrm{Tr}\left(eM_{f^0}\,e'\,M_{\lambda^0}\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right) \\
&\quad + \mathrm{Tr}\left(e'M_{\lambda^0}\,e\,M_{f^0}\,X_k'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right) \\
&\quad + \mathrm{Tr}\left(e'M_{\lambda^0}\,X_k\,M_{f^0}\,e'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right)\Big] \\
=& \frac{1}{\sqrt{NT}}\mathrm{Tr}\left(eP_{f^0}\,e'\,M_{\lambda^0}\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right) \\
&- \frac{1}{\sqrt{NT}}\mathrm{Tr}\left\{\left[ee' - \mathbb{E}\left(ee'\right)\right]\,M_{\lambda^0}\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right\} \\
&- \frac{1}{\sqrt{NT}}\mathrm{Tr}\left[\mathbb{E}\left(ee'\right)\,M_{\lambda^0}\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right] \\
&+ \frac{1}{\sqrt{NT}}\mathrm{Tr}\left(e'P_{\lambda^0}\,e\,M_{f^0}\,X_k'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right) \\
&- \frac{1}{\sqrt{NT}}\mathrm{Tr}\left\{\left[e'e - \mathbb{E}\left(e'e\right)\right]\,M_{f^0}\,X_k'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right\} \\
&- \frac{1}{\sqrt{NT}}\mathrm{Tr}\left[\mathbb{E}\left(e'e\right)\,M_{f^0}\,X_k'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right] \\
&+ \frac{1}{\sqrt{NT}}\mathrm{Tr}\left(e'M_{\lambda^0}\,X_k\,M_{f^0}\,e'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right) \\
=& -\frac{1}{\sqrt{NT}}\mathrm{Tr}\left[\mathbb{E}\left(ee'\right)\,M_{\lambda^0}\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right] \\
&- \frac{1}{\sqrt{NT}}\mathrm{Tr}\left[\mathbb{E}\left(e'e\right)\,M_{f^0}\,X_k'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right] + o_p(1)\,, \\
=& -\kappa^{-1}B_2 - \kappa B_3 + o_p(1)\,,
\end{aligned}$$

Combining these results we obtain

$$\begin{aligned}
\sqrt{NT}\left(\widehat{\beta} - \beta^0\right) &= W_{NT}^{-1}\left(\frac{1}{\sqrt{NT}}C^{(1)} + \frac{1}{\sqrt{NT}}C^{(1)}\right)\,, \\
&\underset{d}{\to} \mathcal{N}\left(-W^{-1}\,\left(\kappa B_1 + \kappa^{-1}B_2 + \kappa B_3\right),\, W^{-1}\,\Omega\,W^{-1}\right)\,,
\end{aligned}$$

which is what we wanted to show. ∎

E Expansions of Projectors and Residuals

The incidental parameter estimators $\widehat{f}$ and $\widehat{\lambda}$ as well as the residuals $\widehat{e}$ enter into the asymptotic bias and variance estimators for the LS estimator $\widehat{\beta}$. To describe the properties of $\widehat{f}$, $\widehat{\lambda}$ and $\widehat{e}$, it is convenient to have asymptotic expansions of the projectors $M_{\widehat{\lambda}}(\beta)$ and $M_{\widehat{f}}(\beta)$ that correspond to the minimizing parameters $\widehat{\lambda}(\beta)$ and $\widehat{f}(\beta)$ in equation (3.3). Note that the minimizing $\widehat{\lambda}(\beta)$ and $\widehat{f}(\beta)$ can be defined for all values of β, not only for the optimal value $\beta = \widehat{\beta}$. The corresponding residuals are $\widehat{e}(\beta) = Y - \beta\cdot X - \widehat{\lambda}(\beta)\,\widehat{f}'(\beta)$.

Theorem E.1. *Under Assumption 1, 3, and 4(i) we have the following expansions*

$$M_{\widehat{\lambda}}(\beta) = M_{\lambda^0} + M^{(1)}_{\widehat{\lambda},e} + M^{(2)}_{\widehat{\lambda},e} - \sum_{k=1}^{K} \left(\beta_k - \beta^0_k\right) M^{(1)}_{\widehat{\lambda},k} + M^{(\text{rem})}_{\widehat{\lambda}}(\beta) \ ,$$

$$M_{\widehat{f}}(\beta) = M_{f^0} + M^{(1)}_{\widehat{f},e} + M^{(2)}_{\widehat{f},e} - \sum_{k=1}^{K} \left(\beta_k - \beta^0_k\right) M^{(1)}_{\widehat{f},k} + M^{(\text{rem})}_{\widehat{f}}(\beta) \ ,$$

$$\widehat{e}(\beta) = M_{\lambda^0}\, e\, M_{f^0} + \widehat{e}^{(1)}_{e} - \sum_{k=1}^{K} \left(\beta_k - \beta^0_k\right) \widehat{e}^{(1)}_{k} + \widehat{e}^{(\text{rem})}(\beta) \ ,$$

where the spectral norms of the remainders satisfy for any series $\eta_{NT} \to 0$

$$\sup_{\{\beta:\left\|\beta-\beta^0\right\|\leq\eta_{NT}\}} \frac{\left\|M^{(\text{rem})}_{\widehat{\lambda}}(\beta)\right\|}{\|\beta-\beta^0\|^2 + (NT)^{-1/2}\,\|e\|\,\|\beta-\beta^0\| \ + (NT)^{-3/2}\,\|e\|^3} = \mathcal{O}_p\left(1\right) \ ,$$

$$\sup_{\{\beta:\left\|\beta-\beta^0\right\|\leq\eta_{NT}\}} \frac{\left\|M^{(\text{rem})}_{\widehat{f}}(\beta)\right\|}{\|\beta-\beta^0\|^2 + (NT)^{-1/2}\,\|e\|\,\|\beta-\beta^0\| \ + (NT)^{-3/2}\,\|e\|^3} = \mathcal{O}_p\left(1\right) \ ,$$

$$\sup_{\{\beta:\left\|\beta-\beta^0\right\|\leq\eta_{NT}\}} \frac{\left\|\widehat{e}^{(\text{rem})}(\beta)\right\|}{(NT)^{1/2}\|\beta-\beta^0\|^2 + \|e\|\,\|\beta-\beta^0\| + (NT)^{-1}\|e\|^3} = \mathcal{O}_p\left(1\right) \ ,$$

and we have $\text{rank}(\widehat{e}^{(\text{rem})}(\beta)) \leq 7R$*, and the expansion coefficients are given by*

$$\begin{aligned}
M^{(1)}_{\widehat{\lambda},e} &= -\,M_{\lambda^0}\, e\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime} \ - \ \lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\,e'\,M_{\lambda^0} \ , \\
M^{(1)}_{\widehat{\lambda},k} &= -\,M_{\lambda^0}\, X_k\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime} \ - \ \lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\,X_k'\,M_{\lambda^0} \ , \\
M^{(2)}_{\widehat{\lambda},e} &= M_{\lambda^0}\, e\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime}\, e\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime} \\
&\quad + \lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\,e'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\,e'\,M_{\lambda^0} \\
&\quad - M_{\lambda^0}\, e\, M_{f^0}\, e'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime} \\
&\quad - \lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\, e\, M_{f^0}\, e'\, M_{\lambda^0} \\
&\quad - M_{\lambda^0}\, e\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\,e'\,M_{\lambda^0} \\
&\quad + \lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\,e'\,M_{\lambda^0}\, e\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime} \ ,
\end{aligned}$$

analogously

$$\begin{aligned}
M^{(1)}_{\widehat{f},e} &= \ -\,M_{f^0}\, e'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}f^{0\prime} \ - \ f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\, e\, M_{f^0} \ , \\
M^{(1)}_{\widehat{f},k} &= \ -\,M_{f^0}\, X_k'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}f^{0\prime} \ - \ f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\, X_k\, M_{f^0} \ , \\
M^{(2)}_{\widehat{f},e} &= M_{f^0}\, e'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}f^{0\prime}\, e'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}f^{0\prime} \\
&\quad + f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\, e\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\, e\, M_{f^0} \\
&\quad - M_{f^0}\, e'\, M_{\lambda^0}\, e\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime} \\
&\quad - f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\, e'\, M_{\lambda^0}\, e\, M_{f^0} \\
&\quad - M_{f^0}\, e'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\, e\, M_{f^0} \\
&\quad + f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\, e\, M_{f^0}\, e'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}f^{0\prime} \ ,
\end{aligned}$$

and finally

$$\begin{aligned}
\widehat{e}_k^{(1)} &= M_{\lambda^0}\, X_k\, M_{f^0}\ , \\
\widehat{e}_e^{(1)} &= -M_{\lambda^0}\, e\, M_{f^0}\, e'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime}f^0)^{-1}\, f^{0\prime} \\
&\quad - \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime}f^0)^{-1}\, f^{0\prime}\, e'\, M_{\lambda^0}\, e\, M_{f^0} \\
&\quad - M_{\lambda^0}\, e\, f^0\, (f^{0\prime}f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\, e\, M_{f^0}\ .
\end{aligned}$$

Proof. The general expansion of $M_{\widehat{\lambda}}(\beta)$ is given Moon and Weidner (2013), and in the theorem we just make this expansion explicit up to a particular order. The result for $M_{\widehat{f}}(\beta)$ is just obtained by symmetry ($N \leftrightarrow T$, $\lambda \leftrightarrow f$, $e \leftrightarrow e'$, $X_k \leftrightarrow X_k'$). For the residuals $\widehat{e}$ we have

$$\widehat{e} = M_{\widehat{\lambda}} \left(Y - \sum_{k=1} \widehat{\beta}_k\, X_k \right) = M_{\widehat{\lambda}} \left[e - \left(\widehat{\beta} - \beta^0\right) \cdot X + \lambda^0 f^{0\prime} \right]\ ,$$

and plugging in the expansion of $M_{\widehat{\lambda}}$ gives the expansion of $\widehat{e}$. We have $\widehat{e}(\beta) = A_0 + \lambda^0 f^{0\prime} - \widehat{\lambda}(\beta)\widehat{f}'(\beta)$, where $A_0 = e - \sum_k (\beta_k - \beta_k^0) X_k$. Therefore $\widehat{e}^{(\mathrm{rem})}(\beta) = A_1 + A_2 + A_3$ with $A_1 = A_0 - M_{\lambda^0}\, A_0\, M_{f^0}$, $A_2 = \lambda^0 f^{0\prime} - \widehat{\lambda}(\beta)\widehat{f}'(\beta)$, and $A_3 = -\widehat{e}_e^{(1)}$. We find $\mathrm{rank}(A_1) \leq 2R$, $\mathrm{rank}(A_2) \leq 2R$, $\mathrm{rank}(A_3) \leq 3R$, and thus $\mathrm{rank}(\widehat{e}^{(\mathrm{rem})}(\beta)) \leq 7R$, as stated in the theorem. ∎

Having expansions for $M_{\widehat{\lambda}}(\beta)$ and $M_{\widehat{f}}(\beta)$ we also have expansions for $P_{\widehat{\lambda}}(\beta) = \mathbb{I}_N - M_{\widehat{\lambda}}(\beta)$ and $P_{\widehat{f}}(\beta) = \mathbb{I}_T - M_{\widehat{f}}(\beta)$. The reason why we give expansions of the projectors and not expansions of $\widehat{\lambda}(\beta)$ and $\widehat{f}(\beta)$ directly is that for the latter we would need to specify a normalization, while the projectors are independent of any normalization choice. An expansion for $\widehat{\lambda}(\beta)$ can for example be defined by $\widehat{\lambda}(\beta) = P_{\widehat{\lambda}}(\beta)\lambda^0$, in which case the normalization of $\widehat{\lambda}(\beta)$ is implicitly defined by the normalization of λ^0.

F Consistency Proof for Bias and Variance Estimators (Theorem 4.4)

Corollary F.1. *Under the Assumptions of Theorem 4.3 we have* $\sqrt{NT}\left(\widehat{\beta} - \beta^0\right) = \mathcal{O}_p(1)$.

This corollary directly follows from Theorem 4.3.

Corollary F.2. *Under the Assumptions of Theorem 4.4 we have*

$$\begin{aligned}
\left\| P_{\widehat{\lambda}} - P_{\lambda^0} \right\| &= \left\| M_{\widehat{\lambda}} - M_{\lambda^0} \right\| = \mathcal{O}_p(N^{-1/2})\ , \\
\left\| P_{\widehat{f}} - P_{f^0} \right\| &= \left\| M_{\widehat{f}} - M_{f^0} \right\| = \mathcal{O}_p(T^{-1/2})\ .
\end{aligned}$$

Proof. Using $\|e\| = \mathcal{O}_p(N^{1/2})$ and $\|X_k\| = \mathcal{O}_p(N)$ we find that the expansion terms in Theorem E.1 satisfy

$$\left\| M_{\widehat{\lambda},e}^{(1)} \right\| = \mathcal{O}_p(N^{-1/2})\ , \qquad \left\| M_{\widehat{\lambda},e}^{(2)} \right\| = \mathcal{O}_p(N^{-1})\ , \qquad \left\| M_{\widehat{\lambda},k}^{(1)} \right\| = \mathcal{O}_p(1)\ .$$

Together with corollary F.1 the result for $\left\| M_{\widehat{\lambda}} - M_{\lambda^0} \right\|$ immediately follows. In addition we have $P_{\widehat{\lambda}} - P_{\lambda^0} = -M_{\widehat{\lambda}} + M_{\lambda^0}$. The proof for $M_{\widehat{f}}$ and $P_{\widehat{f}}$ is analogous. ∎

Lemma F.3. *Under the Assumptions of Theorem 4.4 we have*

$$A_0 \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2 \left(\mathfrak{X}_{it}\mathfrak{X}_{it}' - \mathcal{X}_{it}\mathcal{X}_{it}'\right) = o_p(1) \ ,$$

$$A_1 \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2 \left(\mathcal{X}_{it}\mathcal{X}_{it}' - \widehat{\mathcal{X}}_{it}\widehat{\mathcal{X}}_{it}'\right) = o_p(1) \ ,$$

$$A_2 \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \left(e_{it}^2 - \widehat{e}_{it}^2\right) \widehat{\mathcal{X}}_{it}\widehat{\mathcal{X}}_{it}' = o_p(1) \ .$$

Lemma F.4. *Let $\widehat{f}$ and f^0 be normalized as $\widehat{f}'\widehat{f}/T = \mathbb{I}_R$ and $f^{0\prime}f^0/T = \mathbb{I}_R$. Then, under the assumptions of Theorem 4.4, there exists an $R \times R$ matrices $H = H_{NT}$ such that*[22]

$$\left\|\widehat{f} - f^0 H\right\| = O_p(1) \ , \qquad \left\|\widehat{\lambda} - \lambda^0 \left(H'\right)^{-1}\right\| = O_p(1) \ .$$

Furthermore

$$\left\|\widehat{\lambda}\left(\widehat{\lambda}'\widehat{\lambda}\right)^{-1}\left(\widehat{f}'\widehat{f}\right)^{-1}\widehat{f}' - \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right\| = O_p\left(N^{-3/2}\right) \ .$$

Lemma F.5. *Under the Assumptions of Theorem 4.4 we have*

$$\text{(i)} \qquad N^{-1}\left\|\mathbb{E}(e'X_k \,|\, \mathcal{C}) - \left(\widehat{e}'\,X_k\right)^{\text{truncR}}\right\| = o_p(1) \ ,$$

$$\text{(ii)} \qquad N^{-1}\left\|\mathbb{E}(e'e) - \left(\widehat{e}'\,\widehat{e}\right)^{\text{truncD}}\right\| = o_p(1) \ ,$$

$$\text{(iii)} \qquad T^{-1}\left\|\mathbb{E}(ee') - \left(\widehat{e}\,\widehat{e}'\right)^{\text{truncD}}\right\| = o_p(1) \ .$$

Lemma F.6. *Under the Assumptions of Theorem 4.4 we have*

$$\text{(i)} \qquad N^{-1}\left\|\left(\widehat{e}'\,X_k\right)^{\text{truncR}}\right\| = \mathcal{O}_p(MT^{1/8}) \ ,$$

$$\text{(ii)} \qquad N^{-1}\left\|\left(\widehat{e}'\,\widehat{e}\right)^{\text{truncD}}\right\| = \mathcal{O}_p(1) \ ,$$

$$\text{(iii)} \qquad T^{-1}\left\|\left(\widehat{e}\,\widehat{e}'\right)^{\text{truncD}}\right\| = \mathcal{O}_p(1) \ .$$

The proof of the above lemmas is given in the supplementary material. Using these lemmas we can now prove Theorem 4.4.

Proof of Theorem 4.4, Part I: show $\widehat{W} = W + o_p(1)$.
Using $|\text{Tr}(C)| \leq \|C\| \,\text{rank}(C)$ and corollary F.2 we find

$$\begin{aligned}
&\left|\widehat{W}_{k_1k_2} - W_{NT,k_1k_2}\right| \\
&\quad = \left|(NT)^{-1}\text{Tr}\left[\left(M_{\widehat{\lambda}} - M_{\lambda^0}\right) X_{k_1} M_{\widehat{f}} X_{k_2}'\right] + (NT)^{-1}\text{Tr}\left[M_{\lambda^0} X_{k_1}\left(M_{\widehat{f}} - M_{f^0}\right) X_{k_2}'\right]\right. \\
&\quad \leq \frac{2R}{NT}\left\|M_{\widehat{\lambda}} - M_{\lambda^0}\right\| \|X_{k_1}\| \|X_{k_2}\| \frac{2R}{NT}\left\|M_{\widehat{f}} - M_{f^0}\right\| \|X_{k_1}\| \|X_{k_2}\| \\
&\quad = \frac{2R}{NT}\mathcal{O}_p(N^{-1})\mathcal{O}_p(NT) + \frac{2R}{NT}\mathcal{O}_p(T^{-1})\mathcal{O}_p(NT) \\
&\quad = o_p(1) \ .
\end{aligned}$$

[22] We consider a limit $N, T \to \infty$ and for different N, T different H-matrices can be chosen, but we write H instead of H_{NT} to keep notation simple.

Thus we have $\widehat{W} = W_{NT} + o_p(1) = W + o_p(1)$. ∎

Proof of Theorem 4.4, Part II: show $\widehat{\Omega} = \Omega + o_p(1)$.
Let $\Omega_{NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(e_{it}^2\right)\mathcal{X}_{it}\mathcal{X}_{it}'$ and $\Omega_{NT}^{*} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2\,\mathfrak{X}_{it}\,\mathfrak{X}_{it}'$. As already stated in equation (D.1) we can use part (i) and (j) of Lemma D.1 to show that $\Omega_{NT} = \mathbb{E}\left(\Omega_{NT}^{*}\,\middle|\,\mathcal{C}\right) + o_p(1)$. Using cross-sectional independence conditional on $\mathcal{C}$ we find that

$$
\begin{aligned}
&\operatorname{Var}\left(\Omega_{NT,k_1k_2}^{*}\,\middle|\,\mathcal{C}\right)\\
&= \frac{1}{(NT)^2}\sum_{i,j=1}^{N}\sum_{t,\tau=1}^{T}\Big[\mathbb{E}\left(e_{it}^2\mathfrak{X}_{k_1,it}\mathfrak{X}_{k_2,it}\,e_{j\tau}^2\mathfrak{X}_{k_1,j\tau}\mathfrak{X}_{k_2,j\tau}\,\middle|\,\mathcal{C}\right)\\
&\qquad\qquad - \mathbb{E}\left(e_{it}^2\mathfrak{X}_{k_1,it}\mathfrak{X}_{k_2,it}\,\middle|\,\mathcal{C}\right)\mathbb{E}\left(e_{j\tau}^2\mathfrak{X}_{k_1,j\tau}\mathfrak{X}_{k_2,j\tau}\,\middle|\,\mathcal{C}\right)\Big]\\
&= \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t,\tau=1}^{T}\Big[\mathbb{E}\left(e_{it}^2\mathfrak{X}_{k_1,it}\mathfrak{X}_{k_2,it}\,e_{i\tau}^2\mathfrak{X}_{k_1,i\tau}\mathfrak{X}_{k_2,i\tau}\,\middle|\,\mathcal{C}\right)\\
&\qquad\qquad - \mathbb{E}\left(e_{it}^2\mathfrak{X}_{k_1,it}\mathfrak{X}_{k_2,it}\,\middle|\,\mathcal{C}\right)\mathbb{E}\left(e_{i\tau}^2\mathfrak{X}_{k_1,i\tau}\mathfrak{X}_{k_2,i\tau}\,\middle|\,\mathcal{C}\right)\Big]\\
&= \frac{1}{(NT)^2}\sum_{i=1}^{N}\left\{\sum_{t,\tau=1}^{T}\mathbb{E}\left(e_{it}^2\mathfrak{X}_{k_1,it}\mathfrak{X}_{k_2,it}\,e_{i\tau}^2\mathfrak{X}_{k_1,i\tau}\mathfrak{X}_{k_2,i\tau}\,\middle|\,\mathcal{C}\right) - \left[\sum_{t=1}^{T}\mathbb{E}\left(e_{it}^2\mathfrak{X}_{k_1,it}\mathfrak{X}_{k_2,it}\,\middle|\,\mathcal{C}\right)\right]^2\right\}\\
&\leq \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t,\tau=1}^{T}\mathbb{E}\left(e_{it}^2\mathfrak{X}_{k_1,it}\mathfrak{X}_{k_2,it}\,e_{i\tau}^2\mathfrak{X}_{k_1,i\tau}\mathfrak{X}_{k_2,i\tau}\,\middle|\,\mathcal{C}\right)\\
&\leq \frac{1}{N}\sqrt{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,\tau=1}^{T}\mathbb{E}\left(e_{it}^4e_{i\tau}^4\,\middle|\,\mathcal{C}\right)\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,\tau=1}^{T}\mathbb{E}\left(\mathfrak{X}_{k_1,it}^2\mathfrak{X}_{k_2,it}^2\,\mathfrak{X}_{k_1,i\tau}^2\mathfrak{X}_{k_2,i\tau}^2\,\middle|\,\mathcal{C}\right)}\\
&= \frac{1}{N}\,\mathcal{O}(1) = o(1)\,,
\end{aligned}
$$

where we used that both e and X_k have uniformly bounded 8'th moments. This shows that $\Omega_{NT}^{*} - \mathbb{E}\left(\Omega_{NT}^{*}\,\middle|\,\mathcal{C}\right) = o_p(1)$. We also have $\Omega_{NT}^{*} - \widehat{\Omega} = A_0 + A_1 + A_2 = o_p(1)$, where A_0, A_1 and A_2 are defined in Lemma F.3, and the lemmas states that A_0, A_1 and A_2 are all $o_p(1)$. Combining the above we thus conclude that $\widehat{\Omega} = \Omega + o_p(1)$. ∎

Proof of Theorem 4.4, Part III: show $\widehat{B}_1 = B_1 + o_p(1)$.
Let $B_{1,k,NT} = N^{-1}\operatorname{Tr}\left[P_{f^0}\,\mathbb{E}\left(e'\,X_k\,\middle|\,\mathcal{C}\right)\right]$, According to Assumption 6 we have $B_{1,k} = B_{1,k,NT} + o_p(1)$. What is left to show is that $B_{1,k,NT} = \widehat{B}_{1,k} + o_p(1)$. Using $|\operatorname{Tr}(C)| \leq \|C\|\operatorname{rank}(C)$ we

find

$$\begin{aligned}\left|B_{1,k,NT}-\widehat{B}_1\right| &= \left|\mathbb{E}\left[\frac{1}{N}\mathrm{Tr}(P_{f^0}\,e'\;X_k)\,\middle|\,\mathcal{C}\right]-\frac{1}{N}\mathrm{Tr}\left[P_{\widehat{f}}\left(\widehat{e}'X_k\right)^{\mathrm{truncR}}\right]\right| \\ &\leq \left|\frac{1}{N}\mathrm{Tr}\left[\left(P_{f^0}-P_{\widehat{f}}\right)\left(\widehat{e}'X_k\right)^{\mathrm{truncR}}\right]\right| \\ &\quad+\left|\frac{1}{N}\mathrm{Tr}\left\{P_{f^0}\left[\mathbb{E}\left(e'\;X_k\,|\,\mathcal{C}\right)-\left(\widehat{e}'X_k\right)^{\mathrm{truncR}}\right]\right\}\right| \\ &\leq \frac{2R}{N}\left\|P_{f^0}-P_{\widehat{f}}\right\|\left\|\left(\widehat{e}'X_k\right)^{\mathrm{truncR}}\right\| \\ &\quad+\frac{R}{N}\left\|P_{f^0}\right\|\left\|\mathbb{E}\left(e'\;X_k\,|\,\mathcal{C}\right)-\left(\widehat{e}'X_k\right)^{\mathrm{truncR}}\right\|\;.\end{aligned}$$

We have $\left\|P_{f^0}\right\|=1$. We now apply Lemmas F.5, F.2 and F.6 to find

$$\left|B_{1,k,NT}-\widehat{B}_1\right| = N^{-1}\left(\mathcal{O}_p(N^{-1/2})\mathcal{O}_p(MNT^{1/8})+o_p(N)\right)=o_p(1)\;.$$

This is what we wanted to show. ∎

Proof of Theorem 4.4, final part: show $\widehat{B}_2=B_2+o_p(1)$ and $B_3=B_3+o_p(1)$.
Define

$$B_{2,k,NT}=\frac{1}{T}\mathrm{Tr}\left[\mathbb{E}\left(ee'\right)\;M_{\lambda^0}\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right]\;.$$

According to Assumption 6 we have $B_{2,k}=B_{2,k,NT}+o_p(1)$. What is left to show is that $B_{2,k,NT}=\widehat{B}_{2,k}+o_p(1)$. We have

$$\begin{aligned}B_{2,k}-\widehat{B}_{2,k} =&\frac{1}{T}\mathrm{Tr}\left[\mathbb{E}\left(ee'\right)\;M_{\lambda^0}\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right] \\ &-\frac{1}{T}\mathrm{Tr}\left[\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\;M_{\widehat{\lambda}}\,X_k\,\widehat{f}\left(\widehat{f}'\widehat{f}\right)^{-1}\left(\widehat{\lambda}'\widehat{\lambda}\right)^{-1}\widehat{\lambda}'\right] \\ =&\frac{1}{T}\mathrm{Tr}\left[\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\;M_{\widehat{\lambda}}\,X_k\,\left(f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}-\widehat{f}\left(\widehat{f}'\widehat{f}\right)^{-1}\left(\widehat{\lambda}'\widehat{\lambda}\right)^{-1}\widehat{\lambda}'\right)\right] \\ &+\frac{1}{T}\mathrm{Tr}\left[\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\;\left(M_{\lambda^0}-M_{\widehat{\lambda}}\right)\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right] \\ &+\frac{1}{T}\mathrm{Tr}\left\{\left[\mathbb{E}\left(ee'\right)-\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\right]\;M_{\lambda^0}\,X_k\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right\}\;.\end{aligned}$$

Using $\left|\mathrm{Tr}\left(C\right)\right|\leq\left\|C\right\|\mathrm{rank}\left(C\right)$ (which is true for every square matrix C, see the supplementary material) we find

$$\begin{aligned}\left|B_{2,k}-\widehat{B}_{2,k}\right| \leq&\frac{R}{T}\left\|\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\right\|\left\|X_k\right\|\left\|f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}-\widehat{f}\left(\widehat{f}'\widehat{f}\right)^{-1}\left(\widehat{\lambda}'\widehat{\lambda}\right)^{-1}\widehat{\lambda}'\right\| \\ &+\frac{R}{T}\left\|\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\right\|\left\|M_{\lambda^0}-M_{\widehat{\lambda}}\right\|\left\|X_k\right\|\left\|f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right\| \\ &+\frac{R}{T}\left\|\mathbb{E}\left(ee'\right)-\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\right\|\left\|X_k\right\|\left\|f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right\|\;.\end{aligned}$$

Here we used $\left\|M_{f^0}\right\| = \left\|M_{\widehat{f}}\right\| = 1$. Using $\|X_k\| = \mathcal{O}_p(\sqrt{NT})$, and applying Lemmas F.2, F.4, F.5 and F.6, we now find

$$\begin{aligned}\left|B_{2,k} - \widehat{B}_{2,k}\right| = T^{-1}\Big[&\mathcal{O}_p(T)\,\mathcal{O}_p((NT)^{1/2})\,\mathcal{O}_p(N^{-3/2}) \\ &+ \mathcal{O}_p(T)\,\mathcal{O}_p(N^{-1/2})\,\mathcal{O}_p((NT)^{1/2})\,\mathcal{O}_p((NT)^{-1/2}) \\ &+ o_p(T)\,\mathcal{O}_p((NT)^{1/2})\,\mathcal{O}_p((NT)^{-1/2})\Big] = o_p(1)\ .\end{aligned}$$

This is what we wanted to show. The proof of $\widehat{B}_3 = B_3 + o_p(1)$ is analogous. ∎

G Proofs for Section 5 (Testing)

Proof of Theorem 5.1. Using the expansion for $L_{NT}(\beta)$ in Lemma A.1 of Moon and Weidner (2013) we find for the derivative (the sign convention $\epsilon_k = \beta_k^0 - \beta_k$ results in the minus sign below)

$$\begin{aligned}\frac{\partial L_{NT}}{\partial \beta_k} &= -\frac{1}{NT}\sum_{g=2}^{\infty} g \sum_{\kappa_1=0}^{K}\sum_{\kappa_2=0}^{K}\cdots\sum_{\kappa_{g-1}=0}^{K} \epsilon_{\kappa_1}\,\epsilon_{\kappa_2}\,\ldots\,\epsilon_{\kappa_{g-1}}\,L^{(g)}\left(\lambda^0,\,f^0,\,X_k,\,X_{\kappa_1},\ldots,X_{\kappa_{g-1}}\right) \\ &= \left[2W_{NT}(\beta-\beta^0)\right]_k - \frac{2}{\sqrt{NT}}C_{NT,k} + \frac{1}{NT}\nabla R_{1,NT,k} + \frac{1}{NT}\nabla R_{2,NT,k}\ ,\end{aligned}$$

where

$$\begin{aligned}W_{NT,k_1k_2} &= \frac{1}{NT}\,L^{(2)}\left(\lambda^0,\,f^0,\,X_{k_1},\,X_{k_2}\right)\ , \\ C_{NT,k} &= \frac{1}{2\sqrt{NT}}\sum_{g=2}^{G_e} g\,(\epsilon_0)^{g-1}\,L^{(g)}\left(\lambda^0,\,f^0,\,X_k,\,X_0,\ldots,X_0\right) \\ &= \sum_{g=2}^{G_e}\frac{g}{2\sqrt{NT}}\,L^{(g)}\left(\lambda^0,\,f^0,\,X_k,\,e,\ldots,e\right)\ ,\end{aligned}$$

and

$$\begin{aligned}\nabla R_{1,NT,k} &= -\sum_{g=G_e+1}^{\infty} g\,(\epsilon_0)^{g-1}\,L^{(g)}\left(\lambda^0,\,f^0,\,X_k,\,X_0,\ldots,X_0\right)\ , \\ &= -\sum_{g=G_e+1}^{\infty} g\,L^{(g)}\left(\lambda^0,\,f^0,\,X_k,\,e,\ldots,e\right)\ , \\ \nabla R_{2,NT,k} &= -\sum_{g=3}^{\infty} g\sum_{r=1}^{g-1}\binom{g-1}{r}\sum_{k_1=1}^{K}\cdots\sum_{k_r=1}^{K}\epsilon_{k_1}\,\ldots\,\epsilon_{k_r}\,(\epsilon_0)^{g-r-1} \\ &\qquad\qquad L^{(g)}\left(\lambda^0,\,f^0,\,X_k,\,X_{k_1},\ldots,X_{k_r},X_0,\ldots,X_0\right)\ . \\ &= -\sum_{g=3}^{\infty} g\sum_{r=1}^{g-1}\binom{g-1}{r}\sum_{k_1=1}^{K}\cdots\sum_{k_r=1}^{K}(\beta_{k_1}^0-\beta_{k_1})\,\ldots\,(\beta_{k_r}^0-\beta_{k_r}) \\ &\qquad\qquad L^{(g)}\left(\lambda^0,\,f^0,\,X_k,\,X_{k_1},\ldots,X_{k_r},e,\ldots,e\right)\ .\end{aligned}$$

The above expressions for W_{NT} and C_{NT} are equivalent to their definitions given in theorem 4.1. Using the bound on $L^{(g)}$ we find[23]

$$
\begin{aligned}
|\nabla R_{1,NT,k}| &\leq c_0\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \sum_{g=G_e+1}^{\infty} g^2 \left(\frac{c_1\|e\|}{\sqrt{NT}}\right)^{g-1} \\
&\leq 2\, c_0\, (1+G_e)^2\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \left(\frac{c_1\|e\|}{\sqrt{NT}}\right)^{G_e} \left[1-\left(\frac{c_1\|e\|}{\sqrt{NT}}\right)\right]^{-3} = o_p(\sqrt{NT})\,, \\
|\nabla R_{2,NT,k}| &\leq c_0\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \sum_{g=3}^{\infty} g^2 \sum_{r=1}^{g-1} \binom{g-1}{r} c_1^{g-1} \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta_k^0| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}}\right) \\
&\qquad\qquad \times \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta_k^0| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right)^{g-2} \\
&\leq c_0\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \sum_{g=3}^{\infty} g^3\, (4c_1)^{g-1} \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta_k^0| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}}\right) \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta_{\widetilde{k}}^0| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right)^{g-2} \\
&\leq c_2\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta_k^0| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}}\right) \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta_{\widetilde{k}}^0| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right)\,,
\end{aligned}
$$

where $c_0 = 8Rd_{\max}(\lambda^0, f^0)/2$ and $c_1 = 16d_{\max}(\lambda^0, f^0)/d_{\min}^2(\lambda^0, f^0)$ both converge to a constants as $N, T \to \infty$, and the very last inequality is only true if $4c_1 \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta_{\widetilde{k}}^0| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right) < 1$, and $c_2 > 0$ is an appropriate positive constant. To show $\nabla R_{1,NT,k} = o_p(NT)$ we used Assumption 3*. From the above inequalities we find for $\eta_{NT} \to \infty$

$$
\begin{aligned}
\sup_{\{\beta: \|\beta-\beta^0\| \leq \eta_{NT}\}} \frac{\|\nabla R_{1,NT}(\beta)\|}{\sqrt{NT}} &= o_p(1)\,, \\
\sup_{\{\beta: \|\beta-\beta^0\| \leq \eta_{NT}\}} \frac{\|\nabla R_{2,NT}(\beta)\|}{NT\, \|\beta-\beta^0\|} &= o_p(1)\,.
\end{aligned}
$$

Thus $R_{NT}(\beta) = R_{1,NT}(\beta) + R_{2,NT}(\beta)$ satisfies the bound in the theorem. ∎

Proof of Theorem 5.2. Using Theorem 4.3 it is straightforward to show that WD_{NT}^* has limiting distribution χ_r^2.

For the LR test we have to show that the estimator $\widehat{c} = (NT)^{-1}\mathrm{Tr}(\widehat{e}(\widehat{\beta})\,\widehat{e}'(\widehat{\beta}))$ is consistent for $c = \mathbb{E}e_{it}^2$. As already noted in the main text we have $\widehat{c} = L_{NT}\left(\widehat{\beta}\right)$, and using our expansion and $\sqrt{NT}$-consistency of $\widehat{\beta}$ we immediately obtain

$$
\widehat{c} = \frac{1}{NT}\,\mathrm{Tr}(M_{\lambda^0} e M_{f^0} e') + o_p(1)\,.
$$

Alternatively, one could use the expansion of $\widehat{e}$ in Theorem E.1 to show this. From the above

[23] Here we use $\binom{n}{k} \leq 4^n$.

result we find

$$\left|\widehat{c} - \frac{1}{NT}\mathrm{Tr}(ee')\right| = \frac{1}{NT}\left|\mathrm{Tr}(P_{\lambda^0} e M_{f^0} e') + \mathrm{Tr}(e P_{f^0} e')\right| + o_p(1)$$
$$\leq \frac{2R}{NT}\|e\|^2 + o_p(1) = o_p(1)\ .$$

By the weak law of large numbers we thus have

$$\widehat{c} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2 + o_p(1) = c + o_p(1)\ ,$$

i.e. $\widehat{c}$ is indeed consistent for c. Having this one immediately obtains the result for the limiting distribution of LR^*_{NT}.

For the LM test we first want to show that equation (5.3) holds. Using the expansion of $\widehat{e}$ in Theorem E.1 one obtains

$$\begin{aligned}
\sqrt{NT}(\widetilde{\nabla}\mathcal{L}_{NT})_k &= -\frac{2}{\sqrt{NT}}\,\mathrm{Tr}\left(X_k'\widetilde{e}\right) \\
&= \left[2\sqrt{NT}\,W_{NT}\,\left(\widetilde{\beta}-\beta^0\right)\right]_k + \frac{2}{NT}C^{(1)}(\lambda^0,f^0,X_k,e) + \frac{2}{NT}C^{(2)}(\lambda^0,f^0,X_k,e) \\
&\qquad - \frac{2}{\sqrt{NT}}\,\mathrm{Tr}\left(X_k'\widetilde{e}^{(\mathrm{rem})}\right) \\
&= \left[2\sqrt{NT}\,W_{NT}\,\left(\widetilde{\beta}-\beta^0\right) + \frac{2}{NT}C_{NT}\right]_k + o_p(1) \\
&= \sqrt{NT}\left[\nabla L_{NT}(\widetilde{\beta})\right]_k + o_p(1)\,,
\end{aligned}$$

which is what we wanted to show. Here we used that $|\mathrm{Tr}\left(X_k'\widetilde{e}^{(\mathrm{rem})}\right)| \leq 7R\|X_k\|\|\widetilde{e}^{(\mathrm{rem})}\| = \mathcal{O}_p(N^{3/2})$. Note that $\|X_k\| = \mathcal{O}_p(N)$, and Theorem E.1 and $\sqrt{NT}$-consistency of $\widetilde{\beta}$ imply $\|\widetilde{e}^{(\mathrm{rem})}\| = \mathcal{O}_p(\sqrt{N})$. We also used the expression for $\nabla L_{NT}(\widetilde{\beta})$ given in Theorem 5.1, and the bound on $\nabla R_{NT}(\beta)$ given there.

We now use equation (5.4) and $\widetilde{W} = W + o_p(1)$, $\widetilde{\Omega} = \Omega + o_p(1)$, and $\widetilde{B} = B + o_p(1)$ to obtain

$$LM^*_{NT} \underset{d}{\longrightarrow} (C-B)'W^{-1}H'(HW^{-1}\Omega W^{-1}H')^{-1}HW^{-1}(C-B)\ .$$

Under H_0 we thus find $LM^*_{NT} \to_d \chi^2_r$. ∎

Tables with Simulation Results

		$\rho^0 = 0.3$			$\rho^0 = 0.9$		
		OLS	FLS	BC-FLS	OLS	FLS	BC-FLS
$T = 5$, $M = 2$	bias	0.1232	-0.1419	-0.0713	0.0200	-0.3686	-0.2330
	std	0.1444	0.1480	0.0982	0.0723	0.1718	0.1301
	rmse	0.1898	0.2050	0.1213	0.0750	0.4067	0.2669
$T = 10$, $M = 3$	bias	0.1339	-0.0542	-0.0201	0.0218	-0.1019	-0.0623
	std	0.1148	0.0596	0.0423	0.0513	0.1094	0.0747
	rmse	0.1764	0.0806	0.0469	0.0557	0.1495	0.0973
$T = 20$, $M = 4$	bias	0.1441	-0.0264	-0.0070	0.0254	-0.0173	-0.0085
	std	0.0879	0.0284	0.0240	0.0353	0.0299	0.0219
	rmse	0.1687	0.0388	0.0250	0.0434	0.0345	0.0235
$T = 40$, $M = 5$	bias	0.1517	-0.0130	-0.0021	0.0294	-0.0057	-0.0019
	std	0.0657	0.0170	0.0160	0.0250	0.0105	0.0089
	rmse	0.1654	0.0214	0.0161	0.0386	0.0119	0.0091
$T = 80$, $M = 6$	bias	0.1552	-0.0066	-0.0007	0.0326	-0.0026	-0.0006
	std	0.0487	0.0112	0.0109	0.0179	0.0056	0.0053
	rmse	0.1627	0.0130	0.0109	0.0372	0.0062	0.0053

Table 1: Simulation results for the AR(1) model described in the main text with $N = 100$, $\rho_f = 0.5$, $\sigma_f = 0.5$, and different values of T (with corresponding bandwidth M) and true AR(1) coefficient ρ^0. The OLS estimator, the LS estimator with factors (FLS, computed with correct $R = 1$), and corresponding bias corrected LS estimator with factors (BC-FLS) were computed for 10,000 simulation runs. The table lists the mean bias, the standard deviation (std), and the square root of the mean square error (rmse) for the three estimators.

		$\rho^0 = 0.3$			$\rho^0 = 0.9$		
		OLS	FLS	BC-FLS	OLS	FLS	BC-FLS
$T = 5$, $M = 2$	bias	0.1239	-0.5467	-0.3721	0.0218	-0.9716	-0.7490
	std	0.1454	0.1528	0.1299	0.0731	0.1216	0.1341
	rmse	0.1910	0.5676	0.3942	0.0763	0.9792	0.7609
$T = 10$, $M = 3$	bias	0.1343	-0.1874	-0.1001	0.0210	-0.4923	-0.3271
	std	0.1145	0.1159	0.0758	0.0518	0.1159	0.0970
	rmse	0.1765	0.2203	0.1256	0.0559	0.5058	0.3412
$T = 20$, $M = 4$	bias	0.1451	-0.0448	-0.0168	0.0255	-0.1822	-0.1085
	std	0.0879	0.0469	0.0320	0.0354	0.0820	0.0528
	rmse	0.1696	0.0648	0.0362	0.0436	0.1999	0.1207
$T = 40$, $M = 5$	bias	0.1511	-0.0161	-0.0038	0.0300	-0.0227	-0.0128
	std	0.0663	0.0209	0.0177	0.0250	0.0342	0.0225
	rmse	0.1650	0.0264	0.0181	0.0390	0.0410	0.0258
$T = 80$, $M = 6$	bias	0.1550	-0.0072	-0.0011	0.0325	-0.0030	-0.0010
	std	0.0488	0.0123	0.0115	0.0182	0.0064	0.0057
	rmse	0.1625	0.0143	0.0116	0.0372	0.0071	0.0058

Table 2: Same DGP as Table 1, but misspecification in number of factors R is present. The true number of factors is $R = 1$, but the FLS and BC-FLS are calculated with $R = 2$.

	$M=1$	$M=2$	$M=3$	$M=4$	$M=5$	$M=6$	$M=7$	$M=8$
$\rho^0=0$	0.889	0.832	0.791	0.754	0.720	0.689	0.660	0.633
$\rho^0=0.3$	0.752	0.806	0.778	0.742	0.708	0.677	0.648	0.621
$\rho^0=0.6$	0.589	0.718	0.728	0.704	0.674	0.644	0.616	0.590
$\rho^0=0.9$	0.299	0.428	0.486	0.510	0.519	0.516	0.508	0.495

Table 3: Simulation results for the AR(1) model with $N=100$, $T=20$, $\rho_f=0.5$, and $\sigma_f=0.5$. For different values of the AR(1) coefficient ρ^0 and of the bandwidth M, we give the fraction of the LS estimator bias that is accounted for by the bias correction, i.e. the fraction $\sqrt{NT}\,\mathbb{E}(\widehat{\beta}-\beta)/\mathbb{E}(\widehat{W}^{-1}\widehat{B})$, computed over 10,000 simulation runs. Here and in all following tables it is assumed that $R=1$ is correctly specified.

		BC-FLS for $\rho^0=0.3$			BC-FLS for $\rho^0=0.9$		
		M=2	M=5	M=8	M=2	M=5	M=8
$T=20$	bias	-0.0056	-0.0082	-0.0100	-0.0100	-0.0083	-0.0089
	std	0.0239	0.0241	0.0247	0.0253	0.0212	0.0208
	rmse	0.0245	0.0255	0.0266	0.0272	0.0228	0.0227
$T=40$	bias	-0.0017	-0.0023	-0.0030	-0.0024	-0.0019	-0.0018
	std	0.0159	0.0159	0.0159	0.0095	0.0089	0.0085
	rmse	0.0160	0.0161	0.0162	0.0098	0.0091	0.0087

Table 4: Same specification as Table 1. We only report the properties of the bias corrected LS estimator, but for multiple values of the bandwidth parameter M and two different values for T. Results were obtained using 10,000 simulation runs.

		$\rho_f=0.3$			$\rho_f=0.7$		
		OLS	FLS	BC-FLS	OLS	FLS	BC-FLS
$\sigma_f=0$	bias	-0.0007	-0.0076	-0.0043	-0.0004	-0.0074	-0.0041
	std	0.0182	0.0332	0.0243	0.0178	0.0331	0.0242
	rmse	0.0182	0.0340	0.0247	0.0178	0.0339	0.0245
$\sigma_f=0.2$	bias	0.0153	-0.0113	-0.0032	0.0474	-0.0291	-0.0071
	std	0.0251	0.0303	0.0229	0.0382	0.0387	0.0272
	rmse	0.0294	0.0323	0.0231	0.0609	0.0484	0.0281
$\sigma_f=0.5$	bias	0.0567	-0.0137	-0.0041	0.1491	-0.0403	-0.0126
	std	0.0633	0.0260	0.0207	0.0763	0.0298	0.0226
	rmse	0.0850	0.0294	0.0211	0.1675	0.0501	0.0259

Table 5: Simulation results for the AR(1) model with $N=100$, $T=20$, $M=4$, and $\rho^0=0.6$. The three different estimators were computed for 10,000 simulation runs, and the mean bias, standard deviation (std), and root mean square error (rmse) are reported.

		size			size		
		WD	LR	LM	WD^*	LR^*	LM^*
$\rho^0 = 0$	$N = 100,\ T = 20,\ M = 4$	0.219	0.214	0.192	0.066	0.062	0.056
	$N = 400,\ T = 80,\ M = 6$	0.199	0.198	0.195	0.055	0.054	0.054
	$N = 400,\ T = 20,\ M = 4$	0.560	0.556	0.532	0.089	0.088	0.076
	$N = 1600,\ T = 80,\ M = 6$	0.593	0.591	0.586	0.056	0.055	0.055
$\rho^0 = 0.6$	$N = 100,\ T = 20,\ M = 4$	0.326	0.311	0.272	0.098	0.091	0.077
	$N = 400,\ T = 80,\ M = 6$	0.260	0.255	0.248	0.056	0.053	0.057
	$N = 400,\ T = 20,\ M = 4$	0.591	0.582	0.552	0.174	0.167	0.136
	$N = 1600,\ T = 80,\ M = 6$	0.666	0.663	0.656	0.060	0.058	0.059

Table 6: Simulation results for the AR(1) model with $\rho_f = 0.5$ and $\sigma_f = 0.5$. For the different values of ρ^0, N, T and M we test the hypothesis $H_0 : \rho = \rho^0$ using the uncorrected and bias corrected Wald, LR and LM test and nominal size 5%. The size of the different tests is reported, based on 10,000 simulation runs.

			power			power		
			WD	LR	LM	WD^*	LR^*	LM^*
$\rho^0 = 0$	$N = 100,\ T = 20,\ M = 4$	H_a^{left}	0.094	0.089	0.076	0.128	0.123	0.121
		H_a^{right}	0.526	0.515	0.487	0.235	0.227	0.206
	$N = 400,\ T = 80,\ M = 6$	H_a^{left}	0.066	0.064	0.063	0.154	0.151	0.153
		H_a^{right}	0.549	0.545	0.540	0.194	0.191	0.190
	$N = 400,\ T = 20,\ M = 4$	H_a^{left}	0.306	0.305	0.284	0.100	0.097	0.096
		H_a^{right}	0.791	0.787	0.769	0.309	0.305	0.279
	$N = 1600,\ T = 80,\ M = 6$	H_a^{left}	0.254	0.253	0.248	0.128	0.127	0.129
		H_a^{right}	0.871	0.869	0.866	0.225	0.224	0.224
$\rho^0 = 0.6$	$N = 100,\ T = 20,\ M = 4$	H_a^{left}	0.192	0.180	0.147	0.184	0.171	0.171
		H_a^{right}	0.619	0.605	0.563	0.335	0.318	0.294
	$N = 400,\ T = 80,\ M = 6$	H_a^{left}	0.081	0.079	0.076	0.184	0.195	0.200
		H_a^{right}	0.680	0.675	0.668	0.335	0.262	0.267
	$N = 400,\ T = 20,\ M = 4$	H_a^{left}	0.421	0.412	0.378	0.184	0.160	0.150
		H_a^{right}	0.792	0.787	0.765	0.335	0.426	0.399
	$N = 1600,\ T = 80,\ M = 6$	H_a^{left}	0.318	0.314	0.307	0.200	0.169	0.172
		H_a^{right}	0.912	0.911	0.908	0.268	0.316	0.320

Table 7: As Table 6, but we report the power for testing the alternatives $H_a^{\text{left}} : \rho = \rho^0 - (NT)^{-1/2}$ and $H_a^{\text{right}} : \rho = \rho^0 + (NT)^{-1/2}$.

			size corrected power			size corrected power		
			WD	LR	LM	WD^*	LR^*	LM^*
$\rho^0 = 0$	$N = 100,\ T = 20,\ M = 4$	H_a^{left}	0.010	0.011	0.010	0.105	0.104	0.112
		H_a^{right}	0.211	0.208	0.206	0.199	0.197	0.193
	$N = 400,\ T = 80,\ M = 6$	H_a^{left}	0.008	0.008	0.008	0.143	0.143	0.145
		H_a^{right}	0.236	0.237	0.235	0.181	0.182	0.181
	$N = 400,\ T = 20,\ M = 4$	H_a^{left}	0.008	0.008	0.009	0.055	0.052	0.062
		H_a^{right}	0.187	0.185	0.181	0.210	0.208	0.208
	$N = 1600,\ T = 80,\ M = 6$	H_a^{left}	0.005	0.005	0.005	0.119	0.119	0.120
		H_a^{right}	0.226	0.227	0.225	0.213	0.213	0.212
$\rho^0 = 0.6$	$N = 100,\ T = 20,\ M = 4$	H_a^{left}	0.014	0.014	0.016	0.114	0.115	0.127
		H_a^{right}	0.196	0.193	0.196	0.233	0.234	0.231
	$N = 400,\ T = 80,\ M = 6$	H_a^{left}	0.005	0.005	0.005	0.185	0.187	0.184
		H_a^{right}	0.288	0.288	0.288	0.248	0.252	0.247
	$N = 400,\ T = 20,\ M = 4$	H_a^{left}	0.013	0.016	0.015	0.040	0.039	0.051
		H_a^{right}	0.128	0.127	0.126	0.206	0.201	0.209
	$N = 1600,\ T = 80,\ M = 6$	H_a^{left}	0.005	0.005	0.005	0.153	0.153	0.154
		H_a^{right}	0.236	0.236	0.238	0.291	0.291	0.291

Table 8: As Table 7, but we report the size corrected power.