Department of Economics

Athens University of Economics and Business

WORKING PAPER no. 18-2023

Time Varying Three Pass Regression Filter

Yiannis Dendramis, George Kapetanios and Massimiliano Marcellino

October 2023

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5

The Working Papers in this series circulate mainly for early presentation and discussion, as well as for the information of the Academic Community and all interested in our current research activity.

The authors assume full responsibility for the accuracy of their paper as well as for the opinions expressed therein.

Time Varying Three Pass Regression Filter

Yiannis Dendramis[1] George Kapetanios[2]
Massimiliano Marcellino[3*]

[1] *Athens University of Economics and Business, ydendramis@aueb.gr*
[2] *King's College London, george.kapetanios@kcl.ac.uk*
[3] *Bocconi University, massimiliano.marcellino@unibocconi.it*

October 2023

Abstract

We propose non parametric estimators for the three pass regression filter for factor extraction from large dimensional datasets when the factor loadings, the proxy and the target equation parameters are allowed to vary stochastically over time. We provide theoretically optimal and empirically efficient solutions for the choice of bandwidth of the kernel-based estimators. Moreover, we prove consistency of the associated forecasts when both the time and the cross section dimensions of our dataset become large. We also link our proposals with the time varying parameter constrained least squares estimator and with the time varying partial least squares method, and show that these are special cases of our approach. We asses the finite sample performance of our approach by an extensive set of Monte Carlo experiments, also comparing it with other alternatives proposed in the literature. Finally, we illustrate the empirical advantages of our approach in an out of sample forecasting exercise, using a large panel of macroeconomic series to predict key variables of interest.

Keywords: Factor model, Principal components, Constrained least squares, Partial least squares, Forecasting, Parameter time variation, Kernel estimation.

JEL classification: C13; C22; C51.

*We would like to thank the participants of the 9th Annual Conference of the International Association for Applied Econometrics (IAAE), and the 5th Vienna Workshop on High-Dimensional Times Series in Macroeconomics and Finance

1 Introduction

Factor models can summarize efficiently information from large data sets and have received extensive attention in the empirical macro and forecasting literature, starting from the seminal contributions by Stock and Watson (2002a), Stock and Watson (2002b), Bai (2003), Forni *et al.* (2000), Forni *et al.* (2005) (see also Stock and Watson (2011), Stock and Watson (2015), for reviews). Stock and Watson (2009) and Bates *et al.* (2013) argue that when the factor loadings undergo small instabilities, the factor estimates obtained via the conventional principal component analysis (PCA) are still consistent. Simulation and empirical support for this proposition are provided by Banerjee *et al.* (2008), who find robustness of PCA also when there are limited changes in the dynamics of the factors. However, since macroeconomic datasets typically span a long time period, it is restrictive to assume that the factor loadings and their dynamic evolution are time-invariant or undergo negligible changes during the whole period.

The problem of structural changes in factor loadings has received a great deal of attention in recent years. Among the relevant contributions, some focus on detecting and modeling a small number of large breaks, e.g., Chen *et al.* (2014), Breitung and Eickmeier (2011), Cheng *et al.* (2016), Ma and Su (2018), Massacci (2017), others assume either slow changes in the loadings, e.g., Su and Wang (2017),[1] or propose a modified Kalman filter based estimation of a fully parametric time-varying factor model, as in Eickmeier *et al.* (2015).

Another strand of the factor literature focused on improving the efficiency of factor estimators and associated forecasts, by focusing on a relevant subset of the available large set of variables, or factors. A first possibility is to pre-select the indicators that are most correlated with the target variable of interest, see for example Boivin and Ng (2006) and Groen and Kapetanios (2016). A more efficient solution is the three-pass regression filter (3PRF) developed by Kelly and Pruitt (2015), (KP), which is computationally efficient as it is based on a set of simple OLS regressions and has a number of (asymptotic) optimality properties, performs well in finite samples compared to more complex alternatives, and produces good nowcasts and short-term forecasts for a variety of macroeconomic and financial variables, see KP. Basically, contrary to the method of principal components, 3PRF assures that the estimated factors are those most relevant for predicting the target variable of interest.

[1] Su and Wang (2017) use a local kernel estimator in the time dimension to study gradual changes in a PCA framework. This approach excludes sudden large changes and exploits only the data in a local neighborhood of a particular time observation. Most importantly, they assume a deterministic parametrization for the loadings which can be questionable for economic data.

The 3PRF has been later extended to deal with mixed frequency data by Hepenstrick *et al.* (2019) and to allow for Markov Switching in the factor loadings and dynamics by Guérin *et al.* (2020). Yet, as Hansen (2001) points out, "it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect". Hence, it seems more realistic to assume smooth changes rather than abrupt change.

The type of parameter time variation indeed is very important, as parametric methods are only optimal (and often only consistent) if the assumptions on the type of time variation hold, but often testing these assumptions is difficult, as test statistics (e.g., Andrews (1993), Corradi and Swanson (2014)) tend to have low power when applied with the rather short time series typically available in macroeconomic applications. Hence, there has been more and more interest in non-parametric modelling parameter time-variation. While early studies, such as Robinson (1989), Robinson (1991), and Chen and Hong (2012) assumed smooth deterministic changes, the more recent literature permits to have persistent stochastic time variation and studies the properties of kernel-based estimators in this context, see e.g. Giraitis *et al.* (2014), Giraitis *et al.* (2018), Giraitis *et al.* (2021).

In this paper, we assume stochastic non-parametric time variation in the parameters of factor models, introduce kernel-based estimators in the context of the 3PRF for factor estimation and forecasting, establish the limiting distributions of the estimated factors and factor loadings under the standard large N and large T framework, and prove consistency of the associated forecasts. We also link our proposed method with the time varying parameter constrained least squares estimator and with the time varying partial least squares method, and show that these are special cases of our approach. Finally, we propose a BIC-type information criterion to determine the number of common factors in this context, and suggest specific cross validation methods to select the bandwidths of the kernel based estimators

We then assess the finite sample performance of our approach by an extensive set of Monte Carlo experiments, also comparing it with other alternatives proposed in the literature, ranging from the standard 3PRF and PCA analysis, to more sophisticated PCA based approaches. Overall, the relative performance of the TV-3PRF is quite satisfactory, with gains up to 25-30% with respect to 3PRF, also for rather small sizes with comparable N and T dimensions ($N = 100$, $T = 100$).

Finally, we illustrate the empirical advantages of our method in an out of sample forecasting exercise, using a large panel of US macroeconomic series to predict key variables of interest, such as the Federal Funds Rate, employment, hours worked, housing starts, or the USD/pound exchange rate. The gains of the TV-3PRF are generally confirmed also in this setting, making it not only an interesting theoretical contribution but also an additional tool for applied econometrics and empirical economic analysis.

The rest of the paper is structured as follows. Section 2 describes the time-varying 3PRF methods and the associated assumptions. Section 3 presents our theoretical contributions and the relations with other popular methods in the literature. Section 4 discusses the Monte Carlo results. Section 5 presents the empirical application. Section 6 summarises the main results and concludes. All proofs are relegated to the Appendix.

2 The Model and the Estimators

In this section we discuss, in turn, the model, the estimation procedure, and the assumptions. The following section studies the properties of the estimators.

2.1 The model

Our set-up extends that of Kelly and Pruitt (2015) who first introduced the fixed parameter three-pass regression filter (3PRF). We consider a target variable (y) that we wish to forecast, by the use of a large set of predictors (x). Since the number of predictors (N) is large or even larger than the number of time series observations (T), we assume a stochastic loading, approximate factor model for the data. Yet, not all factors drive the target variable, but a subset of them includes predictive information. We identify the set of relevant factors (f) by the use of (M) proxy variables (z), which are driven by the same factors as the target variable. Formally, the stochastic parameter 3PRF model is described by the following set of equations

$$y_{t+1} = \beta_{0t} + \beta_t' F_t + \eta_{t+1} \tag{1}$$

$$z_t = \lambda_{0t} + \Lambda_t F_t + \omega_t \tag{2}$$

$$x_t = \phi_{0t} + \Phi_t F_t + \varepsilon_t \tag{3}$$

where y_{t+1} is the variable that we wish to forecast (target variable); the full set of latent factors is $F_t = (f_t', g_t')'$, where f_t is the set of $K_f > 0$ relevant (for y_{t+1}) factors, and g_t is the set of $K_g > 0$ factors that do not drive y_{t+1}, but explain a portion of the large dataset x_t. According to this, we assume that the stochastic parameters of the equation (1) are $\beta_t = \left(\beta_{ft}', \beta_{gt}'\right)'$, $\beta_{ft}' \neq 0$, $\beta_{gt}' = 0$, $\Lambda_t = (\Lambda_{ft}, \Lambda_{gt})$, for the proxy equation and $\Phi_t = (\Phi_{f,t}, \Phi_{g,t})$, for the large dataset equation, as well as β_{0t}, λ_{0t}, ϕ_{0t} that correspond to the stochastic intercepts. Since the set of M proxies in the vector z_t, and the target variable

y_{t+1}, are driven by the same factors f_t, we set $\Lambda_{gt} = \mathbf{0}$ and require Λ_{ft} to have full rank (see assumptions below for more details). We can allow $M << \min(N, T)$. The full set of $K = K_f + K_g$ factors in F_t drive the large set of predictors x_t according to the stochastic loading matrix Φ_t and the intercept ϕ_{0t}.

Equations (1)-(3) define the factor structure. As mentioned, the target's factor loadings, $\beta_t = \left(\beta'_{ft}, 0'\right)'$, allow the target to depend on a strict subset of the factors driving the predictors. We refer to this subset as the relevant factors, which are denoted f_t . The irrelevant factors, g_t, do not influence the forecast target but may drive the cross section of predictive variables x_t . The proxies z_t are driven by the relevant factors. In addition, each variable has an idiosyncratic component, whose properties are specified below.

2.2 The time-varying three pass regression filter

Our estimation algorithm is based on the use of kernels that are associated with weights, $k_{H,ts} = K\left(\frac{t-s}{H}\right)$, $K_{H,t} = \sum_{s=1}^{T} k_{H,ts}$, where $K(.)$ is generally specified as a probability density function. The bandwidth H governs the relative magnitude of the weights, with $H \to \infty$ and $H = o(T)$. Popular choices for $K(.)$ include the normal density kernel, $K(u) = \exp\left(-\frac{1}{2}u^2\right)$, the rolling window kernel, $K(u) = I(0 \leq u \leq 1)$, and the exponential weighted moving average (EWMA) kernel, $K(u) = \exp(-u)$, for $u \in [0, \infty)$.

Given the model in equations (1)-(3) the TV-3PRF consists of the following three steps.

Step 1: Run the time series regression of each predictor x_{it} on the M proxies z_t,

$$x_{it} = \phi_{0it} + z_t'\Phi_{it} + \epsilon_{it}, t = 1, .., T \tag{4}$$

and retain the i-th predictor loading estimates $\widehat{\Phi}_{it}$, for $t = 1, .., T$ as

$$\widehat{\Phi}_{it} = \left(\sum_{j=1}^{T} k_{H,tj}\left(z_j - \bar{z}_t\right)\left(z_j - \bar{z}_t\right)'\right)^{-1}\left(\sum_{j=1}^{T} k_{H,tj}\left(z_j - \bar{z}_t\right)\left(x_{ij} - \bar{x}_{it}\right)'\right) \tag{5}$$

with $\bar{z}_t = K_{H,t}^{-1}\sum_{l=1}^{T} k_{H,lt} z_l$, and $\bar{x}_t = K_{H,t}^{-1}\sum_{l=1}^{T} k_{H,lt} x_l$. Collect the estimates of factor loadings, for all variables and each time period, in the matrix $\widehat{\Phi}_t = \left(\widehat{\Phi}_{1t}, ..., \widehat{\Phi}_{Nt}\right)'$.

Step 2: For each time period t, run the cross sectional regression, of $x_t = (x_{1t}, .., x_{Nt})'$ on $\widehat{\Phi}_t$, so that each equation is

$$x_{it} = \phi_{0it} + \widehat{\Phi}'_{it} F_t + \varepsilon_{it}, i = 1, .., N \tag{6}$$

and retain the obtained estimates in the $M \times 1$ vector $\widehat{F}_t$ obtained as

$$\widehat{F}_t = \left(\widehat{\Phi}'_t J_N \widehat{\Phi}_t\right)^{-1} \widehat{\Phi}'_t J_N x_t \tag{7}$$

where $J_N \equiv I_N - \frac{1}{N} i_N i_N'$, I_N is the N-dimensional identity matrix, and i_N an N-vector of ones.

Step 3: Run the time series regression of the target variable on the estimated factors, that is,

$$y_{t+1} = \beta_{0t} + \widehat{F}_t' \beta_t + \eta_{t+1}, \, t = 1, .., T \tag{8}$$

where β_{0t} and β_t are estimated by

$$\widehat{\beta}_t = \left(\sum_{j=1}^{T} k_{L,tj} \left(\widehat{F}_j - \overline{\widehat{F}}_t \right) \left(\widehat{F}_j - \overline{\widehat{F}}_t \right)' \right)^{-1} \left(\sum_{j=1}^{T} k_{L,tj} \left(\widehat{F}_j - \overline{\widehat{F}}_t \right) \left(y_{j+1} - \overline{y}_t \right) \right) \tag{9}$$

$$\widehat{\beta}_{0t} = \overline{y}_t - \overline{\widehat{F}}_t \widehat{\beta}_t$$

with $\overline{\widehat{F}}_t = K_{L,t}^{-1} \sum_{j=1}^{T} k_{L,jt} \widehat{F}_j$, $\overline{y}_t = K_{L,t}^{-1} \sum_{j=1}^{T} k_{L,jt} y_{j+1}$, $\widehat{F}_t$ is given by (7), and L is a bandwidth parameter that can be different from H (used in step 1).

The TV-3PRF defined above extracts factors by first running N separate time series regressions, one for each predictor and all periods $t = 1, .., T$. Here, the regression coefficients describe the varying sensitivity of the predictors to the factors that drive the z variables, and correspond to factor loadings. In the second pass, we run T cross sectional regressions of all the variables on the loadings estimated in the first step, and the estimated coefficients for each period t estimate the relevant factors for that period. The most important requirement is that the proxies' common components span the space of the relevant factors. Or, put it another way, proxies and target have the same source of variation, after netting out the effect of the time varying parameters. In the final step, we run a single time series forecasting regression of the target variable y_{t+1} on the second-pass estimated factors $\widehat{F}_t$, which could be also denoted as $\widehat{f}_t$ as they are only the relevant ones. Intuitively, since the relevant factor space is spanned by $\widehat{F}_t$, the third stage regression delivers consistent forecasts, in the sense that $\widehat{y}_{t+1} = \widehat{\beta}_{0t} + \widehat{F}_t' \widehat{\beta}_t$ converges to $\beta_{0t} + F_t' \beta_t$. We will show this formally below.

To conclude, note that if all the factors are relevant, the 3PRF factors boil down to the common principal component based factor estimators of Stock and Watson (2002a), Stock and Watson (2002b). In this context, an $M = K$ dimensional subset of x could be used as proxies, and the TV-3PRF factors provide a general extension of the principal component factor analysis to the time-varying case, assuming stochastic loadings.

2.3 Assumptions

To fully characterise the model described by equations (1), (2), (3), we introduce the following definitions and assumptions. These provide the groundwork for the theoretical results developed in the next sections. For a generic random variable r_{jt}, we write $r_{jt} \in \Upsilon(\rho)$, $\rho > 0$, to denote a thin-tailed distribution for r_{jt}:

$$\underset{j,t}{max} E \exp(\alpha |r_{jt}|^{\rho}) < \infty, \text{ for some } \alpha > 0, \tag{10}$$

which implies

$$P\left(|r_{jt}| \geq \zeta\right) \leq c_0 \exp\left(-c_1 \zeta^{\rho}\right), c_0, c_1 > 0. \tag{11}$$

Also, with $r_{jt} \in \Xi(\theta)$, $\theta > 2$ we denote a heavy tailed distributed random variable for which it holds that:

$$\underset{j,t}{max} E|r_{jt}|^{\theta} < \infty, \tag{12}$$

which implies

$$P\left(|r_{jt}| \geq \zeta\right) \leq c_1 \zeta^{-\theta}, c_1 > 0. \tag{13}$$

Moreover, the generic centered stochastic process $r_{jt} - E(r_{jt})$ is strong (α-) mixing (but not necessairily stationary) with mixing coefficients α_k,[2] such that for some $0 < \phi < 1$, and $c > 0$, $\alpha_k \leq c\phi^k$, $k \geq 1$.

Next, we characterize the factors, the stochastic parameter time variation, and the residuals in model (1)-(3).

Assumption 1 *1. For any (i, j) the processes $(F_{it} - E(F_{it}))$, $(F_{it}\eta_t)$, $(F_{it}\varepsilon_{jt})$, $(F_{it}\omega_{jt})$, (η_t), (ε_{jt}), (ω_{jt}) are strong (α-) mixing process. In addition, F_t is allowed to be heavy tailed distributed, i.e. $F_t \in \Xi(\theta)$, $\theta > 8$ and $E(F_t) = \mu$, $cov(F_t) = \Delta_F$.*

2. Let Φ_{jt} be the vector of loadings of the j-th series in x_t. Then $N^{-1}\sum_{j=1}^{N}\Phi_{jt} \underset{T\to\infty}{\overset{p}{\to}} \overline{\Phi} < \infty$, $N^{-1}\Phi_t' J_N \Phi_t \underset{N\to\infty}{\overset{p}{\to}} \Delta_{\Phi}$, $N^{-1}\Phi_t' J_N \phi_{0t} \underset{N\to\infty}{\overset{p}{\to}} P_1$. Moreover, the elements of $\Phi_t = \left(\Phi_{jm,t}\right)$ and $\phi_{0t} = \left(\phi_{0jt}\right)$ satisfy

$$\left|\phi_{0jt} - \phi_{0js}\right| = C\left(\frac{|t-s|}{T}\right)^{1/2} \phi_{0jts}, \tag{14}$$

$$\left|\Phi_{jm,t} - \Phi_{jm,s}\right| = C\left(\frac{|t-s|}{T}\right)^{1/2} \Phi_{jm,ts},$$

for all $j = 1,..,N$, $m = 1,..,K$, $1 \leq s,t \leq T$, and ϕ_{0jt}, $\Phi_{jm,t}$, $\Phi_{jm,ts}$, $\phi_{0jts} \in \Upsilon(\rho)$, for some $C, \rho > 0$.

3. $E(\varepsilon_{it}) = 0$, $E\left(|\varepsilon_{it}|^8\right) \leq C$ with $C < \infty$. It also holds that $E(F_t\varepsilon_{it}) = 0$, $E|F_t\varepsilon_{it}|^r < C$, $E(\Phi_{is}\varepsilon_{it}) = 0$ for all s, t, and $E\left(|\Phi_{is}\varepsilon_{it}|^r\right) < C$ with $r > 2$.

[2]Let $_{j},_{..,\infty}$ denote σ-fields generated by $(r_t, t \leq j)$ and $(r_t, t \geq j)$ respectively. Define the α-mixing coefficient as $\alpha_k = \sup_j \sup_{A \in_{-\infty}^{j}, B \in_{j+k}^{\infty}} |P(A)P(B) - P(A \cap B)|$.

4. $E(\omega_t) = 0$, $E\|\omega_t\|^4 < C$, $K_t^{-1/2}\sum_{s=1}^{T} k_{t,s}\omega_s = O_p(1)$, $cov(\omega_t) = \Delta_\omega$

5. $E(\eta_{t+1}|\eta_t, F_t, y_{t-1}, F_{t-1}, ...) = 0$, $E\left(\eta_{t+1}^2|\eta_t, F_t, y_{t-1}, F_{t-1}, ...\right) = \sigma_\eta^2$ *and* $\varepsilon_{\eta,t+1} = \sigma_\eta^{-1}\eta_{t+1}$. *Also,* $E\left(\eta_{t+1}^4\right) \leq C$, η_{t+1} *is independent of* $\left(\Phi_{ij,t}\right)$, *and* (ε_{it}) *with* $E(F_t\eta_{t+1}) = 0$, $E|F_t\eta_{t+1}|^r < C < \infty$, *for all* t, *with* $r > 2$.

6. The parameters β_{0t}, β_{jt} *of the target equation evolve stochastically over time with*

$$|\beta_{0t} - \beta_{0s}| = C\left(\frac{|t-s|}{T}\right)^{1/2}\beta_{0ts}, \tag{15}$$

$$\left|\beta_{jt} - \beta_{js}\right| = C\left(\frac{|t-s|}{T}\right)^{1/2}\beta_{jts},$$

where $\beta_{0t}, \beta_{0s}, \beta_{jt}, \beta_{js}, \beta_{jts}, \beta_{0ts} \in \Upsilon(\rho)$, *for some* $\rho > 0$.

7. The elements of $\Lambda_t = \left(\Lambda_{mj,t}\right)$ *and* $\lambda_{0t} = \left(\lambda_{0jt}\right)$ *satisfy*

$$\left|\lambda_{0jt} - \lambda_{0js}\right| = C\left(\frac{|t-s|}{T}\right)^{1/2}\lambda_{0jts}, \tag{16}$$

$$\left|\Lambda_{mj,t} - \Lambda_{mj,s}\right| = C\left(\frac{|t-s|}{T}\right)^{1/2}\Lambda_{mj,ts},$$

for all $j = 1, .., K$, $m = 1, .., M$ *and* λ_{0jt}, λ_{0jts}, $\Lambda_{mj,t}$, $\Lambda_{mj,ts} \in \Upsilon(\rho)$, *for some* $\rho > 0$.

Assumption 2 *For* $C < \infty$ *and any* i, j, t, s, m, v:

1. $E\left(\varepsilon_{it}\varepsilon_{js}\right) = \sigma_{ijts}$, $|\sigma_{ijts}| \leq \overline{\sigma}_{ij}$, *and* $|\sigma_{ij,ts}| \leq \tau_{ts}$ *and*

(a) $N^{-1}\sum_{i,j=1}^{N}\overline{\sigma}_{ij} \leq C$ *(b)* $N^{-1}\sum_{i,s}|\sigma_{ii,ts}| \leq C$
(c) $K_l^{-1}\sum_{t,s}^{T}\tau_{ts} \leq C$, *for all* l *(d)* $N^{-1}K_l^{-1}\sum_{i,j,t,s}|\sigma_{ij,ts}| \leq C$, *for all* l

2. $E\left|N^{-1/2}K_s^{-1/2}\sum_{u=1}^{T}\sum_{i=1}^{N} k_{su}\left[\varepsilon_{it}\varepsilon_{iu} - E\left(\varepsilon_{it}\varepsilon_{iu}\right)\right]\right|^2 \leq C$.

3. $E\left|K_s^{-1/2}\sum_{t=1}^{T} k_{st}F_{m,t}\omega_{v,t}\right|^2 \leq C$, *for the* m, v *elements of vectors* F_t, ω_t.

4. $E\left|K_s^{-1/2}\sum_{t=1}^{T} k_{st}\varepsilon_{it}\omega_{m,t}\right|^2 \leq C$, *for all* $s = 1, .., T$.

Assumption 3 $\Delta_\Phi = I$ *and* $P_1 = 0$, Δ_F *is diagonal, positive definite, with each diagonal element unique.*

Assumption 4 $\Lambda_t = \left[\Lambda_{ft}\ \mathbf{0}\right]$, *and* Λ_{ft} *is non singular.*

Assumption 5 *The bandwidth parameters* H, L *and the cross section size* N *satisfy*

$$c_0 N^{\varepsilon} \leq H, L = o\left(\frac{T}{(\log T)^{\nu}}\right) = o(1), \text{ with } \nu = \frac{\rho+2}{2\rho}, \tag{17}$$

$$\varepsilon > \frac{8}{\theta - 4} > 0, \theta > 8, \rho > 0.$$

The assumptions above are similar to those commonly used in the literature on factor models (see e.g. Stock and Watson (2002a), Stock and Watson (2002b), Bai (2003), Kelly and Pruitt (2015)), but amended to account for the stochastic time variation in the parameters (see e.g. Giraitis *et al.* (2014), Giraitis *et al.* (2018), Giraitis *et al.* (2021)).

In Assumption 1, we require factors and loadings to be cross-sectionally regular in that they have well-behaved covariance matrices for large T and N, respectively. Extending the work of Kelly and Pruitt (2015), we allow for temporal dependence in the factors F_t, through the strong (α-) mixing condition. We also assume that the factors are systematic, in the sense that they affect an infinite number of cross sectional units x_t. This is captured by the full rank assumption of Δ_Φ. We also impose the identification assumption that the second moments of the factors do not vary over time. Note that this is not limiting the generality of our model. For instance, replacing the factors F_t by $\widetilde{F}_t = F_t cov_t(F_t)^{-1/2} \Delta_F^{1/2}$, and the loadings Φ_t by $\widetilde{\Phi}_t = \Phi_t cov_t(F_t)^{1/2} \Delta_F^{-1/2}$, we can ensure that the assumption is satisfied even in the case $cov_t(F_t)$ changes over time.

Moreover, we allow all parameters of the model to vary over time through the assumptions given in equations (14), (15), (16). An example of a process that satisfies this set of equations is the bounded random walk model, other examples are provided in Giraitis *et al.* (2014), Giraitis *et al.* (2018). As discussed, we do not require a specific parametric form for the evolution in Φ_t, Λ_t, β_t, ϕ_{0t}, λ_{0t}, β_t, β_{0t} but allow for general stochastic relationships among model parameters at different points of time. This permits a quite flexible specification, which nests the constant parameter 3PRF, while still preserving good properties for the parameter estimators, as we will see. Furthermore, when all factors of x_t drive the target (and proxies) variable, the model nests the time varying PCA model developed in Su and Wang (2017), but for deterministic processes of the loadings evolution.

In Assumption 2 we allow for some cross sectional correlation and serial dependence (including GARCH effects) among ε_{it} (see also Stock and Watson (2002a)). Moreover, we allow for some limited dependence among the errors of the equations for the proxies and the factors and the idiosyncratic shocks.

Assumption 3 sets the identification conditions. We require the covariance of the predictor loadings to be the identity matrix ($\Delta_\Phi = I$), and the factors to be orthogonal to one another. As with principal components, the particular normalization is not important. We ultimately estimate a vector space spanned by the factors, and this space does not depend upon the choice of normalization.

Assumption 4 states that, for all time periods $t = 1, .., T$, the proxies (i) have zero loading on irrelevant factors, (ii) have linearly independent loadings on the relevant factors, and (iii) their number is equal to the number of relevant factors. Combined with the normalization assumption, this says that the common component of the proxies spans the relevant factor space, and that none of the proxy variation is due to irrelevant factors.

Assumption 5 is a technical requirement on the bandwidth parameters H, L which governs how much information we want to use from the neighboring periods or, put it in another way, our prior assumption about the smoothness of the model parameters, as a function of time. All asymptotic results rely on this condition on H and L.

3 Theoretical Properties of Forecasts and Estimators

With the assumptions introduced in the previous section in place, we now derive the asymptotic properties of the time varying three-pass regression filter. Our proofs build upon the seminal works of Stock and Watson (2002a), Bai (2003), Bai and Ng (2002), Bai and Ng (2006), and Kelly and Pruitt (2015), combined with the theory on stochastic parameter time variation developed in Giraitis *et al.* (2014), Giraitis *et al.* (2018), Giraitis *et al.* (2021), inter alia. The Theorems reported in the main text are our central new results, proofs and additional results are provided in the Appendix.

3.1 Consistency results for the optimal forecast

Assumptions 1-5 contain the general conditions that are sufficient for consistency of the optimal forecasts. We set $S_{zx,H,t} = K_{H,t}^{-1} \sum_{j=1}^{T} k_{H,tj} \left(z_j - \overline{z}_t\right) \left(x_j - \overline{x}_t\right)'$, for $\overline{x}_t = K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} x_s$, $\overline{z}_t = K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} z_s$, for generic vectors x_t, z_t, and bandwidth H. Also, for generic bandwidth L, we define $S_{xy,L,t} = K_{L,t}^{-1} \sum_{j=1}^{T} k_{L,tj} \left(x_j - \overline{x}_t\right) \left(y_{j+1} - \overline{y}_t\right)'$, for $\overline{x}_t = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} x_s$, $\overline{y}_t = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} y_{s+1}$. The next Theorem says that the TV-3PRF is consistent, in the sense that the difference between the feasible and the infeasible forecast vanishes.

Theorem 1 *Let Assumptions 1-5 hold. Then,*

(i) The time varying three pass regression filter is consistent for the infeasible best forecast, that is

$$\widehat{y}_{t+1} = \widehat{\beta}_{0t} + \widehat{F}_t' \widehat{\beta}_t \underset{N,T \to \infty}{\overset{p}{\to}} \beta_{0t} + F_t' \beta_t. \tag{18}$$

(ii) The target equation of the time varying three pass regression filter can be written as

$$\widehat{y}_{t+1} = \overline{y}_t + \left(x_t - \overline{x}_t\right)' \widehat{a}_t + o_p\left(1\right), \tag{19}$$

where $\widehat{a}_t = J_N S_{zx,H,t}' \left(S_{zx,H,t} J_N S_{xx,L,t} J_N S_{zx,H,t}'\right)^{-1} S_{zx,H,t} J_N S_{xy,L,t}$.

(iii) Also, let $\widehat{a}_{it}$ be the i-th element of $\widehat{a}_t$. Then, for all $i = 1, .., N$,

$$N\widehat{a}_{it} \underset{N,T\rightarrow\infty}{\overset{p}{\rightarrow}} \left(\Phi_{it} - \overline{\Phi}_t\right)' \beta_t, \tag{20}$$

where $\left(\Phi_{it} - \overline{\Phi}_t\right)' = S_i J_N \Phi_t$ with S_i being a $1 \times N$ selector vector, with the i-th element equal to 1 and the remaining elements equal to zero.

In the Appendix we also provide probability limits for loadings $\widehat{\Phi}_t$, factors $\widehat{F}_t$ and third stage predictive coefficients $\widehat{\beta}_t$. The developed bounds depend on the kernel approximation and the number of cross-sectional and time-series observations, and in particular on Assumption 5. Lemma A3 given in the Appendix, also provides these results for model (1)-(3) as N, T increase to infinity, according to Assumption 5.

When the parameters are constant, this theorem reduces to the corresponding one in Kelly and Pruitt (2015). Moreover, when there are no irrelevant factors so that the 3PRF and PCA deliver the same factors (possibly up to rotation), Theorem 1 generalizes the consistency of the feasible PC based forecasts obtained in the constant parameter case by Stock and Watson (2002a).

The element (ii) of Theorem 1 provides an alternative characterization of the fitted target equation. This will be our main vehicle to derive asymptotic distributions as well as to estimate the degrees of freedom. In this representation, $\widehat{a}_t$ can be interpreted as the coefficients of the individual predictors x_t.

As noted by KP, but in our time varying framework, $\widehat{a}_t$ links the factors to the target variable via the large set of available indicators. Therefore, the probability limit of $\widehat{a}_t$ (element (iii) of Theorem 1) is the product of the loadings of x_t on the relevant factors f_t ($\left(\Phi_{it} - \overline{\Phi}_t\right)'$) and of those of y_t on the same factors (β_t). Hence, $\widehat{a}_t$ can be also interpreted as a constrained least squares coefficient estimator, and we will explore the implications of this observation later in the paper. It is also interesting to note that $\widehat{a}_t$ is multiplied by N to obtain its (non-degenrate) limit. As in KP, this is due to the fact that the dimension of $\widehat{a}_t$ is the same as the number of predictors, N, so that when N grows the contribution of each predictor in x_t goes to zero.

Next, we consider the issue of the selection of the proxies. If there is a single relevant factor, then the target variable itself, y, can be used as proxy. Following KP, we will refer to this as the target-proxy approach. In a case like this, the resulting 3PRF factor can be quite different from the first principal component, which instead aims at maximizing the explanatory power for x.

A multiple factor extention of the above target-proxy approach is to use the following automatic proxy selection algorithm, which was proposed by KP and also links the 3PRF to partial least squares (PLS), as we will see in more details in the next subsections. This method for proxy selection could be rather helpful in situations where economic theory does not provide enough insights for choosing proxies.

Algorithm 1 Automatic proxy selection

Step 1 Initialize $z_{0,t} = y_{t+1}$, for $k = 1, .., M$ do the following

Step 2 Define the k-th automatic proxy to be $z_{k-1,t}$. Stop if $k = M$; otherwise proceed

Step 3 Compute the TV-3PRF for target y_{t+1} using cross section x_t and statistical proxies 1 to k. Denote the resulting fitted values $\widehat{y}_{t+1}$.

Step 4 Calculate $z_{k,t} = y_{t+1} - \widehat{y}_{t+1}$

The following Corollary is a direct extension of the corresponding one in KP.

Corollary 1 *Let Assumptions 1-5 hold, except those related with the proxy equation and its error. Then the automatic proxy three pass regression filter forecast of y automatically satisfies assumptions 1.4, 2.3, 2.4 when $M = K_f$. As a result the automatic proxy is consistent according to Theorem 1.*

When we know the number of relevant factors, Algorithm 1 constructs proxies that satisfy the relevance criterion in assumption 4. When $K_f = 1$, the Corollary 1 proves that the target proxy is consistent. When $K_f > 1$, the target-proxy does not extract enough factors to attain the infeasible best. In this case, the automatic proxy selection can be used M-times, and Corollary 1 establishes consistency.

We conclude this subsection with a derivation of the degrees of freedom (DoF) of the TV-3PRF, to be used later on to design information criteria for the selection of the number of factors. To do so, we employ the generalized definition of DoF proposed by Efron (2004). According to this, we derive the the DoF as the sensitivity of the fitted values $\hat{y}$, seen solely as functions of y (see also Krämer and Sugiyama (2011)). The next theorem provides the result.

Theorem 2 *Let Assumptions 1-5 hold and define the time t weighted Krylov sequence as $W^{(t)} = \left\{ J_N S_{xy,H,t}, J_N S_{xx,H,t} J_N S_{xy,H,t}, .., (J_N S_{xx,H,t})^{K-1} J_N S_{xy,H,t} \right\}$. The K-factor, auto-proxy TV-3PRF forecast can be implemented in one step as*

$$\widehat{y}_{t+1} = \overline{y}_{L,t} + (x_t - \overline{x}_{L,t})'\widehat{\alpha}_t, \tag{21}$$

with

$$\widehat{a}_t = W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}S_{xy,L,t}. \tag{22}$$

Then, the Efron (2004), DoF is

$$DoF_{H,L,K} = \sum_{s=1}^{T}\frac{k_{L,ss}}{K_{L,s}} + \sum_{s=1}^{T}\left(x_s - \overline{x}_{L,s}\right)'\frac{\partial a_s}{\partial y_{s+1}}, \tag{23}$$

where $\frac{\partial a_t}{\partial y_{t+1}}$ *is the t-th row of the* $T \times N$ *matrix*

$$d\widehat{a}_t = \left(S'_{xy,L,t}\otimes I_N + I_N \otimes S'_{xy,L,t}\right)H^{(t)}_{(1)}\otimes\left(I_N - H^{(t)}S_{xx,L,t}\right)K^{-1}_{H,t}U'_K\Lambda^{(t)}_H\left(I_T - \frac{1}{K_{H,t}}1_{TT}\Lambda^{(2,t)}_L\right)dy \tag{24}$$

$$+K^{-1}_{L,t}H^{(t)}\widetilde{X}^{(t)\prime}_L\Lambda^{(t)}_L\left(I_T - \frac{1}{K_t}1_{TT}\Lambda^{(2,t)}_L\right)dy$$

and $H^{(t)}_{(1)} = W^{(t)}\left(W^{(t)\prime}S^{(t)}_{xx,L,t}W^{(t)}\right)^{-1}$, $U_K = \left\{\widetilde{X}^{(t)}_H J_N, \widetilde{X}^{(t)}_H J_N S_{xx,H,t}J_N, .., \widetilde{X}^{(t)}_H J_N\left(S_{xx,H,t}J_N\right)^{K-1}\right\}$, $1_{TT} = 1_T 1'_T$, $\Lambda^{(2,t)}_L = \Lambda^{(t)}_L\Lambda^{(t)}_L$, $\Lambda^{(t)}_H = diag\left(\sqrt{k_{H,1t}}, \sqrt{k_{H,2t}}, .., \sqrt{k_{H,Tt}}\right)$, $H^{(t)} = H^{(t)}_{(1)}W^{(t)\prime}$, $\widetilde{X}^{(s)} = \left[k^{1/2}_{s,1}\widetilde{x}_{s,1}, .., k^{1/2}_{s,T}\widetilde{z}_{s,T}\right]'$, *for* $\widetilde{x}_{s,t} = x_t - \overline{x}_s$, *with* $\overline{x}_t = K^{-1}_t\sum_{s=1}^{T}k_{t,l}x_l$.

In empirical applications, the DoF can be used, together with an information criterion (IC), to decide on the number of factors and the first and third step bandwidth parameters H, L. All ICs are functions of the Degrees of Freedom (DoF) parameter. For instance, the Bayesian IC (BIC) for a model with K factors can be written as

$$BIC\left(K\right) : \|\widehat{y}_K - y\|^2 + \log T\sigma^2_\eta DoF\left(K\right), \tag{25}$$

where σ^2_η is the noise level which can be estimated from the residuals $\widehat{\eta}_{t+1}$. Hence, the optimal H, L, K can be chosen as the value that minimizes $BIC\left(K\right)$. Bai and Ng (2002) also use information criteria to determine the number of factors by focusing on penalty terms that account for large N, T dimensions, rather than the Degrees of Freedom, which they set equal to the number of factors.

3.2 Asymptotic distribution of the optimal forecast

In this section we provide theoretical results on the asymptotic distributions of the optimal forecasts and underlying quantities. These are mainly based on the asymptotic distributions of three quantities: $K^{-1}_t\sum_{s=1}^{T}k_{t,s}F_s\eta_{s+1}$, $K^{-1}_t\sum_{s=1}^{T}k_{t,s}F_s\varepsilon_{is}$, and $N^{-1}\sum_{i=1}^{N}\Phi_{is}\varepsilon_{it}$. Under Assumption 1, the limiting distributions of these quantities are provided in the Theorem 3, elements (i), (ii), (iii) below.

Theorem 3 *Let Assumption 1-5 hold then,*

(i) Let $S_{F\eta,t} = K_t^{-1} \sum_{s=1}^{T} k_{t,s} F_s \eta_{s+1}$ *,* $\Gamma_{F\eta,t} = \lim_{T\to\infty} K_{2,t}^{-1} \sum_{s=1}^{T} k_{t,s}^2 E\left(F_s F_s' \eta_{t+1}^2\right)$*, then*

$$\frac{K_t}{K_{2,t}^{1/2}} \Gamma_{F\eta,t}^{-1/2} S_{F\eta,t} \xrightarrow{d} N(0, I). \tag{26}$$

(ii) Let $C_{F\varepsilon,it} = K_t^{-1} \sum_{s=1}^{T} k_{t,s} F_s \varepsilon_{is}$*, Let* $\Gamma_{F\varepsilon,it} = \lim_{T\to\infty} K_{2,t}^{-1} \sum_{s=1}^{T} \sum_{r=1}^{T} k_{t,s} k_{t,r} E\left(F_s F_r' \varepsilon_{is} \varepsilon_{ir}\right)$*, then*

$$\frac{K_t}{K_{2,t}^{1/2}} \Gamma_{F\varepsilon,it}^{-1/2} C_{F\varepsilon,it} \xrightarrow{d} N(0, I). \tag{27}$$

(iii) Let $\Gamma_{\Phi\varepsilon,st} = p\lim N^{-1} \sum_{i=1}^{N} E\left(\Phi_{is} \Phi_{is}' \varepsilon_{it}^2\right)$*,* $C_{\phi\varepsilon,ist} = \frac{1}{N} \sum_{i=1}^{N} \Phi_{is} \varepsilon_{it}$*. For every* $s, t = 1, .., T$ *it holds that*

$$N^{1/2} \Gamma_{\Phi\varepsilon,st}^{-1/2} C_{\phi\varepsilon,ist} \xrightarrow{d} N(0, I). \tag{28}$$

(iv) Let $\widehat{a}_t$ *as in (19), and* $\widetilde{a}_{it} = S_i G_{a,t} \beta_t$*, where* S_i *is a selector vector, then*

$$\frac{\sqrt{K_t} N \left(\widehat{a}_{it} - \widetilde{a}_{it}\right)}{A_{it}} \xrightarrow{d} N(0, 1). \tag{29}$$

where, $G_{a,t} = J_N S_{zx,H,t}' \left(\frac{1}{N^2} S_{zx,H,t} J_N S_{xx,L,t} J_N S_{zx,H,t}'\right)^{-1} \frac{1}{N} S_{zx,H,t} J_N S_{xf,L,t}$ *and* A_{it}^2 *is the i-th diagonal element of* $\widehat{Avar}\left(\widehat{a}_{it}\right)$*, with*

$$\widehat{Avar}\left(\widehat{a}_{it}\right) = \Omega_{a,t} \left(K_{L,t}^{-1} \sum_{j=1}^{T} k_{L,tj} \widehat{\eta}_{j+1}^2 \left(x_j - \overline{x}_t\right) \left(x_j - \overline{x}_t\right)' \right) \Omega_{a,t}', \tag{30}$$

for $\Omega_{a,t} = J_N S_{zx,H,t}' \left(N^{-2} S_{zx,H,t} J_N S_{xx,L,t} J_N S_{zx,H,t}'\right)^{-1} N^{-1} S_{zx,H,t} J_N$*, and*
$S_{xf,L,t} = K_{L,t}^{-1} \sum_{j=1}^{T} k_{L,tj} \left(x_j - \overline{x}_t\right) \left(F_j - \overline{F}_t\right)'$*, for* $\overline{F}_t = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} F_s$*.*

(v) The automatic proxy three pass regression filter forecast given by Algorithm 1 is asymptotically normal.

The element (iv) of Theorem 3 establishes the asymptotic distribution of the coefficients $\widehat{a}_t$ which can then be used to construct the optimal forecast $\widehat{y}_{t+1}$. Moreover, it can be employed to derive feasible t- and F-statistics for inference on a_t, for example to decide whether or not a single component of x_t has predictive content for y_{t+1}. Note that the matrix $G_{a,t}$ appears in $\widetilde{a}_{it}$ because factors are only identified up to an orthonormal rotation. Next, using the results for $\widehat{a}_t$, we can derive the asymptotic distribution of the TV-3PRF forecast, which generalizes element (v) of Theorem 3.

Theorem 4 *Let Assumptions 1-5 hold. Then*

(i) For $E_t y_{t+1} = \beta_{0t} + F_t' \beta_t$ *it holds that*

$$\frac{\sqrt{K_{L,t}}\left(\widehat{y}_{t+1} - E_t y_{t+1}\right)}{Q_t} \xrightarrow{d} N\left(0,1\right), \tag{31}$$

where Q_t^2 *is the i-th diagonal element of* $N^{-2}\left(x_t - \overline{x}_{L,t}\right)' \widehat{Avar}\left(\widehat{a}_t\right)\left(x_t - \overline{x}_{L,t}\right)$.

(ii) Let $\Sigma_{\beta,t} = \Sigma_{z,t}^{-1}\Gamma_{F\eta,t}\Sigma_{z,t}^{-1}$, $\Sigma_{z,t} = \Lambda_t \Delta_F \Lambda_t + \Delta_\omega$ *and* $G_{\beta,t} = \widehat{\beta}_{1,t}^{-1}\widehat{\beta}_{2,t}\widehat{\beta}_{3,t}^{-1} N^{-1} S_{zx,H,t} J_N S_{xf,L,t}$, *where* $\widehat{\beta}_{1,t}$, $\widehat{\beta}_{2,t}$, $\widehat{\beta}_{3,t}$, *given in Lemma A5 of the Appendix. Then*

$$\sqrt{K_t}\left(\widehat{\beta}_t - G_{\beta,t}\beta_t\right) \xrightarrow{d} N\left(0, \Sigma_{\beta,t}\right), \tag{32}$$

where $\widehat{Avar}\left(\widehat{\beta}_t\right) = \Omega_{\beta,t} K_{L,t}^{-1} \sum_{j=1}^{T} k_{L,tj}\widehat{\eta}_{j+1}^2 \left(\widehat{F}_j - \overline{\widehat{F}}_t\right)\left(\widehat{F}_j - \overline{\widehat{F}}_t\right)' \Omega_{\beta,t}$, *and*

$$\Omega_{\beta,t} = \left(K_{L,t}^{-1} \sum_{j=1}^{T} k_{L,tj}\left(\widehat{F}_j - \overline{\widehat{F}}_t\right)\left(\widehat{F}_j - \overline{\widehat{F}}_t\right)'\right)^{-1}$$

This Theorem shows that the optimal TV-3PRF forecast is asymptotically normal, as in the case of the static 3PRF, and it provides a standard error estimator for constructing predictive intervals that depends on $\widehat{Avar}\left(\widehat{a}_t\right)$, the estimated variance covariance matrix of $\widehat{a}_t$ (rather than on $\widehat{Avar}\left(\widehat{F}\right)$ and $\widehat{Avar}\left(\widehat{\beta}\right)$, which would be slightly more complex to derive). The second element of this Theorem (element (ii)), provides the asymptotic distribution of the predictive loadings on the latent factors (β_t) and a consistent estimator of their asymptotic covariance matrix, which can be relevant for inference on the factor coefficients when the forecasting equation is written as in (1).

Finally, we derive the asymptotic distribution of the estimated TV-3PRF factors.

Theorem 5 *Let* $H_{0,t} = \widehat{F}_{A,t}\widehat{F}_{B,t}^{-1} N^{-1} K_{H,t}^{-1} \widetilde{Z}_H^{(t)\prime} \widetilde{X}_H^{(t)} J_N \phi_{0t}$, *and* $H_t = \widehat{F}_{A,t}\widehat{F}_{B,t}^{-1} N^{-1} K_{H,t}^{-1} \widetilde{Z}_H^{(t)\prime} \widetilde{X}_H^{(t)} J_N \Phi_t F_t$, *as in (106), (110). Under Assumptions 1-5 we have for every t,*

(i) if $\frac{\sqrt{N}}{K_{H,t}} \to 0$, *then*

$$\sqrt{N}\left(\widehat{F}_t - \left(H_{0,t} + H_t F_t\right)\right) \xrightarrow{d} N\left(0, \Sigma_{F,t}\right),$$

where $\Sigma_{F,t} = \left(\Lambda_t \Delta_F \Lambda_t' + \Delta_{\omega,t}\right)\left(\Lambda_t \Delta_F \Delta_F \Lambda_t'\right)^{-1} \Lambda_t \Delta_F \Gamma_{F\varepsilon,t} \Delta_F \Lambda_t' \left(\Lambda_t \Delta_F \Delta_F \Lambda_t'\right)^{-1}\left(\Lambda_t \Delta_F \Lambda_t' + \Delta_\omega\right)$.

(ii) if $\liminf \sqrt{N}/K_{H,t} \geq \tau \geq 0$, *then*

$$K_{H,t}\left(\widehat{F}_t - \left(H_{0,t} + H_t F_t\right)\right) = O_p\left(1\right).$$

In Theorems 4 and 5 the matrices $G_{\beta,t}$ and H_t exist since we are estimating a vector space. As in KP, we do not provide an estimator for $\Sigma_{F,t}$ since the presence of irrelevant factors complicates its derivation. This is also another reason for using the representation of the forecast equation in terms of $\widehat{a}_t$ and x_t to derive the asymptotic distribution of the TV-3PRF forecast.

3.3 Relation with Partial and Constrained Least Squares

The method of partial least squares (PLS) is a special case of the fixed parameter three pass regression filter. In particular, KP showed that the PLS forecasts are identical to those from the fixed parameter 3PRF, when (i) the predictors are first demeaned and variance standardized, (ii) the first two regressions run without the constant terms, and (iii) proxies are automatically selected. The fixed parameter PLS forecast (see Wold (1975)) is constructed with the following steps.

Algorithm 2 Fixed parameter Partial Least Squares

Step 1 For $i = 1, .., N$, set the scalar $\widehat{\phi}_i = x_i'y$ and $\widehat{\Phi} = \left(\widehat{\phi}_1, \widehat{\phi}_2, ..., \widehat{\phi}_N\right)'$

Step 2 Set $\widehat{u}_t = x_t'\widehat{\Phi}$, and $\widehat{u} = (\widehat{u}_1, \widehat{u}_2, ..., \widehat{u}_T)'$

Step 3 Run the predictive regression of y on $\widehat{u}$

For one factor, and in one step, Algorithm 2 can be represented as $\widehat{y}^{PLS}_{t+1} = \overline{y} + x_t'\widehat{\alpha}_1$, for $\widehat{\alpha}_1 = \left(X'y\left(y'XX'XX'y\right)^{-1}y'XX'y\right)$. For m factors, it can be shown that the PLS estimate becomes $\widehat{\alpha}_m = W\left(W'SW\right)^{-1}W's$, for $S = X'X$, $s = X'y$ and W is the Krylov sequence $\left\{s, Ss, ..., S^{m-1}s\right\}$.

Inspection of Theorem 2 allows us to infer the relationship between $\widehat{y}^{PLS}_{t+1}$ and $\widehat{y}_{t+1}$ implied by the TV-3PRF, implemented via the automatic proxy approach. In particular, the autoproxy estimate of $\widehat{\alpha}_t$ given by equation (22), implies that the one step, multiple factor, autoproxy estimate of TV-3PRF is identical to a time varying partial least squares method, when we first demean and standardize predictors by their kernel based estimates.

To understand the relationship between the TV-3PRF and (TV-) constrained least squares, we note that equation (19) allow us to infer the contribution of the i-th predictor x_{it}, when it is combined with the remaining predictors. To this end, $\widehat{a}_t$ is a projection coefficient relating y_{t+1} to x_t under the constraint that irrelevant factors do not influence forecasts. As in the fixed parameter case, the TV-3PRF forecast may be derived as the solution to a constrained least squares problem, as shown in the next theorem.

Theorem 6 *Let Assumptions 1-5 hold. The time varying three pass regression filter's implied N-dimensional predictive coefficient $\widehat{a}_t$, given in equation (19), is the solution to the problem*

$$\arg\min_{a_{0t},a_t} \sum_{s=1}^{T} k_{L,st} \left(y_{s+1} - a_{0t} - x_s' a_t\right)^2 \tag{33}$$

subject to the constraint

$$(I - W_{xz,H,t} \left(S_{zx,H,t} W_{xz,H,t}\right)^{-1} W_{xz,H,t}') a_t = 0,$$

where $W_{xz,H,t} = J_N S_{xz,H,t}$

The constraint in Theorem 6 has an intuitive explanation. Premultipling by $(x_t - \overline{x}_{H,t})'$ we have that $((x_t - \overline{x}_t)' - (x_t - \overline{x}_t)' W_{xz,H,t} \left(S_{zx,H,t} W_{xz,H,t}\right)^{-1} W_{xz,H,t}') a_t = 0$. Notice that, as in the proof of (19), it holds that

$$(x_t - \overline{x}_t)' W_{xz,H,t} \left(S_{zx,H,t} W_{xz,H,t}\right)^{-1} S_{zz,H,t} = \widehat{F}_t - K_H^{-1} \sum_{s=1}^{T} k_{H,st} \widehat{F}_{H,s} + o_p(1)$$

Then the constraint becomes

$$\begin{aligned} (x_t - \overline{x}_t) - \widehat{\Phi}_t \left(\widehat{F}_t - \overline{\widehat{F}}_t\right) &= \widehat{\varepsilon}_t \\ &\approx \varepsilon_t + \Phi_{gt} g_t, \end{aligned}$$

as in Lemma A6 and where $g_t$are the irrelevant factors. Since the covariance of a_t,ε_t is zero, the product of a_t with the target irrelevant common component $\Phi_{gt} g_t$ is equal to zero. This constraint ensures that factors irrelevant to y_{t+1} drop out of the TV-3PRF forecast.

3.4 Selection of the Tuning Parameters

Application of the TV-3PRF requires the selection of the bandwidth parameter for the passes 1 and 3 of the estimation algorithm. An information criterion combined with the degrees of freedom discussed above provides an avenue for selecting this parameter. Another possibility is the use of a cross validation method based on the mean squared forecast error (MSFE).

Pass 1 and 3 of the TV-3PRF depend on bandwidths H and L, respectively. These two parameters can be obtained as the optimization of an end of sample cross validation scheme, as follows,

$$\{H, L\} = \arg\min \frac{1}{\kappa} \sum_{j=T-\kappa+1}^{T} \left(y_j - \widehat{\beta}_{0j-1} + \widehat{\beta}_{j-1}' \widehat{F}_{j-1}\right)^2, \tag{34}$$

where κ is the number of observations over which the MSFE is computed. In objective (34), the H parameter does not vary with $i = 1, .., N$. This reduces the flexibility of our approach to capture varying degrees of time variation in loadings processes, but facilitates the implementation of the TV-3PRF by selecting only two parameters. It also leads to faster computation, especially when N is large. Allowing the bandwidth H to vary over i (i.e. H_i), in (34) can be a viable option only when N is small enough, as in this case objective (34) will need to be optimized over $N + 1$ bandwidth parameters ($\{H_i\}_{i=1}^N, L$).

To allow for different bandwidths across the large set of N series, without exploding the computational burden, we propose a parametrisation of H_i that depends on the volatility of series x_{it}. To this end, let σ_i denote the volatility of the i-th series, and define the adjustment factor for series i as

$$c_i = 1 + \frac{\sigma_i - \min(\sigma_i)}{\max(\sigma_i) - \min(\sigma_i)}(d_H - 1), for d_H \in (0, 1), i = 1, .., N. \tag{35}$$

The c_i maps the volatilities of x_{it} on the interval $[d_H, 1]$. The highest the σ_i, the lowest the c_i. Setting $H_i = c_i H$, allows us to link individual bandwidths H_i with the global bandwidth parameter H and the individual adjustment factors c_i. When we choose $d_H = 1$, there is no heterogeneity for the bandwidths ($H_i = H$). As we deviate from 1, we allow for higher degree of heterogeneity across the bandwidths H_i. At the end, we can choose a specific value for d_H (e.g. $d_H = 0.7$) and optimize objective (34), or include the d_H in the objective of an end of sample cross validation, that is

$$\{d_H, H, L\} = \arg\min \frac{1}{\kappa} \sum_{j=T-\kappa+1}^{T} \left(y_j - \widehat{\beta}_{0j-1} + \widehat{\beta}'_{j-1}\widehat{F}_{j-1} \right)^2. \tag{36}$$

In our Monte Carlo section we fix $d_H = 0.7$ while in our empirical section we also choose it through cross validation.

4 Monte Carlo Evaluation

4.1 Design

We generate data according to the following model

$$x_t = \phi_{f,t} f_t + \phi_{g,t} g_t + \kappa \varepsilon_t \tag{37}$$

$$y_{t+1} = f_t + \eta_t \tag{38}$$

with $\eta_t \sim N\left(0, \sigma_\eta^2\right)$, where σ_η^2 is adjusted such that the infeasible best forecast has an R^2 of 50%. This R^2 was also chosen by Kelly and Pruitt (2015). We consider two sample sizes with $T = \{200, 100\}$, and $N = \{200, 100\}$. The relevant (f_t) and irrelevant (g_t) factors are generated as

$$f_{jt} = \rho_f f_{jt-1} + u_{jt}, u_{jt} \sim N(0,1) \tag{39}$$

$$g_{lt} = \rho_g g_{lt} + v_{lt}, v_{lt} \sim N\left(0, \sigma_l^2\right) \tag{40}$$

for $j = 1, .., k_f$, $l = 1, .., k_g$. We choose $k_f = 1$, and $k_g = 4$ and σ_l^2, $l = 2, .., 4$, such that the irrelevant factors are dominant, in the sense that they have variances 1.25, 1.75, 2.25 and 2.75 times larger than the relevant factor. These choices make the relevant factor estimation rather difficult, as the factor that drives y_{t+1} explains, overall, a small fraction of variability of x.

For the idiosyncratic terms, we assume autoregressive dynamics

$$\varepsilon_{it} = a\varepsilon_{it-1} + \varsigma_{it}$$

and cross sectional correlation specified via

$$\varsigma_t = (1 + d^2)\nu_{i,t} + d\nu_{i-1,t} + d\nu_{i+1,t},$$

where $\nu_{i+1,t}$ is standard normal and the cross sectional parameter takes values of 0 and 1.

The κ parameter in the large dataset equation (37) controls the strength of the factor model, such that the factor component $\phi_{f,t} f_t + \phi_{g,t} g_t$ explains 10% or 30% of the variation of x_{it} on average, as measured by

$$R_i^2 = \frac{\widehat{Var_i(\phi_{f,i,t} f_t + \phi_{g,i,t} g_t)}}{\widehat{Var_i(x_{it})}}. \tag{41}$$

The factor loadings for the five factors $r = \{f, g_1, .., g_4\}$, are generated according to the following models

$$\phi_{rit} = \frac{u_{rit}}{\sqrt{t}}, \tag{42}$$

$$\phi_{rit} = u_{rit}, \tag{43}$$

$$\phi_{rit} = \varepsilon_{ri}, \tag{44}$$

with $u_{rit} = u_{rit-1} + \varepsilon_{rit}^u$, $\varepsilon_{rit}^u \sim N(0,1)$, $\varepsilon_{ri} \sim N(0,1)$, for $r = \{f, g_1, .., g_4\}$. The loadings are centered on zero to have a balanced contribution of the common component across the time dimension.

The process defined in (42) is a bounded random walk process, that satisfies the set of assumptions discussed in Section 2.3, as discussed in Giraitis *et al.* (2018). The second process considered, (43), is a random walk process that is not covered by the theory, but since this is a common assumption in the empirical macroeconomic literature we will assess the performance of TV-3PRF also in this case. The third process, defined in (44), has fixed loadings.

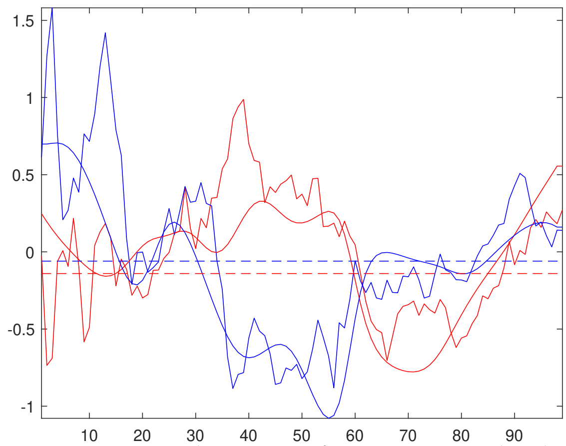

Figure 1: The more volatile lines are two realizations of time varying loadings generated by the bounded random walk model (42). The smoothed curved lines are the TV-3PRF estimates of loadings, and the dashed lines are the fixed parameter 3PRF estimated loadings.

In Figure 1 we present two realizations of the loadings process (42), which generate the large dataset x_{it}, their kernel based time varying estimates, as well as the fixed parameter 3PRF loadings estimates. The fixed parameter estimates of loadings capture only on average the path of the true loading process. On the other hand, our proposed kernel based estimates are able to capture the upward and downward trends of the true process. As we will see, this will imply a better forecasting performance

In our Monte Carlo exercise we compare the performance of the TV-3PRF with that of the standard 3PRF (benchmark), standard principal component (PCA), and two targeted predictors PCA methods (see Bai and Ng (2008)). In the first one, we first reduce the N-dimension of the large dataset x by choosing the 30 most relevant regressors identified by regressing y_{t+1} on x_t by the LARS algorithm (see Efron *et al.* (2004)); then the reduced set of regressors is used to extract factors, through PCA (PCA-lars). In the second targeted predictors approach, the target variable y_{t+1} is regressed on each x_{it}, and the regressor x_{it} is considered as relevant when its coefficient is significant at the 10% level (PCA-ht(10%)). All the various methods are used to forecast the last observation of the target variable, and average (across all simulations) relative mean squared forecast errors are computed.

Time Varying Loadings, MSFE, DGP given by (42)																		
ρ_f	0	0.3	0.3	0.3	0.3	0.9	0.9	0.9	0.9	0	0.3	0.3	0.3	0.3	0.9	0.9	0.9	0.9
ρ_g	0	0.9	0.9	0.9	0.9	0.3	0.3	0.3	0.3	0	0.9	0.9	0.9	0.9	0.3	0.3	0.3	0.3
β	0	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1
a	0	0.3	0.3	0.9	0.9	0.3	0.3	0.9	0.9	0	0.3	0.3	0.9	0.9	0.3	0.3	0.9	0.9
	R^2 =0.1, N =100, T =100									R^2 =0.1, N =200, T =200								
$pca(5)$	0.84	0.87	0.88	0.92	1	0.99	0.93	1.07	1.05	0.83	0.87	0.88	0.93	0.95	0.94	0.91	1.18	1.12
$pca(1)$	0.8	0.82	**0.85**	**0.86**	0.99	0.98	0.9	1.28	1.19	0.81	0.85	0.86	**0.92**	0.94	0.93	0.88	1.25	1.21
$pca\text{-}lars(1)$	0.82	0.84	0.85	0.87	0.99	0.99	0.92	1.16	1.16	0.84	0.88	0.91	0.94	0.95	1.01	0.92	1.1	1.12
$pca\text{-}ht(10\%)(1)$	0.91	0.91	0.93	0.92	1	0.99	0.92	1.16	1.08	0.87	0.9	0.91	0.95	0.96	0.97	0.93	1.14	1.1
$3prf(1)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	0.8	0.81	0.87	0.95	0.96	**0.87**	0.87	**0.87**	**0.88**	0.76	0.81	0.82	0.97	**0.93**	**0.78**	**0.76**	0.9	0.92
$tv\text{-}3prf(1, ic_{H,L})$	0.8	0.82	0.88	0.94	**0.94**	0.9	0.9	0.95	0.94	**0.73**	**0.8**	**0.8**	0.96	0.94	0.82	0.8	0.94	0.95
$tv\text{-}3prf_{H_i}(1)$	**0.79**	**0.8**	0.87	0.99	0.97	0.87	**0.86**	0.87	0.88	0.76	0.82	0.82	0.96	0.93	0.78	0.77	**0.88**	**0.91**
	R^2 =0.3, N =100, T =100									R^2 =0.3, N =200, T =200								
$pca(5)$	0.87	0.87	0.86	0.91	0.97	0.95	0.94	1.11	1.08	0.84	0.92	0.88	0.96	0.94	0.96	0.96	1.15	1.11
$pca(1)$	0.81	0.82	0.84	**0.85**	0.92	0.91	0.9	1.2	1.15	0.81	0.88	0.85	0.95	0.92	0.95	0.97	1.25	1.1
$pca\text{-}lars(1)$	0.83	0.84	0.86	0.86	0.91	0.94	0.94	1.13	1.13	0.82	0.9	0.85	0.96	0.93	0.94	0.99	1.11	1.07
$pca\text{-}ht(10\%)(1)$	0.91	0.88	0.91	0.91	0.96	0.94	0.95	1.1	1.07	0.85	0.93	0.87	0.95	0.95	0.98	0.97	1.16	1.09
$3prf(1)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	0.72	**0.73**	0.75	0.86	0.86	**0.78**	0.77	**0.92**	0.97	0.67	0.69	0.69	0.82	**0.86**	0.73	0.74	**0.87**	**0.81**
$tv\text{-}3prf(1, ic_{H,L})$	**0.69**	0.74	**0.71**	0.88	**0.84**	0.82	0.79	0.93	**0.91**	**0.64**	**0.65**	**0.66**	0.82	0.87	**0.69**	**0.72**	0.89	0.87
$tv\text{-}3prf_{H_i}(1)$	0.73	0.73	0.75	0.85	0.86	0.8	**0.76**	0.94	0.92	0.7	0.71	0.7	**0.81**	0.86	0.73	0.76	0.89	0.81

Table 1: Reported results are MSFE of each model over 1000 simulations. The $pca(5)$ corresponds to the PCA, five factors method, the $pca\text{-}lars(1)$ corresponds to the targeted predictors, using $lars$ algorithm, one factor model, the $pca\text{-}ht(10\%, 1)$ corresponds to the targeted predictors, using t-test at 10% significance level, one factor model. The $3prf$ corresponds to the fixed parameter, one autoproxy factor, three pass regression filter, and $tv\text{-}3prf$, is the time varying parameter, one autoproxy factor, three pass regression filter, where the bandwidth is selected through cross validation (see $tv - 3prf(1)$ for (34) and $tv\text{-}3prf_{H_i}(1)$ for (36)), or through the BIC criterion (see $tv\text{-}3prf(1, ic_{H,L})$). The best performing is in red.

Time Varying Loadings, MSFE, DGP given by (43)																		
ρ_f	0	0.3	0.3	0.3	0.3	0.9	0.9	0.9	0.9	0	0.3	0.3	0.3	0.3	0.9	0.9	0.9	0.9
ρ_g	0	0.9	0.9	0.9	0.9	0.3	0.3	0.3	0.3	0	0.9	0.9	0.9	0.9	0.3	0.3	0.3	0.3
β	0	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1
a	0	0.3	0.3	0.9	0.9	0.3	0.3	0.9	0.9	0	0.3	0.3	0.9	0.9	0.3	0.3	0.9	0.9
	R^2 =0.1, N =100, T =100									R^2 =0.1, N =200, T =200								
$pca(5)$	0.85	0.88	0.88	0.96	0.94	1.01	0.98	1.06	1.1	0.79	0.8	0.78	0.98	0.95	0.97	0.91	1.1	1.15
$pca(1)$	0.8	0.86	0.83	0.91	**0.9**	0.94	0.97	1.16	1.18	0.76	0.77	0.77	0.99	0.94	0.94	0.89	1.19	1.2
$pca\text{-}lars(1)$	0.82	0.89	0.85	0.94	0.92	0.99	0.95	1.07	1.14	0.79	0.81	0.79	0.96	0.97	0.99	0.96	1.05	1.08
$pca\text{-}ht(10\%)(1)$	0.86	0.97	0.89	0.97	0.94	1	0.99	1.07	1.12	0.81	0.84	0.86	0.99	0.97	1	0.95	1.09	1.11
$3prf(1)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	**0.78**	**0.77**	0.8	**0.9**	0.96	**0.8**	**0.86**	**0.94**	**0.9**	0.65	0.68	0.68	**0.9**	0.94	**0.72**	**0.75**	**0.92**	0.86
$tv\text{-}3prf(1, ic_{H,L})$	0.78	0.77	0.83	0.92	0.93	0.87	0.89	0.94	0.94	**0.62**	**0.67**	**0.67**	0.92	**0.93**	0.75	0.79	0.95	0.91
$tv\text{-}3prf_{H_i}(1)$	0.79	0.77	**0.78**	0.92	0.97	0.8	0.86	0.95	0.9	0.67	0.68	0.67	0.91	0.95	0.73	0.76	0.95	**0.85**
	R^2 =0.3, N =100, T =100									R^2 =0.3, N =200, T =200								
$pca(5)$	0.78	0.84	0.84	0.99	0.93	0.92	0.92	1.09	1.09	0.75	0.82	0.84	0.88	0.94	0.88	0.9	1.18	1.1
$pca(1)$	0.7	0.79	0.77	0.92	0.92	0.86	0.92	1.21	1.25	0.69	0.77	0.79	0.85	0.92	0.91	0.88	1.19	1.12
$pca\text{-}lars(1)$	0.74	0.81	0.79	0.91	0.94	0.89	0.93	1.16	1.21	0.73	0.79	0.83	0.86	0.95	0.91	0.94	1.13	1.1
$pca\text{-}ht(10\%)(1)$	0.83	0.88	0.91	0.95	0.96	0.96	0.96	1.12	1.15	0.79	0.83	0.88	0.93	0.96	0.93	0.95	1.12	1.05
$3prf(1)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	0.6	0.67	**0.62**	0.78	0.78	0.7	0.77	0.94	0.95	0.56	0.59	0.61	**0.72**	0.76	0.69	0.66	**0.84**	0.83
$tv\text{-}3prf(1, ic_{H,L})$	**0.59**	0.66	0.64	0.8	0.79	0.71	0.76	**0.9**	**0.93**	**0.51**	**0.57**	**0.59**	0.73	**0.75**	**0.67**	**0.64**	0.85	**0.82**
$tv\text{-}3prf_{H_i}(1)$	0.6	**0.65**	0.64	**0.74**	**0.77**	**0.69**	**0.75**	0.99	0.99	0.55	0.6	0.61	0.72	0.76	0.7	0.67	0.87	0.86

Table 2: See notes in Table 1

Constant Loadings, MSFE, DGP given by (44)																		
ρ_f	0	0.3	0.3	0.3	0.3	0.9	0.9	0.9	0.9	0	0.3	0.3	0.3	0.3	0.9	0.9	0.9	0.9
ρ_g	0	0.9	0.9	0.9	0.9	0.3	0.3	0.3	0.3	0	0.9	0.9	0.9	0.9	0.3	0.3	0.3	0.3
β	0	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1
a	0	0.3	0.3	0.9	0.9	0.3	0.3	0.9	0.9	0	0.3	0.3	0.9	0.9	0.3	0.3	0.9	0.9
	$R^2=0.1, N=100, T=100$									$R^2=0.1, N=200, T=200$								
$pca(5)$	1.04	1.06	1.06	1.12	1.06	1.08	1.12	1.14	1.09	**0.97**	1.02	1.14	1.22	1.2	**0.97**	1.1	1.21	1.2
$pca(1)$	1.17	1.17	1.09	1.12	1.02	1.23	1.26	1.37	1.25	1.41	1.45	1.31	1.21	1.19	1.35	1.33	1.38	1.36
$pca\text{-}lars(1)$	1.15	1.13	1.13	1.11	1.03	1.14	1.18	1.22	1.2	1.26	1.19	1.14	1.17	1.18	1.24	1.2	1.1	1.09
$pca\text{-}ht(10\%)(1)$	1.05	1.07	1.1	1.07	1.03	1.02	1.1	1.23	1.17	1.17	1.19	1.14	1.15	1.16	1.12	1.15	1.21	1.19
$3prf(1)$	**1**	1	**1**	**1**	**1**	1	**1**	1	1	1	**1**	**1**	**1**	**1**	1	**1**	1	1
$tv\text{-}3prf(1)$	1.09	1.08	1.09	1.07	1.08	1.02	1.1	1	0.99	1.26	1.26	1.22	1.12	1.19	1.1	1.13	**0.98**	**0.99**
$tv\text{-}3prf(1, ic_{H,L})$	1.01	**0.99**	1	1	1.01	**0.99**	1	**0.97**	**0.98**	1	1	1	1.02	1.02	1.01	1	1	1
$tv\text{-}3prf_{H_i}(1)$	1.11	1.07	1.11	1.12	1.08	1.05	1.08	1.01	0.99	1.27	1.27	1.24	1.2	1.25	1.11	1.15	1.01	1.03
	$R^2=0.3, N=100, T=100$									$R^2=0.3, N=200, T=200$								
$pca(5)$	**0.93**	**0.93**	**0.97**	1.26	1.29	**0.9**	**0.99**	1.14	1.14	**0.92**	**0.96**	**0.94**	1.31	1.38	**0.92**	**0.96**	1.19	1.21
$pca(1)$	1.56	1.46	1.55	1.4	1.44	1.64	1.56	1.46	1.38	1.84	1.64	1.65	1.51	1.54	1.81	1.74	1.66	1.58
$pca\text{-}lars(1)$	1.43	1.37	1.27	1.32	1.36	1.36	1.33	1.2	1.23	1.11	1.14	1.17	1.15	1.2	1.17	1.12	1.05	1.07
$pca\text{-}ht(10\%)(1)$	1.31	1.26	1.24	1.3	1.33	1.31	1.24	1.18	1.16	1.59	1.44	1.5	1.38	1.44	1.56	1.5	1.24	1.23
$3prf(1)$	1	1	1	**1**	**1**	1	1	**1**	1	1	1	1	**1**	**1**	1	1	**1**	**1**
$tv\text{-}3prf(1)$	1.29	1.17	1.23	1.09	1.15	1.21	1.29	1.1	1.12	1.44	1.28	1.3	1.21	1.15	1.32	1.32	1.16	1.09
$tv\text{-}3prf(1, ic_{H,L})$	1	1	1	1.01	1.02	1.01	1	1	**0.98**	1	1	1	1.01	1.01	1	1	1	1
$tv\text{-}3prf_{H_i}(1)$	1.31	1.2	1.21	1.15	1.24	1.26	1.34	1.11	1.13	1.45	1.32	1.35	1.28	1.23	1.38	1.36	1.19	1.13

Table 3: See notes in Table 1

4.2 Results

Tables 1-3 present MSFE results for, respectively, the loadings specifications in (42), (43) and (44). Table 4 presents the numbers of factors determined by the BIC-type information criterion in (25) for the loadings specifications in (42). Some interesting findings emerge from these tables.

First, focusing on the bounded random walk evolution of loadings (see Table 1), we observe that the TV-3PRF methods are the best performing providing benefits in the 10%-40% range. As we increase the factor strength (R^2 from 10% to 30%) our approach is able to forecast better, compared to the 3PRF, as expected. The three versions of TV-3PRF perform almost equivalently in this case. Selecting the bandwidths (H, L) through information criteria is appropriate in all cases, while this approach seems to benefit more from the strong factor model, than the cross validation methods for selecting bandwidths. The PCA methods perform worse than 3PRF in terms of $MSFE$, except when the factors are highly persistent. The two penalized regression versions seem to perform equaly well, and provide some slight advantages over the standard PCA methods. The above results are not systematically affected by the the choices of ρ_f, ρ_g, β, α that control the autocorrelation and cross sectional dependence of the generated samples.

Number of factors Selected by IC, DGP given by (42)									
ρ_f	0	0.3	0.3	0.3	0.3	0.9	0.9	0.9	0.9
ρ_g	0	0.9	0.9	0.9	0.9	0.3	0.3	0.3	0.3
β	0	0	1	0	1	0	1	0	1
a	0	0.3	0.3	0.9	0.9	0.3	0.3	0.9	0.9
$R^2 = 0.1$									
N = 100, T = 100	1.21	1.08	1.06	1.08	1.1	1.03	1.04	1.06	1.07
N = 200, T = 200	1.04	1.01	1.01	1.03	1.05	1.04	1.06	1.01	1.02
N = 200, T = 100	2.67	1.43	1.12	1.1	1.1	1.21	1.05	1.04	1.04
N = 100, T = 200	1.02	1.01	1.02	1.04	1.06	1.03	1.01	1.02	1.05
$R^2 = 0.3$									
N = 100, T = 100	1.16	1.11	1.08	1.11	1.12	1.06	1.05	1.02	1.03
N = 200, T = 200	1.04	1.03	1.02	1.04	1.05	1.01	1.01	1.01	1.01
N = 200, T = 100	1.52	1.14	1.09	1.08	1.08	1.13	1.05	1.01	1.02
N = 100, T = 200	1.04	1.03	1.04	1.06	1.07	1.03	1.01	1.01	1.02

Table 4: Average number of factors selected for the stochastically evolving over time loadings processes (42). The data are generated using one relevant factor.

Second, for the random walk loadings (see Table 2), the results are similar to those discussed above for the bounded random walk loadings, a remarkable result as this case is not covered by the theory but is often used in empirical analyses.

Third, focusing on the fixed parameter loadings case (see Table 3), we observe that the 3PRF method is now generally the best performing. Interestingly, the TV-3PRF, with bandwidths (H, L) selected using the information criterion, performs equally well in terms of $MSFE$. This indicates the robustness of this version of TV-3PRF to capture general loading dynamics.

Finally, Table 4 highlights the good performance of the BIC-type criterion in equation (25) for selecting the number of factors. In fact, on average, the selected number of factors is very close to 1 across all parameterizations and sample sizes, even more so when the explanatory power of the factors is 0.3 rather than 0.1. Results in the Appendix show that this finding remains broadly unchanged even when the loadings follow random walk processes.

In summary, this extensive analysis based on simulated data supports the use of the TV-3PRF as a proper method to handle the case of generic time variation in the factor loadings. Its performance is systematically better than constant parameters 3PRF and PCA, and generally also of more elaborate versions of PCA, which makes it a particularly interesting tool for empirical analysis of macroeconomic data.

5 Empirical Application

As an empirical illustration of our proposal we use the FRED-QD dataset (McCracken and Ng (2016)) to forecast a set of key U.S. macroeconomic variables [3]. Our dataset starts at Q1-1960 and ends at Q3-2022, for a total of 251 quarterly observations. The factors are extracted from a large set of 206 macroeconomic and financial time series included in FRED-QD dataset.

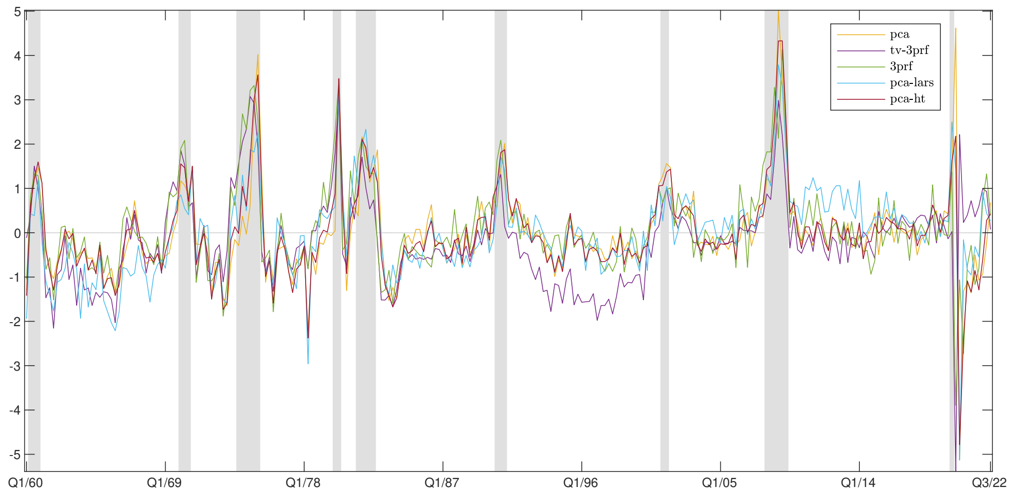

Figure 2: Factor estimates across the different methods considered. For the TV-3PRF and 3PRF method, the GDP growth is used as the target and proxy variable.

First, we report results from an insample illustrative exercise. Excluding the GDP growth rate series, from the large dataset x_{it} we estimate one factor ($\widehat{f}_t$) and its associated loadings ($\widehat{\Phi}_{it}$) using all the methods considered in the Monte Carlo section. In the 3PRF and TV-3PRF methods the GDP is used as a proxy variable (autoproxy). The estimated factor is presented in Figure 2. This figure shows that the factors extracted from all the methods considered are relatively similar and follow closely the US business cycle, as captured by the NBER recession periods (highlighted in grey in Figure 2). To better understand which approach provides the best insample fit we compute their insample, mean squared errors (MSE). To this end, we regress the target variable, GDP growth, on the estimated factor and compute the MSE using the residuals of this regression ($\widehat{\eta}_t$ in equation (1)). The the PCA factor delivers the highest insample MSE (0.80), followed by the hard thresholding PCA factor (0.74), the LARS PCA factor (0.71), the 3PRF factor (0.51) and TV-3PRF factor (0.36). This exercise suggests that our approach provides important information about GDP growth, that is not necessarily reflected in the competing approaches.

[3] Data details and series transformations are available online at FRED-QD description.

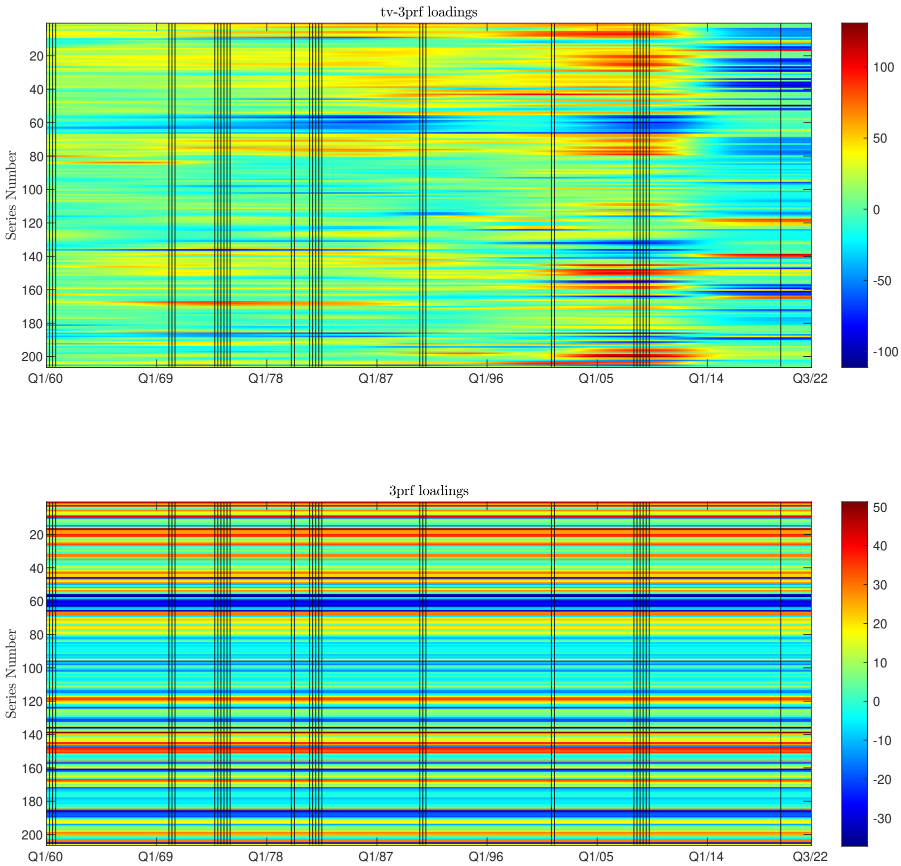

Figure 3: Factor loadings across time for the TV-3PRF and 3PRF methods. The vertical stripes indicate periods of NBER recession.

As we have discussed, an attractive feature of our TV-3PRF method is its ability to capture general dynamics of the loadings processes. Figure 3 shows the heat map of the estimated loadings across the time span (horizontal axis) and the large set of series (vertical axis). For comparison, we include the loadings of the 3PRF method, that remain constant across time. Some relevant considerations can be drawn from this figure. First, there is substantial time variation in all the loading processes, providing evidence in favor of our TV-3PRF method. Second, as in factor estimates (see Figure 2), the timing of loading shifts seems to coincide with the business cycle phases, as captured by the NBER recession dating. Macroeconomic series associated with substantial time variation include the consumer loans, the number of employed persons, the total reserves, and the industrial production, which can be directly related to the state of the business cycle.[4] Overall, our results suggest that the assumption of constant factor loadings often employed when studying this dataset is likely to be too restrictive.

[4]In Appendix 7.5.3 we present the heatmap of series associated with high variability loadings

Moving to the out of sample forecasting exercise, we focus on macroeconomic aggregates that receive considerably attention in the literature and policy-making circles. The series are presented in Table 5, and include GDP growth, industrial production, unemployment rate, inflation indices, exchange rates, and several others. The evaluation period starts at Q1-1985, including more than 150 quarterly observations. Following the approach of Bai and Ng (2008) and Stock and Watson (2012), any variable that we eventually target is removed from the set of predictors. Before forecasting each target, we first transform the data by partialing the large set predictors and the target with respect to a constant and q lags of the target. The lag length q is selected by the standard Bayesian Information Criterion. This allows us to abstract from autoregressive model selection issues and focus on the factor extraction, which is the main goal of our paper. At the end of this procedure, we compute one and four step ahead direct forecasts using all methods considered in the Monte Carlo section.

The forecasting performance of the alternative models is evaluated relative to that of the standard 3PRF, using the relative mean squared forecast error (r-MSFE). For each model m and target series s, it is:

$$r\text{-}RMSFE_{(m,s)} = \frac{\sum_{t=t_0}^{T} \left(e_t^{(m,s)} \right)^2}{\sum_{t=t_0}^{T} \left(e_t^{(3prf,s)} \right)^2}, \tag{45}$$

where $e_t^{(m,s)} = y_t^{(s)} - \widehat{y}_t^{(m,s)}$ is the one or four step ahead forecast error of model m for series s, and $e_t^{(3prf,s)} = y_t^{(s)} - \widehat{y}_t^{(3prf,s)}$ is the counterpart for the benchmark 3PRF model. When the $r\text{-}MSFE_{(m,s)}$ is less than one, model m out performs the benchmark 3PRF for macroeconomic variable s. To assess the statistical significance of the MSFE differentials, we use the common Diebold and Mariano (1995) test (henceforth DM test).[5] To assess whether the relative performance of a model is stable over time, we implement the forecast fluctuation test developed by Giacomini and Rossi (2010) (henceforth GR test). The forecast fluctuation test measures the relative local forecasting performance for the two models. In contrast to the DM test, which measures the global performance over the forecasting horizon, the GR test assesses the stability of the relative performance over the entire time path. The test statistic is equivalent to the DM statistic computed over rolling windows of forecasts. Both DM and GR tests are computed relatively to the 3PRF method.

The tuning parameters of the TV-3PRF are calibrated by the two cross validation schemes ((34) and (36)). As an alternative, we use the information criterion (25) to select both the

[5]Since our out of sample period contains a large number of observations, the small sample correction of the DM test proposed by Harvey *et al.* (1997) is not necessary, as it would lead to very minor modifications of the original DM statistic.

Series Names: Macroeconomic Forecasting	
TBILL	3-Month Treasury Bill: Secondary Market Rate
IP-DCG	Industrial Production: Durable Consumer Goods
AE-WT	All Employees: Wholesale Trade (Thousands of Persons)
FEDFUNDS	Effective Federal Funds Rate
PCE-FSI	Personal consumption expenditures: Financial services and insurance
PFI	Real private fixed investment: Nonresidential
CPI-LFE	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy
IP-M	Industrial Production: Manufacturing
UNRATE	Unemployment Rate less than 27 weeks
AHEP	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing
CP	3-Month AA Financial Commercial Paper Rate
CLAIMS	Initial Claims
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started
CUMFNS	Capacity Utilization: Manufacturing
GCE	Real Government Consumption Expenditures and Gross Investment: Federal
PAYEMS	All Employees: Total nonfarm
EXPORTS	Real Exports of Goods and Services
PCE-FE	Personal Consumption Expenditures Excluding Food and Energy
GS10	10-Year Treasury Constant Maturity Rate
EXR-SUS	Switzerland / U.S. Foreign Exchange Rate
UNRATE	Civilian Unemployment Rate
GDP	Real Gross Domestic Product
M2REAL	Real M2 Money Stock
EXR-USUK	U.S. / U.K. Foreign Exchange Rate
FPI	Real private fixed investment

Table 5: Macroeconomic series used in the Out of sample forecasting Exercise. All series have been transformed to stationary according to the recommendations of the FRED-MD database.

bandwidths (H, L), and the number of factors (labeled $tv - 3prf(ic)$ in the Tables). For comparison, we also report results for the PCA method, when the number of factors is selected by information criteria (see Bai and Ng (2002)).

In Table 6 we present the outcome of the forecast evaluation. Overall, across all the predicted 25 variables and 2 horizons, the TV-3PRF models obtain the best forecasting results in 23 out of 50 cases, while they are among the best three performing models, in 37 cases. The 3PRF is best performing in 4 cases and the targeted predictors PCA methods are the best performing in 5 cases. In the remainder of the cases, the two versions of the standard PCA method perform best. More specifically, for 1-quarter ahead forecasting, the TV-3PRF methods are the best performing in 17 out of 25 cases. The worse performance at the longer forecast horizon is likely due to the use of the direct approach that, while convenient in the large dataset context as it avoids the need of predicting a large number of variables, weakens the relationship between target and explanatory variables. The most significant improvements in forecasting with TV-3PRF are for the TBill, the Industrial Production indexes (IP-DCG and IP-M), the Fed Funds rate, and the Commercial Paper. In these cases, the TV-3PRF forecasts are also significantly different from their 3PRF counterparts in terms of DM and/or GW tests. It is interesting to note that the two cross validation versions of TV-3PRF perform almost equivalently most of the times while, as in the Monte Carlo section, the BIC-type specification of the TV-3PRF provides significant advantages in some cases (e.g., for IP-DCG, AE-WT, IP-M).

In Table 7 we repeat the evaluation focusing on the NBER recession periods only, to examine whether this leads to any substantial changes in the model rankings. Overall, the modifications are limited and point to an even slightly better performance of the TV-3PRF during recessionary periods. In particular, now the TV-3PRF performs best in 27 cases, and the 3PRF in 3 cases. As in the full evaluation period and in the Monte Carlo simulations, the BIC-type based specification of the TV-3PRF performs particularly well.

Finally, results reported in the Appendix show that when the sample ends in 2019Q4, so that the Covid-19 period is excluded, the TV-3PRF methods are best for a lower number of cases, 16 out of 50. Hence, the allowed time variation performed particularly well during the Covid-19 period.

6 Conclusions

Parameter time variation is pervasive in econometric models, due to the many sources of changes in economic relationships. This is typically addressed by making specific assumptions on the type of parameter evolution, for example assuming autoregressive or random walk evolution of the parameters. Yet, since parameter time variation is unobservable, any specific parametric model is likely mis-specified. Hence, in this paper, we consider stochastic non-parametric time variation in the parameters, focusing on the case of the loadings in factor models for large datasets, which has attracted considerable interest in the recent empirical macroeconomic and financial literature.

We have introduced kernel-based estimators in the context of the time-varying three pass regression filter, an efficient estimation and forecasting algorithm that permits to target the estimated factors to a specific variable of interest. We have established consistency of the resulting forecast for the unfeasible forecast based on true factors and parameters, together with the the limiting distributions of the estimated factors and factor loadings under the standard large N and large T framework. We have also linked our proposed method with the time varying parameter constrained least squares estimator and with the time varying partial least squares method, and shown that these are special cases of our approach. Finally, we have proposed a BIC-type information criterion to determine the number of common factors in this context, and specific cross validation methods to select the bandwidths of the kernel based estimators.

We have used an extensive set of simulation exercises to asses the finite sample performance of our approach, in comparison to the standard 3PRF and PCA, and also to a variety of more sophisticated PCA based methods. Overall, the TV-3PRF performs quite well, even for rather small samples (N=100, T=100).

Out of Sample Macroeconomic Forecasting										
h	1	4	1	4	1	4	1	4	1	4
	TBILL		IP-DCG		AE-WT		FEDFUNDS		PCE-FSI	
pca(ic)	1.55	1.06	1.21	1.21	0.99	1	1.45	1.09	0.92	1.1
pca(1)	0.97	0.66	0.94	0.96	0.99	1.04	1.07	0.71	0.97	0.78
pca-lars(1)	1.2	0.81	0.92	1.1	0.94	1.06	1.55	0.78	0.96	0.96
pca-ht(10%)(1)	0.95	0.77	1.12	0.97	0.99	1.02	1.02	0.82	0.92	0.84
3prf(1)	1	1	1	1	1	1	1	1	1	1
tv-3prf(1)	$0.96_{\diamond\diamond}$	0.75	0.72	1.11	0.81	$1.09_{\diamond}$	0.83	0.83	0.88	1.06
tv-3prf(ic)	$0.94_{\diamond\diamond}$	$1.17_{\diamond\diamond}$	$0.91_{\diamond\diamond}$	0.93	0.86	$0.97_{\diamond\diamond}$	$0.85^{*}_{\diamond\diamond}$	$1.02_{\diamond\diamond}$	1.01	$0.92_{\diamond\diamond}$
tv-3prf$_{H_i}$(1)	$0.81_{\diamond\diamond}$	0.8	0.7	$1.15_{\diamond}$	0.81	1.11	0.79	0.87	0.89	1.07
	PFI		CPI-LFE		IP-M		UNRATE		AHEP	
pca(ic)	1.15	1.07	1.01	0.82	1	1.01	0.99	1.13	0.95	0.91
pca(1)	0.97	1.01	0.9	0.92	1.11	1.03	1.17	1.02	0.81	0.88
pca-lars(1)	1.04	1.19	1.06	1	1.02	1.14	1.07	1.16	0.83	0.91
pca-ht(10%)(1)	0.99	1	0.9	0.95	0.99	1.03	1.37	1.08	0.9	0.94
3prf(1)	1	1	1	1	1	1	1	1	1	1
tv-3prf(1)	0.89^{*}	1.22	$0.9_{\diamond\diamond}$	0.91	0.91	1.23	0.92	1.18	0.92	0.9
tv-3prf(ic)	$0.92^{**}_{\diamond}$	1.01	$0.83^{*}_{\diamond\diamond}$	$0.98_{\diamond\diamond}$	$0.86_{\diamond\diamond}$	$0.97_{\diamond\diamond}$	1.04	1	$0.91^{**}_{\diamond\diamond}$	$0.91_{\diamond\diamond}$
tv-3prf$_{H_i}$(1)	0.88^{*}	1.14	$0.93_{\diamond\diamond}$	$0.93_{\diamond}$	0.92	$1.24_{\diamond}$	0.92	1.21	0.98	0.9
	CP		CLAIMS		HOUST		CUMFNS		GCE	
pca(ic)	1.53	1.06	0.95	1.07	1.02	1.18	1.11	1.02	0.92	0.98
pca(1)	0.99	0.79	1.09	1.02	1.01	0.88	1.18	1.16	0.87	0.98
pca-lars(1)	1.34	0.96	1.08	1.06	1.1	0.82	0.96	1.08	0.88	0.98
pca-ht(10%)(1)	0.99	0.85	1.12	1.03	1.01	0.91	0.85	0.98	0.89	0.99
3prf(1)	1	1	1	1	1	1	1	1	1	1
tv-3prf(1)	$0.92_{\diamond\diamond}$	0.95	0.92	0.98	0.93	1.08	0.93	$0.97_{\diamond}$	0.94	1.16
tv-3prf(ic)	$0.9_{\diamond\diamond}$	$1.04_{\diamond\diamond}$	0.99	0.99	0.97	1	$0.83_{\diamond}$	$0.99_{\diamond\diamond}$	0.87^{**}	1.01
tv-3prf$_{H_i}$(1)	$0.87_{\diamond\diamond}$	0.92	0.95	0.99	0.96	1.06	0.93	$0.99_{\diamond\diamond}$	0.97	1.15
	PAYEMS		EXPORTS		PCE-FE		GS10		EXR-SUS	
pca(ic)	0.99	1	1.03	1.04	0.97	1.03	1.04	0.89	0.94	1.02
pca(1)	1.21	1.05	1	0.99	0.95	1	0.86	0.84	0.85	0.9
pca-lars(1)	1.18	1.19	1.01	1.01	1.13	1.04	0.88	0.82	0.87	0.89
pca-ht(10%)(1)	1.29	1	1.01	1.01	0.95	1	0.89	0.93	0.91	0.92
3prf(1)	1	1	1	1	1	1	1	1	1	1
tv-3prf(1)	0.97	$1.12_{\diamond\diamond}$	0.97	1.13	0.98	1.09	1	0.91	1.14	1.19
tv-3prf(ic)	$1.06_{\diamond}$	$0.98_{\diamond\diamond}$	0.98	1	0.93	1	0.98	$0.98_{\diamond}$	1.09	1.12
tv-3prf$_{H_i}$(1)	1.03	$1.18_{\diamond\diamond}$	1	$1.13_{\diamond\diamond}$	0.97	1.1	0.95	0.95	1.13	1.2
	UNRATE		GDP		M2REAL		EXR-USUK		FPI	
pca(ic)	1.1	1.09	1.21	1.05	0.93	1.05	0.93	1.05	1.12	1.14
pca(1)	1.3	1.12	1.19	0.98	1	0.93	0.86	0.95	1.14	0.98
pca-lars(1)	1.19	0.99	1.01	1.08	0.99	1	0.91	1.01	0.85	1.04
pca-ht(10%)(1)	1.07	1.09	1.13	1.07	1	0.94	0.93	0.97	1.12	1.02
3prf(1)	1	1	1	1	1	1	1	1	1	1
tv-3prf(1)	1.3	1.06	1.05	$1.17_{\diamond}$	1.1	$1.07_{\diamond\diamond}$	$1.1_{\diamond\diamond}$	1.08	0.97	1.29
tv-3prf(ic)	1.02	0.99	0.98	0.99	1.12	$0.97_{\diamond\diamond}$	1.03	1.02	0.96^{**}	0.99
tv-3prf$_{H_i}$(1)	1.36	1.13	1.01	1.19	1.12	$1.1_{\diamond\diamond}$	1.16	1.09	0.97	1.24

Table 6: MSFE results for h=1,4 periods ahead forecasts. The results are relative to the MSFE of the 3PRF, one autoproxy factor model ($3prf(1)$). In red you can see the best performing method, in blue the second best and in pink the third best. The * (**) denotes statistically different forecasts from the $3prf(1)$ model at the 10% (5%) significance level, according to the Diebold and Mariano test. The $\diamond$ ($\diamond\diamond$) denotes statistically different forecasts from the $3prf(1)$ model at the 10% (5%) significance level according to the forecast fluctuation test. The forecasting period starts at Q4-1984 and ends at Q1-2022, meaning that we forecast 151 quarterly observations. The estimation period starts at Q1-1960. Table 5 presents details about the reference names of the series.

Out of Sample Macroeconomic Forecasting (NBER crises periods)										
h	1	4	1	4	1	4	1	4	1	4
	TBILL		IP-DCG		AE-WT		FEDFUNDS		PCE-FSI	
$pca(ic)$	2.52	1.27	0.65	0.98	1.01	0.85	2.21	1.4	0.67	1.16
$pca(1)$	1.04	1.22	0.99	1.08	1.16	1.35	1.09	1.21	0.72	0.73
$pca\text{-}lars(1)$	1.34	1.43	0.57	1.13	0.64	1.08	1.89	1.15	0.63	1.14
$pca\text{-}ht(10\%)(1)$	0.97	1.18	1.2	1.08	1.04	1.26	1	1.06	0.55	0.77
$3prf(1)$	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	0.48	1.29	0.43	0.98	0.62	0.95	0.52	1.3	0.64	0.8
$tv\text{-}3prf(ic)$	0.77	0.99	0.88	0.93	0.87	0.82	0.84	0.95	1.02	0.92
$tv\text{-}3prf_{H_i}(1)$	0.46	1.38	0.39	0.93	0.63	0.98	0.56	1.09	0.63	0.77
	PFI		CPI-LFE		IP-M		UNRATE		AHEP	
$pca(ic)$	0.94	1.1	1.03	0.96	0.69	1.01	0.78	1.02	0.93	1.1
$pca(1)$	0.82	1.18	0.7	0.81	1.48	1.16	1.5	1.34	0.84	0.96
$pca\text{-}lars(1)$	0.86	1.11	0.8	1.01	0.79	1.07	0.59	1.16	0.89	0.94
$pca\text{-}ht(10\%)(1)$	0.91	1.02	0.72	0.83	1.07	1.12	1.75	1.09	1.04	0.94
$3prf(1)$	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	1.06	1.19	0.39	0.83	0.73	1.12	0.77	1.19	0.92	0.63
$tv\text{-}3prf(ic)$	0.92	0.97	0.41	0.95	0.96	1.01	0.98	1	0.9	0.9
$tv\text{-}3prf_{H_i}(1)$	1.01	1.14	0.38	0.81	0.72	1.12	0.7	1.19	1.22	0.68
	CP		CLAIMS		HOUST		CUMFNS		GCE	
$pca(ic)$	2.29	1.49	0.74	0.98	0.86	0.99	0.75	0.95	0.41	0.84
$pca(1)$	0.86	1.48	1.18	1.06	0.79	0.9	2.03	1.09	0.45	0.91
$pca\text{-}lars(1)$	1.07	1.04	0.93	1.02	0.94	0.86	0.68	1.16	0.44	0.73
$pca\text{-}ht(10\%)(1)$	0.82	1.17	1.34	1.07	0.81	0.92	0.77	1.04	0.62	0.95
$3prf(1)$	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	0.74	1.57	0.53	1.08	0.18	0.86	0.86	1.06	0.67	0.63
$tv\text{-}3prf(ic)$	0.68	0.96	0.88	1.01	0.87	0.96	1.07	1.02	0.29	0.49
$tv\text{-}3prf_{H_i}(1)$	0.78	1.11	0.54	1.02	0.24	0.82	0.86	1.07	0.31	0.62
	PAYEMS		EXPORTS		PCE-FE		GS10		EXR-SUS	
$pca(ic)$	0.8	0.91	0.95	1.16	0.88	0.98	1.24	0.98	0.89	0.95
$pca(1)$	1.22	1.33	1.08	1.11	0.8	1	0.38	0.9	0.68	0.94
$pca\text{-}lars(1)$	0.72	1.09	1.09	1.06	1.2	1.05	0.38	0.87	0.69	0.92
$pca\text{-}ht(10\%)(1)$	0.82	1.17	1.11	1.09	0.81	1.01	0.47	1	0.91	0.91
$3prf(1)$	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	0.93	0.98	1.27	1.24	1.14	1.05	0.81	0.77	1.44	0.76
$tv\text{-}3prf(ic)$	0.78	0.92	1.07	1.02	0.97	1	0.67	0.86	1.41	0.79
$tv\text{-}3prf_{H_i}(1)$	0.9	1.04	1.25	1.26	1.19	1.05	0.82	0.78	1.32	1.22
	UNRATE		GDP		M2REAL		EXR-USUK		FPI	
$pca(ic)$	0.4	1.06	0.83	1.07	1.1	1.04	1.07	1.04	0.63	1.15
$pca(1)$	2.01	1.33	1.08	1.31	0.98	1.03	0.92	1.02	1.27	1.15
$pca\text{-}lars(1)$	0.89	0.97	0.82	1.35	1.09	1.15	1.03	1	0.63	1.13
$pca\text{-}ht(10\%)(1)$	0.96	1.05	0.66	1.25	0.98	1.03	1.07	1.01	1.34	1.14
$3prf(1)$	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	0.86	1.2	1.16	1.08	1.88	0.86	1.04	0.93	0.53	0.95
$tv\text{-}3prf(ic)$	0.94	0.99	0.61	0.98	1.32	1.04	0.89	1.07	0.89	0.97
$tv\text{-}3prf_{H_i}(1)$	0.9	1.23	0.72	1.06	1.95	0.89	1.11	1.01	0.51	0.92

Table 7: MSFE results for the NBER crises periods for h=1, 4 periods ahead forecasts.

Finally, to illustrate the empirical applicability of our method, we have conducted an out of sample forecasting exercise, using a large panel of US macroeconomic series to predict key variables of interest. This exercise confirmed the gains from the use of TV-3PRF, making it not only a relevant theoretical contribution but also a useful additional tool for applied econometric and statistical analyses.

References

AMEMIYA, T. (1985). *Advanced Econometrics*. Harvard University Press.

ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61** (4), 821–856.

BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, **71** (1), 135–171.

— and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70** (1), 191–221.

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

— and — (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, **146** (2), 304–317.

BANERJEE, A., MARCELLINO, M. and MASTEN, I. (2008). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. In *Forecasting in the presence of structural breaks and model uncertainty*, vol. 3, Emerald Group Publishing Limited, pp. 149–194.

BATES, B. J., PLAGBORG-MØLLER, M., STOCK, J. H. and WATSON, M. W. (2013). Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*, **177** (2), 289–304.

BOIVIN, J. and NG, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, **132** (1), 169–194.

BREITUNG, J. and EICKMEIER, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, **163** (1), 71–84.

CHEN, B. and HONG, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, **80** (3), 1157–1183.

CHEN, L., DOLADO, J. J. and GONZALO, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, **180** (1), 30–48.

CHENG, X., LIAO, Z. and SCHORFHEIDE, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, **83** (4), 1511–1543.

CORRADI, V. and SWANSON, N. R. (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics*, **182** (1), 100–118.

DENDRAMIS, Y., GIRAITIS, L. and KAPETANIOS, G. (2021). Estimation of time-varying covariance matrices for large datasets. *Econometric Theory*, **37** (6), 1100–1134.

DIEBOLD, F. X. and MARIANO, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.

EFRON, B. (2004). The estimation of prediction error: covariance penalties and cross-validation. *Journal of the American Statistical Association*, **99** (467), 619–632.

—, HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *The Annals of Statistics*, **32** (2), 407–451.

EICKMEIER, S., LEMKE, W. and MARCELLINO, M. (2015). Classical time varying factor-augmented vector auto-regressive models—estimation, forecasting and structural analysis. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 493–533.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, **82** (4), 540–554.

—, —, — and — (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American statistical association*, **100** (471), 830–840.

GIACOMINI, R. and ROSSI, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, **25**, 595–620.

GIRAITIS, L., KAPETANIOS, G. and MARCELLINO, M. (2021). Time-varying instrumental variable estimation. *Journal of Econometrics*, **224** (2), 394–415.

—, — and YATES, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, **179** (1), 46–65.

—, — and — (2018). Inference on multivariate heteroscedastic time varying random coefficient models. *Journal of Time Series Analysis*, **39** (2), 129–149.

GROEN, J. J. and KAPETANIOS, G. (2016). Revisiting useful approaches to data-rich macroeconomic forecasting. *Computational Statistics Data Analysis*, **100**, 221–239.

GUÉRIN, P., LEIVA-LEON, D. and MARCELLINO, M. (2020). Markov-switching three-pass regression filter. *Journal of Business & Economic Statistics*, **38** (2), 285–302.

HANSEN, B. E. (2001). The new econometrics of structural change: Dating breaks in us labor productivity. *Journal of Economic perspectives*, **15** (4), 117–128.

HARVEY, D., LEYBOURNE, S. and NEWBOLD, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, **13**, 281—291.

HEPENSTRICK, C., MARCELLINO, M. *et al.* (2019). Forecasting gross domestic product growth with large unbalanced data sets: the mixed frequency three-pass regression filter. *Journal of the Royal Statistical Society Series A*, **182** (1), 69–99.

KELLY, B. and PRUITT, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, **186** (2), 294–316.

KRÄMER, N. and SUGIYAMA, M. (2011). The degrees of freedom of partial least squares regression. *Journal of the American Statistical Association*, **106** (494), 697–705.

MA, S. and SU, L. (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of econometrics*, **207** (1), 1–29.

MASSACCI, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics*, **197** (1), 101–129.

MCCRACKEN, M. W. and NG, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, **34** (4), 574–589.

ROBINSON, P. M. (1989). *Nonparametric Estimation of Time-Varying Parameters*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 253–264.

— (1991). Time-varying nonlinear regression. In P. Hackl and A. H. Westlund (eds.), *Economic Structural Change*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 179–190.

STOCK, J. H. and WATSON, M. (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, **173**, 205.

— and WATSON, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, **97** (460), 1167–1179.

— and — (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, **20** (2), 147–162.

— and — (2011). Dynamic factor models.

— and — (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, **30** (4), 481–493.

— and — (2015). Factor models for macroeconomics.

SU, L. and WANG, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, **198** (1), 84–101.

WHITE, H. (2014). *Asymptotic theory for econometricians*. Academic press.

WOLD, H. (1975). Soft modelling by latent variables: the non-linear iterative partial least squares (nipals) approach. *Journal of Applied Probability*, **12** (S1), 117–142.

7 Appendix

In this appendix we include the proofs of the Theorems presented in the main text of the paper, together with a set of additional results needed for the proofs. Furthermore, we include additional Monte Carlo evidence on the usefullness of our TV-3PRF.

7.1 Kernel weights

Our estimation strategy for the time varying coefficients is based on the use of kernels $K(x) \geq 0$ with weights

$$k_{H,ts} = K\left(\frac{t-s}{H}\right), \text{ and } K_{H,t} = \sum_{s=1}^{T} k_{H,ts}, \tag{46}$$

where H is the bandwidth such that $H \to \infty$, with rate $H = o(T)$. In what follows, whenever possible, we drop the subscript H in $k_{H,t,s}$ and $K_{H,t}$, (i.e. write $k_{t,s}$, and K_t respectively) for notational convenience.

We consider kernels that satisfy the following conditions

$$k_{ts} \leq C\left(1 + \left(\frac{|t-s|}{H}\right)^{\nu}\right)^{-1} \tag{47}$$
$$|k_{t,s} - k_{t,s+1}| \leq CH^{-1}\left(1 + \left(\frac{|t-s|}{H}\right)^{\nu}\right)^{-1}$$

for $\nu > 3$.

7.2 Notation

To facilitate exposition, we define the following kernel weighted versions of data, factors, and idiosyncratic components. That is, let $\widetilde{y}^{(t)} = \left[k_{t,1}^{1/2}\widetilde{y}_{t,2}, .., k_{t,T}^{1/2}\widetilde{y}_{t,T+1}\right]'$, $\widetilde{y}_{t,s} = y_s - \overline{y}_t$, $\overline{y}_t = K_t^{-1}\sum_{l=1}^{T} k_{t,l}y_{l+1}$, and $\widetilde{\eta}^{(t)} = \left[k_{t,1}^{1/2}\widetilde{\eta}_{t,2}, .., k_{t,T}^{1/2}\widetilde{\eta}_{t,T+1}\right]'$, $\widetilde{\eta}_{t,s} = \eta_s - \overline{\eta}_t$, with $\overline{\eta}_t = K_t^{-1}\sum_{s=1}^{T} k_{t,l}\eta_{l+1}$. Also, $\widetilde{Z}^{(s)} = \left[k_{s,1}^{1/2}\widetilde{z}_{s,1}, .., k_{s,T}^{1/2}\widetilde{z}_{s,T}\right]'$, for $\widetilde{z}_{s,t} = z_t - \overline{z}_s$, with $\overline{z}_t = K_t^{-1}\sum_{s=1}^{T} k_{t,l}z_l$ and $z_t^{(s)} = k_{s,t}^{1/2}z_t$. We can define analogous weighted matrices for F_t, x_t, ω_t, ε_t. In practice, all these matrices depend on a bandwidth parameter H (or L), through the kernel weigting (46). Then, the matrix $\widetilde{Z}^{(s)}$ is a function of the bandwidth H, that is $\widetilde{Z}_H^{(s)}$. When this does not introduce any confusion, we drop the bandwidth subscript.

When possible, we will use the summations $S_{zx,t} = K_t^{-1} \sum_{j=1}^{T} k_{tj} \left(z_j - \bar{z}_t\right) \left(x_j - \bar{x}_t\right)'$, and $S_{xy,t} = K_t^{-1} \sum_{j=1}^{T} k_{tj} \left(x_j - \bar{x}_t\right) \left(y_{j+1} - \bar{y}_t\right)'$, which also depend on bandwidths H or L (i.e. $S_{zx,H,t}$, $S_{xy,L,t}$).

7.3 Main Results

Here we present the main theoretical results of the paper. The proofs are derived using the auxiliary results given afterward in section 7.4. As before, the notation used here follows the descriptions in section 7.2.

7.3.1 Proof of Theorem 1

Proof of (i): By Lemma A4 and Lemma A5 we have

$$\widehat{F}_t' \widehat{\beta}_t \underset{N,T\to\infty}{\overset{p}{\to}} \left(\Lambda_t \Delta_F P_1 + \Lambda_t \Delta_F \Delta_\Phi F_t\right)' \left(\Lambda_t \Delta_F \Delta_\Phi \Delta_F \Lambda_t'\right)^{-1} \left(\Lambda_t \Delta_F \Lambda_t' + \Delta_\omega\right) \tag{48}$$

$$\times \left(\Lambda_t \Delta_F \Lambda_t' + \Delta_\omega\right)^{-1} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \Lambda_t' \left(\Lambda_t \Delta_F \Delta_\Phi \Delta_F \Delta_\Phi \Delta_F \Lambda_t'\right)^{-1} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \beta_t \tag{49}$$

$$= \left(\Lambda_t \Delta_F P_1 + \Lambda_t \Delta_F \Delta_\Phi F_t\right)' \left(\Lambda_t \Delta_F \Delta_\Phi \Delta_F \Delta_\Phi \Delta_F \Lambda_t'\right)^{-1} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \beta_t. \tag{50}$$

By Assumptions 3, 4, we have $P_1 = 0$, Δ_F is diagonal and positive definite, $\Lambda_t = [\Lambda_{f_t}, 0]$ with Λ_{f_t} non singular, and $\Delta_\Phi = I$. Then

$$\widehat{F}_t' \widehat{\beta}_t \underset{N,T\to\infty}{\overset{p}{\to}} F_t' \Delta_\Phi \Delta_F \Lambda_t' \left(\Lambda_t \Delta_F \Delta_\Phi \Delta_F \Delta_\Phi \Delta_F \Lambda_t'\right)^{-1} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \beta_t \tag{51}$$

$$= F_t' \Delta_\Phi \Delta_F \Lambda_t' \left(\Delta_\Phi \Delta_F \Lambda_t'\right)^{-1} \left(\Lambda_t \Delta_F \Delta_\Phi \Delta_F\right)^{-1} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \beta_t$$

$$= F_t' \beta_t.$$

Noticing that

$$\widehat{y}_{t+1} = \widehat{\beta}_{0t} + \widehat{F}_t' \widehat{\beta}_t \tag{52}$$

and

$$\widehat{\beta}_{0t} = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} y_{s+1} - K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \widehat{F}_s' \widehat{\beta}_{L,t}, \tag{53}$$

it follows that

$$
\begin{aligned}
\widehat{y}_{t+1} &= K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}y_{s+1} - K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\widehat{F}_s'\widehat{\beta}_t + \widehat{F}_t\widehat{\beta}_t \\
&= K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\beta_{0s} + K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}F_s'\beta_s + K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\eta_{s+1} - K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\widehat{F}_s'\widehat{\beta}_t + \widehat{F}_t'\widehat{\beta}_t \\
&= K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\beta_{0s}-\beta_{0t}\right) + \beta_{0t} + K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\eta_{s+1} + K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}F_s'\beta_s \\
&+ K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\widehat{F}_s'\left(\widehat{\beta}_s-\widehat{\beta}_t\right) - K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\widehat{F}_s'\widehat{\beta}_s + \widehat{F}_t'\widehat{\beta}_t \\
&\underset{N,T\to\infty}{\overset{p}{\to}} \beta_{0t} + F_t'\beta_t
\end{aligned}
\tag{54}
$$

since $\widehat{F}_t'\widehat{\beta}_t \underset{N,T\to\infty}{\overset{p}{\to}} F_t'\beta_t, \forall t$, $K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\eta_{s+1} = o\left(1\right)$ by Lemma A1, the $K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\beta_{0s}-\beta_{0t}\right) = O_p\left(\sqrt{\frac{L}{T}}\right)$ by Assumption 1 and Corollary 9b in Dendramis *et al.* (2021). For the term $K_t^{-1}\sum_{s=1}^{T}k_{t,s}\widehat{F}_s'\left(\widehat{\beta}_s-\widehat{\beta}_t\right)$, notice that by Lemma A5 it is

$$
\widehat{\beta}_t \to \left(\Lambda_t\Delta_F\Lambda_t'+\Delta_\omega\right)^{-1}\Lambda_t\Delta_F\beta_t = G_t^{-1}\Lambda_t\Delta_F\beta_t,
$$

with $G_t = \left(\Lambda_t\Delta_F\Lambda_t'+\Delta_\omega\right)$. Moreover, it is

$$
\begin{aligned}
\widehat{\beta}_s-\widehat{\beta}_t &= G_s^{-1}\Lambda_s\Delta_F\beta_s - G_t^{-1}\Lambda_t\Delta_F\beta_t = G_s^{-1}\Lambda_s\Delta_F\beta_s - G_s^{-1}\Lambda_s\Delta_F\beta_t + G_s^{-1}\Lambda_s\Delta_F\beta_t - G_t^{-1}\Lambda_t\Delta_F\beta_t \\
&= G_s^{-1}\Lambda_s\Delta_F\left(\beta_s-\beta_t\right) + \left(G_s^{-1}\Lambda_s\Delta_F - G_t^{-1}\Lambda_t\Delta_F\right)\beta_t \\
&= G_s^{-1}\Lambda_s\Delta_F\left(\beta_s-\beta_t\right) + \left(G_s^{-1}\Lambda_s\Delta_F - G_s^{-1}\Lambda_t\Delta_F + G_s^{-1}\Lambda_t\Delta_F - G_t^{-1}\Lambda_t\Delta_F\right)\beta_t \\
&= G_s^{-1}\Lambda_s\Delta_F\left(\beta_s-\beta_t\right) + G_s^{-1}\left(\Lambda_s-\Lambda_t\right)\Delta_F\beta_t + \left(G_s^{-1}-G_t^{-1}\right)\Lambda_t\Delta_F\beta_t.
\end{aligned}
$$

Furthermore, since

$$
\left\|G_s^{-1}-G_t^{-1}\right\| = \left\|G_t^{-1}\left(G_t-G_s\right)G_s^{-1}\right\| \leq \left\|G_t^{-1}\right\|\left\|G_t-G_s\right\|\left\|G_s^{-1}\right\| \tag{55}
$$

and

$$
\begin{aligned}
\left\|G_t-G_s\right\| &\leq \left\|\Lambda_t\Delta_F\Lambda_t' - \Lambda_s\Delta_F\Lambda_s'\right\| = \left\|\Lambda_t\Delta_F\Lambda_t' - \Lambda_t\Delta_F\Lambda_s' + \Lambda_t\Delta_F\Lambda_s' - \Lambda_s\Delta_F\Lambda_s'\right\| \\
&\leq \left\|\Lambda_t\right\|\left\|\Delta_F\right\|\left\|\Lambda_t'-\Lambda_s'\right\| + \left\|\Lambda_t-\Lambda_s\right\|\left\|\Delta_F\right\|\left\|\Lambda_s'\right\|,
\end{aligned}
\tag{56}
$$

we have that

$$\left\| K_t^{-1} \sum_{s=1}^{T} k_{t,s} \widehat{F}_s' \left(\widehat{\beta}_s - \widehat{\beta}_t \right) \right\| \leq$$

$$\leq \max_s E \left\| \widehat{F}_s \right\| \left\| K_t^{-1} \sum_{s=1}^{T} k_{t,s} \left(G_s^{-1} \Lambda_s \Delta_F \left(\beta_s - \beta_t \right) + G_s^{-1} \left(\Lambda_s - \Lambda_t \right) \Delta_F \beta_t + \left(G_s^{-1} - G_t^{-1} \right) \Lambda_t \Delta_F \beta_t \right) \right\|$$

$$\leq C_1 \max_s E \left\| G_s^{-1} \right\| \max_s \left\| \Lambda_s \right\| \left\| \Delta_F \right\| K_t^{-1} \sum_{s=1}^{T} k_{t,s} \left\| \left(\beta_s - \beta_t \right) \right\|$$

$$+ C_1 \max_s E \left\| G_s^{-1} \right\| \max_t E \left\| \beta_t \right\| \left\| \Delta_F \right\| K_t^{-1} \sum_{s=1}^{T} k_{t,s} \left\| \left(\Lambda_s - \Lambda_t \right) \right\|$$

$$+ C_1 \max_t E \left\| \Lambda_t \right\| \max_t \left\| \Delta_F \right\| \max_t E \left\| \beta_t \right\| \max_t \left\| G_t^{-1} \right\|^2 K_t^{-1} \sum_{s=1}^{T} k_{t,s} \left\| G_t - G_s \right\|$$

$$= O_p \left(\sqrt{\frac{L}{T}} \right)$$

since $\max_s E \left\| \widehat{F}_s \right\| < \infty$, as all elements in (102) have bounded norm, $\max_s E \left\| G_s^{-1} \right\| < \infty$, for the same reason, $\max_s \left\| \Lambda_s \right\| < \infty$, $\left\| \Delta_F \right\| < \infty$, $\max_t E \left\| \beta_t \right\| < \infty$, by Assumption 1, $K_t^{-1} \sum_{s=1}^{T} k_{t,s} \left\| \beta_s - \beta_t \right\| = O_p \left(\sqrt{\frac{H}{T}} \right)$, $K_t^{-1} \sum_{s=1}^{T} k_{t,s} \left\| \left(\Lambda_s - \Lambda_t \right) \right\| = O_p \left(\sqrt{\frac{H}{T}} \right)$, and $K_t^{-1} \sum_{s=1}^{T} k_{t,s} \left\| G_t - G_s \right\| = O_p \left(\sqrt{\frac{H}{T}} \right)$ by (56), Assumption 1 and Corollary 9b in Dendramis *et al.* (2021), and this proves the result.

Proof of (ii): Notice that from equation (128) we have that

$$\widehat{\beta}_t = S_{zz,H,t}^{-1} S_{zx,H,t} J_N S_{zx,H,t}' \left(S_{zx,H,t} J_N S_{xx,L,t} J_N S_{zx,H,t}' \right)^{-1} S_{zx,H,t} J_N S_{xy,L,t} + o_p \left(1 \right). \quad (57)$$

As in (104), we set

$$\widehat{F}_t = \widehat{F}_{(1),t} x_t \quad (58)$$

$$= S_{zz,H,t} \left(S_{zx,H,t} J_N S_{zx,H,t}' \right)^{-1} S_{zx,H,t} J_N x_t. \quad (59)$$

Then,

$$\widehat{F}_t - \overline{\widehat{F}}_t = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \left(x_t - x_s \right) \right)' + K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\left(\widehat{F}_{(1),t} - \widehat{F}_{(1),s} \right) x_s \right)'$$

$$= \left(x_t - \overline{x}_t \right)' \widehat{F}_{(1),t}' + o_p \left(1 \right) \quad (60)$$

due to equation (127). Hence, we can write

$$\widehat{y}_{t+1} = \overline{y}_t + \left(\widehat{F}_t - \overline{\widehat{F}}_t\right)' \widehat{\beta}_t, \tag{61}$$

which, together with (57), (60) and (104), implies

$$\widehat{y}_{t+1} = \overline{y}_t + (x_t - \overline{x}_t)' \widehat{a}_t + o_p(1), \tag{62}$$

where $\widehat{a}_t = J_N S'_{zx,H,t} \left(S_{zx,H,t} J_N S_{xx,L,t} J_N S'_{zx,H,t}\right)^{-1} S_{zx,H,t} J_N S_{xy,L,t}$.

Proof of (iii): Let S_i be a $1 \times N$ selector vector such that $S_i \widehat{a}_t = \widehat{a}_{it}$, then by algebra analogous to that presented in Lemma A5 we have

$$S_i \widehat{a}_t \underset{N,T\to\infty}{\overset{p}{\to}} N^{-1} S_i J_N \Phi_t \Delta_F \Lambda'_t \left(\Lambda_t \Delta_F \Delta_{\Phi_t} \Delta_F \Delta_{\Phi_t} \Delta_F \Lambda'_t\right)^{-1} \Lambda_t \Delta_F \Delta_{\Phi_t} \Delta_F \beta_t{}'.$$

This, by Assumption 3, Assumption 4 and Lemma A5, reduces to

$$N \widehat{a}_{it} \underset{N,T\to\infty}{\overset{p}{\to}} \left(\Phi_{it} - \overline{\Phi}_t\right)' \beta_t,$$

since $S_i J_N \Phi_t \underset{N,T\to\infty}{\overset{p}{\to}} (\Phi_{it} - \overline{\Phi}_t)'$, $N^{-1} \sum_{j=1}^N \Phi_{jt} \underset{T\to\infty}{\overset{p}{\to}} \overline{\Phi}_t$.

7.3.2 Proof of Corollary 1

When $K_f = 1$, the result follows directly from Theorem 1 by noting that the loading of y_{t+1} on F_t is $\beta_t = (\beta'_{ft}, 0')'$. Therefore, the target-proxy satisfies the condition of Assumption 4 and the consistency result follows. When $K_f > 1$, let $\hat{y}^{(1)}$ denote the 1 automatic proxy TV-3PRF forecast. Then, because of derivations in (54) and Lemma A3, we have

$$\begin{aligned} r_t^{(1)} &= y_t - \hat{y}_t^{(1)} && (63)\\ &= \beta_{0t} + F'_t \beta_t + \eta_t - \widehat{\beta}_{0t} - \widehat{F}'_t \widehat{\beta}_t && (64)\\ &= F'_t(\beta_t - \Phi_t'^{(1)}{}_t \widetilde{F}^{(t)} \beta_t) + \eta_t - \varepsilon_t \Omega_t^{(1)} \eta_t && (65) \end{aligned}$$

with

$$\Omega_t^{(1)} = S'_{zx,H,t} J_N (S_{zx,H,t} J_N S_{xx,L,t} J_N S'^{-1}_{zx,H,t} S_{zx,H,t} J_N \widetilde{X}^{(t)\prime}, \tag{66}$$

where $z_t = y_t$ in the autoproxy case that we consider here. Since $\beta_t = (\beta'_{ft}, 0')'$, the target y_t has zero covariance with the irrelevant factors, and also $\hat{y}_t^{(1)}$, $r_t^{(1)}$ have zero covariance with the irrelevant factors. This means that $r_t^{(1)}$ has zero loadings on the irrelevant factors.

Suppose we have $k < K_f$ automatically selected proxies, with factor loadings $[\Lambda_{kft}, 0]$, and Λ_{kft} is of dimension $k \times K_f$ with full row rank. Using the same arguments as in the $k = 1$ case, we can show that the residual $r_t^{(k)}$ has zero loading on the irrelevant factors. We further need to show that the $r_t^{(k)}$ factor loading on relevant factors is linearly independent of the rows of Λ_{kft}. From equation (63), we observe that these loadings take the form $\beta_t - \Phi_t'^{(1)}{}_t\widetilde{F}^{(t)}\beta_t$. Next, project the relevant factor loading of $r_t^{(k)}$ onto the column space of Λ_{kft}. For linear independence this projection must result a non zero residual. To this end, the residual is given by $(I - \Lambda'_{kft}(\Lambda_{kft}\Lambda'^{-1}_{kft}\Lambda_{kft})(\beta_t - \Phi_t'^{(1)}{}_t\widetilde{F}^{(t)}\beta_t)$, which equals zero when $\Lambda'_{kft}(\Lambda_{kft}\Lambda'^{-1}_{kft}\Lambda_{kft}) = I$. The last statement is true when $k = K_f$. Therefore, the proxies satisfy Assumption 4.

To verify assumptions that correspond to the error of the proxy equation (ω_t), notice that each automatic proxy is a forecast error $y_{t+1} - \widehat{y}_{t+1}$, where $\widehat{y}_{t+1}$ is expressed as

$$\widehat{y}_{t+1} = N^{-1}\alpha_t' x_t, \tag{67}$$

meaning that we can write the automatic proxy as $z_t = b_t f + \omega_t$ with $\omega_t = \eta_{t+1} + N^{-1}a_t'\varepsilon_t$. The η_t component directly satisfies 1.4, 2.3 and 2.4 by Assumption 1.5. Also the $N^{-1}a_t'\varepsilon_t$ component satisfy 1.4, 2.3 and 2.4, due to Assumptions 1.3, 2.2 and Theorem 3(ii). Then, all conditions of Theorem 1 and 5 are satisfied.

7.3.3 Proof of Theorem 2

Inspection of equation (19), allows us to derive the autoproxy TV-3PRF estimator in one step as

$$\widehat{y}_{t+1} = \overline{y}_{L,t} + (x_t - \overline{x}_{L,t})'\,\widehat{a}_t + o_p(1), \tag{68}$$

where

$$\widehat{a}_t = W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}S_{xy,L,t} \tag{69}$$

and $W^{(t)} = \left[J_N S_{xy,H,t},\ J_N S_{xx,H,t} J_N S_{xy,H,t},\ ..,\ (J_N S_{xx,H,t})^{K-1} J_N S_{xy,H,t}\right]$ is the Krylov subspaces sequence, for K factors extracted.

To this end, we use the generalized definition of degrees of freedom, proposed by Efron (2004), in which

$$DoF = E\left[trace\left(\frac{\partial\widehat{y}}{\partial y}\right)\right]. \tag{70}$$

Then, total differential of the estimator $\widehat{a}_t$ is given by

$$\begin{aligned}
d\widehat{a}_t &= dW^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}S_{xy,L,t} \\
&- W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}dW^{(t)\prime}S_{xx,L,t}W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}S_{xy,L,t} \\
&- W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}S_{xx,L,t}dW^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}S_{xy,L,t} \\
&+ W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}dW^{(t)\prime}S_{xy,L,t} + H^{(t)}dS_{xy,L,t},
\end{aligned} \tag{71}$$

where

$$H^{(t)} = W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}. \tag{72}$$

Then, we have

$$\begin{aligned}
d\widehat{a}_t &= dW^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}S_{xy,L,t} \\
&- W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}dW^{(t)\prime}S_{xx,L,t}H^{(t)}S_{xy,L,t} \\
&- H^{(t)}S_{xx,L,t}dW^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}W^{(t)\prime}S_{xy,L,t} \\
&+ W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}dW^{(t)\prime}S_{xy,L,t} + K_{L,t}^{-1}H^{(t)}\widetilde{X}_L^{(t)\prime}d\widetilde{y}_L^{(t)}.
\end{aligned} \tag{73}$$

Next, since $vec\left(ABC\right) = \left(C' \otimes A\right)vecB$, we have

$$\begin{aligned}
d\widehat{a}_t &= \left\{S'_{xy,L,t}W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1} \otimes I_N\right\}vec\left(dW^{(t)}\right) \\
&- \left\{S'_{xy,L,t}H^{(t)}S_{xx,L,t} \otimes W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}\right\}vec\left(dW^{(t)\prime}\right) \\
&- \left\{S'_{xy,L,t}W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1} \otimes H^{(t)}S_{xx,L,t}\right\}vec\left(dW^{(t)}\right) \\
&+ S'_{xy,L,t} \otimes W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}vec\left(dW^{(t)\prime}\right) \\
&+ K_{L,t}^{-1}H^{(t)}\widetilde{X}_L^{(t)\prime}d\widetilde{y}_L^{(t)}
\end{aligned} \tag{74}$$

and

$$\begin{aligned}
d\widehat{a}_t &= \left\{S'_{xy,L,t}W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1} \otimes \left(I_N - H^{(t)}S_{xx,L,t}\right)\right\}vec\left(dW^{(t)}\right) \\
&+ \left\{S'_{xy,L,t}\left(I_N - H^{(t)}S_{xx,L,t}\right) \otimes W^{(t)}\left(W^{(t)\prime}S_{xx,L,t}W^{(t)}\right)^{-1}\right\}vec\left(dW^{(t)\prime}\right) \\
&+ K_{L,t}^{-1}H^{(t)}\widetilde{X}_L^{(t)\prime}d\widetilde{y}_L^{(t)}.
\end{aligned} \tag{75}$$

Notice that C is $p \times 1$, B is $p \times m$, A is $p \times m$ then $(C' \otimes B)\, vec\,(A') = (B \otimes C')\, vec\,(A) = BA'C$ and

$$d\widehat{a}_t = \left\{ S'_{xy,L,t} W^{(t)} \left(W^{(t)\prime} S_{xx,L,t} W^{(t)} \right)^{-1} \otimes \left(I_N - H^{(t)} S_{xx,L,t} \right) \right\} vec\left(dW^{(t)} \right)$$
$$+ \left\{ W^{(t)} \left(W^{(t)\prime} S_{xx,L,t} W^{(t)} \right)^{-1} \otimes S'_{xy,L,t} \left(I_N - H^{(t)} S^{(t)} \right) \right\} vec\left(dW^{(t)} \right)$$
$$+ K_{L,t}^{-1} H^{(t)} \widetilde{X}_L^{(t)\prime} d\widetilde{y}_L^{(t)}, \tag{76}$$

or, equivalently,

$$d\widehat{a}_t = \left\{ \begin{array}{c} S'_{xy,L,t} W^{(t)} \left(W^{(t)\prime} S_{xx,L,t} W^{(t)} \right)^{-1} \otimes \left(I_N - H^{(t)} S_{xx,L,t} \right) + \\ + W^{(t)} \left(W^{(t)\prime} S_{xx,L,t} W^{(t)} \right)^{-1} \otimes S'_{xy,L,t} \left(I_N - H^{(t)} S^{(t)} \right) \end{array} \right\} vec\left(dW^{(t)} \right)$$
$$+ K_{L,t}^{-1} H^{(t)} \widetilde{X}_L^{(t)\prime} d\widetilde{y}_L^{(t)}$$
$$= \left(S'_{xy,L,t} \otimes I_N + I_N \otimes S'_{xy,L,t} \right) \left[W^{(t)} \left(W^{(t)\prime} S_{xx,L,t} W^{(t)} \right)^{-1} \otimes \left(I_N - H^{(t)} S_{xx,L,t} \right) \right] vec\left(dW^{(t)} \right)$$
$$+ K_{L,t}^{-1} H^{(t)} \widetilde{X}_L^{(t)\prime} d\widetilde{y}_L^{(t)}. \tag{77}$$

From the definition of $W^{(t)}$ we have

$$vec\left(W^{(t)} \right) = K_{H,t}^{-1} U'_K \widetilde{y}_H^{(t)}, \tag{78}$$

where $U_K = \left\{ \widetilde{X}_H^{(t)} J_N, \widetilde{X}_H^{(t)} J_N S_{xx,H,t} J_N, .., \widetilde{X}_H^{(t)} J_N \left(S_{xx,H,t} J_N \right)^{K-1} \right\}$ and $vec\left(dW^{(t)} \right) = dvec\left(W^{(t)} \right) = K_{H,t}^{-1} U'_K d\widetilde{y}^{(t)}$

Also, since $\widetilde{y}_H^{(t)} = \Lambda_H^{(t)} \left(y - \overline{y}_{H,t} 1_T \right)$, with $\Lambda_H^{(t)} = diag\left(\sqrt{k_{H,1t}}, \sqrt{k_{H,2t}}, .., \sqrt{k_{HTt}} \right)$, we have that $\widetilde{y}_H^{(t)} = \Lambda_H^{(t)} \left(I_T - \frac{1}{K_t} 1_T 1'_T \Lambda_H^{(t)} \Lambda_H^{(t)} \right) y$ and also

$$d\widetilde{y}_H^{(t)} = \Lambda_H^{(t)} \left(I_T - \frac{1}{K_t} 1_T 1'_T \Lambda_H^{(t)} \Lambda_H^{(t)} \right) dy. \tag{79}$$

Combining this information with the previous derivations, we obtain:

$$d\widehat{a}_t = \left(S'_{xy,L,t} \otimes I_N + I_N \otimes S'_{xy,L,t} \right) \left[H_{(1)}^{(t)} \otimes \left(I_N - H^{(t)} S_{xx,L,t} \right) \right] K_{H,t}^{-1} U'_K \Lambda_H^{(t)} \left(I_T - \frac{1}{K_{H,t}} 1_{TT} \Lambda_L^{(2,t)} \right) dy \tag{80}$$

$$+ K_{L,t}^{-1} H^{(t)} \widetilde{X}_L^{(t)\prime} \Lambda_L^{(t)} \left(I_T - \frac{1}{K_t} 1_{TT} \Lambda_L^{(2,t)} \right) dy,$$

where $H_{(1)}^{(t)} = W^{(t)} \left(W^{(t)\prime} S_{xx}^{(t)} W^{(t)} \right)^{-1}$, $1_{TT} = 1_T 1'_T$, $\Lambda_L^{(2,t)} = \left(\Lambda_L^{(t)} \right)^2$. Writing equation (68) in $T \times 1$ vector form as

$$\widehat{y} = \overline{y}_L + diag\left((x_1 - \overline{x}_{L1})', (x_2 - \overline{x}_{L2})', ..., (x_T - \overline{x}_{LT})' \right) vec\left(\widehat{A} \right) \tag{81}$$

where x_t is a $N \times 1$ vector of the N regressors at time t, $\widehat{A}$ is the $N \times T$ matrix $\widehat{A} = [\widehat{a}_1, \widehat{a}_2, .., \widehat{a}_T]$, and the $diag\left((x_1 - \overline{x}_{L1})', (x_2 - \overline{x}_{L2})', ..., (x_T - \overline{x}_{LT})'\right)$ is of dimension $T \times TN$, we finally obtain:

$$DoF(K) = \sum_{s=1}^{T} \frac{k_{s,s}}{K_s} + trace\left(diag((x_1 - \overline{x}_1)', (x_2 - \overline{x}_2)', ..., (x_T - \overline{x}_T)') \frac{\partial vec\left(\widehat{A}\right)}{\partial y}\right)$$
$$= \sum_{s=1}^{T} \frac{k_{L,ss}}{K_{L,s}} + \sum_{s=1}^{T} (x_s - \overline{x}_{L,s})' \frac{\partial a_s}{\partial y_{s+1}},$$

where $\frac{\partial a_t}{\partial y_{t+1}}$ is the t-th row of the $T \times N$ matrix given in (80).

7.3.4 Proof of Theorem 3

Proof of (i): In view of the Cramer-Wold device, it suffices to verify that for any vector ν with $\nu'\nu = 1$ the scalar random variable $p_{T,t} := \frac{K_t}{K_{2,t}^{1/2}} \nu' \Gamma_{F\eta,t}^{-1/2} S_{F\eta,t}$, has the property

$$p_{T,t} \xrightarrow{d} N(0,1). \tag{82}$$

Notice that for $\xi_{t_s} := K_{2,t}^{-1/2} \nu' \Gamma_{F\eta,t}^{-1/2} F_s \eta_{s+1}$ we have that

$$p_{T,t} = \sum_{s=1}^{T} k_{t,s} \xi_{t_s}.$$

Also, $\nu' \Gamma_{F\eta,t}^{-1/2} F_s \eta_{s+1}$ is a strong mixing sequence since F_s, η_{s+1} are both strong mixing sequences. Further, $E\left(\nu' \Gamma_{F\eta,t}^{-1/2} F_s \eta_{s+1}\right) = 0$, as $E\left(F_s \eta_{s+1}\right) = 0$ and $E\left|\nu' \Gamma_{F\eta,t}^{-1/2} F_s \eta_{s+1}\right|^r < M < \infty$ for all s and $E\left|F_s \eta_{s+1}\right|^r < M < \infty$ for $r > 2$. Also notice that

$$var\left(K_{2,t}^{-1/2} \sum_{s=1}^{T} k_{t,s} v' \Gamma_{F\eta,t}^{-1/2} F_s \eta_{s+1}\right) = \nu' \Gamma_{F\eta,t}^{-1/2} var\left(K_{2,t}^{-1/2} \sum_{s=1}^{T} k_{t,s} F_s \eta_{s+1}\right) \Gamma_{F\eta,t}^{-1/2} \nu$$
$$= \nu' \Gamma_{F\eta,t}^{-1/2} K_{2,t}^{-1} \sum_{s=1}^{T} k_{t,s}^2 E\left(F_s F_s' \eta_{s+1}^2\right) \Gamma_{F\eta,t}^{-1/2} \nu \to 1.$$

The result then follows by White (2014), Theorem 5.20.

Proof of (ii): As in element (i), the scalar random variable $p_{T,t} := \frac{K_t}{K_{2,t}^{1/2}} \nu' \Gamma_{F\varepsilon,it}^{-1/2} C_{F\varepsilon,it}$ has the property

$$p_{T,t} \xrightarrow{d} N(0,1). \tag{83}$$

Notice that, for $\xi_{t_s} := K_{2,t}^{-1/2} \nu' \Gamma_{F\varepsilon,it}^{-1/2} F_s \varepsilon_{is}$, we have that

$$p_{T,t} = \sum_{s=1}^{T} k_{t,s} \xi_{t_s}. \tag{84}$$

Also, $\nu' \Gamma_{F\varepsilon,it}^{-1/2} F_s \varepsilon_{is}$ is a strong mixing sequence since F_s, ε_{is} are both strong mixing sequences. Further, $E\left(\nu' \Gamma_{F\varepsilon,it}^{-1/2} F_s \varepsilon_{is}\right) = 0$, as $E\left(F_s \varepsilon_{is}\right) = 0$ and $E\left|\nu' \Gamma_{F\varepsilon,it}^{-1/2} F_s \varepsilon_{is}\right|^r < M < \infty$ for all s since $E\left|F_s \varepsilon_{is}\right|^r < M < \infty$, for $r > 2$. Moreover,

$$\begin{aligned}
\widetilde{\sigma}^2 &:= var\left(K_{2,t}^{-1/2} \nu' \Gamma_{F\varepsilon,it}^{-1/2} \sum_{s=1}^{T} k_{t,s} F_s \varepsilon_{is}\right) \\
&= \nu' \Gamma_{F\varepsilon,it}^{-1/2} var\left(K_{2,t}^{-1/2} \sum_{s=1}^{T} k_{t,s} F_s \varepsilon_{is}\right) \Gamma_{F\varepsilon,it}^{-1/2} \nu \\
&= \nu' \Gamma_{F\varepsilon,it}^{-1/2} \left(K_{2,t}^{-1} \sum_{s=1}^{T} \sum_{r=1}^{T} k_{t,s} k_{t,r} E\left(F_s F_r' \varepsilon_{is} \varepsilon_{ir}\right)\right) \Gamma_{F\varepsilon,it}^{-1/2} \nu \to 1.
\end{aligned}$$

The result follows from White (2014), Theorem 5.20

Proof of (iii): It suffices to verify that for any vector ν with $\nu'\nu = 1$, the scalar random variable $p_{N,st} := N^{1/2} \nu' \Gamma_{\Phi\varepsilon,st}^{-1/2} C_{\phi\varepsilon,ist}$, has the property

$$p_{N,st} \xrightarrow{d} N(0,1). \tag{85}$$

Defining $\xi_{ist} := N^{-1/2} \nu' \Gamma_{\Phi\varepsilon,st}^{-1/2} \Phi_{is} \varepsilon_{it}$, then

$$p_{N,st} = \sum_{i=1}^{N} \xi_{ist} = N^{-1/2} \sum_{i=1}^{N} \nu' \Gamma_{\Phi\varepsilon,st}^{-1/2} \Phi_{is} \varepsilon_{it}. \tag{86}$$

Also, $\nu' \Gamma_{\Phi\varepsilon,t}^{-1/2} \Phi_{it} \varepsilon_{it}$ is a strong mixing sequence since Φ_{is} is a triangular array of vectors and ε_{it} is a strong mixing sequence. Further $E\left(\nu' \Gamma_{\Phi\varepsilon,st}^{-1/2} \Phi_{is} \varepsilon_{it}\right) = 0$ since $E\left(\Phi_{is} \varepsilon_{it}\right) = 0$ and $E\left(\left|\nu' \Gamma_{\Phi\varepsilon,st}^{-1/2} \Phi_{is} \varepsilon_{it}\right|^r\right) < M < \infty$ for all i since $E\left(\left|\Phi_{is} \varepsilon_{it}\right|^r\right) < M < \infty$, $r > 2$. Finally,

$$\begin{aligned}
\widetilde{\widetilde{\sigma}}^2 &:= var\left(N^{-1/2} \sum_{i=1}^{N} \nu' \Gamma_{\Phi\varepsilon,st}^{-1/2} \Phi_{is} \varepsilon_{it}\right) \\
&= \nu' \Gamma_{\Phi\varepsilon,st}^{-1/2} var\left(N^{-1/2} \sum_{i=1}^{N} \Phi_{is} \varepsilon_{it}\right) \Gamma_{\Phi\varepsilon,st}^{-1/2} \nu \\
&= \nu' \Gamma_{\Phi\varepsilon,st}^{-1/2} \left(N \sum_{i=1}^{N} \sum_{j=1}^{N} E\left(\Phi_{is} \Phi_{js} \varepsilon_{it} \varepsilon_{jt}\right)\right) \Gamma_{\Phi\varepsilon,st}^{-1/2} \nu \to 1.
\end{aligned}$$

The result follows from White (2014), Theorem 5.20.

*Proof of (iv):*Notice that

$$
\begin{aligned}
N\widehat{a}_{it} - N\widetilde{a}_{it} &= NS_iJ_NS'_{zx,H,t}\left(S_{zx,H,t}J_NS_{xx,L,t}J_NS'_{zx,H,t}\right)^{-1}S_{zx,H,t}J_NS_{xy,L,t} \\
&\quad - NS_iJ_NS'_{zx,H,t}\left(\frac{1}{N^2}S_{zx,H,t}J_NS_{xx,L,t}J_NS'_{zx,H,t}\right)^{-1} \\
&\quad \times \frac{1}{N}S_{zx,H,t}J_N\frac{1}{K_{L,t}}\widetilde{X}_L^{(t)\prime}\widetilde{F}_L^{(t)}\beta_t
\end{aligned}
$$

$$
\begin{aligned}
&= NS_iJ_NS'_{zx,H,t}\left(S_{zx,H,t}J_NS_{xx,L,t}J_NS'_{zx,H,t}\right)^{-1}S_{zx,H,t}J_N\frac{1}{K_{L,t}}\widetilde{X}_L^{(t)\prime}\widetilde{F}_L^{(t)}\beta_t \\
&+ NS_iJ_NS'_{zx,H,t}\left(S_{zx,H,t}J_NS_{xx,L,t}J_NS'_{zx,H,t}\right)^{-1}S_{zx,H,t}J_N\frac{1}{K_{L,t}}\widetilde{X}_L^{(t)\prime}\widetilde{\eta}_{L,t} \\
&- NS_iJ_NS'_{zx,H,t}\left(\frac{1}{N^2}S_{zx,H,t}J_NS_{xx,L,t}J_NS'_{zx,H,t}\right)^{-1} \\
&\times \frac{1}{N}S_{zx,H,t}J_N\frac{1}{K_{L,t}}\widetilde{X}_L^{(t)\prime}\widetilde{F}_L^{(t)}\beta_t
\end{aligned}
$$

$$
= S_iJ_NS'_{zx,H,t}\left(N^{-2}S_{zx,H,t}J_NS_{xx,L,t}J_NS'_{zx,H,t}\right)^{-1}N^{-1}S_{zx,H,t}J_N\frac{1}{K_{L,t}}\widetilde{X}_L^{(t)\prime}\widetilde{\eta}_{L,t}.
$$

The result follows from Lemma A8 since $\widehat{\eta}_t = \eta_t + o_p(1)$.

Proof of (v): See proof of Corollary 1

7.3.5 Proof of Theorem 4

Proof of (i): First, set $\widetilde{y}_{t+1} = \overline{y}_{L,t} + (x_t - \overline{x}_{L,t})' G_{a,t}\beta_t$ and since $\widehat{y}_{t+1} = \overline{y}_{L,t} + (x_t - \overline{x}_{L,t})'\,\widehat{a}_t + o_p(1)$, we have that

$$
\begin{aligned}
\sqrt{K_{L,t}}\left(\widehat{y}_{t+1} - \widetilde{y}_{t+1}\right) &= \sqrt{K_{L,t}}\left((x_t - \overline{x}_{L,t})'\,\widehat{a}_t - (x_t - \overline{x}_{L,t})' G_{a,t}\beta_t\right) \\
&= N^{-1}(x_t - \overline{x}_{L,t})'\sqrt{K_{L,t}}N\left(\widehat{a}_t - \widetilde{a}_t\right),
\end{aligned}
$$

for $\widetilde{a}_{it} = S_iG_{a,t}\beta_t$, which implies that

$$
\frac{\sqrt{K_{L,t}}\left(\widehat{y}_{t+1} - \widetilde{y}_{t+1}\right)}{Q_t} \xrightarrow{d} N(0,1),
$$

where $Q_t^2 = N^{-2}(x_t - \overline{x}_{L,t})'\,\widehat{Avar}(\widehat{a}_t)(x_t - \overline{x}_{L,t})$. Also, by Theorem 1 we have

$$
\widetilde{y}_{t+1} = \left(\overline{y}_t + (x_t - \overline{x}_t)' G_{a,t}\beta_t\right) \underset{N,T\to\infty}{\xrightarrow{p}} E_t y_{t+1},
$$

and by Slutsky's Theorem convergence in distribution follows.

Proof of (ii): Notice that

$$\begin{aligned}\left(\widehat{\beta}_t - G_{\beta,t}\beta_t\right) &= \widehat{\beta}_{1,t}^{-1}\widehat{\beta}_{2,t}\widehat{\beta}_{3,t}^{-1}N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N K_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{y}_L^{(t)} \\ &- \widehat{\beta}_{1,t}^{-1}\widehat{\beta}_{2,t}\widehat{\beta}_{3,t}^{-1}N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N K_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{F}_L^{(t)}\beta_t \\ &+ o_p(1) \\ &= \widehat{\beta}_{1,t}^{-1}\widehat{\beta}_{2,t}\widehat{\beta}_{3,t}^{-1}N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N K_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{\eta}_L^{(t)}.\end{aligned}$$

Then,

$$\begin{aligned}K_{L,t}^{1/2}\left(\widehat{\beta}_t - G_{\beta,t}\beta_t\right) &= \widehat{\beta}_{1,t}^{-1}\widehat{\beta}_{2,t}\widehat{\beta}_{3,t}^{-1}N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N K_{L,t}^{-1/2}\widetilde{X}_L^{(t)\prime}\widetilde{\eta}_L^{(t)} \\ &\xrightarrow{d} N\left(0, \Sigma_{\beta,t}\right)\end{aligned}$$

by Lemma A8, and following the results in Lemma A4 and Lemma A5, we have that

$$\Sigma_{\beta,t} = \Psi_{\beta,t}\Gamma_{F\eta,t}\Psi_{\beta,t}',$$

where

$$\Psi_{\beta,t} = \Sigma_{z,t}^{-1}\left(\Lambda_t\Delta_F\Delta_{\Phi_t}\Delta_F\Lambda_t'\right)\left(\Lambda_t\Delta_F\Delta_{\Phi_t}\Delta_F\Delta_{\Phi_t}\Delta_F\Lambda_t'\right)^{-1}\Lambda_t\Delta_F\Delta_\Phi$$

with $\Sigma_{z,t} = \Lambda_t\Delta_F\Lambda_t' + \Delta_\omega$ and

$$\begin{aligned}\widehat{\beta}_{1,t} &= K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{Z}_H^{(t)} \xrightarrow[N,T\to\infty]{p} \Lambda_t\Delta_F\Lambda_t' + \Delta_\omega \\ \widehat{\beta}_{2,t} &= N^{-1}K_{H,t}^{-2}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\widetilde{X}_H^{(t)\prime}\widetilde{Z}_H^{(t)} \xrightarrow[N,T\to\infty]{p} \Lambda_t\Delta_F\Delta_{\Phi_t}\Delta_F\Lambda_t' \\ \widehat{\beta}_{3,t} &= N^{-2}K_{H,t}^{-2}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N K_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{X}_L^{(t)}J_N\widetilde{X}_H^{(t)\prime}\widetilde{Z}_H^{(t)} \xrightarrow[N,T\to\infty]{p} \Lambda_t\Delta_F\Delta_\Phi\Delta_F\Delta_\Phi\Delta_F\Lambda_t'.\end{aligned}$$

By Assumptions 1-4, we have that

$$\Sigma_{\beta,t} = \Sigma_{z,t}^{-1}\Gamma_{F\eta,t}\Sigma_{z,t}^{-1}.$$

To show that $\widehat{Avar}\left(\widehat{\beta}_t\right)$ is a consistent estimator of $\Sigma_{\beta,t}$, notice that

$$\begin{aligned}K_{L,t}^{1/2}\left(\widehat{\beta}_t - G_{\beta,t}\beta_t\right) &= \widehat{\beta}_{1,t}^{-1}\widehat{\beta}_{2,t}\widehat{\beta}_{3,t}^{-1}N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N K_{L,t}^{-1/2}\widetilde{X}_L^{(t)\prime}\widetilde{\eta}_L^{(t)} + o_p(1) \\ &= \left(K_{L,t}^{-1}\widehat{\widetilde{F}}_L^{(t)\prime}\widehat{\widetilde{F}}_L^{(t)}\right)^{-1}K_{L,t}^{-1/2}\widehat{\widetilde{F}}_L^{(t)\prime}\widetilde{\eta}_L^{(t)} + o_p(1),\end{aligned}$$

which implies that the asymptotic variance of $\widehat{\beta}_t$ equals to the probability limit of

$$\left(K_{L,t}^{-1}\widehat{\widetilde{F}}_L^{(t)\prime}\widehat{\widetilde{F}}_L^{(t)}\right)^{-1}K_{L,t}^{-1}\widehat{\widetilde{F}}_L^{(t)\prime}\widetilde{\eta}_L^{(t)}\widetilde{\eta}_L^{(t)}\widehat{\widetilde{F}}_L^{(t)\prime}\left(K_{L,t}^{-1}\widehat{\widetilde{F}}_L^{(t)\prime}\widehat{\widetilde{F}}_L^{(t)}\right)^{-1}. \tag{87}$$

Assumpion 1 and Lemma A7 imply that (87) equals to

$$K_{L,t}^{-1}\widetilde{\widehat{F}}_{L}^{(t)\prime}\widetilde{\eta}_{L}^{(t)}\widetilde{\eta}_{L}^{(t)}\widetilde{\widehat{F}}_{L}^{(t)\prime} \underset{N,T\to\infty}{\overset{p}{\to}} K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\eta_{s+1}-\overline{\eta}_{t}\right)^{2}k_{L,ts}\left(\widehat{F}_{s}-\overline{F}_{L,t}\right)\left(\widehat{F}_{s}-\overline{F}_{L,t}\right)'$$

$$\underset{N,T\to\infty}{\overset{p}{\to}} K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\eta_{s+1}^{2}k_{L,ts}\left(\widehat{F}_{s}-\overline{F}_{L,t}\right)\left(\widehat{F}_{s}-\overline{F}_{L,t}\right)'.$$

By Theorem 1, we have that $\eta_t = \widehat{\eta}_t + o_p(1)$, $\overline{\eta}_t = o_p(1)$ which implies that (87) and $\widehat{Avar}\left(\widehat{\beta}_t\right)$ share the same probability limit and therefore $\widehat{Avar}\left(\widehat{\beta}_t\right)$ is a consistent estimator for $\Sigma_{\beta,t}$.

7.3.6 Proof of Theorem 5

First, note that

$$\left(\widehat{F}_t-\left(H_{0,t}+H_tF_t\right)\right)=\widehat{F}_{A,t}\widehat{F}_{B,t}^{-1}N^{-1}K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)}J_N\varepsilon_t.$$

Using Lemma A9, we have that

$$\sqrt{N}\left(\widehat{F}_t-\left(H_{0,t}+H_tF_t\right)\right)\overset{d}{\to}N\left(0,\Sigma_{F,t}\right),$$

with $\Sigma_{F,t}$ derived from results in Lemma A9, Lemma A4 and assumption 3,4.

7.3.7 Proof of Theorem 6

An equivalent expression is

$$\begin{aligned}&\arg\min_{a_t}\sum_{s=1}^{T}k_{L,s,t}\left(y_{s+1}-\overline{y}_{L,t}-\left(x_s-\overline{x}_{L,t}\right)'a_t\right)^2 \qquad (88)\\&=\arg\min_{a_t}\left(\widetilde{y}_L^{(t)}-\widetilde{X}_L^{(t)}a_t\right)'\left(\widetilde{y}_L^{(t)}-\widetilde{X}_L^{(t)}a_t\right).\end{aligned}$$

Then, a_t can be estimated by the minimization of the weighted sum of squared errors. Notice that since the constraint is of the form $Q_t'a_t = 0$, by Amemiya (1985) (Section 1.4), the solution is given as

$$\begin{aligned}\widehat{a}_t&=R_t\left(R_t'S_{xx,L,t}R_t\right)^{-1}R_t'S_{xy,L,t}\\S_{xx,L,t}&=K_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{X}_L^{(t)}\\S_{xy,L,t}&=K_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{y}_L^{(t)},\end{aligned}$$

where R_t is such that $R_t'Q_t = 0$ and $[R_t', Q_t]$ is non singular. The result follows when $R_t = W_{xz,H,t}$, for which we have

$$\widehat{a}_t=W_{xz,H,t}\left(W_{xz,H,t}'S_{xx,L,t}W_{xz,H,t}\right)^{-1}W_{xz,H,t}'S_{xy,L,t}.$$

7.4 Auxiliary Lemmas

The results presented in this part of the Appendix are necessary for the proofs of the main theoretical results given in section 7.3. The notation used here follows that introduced in section 7.2.

Lemma A1 *Let Assumptions 1,2 and 5 hold then for* $\delta_{NK_s} = \min\left(\sqrt{N}, \sqrt{K_s}\right)$ *and all* $r = 1, .., T$.

1. $E\left|N^{-1/2}K_s^{-1/2}\sum_{i,u} k_{su}F_{mu}\left[\varepsilon_{iu}\varepsilon_{it} - \sigma_{ii,ut}\right]\right|^2 \leq M$
2. $E\left|N^{-1/2}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T} k_{su}\omega_{mu}\left[\varepsilon_{iu}\varepsilon_{it} - \sigma_{ii,ut}\right]\right|^2 \leq M$
3. (a) $N^{-1/2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{j=1}^{N} k_{st}\varepsilon_{jt} = O_p(1)$
   (b) $N^{-1/2}\sum_{j=1}^{N}\varepsilon_{jt} = O_p(1)$
   (c) $K_s^{-1/2}\sum_{t=1}^{T} k_{st}\varepsilon_{jt} = O_p(1)$
4. $K_s^{-1/2}\sum_{t=1}^{T} k_{st}\eta_{t+1} = O_p(1)$
5. $K_s^{-1/2}\sum_{t=1}^{T} k_{st}\varepsilon_{it}\eta_{t+1} = O_p(1)$
6. $N^{-1/2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N} k_{st}\varepsilon_{it}\eta_{t+1} = O_p(1)$
7. $N^{-1}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{mir}k_{st}\varepsilon_{it}F_{vt} = O_p(1)$
8. $N^{-1}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{mir}k_{st}\varepsilon_{it}\omega_{vt} = O_p(1)$
9. $N^{-1/2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{mir}k_{ts}\varepsilon_{it}\eta_{t+1} = O_p(1)$
10. $N^{-1}K_s^{-1/2}\sum_{i,u} k_{ts}\varepsilon_{iu}\varepsilon_{it} = O\left(\delta_{NK_S}^{-1}\right)$
11. $N^{-1}K_s^{-3/2}\sum_{t=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N} k_{su}\varepsilon_{iu}\eta_{u+1}k_{s,t}\varepsilon_{it} = O\left(\delta_{NK_S}^{-1}\right)$
12. $K_s^{-1/2}N^{-1}\sum_{u=1}^{T}\sum_{i=1}^{N} k_{su}F_{mu}\varepsilon_{iu}\varepsilon_{it} = O_p\left(\delta_{N,K_s}^{-1}\right)$
13. $K_s^{-1/2}N^{-1}\sum_{u=1}^{T}\sum_{i=1}^{N} k_{su}\omega_{mu}\varepsilon_{iu}\varepsilon_{it} = O_p\left(\delta_{N,K_s}^{-1}\right)$
14. $N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N} k_{su}\varepsilon_{iu}\eta_{u+1}k_{st}\varepsilon_{it}F_{mt} = O_p(1)$

7.4.1 Proof of Lemma A1

In the following we use the result

$$K_s^{-1}\sum_{t=1}^{T} k_{s,t}^2 = O(1) \tag{89}$$

from 6.1 of Giraitis *et al.* (2014).

1. $E\left|N^{-1/2}K_s^{-1/2}\sum_{i,u} k_{s,u}F_{mu}\left[\varepsilon_{iu}\varepsilon_{it} - \sigma_{ii,ut}\right]\right|^2$
$= E\left|N^{-1}K_s^{-1}\sum_{i,u,j,v} k_{s,u}F_{mu}\left[\varepsilon_{iu}^{(s)}\varepsilon_{it}^{(s)} - \sigma_{ii,ut}\right]k_{s,v}F_{mv}\left[\varepsilon_{jv}\varepsilon_{jt} - \sigma_{jj,vt}\right]\right|$

$\leq \max_{uv} E\left|F_{mu}F_{mv}\right| E\left|N^{-1}K_s^{-1}\sum_{i,u,j,v} k_{s,u}k_{s,v}\left[\varepsilon_{iu}\varepsilon_{it}-\sigma_{ii,ut}\right]\left[\varepsilon_{jv}\varepsilon_{jt}-\sigma_{jj,vt}\right]\right|$
$\leq \max_{uv} E\left|F_{mu}F_{mv}\right| E\left|N^{-1/2}K_s^{-1/2}\sum_{i,u} k_{s,u}\left[\varepsilon_{iu}\varepsilon_{it}-\sigma_{ii,ut}\right]\right|^2$
because of Assumptions 1 and 2.

2. As in item 1.

3. (a) $E\left|N^{-1/2}K_s^{-1/2}\sum_{t=1}^T\sum_{j=1}^N k_{s,t}\varepsilon_{jt}\right|^2 =$
$= E\left(N^{-1}K_s^{-1}\sum_{j=1}^N\sum_{t=1}^T\sum_{i=1}^N\sum_{l=1}^T k_{s,t}\varepsilon_{jt}k_{s,l}\varepsilon_{il}\right) =$
$=N^{-1}K_s^{-1}\sum_{j=1}^N\sum_{t=1}^T\sum_{i=1}^N\sum_{l=1}^T k_{s,t}k_{s,l}\sigma_{ijtl}$
$\leq N^{-1}K_s^{-1}\sum_{j=1}^N\sum_{t=1}^T\sum_{i=1}^N\sum_{l=1}^T k_{s,t}k_{s,l}\left|\sigma_{ijtl}\right|$
$= O_p(1)$
by Assumption 2.1 (d).

(b) $E\left|N^{-1/2}\sum_{j=1}^N\varepsilon_{jt}\right|^2 = E\left(N^{-1}\sum_{j=1}^N\sum_{i=1}^N\varepsilon_{jt}\varepsilon_{it}\right)$
$= N^{-1}\sum_{j=1}^N\sum_{i=1}^N\sigma_{ij,tt} \leq N^{-1}\sum_{j=1}^N\sum_{i=1}^N\left|\sigma_{ij,tt}\right| = O_p(1)$
by Assumption 2.1 (a).

(c) $E\left|K_s^{-1/2}\sum_{t=1}^T k_{s,t}\varepsilon_{jt}\right|^2 = K_s^{-1}E\left(\sum_{t=1}^T\sum_{u=1}^T k_{s,t}\varepsilon_{jt}k_{s,u}\varepsilon_{ju}\right)$
$= K_s^{-1}\sum_{t=1}^T\sum_{u=1}^T k_{s,t}k_{s,u}\sigma_{jj,tu} \leq K_s^{-1}\sum_{t=1}^T\sum_{u=1}^T k_{s,t}k_{s,u}\left|\sigma_{jj,tu}\right| = O_p(1)$
by Assumption 2.1 (c).

4. $E\left|K_s^{-1/2}\sum_{t=1}^T k_{s,t}\eta_{t+1}\right|^2 = E\left(K_s^{-1}\sum_{t=1}^T\sum_{l=1}^T k_{s,t}\eta_{t+1}\eta_{l+1}k_{s,l}\right)$
$= K_s^{-1}\sum_{t=1}^T k_{s,t}^2 E\left(\eta_{t+1}^2\right) = O_p(1)$
by Assumptions 1.5 and (89).

5. $E\left|K_s^{-1/2}\sum_{t=1}^T k_{ts}\varepsilon_{it}\eta_{t+1}\right|^2 = E\left(K_s^{-1}\sum_{l=1}^T\sum_{t=1}^T k_{ts}k_{tl}\varepsilon_{it}\eta_{t+1}\varepsilon_{il}\eta_{l+1}\right)$
$= K_s^{-1}\sum_{t=1}^T k_{ts}k_{ts}E\left(\varepsilon_{it}^2\right)E\left(\eta_{t+1}^2\right) = K_s^{-1}\sum_{t=1}^T k_{ts}k_{ts}\sigma_{ii,tt}E\left(\eta_{t+1}^2\right)$
$\leq K_s^{-1}\sum_{t=1}^T k_{ts}^2\bar{\sigma}_{ii}E\left(\eta_{t+1}^2\right)$
by Assumptions 1.5, 2.1 and (89).

6. $E\left|N^{-1/2}K_s^{-1/2}\sum_{t=1}^T\sum_{i=1}^N k_{ts}\varepsilon_{it}\eta_{t+1}\right|^2$
$= E\left(N^{-1}K_s^{-1}\sum_{t=1}^T\sum_{i=1}^N\sum_{l=1}^T\sum_{j=1}^N k_{ts}k_{tl}\varepsilon_{it}\eta_{t+1}\varepsilon_{jl}\eta_{l+1}\right)$
$= N^{-1}\sum_{i=1}^N\sum_{j=1}^N K_s^{-1}\sum_{t=1}^T k_{s,t}^2\sigma_{ij,tt}E\left(\eta_{t+1}^2\right)$
$\leq N^{-1}\sum_{i=1}^N\sum_{j=1}^N\bar{\sigma}_{ij}K_s^{-1}\sum_{t=1}^T k_{s,t}^2 E\left(\eta_{t+1}^2\right) = O_p(1),$
by Assumptions 1.5, 2.1 and (89).

7. By the Cauchy inequality $N^{-1}K_s^{-1/2}\sum_{t=1}^T\sum_{i=1}^N \phi_{mir}k_{st}\varepsilon_{it}F_{vt}$

$= N^{-1}\sum_{i=1}^{N}\phi_{mir}\left(K_s^{-1/2}\sum_{t=1}^{T}k_{st}\varepsilon_{it}F_{vt}\right)$

$\leq\left(N^{-1}\sum_{i=1}^{N}\phi_{mir}^2\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}\left(K_s^{-1/2}\sum_{t=1}^{T}k_{st}\varepsilon_{it}F_{vt}\right)^2\right)^{1/2} = O_p(1)$

by Assumption 1.2 and Theorem 3.

8. $N^{-1}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{mir}k_{st}\varepsilon_{it}\omega_{vt} = N^{-1}\sum_{i=1}^{N}\phi_{mir}\left(K_s^{-1/2}\sum_{t=1}^{T}\varepsilon_{it}^{(s)}\omega_{vt}^{(s)}\right)$

$\leq\left(N^{-1}\sum_{i=1}^{N}\phi_{mir}^2\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}\left(K_s^{-1/2}\sum_{t=1}^{T}k_{st}\varepsilon_{it}\omega_{vt}\right)^2\right)^{1/2} = O_p(1)$

by Assumptions 1.2 and 2.4.

9. $E\left|N^{-1/2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{mir}\varepsilon_{it}^{(s)}\eta_{t+1}^{(s)}\right|^2$

$= E\left(N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{l=1}^{T}\sum_{j=1}^{N}\phi_{mir}\varepsilon_{it}^{(s)}\eta_{t+1}^{(s)}\phi_{mjr}\varepsilon_{jl}^{(s)}\eta_{l+1}^{(s)}\right)$

$= N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left|\phi_{mir}\phi_{mjr}k_{ts}\varepsilon_{it}\varepsilon_{jt}k_{ts}\left(\eta_{t+1}^{(s)}\right)^2\right|$

$= K_s^{-1}\sum_{t=1}^{T}k_{ts}E\left(\eta_t^2\right)k_{ts}E\left(N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{mir}\phi_{mjr}\varepsilon_{it}\varepsilon_{jt}\right)$

$= K_s^{-1}\sum_{t=1}^{T}k_{ts}^2E\left(\eta_t^2\right)E\left(N^{-1/2}\sum_{i=1}^{N}\phi_{mir}\varepsilon_{it}\right)^2 = O_p(1)$

by Assumption 1.5, Theorem 3 and (89).

10. $N^{-1}K_s^{-1/2}\sum_{i,u}k_{su}\varepsilon_{iu}\varepsilon_{it} =$

$= N^{-1}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{su}\left(\varepsilon_{iu}\varepsilon_{it}-\sigma_{ii,ut}\right)$

$+N^{-1}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{su}\sigma_{ii,ut}$

$\leq N^{-1}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{su}\left(\varepsilon_{iu}\varepsilon_{it}-\sigma_{ii,ut}\right)$

$+\max_u\left(k_{su}\right)N^{-1}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T}\sigma_{ii,ut}$

$= O_p\left(N^{-1/2}\right)+O_p\left(K_s^{-1/2}\right) = O\left(\delta_{NK_s}^{-1}\right)$

by Assumptions 2.1 (b) and 2.2.

11. $N^{-1}K_s^{-3/2}\sum_{t=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}k_{us}\varepsilon_{iu}\eta_{u+1}k_{st}\varepsilon_{it}$

$\leq K_s^{-1}\sum_{u}^{T}k_{su}^{1/2}\eta_{u+1}\left(k_{su}^{1/2}N^{-1}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}k_{st}\varepsilon_{iu}\varepsilon_{it}\right)$

$\leq\left(K_s^{-1}\sum_{u}^{T}k_{su}\eta_{u+1}^2\right)^{1/2}\left(K_s^{-1}\sum_{u}^{T}k_{su}\left(N^{-1}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}k_{st}\varepsilon_{iu}\varepsilon_{it}\right)^2\right)^{1/2}$

$= O\left(\delta_{NK_S}^{-1}\right)$

by Assumption 1.5 and item 10.

12. $K_s^{-1/2}N^{-1}\sum_{u=1}^{T}\sum_{i=1}^{N}k_{su}F_{mu}\varepsilon_{iu}\varepsilon_{it}$

$=N^{-1}K_s^{-1/2}\sum_{u=1}^{T}\sum_{i=1}^{N}k_{su}F_{mu}\left(\varepsilon_{iu}\varepsilon_{it}-\sigma_{ii,ut}\right)$

$+N^{-1}K_s^{-1/2}\sum_{u=1}^{T}\sum_{i=1}^{N}k_{su}F_{mu}\sigma_{ii,ut}$

$\leq N^{-1}K_s^{-1/2}\sum_{u=1}^{T}\sum_{i=1}^{N}k_{su}F_{mu}\left(\varepsilon_{iu}\varepsilon_{it}-\sigma_{ii,ut}\right)$

$+\max_t\left(k_{st}\right)K_s^{-1/2}\max_u\left(\left|F_{mu}\right|\right)N^{-1}\sum_{u=1}^{T}\sum_{i=1}^{N}\sigma_{ii,ut}$

$=O_p\left(N^{-1/2}\right)+O_p\left(K_s^{-1/2}\right)$
by Assumption 2.1(b) and item 1.

13. $K_s^{-1/2}N^{-1}\sum_{u=1}^{T}\sum_{i=1}^{N}k_{su}\omega_{mu}\varepsilon_{iu}\varepsilon_{it}$
$=N^{-1}K_s^{-1/2}\sum_{u=1}^{T}\sum_{i=1}^{N}k_{su}\omega_{mu}\left(\varepsilon_{iu}\varepsilon_{it}-\sigma_{ii,ut}\right)$
$+N^{-1}K_s^{-1/2}\sum_{u=1}^{T}\sum_{i=1}^{N}k_{su}\omega_{mu}\sigma_{ii,ut}$
$\leq N^{-1}K_s^{-1/2}\sum_{u=1}^{T}\sum_{i=1}^{N}k_{su}\omega_{mu}\left(\varepsilon_{iu}\varepsilon_{it}-\sigma_{ii,ut}\right)$
$+\max_t\left(k_{st}\right)K_s^{-1/2}\max_u\left(|\omega_{mu}|\right)N^{-1}\sum_{u=1}^{T}\sum_{i=1}^{N}\sigma_{ii,ut}$
$=O_p\left(N^{-1/2}\right)+O_p\left(K_s^{-1/2}\right)$
by Assumption 2.1 (b) and item 2.

14. $N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}k_{su}\varepsilon_{iu}\eta_{u+1}k_{st}\varepsilon_{it}F_{mt}$
$=N^{-1}\sum_{i=1}^{N}\left(K_s^{-1/2}\sum_{t=1}^{T}k_{st}\varepsilon_{it}F_{mt}\right)\left(K_s^{-1/2}\sum_{u}^{T}k_{us}\varepsilon_{iu}\eta_{u+1}\right)$
$\leq\left(N^{-1}\sum_{i=1}^{N}\left(K_s^{-1/2}\sum_{t=1}^{T}k_{st}\varepsilon_{it}F_{mt}\right)^2\right)^{1/2}\left(N^{-1}\sum_{i=1}^{N}\left(K_s^{-1/2}\sum_{u}^{T}k_{us}\varepsilon_{iu}\eta_{u+1}\right)^2\right)^{1/2}$
$=O_p\left(1\right)$,
by Theorem 3 and item 5.

Lemma A2 *Let Assumptions 1,2 and 5 hold. Then, for* $r=1,..,T$ *and* $\delta_{NK_s}=\min\left(N^{1/2},K_s^{1/2}\right)$, *we have*

1. $K_s^{-1/2}\widetilde{F}^{(s)\prime}\widetilde{\omega}^{(s)}=O_p\left(1\right)$
2. $K_s^{-1/2}\widetilde{F}^{(s)\prime}\widetilde{\eta}^{(s)}=O_p\left(1\right)$
3. $K_s^{-1/2}\widetilde{\varepsilon}^{(s)\prime}\widetilde{\eta}^{(s)}=O_p\left(1\right)$ *(a single element)*
4. $N^{-1/2}\varepsilon_r'J_N\Phi_t=O_p\left(1\right)$
5. $N^{-1}K_s^{-1}\Phi_r'J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{F}^{(s)}=O_p\left(\delta_{NK_s}^{-1}\right)$
6. $N^{-1}K_s^{-1/2}\Phi_r'J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{\omega}^{(s)}=O_p\left(1\right)$
7. $N^{-1/2}K_s^{-1/2}\Phi_rJ_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{\eta}^{(s)}=O_p\left(K_s^{-1/2}\right)=O_p\left(1\right)$
8. $N^{-1}K_s^{-3/2}\widetilde{F}^{(s)\prime}\widetilde{\varepsilon}^{(s)}J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{F}^{(s)}=O_p\left(\delta_{N,K_s}^{-1}\right)$
9. $N^{-1}K_s^{-3/2}\widetilde{\omega}^{(s)\prime}\widetilde{\varepsilon}^{(s)}J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{F}^{(s)}=O_p\left(\delta_{N,K_s}^{-1}\right)$
10. $N^{-1}K_s^{-3/2}\widetilde{\omega}^{(s)\prime}\widetilde{\varepsilon}^{(s)}J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{\omega}^{(s)}=O_p\left(\delta_{N,K_s}^{-1}\right)$
11. $N^{-1}K_s^{-1/2}\widetilde{F}^{(s)\prime}\widetilde{\varepsilon}^{(s)}J_N\varepsilon_t=O_p\left(\delta_{NK_s}^{-1}\right)$
12. $N^{-1}K_s^{-1/2}\widetilde{\omega}^{(s)\prime}\widetilde{\varepsilon}^{(s)}J_N\varepsilon_t=O_p\left(\delta_{NK_s}^{-1}\right)$
13. $N^{-1}K_s^{-3/2}\widetilde{\eta}^{(s)\prime}\widetilde{\varepsilon}^{(s)}J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{F}^{(s)}=O_p\left(\delta_{NK_s}^{-1}\right)$
14. $N^{-1}K_s^{-3/2}\widetilde{\eta}^{(s)\prime}\widetilde{\varepsilon}^{(s)}J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{\omega}^{(s)}=O_p\left(\delta_{NK_s}^{-1}\right)$

7.4.2 Proof of Lemma A2

1. It is $K_s^{-1/2}F^{(s)\prime}\omega^{(s)} = K_s^{-1/2}\sum_{t=1}^{T}F_t^{(s)}\omega_t^{(s)\prime} = K_s^{-1/2}\sum_{t=1}^{T}k_{ts}F_t\omega_t'$.
Then, $K_s^{-1/2}\widetilde{F}^{(s)\prime}\widetilde{\omega}^{(s)} = K_s^{-1/2}\sum_{t=1}^{T}F_t^{(s)}\omega_t^{(s)\prime} - \left(K_s^{-1}\sum_{t=1}^{T}k_{t,s}F_t\right)\left(K_s^{-1/2}\sum_{t=1}^{T}k_{t,s}\omega_t'\right)$
$= O_p(1)$, by Assumptions 1.1 and 1.4 and 2.3.

2. $K_s^{-1/2}\widetilde{F}^{(s)\prime}\widetilde{\eta}^{(s)} = K_s^{-1/2}\sum_{s=1}^{T}F_t^{(s)}\eta_t^{(s)\prime} - \left(K_s^{-1}\sum_{s=1}^{T}k_{t,s}F_s\right)\left(K_s^{-1/2}\sum_{s=1}^{T}k_{t,s}\eta_{s+1}\right)$
$=O_p(1)$, by Assumption 2.1, Theorem 3 and Lemma A1.4.

3. $K_s^{-1/2}\widetilde{\varepsilon}^{(s)\prime}\widetilde{\eta}^{(s)} = K_s^{-1/2}\sum_{t=1}^{T}k_{t,s}\varepsilon_t\eta_{t+1} - \left(K_s^{-1}\sum_{t=1}^{T}k_{t,s}\varepsilon_t\right)\left(K_s^{-1/2}\sum_{t=1}^{T}k_{t,s}\eta_{t+1}\right)$
$= O_p(1)$, by Lemmas A1.4, A1.5 and Assumption 1.3.

4. The m-th element of $N^{-1/2}\varepsilon_r' J_N \Phi_t$ is $N^{-1/2}\sum_{i=1}^{N}\varepsilon_{ir}\phi_{mi,t} - \left(N^{-1/2}\sum_{i=1}^{N}\varepsilon_{it}\right)\left(N^{-1}\sum_{i=1}^{N}\phi_{mi,t}\right)$
$= O_p(1)$ by Assumption 1.2, Lemma A1.3 and Theorem 3.

5. We have $N^{-1}K_s^{-1}\Phi_r' J_N \widetilde{\varepsilon}^{(s)\prime}\widetilde{F}^{(s)} = N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{ir}\varepsilon_{it}^{(s)}F_t^{(s)\prime}$
$-N^{-1}K_s^{-2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}F_l^{(s)\prime}$
$-N^{-2}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}\varepsilon_{jt}^{(s)}F_t^{(s)\prime}$
$+N^{-2}K_s^{-2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}F_l^{(s)\prime}$.

The m_1,m_2 element of the above is
$-N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{m_1 ir}\varepsilon_{it}^{(s)}F_{m_2 t}^{(s)} = O_p\left(K_s^{-1/2}\right)$ by Lemma A1.7.

$-N^{-2}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{m_1 ir}\varepsilon_{jt}^{(s)}F_{m_2 t}^{(s)} =$
$= \left(N^{-1}\sum_{i=1}^{N}\phi_{m_1 ir}\right)\left(N^{-1}\sum_{j=1}^{N}K_s^{-1}\sum_{t=1}^{T}\varepsilon_{jt}^{(s)}F_{m_2 t}^{(s)}\right)$
$= O_p(1)\times O_p\left(K_s^{-1/2}\right) = O_p\left(K_s^{-1/2}\right)$, by Assumption 1.2 and Theorem 3.

$-N^{-1}K_s^{-2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\phi_{m_1 ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}F_{m_2 l}^{(s)\prime} =$
$= \left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{m_2 l}^{(s)\prime}\right)\left(N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{m_1 ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}\right)$
but $K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{m_2 l}^{(s)\prime} = O_p(1)$ by Assumption 1.1
and $N^{-1/2}K_s^{-1}\sum_{t=1}^{T}\left(N^{-1/2}\sum_{i=1}^{N}\phi_{m_1 ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}\right) =$
$= N^{-1/2}K_s^{-1}\sum_{t=1}^{T}k_{s,t}\left(N^{-1/2}\sum_{i=1}^{N}\phi_{m_1 ir}\varepsilon_{it}\right) = O_p\left(N^{-1/2}\right)$ by Theorem 3.
So, overall, $N^{-1}K_s^{-2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}F_l^{(s)\prime} = O_p\left(N^{-1/2}\right)$.

$-N^{-2}K_s^{-2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}F_l^{(s)\prime}$
$= N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_l^{(s)\prime}\right)\left(N^{-1}\sum_{i=1}^{N}\phi_{ir}\right) = O_p\left(N^{-1/2}K_s^{-1/2}\right)$

by Assumptions 1.1, 1.2 and Lemma A1.3.

Summing up the various terms, we have that A2.5 is $O_p\left(\delta_{N,K_s}^{-1}\right)$.

6. $\left(N^{-1}K_s^{-1/2}\Phi_r' J_N \widetilde{\varepsilon}^{(s)\prime}\widetilde{\omega}^{(s)}\right) = N^{-1}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{ir}\varepsilon_{it}^{(s)}\omega_t^{(s)\prime}$
$-N^{-1}K_s^{-3/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}\omega_l^{(s)\prime}$
$-N^{-2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}\varepsilon_{jt}^{(s)}\omega_t^{(s)\prime}$
$+N^{-2}K_s^{-3/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}\omega_l^{(s)\prime}$.

Notice that
-$N^{-1}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{ir}\varepsilon_{it}^{(s)}\omega_t^{(s)\prime} = O_p(1)$, by Lemma A1.8.

-$N^{-2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}\varepsilon_{jt}^{(s)}\omega_t^{(s)\prime} = \left(N^{-1}\sum_{i=1}^{N}\phi_{ir}\right)\left(N^{-1}K_s^{-1/2}\sum_{t=1}^{T}\sum_{j=1}^{N}\varepsilon_{jt}^{(s)}\omega_t^{(s)\prime}\right)$
$= O_p(1)$ by Assumptions 1.2 and 2.4.

-$N^{-1}K_s^{-3/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}\omega_l^{(s)\prime}$
$= \left(K_s^{-1/2}\sum_{l=1}^{T}k_{s,l}^{1/2}\omega_l^{(s)}\right)\left(N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}\right)$
$= O_p\left(N^{-1/2}\right) = O_p(1)$ by Assumptions 1.4 and Theorem 3.

-$N^{-2}K_s^{-3/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}\omega_l^{(s)\prime}$
$= \left(N^{-1}\sum_{i=1}^{N}\phi_{ir}\right)\left(K_s^{-1/2}\sum_{l=1}^{T}k_{s,l}^{1/2}\omega_{m_1 l}^{(s)\prime}\right)\left(N^{-1}K_s^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}\right)$
$= O_p(1)\times O_p(1)\times O_p\left(N^{-1/2}K_s^{-1/2}\right) = O_p\left(N^{-1/2}K_s^{-1/2}\right)$
by Assumptions 1.4, 1.2 and Lemma A1.3.

Then, overall, $N^{-1}K_t^{-1/2}\Phi_t' J_N \widetilde{\varepsilon}^{(s)}\widetilde{\omega}^{(s)} = O_p(1)$.

7. $\left(N^{-1/2}K_s^{-1/2}\Phi_r J_N \widetilde{\varepsilon}^{(s)\prime}\widetilde{\eta}^{(s)}\right) = N^{-1/2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{ir}\varepsilon_{it}^{(s)\prime}\eta_{t+1}^{(s)}$
$-N^{-1/2}K_s^{-3/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)\prime}k_{s,l}^{1/2}\eta_{l+1}^{(s)}$
$-N^{-3/2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}\varepsilon_{jt}^{(s)\prime}\eta_{t+1}^{(s)}$
$+N^{-3/2}K_s^{-3/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)\prime}k_{s,l}^{1/2}\eta_{l+1}^{(s)}$.

Notice that
-$N^{-1/2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{ir}\varepsilon_{it}^{(s)\prime}\eta_{t+1}^{(s)} = O_p(1)$, by Lemma A1.9.
-$N^{-1/2}K_s^{-3/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)\prime}k_{s,l}^{1/2}\eta_{l+1}^{(s)} =$
$= \left(K_s^{-1/2}\sum_{l=1}^{T}k_{s,l}^{1/2}\eta_{l+1}^{(s)}\right)N^{-1/2}K_s^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi_{ir}k_{s,t}^{1/2}\varepsilon_{it}^{(s)\prime} = O_p(1)$
by Theorem 3 and Lemma A1.4.
-$N^{-3/2}K_s^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi_{ir}\varepsilon_{jt}^{(s)\prime}\eta_{t+1}^{(s)}$

$= N^{-1} \sum_{i=1}^{N} \phi_{ir} \left(N^{-1/2} K_s^{-1/2} \sum_{t=1}^{T} \sum_{j=1}^{N} \varepsilon_{jt}^{(s)\prime} \eta_{t+1}^{(s)} \right) = O_p(1)$
by Assumptions 1.2 and Lemma A1.6.
$-N^{-3/2} K_s^{-3/2} \sum_{t=1}^{T} \sum_{l=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi_{ir} k_{s,t}^{1/2} \varepsilon_{jt}^{(s)\prime} k_{s,l}^{1/2} \eta_{l+1}^{(s)} =$
$\left(N^{-1} \sum_{i=1}^{N} \phi_{ir} \right) \left(K_s^{-1/2} \sum_{l=1}^{T} k_{s,l}^{1/2} \eta_{l+1}^{(s)} \right) K_s^{-1/2} \left(N^{-1/2} K_s^{-1/2} \sum_{t=1}^{T} \sum_{j=1}^{N} k_{s,t}^{1/2} \varepsilon_{jt}^{(s)\prime} \right)$
$= O_p\left(K_s^{-1/2}\right)$
by Assumptions 1.2, Lemma A1.3 and A1.4.
So, overall, $N^{-1/2} K_s^{-1/2} \Phi_r J_N \widetilde{\varepsilon}^{(s)\prime} \widetilde{\eta}^{(s)}$ is $O_p(1)$.

8. For the m_1, m_2 generic element $N^{-1} K_s^{-3/2} \widetilde{F}^{(s)} \widetilde{\varepsilon}^{(s)} J_N \widetilde{\varepsilon}^{(s)\prime} \widetilde{F}^{(s)}$ we have
$= N^{-1} K_s^{-3/2} \sum_{t=1}^{T} \sum_{u}^{T} \sum_{i=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} \varepsilon_{it}^{(s)} F_{m_2,t}^{(s)}$
$-N^{-1} K_s^{-5/2} \sum_{t=1}^{T} \sum_{l=1}^{T} \sum_{u}^{T} \sum_{i=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} k_{s,t}^{1/2} \varepsilon_{it}^{(s)} k_{s,l}^{1/2} F_{m_2,l}^{(s)}$
$-N^{-2} K_s^{-3/2} \sum_{t=1}^{T} \sum_{u}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} \varepsilon_{jt}^{(s)} F_{m_2,t}^{(s)}$
$+N^{-2} K_s^{-5/2} \sum_{t=1}^{T} \sum_{l=1}^{T} \sum_{u}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} k_{s,t}^{1/2} \varepsilon_{jt}^{(s)} k_{s,l}^{1/2} F_{m_2,l}^{(s)}$
$-N^{-1} K_s^{-5/2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{u=1}^{T} \sum_{v=1}^{T} k_{s,u}^{1/2} \varepsilon_{iu}^{(s)} k_{s,v}^{1/2} F_{m_1,v}^{(s)} \varepsilon_{it}^{(s)} F_{m_2,t}^{(s)}$
$+N^{-1} K_s^{-7/2} \sum_{t=1}^{T} \sum_{l=1}^{T} \sum_{i=1}^{N} \sum_{u=1}^{T} \sum_{v=1}^{T} k_{s,u}^{1/2} \varepsilon_{iu}^{(s)} k_{s,v}^{1/2} F_{m_1,v}^{(s)} k_{s,t}^{1/2} \varepsilon_{it}^{(s)} k_{s,l}^{1/2} F_{m_2,l}^{(s)}$
$+N^{-2} K_s^{-5/2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{u=1}^{T} \sum_{v=1}^{T} k_{s,u}^{1/2} \varepsilon_{iu}^{(s)} k_{s,v}^{1/2} F_{m_1,v}^{(s)} \varepsilon_{jt}^{(s)} F_{m_2,t}^{(s)}$
$-N^{-2} K_s^{-7/2} \sum_{t=1}^{T} \sum_{l=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{u=1}^{T} \sum_{v=1}^{T} k_{s,u}^{1/2} \varepsilon_{iu}^{(s)} k_{s,v}^{1/2} F_{m_1,v}^{(s)} k_{s,t}^{1/2} \varepsilon_{jt}^{(s)} k_{s,l}^{1/2} F_{m_2,l}^{(s)}$

Then, 8.(i),
$N^{-1} K_s^{-3/2} \sum_{t=1}^{T} \sum_{u}^{T} \sum_{i=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} \varepsilon_{it}^{(s)} F_{m_2,t}^{(s)} =$
$= K_s^{-1/2} N^{-1} \sum_{i=1}^{N} \left(K_s^{-1/2} \sum_{u}^{T} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} \right) \left(K_s^{-1/2} \sum_{t=1}^{T} \varepsilon_{it}^{(s)} F_{m_2,t}^{(s)} \right) = O_p\left(K_s^{-1/2}\right)$ by Theorem 3.

8.(ii)
$N^{-1} K_s^{-5/2} \sum_{t=1}^{T} \sum_{l=1}^{T} \sum_{u}^{T} \sum_{i=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} k_{s,t}^{1/2} \varepsilon_{it}^{(s)} k_{s,l}^{1/2} F_{m_2,l}^{(s)} =$
$= \left(K_s^{-1} \sum_{l=1}^{T} k_{s,l}^{1/2} F_{m_2,l}^{(s)} \right) N^{-1} K_s^{-3/2} \sum_{t=1}^{T} \sum_{u}^{T} \sum_{i=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} k_{s,t}^{1/2} \varepsilon_{it}^{(s)}$
$= \left(K_s^{-1} \sum_{l=1}^{T} k_{s,l}^{1/2} F_{m_2,l}^{(s)} \right) K_s^{-1} \sum_{t=1}^{T} \left(N^{-1} K_s^{-1/2} \sum_{u}^{T} \sum_{i=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} k_{s,t}^{1/2} \varepsilon_{it}^{(s)} \right)$
$\leq \left(K_s^{-1} \sum_{l=1}^{T} k_{s,l}^{1/2} F_{m_2,l}^{(s)} \right) K_s^{-1} \sum_{t=1}^{T} k_{st} \left(N^{-1} K_s^{-1/2} \sum_{u}^{T} \sum_{i=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} \varepsilon_{it} \right)$
$= O_p\left(\delta_{N,T}^{-1}\right)$
by Assumption 1.1, Theorem 3 and Lemma A1.12.

8.(iii)
$N^{-2} K_s^{-3/2} \sum_{t=1}^{T} \sum_{u}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} \varepsilon_{jt}^{(s)} F_{m_2,t}^{(s)}$
$= K_s^{-1/2} \left(N^{-1} K_s^{-1/2} \sum_{u}^{T} \sum_{i=1}^{N} \varepsilon_{iu}^{(s)} F_{m_1,u}^{(s)} \right) \left(N^{-1} K_s^{-1/2} \sum_{t=1}^{T} \sum_{j=1}^{N} \varepsilon_{jt}^{(s)} F_{m_2,t}^{(s)} \right)$
$= O_p\left(K_s^{-1/2}\right)$

by Theorem 3.

8.(iv)
$N^{-2}K_s^{-5/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{iu}^{(s)}F_{m_1,u}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}F_{m_2,l}^{(s)}$
$=\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{m_2,l}^{(s)}\right)\left(N^{-1}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u}^{T}\varepsilon_{iu}^{(s)}F_{m_1,u}^{(s)}\right)\left(N^{-1}K_s^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}\right)$
$=O_p(1)\,O_p(1)\,O_p\left(K_s^{-1/2}N^{-1/2}\right)=O_p\left(K_s^{-1/2}N^{-1/2}\right)$
by Assumption 1.1, Theorem 3 and Lemma A1.3.

8.(v)
$N^{-1}K_s^{-5/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}F_{m_1,v}^{(s)}\varepsilon_{it}^{(s)}F_{m_2,t}^{(s)}$
$=\left(K_s^{-1}\sum_{v=1}^{T}k_{s,v}^{1/2}F_{m_1,v}^{(s)}\right)\left(K_s^{-1}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}\right)\left(N^{-1}K_s^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^{(s)}F_{m_2,t}^{(s)}\right)$
$=O_p\left(K_s^{-1/2}\right)$
by Assumption 1.1, Theorem 3 and Lemma A1.3.

8.(vi)
$N^{-1}K_s^{-7/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}F_{m_1,v}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}F_{m_2,l}^{(s)}=$
$=\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{m_2,l}^{(s)}\right)\left(K_s^{-1}\sum_{v=1}^{T}k_{s,v}^{1/2}F_{m_1,v}^{(s)}\right)$
$K_s^{-1}\sum_{t=1}^{T}K_s^{-1/2}N^{-1}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}$
$\leq\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{m_2,l}^{(s)}\right)\left(K_s^{-1}\sum_{v=1}^{T}k_{s,v}^{1/2}F_{m_1,v}^{(s)}\right)$
$K_s^{-1}\sum_{t=1}^{T}k_{st}\left(K_s^{-1/2}N^{-1}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{su}\varepsilon_{iu}\varepsilon_{it}\right)$
$=O_p\left(\delta_{N,K_S}^{-1}\right)$
by Assumption 1.1 and Lemma A1.10.

8.(vii)
$N^{-2}K_s^{-5/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}F_{m_1,v}^{(s)}\varepsilon_{jt}^{(s)}F_{m_2,t}^{(s)}$
$=O_p\left(N^{-1/2}K_s^{-1/2}\right)$
see 8(iv).

8.(viii)
$N^{-2}K_s^{-7/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}F_{m_1,v}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}F_{m_2,l}^{(s)}$
$=\left(K_s^{-1}\sum_{v=1}^{T}k_{s,v}^{1/2}F_{m_1,v}^{(s)}\right)\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{m_2,l}^{(s)}\right)$
$K_s^{-1/2}N^{-1}\left(N^{-1/2}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}\right)\left(N^{-1/2}K_s^{-1/2}\sum_{j=1}^{N}\sum_{t=1}^{T}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}\right)$
$=O_p\left(K_s^{-1/2}N^{-1}\right)$
by Assumption 1.1 and Lemma A1.3.
So, overall, summing up these terms we have that the item in 8 is $O_p\left(\delta_{N,K_s}^{-1}\right)$.

9. The proof is as in 8, replacing $\omega^{(s)}$ for $F^{(s)}$, Assumption 1.1 for Theorem 3, and Lemma A1.12 for A1.13.

10. For the m_1, m_2 generic element of $N^{-1}K_s^{-3/2}\widetilde{\omega}^{(s)\prime}\widetilde{\varepsilon}^{(s)}J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{\omega}^{(s)}$ we have

$$N^{-1}K_s^{-3/2}\sum_{t=1}^T\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}\varepsilon_{it}^{(s)}\omega_{m_2t}^{(s)}$$
$$-N^{-1}K_s^{-5/2}\sum_{t=1}^T\sum_{l=1}^T\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}\omega_{m_2l}^{(s)}$$
$$-N^{-2}K_s^{-3/2}\sum_{t=1}^T\sum_u^T\sum_{i=1}^N\sum_{j=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}\varepsilon_{jt}^{(s)}\omega_{m_2t}^{(s)}$$
$$+N^{-2}K_s^{-5/2}\sum_{t=1}^T\sum_{l=1}^T\sum_u^T\sum_{i=1}^N\sum_{j=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}\omega_{m_2l}^{(s)}$$
$$-N^{-1}K_s^{-5/2}\sum_{t=1}^T\sum_{i=1}^N\sum_{u=1}^T\sum_{v=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\omega_{m_1v}^{(s)}\varepsilon_{it}^{(s)}\omega_{m_2t}^{(s)}$$
$$+N^{-1}K_s^{-7/2}\sum_{t=1}^T\sum_{l=1}^T\sum_{i=1}^N\sum_{u=1}^T\sum_{v=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\omega_{m_1v}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}\omega_{m_2l}^{(s)}$$
$$+N^{-2}K_s^{-5/2}\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{u=1}^T\sum_{v=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\omega_{m_1v}^{(s)}\varepsilon_{jt}^{(s)}\omega_{m_2t}^{(s)}$$
$$-N^{-2}K_s^{-7/2}\sum_{t=1}^T\sum_{l=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{u=1}^T\sum_{v=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\omega_{m_1v}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}\omega_{m_2l}^{(s)}$$

Focusing on each element of the summation we have

10.(i),

$$N^{-1}K_s^{-3/2}\sum_{t=1}^T\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}\varepsilon_{it}^{(s)}\omega_{m_2t}^{(s)} =$$
$$= K_s^{-1/2}N^{-1}\sum_{i=1}^N\left(K_s^{-1/2}\sum_u^T\varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}\right)\left(K_s^{-1/2}\sum_{t=1}^T\varepsilon_{it}^{(s)}\omega_{m_2t}^{(s)}\right) = O_p\left(K_s^{-1/2}\right)$$

by Assumption 2.4.

10.(ii)

$$N^{-1}K_s^{-5/2}\sum_{t=1}^T\sum_{l=1}^T\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}\omega_{m_2l}^{(s)}=$$
$$=\left(K_s^{-1}\sum_{l=1}^T k_{s,l}^{1/2}\omega_{m_2l}^{(s)}\right)N^{-1}K_s^{-3/2}\sum_{t=1}^T\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}$$
$$=\left(K_s^{-1}\sum_{l=1}^T k_{s,l}^{1/2}\omega_{m_2l}^{(s)}\right)K_s^{-1}\sum_{t=1}^T\left(N^{-1}K_s^{-1/2}\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}\right)$$
$$\leq K_s^{-1/2}\left(K_s^{-1/2}\sum_{l=1}^T k_{s,l}^{1/2}\omega_{m_2l}^{(s)}\right)K_s^{-1}\sum_{t=1}^T k_{st}\left(N^{-1}K_s^{-1/2}\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}\varepsilon_{it}\right)$$
$$=O_p\left(K_s^{-1/2}\right)O_p\left(\delta_{N,T}^{-1}\right)=O_p\left(\delta_{N,T}^{-1}\right)$$

by Assumption 1.4, and Lemma A1.13.

10.(iii)

$$N^{-2}K_s^{-3/2}\sum_{t=1}^T\sum_u^T\sum_{i=1}^N\sum_{j=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}\varepsilon_{jt}^{(s)}\omega_{m_2t}^{(s)}$$
$$=K_s^{-1/2}\left(N^{-1}K_s^{-1/2}\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}\right)\left(N^{-1}K_s^{-1/2}\sum_{t=1}^T\sum_{j=1}^N \varepsilon_{jt}^{(s)}\omega_{m_2t}^{(s)}\right)$$
$$=O_p\left(K_s^{-1/2}\right)$$

by Assumption 2.4.

10.(iv)

$$N^{-2}K_s^{-5/2}\sum_{t=1}^T\sum_{l=1}^T\sum_u^T\sum_{i=1}^N\sum_{j=1}^N \varepsilon_{iu}^{(s)}\omega_{m_1u}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}\omega_{m_2l}^{(s)}$$

$$= K_s^{-1/2}\left(K_s^{-1/2}\sum_{l=1}^{T}k_{s,l}^{1/2}\omega_{m_2 l}^{(s)}\right)\left(N^{-1}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u}^{T}\varepsilon_{iu}^{(s)}\omega_{m_1 u}^{(s)}\right)\left(N^{-1}K_s^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}\right)$$
$$= O_p\left(K_s^{-1/2}\right)O_p\left(1\right)O_p\left(K_s^{-1/2}N^{-1/2}\right) = O_p\left(K_s^{-1/2}N^{-1/2}\right)$$
by Assumption 1.4, 2.4 and Lemma A1.3.

10.(v)
is $O_p\left(\delta_{N,T}^{-1}\right)$ as in 10(ii).

10.(vi)
$$N^{-1}K_s^{-7/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\omega_{m_1 v}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}\omega_{m_2 l}^{(s)} =$$
$$=\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}\omega_{m_2 l}^{(s)}\right)\left(K_s^{-1}\sum_{v=1}^{T}k_{s,v}^{1/2}\omega_{m_1 v}^{(s)}\right)$$
$$K_s^{-1}\sum_{t=1}^{T}K_s^{-1/2}N^{-1}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}$$
$$\leq\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}\omega_{m_2 l}^{(s)}\right)\left(K_s^{-1}\sum_{v=1}^{T}k_{s,v}^{1/2}\omega_{m_1 v}^{(s)}\right)$$
$$K_s^{-1}\sum_{t=1}^{T}k_{st}\left(K_s^{-1/2}N^{-1}\sum_{i=1}^{N}\sum_{u=1}^{T}\varepsilon_{iu}^{(s)}\varepsilon_{it}^{(s)}\right)$$
$$= O_p\left(K_s^{-1}\right)O_p\left(\delta_{N,K_S}^{-1}\right) = O_p\left(\delta_{N,K_S}^{-1}\right)$$
by Assumption 1.4 and Lemma A1.10.

10.(vii)
is $O_p\left(N^{-1/2}K_s^{-1/2}\right)$ as in item (iv).

10.(viii)
$$N^{-2}K_s^{-7/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\omega_{m_1 v}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}\omega_{m_2 l}^{(s)}$$
$$= K_s^{-1/2}\left(K_s^{-1/2}\sum_{v=1}^{T}k_{s,v}^{1/2}\omega_{m_1 v}^{(s)}\right)K_s^{-1/2}\left(K_s^{-1/2}\sum_{l=1}^{T}k_{s,l}^{1/2}\omega_{m_2 l}^{(s)}\right)$$
$$K_s^{-1/2}N^{-1}\left(N^{-1/2}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}\right)\left(N^{-1/2}K_s^{-1/2}\sum_{j=1}^{N}\sum_{t=1}^{T}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}\right)$$
$$= O_p\left(K_s^{-1}\right)O_p\left(K_s^{-1/2}N^{-1}\right) = O_p\left(K_s^{-1/2}N^{-1}\right)$$
by Assumption 1.4 and Lemma A1.3.
So, overall, summing up these terms we have that the item in 10 is $O_p\left(\delta_{N,K_s}^{-1}\right)$.

11. For the m-th element we have the following expression
$$K_s^{-1/2}N^{-1}\sum_{u=1}^{T}\sum_{i=1}^{N}F_{mu}^{(s)}\varepsilon_{iu}^{(s)}\varepsilon_{it}$$
$$-K_s^{-1/2}N^{-2}\sum_{u=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}F_{mu}^{(s)}\varepsilon_{iu}^{(s)}\varepsilon_{jt}$$
$$-K_s^{-3/2}N^{-1}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{l=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)\prime}k_{s,l}^{1/2}F_{ml}^{(s)}\varepsilon_{it}$$
$$+K_s^{-3/2}N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{l=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,l}^{1/2}F_{ml}^{(s)}\varepsilon_{jt}$$

We then have
$-K_s^{-1/2}N^{-1}\sum_{u=1}^{T}\sum_{i=1}^{N}F_{mu}^{(s)}\varepsilon_{iu}^{(s)}\varepsilon_{it} = O_p\left(\delta_{N,K_s}^{-1}\right)$ by Lemma A1.12.

$-K_s^{-1/2}N^{-2}\sum_{u=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}F_{mu}^{(s)}\varepsilon_{iu}^{(s)}\varepsilon_{jt}$
$= N^{-1/2}\left(N^{-1/2}\sum_{j=1}^{N}\varepsilon_{jt}\right)\left(N^{-1}K_s^{-1/2}\sum_{u=1}^{T}\sum_{i=1}^{N}F_{mu}^{(s)}\varepsilon_{iu}^{(s)}\right)$
$=O_p\left(N^{-1/2}\right)$
by Theorem 3 and Lemma A1.3.

$-K_s^{-3/2}N^{-1}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{l=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)\prime}k_{s,l}^{1/2}F_{ml}^{(s)}\varepsilon_{it}$
$=\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{ml}^{(s)}\right)K_s^{-1/2}N^{-1}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)\prime}\varepsilon_{it}$
$\leq\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{ml}^{(s)}\right)\left(K_s^{-1/2}N^{-1}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)\prime}\varepsilon_{it}\right)$
$=O_p\left(1\right)O_p\left(\delta_{N,K_s}^{-1}\right)=O_p\left(\delta_{N,K_s}^{-1}\right)$
by Assumption 1.1 and Lemma A1.10.

$-K_s^{-3/2}N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{l=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,l}^{1/2}F_{ml}^{(s)}\varepsilon_{jt}$
$=\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{ml}^{(s)}\right)\left(N^{-1/2}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}\right)\left(N^{-1/2}\sum_{j=1}^{N}\varepsilon_{jt}\right)N^{-1/2}$
$=O_p\left(N^{-1/2}\right)$
by Assumption 1.1 and Lemma A1.3.

So, overall, item 11 is $O_p\left(\delta_{N,K_s}^{-1}\right)$

12. The proof is as in 11, replacing $\omega^{(s)}$ for $F^{(s)}$ and Assumption 2.4 for Theorem 3.

13. For the m_2 generic element of $N^{-1}K_s^{-3/2}\widetilde{\eta}^{(s)}\widetilde{\varepsilon}^{(s)}J_N\widetilde{\varepsilon}^{(s)\prime}\widetilde{F}^{(s)}$ we have
$N^{-1}K_s^{-3/2}\sum_{t=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}\varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}\varepsilon_{it}^{(s)}F_{m_2t}^{(s)}$
$-N^{-1}K_s^{-5/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}\varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}F_{m_2l}^{(s)}$
$-N^{-2}K_s^{-3/2}\sum_{t=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}\varepsilon_{jt}^{(s)}F_{m_2t}^{(s)}$
$+N^{-2}K_s^{-5/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}F_{m_2l}^{(s)}$
$-N^{-1}K_s^{-5/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\eta_{v+1}^{(s)}\varepsilon_{it}^{(s)}F_{m_2t}^{(s)}$
$+N^{-1}K_s^{-7/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\eta_{v+1}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}F_{m_2l}^{(s)}$
$+N^{-2}K_s^{-5/2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\eta_{v+1}^{(s)}\varepsilon_{jt}^{(s)}F_{m_2t}^{(s)}$
$-N^{-2}K_s^{-7/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\eta_{v+1}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}F_{m_2l}^{(s)}$

Then, 13.(i),
$N^{-1}K_s^{-3/2}\sum_{t=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}\varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}\varepsilon_{it}^{(s)}F_{m_2t}^{(s)}=$
$=O_p\left(K_s^{-1/2}\right)$ by Lemma A1.14.

13.(ii)
$N^{-1}K_s^{-5/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{u}^{T}\sum_{i=1}^{N}\varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}F_{m_2l}^{(s)}=$

$$=\left(K_s^{-1}\sum_{l=1}^T k_{s,l}^{1/2}F_{m_2l}^{(s)}\right)N^{-1}K_s^{-3/2}\sum_{t=1}^T\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}$$
$=O_p\left(\delta_{N,K_s}^{-1}\right)$ by Assumption 1.1, and Lemma A1.11.

13.(iii)
$$N^{-2}K_s^{-3/2}\sum_{t=1}^T\sum_u^T\sum_{i=1}^N\sum_{j=1}^N \varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}\varepsilon_{jt}^{(s)}F_{m_2t}^{(s)}$$
$$=K_s^{-1/2}N^{-1/2}\left(N^{-1/2}K_s^{-1/2}\sum_u^T\sum_{i=1}^N \varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}\right)\left(N^{-1}K_s^{-1/2}\sum_{t=1}^T\sum_{j=1}^N \varepsilon_{jt}^{(s)}F_{m_2t}^{(s)}\right)$$
$$=O_p\left(K_s^{-1/2}N^{-1/2}\right)$$
by Theorem 3 and Lemma A1.6.

13.(iv)
$$N^{-2}K_s^{-5/2}\sum_{t=1}^T\sum_{l=1}^T\sum_u^T\sum_{i=1}^N\sum_{j=1}^N \varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}F_{m_2l}^{(s)}$$
$$=\left(K_s^{-1}\sum_{l=1}^T k_{s,l}^{1/2}F_{m_2l}^{(s)}\right)\left(N^{-1/2}K_s^{-1/2}\sum_{i=1}^N\sum_u^T \varepsilon_{iu}^{(s)}\eta_{u+1}^{(s)}\right)N^{-1/2}\left(N^{-1}K_s^{-1}\sum_{j=1}^N\sum_{t=1}^T k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}\right)$$
$$=O_p(1)\,O_p(1)\,O_p\left(K_s^{-1/2}N^{-1/2}\right)=O_p\left(K_s^{-1/2}N^{-1}\right)$$
by Assumption 1.1, and Lemma A1.3, A1.6.

13.(v)
$$N^{-1}K_s^{-5/2}\sum_{t=1}^T\sum_{i=1}^N\sum_{u=1}^T\sum_{v=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\eta_{v+1}^{(s)}\varepsilon_{it}^{(s)}F_{m_2t}^{(s)}$$
$$=\left(K_s^{-1}\sum_{v=1}^T k_{s,v}^{1/2}\eta_{v+1}^{(s)}(m_1)\right)\left(K_s^{-1}\sum_{u=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}\right)\left(N^{-1}K_s^{-1/2}\sum_{i=1}^N\sum_{t=1}^T \varepsilon_{it}^{(s)}F_{m_2t}^{(s)}\right)$$
$$=K_s^{-1/2}\left(K_s^{-1/2}\sum_{v=1}^T k_{s,v}^{1/2}\eta_{v+1}^{(s)}(m_1)\right)\left(\max_s(k_{s,u})\,K_s^{-1}\sum_{u=1}^T N^{-1}K_s^{-1/2}\sum_{i=1}^N\sum_{t=1}^T \varepsilon_{it}^{(s)}F_{m_2t}^{(s)}\varepsilon_{iu}\right)$$
$$=O_p\left(K_s^{-1/2}\right)O_p\left(\delta_{N,K_s}^{-1}\right)=O_p\left(\delta_{N,K_s}^{-1}\right)$$
by Theorem 3 and Lemma A1.12 and A1.4.

13.(vi)
$$N^{-1}K_s^{-7/2}\sum_{t=1}^T\sum_{l=1}^T\sum_{i=1}^N\sum_{u=1}^T\sum_{v=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\eta_v^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}k_{s,l}^{1/2}F_{m_2l}^{(s)}=$$
$$=\left(K_s^{-1}\sum_{l=1}^T k_{s,l}^{1/2}F_{m_2l}^{(s)}\right)\left(K_s^{-1}\sum_{v=1}^T k_{s,v}^{1/2}\eta_v^{(s)}\right)$$
$$K_s^{-1}\sum_{t=1}^T K_s^{-1/2}N^{-1}\sum_{i=1}^N\sum_{u=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,t}^{1/2}\varepsilon_{it}^{(s)}$$
$$\leq\left(K_s^{-1}\sum_{l=1}^T k_{s,l}^{1/2}F_{m_2l}^{(s)}\right)K_s^{-1/2}\left(K_s^{-1/2}\sum_{v=1}^T k_{s,v}^{1/2}\eta_v^{(s)}\right)$$
$$K_s^{-1}k_{st}\sum_{t=1}^T\left(K_s^{-1/2}N^{-1}\sum_{i=1}^N\sum_{u=1}^T \varepsilon_{iu}^{(s)}\varepsilon_{it}^{(s)}\right)$$
$$=O_p\left(K_s^{-1/2}\right)O_p\left(\delta_{N,K_S}^{-1}\right)=O_p\left(\delta_{N,K_S}^{-1}\right)$$
by Assumption 1.1 and Lemma A1.10 and A1.4.

13.(vii)
$$N^{-2}K_s^{-5/2}\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{u=1}^T\sum_{v=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\eta_v^{(s)}\varepsilon_{jt}^{(s)}F_{m_2t}^{(s)}$$
$$=\left(N^{-1/2}K_s^{-1/2}\sum_{i=1}^N\sum_{u=1}^T k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}\right)\left(K_s^{-1/2}\sum_{v=1}^T k_{s,v}^{1/2}\eta_v^{(s)}\right)K_s^{-1}N^{-1/2}\left(N^{-1}K_s^{-1/2}\sum_{j=1}^N\sum_{t=1}^T \varepsilon_{jt}^{(s)}F_{m_2t}^{(s)}\right)$$

$= O_p\left(N^{-1/2}K_s^{-1}\right)$
by Theorem 3 and Lemma A1.3 and A1.4.

13.(viii)
$N^{-2}K_s^{-7/2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{v=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}k_{s,v}^{1/2}\eta_{v+1}^{(s)}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}k_{s,l}^{1/2}F_{m_2l}^{(s)}$
$= K_s^{-1/2}\left(K_s^{-1/2}\sum_{v=1}^{T}k_{s,v}^{1/2}\eta_{v+1}^{(s)}\right)\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}^{1/2}F_{m_2l}^{(s)}\right)$
$K_s^{-1/2}N^{-1}\left(N^{-1/2}K_s^{-1/2}\sum_{i=1}^{N}\sum_{u=1}^{T}k_{s,u}^{1/2}\varepsilon_{iu}^{(s)}\right)\left(N^{-1/2}K_s^{-1/2}\sum_{j=1}^{N}\sum_{t=1}^{T}k_{s,t}^{1/2}\varepsilon_{jt}^{(s)}\right)$
$= O_p\left(K_s^{-1}N^{-1}\right) \leq O_p\left(K_s^{-1/2}N^{-1}\right).$
So, overall, the item in 13 is $O_p\left(\delta_{N,K_s}^{-1}\right)$.

14. As in item 13, replace Lemma A1.14 with A1.15, A1.12 with A1.13 and Assumption 2.4 for Theorem 3.

Lemma A3 *Let Assumptions 1,2 and 5 hold. Then,*

$$\left\|K_s^{-1}\widetilde{X}^{(s)\prime}\widetilde{Z}^{(s)} - K_s^{-1}\Phi_s\widetilde{F}^{(s)\prime}\widetilde{F}^{(s)}\Lambda_s' - K_s^{-1}\widetilde{\varepsilon}^{(s)\prime}\widetilde{\omega}^{(s)}\right\| = O_p(\lambda)$$
$$\left\|K_s^{-1}\widetilde{X}^{(s)\prime}\widetilde{X}^{(s)} - K_s^{-1}\Phi_s\widetilde{F}^{(s)\prime}\widetilde{F}^{(s)}\Phi_s - K_s^{-1}\widetilde{\varepsilon}^{(s)\prime}\widetilde{\varepsilon}^{(s)}\right\| = O_p(\lambda)$$
$$\left\|K_s^{-1}\widetilde{X}^{(s)\prime}\widetilde{y}^{(s)} - K_s^{-1}\Phi_s\widetilde{F}^{(s)\prime}\widetilde{F}^{(s)}\beta_s - K_s^{-1}\widetilde{\eta}^{(s)\prime}\widetilde{\varepsilon}^{(s)}\right\| = O_p(\lambda)$$
$$\left\|K_s^{-1}\widetilde{Z}^{(s)\prime}\widetilde{Z}^{(s)} - K_s^{-1}\Lambda_s\widetilde{F}^{(s)\prime}\widetilde{F}^{(s)}\Lambda_s' - K_s^{-1}\widetilde{\omega}^{(s)\prime}\widetilde{\omega}^{(s)}\right\| = O_p(\lambda)$$

where $\lambda = \kappa(\log N)^{\nu}\max\left(H^{-1/2}, \left(\frac{H}{T}\right)^{1/2}\right) = o_p(1)$, $\nu = \frac{\rho+2}{2\rho}$, $\rho, \kappa > 0$.

Proof of Lemma A3 Let $K_s = K_{H,s}$, $k_{s,t} = k_{H,st}$ and notice that
$K_s^{-1}\widetilde{X}^{(s)\prime}\widetilde{X}^{(s)} = K_s^{-1}\sum_{t=1}^{T}k_{s,t}\left(x_t - \overline{x}_s\right)\left(x_t - \overline{x}_s\right)'$, with

$$x_t = \phi_{0t} + \Phi_t F_t + \varepsilon_t, \text{ and}$$
$$\overline{x}_s = K_s^{-1}\sum_{l=1}^{T}k_{sl}x_l = K_s^{-1}\sum_{l=1}^{T}k_{sl}\phi_{0l} + K_s^{-1}\sum_{l=1}^{T}k_{sl}\Phi_l F_l + K_s^{-1}\sum_{s=1}^{T}k_{s,l}\varepsilon_l, \tag{90}$$

which implies that for all $t = 1, .., T$ and any s it holds

$$x_t - \overline{x}_s = \phi_{0t} - K_s^{-1}\sum_{l=1}^{T}k_{s,l}\phi_{0l} + \Phi_t F_t - \Phi_s F_t + \Phi_s F_t - \Phi_s\overline{F}_s + \Phi_s\overline{F}_s - K_s^{-1}\sum_{l=1}^{T}k_{s,l}\Phi_l F_l + \varepsilon_t - \overline{\varepsilon}_s \tag{91}$$

$$= \left(\phi_{0t} - \overline{\phi}_{0s}\right) + \left(\Phi_t - \Phi_s\right)F_t + \Phi_s\left(F_t - \overline{F}_s\right) + K_s^{-1}\sum_{l=1}^{T}k_{s,l}\left(\Phi_s - \Phi_l\right)F_l + \left(\varepsilon_t - \overline{\varepsilon}_s\right)$$
$$= j_1 + j_2 + j_3 + j_4 + j_5.$$

Without loss of generality we assume that the number of factors $K = 1$. When $K > 1$ the i-th element of $\Phi_l F_l$ is comprised of terms $\sum_{l=1}^{K} \Phi_{il} F_l$, whose bound equals the $K = 1$ bound when the number of factors K is finite.

Then, $K_s^{-1} \sum_{t=1}^{T} k_{s,t} (x_t - \overline{x}_s)(x_t - \overline{x}_s)'$ can be decomposed as
$K_s^{-1} \sum_{t=1}^{T} k_{s,t} (j_1 + j_2 + j_3 + j_4 + j_5)(j_1 + j_2 + j_3 + j_4 + j_5)'$
$=\Omega_s + K_s^{-1} \sum_{t=1}^{T} k_{s,t} (j_3 + j_5)(j_3 + j_5)'$

with $\Omega_s = K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_1 j_1' + K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_1 j_2' + K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_1 j_4'$
$+K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_2 j_1' + K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_2 j_2' + K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_2 j_4'$
$+K_s^{-1} \sum_{t=1}^{T} k_{s,t} (j_1 + j_2 + j_4) j_3' + K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_3 (j_1 + j_2 + j_4)'$
$+K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_4 j_1' + K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_4 j_2' + K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_4 j_4'$
$+K_s^{-1} \sum_{t=1}^{T} k_{s,t} j_5 (j_1 + j_2 + j_4)' + K_s^{-1} \sum_{t=1}^{T} k_{s,t} (j_1 + j_2 + j_4) j_5'$

We have that $\lambda = o_p(1)$ and we will show that

$$\left\| K_s^{-1} \sum_{t=1}^{T} k_{s,t} (x_t - \overline{x}_s)(x_t - \overline{x}_s)' - K_s^{-1} \sum_{t=1}^{T} k_{s,t} (j_3 + j_5)(j_3 + j_5)' \right\| = O_p(\lambda). \tag{92}$$

Notice that from the property of spectral norm of a symmetric matrix

$$\left\| K_s^{-1} \sum_{t=1}^{T} k_{s,t} (x_t - \overline{x}_s)(x_t - \overline{x}_s)' - K_s^{-1} \sum_{t=1}^{T} k_{s,t} (j_3 + j_5)(j_3 + j_5)' \right\| = \|\Omega_s\| \leq \max_i \sum_{k=1}^{N} \left| \Omega_s^{i,k} \right|, \tag{93}$$

where $\Omega_s^{i,k}$ is the i,k-element of Ω_s. When the following probability limit (94) holds,

$$\max_{i,k} \Pr\left(\left| \Omega_s^{i,k} \right| > \lambda \right) = o_p\left(N^{-2}\right), \tag{94}$$

we have that

$$\Pr\left(\max_{i=1,..,N} \sum_{k=1}^{N} \left| \Omega_s^{i,k} \right| > \lambda \right) \leq \Pr\left(\sum_{i,k=1}^{N} \left| \Omega_s^{i,k} \right| > \lambda \right) \leq \sum_{i,k=1}^{N} \Pr\left(\left| \Omega_s^{i,k} \right| > \lambda \right)$$
$$\leq N^2 \max_{i,k=1,..,N} \Pr\left(\left| \Omega_s^{i,k} \right| > \lambda \right) = o(1),$$

which proves (92), when (94) holds.

To prove (94), first notice that $\Omega_s^{i,k}$ is comprised of terms $J^{i,k}_{\{i_1,i_2,..\},\{j_1,j_2,..\}}$ with $J^{i,k}_{\{i_1,i_2,..\},\{j_1,j_2,..\}}$ being the i,k-th element of $J_{\{i_1,i_2,..\},\{j_1,j_2,..\}} = K_s^{-1} \sum_{t=1}^{T} k_{st} \left(j_{i_1} + j_{i_2} + ... \right) \left(j_{j_1} + j_{j_2} + ... \right)'$.

Then, $\Pr\left(\left| \Omega_s^{i,k} \right| > \lambda \right) =$

$$\Pr\left(\left| J_{1,1}^{i,k} + J_{1,2}^{i,k} + J_{1,4}^{i,k} + J_{2,1}^{i,k} + J_{2,2}^{i,k} + J_{2,4}^{i,k} + J_{4,1}^{i,k} + J_{4,2}^{i,k} + J_{4,4}^{i,k} + J_{\{1,2,4\},3}^{i,k} + J_{3,\{1,2,4\}}^{i,k} + J_{\{1,2,4\},5}^{i,k} + J_{5,\{1,2,4\}}^{i,k} \right| > \lambda \right)$$

$\leq \Pr\left(\left|J^{j,k}_{124} + J^{i,k}_{\{1,2,4\},3}\right| > \frac{1}{4}\lambda\right) + \Pr\left(\left|J^{i,k}_{3,\{1,2,4\}}\right| > \frac{1}{4}\lambda\right)$
$+ \Pr\left(\left|J^{i,k}_{\{1,2,4\},5}\right| > \frac{1}{4}\lambda\right) + \Pr\left(\left|J^{i,k}_{5,\{1,2,4\}}\right| > \frac{1}{4}\lambda\right)$, where $J^{j,k}_{124} = J^{i,k}_{1,1} + J^{i,k}_{1,2} + J^{i,k}_{1,4} + J^{i,k}_{2,1} + J^{i,k}_{2,2} + J^{i,k}_{2,4} + J^{i,k}_{4,1} + J^{i,k}_{4,2} + J^{i,k}_{4,4}$

For the term $\Pr\left(\left|J^{i,k}_{\{1,2,4\},3}\right| > \frac{1}{4}\lambda\right)$ we have that by Lemma C1 in Dendramis *et al.* (2021), $\max_t E\left|\Phi_{it}F_t\right| < \infty$, $\max_t E\left|F_t\right| < \infty$, $\left|\Phi_{ks}\right| < \infty$ $\max_t \left|\phi_{0it}\right| < \infty$, and
$K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left|\left(\left(\phi_{0it} - \overline{\phi}_{0is}\right) + \left(\Phi_{it} - \Phi_{is}\right)F_t + K_s^{-1}\sum_{l=1}^{T} k_{s,l}\left(\Phi_{is} - \Phi_{il}\right)F_l\right)\left(F_t - \overline{F}_s\right)\Phi_{ks}\right|$
$\leq 2\left|\Phi_{ks}\right| \max_t \left|F_t\right| K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left|\phi_{0it} - \overline{\phi}_{0is}\right|$
$+2\left|\Phi_{ks}\right| \max_t \left|F_t\right| K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left|\Phi_{it}F_t - \Phi_{it}EF\right|$
$+2\left|\Phi_{ks}\right| \max_t \left|F_t\right| K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left|\Phi_{it} - \Phi_{is}\right| EF$
$+2\left|\Phi_{ks}\right| \max_t \left|F_t\right| K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left|\Phi_{is}\right|\left|EF - F_t\right|$
$+2\left|\Phi_{ks}\right| \max_t \left|F_t\right| K_s^{-1}\sum_{l=1}^{T} k_{s,l}\left(\Phi_{is} - \Phi_{il}\right)F_l$
$= u_1 + u_2 + u_3 + u_4 + u_5$
When it is true that

$$K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left|\phi_{0it} - \overline{\phi}_{0is}\right| = O_p\left(\sqrt{\frac{H}{T}}\right), \tag{95}$$

$$K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left|\Phi_{it} - \Phi_{is}\right|\left|EF\right| = O_p\left(\sqrt{\frac{H}{T}}\right), \tag{96}$$

$$K_s^{-1}\sum_{l=1}^{T} k_{s,l}\left|\left(\Phi_{is} - \Phi_{il}\right)F_l\right| = O_p\left(\sqrt{\frac{H}{T}}\right), \tag{97}$$

we have that

$$\Pr\left(\left|J^{i,k}_{\{1,2,4\},3}\right| > \frac{1}{4}\lambda\right) \leq \Pr\left(\left|u_1 + u_2 + u_3 + u_4 + u_5\right| > \frac{1}{4}\lambda\right)$$
$$\leq \Pr\left(\left|u_{135} + u_2\right| > \frac{1}{8}\lambda\right) + \Pr\left(\left|u_4\right| > \frac{1}{8}\lambda\right),$$

with $u_{135} = u_1 + u_3 + u_5$. Then,

$$\Pr\left(\left|u_4\right| > \frac{1}{8}\lambda\right) \leq \Pr\left(\frac{H^{1/2}}{K_s}\left|H^{-1/2}\sum_{t=1}^{T} k_{s,t}\left(F_t - EF\right)\right| > \frac{1}{8C_1}\lambda\right)$$
$$\leq \Pr\left(\left|H^{-1/2}\sum_{t=1}^{T} k_{s,t}\left(F_t - EF\right)\right| > \eta\right),$$

with $\eta := \frac{1}{8C_1}\lambda H^{1/2}$ and there exists $a_1, a_2 > 0$, such that $a_1 H \leq K_s \leq a_2 H$, for all s, that is $\frac{K_s}{H^{1/2}} \geq \alpha_1 H^{1/2}$. Then, using (45) of Lemma 2 of Dendramis *et al.* (2021), we obtain

$$\Pr\left(\left|H^{-1/2}\sum_{t=1}^{T} k_{s,t}\left(F_t - EF\right)\right| > \eta\right) \leq g_H\left(2, \theta', c, \eta\right),\ 2 < \theta' < \theta/2 \tag{98}$$

with

$$g_H\left(\gamma_1,\theta',c,\eta\right)=c_0\left\{\exp\left(-c_1\eta^{\gamma_1}\right)+\eta^{-\theta'}H^{-\left(\frac{\theta'}{2}-1\right)}\right\}. \tag{99}$$

Notice that for $\lambda=\kappa\left(\log N\right)^{\nu}\max\left(H^{-1/2},\left(\frac{H}{T}\right)^{1/2}\right)$, for a $\kappa>0$, we have that $\eta\geq a_1\kappa\left(\log N\right)^{\nu}$, $a_1>0$, and for $\gamma_1=2$, $\left(\left(\log N\right)^{\nu}\right)^2>\log N$, when $\nu=\frac{\rho+2}{2\rho}$. Also, it is true that under the specified condition on λ, we have $N^2=o\left(H^{\frac{\theta'}{2}-1}\right)$, and when $c_1\left(a_1\kappa\right)^2>2$ we have that $g_H\left(2,\theta',c,\eta\right)=o\left(N^{-2}\right)$. We conclude that $\Pr\left(|u_4|>\frac{1}{8}\lambda\right)=o\left(N^{-2}\right)$.

For the term $\Pr\left(|u_{135}+u_2|>\frac{1}{8}\lambda\right)$, let

$q_{135}=CH^{-1/2}\sum_{t=1}^{T}k_{s,t}\left|\phi_{0it}-\overline{\phi}_{0is}\right|+CH^{-1/2}\sum_{t=1}^{T}k_{s,t}\left|\Phi_{it}-\Phi_{is}\right|+CH^{-1/2}\sum_{l=1}^{T}k_{s,l}\left|\left(\Phi_{is}-\Phi_{il}\right)F_l\right|$

$q_2=C\,H^{-1/2}\sum_{t=1}^{T}k_{s,t}\left|\Phi_{it}F_t-\Phi_{it}EF\right|$

$$\begin{aligned}\Pr\left(|u_{135}+u_2|>\frac{1}{8}\lambda\right)&\leq\Pr\left(\frac{H^{1/2}}{K_s}|q_{135}+q_2|>\frac{1}{8}\lambda\right)\\&\leq\Pr\left(|q_{135}+q_2|>\frac{1}{8}\eta_1\right),\end{aligned}$$

with $\eta_1:=\frac{1}{8}\lambda H^{1/2}$. Notice that when (95), (96), (97) are true, we have that $q_{135}=O_p\left(\frac{H}{\sqrt{T}}\right)$, since

$$\begin{aligned}|q_{135}|&\leq CH^{-1/2}\sum_{t=1}^{T}k_{s,t}\left|\phi_{0it}-\overline{\phi}_{0is}\right|+CH^{-1/2}\sum_{t=1}^{T}k_{s,t}\left|\Phi_{it}-\Phi_{is}\right|+CH^{-1/2}\sum_{t=1}^{T}k_{s,t}\left|\left(\Phi_{is}-\Phi_{il}\right)F_l\right|\\&=C\frac{K_s}{H^{1/2}}\left(\frac{1}{K_s}\sum_{t=1}^{T}k_{s,t}\left|\phi_{0it}-\overline{\phi}_{0is}\right|+\frac{1}{K_s}\sum_{t=1}^{T}k_{s,t}\left|\Phi_{it}-\Phi_{is}\right|+\frac{1}{K_s}\sum_{t=1}^{T}k_{s,t}\left|\left(\Phi_{is}-\Phi_{il}\right)F_l\right|\right)\\&=O_p\left(\frac{H}{\sqrt{T}}\right).\end{aligned}$$

Since there exist $a_1,a_2>0$, such that $a_1H\leq K_s\leq a_2H$, for all s, that is $\frac{K_s}{H^{1/2}}\leq\alpha_2H^{1/2}$. Also, notice that

$$\frac{\eta_1}{|q_{135}|}\geq\frac{\frac{1}{8}\kappa\left(\log N\right)^{\nu}\frac{H}{T^{1/2}}}{C\frac{H}{T^{1/2}}}\geq16,\nu=\frac{\rho+2}{2\rho},$$

for any $s>0$ and suitable κ. Then, since $|q_{135}|<\eta_1/16$

$$\begin{aligned}\Pr\left(|q_{135}+q_2|>\frac{1}{8}\eta_1\right)&\leq\Pr\left(|q_2|>\frac{1}{8}\eta_1-|q_{135}|\right)\\&\leq\Pr\left(|q_2|>\frac{1}{16}\eta_1\right).\end{aligned}$$

Thus, by claim (59) of Lemma 4 in Dendramis *et al.* (2021), we have that

$$\Pr\left(|q_2| > \frac{1}{16}\eta_1\right) \leq g_H\left(\gamma_1, \theta', c, \eta_1\left(1 \wedge d_H\right)\right),$$
$$2 < \theta' < \theta/2, d_H = \frac{T^{1/2}}{H}, \gamma_1 = \frac{2\rho}{2+\rho},$$

with $a \wedge b = \min(a, b)$. For $H > T^{1/2}$, we have $1 \wedge d_H = d_H$, and

$$\eta_1\left(1 \wedge d_H\right) \geq \frac{1}{8}\kappa\left(\log N\right)^{\nu} \frac{H}{T^{1/2}} \frac{T^{1/2}}{H} \geq C\kappa\left(\log N\right)^{\nu}.$$

Also, noticing that $\nu\gamma_1 = 1$, it is

$$\begin{aligned} g_H\left(\gamma_1, \theta', c, \eta_1\left(1 \wedge d_H\right)\right) &\leq c_0\left\{\exp\left(-c_1\left(C\kappa\left(\log N\right)^{\nu}\right)^{\gamma_1}\right) + \left(C\kappa\left(\log N\right)^{\nu}\right)^{-\theta'} H^{-\left(\frac{\theta'}{2}-1\right)}\right\} \\ &\leq c_0\left\{\exp\left(-c_1\left(C\kappa\right)^{\gamma_1}\log N\right) + \frac{1}{\left(C\kappa\left(\log N\right)^{\nu}\right)^{\theta'}} \frac{1}{H^{\left(\frac{\theta'}{2}-1\right)}}\right\} = o\left(N^{-2}\right), \end{aligned}$$

when κ is selected such that $\left(C\kappa\right)^{\gamma_1} > 2$, θ' is selected close enough to $\theta/2$, we have $N^2 = o\left(H^{\frac{\theta'}{2}-1}\right)$.

To prove (95), (96), (97) notice that $\phi_{0it} - \overline{\phi}_{0is} = \left(\phi_{0it} - \phi_{0is}\right) + \left(\phi_{0is} - \overline{\phi}_{0is}\right)$ where $\phi_{0is} - \overline{\phi}_{0is} = K_s^{-1}\sum_{l=1}^{T} k_{s,l}\left(\phi_{0is} - \phi_{0il}\right) = O_p\left(\sqrt{\frac{H}{T}}\right)$, by Lemma C3 in Dendramis *et al.* (2021), implying that

$$\phi_{0it} - \overline{\phi}_{0is} = \left(\phi_{0it} - \phi_{0is}\right) + O_p\left(\sqrt{\frac{H}{T}}\right). \tag{100}$$

Then,

$$\begin{aligned} K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left|\phi_{0it} - \overline{\phi}_{0is}\right| &= K_s^{-1}\sum_{t=1}^{T} k_{s,t}\left(\phi_{0it} - \phi_{0is}\right) + O_p\left(\sqrt{\frac{H}{T}}\right) \\ &= O_p\left(\sqrt{\frac{H}{T}}\right), \end{aligned}$$

by Lemma C3 in Dendramis *et al.* (2021). The same limit holds for (96) since $|EF| < \infty$. Also, for (97) it holds that

$$K_s^{-1}\sum_{l=1}^{T} k_{s,l}\left(\Phi_{is} - \Phi_{il}\right) F_l = O_p\left(\sqrt{\frac{H}{T}}\right), \tag{101}$$

by Corollary 9(b) in Dendramis *et al.* (2021).

For the term $\Pr\left(\left|J_{\{1,2,4\},5}^{i,k}\right| > \frac{1}{4}\lambda\right)$ notice that since $\max_t |\varepsilon_t| < \infty$,

$K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\left(\left(\phi_{0it}-\overline{\phi}_{0is}\right)+\left(\Phi_{it}-\Phi_{is}\right)F_t+K_s^{-1}\sum_{l=1}^T k_{s,l}\left(\Phi_{is}-\Phi_{il}\right)F_l\right)\left(\varepsilon_{kt}-\overline{\varepsilon}_{ks}\right)\right|$
$\leq 2\max_t|\varepsilon_t|\,K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\phi_{0it}-\overline{\phi}_{0is}\right|$
$+2\max_t|\varepsilon_t|\,K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\Phi_{it}F_t-\Phi_{it}EF\right|$
$+2\max_t|\varepsilon_t|\,K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\Phi_{it}-\Phi_{is}\right|EF$
$+2\max_t|\varepsilon_t|\,K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\Phi_{is}\right|\left|EF-F_t\right|$
$+2\max_t|\varepsilon_t|\,K_s^{-1}\sum_{l=1}^T k_{s,l}\left(\Phi_{is}-\Phi_{il}\right)F_l$

Following the same arguments as in the term $J_{\{1,2,4\},3}^{i,k}$, we can prove that $\left|J_{\{1,2,4\},5}^{i,k}\right|=o_p\left(N^{-2}\right)$. The same bound holds for the terms $J_{3,\{1,2,4\}}^{i,k}$, $J_{5,\{1,2,4\}}^{i,k}$.

For the term $\Pr\left(\left|J_{124}^{j,k}+J_{\{1,2,4\},3}^{i,k}\right|>\frac{1}{4}\lambda\right)$, notice that

$$\begin{aligned}\Pr\left(\left|J_{124}^{j,k}+J_{\{1,2,4\},3}^{i,k}\right|>\frac{1}{4}\lambda\right)&\leq\Pr\left(\left|J_{124}^{j,k}\right|+\left|J_{\{1,2,4\},3}^{i,k}\right|>\frac{1}{4}\lambda\right)\\&\leq\Pr\left(\left|J_{\{1,2,4\},3}^{i,k}\right|>\frac{1}{4}\lambda-\left|J_{124}^{j,k}\right|\right)\end{aligned}$$

if we show that $\left|J_{124}^{j,k}\right|=O_p\left(\sqrt{\frac{H}{T}}\right)$. Then, notice that

$$\frac{\frac{1}{4}\lambda}{\left|J_{124}^{j,k}\right|}\geq\frac{\frac{1}{4}\kappa\left(\log N\right)^{\nu}\left(\frac{H}{T}\right)^{1/2}}{\left(\frac{H}{T}\right)^{1/2}}\geq 2$$

for suitable κ. Then, $\left|J_{124}^{j,k}\right|\leq\frac{1}{8}\lambda$ and

$$\Pr\left(\left|J_{124}^{j,k}+J_{\{1,2,4\},3}^{i,k}\right|>\frac{1}{4}\lambda\right)\leq\Pr\left(\left|J_{\{1,2,4\},3}^{i,k}\right|>\frac{1}{8}\lambda\right)=o_p\left(1\right).$$

To show that $\left|J_{124}^{j,k}\right|=O_p\left(\sqrt{\frac{H}{T}}\right)$, first notice that for the $J_{1,1}^{i,k}$ we have
$\left|K_s^{-1}\sum_{t=1}^T k_{s,t}\left(\phi_{0it}-\overline{\phi}_{0is}\right)\left(\phi_{0kt}-\overline{\phi}_{0ks}\right)\right|$
$\leq\left|K_s^{-1}\sum_{t=1}^T k_{s,t}\left(\phi_{0it}-\phi_{0is}\right)\left(\phi_{0kt}-\phi_{0ks}\right)\right|+\left|2K_s^{-1}\sum_{t=1}^T k_{s,t}\left(\phi_{0it}-\phi_{0is}\right)\left(\phi_{0ks}-\overline{\phi}_{0ks}\right)\right|$
$+\left|K_s^{-1}\sum_{t=1}^T k_{s,t}\left(\phi_{0is}-\overline{\phi}_{0is}\right)\left(\phi_{0ks}-\overline{\phi}_{0ks}\right)\right|$
$\leq\max_t\left|\phi_{0it}\right|K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\phi_{0kt}-\phi_{0ks}\right|+\left|\phi_{0is}\right|K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\phi_{0kt}-\phi_{0ks}\right|$
$+\max_t\left|\phi_{0it}\right|2K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\left(\phi_{0ks}-\overline{\phi}_{0ks}\right)\right|+\left|\phi_{0is}\right|2K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\left(\phi_{0ks}-\overline{\phi}_{0ks}\right)\right|$
$+\left|\phi_{0is}\right|K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\phi_{0ks}-\overline{\phi}_{0ks}\right|+\max_s\left|\overline{\phi}_{0is}\right|K_s^{-1}\sum_{t=1}^T k_{s,t}\left|\phi_{0ks}-\overline{\phi}_{0ks}\right|$
$=O_p\left(\sqrt{\frac{H}{T}}\right)$, by (100), Lemma C3 in DGK21, and $\max_t\left|\phi_{0it}\right|=O\left(1\right)$.

For $J_{1,2}^{i,k}$ we have

$\left|K_s^{-1}\sum_{t=1}^{T}k_{s,t}\left(\phi_{0it}-\overline{\phi}_{0is}\right)F_t\left(\Phi_{kt}-\Phi_{ks}\right)\right|$
$\leq \max_t\left|\phi_{0it}\right|K_s^{-1}\sum_{t=1}^{T}k_{s,t}\left|F_t\left(\Phi_{kt}-\Phi_{ks}\right)\right|+\left|\overline{\phi}_{0is}\right|\left|K_s^{-1}\sum_{t=1}^{T}k_{s,t}F_t\left(\Phi_{kt}-\Phi_{ks}\right)\right|=O_p\left(\sqrt{\frac{H}{T}}\right)$
by (101).

For the term $K_s^{-1}\sum_{t=1}^{T}k_{s,t}j_1j_4'$ we have
$\left|K_s^{-1}\sum_{t=1}^{T}k_{s,t}\left(\phi_{0it}-\overline{\phi}_{0is}\right)\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}\left(\Phi_s-\Phi_l\right)F_l\right)\right|$
$\leq \left|K_s^{-1}\sum_{t=1}^{T}k_{s,t}\left(\phi_{0it}-\overline{\phi}_{0is}\right)\right|O_p(1)\leq O_p\left(\sqrt{\frac{H}{T}}\right)$, by (101) and the steps of the proof for the element $K_s^{-1}\sum_{t=1}^{T}k_{s,t}j_1j_1'$.

For the term $K_s^{-1}\sum_{t=1}^{T}k_{s,t}j_2j_4'$ we have
$\left|K_s^{-1}\sum_{t=1}^{T}k_{s,t}\left(\left(\Phi_{it}-\Phi_{is}\right)F_t\right)\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}\left(\Phi_{ks}-\Phi_{kl}\right)F_l\right)\right|$
$\leq \left|K_s^{-1}\sum_{t=1}^{T}k_{s,t}\left(\left(\Phi_{it}-\Phi_{is}\right)F_t\right)\right|\left|\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}F_l\left(\Phi_{ks}-\Phi_{kl}\right)\right)\right|$
$\leq O_p\left(\sqrt{\frac{H}{T}}\right)$, by (101)

For the term $K_s^{-1}\sum_{t=1}^{T}k_{s,t}j_4j_4'$ we have
$\left|\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}\left(\Phi_{ks}-\Phi_{kl}\right)F_l\right)\left(K_s^{-1}\sum_{l=1}^{T}k_{s,l}\left(\Phi_{ks}-\Phi_{kl}\right)F_l\right)\right|\leq O_p\left(\sqrt{\frac{H}{T}}\right)$, by (101), and this suffices to derive the reported bound of the Lemma.

With analogous arguments we can derive the other results of Lemma A3.

Lemma A4 *Let Assumptions 1, 2 and 5 hold. The probability limits of* $\widehat{F}_t=\left(\widehat{\Phi}_t'J_N\widehat{\Phi}_t\right)^{-1}\widehat{\Phi}_t'J_Nx_t$, *and* $\widehat{\Phi}_t'=\left(\widetilde{Z}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\right)^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}$ *are*

$$\widehat{F}_t\underset{N,T\to\infty}{\overset{p}{\to}}\left(\Lambda_t\Delta_F\Lambda_t'+\Delta_\omega\right)\left(\Lambda_t\Delta_F\Delta_\Phi\Delta_F\Lambda_t'\right)^{-1}\left(\Lambda_t\Delta_FP_1+\Lambda_t\Delta_F\Delta_\Phi F_t\right),\tag{102}$$

$$\widehat{\Phi}_t'\underset{N,T\to\infty}{\overset{p}{\to}}\left(\Lambda_t\Delta_F\Lambda_t'+\Delta_\omega\right)^{-1}\Lambda_t\Delta_F\Phi_t'.\tag{103}$$

Proof of Lemma A4 Let $\widehat{F}_{(1),t}=\left(\widehat{\Phi}_t'J_N\widehat{\Phi}_t\right)^{-1}\widehat{\Phi}_t'J_N$, then we can write

$$\begin{aligned}\widehat{F}_t&=\widehat{F}_{(1),t}x_t&(104)\\&=K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{Z}^{(t)}\left(N^{-1}K_t^{-2}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)}J_N\widetilde{X}^{(t)\prime}\widetilde{Z}^{(t)}\right)^{-1}N^{-1}K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)}J_Nx_t\\&=\widehat{F}_{A,t}\widehat{F}_{B,t}^{-1}\widehat{F}_{C_{(1)},t}x_t.&(105)\end{aligned}$$

By Lemma A3, for the term $\widehat{F}_{A,t}$ we have

$$K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{Z}^{(t)} = K_t^{-1}\Lambda_t\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\Lambda_t' + K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{\omega}^{(t)} \xrightarrow[N,T\to\infty]{p} \Lambda_t\Delta_F\Lambda_t' + \Delta_\omega, \tag{106}$$

with $E\left(K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{\omega}^{(t)}\right) \to \Delta_\omega$. For the term $\widehat{F}_{B,t}$ we have

$$\begin{aligned} N^{-1}K_t^{-2}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)}J_N\widetilde{X}^{(t)\prime}\widetilde{Z}^{(t)} &= N^{-1}\left(K_t^{-1}\Phi_t\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\Lambda_t'\right)' J_N K_t^{-1}\Phi_t\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\Lambda_t' \\ &+ N^{-1}\left(K_t^{-1}\widetilde{\varepsilon}^{(t)\prime}\widetilde{\omega}^{(t)}\right)' J_N K_t^{-1}\Phi_t\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\Lambda_t' \qquad (107) \\ &+ N^{-1}\left(K_t^{-1}\Phi_t\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\Lambda_t'\right)' J_N K_t^{-1}\widetilde{\varepsilon}^{(t)\prime}\widetilde{\omega}^{(t)} \qquad (108) \\ &+ N^{-1}\left(K_t^{-1}\widetilde{\varepsilon}^{(t)\prime}\widetilde{\omega}^{(t)}\right)' J_N K_t^{-1}\widetilde{\varepsilon}^{(t)\prime}\widetilde{\omega}^{(t)} \qquad (109) \\ &\xrightarrow[N,T\to\infty]{p} \Lambda_t\Delta_F\Delta_\Phi\Delta_F\Lambda_t', \qquad (110) \end{aligned}$$

due to results in Lemma A2.6 and Lemma A3 . For the term $\widehat{F}_{C_{(1)},t}x_t$ we have that

$$\begin{aligned} N^{-1}K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)}J_N x_t &= N^{-1}\left(K_t^{-1}\Phi_t\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\Lambda_t'\right)' J_N\phi_{0t} \qquad (111) \\ & N^{-1}\left(K_t^{-1}\Phi_t\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\Lambda_t'\right)' J_N\Phi_t F_t \qquad (112) \\ & N^{-1}\left(K_t^{-1}\Phi_t\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\Lambda_t'\right)' J_N\varepsilon_t \qquad (113) \\ & N^{-1}\left(K_t^{-1}\widetilde{\varepsilon}^{(t)\prime}\widetilde{\omega}^{(t)}\right)' J_N\phi_{0t} \qquad (114) \\ & N^{-1}\left(K_t^{-1}\widetilde{\varepsilon}^{(t)\prime}\widetilde{\omega}^{(t)}\right)' J_N\Phi_t F_t \qquad (115) \\ &+ N^{-1}\left(K_t^{-1}\widetilde{\varepsilon}^{(t)\prime}\widetilde{\omega}^{(t)}\right)' J_N\varepsilon_t \qquad (116) \\ &\xrightarrow[N,T\to\infty]{p} \Lambda_t\Delta_F P_1 + \Lambda_t\Delta_F\Delta_\Phi F_t. \qquad (117) \end{aligned}$$

The final result is obtained via the continuous mapping Theorem.

For the term $\widehat{\Phi}_t$, we have that $\widehat{\Phi}_t' = \left(K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{Z}^{(t)}\right)^{-1} K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)}$ with $K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{Z}^{(t)} \xrightarrow[N,T\to\infty]{p} \Lambda_t\Delta_F\Lambda_t' + \Delta_\omega$. Also,

$$K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)} = \left(K_s^{-1}\Phi_s\widetilde{F}^{(s)\prime}\widetilde{F}^{(s)}\Lambda_s' + K_s^{-1}\widetilde{\varepsilon}^{(s)\prime}\widetilde{\omega}^{(s)}\right)' \xrightarrow[N,T\to\infty]{p} \Lambda_t\Delta_F\Phi_t', \tag{118}$$

by Theorem 3 , Lemma A2.1, and Assumption 2. Then, the final result is obtained via the continuous mapping Theorem.

Lemma A5 *Let Assumptions 1, 2 and 5 hold. For the third stage predictive coefficient* $\widehat{\beta}_t$ *we have*

$$\widehat{\beta}_t \underset{N,T\to\infty}{\overset{p}{\to}} \left(\Lambda_t \Delta_F \Lambda_t' + \Delta_\omega\right)^{-1} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \Lambda_t' \left(\Lambda_t \Delta_F \Delta_\Phi \Delta_F \Delta_\Phi \Delta_F \Lambda_t'\right)^{-1} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \beta_t. \quad (119)$$

Proof of Lemma A5 The third stage estimator of the target equation is given by

$$\widehat{\beta}_t = \left(K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_s - \overline{\widehat{F}}_t\right) \left(\widehat{F}_s - \overline{\widehat{F}}_t\right)'\right)^{-1} \quad (120)$$

$$\times \left(K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_s - \overline{\widehat{F}}_t\right) \left(y_{s+1} - \overline{y}_t\right)\right) \quad (121)$$

$$= \Delta_{(1),t}^{-1} \Delta_{(2),t},$$

Then, using the decomposition (104) of $\widehat{F}_t$, we have

$$\Delta_{(1),t} = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left\{ \begin{array}{c} \left(\widehat{F}_{(1),t} \left(x_s - \overline{x}_t\right) + \left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right) x_s + \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right)\right) \\ \times \left(\widehat{F}_{(1),t} \left(x_s - \overline{x}_t\right) + \left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right) x_s + \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right)\right)' \end{array} \right\}$$

$$= K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \left(x_s - \overline{x}_t\right) \left(x_s - \overline{x}_t\right)' \widehat{F}_{(1),t}'\right) + K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \left(x_s - \overline{x}_t\right) x_s' \left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right)'\right)$$
$$+ K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \left(x_s - \overline{x}_t\right) \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right)'\right) + K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right) x_s \left(x_s - \overline{x}_t\right)' \widehat{F}_{(1),t}'\right)$$
$$+ K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right) x_s \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right)'\right) + K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right) \left(x_s - \overline{x}_t\right)' \widehat{F}_{(1),t}'$$
$$+ K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right) x_s' \left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right)' + K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right) x_s x_s' \left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right)'\right)$$
$$+ K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right) \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right)'$$

$$= K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \left(x_s - \overline{x}_t\right) \left(x_s - \overline{x}_t\right)' \widehat{F}_{(1),t}\right) + S_t^{(1)} + S_t^{(2)} + S_t^{(3)} + S_t^{(4)}$$
$$+ S_t^{(5)} + S_t^{(6)} + S_t^{(7)} + S_t^{(8)}.$$

Analogously, for $\Delta_{(2),t}$ we have

$$\Delta_{(2),t} = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_s - \overline{\widehat{F}}_t\right) \left(y_{s+1} - \overline{y}_t\right)$$
$$= K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \left(x_s - \overline{x}_t\right) + \left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t}\right) x_s + \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}}_t\right)\right) \left(y_{s+1} - \overline{y}_t\right)$$

$$= K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),t}\left(x_s-\overline{x}_t\right)\left(y_{s+1}-\overline{y}_t\right)' + K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),s}-\widehat{F}_{(1),t}\right)x_s\left(y_{s+1}-\overline{y}_t\right)'$$

$$+ K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}\overline{x}_t-\overline{\widehat{F}}_t\right)\left(y_{s+1}-\overline{y}_t\right)'$$

$$= K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),t}\left(x_s-\overline{x}_t\right)\left(y_{s+1}-\overline{y}_t\right)' + W_t^{(1)} + W_t^{(2)}.$$

We will prove that $S_t^{(1)}+S_t^{(2)}+S_t^{(3)}+S_t^{(4)}+S_t^{(5)}+S_t^{(6)}+S_t^{(7)}+S_t^{(8)}=o\left(1\right)$ and $W_t^{(1)}+W_t^{(2)}=o\left(1\right)$.

Focusing on the terms $S_t^{(i)}$, $i=1,..,8$ we have that

$S_t^{(1)} = K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}\left(x_s-\overline{x}_t\right)x_s'\left(\widehat{F}_{(1),s}-\widehat{F}_{(1),t}\right)'\right)$
$= K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}x_sx_s'\widehat{F}_{(1),s}'-\widehat{F}_{(1),t}\overline{x}_tx_s'\widehat{F}_{(1),s}'-\widehat{F}_{(1),t}x_sx_s'\widehat{F}_{(1),t}'+\widehat{F}_{(1),t}\overline{x}_tx_s'\widehat{F}_{(1),t}'\right)$
$= \widehat{F}_{(1),t}\overline{x_tx_t'\widehat{F}_{(1),t}'}-\widehat{F}_{(1),t}\overline{x}_t\overline{x_t'\widehat{F}_{(1),t}'}-\widehat{F}_{(1),t}\overline{x_tx_t'}\widehat{F}_{(1),t}'+\widehat{F}_{(1),t}\overline{x}_t\overline{x}_t'\widehat{F}_{(1),t}'$

$S_t^{(2)} = K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}\left(x_s-\overline{x}_t\right)\left(\widehat{F}_{(1),t}\overline{x}_t-\overline{\widehat{F}}_t\right)'\right)$
$= K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}x_s\overline{x}_t'\widehat{F}_{(1),t}'-\widehat{F}_{(1),t}x_s\overline{\widehat{F}}_t'-\widehat{F}_{(1),t}\overline{x}_t\overline{x}_t'\widehat{F}_{(1),t}'+\widehat{F}_{(1),t}\overline{x}_t\overline{\widehat{F}}_t'\right)$
$= \widehat{F}_{(1),t}\overline{x}_t\overline{x}_t'\widehat{F}_{(1),t}'-\widehat{F}_{(1),t}\overline{x}_t\overline{\widehat{F}}_t'-\widehat{F}_{(1),t}\overline{x}_t\overline{x}_t'\widehat{F}_{(1),t}'+\widehat{F}_{(1),t}\overline{x}_t\overline{\widehat{F}}_t'=0$

$S_t^{(3)} = S_t^{(1)\prime}$

$S_t^{(4)} = K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\left(\widehat{F}_{(1),s}-\widehat{F}_{(1),t}\right)x_s\left(\widehat{F}_{(1),t}\overline{x}_t-\overline{\widehat{F}}_t\right)'\right)$
$= K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),s}x_s\overline{x}_t'\widehat{F}_{(1),t}'-\widehat{F}_{(1),s}x_s\overline{\widehat{F}}_t-\widehat{F}_{(1),t}x_s\overline{x}_t'\widehat{F}_{(1),t}'+\widehat{F}_{(1),t}x_s\overline{\widehat{F}}_t'\right)$
$= \overline{\widehat{F}_{(1),t}x_t}\overline{x}_t'\widehat{F}_{(1),t}'-\overline{\widehat{F}_{(1),t}x_t\widehat{F}_t}'-\widehat{F}_{(1),t}\overline{x}_t\overline{x}_t'\widehat{F}_{(1),t}'+\widehat{F}_{(1),t}\overline{x}_t\overline{\widehat{F}}_t'$

$S_t^{(5)} = S_t^{(2)\prime} = 0$

$S_t^{(6)} = K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}\overline{x}_t-\overline{\widehat{F}}_t\right)x_s'\left(\widehat{F}_{(1),s}-\widehat{F}_{(1),t}\right)'$
$= S_t^{(4)\prime}$

$S_t^{(7)} = K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\left(\widehat{F}_{(1),s}-\widehat{F}_{(1),t}\right)x_sx_s'\left(\widehat{F}_{(1),s}-\widehat{F}_{(1),t}\right)'\right)$
$= K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),s}x_sx_s'\widehat{F}_{(1),s}'-\widehat{F}_{(1),s}x_sx_s'\widehat{F}_{(1),t}'-\widehat{F}_{(1),t}x_sx_s'\widehat{F}_{(1),s}'+\widehat{F}_{(1),t}x_sx_s'\widehat{F}_{(1),t}'\right)$
$=\overline{\widehat{F}_{(1),t}x_tx_t'\widehat{F}_{(1),t}'}-\overline{\widehat{F}_{(1),t}x_tx_t'}\widehat{F}_{(1),t}'-\widehat{F}_{(1),t}\overline{x_tx_t'\widehat{F}_{(1),t}'}+\widehat{F}_{(1),t}\overline{x_tx_t'}\widehat{F}_{(1),t}'$

$$S_t^{(8)} = K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}\overline{x}_t - \overline{\widehat{F}}_t\right)\left(\widehat{F}_{(1),t}\overline{x}_t - \overline{\widehat{F}}_t\right)'$$
$$= K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}\overline{x}_t\overline{x}_t'\widehat{F}_{(1),t}' - \widehat{F}_{(1),t}\overline{x}_t\overline{\widehat{F}}_t' - \overline{\widehat{F}}_t\overline{x}_t'\widehat{F}_{(1),t}' + \overline{\widehat{F}}_t\overline{\widehat{F}}_t'\right)$$
$$= \widehat{F}_{(1),t}\overline{x}_t\overline{x}_t'\widehat{F}_{(1),t}' - \widehat{F}_{(1),t}\overline{x}_t\overline{\widehat{F}}_t' - \overline{\widehat{F}}_t\overline{x}_t'\widehat{F}_{(1),t}' + \overline{\widehat{F}}_t\overline{\widehat{F}}_t'$$

Then, we have

$$\sum_{i=1}^{8} S_t^{(i)} = \overline{\widehat{F}_{(1),t}x_tx_t'\widehat{F}_{(1),t}'} - \overline{\widehat{F}}_t\overline{\widehat{F}}_t' + \widehat{F}_{(1),t}\overline{x}_t\overline{x}_t'\widehat{F}_{(1),t}' - \widehat{F}_{(1),t}\overline{x_tx_t'}\widehat{F}_{(1),t}'$$
$$= -K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),s}x_sx_s'\widehat{F}_{(1),s}' + K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),s}x_sx_s'\widehat{F}_{(1),t}'$$
$$+K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),s}x_sx_s'\widehat{F}_{(1),t}' - K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),t}x_sx_s'\widehat{F}_{(1),t}'$$
$$-K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),s}x_s\overline{\widehat{F}}_t' + K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),t}x_s\overline{\widehat{F}}_t'$$
$$+K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),t}\overline{x}_tx_s'\widehat{F}_{(1),t}' - K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),t}x_s\overline{\widehat{F}}_t'$$

$$= K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),s}x_s\left(x_s'\widehat{F}_{(1),s}' - x_s'\widehat{F}_{(1),t}'\right) + K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),s}x_s - \widehat{F}_{(1),t}x_s\right)x_s'\widehat{F}_{(1),t}'$$
$$+ K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\left(\widehat{F}_{(1),t}x_s - \widehat{F}_{(1),s}x_s\right)\overline{\widehat{F}}_t' + K_{L,t}^{-1}\sum_{s=1}^{T} k_{L,ts}\widehat{F}_{(1),t}\overline{x}_t\left(x_s'\widehat{F}_{(1),t}' - x_s'\widehat{F}_{(1),s}'\right) \tag{122}$$

We will show that all the above terms are $o_p(1)$. First notice that by the decomposition (104) we have that $\widehat{F}_{(1),t}x_s = \widehat{F}_{A,t}\widehat{F}_{B,t}^{-1}\widehat{F}_{C_{(1)},t}x_s$ and the term $\widehat{F}_{C_{(1)},t}x_s$ is

$$\begin{aligned}
N^{-1}K_t^{-1}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)}J_Nx_s = {} & \Lambda_t\left(K_t^{-1}\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\right)N^{-1}\Phi_t'J_N\left(\phi_{0s} - \phi_{0t}\right) + \Lambda_t\left(K_t^{-1}\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\right)N^{-1}\Phi_t'J_N\phi_{0t} \\
& + \Lambda_t\left(K_t^{-1}\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\right)N^{-1}\Phi_t'J_N\left(\Phi_s - \Phi_t\right)F_s + \Lambda_t\left(K_t^{-1}\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\right)\left(N^{-1}\Phi_t'J_N\Phi_t\right)F_s \\
& + \Lambda_t\left(K_t^{-1}\widetilde{F}^{(t)\prime}\widetilde{F}^{(t)}\right)\left(N^{-1}\Phi_t'J_N\varepsilon_s\right) \\
& + \Lambda_t\left(N^{-1}K_t^{-1}\widetilde{F}^{(t)\prime}\widetilde{\varepsilon}^{(t)}J_N\phi_{0s}\right) \\
& + \Lambda_t\left(N^{-1}K_t^{-1}\widetilde{F}^{(t)\prime}\widetilde{\varepsilon}^{(t)}J_N\Phi_s\right)F_s \\
& + \Lambda_t\left(N^{-1}K_t^{-1}\widetilde{F}^{(t)\prime}\widetilde{\varepsilon}^{(t)}J_N\varepsilon_s\right) \\
& + \left(K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{F}^{(t)\prime}\right)N^{-1}\Phi_t'J_N\left(\phi_{0s} - \phi_{0t}\right) + \left(K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{F}^{(t)\prime}\right)N^{-1}\Phi_t'J_N\phi_{0t} \\
& + \Lambda_t\left(K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{F}^{(t)\prime}\right)N^{-1}\Phi_t'J_N\left(\Phi_s - \Phi_t\right)F_s \\
& + \Lambda_t\left(K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{F}^{(t)\prime}\right)\left(N^{-1}\Phi_t'J_N\Phi_t\right)F_s \\
& + \left(K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{F}^{(t)\prime}\right)\left(N^{-1}\Phi_t'J_N\varepsilon_s\right) \\
& + \left(N^{-1}K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{\varepsilon}^{(t)}J_N\Phi_s\right)F_s \\
& + N^{-1}K_t^{-1}\widetilde{\omega}^{(t)\prime}\widetilde{\varepsilon}^{(t)}J_N\varepsilon_s + o_p(1)
\end{aligned} \tag{123}$$

$$= \Lambda_t \left(K_t^{-1} \widetilde{F}^{(t)\prime} \widetilde{F}^{(t)} \right) N^{-1} \Phi_t' J_N \left(\phi_{0s} - \phi_{0t} \right) + \Lambda_t \left(K_t^{-1} \widetilde{F}^{(t)\prime} \widetilde{F}^{(t)} \right) N^{-1} \Phi_t' J_N \phi_{0t}$$
$$+ \Lambda_t \left(K_t^{-1} \widetilde{F}^{(t)\prime} \widetilde{F}^{(t)} \right) N^{-1} \Phi_t' J_N \left(\Phi_s - \Phi_t \right) F_s + \Lambda_t \left(K_t^{-1} \widetilde{F}^{(t)\prime} \widetilde{F}^{(t)} \right) \left(N^{-1} \Phi_t' J_N \Phi_t \right) F_s$$
$$+ o_p(1) = O_p(1), \tag{124}$$

by Lemma A2 (items 4, 5, 6, 12), Theorem 3, and Assumption 1. Now, since

$$N^{-1} \Phi_t' J_N \left(\phi_{0s} - \phi_{0t} \right) = N^{-1} \sum_{i=1}^{N} \Phi_{it} \left(\phi_{0is} - \phi_{0it} \right) - \left(N^{-1} \sum_{i=1}^{N} \Phi_{it} \right) \left(N^{-1} \sum_{i=1}^{N} \left(\phi_{0is} - \phi_{0it} \right) \right),$$

we have that

$$K_t^{-1} \sum_{s=1}^{T} k_{ts} N^{-1} \Phi_t' J_N \left(\phi_{0s} - \phi_{0t} \right) = N^{-1} \sum_{i=1}^{N} \Phi_{it} \left(K_t^{-1} \sum_{s=1}^{T} k_{ts} \left(\phi_{0is} - \phi_{0it} \right) \right)$$
$$- \left(N^{-1} \sum_{i=1}^{N} \Phi_{it} \right) \left(N^{-1} \sum_{i=1}^{N} \left(K_t^{-1} \sum_{s=1}^{T} k_{ts} \left(\phi_{0is} - \phi_{0it} \right) \right) \right)$$
$$= O_p \left(\sqrt{\frac{H}{T}} \right),$$

since $N^{-1} \sum_{i=1}^{N} \Phi_{it} = O_p(1)$ and $K_t^{-1} \sum_{s=1}^{T} k_{ts} \left(\phi_{0is} - \phi_{0it} \right) = O_p \left(\sqrt{\frac{H}{T}} \right)$. Also, with analogous steps we can prove that

$$K_t^{-1} \sum_{s=1}^{T} k_{ts} N^{-1} \Phi_t' J_N \left(\Phi_s - \Phi_t \right) = O_p \left(\sqrt{\frac{H}{T}} \right).$$

The expression in (124), the results in Lemma A2, and Assumption 1, imply that

$$K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(N^{-1} K_t^{-1} \widetilde{Z}^{(t)\prime} \widetilde{X}^{(t)} J_N x_s \right) \underset{N,T\to\infty}{\overset{p}{\rightarrow}} \Lambda_t \Delta_F P_1 + \Lambda_t \Delta_F \Delta_\Phi \overline{F}_{L,t} \tag{125}$$

with $\overline{F}_{L,t} = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} F_s$. Then, we have

$$K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} x_s - \widehat{F}_{(1),s} x_s \right)$$
$$= K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \begin{pmatrix} S_{zz,t} \left(N^{-1} S_{zx,t} J_N S_{zx,t}' \right)^{-1} N^{-1} S_{zx,t} J_N x_s \\ -S_{zz,s} \left(N^{-1} S_{zx,s} J_N S_{zx,s}' \right)^{-1} N^{-1} S_{zx,s} J_N x_s \end{pmatrix}$$
$$= K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \begin{pmatrix} \left(S_{zz,t} - S_{zz,s} \right) \left(N^{-1} S_{zx,t} J_N S_{zx,t}' \right)^{-1} N^{-1} S_{zx,t} J_N x_s \\ +S_{zz,s} \left(\left(N^{-1} S_{zx,t} J_N S_{zx,t}' \right)^{-1} - \left(N^{-1} S_{zx,s} J_N S_{zx,s}' \right)^{-1} \right) N^{-1} S_{zx,t} J_N x_s \\ +S_{zz,s} \left(N^{-1} S_{zx,s} J_N S_{zx,s}' \right)^{-1} \left(N^{-1} S_{zx,t} J_N x_s - N^{-1} S_{zx,s} J_N x_s \right) \end{pmatrix}$$

where,

$$\begin{aligned}
&\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\widehat{F}_{(1),t}x_s-\widehat{F}_{(1),s}x_s\right)\right\| \\
&\leq\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(S_{zz,t}-S_{zz,s}\right)\right\|\left\|\left(N^{-1}S_{zx,t}J_NS'_{zx,t}\right)^{-1}\right\|\max_s\left\|N^{-1}S_{zx,t}J_Nx_s\right\| \\
&+\max_s\left\|S_{zz,h_1,s}\right\|\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(N^{-1}S_{zx,t}J_NS'_{zx,t}\right)^{-1}-\left(N^{-1}S_{zx,s}J_NS'_{zx,s}\right)^{-1}\right\|\max_s\left\|N^{-1}S_{zx,t}J_Nx_s\right\| \\
&+\max_s\left\|S_{zz,s}\right\|\max_s\left\|\left(N^{-1}S_{zx,s}J_NS'_{zx,s}\right)^{-1}\right\|\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(N^{-1}S_{zx,t}J_Nx_s-N^{-1}S_{zx,s}J_Nx_s\right)\right\| \\
&=o_p\left(1\right)
\end{aligned}$$

since, for all t we have $\left\|\left(N^{-1}S_{zx,t}J_NS'_{zx,t}\right)^{-1}\right\|=O_p\left(1\right)$, (see equation (110)), by (124) $\max_s\left\|N^{-1}S_{zx,t}J_Nx_s\right\|=O_p\left(1\right)$, and because of (106) $\max_s\left\|S_{zz,s}\right\|=O_p\left(1\right)$.

Also,

$$\begin{aligned}
&\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(S_{zz,t}-S_{zz,s}\right)\right\| && (126)\\
&=\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\Lambda_t\Delta_F\Lambda'_t-\Lambda_s\Delta_F\Lambda'_s\right)\right\|+o_p\left(1\right) \\
&\leq\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\Lambda_t\Delta_F\Lambda'_t-\Lambda_t\Delta_F\Lambda'_s+\Lambda_t\Delta_F\Lambda'_s-\Lambda_s\Delta_F\Lambda'_s\right)\right\|+o_p\left(1\right) \\
&\leq\left\|\Lambda_t\right\|\left\|\Delta_F\right\|\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\Lambda'_t-\Lambda'_s\right)\right\|+\left\|K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\Lambda_t-\Lambda_s\right)\right\|\left\|\Delta_F\right\|\left\|\Lambda'_s\right\| \\
&=O_p\left(\sqrt{\frac{L}{T}}\right)+O_p\left(\sqrt{\frac{L}{T}}\right)=o_p\left(1\right)
\end{aligned}$$

Analogously, we can prove that $K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(N^{-1}S_{zx,t}J_NS'_{zx,t}\right)^{-1}-\left(N^{-1}S_{zx,s}J_NS'_{zx,s}\right)^{-1}=o_p\left(1\right)$.

Also,

$$\left\| K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(N^{-1} S_{zx,t} J_N x_s - N^{-1} S_{zx,s} J_N x_s \right) \right\|$$
$$= \left\| K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\Lambda_t \Delta_F P_1 + \Lambda_t \Delta_F \Delta_\Phi F_s - \left(\Lambda_s \Delta_F P_1 + \Lambda_s \Delta_F \Delta_\Phi F_s \right) \right) \right\| + o_p(1)$$
$$= \left\| K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\Lambda_t \left(\Delta_F P_1 + \Delta_F \Delta_\Phi F_s \right) - \Lambda_s \left(\Delta_F P_1 + \Delta_F \Delta_\Phi F_s \right) \right) \right\| + o_p(1)$$
$$\leq \left\| K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\Lambda_t - \Lambda_s \right) \right\| \left\| \Delta_F P_1 + \Delta_F \Delta_\Phi F_s \right\| = o_p(1),$$

and we conclude that

$$K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} x_s - \widehat{F}_{(1),s} x_s \right) = o_p(1). \tag{127}$$

Moreover, since $\max_s \left\| \widehat{F}_s \right\| = \max_s \left\| x_s' \widehat{F}_{(1),t}' \right\| = O_p(1)$, $\max_s \left\| \widehat{F}_{(1),t} \overline{x}_t \right\| = \left\| K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \widehat{F}_{(1),t} x_s \right\| = O_p(1)$ which, together with (127), imply $\left\| \sum_{i=1}^{8} S_t^{(i)} \right\| \leq \sum_{i=1}^{8} \left\| S_t^{(i)} \right\| = o_p(1)$.

For the term $\Delta_{(2),t}$ we will prove that $W_t^{(1)} + W_t^{(2)} = o_p(1)$.
We have, $W_t^{(1)} = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),s} - \widehat{F}_{(1),t} \right) x_s \left(y_{s+1} - \overline{y}_t \right)'$
$=K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),s} x_s y_{s+1} - \widehat{F}_{(1),s} x_s \overline{y}_t - \widehat{F}_{(1),t} x_s y_{s+1} + \widehat{F}_{(1),t} x_s \overline{y}_t' \right)$
and
$W_t^{(2)} = K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \overline{x}_t - \overline{\widehat{F}_t} \right) \left(y_{s+1} - \overline{y}_t \right)'$
$=K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left(\widehat{F}_{(1),t} \overline{x}_t y_{s+1} - \widehat{F}_{(1),t} \overline{x}_t \overline{y}_t - \overline{\widehat{F}}_t y_{s+1} + \overline{\widehat{F}}_t \overline{y}_t \right).$

Then, $W_t^{(1)} + W_t^{(2)} = \overline{\widehat{F}_{(1),t} x_t y_t} - \widehat{F}_{(1),t} \overline{x_t y_t} + \widehat{F}_{(1),t} \overline{x}_t \overline{y_t} - \overline{\widehat{F}}_t \overline{y_t}$
$=K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \widehat{F}_{(1),t} \left(x_s \overline{y_t} - x_s y_{s+1} \right) - K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \widehat{F}_{(1),s} \left(x_s \overline{y_t} - x_s y_{s+1} \right).$

For the term $\left(\overline{y_t} - y_t \right)$ we have
$K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \left(y_{l+1} - y_{s+1} \right) = K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \left(\beta_{0l} + \beta_l' F_l + \eta_{l+1} - \beta_{0s} - \beta_s' F_s - \eta_{s+1} \right)$
$=K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \left(\beta_{0l} - \beta_{0s} \right) + K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \left(\beta_l' F_l - \beta_s' F_l + \beta_s' F_l - \beta_s' F_s \right)$
$+K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \left(\eta_{l+1} - \eta_{s+1} \right)$
$=K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \left(\beta_{0l} - \beta_{0s} \right) + K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \left(\beta_l' - \beta_s' \right) F_l$
$+K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \beta_s' \left(F_l - F_s \right) + K_{L,s}^{-1} \sum_{l=1}^{T} k_{L,sl} \left(\eta_{l+1} - \eta_{s+1} \right)$
$=O_p(1).$

Moreover, it is $K_{L,s}^{-1}\sum_{l=1}^{T}k_{L,sl}\left(\beta_{0l}-\beta_{0s}\right)=o_p\left(1\right)$ and $K_{L,s}^{-1}\sum_{l=1}^{T}k_{L,sl}\left(\beta_l'-\beta_s'\right)F_l=o_p\left(1\right)$ by Corollary 9b in Dendramis *et al.* (2021), and $K_{L,s}^{-1}\sum_{l=1}^{T}k_{L,sl}\beta_s'\left(F_l-F_s\right)=O_p\left(1\right)$ since F_t-EF_t is a strong a-mixing process and $K_{L,s}^{-1}\sum_{l=1}^{T}k_{L,sl}F_l-F_s=K_{L,s}^{-1}\sum_{l=1}^{T}k_{L,sl}\left(F_l-EF_s+EF_s-F_s\right)$ $=K_{L,s}^{-1}\sum_{l=1}^{T}k_{L,sl}\left(F_l-EF_s\right)+K_{L,s}^{-1}\sum_{l=1}^{T}k_{L,sl}\left(EF_s-F_s\right)=O_p\left(1\right)$ by Lemma 3 in Dendramis *et al.* (2021).

So, overall, by (127) we have that $K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\widehat{F}_{(1),t}-\widehat{F}_{(1),s}\right)x_s\left(\overline{y_{s+1}}-y_{s+1}\right)=o_p\left(1\right)$, implying that $W_t^{(1)}+W_t^{(2)}=o_p\left(1\right)$.

Since both $W_t^{(1)}+W_t^{(2)}=o_p\left(1\right)$ and $\sum_{i=1}^{8}S_t^{(i)}=o\left(1\right)$ we have that

$$\widehat{\beta}_t=\left(K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\left(\widehat{F}_{(1),t}\left(x_s-\overline{x}_t\right)\left(x_s-\overline{x}_t\right)'\widehat{F}_{(1),t}\right)\right)^{-1}\times\left(K_{L,t}^{-1}\sum_{s=1}^{T}k_{L,ts}\widehat{F}_{(1),t}\left(x_s-\overline{x}_t\right)\left(y_{s+1}-\overline{y}_t\right)'\right)+o_p\left(1\right).\tag{128}$$

In matrix form we can define

$$\widetilde{\widehat{F}}^{(t)\prime}=\widehat{F}_{(1),t}\left(k_{L,t1}^{1/2}\left(x_1-\overline{x}_t\right),..,k_{L,tT-1}^{1/2}\left(x_{T-1}-\overline{x}_t\right)\right).$$

Then,

$$\begin{aligned}\widehat{\beta}_t&=\left(\widetilde{\widehat{F}}^{(t)\prime}\widetilde{\widehat{F}}^{(t)}\right)^{-1}\widetilde{\widehat{F}}^{(t)\prime}\widetilde{y}^{(t)}+o_p\left(1\right)\\&=\left(K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\right)^{-1}N^{-1}K_{H,t}^{-2}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\widetilde{X}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\left(N^{-2}K_{H,t}^{-2}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_NK_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{X}_L^{(t)}J_N\widetilde{X}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\right)^{-1}\\&\quad\times N^{-1}K_{H,t}^{-2}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\widetilde{X}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\left(K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\right)^{-1}\\&\quad\times K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\left(N^{-1}K_{H,t}^{-2}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\widetilde{X}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\right)^{-1}N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_NK_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{y}_L^{(t)}\\&=\left(K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\right)^{-1}N^{-1}K_{H,t}^{-2}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\widetilde{X}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\left(N^{-2}K_{H,t}^{-2}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_NK_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{X}_L^{(t)}J_N\widetilde{X}_H^{(t)\prime}\widetilde{Z}_H^{(t)}\right)^{-1}\\&\quad\times N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_NK_{L,t}^{-1}\widetilde{X}_L^{(t)\prime}\widetilde{y}_L^{(t)}\\&=\widehat{\beta}_{1,t}^{-1}\widehat{\beta}_{2,t}\widehat{\beta}_{3,t}^{-1}\widehat{\beta}_{4,t}.\end{aligned}$$

For $\widehat{\beta}_1^{-1}$ and $\widehat{\beta}_2$ we have

$$\widehat{\beta}_{1,t}\underset{N,T\to\infty}{\overset{p}{\to}}\Lambda_t\Delta_F\Lambda_t'+\Delta_\omega,\tag{129}$$

$$\widehat{\beta}_{2,t}\underset{N,T\to\infty}{\overset{p}{\to}}\Lambda_t\Delta_F\Delta_\Phi\Delta_F\Lambda_t'.\tag{130}$$

For $\widehat{\beta}_3^{-1}$ and $\widehat{\beta}_4$ we have [6]

[6]The exact steps followed for $\widehat{\beta}_3^{-1}$ and $\widehat{\beta}_4$ are the same as in Kelly and Pruitt (2015).

$$\widehat{\beta}_{3,t} \underset{N,T\to\infty}{\overset{p}{\to}} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \Delta_\Phi \Delta_F \Lambda_t', \widehat{\beta}_{4,t} \underset{N,T\to\infty}{\overset{p}{\to}} \Lambda_t \Delta_F \Delta_\Phi \Delta_F \beta_t. \tag{131}$$

From the continuous mapping theorem, we have the desired result.

Lemma A6 *Let Assumptions 1-5 hold, and* $\widehat{\varepsilon}_t = x_t - \widehat{\phi}_{0t} - \widehat{\Phi}_t \widehat{F}_t$, *with* $\widehat{\phi}_{0t} = K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} x_s - \widehat{\Phi}_t \left(K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \widehat{F}_s \right)$. *Then, it holds that* $\widehat{\Phi}_t \widehat{F}_{H,t} \underset{N,T\to\infty}{\overset{p}{\to}} \Phi_t S_k F_t$ *and* $\widehat{\varepsilon}_t \underset{N,T\to\infty}{\overset{p}{\to}} \varepsilon_t + \Phi_{gt} g_t - K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \Phi_{gs} g_s$, *for* S_k *a* $K \times K$ *selector matrix that has ones in the first* K_f *main diagonal positions and zeros elsewhere.*

Proof of Lemma A6 By Lemma A4, Assumption 3 and 4, we have that

$$\widehat{\Phi}_t \widehat{F}_t \underset{N,T\to\infty}{\overset{p}{\to}} \Phi_t S_k F_t = \Phi_{ft} f_t, \tag{132}$$

and this implies the stated probability limit for $\widehat{\varepsilon}_t$, since

$$\begin{aligned}
\widehat{\varepsilon}_t &= x_t - K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} x_s + \widehat{\Phi}_t K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \widehat{F}_s - \widehat{\Phi}_t \widehat{F}_t \\
&= \phi_{0t} + \Phi_t F_t + \varepsilon_t - K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \phi_{0s} - K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \Phi_s F_s - K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \varepsilon_s \\
&+ \widehat{\Phi}_t K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \widehat{F}_s - \widehat{\Phi}_t \widehat{F}_t + K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \widehat{\Phi}_s \widehat{F}_s - K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \widehat{\Phi}_s \widehat{F}_s \\
&= \varepsilon_t + K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\phi_{0t} - \phi_{0s} \right) + \Phi_t F_t - \widehat{\Phi}_t \widehat{F}_t - K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \varepsilon_s \\
&+ K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\widehat{\Phi}_t - \widehat{\Phi}_s \right) \widehat{F}_s + K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\widehat{\Phi}_s \widehat{F}_s - \Phi_s F_s \right) \\
&\underset{N,T\to\infty}{\overset{p}{\to}} \varepsilon_t + \Phi_{gt} g_t - K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \Phi_{gs} g_s,
\end{aligned} \tag{133}$$

since $K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\phi_{0t} - \phi_{0s} \right) = O_p \left(\sqrt{\frac{H}{T}} \right)$ by Corollary 9b in Dendramis *et al.* (2021), $K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \varepsilon_s = o(1)$ by Lemma A1, $\Phi_t F_t - \widehat{\Phi}_t \widehat{F}_t \underset{N,T\to\infty}{\overset{p}{\to}} \Phi_{gt} g_t$, while for $G_t = \left(\Lambda_t \Delta_F \Lambda_t' + \Delta_\omega \right)$ we have

$$
\begin{aligned}
\widehat{\Phi}_t - \widehat{\Phi}_s &= \Phi_t \Delta_F \Lambda_t' G_t^{-1} - \Phi_s \Delta_F \Lambda_s' G_s^{-1} \\
&= \Phi_t \Delta_F \Lambda_t' G_t^{-1} - \Phi_t \Delta_F \Lambda_s' G_s^{-1} + \Phi_t \Delta_F \Lambda_s' G_s^{-1} - \Phi_s \Delta_F \Lambda_s' G_s^{-1} \\
&= \Phi_t \Delta_F \left(\Lambda_t' G_t^{-1} - \Lambda_s' G_s^{-1} \right) + \left(\Phi_t - \Phi_s \right) \Delta_F \Lambda_s' G_s^{-1} \\
&= \Phi_t \Delta_F \left(\Lambda_t' G_t^{-1} - \Lambda_t' G_s^{-1} + \Lambda_t' G_s^{-1} - \Lambda_s' G_s^{-1} \right) + \left(\Phi_t - \Phi_s \right) \Delta_F \Lambda_s' G_s^{-1} \\
&= \Phi_t \Delta_F \Lambda_t' \left(G_t^{-1} - G_s^{-1} \right) + \Phi_t \Delta_F \left(\Lambda_t' - \Lambda_s' \right) G_s^{-1} + \left(\Phi_t - \Phi_s \right) \Delta_F \Lambda_s' G_s^{-1}.
\end{aligned}
$$

Then,

$K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\widehat{\Phi}_t - \widehat{\Phi}_s \right) \widehat{F}_s$

$= K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\Phi_t \Delta_F \Lambda_t' \left(G_t^{-1} - G_s^{-1} \right) \right) \widehat{F}_s + K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\Phi_t \Delta_F \left(\Lambda_t' - \Lambda_s' \right) G_s^{-1} \right) \widehat{F}_s$

$+ K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\left(\Phi_t - \Phi_s \right) \Delta_F \Lambda_s' G_s^{-1} \right) \widehat{F}_s$

The above matrix is of $N \times K$ dimension and we will prove that each element $i = 1, .., N$ is $o_p(1)$. Notice that since $\|\phi_{mit}\|$, $\|\Lambda_t\|$, $\|\Delta_F\|$, $\left\| \widehat{F}_t \right\|$, $\left\| G_t^{-1} \right\| = O_p(1)$ for all t, the convergence to zero of the above sum is governed by the terms $d^{(1)} = K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(G_t^{-1} - G_s^{-1} \right)$, $d^{(2)} = K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\Lambda_t' - \Lambda_s' \right)$, and $d^{(3)} = K_{H,t}^{-1} \sum_{s=1}^{T} k_{H,ts} \left(\Phi_{mi,t} - \Phi_{mi,s} \right)$. Then, $d^{(1)} = o_p(1)$ by (56), $d^{(2)}$ and $d^{(3)}$ are $o_p(1)$ by Corollary 9b in Dendramis *et al.* (2021).

Lemma A7 *Let Assumptions 1, 2 and 5 hold. Then*

$$K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \widehat{F}_s \eta_{s+1} = o_p(1). \tag{134}$$

Proof of Lemma A7 By (105) notice that

$$
\begin{aligned}
\left\| K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \widehat{F}_s \eta_{s+1} \right\| &\leq K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \left\| \widehat{F}_{A,s} \widehat{F}_{B,s}^{-1} \widehat{F}_{C,s} \eta_{s+1} \right\| \leq \\
&\leq \max_t \left\| \widehat{F}_{A,t} \right\| \max_t \left\| \widehat{F}_{B,t}^{-1} \right\| \left\| K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} \widehat{F}_{C,s} \eta_{s+1} \right\| \\
&\leq C \left\| K_{L,t}^{-1} \sum_{s=1}^{T} k_{L,ts} F_s \eta_{s+1} \right\| + o(1). \\
&= o_p(1), \text{ as } N, T \to \infty
\end{aligned}
$$

Lemma A8 *Let Assumptions 1, 2 and 5 hold. Then*

$$K_{H,t}^{-1}K_{L,t}^{-1/2}N^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\widetilde{X}_L^{(t)\prime}\widetilde{\eta}_L^{(t)}\xrightarrow{d}N\left(0,\Lambda_t\Delta_F\Delta_\Phi\Gamma_{F\eta,t}\Delta_\Phi\Delta_F\Lambda_t^{\prime}\right), \tag{135}$$

where $\Gamma_{F\eta,t}$ *is defined in Theorem 3.*

Proof of Lemma A8

$$\begin{aligned}
K_{H,t}^{-1}K_{L,t}^{-1}N^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\widetilde{X}_L^{(t)\prime}\widetilde{\eta}_L^{(t)} &= \Lambda_t\left(K_{H,t}^{-1}\widetilde{F}_H^{(t)\prime}\widetilde{F}_H^{(t)}\right)\left(N^{-1}\Phi_t^{\prime}J_N\Phi_t\right)\left(K_{L,t}^{-1}\widetilde{F}_L^{(t)\prime}\widetilde{\eta}_L^{(t)}\right)\\
&+\Lambda_t\left(K_{H,t}^{-1}\widetilde{F}_H^{(t)\prime}\widetilde{F}_H^{(t)}\right)\left(N^{-1}K_{L,t}^{-1}\Phi_t^{\prime}J_N\widetilde{\varepsilon}_L^{(t)}\widetilde{\eta}_L^{(t)}\right)\\
&+\Lambda_t\left(N^{-1}K_{H,t}^{-1}\widetilde{F}_H^{(t)\prime}\widetilde{\varepsilon}_H^{(t)}J_N\Phi_t^{\prime}\right)\left(K_{L,t}^{-1}\widetilde{F}_L^{(t)\prime}\widetilde{\eta}_L^{(t)}\right)\\
&+\Lambda_tN^{-1}K_{H,t}^{-1}\widetilde{F}_H^{(t)\prime}\widetilde{\varepsilon}_H^{(t)}J_NK_{L,t}^{-1}\widetilde{\varepsilon}_L^{(t)}\widetilde{\eta}_L^{(t)}\\
&+\left(K_{H,t}^{-1}\widetilde{\omega}_H^{(t)\prime}\widetilde{F}_H^{(t)}\right)\left(N^{-1}\Phi_t^{\prime}J_N\Phi_t\right)\left(K_{L,t}^{-1}\widetilde{F}_L^{(t)\prime}\widetilde{\eta}_L^{(t)}\right)\\
&+\left(K_{H,t}^{-1}\widetilde{\omega}_H^{(t)\prime}\widetilde{F}_H^{(t)}\right)\left(N^{-1}K_{L,t}^{-1}\Phi_t^{\prime}J_N\widetilde{\varepsilon}_L^{(t)}\widetilde{\eta}_L^{(t)}\right)\\
&+\left(N^{-1}K_{H,t}^{-1}\widetilde{\omega}_H^{(t)\prime}\widetilde{\varepsilon}_H^{(t)}J_N\Phi_t^{\prime}\right)\left(K_{L,t}^{-1}\widetilde{F}_L^{(t)\prime}\widetilde{\eta}_L^{(t)}\right)\\
&+\left(N^{-1}K_{H,t}^{-1}\widetilde{\omega}_H^{(t)\prime}\widetilde{\varepsilon}_H^{(t)}J_NK_{L,t}^{-1}\widetilde{\varepsilon}_L^{(t)}\widetilde{\eta}_L^{(t)}\right).
\end{aligned}$$

For $K_t=\min\left(K_{L,t},K_{H,t}\right)$, we have that the above equals to

$$\begin{aligned}
&O_p\left(K_{L,t}^{-1/2}\right)+O_p\left(K_t^{-1/2}N^{-1/2}\right)+O_p\left(K_t^{-1/2}\delta_{N,K_t}^{-1}\right)+O_p\left(K_t^{-1/2}\delta_{N,K_t}^{-1}\right)\\
&+O_p\left(K_t^{-1}\right)+O_p\left(N^{-1/2}K_t^{-1}\right)+O_p\left(K_t^{-1}\right)+O_p\left(K_t^{-1/2}\delta_{N,K_t}^{-1}\right),
\end{aligned}$$

where the first term dominates and the stated asymptotic distribution is obtained by Theorem 3.

Lemma A9 *Let Assumptions 1, 2 and 5 hold. As* $N,T\to\infty$*, for* $S_{zx,H,t}=K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}$ *we have*
(i) if $\sqrt{N}/K_t\to 0$*, then for every t*

$$N^{-1/2}S_{zx,H,t}J_N\varepsilon_t\xrightarrow{d}N\left(0,\Lambda_t\Delta_F\Gamma_{F\varepsilon,t}\Delta_F\Lambda_t^{\prime}\right);$$

(ii) if $\liminf\sqrt{N}/K_t\geq\tau\geq 0$ *then*

$$N^{-1}\widetilde{Z}^{(t)\prime}\widetilde{X}^{(t)}J_N\varepsilon_t=O_p\left(1\right).$$

Proof of Lemma A9 We have

$$N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\varepsilon_t = N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N x_t - N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\left(\phi_{0t}+\Phi_t F_t\right) \tag{136}$$

$$= \widehat{F}_{C_{(1)},t}x_t - N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\left(\phi_{0t}+\Phi_t F_t\right),$$

where $\widehat{F}_{C_{(1)},t}x_t$ has been defined in Lemma A4. Also,

$$\begin{aligned}
N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\varepsilon_t &= \Lambda_t\left(K_{H,t}^{-1}\widetilde{F}_H^{(t)\prime}\widetilde{F}_H^{(t)}\right)\left(N^{-1}\Phi_t^{\prime}J_N\varepsilon_t\right)+ \\
&+\Lambda_t\left(N^{-1}K_{H,t}^{-1}\widetilde{F}_H^{(t)\prime}\widetilde{\varepsilon}_H^{(t)}J_N\varepsilon_t\right) \\
&+\left(K_{H,t}^{-1}\widetilde{\omega}_H^{(t)\prime}\widetilde{F}_H^{(t)\prime}\right)\left(N^{-1}\Phi_t^{\prime}J_N\varepsilon_t\right) \\
&+N^{-1}K_{H,t}^{-1}\widetilde{\omega}_H^{(t)\prime}\widetilde{\varepsilon}_H^{(t)}J_N\varepsilon_t \\
&= O_p\left(N^{-1/2}\right)+O_p\left(\delta_{N,K_t}^{-1}K_{H,t}^{-1/2}\right) \\
&+O_p\left(K_{H,t}^{-1/2}\right)O_p\left(N^{-1/2}\right) \\
&+O_p\left(\delta_{N,K_t}^{-1}K_t^{-1/2}\right),
\end{aligned}$$

by results in Lemma A2 and Theorem 3.

When $\frac{\sqrt{N}}{K_{H,t}} \to 0$, the first term determines the limiting distribution, in which case the result (i) is obtained by Theorem 3.

Then, if $\frac{\sqrt{N}}{K_{H,t}} \geq \tau \geq 0$, we have $K_{H,t}\left(N^{-1}K_{H,t}^{-1}\widetilde{Z}_H^{(t)\prime}\widetilde{X}_H^{(t)}J_N\varepsilon_t\right) = O_p(1)$, since $\liminf \frac{K_{H,t}}{\sqrt{N}} \leq \frac{1}{\tau} \leq \infty$.

7.5 Auxiliary Monte Carlo and Empirical Results

7.5.1 Further visualization of our Monte Carlo design

In Figure 4, we use a sample generated by equations (37)-(38) and analyse the contribution of the common factor component for the generated large dataset x_t. This is measured by the R^2 on the time (t) and cross sectional (i) dimension. As it becomes visible, our design generates samples with balanced factor contribution across time and cross section, while the κ parameter of (37) can efficiently control the overall contribution of the common component.

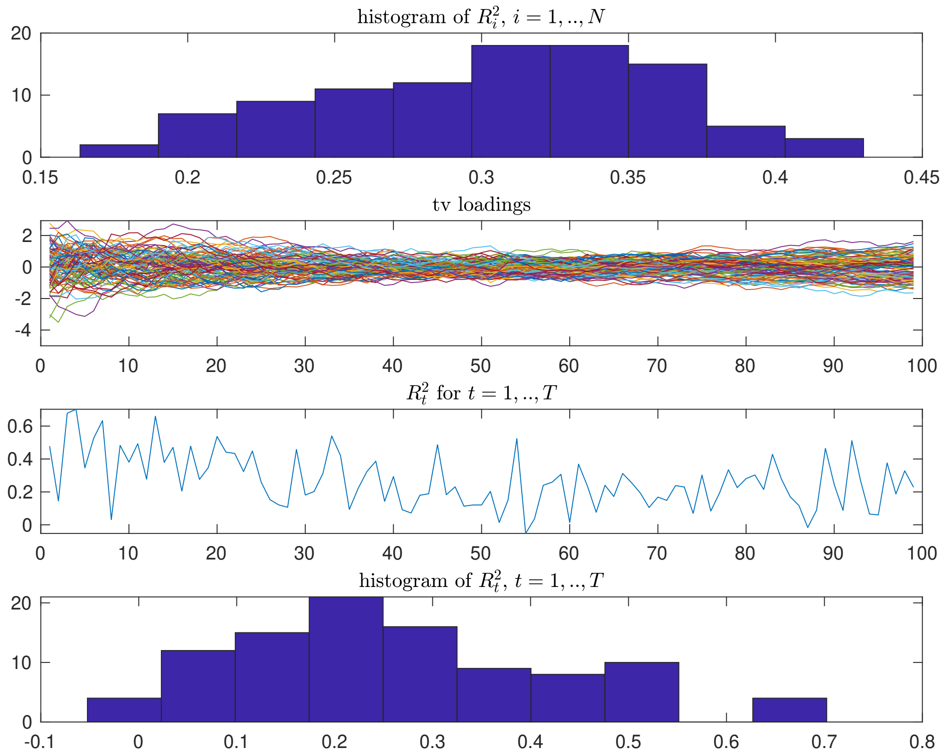

Figure 4: One realization of the data generated by equations (37)-(38). We report the histogram of R_i^2 for all generated series of the large dataset, $i = 1, .., N$, when the loadings process generated by model (42) and the parameter κ of (37) is adjusted to have an $R_i^2 = .3$. The time t R_t^2, is computed using cross sectional data of size N for all $t = 1, .., T$. More specifically, the R_t^2 is measured as $R_t^2 = \frac{\widehat{Var_t(\phi_{f,t} f_t + \phi_{g,t} g_t)}}{\widehat{Var_t(x_{it})}}$

7.5.2 Further Monte Carlo Results

In this section we present additional evidence on the performance of the Information Criterion for selecting the number of factors.

Number of factors Selected by IC, DGP given by (43)									
ρ_f	0	0.3	0.3	0.3	0.3	0.9	0.9	0.9	0.9
ρ_g	0	0.9	0.9	0.9	0.9	0.3	0.3	0.3	0.3
β	0	0	1	0	1	0	1	0	1
a	0	0.3	0.3	0.9	0.9	0.3	0.3	0.9	0.9
R^2 =0.1									
N = 100, T = 100	1.2	1.08	1.06	1.09	1.13	1.03	1.05	1.06	1.08
N = 200, T = 200	1.03	1.02	1.01	1.04	1.05	1.03	1.01	1.02	1.02
N = 200, T = 100	2.47	1.44	1.11	1.1	1.11	1.19	1.05	1.04	1.05
N = 100, T = 200	1.03	1.02	1.03	1.06	1.06	1.08	1.01	1.02	1.04
R^2 =0.3, tv loadings									
N = 100, T = 100	1.17	1.1	1.1	1.14	1.13	1.05	1.04	1.03	1.05
N = 200, T = 200	1.05	1.05	1.03	1.05	1.05	1.02	1.03	1.01	1.02
N = 200, T = 100	1.34	1.14	1.1	1.09	1.1	1.12	1.06	1.02	1.03
N = 100, T = 200	1.06	1.04	1.04	1.07	1.08	1.03	1.03	1.03	1.03

Table 8: Average number of factors selected for the tv loadings process (43)

7.5.3 Additional Visualizations and Tables for the Empirical Section

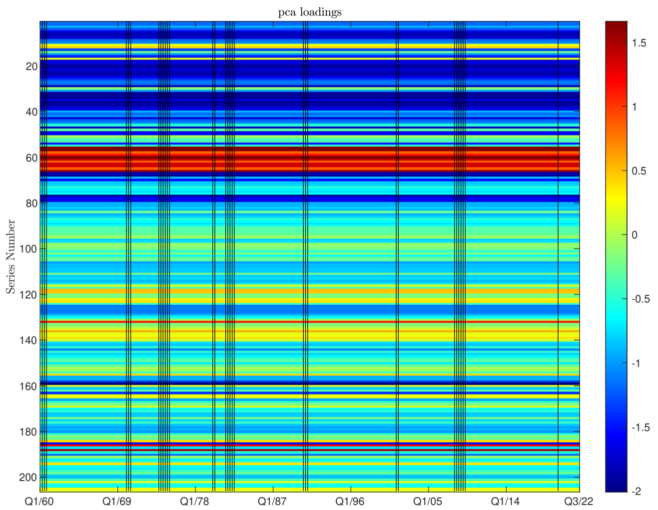

Figure 5: Loadings implied by the PCA method. The vertical stripes indicate periods of NBER recession.

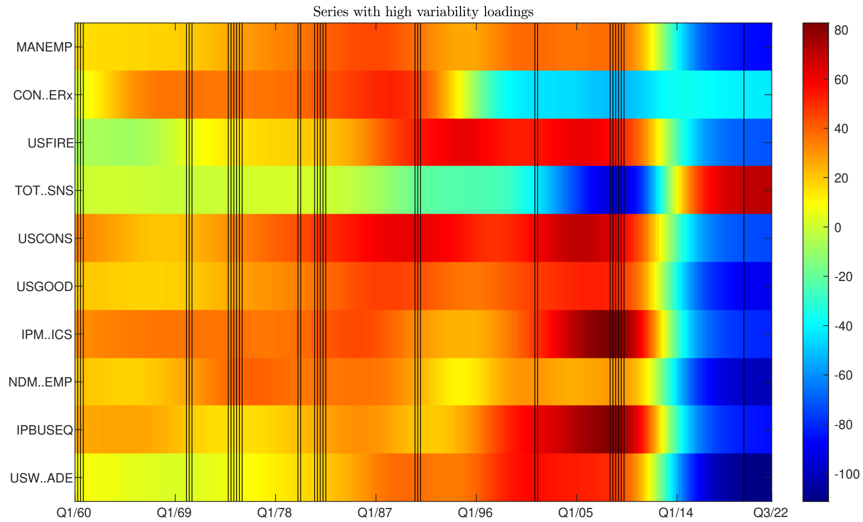

Figure 6: Loadings with the highest variability across time. The vertical stripes indicate periods of NBER recession.

Series with high variability loadings	
MANEMP	All Employees: Manufacturing (Thousands of Persons)
CONSUMERx	Real Consumer Loans at All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE
USFIRE	All Employees: Financial Activities (Thousands of Persons)
TOTRESNS	Total Reserves of Depository Institutions (Billions of Dollars)
USCONS	All Employees: Construction (Thousands of Persons)
USGOOD	All Employees: Goods-Producing Industries (Thousands of Persons)
IPMANSICS	Industrial Production: Manufacturing (SIC) (Index 2012=100)
NDMANEMP	All Employees: Nondurable goods (Thousands of Persons)
IPBUSEQ	Industrial Production: Business Equipment (Index 2012=100)
USWTRADE	All Employees: Wholesale Trade (Thousands of Persons)

Table 9: Series associated with loadings of high variability

Out of Sample Macroeconomic Forecasting:up to 2019Q4										
h	1	4	1	4	1	4	1	4	1	4
	TBILL		IP-DCG		AE-WT		FEDFUNDS		PCE-FSI	
$pca(ic)$	1.58	0.9	**0.79**	1.2	0.99	0.93	1.61	1.08	0.93	1.06
$pca(1)$	**0.94**	0.65	1.07	0.95	1.04	1.07	1	0.72	0.95	0.81
$pca\text{-}lars(1)$	1.07	0.81	0.86	1.06	0.99	1.02	1.34	**0.76**	0.93	**0.92**
$pca\text{-}ht(10\%)(1)$	**0.93**	**0.75**	1.11	**0.97**	1	1.03	0.98	**0.81**	**0.89**	**0.84**
$3prf(1)$	1	1	1	1	1	**1**	1	1	1	1
$tv\text{-}3prf(1)$	1$_{\diamond\diamond}$	**0.74**	**0.8**	1.24	0.92	1.09$_{\diamond}$	**0.88**	0.83	0.88	1.07
$tv\text{-}3prf(ic)$	0.99$_{\diamond\diamond}$	1.19$_{\diamond\diamond}$	0.92$_{\diamond\diamond}$	**0.99**	**0.98**	**0.97**$_{\diamond\diamond}$	**0.92**$_{\diamond\diamond}$	1.05$_{\diamond\diamond}$	1.01	0.97$_{\diamond\diamond}$
$tv\text{-}3prf_{H_i}(1)$	0.84$_{\diamond\diamond}$	0.79	0.78	1.28	**0.92**	1.07	0.85	0.88	**0.89**	1.08
	PFI		CPI-LFE		IP-M		UNRATE		AHEP	
$pca(ic)$	1.05	1.1	1.3	0.89	0.95	**0.98**	**0.99**	1.12	0.95	0.92
$pca(1)$	0.9	1.01	**0.78**	0.7	1.12	0.99	1.26	**0.99**	0.8	0.87
$pca\text{-}lars(1)$	1.01	1.06	0.97	**0.81**	1	1.19	1.01	1.23	**0.83**	**0.9**
$pca\text{-}ht(10\%)(1)$	**0.92**	0.95	**0.81**	**0.77**	1.04	1.03	1.34	0.98	**0.89**	0.93
$3prf(1)$	1	**1**	1	1	1	1	**1**	1	1	1
$tv\text{-}3prf(1)$	0.96	1.25	0.83$_{\diamond\diamond}$	0.86	0.85	0.98	0.98	1.16	0.9	**0.91**
$tv\text{-}3prf(ic)$	**0.95**$_{\diamond}$	**1**	0.67$^{*}_{\diamond\diamond}$	0.96$_{\diamond\diamond}$	**0.92**$^{**}_{\diamond\diamond}$	0.93$_{\diamond\diamond}$	1.01	**0.99**	0.9$^{**}_{\diamond\diamond}$	0.91$_{\diamond\diamond}$
$tv\text{-}3prf_{H_i}(1)$	0.96	1.19	0.84$_{\diamond\diamond}$	0.88$_{\diamond}$	**0.85**	**0.97**$_{\diamond}$	1	1.21	0.95	0.91
	CP		CLAIMS		HOUST		CUMFNS		GCE	
$pca(ic)$	1.56	1.06	**0.94**	1.1	1.04	1.2	0.98	**0.94**	0.88	**0.98**
$pca(1)$	**0.94**	0.78	1.06	0.98	**0.93**	**0.87**	1.38	1.26	0.83	0.97
$pca\text{-}lars(1)$	1.12	0.96	1.04	1.05	1.04	0.8	**0.92**	1.09	**0.85**	**0.98**
$pca\text{-}ht(10\%)(1)$	**0.95**	**0.85**	1.11	**1**	0.94	**0.9**	0.94	1	0.87	0.98
$3prf(1)$	1	1	1	1	1	1	1	1	1	1
$tv\text{-}3prf(1)$	0.99$_{\diamond\diamond}$	0.94	**0.88**	1.14	0.8	1.06	0.87	**0.93**$_{\diamond}$	0.91	1.16
$tv\text{-}3prf(ic)$	0.96$_{\diamond\diamond}$	1.06$_{\diamond\diamond}$	0.98	**0.99**	0.96	0.99	0.96$_{\diamond}$	0.97$_{\diamond\diamond}$	**0.85****	1.04
$tv\text{-}3prf_{H_i}(1)$	0.93$_{\diamond\diamond}$	**0.9**	0.87	1.18	**0.82**	1.04	**0.88**	**0.96**$_{\diamond\diamond}$	0.93	1.13
	PAYEMS		EXPORTS		PCE-FE		GS10		EXR-SUS	
$pca(ic)$	1.01	0.94	1.1	1.04	1.09	1.11	1.02	**0.89**	0.94	0.97
$pca(1)$	1.12	1.07	0.99	0.93	0.91	0.97	0.85	**0.83**	0.88	0.9
$pca\text{-}lars(1)$	**0.92**	1.05	1.02	**0.95**	1.13	1.04	**0.87**	0.81	**0.88**	**0.9**
$pca\text{-}ht(10\%)(1)$	0.99	1.02	1.01	**0.95**	**0.93**	**0.97**	**0.88**	0.91	**0.93**	**0.92**
$3prf(1)$	1	**1**	1	1	1	**1**	1	1	1	1
$tv\text{-}3prf(1)$	**0.93**	1.01$_{\diamond\diamond}$	0.94	1.12	**0.98**	1.15	1	0.9	1.18	1.19
$tv\text{-}3prf(ic)$	0.85$^{**}_{\diamond\diamond}$	**0.95**$_{\diamond\diamond}$	**0.97**	1.01	0.98	1.01	0.98	0.98$_{\diamond}$	1.13	1.12
$tv\text{-}3prf_{H_i}(1)$	0.96	1.04$_{\diamond\diamond}$	**0.94**	1.12$_{\diamond\diamond}$	0.99	1.16	0.94	0.94	1.17	1.21
	UNRATE		GDP		M2REAL		EXR-USUK		FPI	
$pca(ic)$	0.89	1.16	0.9	1.01	0.98	1.1	0.95	**1**	0.96	1.18
$pca(1)$	1.27	1.12	1.32	0.95	0.97	0.87	0.87	0.96	1.08	0.97
$pca\text{-}lars(1)$	1.14	0.98	**0.94**	1.09	**0.97**	1.01	**0.91**	1	0.85	1.05
$pca\text{-}ht(10\%)(1)$	**1.01**	**0.99**	1.2	1.1	**0.97**	**0.9**	**0.93**	**0.98**	1.11	1.01
$3prf(1)$	**1**	1	1	**1**	1	1	1	1	1	**1**
$tv\text{-}3prf(1)$	1.01	1.17	1.23	1.08$_{\diamond}$	1.27	**0.91**$_{\diamond\diamond}$	1.12$_{\diamond\diamond}$	1.1	**0.88**	1.26
$tv\text{-}3prf(ic)$	1.01	**0.99**	**0.99**	**0.99**	1.06	0.95$_{\diamond\diamond}$	1.05	1.04	0.94	**0.99**
$tv\text{-}3prf_{H_i}(1)$	1.06	1.24	1.13	1.1	1.29	0.95$_{\diamond\diamond}$	1.19	1.11	**0.89**	1.2

Table 10: See notes in Table 6. Forecasting stops at 2019Q4

Department of Economics
Athens University of Economics and Business

List of Recent Working Papers

2021

01-21 Historical Cycles of the Economy of Modern Greece From 1821 to the Present, George Alogoskoufis
02-21 Greece Before and After the Euro: Macroeconomics, Politics and the Quest for Reforms, George Alogoskoufis
03-21 Commodity money and the price level, George C. Bitros. Published in: *Quarterly Journal of Austrian Economics*, 2022
04-21 Destabilizing asymmetries in central banking: With some enlightenment from money in classical Athens, George C. Bitros. Published in: *Journal of Economic Asymmetries*, 2021
05-21 Exploring the Long-Term Impact of Maximum Markup Deregulation, Athanasios Dimas and Christos Genakos
06-21 A regularization approach for estimation and variable selection in high dimensional regression models, Y. Dendramis, L. Giraitis, G. Kapetanios
07-21 Tax Competition in the Presence of Environmental Spillovers, Fabio Antoniou, Panos Hatzipanayotou, Michael S. Michael, Nikos Tsakiris
08-21 Firm Dynamics by Age and Size Classes and the Choice of Size Measure, Stelios Giannoulakis and Plutarchos Sakellaris
09-21 Measuring the Systemic Importance of Banks, Georgios Moratis, Plutarchos Sakellaris
10-21 Firms' Financing Dynamics Around Lumpy Capacity Adjustments, Christoph Görtz, Plutarchos Sakellaris, John D. Tsoukalas
11-21 On the provision of excludable public goods General taxes or user prices? George Economides and Apostolis Philippopoulos
12-21 Asymmetries of Financial Openness in an Optimal Growth Model, George Alogoskoufis
13-21 Evaluating the impact of labour market reforms in Greece during 2010-2018 , Georgios Gatopoulos, Alexandros Louka, Ioannis Polycarpou, Nikolaos Vettas
14-21 From the Athenian silver to the bitcoin standard: Private money in a state-enforced free banking model, George C. Bitros
15-21 Ordering Arbitrage Portfolios and Finding Arbitrage Opportunities. Stelios Arvanitis and Thierry Post
16-21 Inconsistency for the Gaussian QMLE in GARCH-type models with infinite variance, Stelios Arvanitis and Alexandros Louka
17-21 Competition and Pass-Through: Evidence from Isolated Markets, Christos Genakos and Mario Pagliero
18-21 Exploring Okun's Law Asymmetry: An Endogenous Threshold LSTR Approach, Dimitris Christopoulos, Peter McAdam and Elias Tzavalis
19-21 Limit Theory for Martingale Transforms with Heavy-Tailed Multiplicative Noise, Stelios Arvanitis and Alexandros Louka
20-21 Optimal taxation with positional considerations, Ourania Karakosta and Eleftherios Zacharias

<u>2022</u>

<u>2023</u>

06-23 What makes for survival? Key characteristics of Greek incubated early-stage startup(per)s during the Crisis: a multivariate and machine learning approach, Ioannis Besis, Ioanna Sapfo Pepelasis and Spiros Paraskevas
07-23 The Twin Deficits, Monetary Instability and Debt Crises in the History of Modern Greece, George Alogoskoufis
08-23 Dealing with endogenous regressors using copulas; on the problem of near multicollinearity, Dimitris Christopoulos, Dimitris Smyrnakis and Elias Tzavalis
09-23 A machine learning approach to construct quarterly data on intangible investment for Eurozone, Angelos Alexopoulos and Petros Varthalitis
10-23 Asymmetries in Post-War Monetary Arrangements in Europe: From Bretton Woods to the Euro Area, George Alogoskoufis, Konstantinos Gravas and Laurent Jacque
11-23 Unanticipated Inflation, Unemployment Persistence and the New Keynesian Phillips Curve, George Alogoskoufis and Stelios Giannoulakis
12-23 Threshold Endogeneity in Threshold VARs: An Application to Monetary State Dependence, Dimitris Christopoulos, Peter McAdam and Elias Tzavalis
13-23 A DSGE Model for the European Unemployment Persistence, Konstantinos Giakas
14-23 Binary public decisions with a status quo: undominated mechanisms without coercion, Efthymios Athanasiou and Giacomo Valletta
15-23 Does Agents' learning explain deviations in the Euro Area between the Core and the Periphery? George Economides, Konstantinos Mavrigiannakis and Vanghelis Vassilatos
16-23 Mild Explocivity, Persistent Homology and Cryptocurrencies' Bubbles: An Empirical Exercise, Stelios Arvanitis and Michalis Detsis
17-23 A network and machine learning approach to detect Value Added Tax fraud, Angelos Alexopoulos, Petros Dellaportas, Stanley Gyoshev, Christos Kotsogiannis, Sofia C. Olhede, Trifon Pavkov

Department of Economics
Athens University of Economics and Business

The Department is the oldest Department of Economics in Greece with a pioneering role in organising postgraduate studies in Economics since 1978. Its priority has always been to bring together highly qualified academics and top quality students. Faculty members specialize in a wide range of topics in economics, with teaching and research experience in world-class universities and publications in top academic journals.

The Department constantly strives to maintain its high level of research and teaching standards. It covers a wide range of economic studies in micro-and macroeconomic analysis, banking and finance, public and monetary economics, international and rural economics, labour economics, industrial organization and strategy, economics of the environment and natural resources, economic history and relevant quantitative tools of mathematics, statistics and econometrics.

Its undergraduate program attracts high quality students who, after successful completion of their studies, have excellent prospects for employment in the private and public sector, including areas such as business, banking, finance and advisory services. Also, graduates of the program have solid foundations in economics and related tools and are regularly admitted to top graduate programs internationally. Three specializations are offered:1. Economic Theory and Policy, 2. Business Economics and Finance and 3. International and European Economics. The postgraduate programs of the Department (M.Sc and Ph.D) are highly regarded and attract a large number of quality candidates every year.

For more information:

https://www.dept.aueb.gr/en/econ/

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5