Efficiency in large dynamic panel models with common factor

Patrick Gagliardini, Christian Gourieroux

EFFICIENCY IN LARGE DYNAMIC PANEL MODELS WITH COMMON FACTOR

Patrick GAGLIARDINI* and Christian GOURIEROUX†

First version: August 2008
This version: February 2009

*University of Lugano and Swiss Finance Institute.
†CREST, CEPREMAP (Paris) and University of Toronto.

Acknowledgements: We thank A. Monfort for helpful comments. The first author gratefully acknowledges financial support of the Swiss National Science Foundation through the NCCR FINRISK network. The second author gratefully acknowledges financial support of NSERC Canada and of the chair AXA/Risk Foundation: "Large Risks in Insurance".

Efficiency in Large Dynamic Panel Models with Common Factor

Abstract

This paper deals with efficient estimation in exchangeable nonlinear dynamic panel models with common unobservable factor. The specification accounts for both micro- and macro-dynamics, induced by the lagged individual observation and the common stochastic factor, respectively. For large cross-sectional and time dimensions, and under a semi-parametric identification condition, we derive the efficiency bound and introduce efficient estimators for both the micro- and macro-parameters. In particular, we show that the fixed effects estimator of the micro-parameter is not only consistent, but also asymptotically efficient. The results are illustrated with the stochastic migration model for credit risk analysis.

Keywords: Nonlinear Panel Model, Factor Model, Exchangeability, Systematic Risk, Efficiency Bound, Semi-parametric Efficiency, Fixed Effects Estimator, Bayesian Statistics, Stochastic Migration, Granularity.

1 Introduction

This paper considers efficient estimation in nonlinear dynamic panel models with common unobservable factor. We focus on exchangeable specifications that are appropriate to analyze a large homogeneous population of individuals featuring rich patterns of serial and cross-sectional dependence. Such a framework is encountered in credit risk applications, where the panel data are the rating histories of a large pool of firms in a given industrial sector and country [1]. The common factor represents a latent macro-variable, such as the sector and country specific business cycle, that introduces dependence across the rating dynamics of the firms, the so-called migration correlation. The purpose is to predict the future risk in a large portfolio of corporate bonds or credit derivatives issued by the firms in the pool.

The model involves both a micro- and a macro-dynamic. Conditional on a given factor path, the individuals are assumed independent and identically distributed, with observations y_{it}, t varying, following a same time-inhomogeneous Markov process for any individual i. The transition density $h(y_{it}|y_{i,t-1}, f_t; \beta)$ at date t depends on the factor value f_t and the unknown parameter β. The micro-dynamic is captured by the lagged individual observation $y_{i,t-1}$ and unknown parameter β. The macro-dynamic is driven by the time-varying stochastic common factor f_t. The latter is unobservable and follows a Markov process with transition density $g(f_t|f_{t-1}; \theta)$, which depends on the unknown parameter θ. When this common factor is integrated out, it introduces both non-Markovian serial dependence within the individual histories, and cross-sectional dependence between individuals.

When the cross-sectional dimension n is fixed and the time dimension T tends to infinity, the Maximum Likelihood (ML) estimators of micro-parameter β and macro-parameter θ are asymptotically normal and efficient. However, this asymptotic scheme is not appropriate for a setting involving very large n and moderately large T, as in credit risk applications. For instance, for corporate rating data the number of firms is typically of order $n \simeq 10,000$, while the number of dates is about $T \simeq 20$ with yearly data. For mortgage

[1] This framework is also encountered in the securitization of a pool of loans (Collateralized Debt Obligations, CDO), or insurance contracts [Insurance Linked Securities (ILS) and longevity bonds].

data, we typically have $n \simeq 100,000 - 1,000,000$ mortgages and $T \simeq 200$ months.

The aim of this paper is to derive the efficiency bound for estimating both the micro-parameter β and the macro-parameter θ, when $n, T \to \infty$ and $T^b/n = O(1)$, for $b > 1$. The derivation has to account for the different rates of increasing information concerning the two types of parameters. First, we show that the efficiency bound for micro-parameter β does not depend on the parametric model defining the macro-dynamic. In particular, this bound coincides with the efficiency bound with known transition of the factor, and also with the semi-parametric efficiency bound when the transition of the factor is left unspecified. Second, a consistent and (semi-)parametrically efficient estimator of the micro-parameter is the ML estimator of β computed as if the factor values are fixed time effects. To get the intuition for these findings, it is useful to remark that our specification with random time effects can be seen as a Bayesian approach, with prior $\prod_{t=1}^{T} g(f_t|f_{t-1};\theta)$ on the factor values [2]. The results above provide an example of the well-known asymptotic equivalence of frequentist and Bayesian methods in large sample, implying the irrelevance of the prior choice. Third, an efficient estimator of the macro-parameter θ is the ML estimator computed by replacing the unobservable factor values with consistent cross-sectional approximations.

In Section 2 we introduce the nonlinear dynamic panel model with common factor. This model includes the Single Risk Factor (SRF) model suggested for the regulation of credit risk in Basel 2. Then, we explain why our specification is not simply a panel model with fixed effects, as usually considered in the econometric literature. The efficiency bound is derived in Section 3. The derivation is based on an asymptotic expansion of the log-likelihood function. For this purpose, the integration of the latent factor is performed along the lines of the Laplace approximation [Jensen (1995)]. If the micro-parameter is semi-parametrically identified, we show that the efficiency bound for micro-parameter β is independent of the parametric specification of the factor dynamics. In Section 4 we introduce efficient estimators of both parameters, that do not involve numerical integration w.r.t. the unobservable factor. We first show that the fixed effects estimator of the micro-parameter is efficient. This estimator is used to derive consistent approximations $\hat{f}_t$ of the factor values. Then, we show that the estimator of the macro-parameter derived from maximizing

[2]See Aigner et al. (1984) for a discussion of this interpretation in a general latent variable setting.

the macro-likelihood after substitution of the factor values f_t by their approximations $\hat{f}_t$ is efficient. Finally, we discuss the link with the granularity adjustment introduced in Pillar 2 of the Basel 2 regulation. In Section 5, the results of the paper are applied to the stochastic migration model used for credit risk analysis. In this model, the observable endogenous variable corresponds to the rating and the common stochastic factor accounts for migration correlation. The patterns of the efficiency bound, and the computation of the efficient estimators are discussed for this example. Section 6 concludes. The proofs of the results are gathered in Appendices A.1-A.7. They include in particular an appropriate Weak Law of Large Numbers (WLLN) and a large deviation theorem for maximum likelihood estimators (see Appendices A.1 and A.2). The regularity conditions are listed in Appendix A.3. The proofs of the technical Lemmas are given in Appendix B on the web-site http://www.istituti.usilu.net/gagliarp/proofsPANEL.htm.

2 Exchangeable nonlinear panel model with common factor

2.1 The model

Let us consider panel data y_{it} for a large homogeneous population of individuals $i = 1, ..., n$ observed at dates $t = 1, ..., T$. We assume a nonlinear dynamic specification with common factor such that:

A.1: *Conditional on a factor path (f_t), the individual histories (y_{it}), $i = 1, ..., n$, are i.i.d. time-inhomogeneous Markov processes of order 1, with transition pdf $h\left(y_{i,t}|y_{i,t-1}, f_t; \beta\right)$ and unknown parameter $\beta \in \mathcal{B}$, where $\mathcal{B} \subset \mathbb{R}^q$.*

A.2: *The factor (f_t) is a Markov process of order 1 in $\mathbb{R}^K$, with transition pdf $g(f_t|f_{t-1}; \theta)$ and unknown parameter $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^p$.*

We denote by β_0 and θ_0 the true values of parameters β and θ, respectively. The common factor f_t is unobservable and has to be integrated out to derive the joint density of observations y_{it}. The latent factor introduces both non-Markovian individual dynamics and dependence across individuals. The distribution is exchangeable, i.e. symmetric w.r.t. the individuals [3]. The focus is on the efficient estimation of both micro-parameter β and macro-parameter θ [4].

Next Assumptions A.3, A.4 and A.5 concern the stationarity and ergodicity properties of the model.

A.3: *The process* $(y_{1,t}, ..., y_{n,t}, f_t)$ *is strictly stationary, for any* $n \in \mathbb{N}$.

A.4: *The process* (f_t) *is strong mixing.*

A.5: *Conditional on the factor path* (f_t), *process* $(y_{i,t})$ *is ergodic and strong mixing with* α*-mixing coefficients* $\alpha_h\left[(f_t)\right], h \in \mathbb{N}$, *for any path* (f_t) P*-a.s., such that* $E\left[\alpha_h\left[(f_t)\right]\right] \leq c_1 \exp\left(-c_2 h^{1/\delta}\right)$, *as* $h \to \infty$, *for some constants* $c_1, c_2, \delta > 0$.

Assumption A.5 requires that the individual processes $(y_{i,t})$ are ergodic and strong mixing, conditional on the factor path. The conditional mixing coefficients can depend on the factor path, but their expectation converges to zero at an exponential rate as lag h increases. Assumption A.5 is similar in spirit to the work in Granger (1980), Granger, Joyeux (1980), Bougerol, Picard (1992). The geometric decay of the integrated mixing coefficients implies that the initial values of the $y_{i,t}$'s have no effect in the long run even after integrating out the factors. Assumptions A.3-A.5 are used to study the asymptotic behavior of nonlinear aggregates of the type:

$$\frac{1}{T}\sum_{t=1}^{T}\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(y_{i,t}, f_t, \beta)\right),$$

as $n, T \to \infty$ such that $T/n \to 0$, where a is a matrix-valued function of individual observation $y_{i,t}$, factor value f_t and micro-parameter β, and φ is a continuous mapping. The precise asymptotic results are provided in Appendix 1. These results are used to derive the asymptotic properties of the estimators introduced in Section 4 [5].

[3]The exchangeability is equivalent to the existence of a factor representation [see e.g. de Finetti (1931), Hewitt, Savage (1955)].

[4]Exchangeable linear panel models are considered in Hjellwig, Tjostheim (1999) and Hansen, Nielsen, Nielsen (2004).

[5]The stationarity and ergodicity Assumptions A.3-A.5 for asymptotic analysis with $n, T \to \infty, T/n \to 0$,

2.2 The Single Risk Factor (SRF) model

The specification introduced in Section 2.1 is motivated by the SRF model introduced by Vasicek (1987), (1991), and recommended for the analysis of credit risk in the second Pillar of Basel 2 [BCBS (2001)], concerning internal models. The objective is to analyze the risk of a portfolio of loans or credit derivatives, included in the balance sheet of a bank or credit institution. These portfolios contain several millions of individual assets and have to be segmented into subportfolios, which are homogeneous by the type of contract (asset) and by the type of borrowers, including at least their ratings among their characteristics. The SRF model is applied to these homogeneous subportfolios separately. The sizes of these subportfolios are still rather large including some 10 thousands of individual loans for mortgages and credit cards, for instance.

The basic Vasicek model is written for firms, but the same approach is applicable to consumers. This model introduces the asset $A_{i,t}$ and liability $L_{i,t}$ as latent variables. Then, the latent model is written on the log-ratio of asset to liability $y^*_{i,t} = \log(A_{i,t}/L_{i,t})$ as:

$$y^*_{i,t} = \alpha + \beta F_t + \sigma u_{i,t},\ t = 1, ..., T, \quad i \in \mathcal{P}_t,$$

where $\mathcal{P}_t$ denotes the set of firms in the portfolio, which are still alive at time t (called Population-at-Risk), and where the common factor F_t and the idiosyncratic factors $u_{i,t}$ are independent standard Gaussian variables. The sensitivity coefficients α, β, σ are independent of the individuals, according to the definition of an ”homogeneous” portfolio. The observed endogenous variable is the default occurrence:

$$y_{i,t} = \mathbb{1}_{A_{i,t}<L_{i,t}} = \mathbb{1}_{y^*_{i,t}<0}.$$

We deduce the probability of default conditional on the common factor:

$$PD_t = P\left[y_{i,t} = 1 | y_{i,t-1} = 0, F_t\right] = \Phi\left[-\left(\alpha/\sigma\right) - \left(\beta/\sigma\right) F_t\right].$$

differ significantly from the hypotheses used with finite n [e.g., see Douc, Moulines, Rydén (2004) for the asymptotic properties of the ML estimator in autoregressive models with Markov regimes, that correspond to the case $n = 1$]. This is because the estimators depend on cross-sectional aggregates of the type $\frac{1}{n}\sum_{i=1}^{n} a(y_{i,t}, f_t, \beta)$, which are functions of the factor path $\underline{f_t}$ but not of the individual observations $y_{i,t}$, $i = 1, \cdots, n$, when $n \to \infty$ (see Appendix 1).

The observed default occurrences are independent with Bernoulli distribution $y_{i,t} \sim \mathcal{B}(1, PD_t)$, conditional on the common factor. This basic model can be extended in various ways by allowing for a dynamics of the common factor, or for a joint analysis of more than two rating levels by means of stochastic migration models describing the transitions between rating classes AAA, AA, ... (see Section 5). The advantage of this specification is to distinguish the idiosyncratic risks $u_{i,t}$, which can be diversified, and the undiversifiable systematic risk F_t.

Finally note that the marginal probability of default is $PD = \Phi\left(-\alpha/\sqrt{\beta^2+\sigma^2}\right)$, whereas the default correlation between any two firms i and j is:

$$\rho = Corr\left(y_{i,t}, y_{j,t}\right) = \frac{\Psi\left(-\alpha/\sqrt{\beta^2+\sigma^2}, -\alpha/\sqrt{\beta^2+\sigma^2}; \rho^*\right) - PD^2}{PD(1-PD)},$$

where $\rho^* = \sqrt{\beta^2/\left(\beta^2+\sigma^2\right)}$ is the correlation between the log asset-to-liability ratios, and $\Psi(.,.;\rho^*)$ denotes the joint cdf of the bivariate standard Gaussian distribution with correlation coefficient ρ^*. In the new regulation, the required capital depends on the values of PD and ρ, that is, indirectly on the values α, β, σ, and is especially sensitive to default correlation. This explains the importance of a simple, robust and efficient estimation of parameter ρ.

2.3 The panel model with fixed effects

The econometric literature on nonlinear panel models with fixed effects [see e.g. Hahn, Newey (2004)] considers specifications such that the variables $y_{i,t}$, $i = 1, 2, ..., n$, $t = 1, ..., T$, are independent with pdf $f\left(y_{i,t}; \alpha_i, \beta\right)$, where α_i is the fixed effects of individual i [6]. The focus of this literature is on the correction of the bias of the ML estimator of β caused by the incidental parameters problem [Neyman, Scott (1948); see also Lancaster (2000) for a review]. The model introduced in Section 2.1 can be seen as a model with fixed time effects instead of fixed individual effects. However, there are important differences between our setting and the fixed effect panel literature:

i) In practice n is much larger than T, and therefore the incidental parameter problem is much less pronounced with fixed time effects than with fixed individual effects. In

[6] See Hahn, Kuersteiner (2004), Arellano, Bonhomme (2006) for extensions to a dynamic setting.

particular, the bias corrections are less important in our setting and even not required if $T/n \to 0$.

ii) The nonlinear panel model with common factor in Section 2.1 is clearly a time series model introduced for prediction purpose. For instance, the SRF model of Basel 2 (Section 2.2) is the basis for determining the distribution of the future portfolio value and the corresponding 1% quantile, called CreditVaR. At the opposite, a model with fixed individual effects is used to get a segmentation of the population in order to get homogeneous segments, i.e. with similar α_i values. For instance, in the credit risk problem, the models with fixed individual effects are typically used to get the homogeneous subportfolios, whereas the SRF model is written for each homogeneous subportfolio to analyse jointly the evolution of their risks.

iii) As a consequence, we are also interested in the filtering of the factor values, in their dynamics, that is, in macro-parameter θ, and in their interpretations.

3 Efficiency bound

3.1 The likelihood function

The joint density of $\underline{y_T} = (y_{i,t}, t = 1, ..., T, i = 1, ..., n)$ and $\underline{f_T} = (f_t, t = 1, ..., T)$ is given by:

$$\begin{aligned} l\left(\underline{y_T}, \underline{f_T}; \beta, \theta\right) &= \prod_{i=1}^{n} \prod_{t=1}^{T} h\left(y_{i,t} | y_{i,t-1}, f_t; \beta\right) \prod_{t=1}^{T} g(f_t | f_{t-1}; \theta) \\ &= l\left(\underline{y_T} | \underline{f_T}; \beta\right) l\left(\underline{f_T}; \theta\right), \text{ (say).} \end{aligned}$$

The density of $\underline{y_T}$ is obtained by integrating out factors $\underline{f_T}$:

$$\begin{aligned} l\left(\underline{y_T}; \beta, \theta\right) &= \int \cdots \int \prod_{t=1}^{T} \prod_{i=1}^{n} h\left(y_{i,t} | y_{i,t-1}, f_t; \beta\right) \prod_{t=1}^{T} g(f_t | f_{t-1}; \theta) \prod_{t=1}^{T} df_t \qquad (3.1) \\ &= \int \cdots \int \exp\left\{ \sum_{t=1}^{T} \sum_{i=1}^{n} \log h\left(y_{i,t} | y_{i,t-1}, f_t; \beta\right) \right\} \prod_{t=1}^{T} g(f_t | f_{t-1}; \theta) \prod_{t=1}^{T} df_t. \end{aligned}$$

This likelihood function involves an integral with a large dimension increasing with T, which complicates the analytical study of the maximum likelihood estimators. However, for large n, the integral with respect to the factor values can be approximated by expanding the integrand around its maximum w.r.t. the factor, along the lines of the Laplace approximation [see e.g. Jensen (1995) for the general setting, and Arellano, Bonhomme (2006) for the use of Laplace approximation in panel models with fixed individual effects]. This expansion yields an integrand of a Gaussian micro-dynamic model. Specifically, let us define for any β the cross-sectional ML estimator of the factor value [7]:

$$\hat{f}_{n,t}(\beta) = \arg\max_{f_t} \sum_{i=1}^{n} \log h\left(y_{i,t}|y_{i,t-1}, f_t; \beta\right). \tag{3.2}$$

Proposition 1. *Under the regularity Assumptions H.1-H.19 in Appendix A.3, the joint density of $\left(\underline{y_T}\right)$ is such that:*

$$\begin{aligned} l\left(\underline{y_T}; \beta, \theta\right) &= \left(\frac{2\pi}{n}\right)^{TK/2} \prod_{t=1}^{T} \left[\det I_{nt}\left(\beta\right)\right]^{-1/2} \prod_{t=1}^{T}\prod_{i=1}^{n} h\left(y_{i,t}|y_{i,t-1}, \hat{f}_{nt}\left(\beta\right); \beta\right) \\ &\quad \prod_{t=1}^{T} g\left(\hat{f}_{nt}\left(\beta\right)|\hat{f}_{n,t-1}\left(\beta\right); \theta\right) \exp\left[\frac{T}{n}\Psi_{nT}\left(\beta, \theta\right)\right], \end{aligned}$$

where:

$$I_{nt}\left(\beta\right) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f_t \partial f_t'}\left(y_{i,t}|y_{i,t-1}, \hat{f}_{nt}\left(\beta\right); \beta\right),$$

$\sup_{\beta\in\mathcal{B},\theta\in\Theta} \Psi_{nT}\left(\beta, \theta\right) = O_p(1)$ *as* $n, T \to \infty$ *such that* $T^b/n = O(1)$, $b > 1$, *and the probability order* O_p *is w.r.t. the true distribution.*

Proof. See Appendix 5. □

From Proposition 1 we deduce an expansion for the (nT-standardized) log-likelihood function of the sample:

$$\mathcal{L}_{nT}\left(\beta, \theta\right) = \frac{1}{nT}\log l\left(\underline{y_T}; \beta, \theta\right).$$

Corollary 2. *The (nT-standardized) log-likelihood function is such that:*

$$\mathcal{L}_{nT}\left(\beta, \theta\right) = \mathcal{L}^*_{nT}\left(\beta\right) + \frac{1}{n}\mathcal{L}_{1,nT}\left(\beta, \theta\right) + \frac{1}{n^2}\mathcal{L}_{2,nT}\left(\beta, \theta\right), \tag{3.3}$$

[7] From a mathematical point of view (see Appendix A.4), the cross-sectional ML estimator $\hat{f}_{n,t}(\beta)$ is defined by optimizing on a well-chosen compact set $\mathcal{F}_n$ generating the entire set $\mathbb{R}^K$, when $n \to \infty$.

where:

$$\mathcal{L}^*_{nT}(\beta) = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1}, \hat{f}_{nt}(\beta);\beta\right), \tag{3.4}$$

$$\mathcal{L}_{1,nT}(\beta,\theta) = -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\log\det I_{nt}(\beta) + \frac{1}{T}\sum_{t=1}^{T}\log g\left(\hat{f}_{nt}(\beta)\,|\hat{f}_{n,t-1}(\beta);\theta\right), \tag{3.5}$$

and $\mathcal{L}_{2,nT}(\beta,\theta) = \Psi_{nT}(\beta,\theta)$.

Function $\mathcal{L}^*_{nT}(\beta)$, called profile log-likelihood function, is the log-likelihood of β concentrated w.r.t. the factor values, as if the latter are nuisance parameters. In Corollary 2, the profile log-likelihood function $\mathcal{L}^*_{nT}(\beta)$ is the leading term in an asymptotic expansion of the log-likelihood function $\mathcal{L}_{nT}(\beta,\theta)$ in powers of $1/n$. The transition density of the factor enters in the term $\mathcal{L}_{1,nT}(\beta,\theta)$ at asymptotic order $1/n$, and is expected to be irrelevant for the efficiency bound of β when $n \to \infty$ (see Section 3.2 for a precise statement). These results are an example of the asymptotic equivalence of frequentist and Bayesian methods in large sample. To get the main intuition, let T be fixed for a moment. Then, our specification with stochastic common factor can be seen as a Bayesian approach w.r.t. to the time effects parameters, with prior density $\prod_{t=1}^{T} g(f_t|f_{t-1};\theta)$ [8]. As the cross-sectional dimension n tends to infinity, it is known from Bayesian statistics that the posterior distribution of the parameter f_t, scaled by $\sqrt{n}$, approaches a normal distribution centered at the ML estimator $\hat{f}_{nt}(\beta)$, for given parameter β. This is why the "Bayesian" posterior density function for β given in (3.1) corresponds, up to a scale factor, to the joint density of $\left(\underline{y_T}\right)$ and $\left(\underline{f_T}\right)$ with f_t replaced by $\hat{f}_{nt}(\beta)$, $t = 1, ..., T$. The irrelevance of the second term in the RHS of (3.3) involving the transition density of the factor corresponds to the irrelevance of the prior distribution in large sample. Thus, the Bayesian log-likelihood $\mathcal{L}_{nT}(\beta,\theta)$ approaches the log-likelihood $\mathcal{L}^*_{nT}(\beta)$, which is the "frequentist" log-likelihood for β concentrated w.r.t. parameters f_t, $t = 1, ..., T$. Our results show that this asymptotic equivalence remains true when $n, T \to \infty$ such that $T^b/n \to 0$, $b > 1$. The additional term in $\mathcal{L}_{1,nT}(\beta,\theta)$ involves the determinant of the Hessian matrix $I_{nt}(\beta)$, which is the Jacobian for a change of variable performed in the Laplace approximation (see the proof of Proposition 1). The term $I_{nt}(\beta)$ corresponds to the term introduced by Cox and Reid (1987) in their modified profile

[8] This prior depends on "hyperparameter" θ and is independent of parameter β.

likelihood to correct the likelihood function after concentration w.r.t. nuisance (incidental) parameters. For the derivation of the semiparametric efficiency bound, the term involving $I_{nt}(\beta)$ is irrelevant when $n \to \infty$ under the semi-parametric identification conditions given below [9].

3.2 Efficiency bound

The ML estimator $\left(\hat{\beta}, \hat{\theta}\right)$ is defined by:

$$\left(\hat{\beta}, \hat{\theta}\right) = \arg\max_{\beta,\theta} \mathcal{L}_{nT}(\beta, \theta). \tag{3.6}$$

Under the regularity conditions listed in Appendix 3, we prove in Appendix 6 that the ML estimator is asymptotically normal:

$$\begin{bmatrix} \sqrt{nT}\left(\hat{\beta} - \beta_0\right) \\ \sqrt{T}\left(\hat{\theta} - \theta_0\right) \end{bmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} B^*_{\beta\beta} & B^*_{\beta\theta} \\ B^*_{\theta\beta} & B^*_{\theta\theta} \end{pmatrix}\right),$$

with different rates of convergence for the micro- and macro-component, that are root-nT and root-T, respectively. The asymptotic variance-covariance matrix $B^* = \begin{pmatrix} B^*_{\beta\beta} & B^*_{\beta\theta} \\ B^*_{\theta\beta} & B^*_{\theta\theta} \end{pmatrix}$ defines the efficiency bound for estimating (β, θ).

To compute the efficiency bound, let us introduce the large sample counterparts of the likelihood terms in the RHS of (3.3).

(i) Let us first consider $\mathcal{L}^*_{nT}(\beta)$. We can define at each date t the pseudo-true factor value:

$$f_t(\beta) = \arg\max_{f} E_0\left[\log h\left(y_{it}|y_{i,t-1}, f; \beta\right)|\underline{f_t}\right],$$

where $E_0\left[.|\underline{f_t}\right]$ denotes the expectation w.r.t. the true conditional distribution of $(y_{i,t}, y_{i,t-1})$ at date t given $\underline{f_t} = \{f_t, f_{t-1}, ...\}$. This function yields the factor value $f_t(\beta)$ that maximizes the limiting cross-sectional log-likelihood at date t, for any given parameter value β. It corresponds to the population counterpart of $\hat{f}_{n,t}(\beta)$ in (3.2) when $n \to \infty$. The pseudo-true factor value $f_t(\beta)$ is a function of both parameter β and information $\underline{f_t}$. Moreover, by

[9] In his discussion of the Cox and Reid (1987) paper, Sweeting (1987) suggests that this correction term can be derived in a Bayesian setting by integrating the nuisance parameters and using a Laplace approximation.

the properties of the Kullback-Leibler discrepancy at true parameter value β_0, the pseudo-true factor value $f_t(\beta_0)$ coincides with the true factor value f_t, P-a.s., for any t. Then, let us define the function:

$$\begin{aligned} \mathcal{L}^*(\beta) &= \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E_0\left[\log h\left(y_{it}|y_{i,t-1}, f_t(\beta);\beta\right)|\underline{f_t}\right] \\ &= E_0\left[\log h\left(y_{it}|y_{i,t-1}, f_t(\beta);\beta\right)\right]. \end{aligned}$$

The assumptions below concern the identification of parameter β.

A.6 (Global semi-parametric identification assumption for β): *The mapping $\beta \to \mathcal{L}^*(\beta)$ is uniquely maximized at the true parameter value β_0.*

A.7 (Local semi-parametric identification assumption for β): *The matrix $I_0^* = -\dfrac{\partial^2\mathcal{L}^*(\beta_0)}{\partial\beta\partial\beta'}$ is positive definite.*

The matrix I_0^* is given by (see Appendix 6.2):

$$I_0^* = E_0\left[I_{\beta\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}I_{f\beta}(t)\right], \tag{3.7}$$

where $I_{\beta\beta}(t)$, $I_{ff}(t)$, $I_{\beta f}(t)$ and $I_{f\beta}(t) = I_{\beta f}(t)'$ denote the blocks of the conditional information matrix at date t:

$$I(t) = E_0\left[-\frac{\partial^2\log h\left(y_{it}|y_{i,t-1}, f_t;\beta_0\right)}{\partial\left(\beta', f'\right)'\partial\left(\beta', f'\right)}|\underline{f_t}\right]. \tag{3.8}$$

Assumptions A.6 and A.7 correspond to identification conditions for parameter β in a semi-parametric setting, in which the transition of the factor f_t is left unconstrained and is treated as an infinite-dimensional parameter. This interpretation is justified by the fact that the criterion $\mathcal{L}^*(\beta)$ is the large sample counterpart of the profile likelihood function $\mathcal{L}^*_{nT}(\beta)$ in (3.4), that is, the likelihood of β concentrated w.r.t. "parameters" f_t, $t = 1, ..., T$. When Assumptions A.6 and A.7 are not met, the identification of parameter β relies on the parametric model $g(f_t|f_{t-1};\theta)$ for the transition of the factor. Intuitively, we would have to distinguish the transformations of vector β that are identified by criterion $\mathcal{L}^*(\beta)$, and the transformations of β that are identified only with the contribution of the parametric model $g(f_t|f_{t-1};\theta)$. This would induce different rates of convergence for these transformations, that are $1/\sqrt{nT}$ and $1/\sqrt{T}$, respectively. The in-depth analysis of this general setting is beyond the scope of this paper.

(ii) Let us now consider the time series component $\mathcal{L}_{1,nT}(\beta,\theta)$ of the log-likelihood. Under Assumptions A.6-A.7 parameter β can be estimated at a rate infinitely faster than θ and the relevant criterion for identification of θ is the mapping $\theta \to \mathcal{L}_1(\beta_0,\theta)$, where $\mathcal{L}_1(\beta_0,\theta)$ is the large sample limit of $\mathcal{L}_{1,nT}(\beta,\theta)$ in (3.5) for $\beta=\beta_0$. We have $\mathcal{L}_1(\beta_0,\theta) = E_0\left[\log g(f_t|f_{t-1};\theta)\right]$, up to a term constant in θ. Thus, the identification assumptions for the macro-parameter are the following:

A.8 (Global identification assumption for θ): *The mapping $\theta \to E_0\left[\log g(f_t|f_{t-1};\theta)\right]$ is uniquely maximized at the true parameter value θ_0.*

A.9 (Local identification assumption for θ): *The matrix $I_{1,\theta\theta} = E_0\left[-\frac{\partial^2 \log g\left(f_t|f_{t-1};\theta_0\right)}{\partial\theta\partial\theta'}\right]$ is positive definite.*

Assumptions A.8 and A.9 correspond to the standard global and local identification conditions for estimating θ in a model with observable factor values.

Proposition 3. *Under Assumptions A.1-A.9 and H.1-H.19, and if $n,T\to\infty$ such that $T^b/n = O(1)$, $b>1$, the efficiency bound for (β,θ) is:*

$$B^* = \begin{pmatrix} B^*_{\beta\beta} & B^*_{\beta\theta} \\ B^*_{\theta\beta} & B^*_{\theta\theta} \end{pmatrix} = \begin{pmatrix} (I_0^*)^{-1} & 0 \\ 0 & I_{1,\theta\theta}^{-1} \end{pmatrix},$$

where:

$$I_0^* = E_0\left[I_{\beta\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} I_{f\beta}(t)\right],$$

and

$$I_{1,\theta\theta} = E_0\left[-\frac{\partial^2 \log g\left(f_t|f_{t-1};\theta_0\right)}{\partial\theta\partial\theta'}\right].$$

Proof. See Appendix 6. □

The zero out-of-diagonal blocks in the efficiency bound imply that parameters β and θ can be considered independently for estimation purpose. This justifies ex-post their interpretation as micro- and macro-parameters, respectively, since parameter β contain no macro-information under Assumptions A.6-A.7. The result in Proposition 3 is a consequence of the expansion of the likelihood function in Corollary 2. Indeed, under identification Assumptions A.6-A.7 and regularity conditions (see Appendix 3), for large n and T the relevant term for estimation of parameter β is $\mathcal{L}^*_{nT}(\beta)$. The corresponding limit log-likelihood function is $\mathcal{L}^*(\beta)$, and the efficiency bound $B^*_{\beta\beta}$ for β is the inverse of the Hessian $I_0^* = -\frac{\partial^2 \mathcal{L}^*(\beta_0)}{\partial\beta\partial\beta'}$. Similarly, the efficiency bound $B^*_{\theta\theta}$ for θ is the inverse of the Hessian $I_{1,\theta\theta} = -\frac{\partial^2 \mathcal{L}_1(\beta_0,\theta_0)}{\partial\theta\partial\theta'}$. Moreover, the (standardized) ML estimators of β and θ are asymptotically independent. Therefore, the efficiency bound $B^*_{\beta\beta}$ for β given in Proposition 3 is the same as the efficiency bound for β with known transition of the factor. Finally, matrix I_0^* in (3.7) is smaller than the information $I_0^{**} = E_0\left[I_{\beta\beta}(t)\right]$ corresponding to the case of observable factor, while matrix $I_{1,\theta\theta}$ is equal to the information for θ with observable factor. Therefore, the unobservability of the factor has no efficiency impact asymptotically for estimating θ, but has an impact for estimating β. This is due to the fact that the factor values can be estimated at rate $1/\sqrt{n}$ (see Section 4.2), a rate which is infinitely faster than the rate $1/\sqrt{T}$ for estimating θ, if $T^b/n = O(1)$, $b > 1$, and infinitely slower than the rate $1/\sqrt{nT}$ for estimating β.

The efficiency bound $B^*_{\beta\beta}$ for parameter β in Proposition 3 is independent of the parametric model $g(f_t|f_{t-1};\theta)$, $\theta \in \mathbb{R}^p$, for the transition of the factor, that is factor distribution free. This suggests that the efficiency result extends to a semi-parametric setting. Specifically, the asymptotic semi-parametric efficiency bound B for β is the efficiency bound for estimating β in the semi-parametric model in which the transition $g(f_t|f_{t-1})$ of the factor is a functional parameter. The semi-parametric efficiency bound B can be computed by using Stein's heuristic. More precisely, let $g_\theta = g(f_t|f_{t-1};\theta)$ be a well-specified parametric model for the transition of f_t with parameter $\theta \in \mathbb{R}^p$ that satisfies Assumptions A.8-A.9 and the regularity conditions H.16-H.19 in Appendix 3, and let $B^*_{\beta\beta}(g_\theta)$ be the corresponding parametric efficiency bound for estimating β.

Definition 1. *The semi-parametric efficiency bound B is defined by:*

$$B = \max_{g_\theta} B^*_{\beta\beta}(g_\theta),$$

where the maximization is performed w.r.t. the well-specified parametric models g_θ for the *transition of f_t that satisfy Assumptions A.8-A.9 and H.16-H.19.*

The result in Proposition 3 shows that $B^*_{\beta\beta}(g_\theta)$ is independent of g_θ. Therefore we deduce:

Corollary 4. *Under Assumptions A.1-A.7 and H.1-H.15, and if $n, T \to \infty$ such that $T^b/n = O(1)$, $b > 1$, the semi-parametric efficiency bound is equal to the parametric efficiency bound:*

$$B = B^*_{\beta\beta} = E_0 \left[I_{\beta\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} I_{f\beta}(t) \right]^{-1}.$$

Thus, any well-specified parametric model g_θ is the least-favorable one in the sense of Chamberlain (1987). The results in Proposition 3 and Corollary 4 show that the knowledge of the parametric model for the transition of the factor, and even the knowledge of the transition itself, are irrelevant for the efficient estimation of micro-parameter β.

3.3 Identification in the SRF model

The SRF model of Section 2.2 is such that $y_{i,t}|F_t \sim \mathcal{B}(1, \Phi[-(\alpha/\sigma) - (\beta/\sigma) F_t])$ and the observations can be summarized by means of the sufficient statistics $\widehat{PD}_t = \dfrac{1}{n}\sum_{i=1}^{n} y_{i,t}$, that are the cross-sectional default frequencies. In a semi-parametric framework, in which the transition of the factor is left unspecified, the micro-parameters α/σ and β/σ are not semi-parametrically identified. The initial factor can be replaced by $f_t = \Phi[-(\alpha/\sigma) - (\beta/\sigma) F_t] = PD_t$, and the model becomes $y_{i,t} \sim \mathcal{B}(1, f_t)$. This corresponds to a degenerate model, which has no longer micro-parameters. The factor values are approximated by $\hat{f}_t = \widehat{PD}_t$. Nevertheless, the default correlation is still semi-parametrically identified and can be consistently estimated at order $1/\sqrt{T}$ by [Gagliardini, Gouriéroux (2005a)]:

$$\hat{\rho} = \frac{\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \left(\widehat{PD}_t - \widehat{PD}\right)^2}{\widehat{PD}\left(1 - \widehat{PD}\right)},$$

where $\widehat{PD} = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \widehat{PD}_t$. Of course, the micro-parameters α/σ and β/σ can be identified when a parametric specification for the factor dynamics is introduced. For instance, the

SRF model considered by Basel 2 is identifiable due to the assumption that the factor values F_t are independent standard normal. We see in Section 5 that the semi-parametric identification of micro-parameters is recovered either when more than two rating levels are considered, or in a two-state framework without absorbing state.

4 Efficient estimators and granularity adjustment

In this Section we introduce asymptotically efficient estimators of the micro- and macro-parameters that are easier to compute than the ML estimators. These estimators rely on the asymptotic expansion of the log-likelihood funtion and do not involve the numerical integration w.r.t. the unobservable factor.

4.1 The fixed effects estimator of the micro-parameter

The asymptotic expansion of the likelihood function in Corollary 2, and the derivation of the efficiency bound in Proposition 3, suggest that the (semi-)parametric efficiency bound for β can be achieved by maximizing the likelihood function $\mathcal{L}_{nT}^{*}(\beta)$, i.e. by computing the fixed effects estimator which considers the f_t values as additional unknown parameters.

Proposition 5. *Under Assumptions A.1-A.7 and H.1-H.15, and if $n, T \to \infty$ such that $T^b/n = O(1)$, $b > 1$, the estimator:*

$$\hat{\beta}_{nT}^{*} = \arg\max_{\beta} \sum_{t=1}^{T} \sum_{i=1}^{n} \log h\left(y_{i,t} | y_{i,t-1}, \hat{f}_{nt}(\beta); \beta\right),$$

is consistent, root-nT asymptotically normal and (semi-)parametrically efficient.

Proof. See Appendix 6. □

The semi-parametric estimator $\hat{\beta}_{nT}^{*}$ achieves the same asymptotic efficiency as a parametric estimator that uses the information on the true transition of (f_t). It is computed by maximizing the likelihood function for β concentrated w.r.t. the factor values. Proposition 5 completes the standard analysis of the incidental parameter problem. If $T \to \infty$ and n is fixed, the fixed effects estimator $\hat{\beta}_{nT}^{*}$ is not consistent. If $n, T \to \infty$ and $T/n \to c > 0$

(say), the fixed effects estimator is consistent, but not efficient [10]. It becomes efficient if $n, T \to \infty$ such that $T^b/n = O(1)$, $b > 1$.

4.2 Approximation of the factor values

The efficient estimator $\hat{\beta}^*_{nT}$ can be used to derive cross-sectional approximations of the factor values [11]. A consistent approximation of the factor value at date t is:

$$\hat{f}_{nT,t} = \hat{f}_{n,t}\left(\hat{\beta}^*_{nT}\right).$$

This approximation tends to f_t at rate $1/\sqrt{n}$. More precisely, we have:

Proposition 6. *Suppose Assumptions A.1-A.7 and H.1-H.15 hold, and let $n, T \to \infty$ such that $T^b/n = O(1)$, $b > 1$. Then:*

i) For any date t, conditional on $\underline{f_t}$ we have:

$$\sqrt{n}\left(\hat{f}_{nT,t} - f_t\right) \xrightarrow{d} N\left(0, I_{ff}(t)^{-1}\right).$$

ii) $\sup_{1 \le t \le T} \left\| \hat{f}_{nT,t} - f_t \right\| = O_p\left(\sqrt{\frac{\log(n)^a}{n}}\right)$, *where* $a = 2a_1 + a_2 + a_3$, *and* $a_1, a_2, a_3 > 0$ *are defined in Assumptions H.12-H.14 in Appendix A.3.*

Proof. See Appendix 6. □

For any given date t, the factor approximation $\hat{f}_{nT,t}$ converges to the true factor value f_t at rate $1/\sqrt{n}$. Since $\hat{\beta}^*_{nT}$ is root-nT consistent, estimator $\hat{f}_{nT,t}$ is asymptotically equivalent to the unfeasible ML estimator $\hat{f}_{n,t}(\beta_0)$ for known micro-parameter β_0. The asymptotic variance $I_{ff}(t)^{-1}$ of $\hat{f}_{nT,t}$ is the inverse of the Fisher information for estimating f_t in the

[10] In such a framework, the bias is negligible with respect to the stochastic term of the expansion. Any crude penalization approach used to eliminate the bias at order $1/n$ [see e.g. Arellano, Hahn (2006), Woutersen (2002), Bester, Hansen (2005), Arellano, Bonhomme (2006)] will have an effect on the dominant stochastic term and generally induce a loss of efficiency.

[11] Consistent approximations of factor values in panel data with large cross-sectional and time dimensions have been proposed in, e.g., Forni, Reichlin (1998), Bai, Ng (2002), Stock, Watson (2002), Forni, Hallin, Lippi, Reichlin (2004), Connor, Hagmann, Linton (2007). All these papers consider linear factor models for the micro-dynamics.

cross-section at date t with known β_0. Since the factor value f_t is stochastic, the convergence rate for $\sup_{1\leq t\leq T}\left\|\hat{f}_{nT,t}-f_t\right\|$ has to be adjusted by a logarithmic factor. Proposition 6 ii) is derived by using a large deviation bound for ML estimators conditional on the factor path $\underline{f_t}$, and then integrating out $\underline{f_t}$ [see also Appendix A.4 for a general result on the uniform convergence rate of $\hat{f}_{n,t}(\beta)$].

4.3 Efficient estimator of the macro-parameter

The consistent approximations of the factor values $\hat{f}_{nT,t}$ can be used to derive an approximation of the macro-likelihood function:

$$\sum_{t=1}^{T}\log g\left(\hat{f}_{nT,t}|\hat{f}_{nT,t-1};\theta\right).$$

By maximizing this approximate likelihood w.r.t. θ, we get an efficient estimator of the macro-parameter.

Proposition 7. *Under Assumptions A.1-A.9 and H.1-H.19, and if $n,T\to\infty$ such that $T^b/n=O(1)$, $b>1$, the estimator:*

$$\hat{\theta}_{nT}=\arg\max_{\theta}\sum_{t=1}^{T}\log g\left(\hat{f}_{nT,t}|\hat{f}_{nT,t-1};\theta\right),$$

is root-T asymptotically normal and efficient.

Proof. This follows from Proposition 6 ii) by using Assumption H.16-H.19, condition $T^b/n=O(1)$, $b>1$, and standard asymptotic arguments for extremum estimators [see Connor, Hagmann, Linton (2007) for a similar result in a semi-parametric model with linear factor structure and nonlinear factor dynamics]. $\square$

Estimator $\hat{\theta}_{nT}$ is asymptotically equivalent to the unfeasible ML estimator $\hat{\theta}_T^{**}=\arg\max_{\theta}\sum_{t=1}^{T}\log g\left(f_t|f_{t-1};\theta\right)$ that uses the true factor values. As already noted in Section 3, replacing the true factor values by their root-n consistent approximations has no effect asymptotically for estimating θ at rate root-T, if $T^b/n=O(1)$, $b>1$. Since Propositions 5 and 7 show that estimators $\hat{\beta}^*_{nT}$ and $\hat{\theta}_{nT}$ achieve the efficiency bounds for parameters β and θ, respectively, then the joint estimator $\left(\hat{\beta}^*_{nT},\hat{\theta}_{nT}\right)$ is also asymptotically efficient [see Gouriéroux, Monfort (1995)].

4.4 Granularity adjustment

Let us now discuss the relationship between the estimators of the micro- and macro-parameters derived in Sections 4.1 and 4.3, respectively, and the granularity adjustments introduced for Pillar 2 of the Basel 2 regulation [see e.g. Gordy, Lutkebohmert (2007)]. The estimators $\left(\hat{\beta}^*_{nT}, \hat{\theta}_{nT}\right)$ in Propositions 5 and 7 are asymptotically equivalent to the estimators $\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right)$ obtained by maximizing the approximate log-likelihood function:

$$\mathcal{L}^{\text{CSA}}_{nT}(\beta, \theta) = \mathcal{L}^*_{nT}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta, \theta),$$

which admits a closed form expression (that is, without integrals w.r.t. the factor values). These estimators are called cross-sectional asymptotic (CSA) estimators in the recent literature on granularity adjustment [Gouriéroux, Jasiak (2008)]. The expansion can also be considered up to order $1/n^2$. This expansion provides a more accurate approximation of the log-likelihood function:

$$\mathcal{L}^{\text{GA}}_{nT}(\beta, \theta) = \mathcal{L}^*_{nT}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta, \theta) + \frac{1}{n^2}A_{nT}(\beta, \theta), \tag{4.1}$$

where A_{nT} is given in (A.6) in Appendix 5. This second-order approximation of the likelihood function admits also a closed form expression, and its optimization provides a more accurate approximation of the unfeasible ML estimator. This estimator, called granularity adjusted (GA) ML estimator, is defined by:

$$\left(\tilde{\beta}^{\text{GA}}_{nT}, \tilde{\theta}^{\text{GA}}_{nT}\right) = \underset{\beta,\theta}{\text{argmax}}\mathcal{L}^{\text{GA}}_{nT}(\beta, \theta). \tag{4.2}$$

It is easily checked that an estimator asymptotically equivalent to the GAML estimator up to order $1/n^2$ is:

$$\begin{aligned}\begin{pmatrix} \hat{\beta}^{\text{GA}}_{nT} \\ \hat{\theta}^{\text{GA}}_{nT} \end{pmatrix} = \begin{pmatrix} \hat{\beta}^*_{nT} \\ \hat{\theta}_{nT} \end{pmatrix} &+ \left(-\frac{\partial^2 \mathcal{L}^{\text{CSA}}_{nT}\left(\hat{\beta}^*_{nT}, \hat{\theta}_{nT}\right)}{\partial(\beta', \theta')'\partial(\beta', \theta')}\right)^{-1} \left[\frac{1}{n}\frac{\partial \mathcal{L}_{1,nT}\left(\hat{\beta}^*_{nT}, \hat{\theta}_{nT}\right)}{\partial(\beta', \theta')'}\right. \\ &\left. + \frac{1}{n^2}\frac{\partial A_{nT}\left(\hat{\beta}^*_{nT}, \hat{\theta}_{nT}\right)}{\partial(\beta', \theta')'}\right].\end{aligned}$$

The difference between the GAML estimators and the estimators $\left(\hat{\beta}^*_{nT}, \hat{\theta}_{nT}\right)$ gives the closed form expression of the granularity adjustment.

5 Stochastic migration model

5.1 The model

The stochastic migration model has been introduced to analyze the dynamics of corporate ratings and is a basic element for the prediction of future credit risk in a homogeneous pool of credits [e.g., Gupton et al (1997), Gordy, Heitfield (2002), Gagliardini, Gouriéroux (2005b), Feng et al (2008)]. A basic stochastic migration model is the ordered qualitative model with one factor, which extends the SRF model of Section 2.2 to more than two alternatives. Let us denote by $y_{i,t}$, t varying, the sequence of ratings for corporate i. The possible ratings are $k = 1, 2, ..., K$, say [12]. The micro-dynamic model specifies the transition matrices with elements depending on the factor value:

$$\pi_{lk,t} = P\left[y_{i,t} = k | y_{i,t-1} = l, f_t\right] = G\left(\frac{a_k - \alpha_l f_t - \gamma_l}{\sigma_l}\right) - G\left(\frac{a_{k-1} - \alpha_l f_t - \gamma_l}{\sigma_l}\right),$$

where $a_1 < a_2 < ... < a_{K-1}$ and $\alpha_l, \gamma_l, \sigma_l$, $l = 1, ..., K$ are unknown micro-parameters, and $a_0 = -\infty$, $a_K = +\infty$. Function G is the cdf of a probability distribution, that corresponds to the standard normal distribution for the Probit model, where $G(x) = \Phi(x)$, and to the logistic distribution for the Logit model, where $G(x) = 1/\left(1 + e^{-x}\right)$. The ratios $\left(a_k - \alpha_l f_t - \gamma_l\right)/\sigma_l$ in the above transition probabilities allow to identify semiparametrically the micro-parameters and the factor values up to location and scale transformations. For semiparametric identification (see Assumptions A.6-A.7), we impose the constraints $a_1 = 0$, $\sigma_1 = 1$, $\gamma_1 = 0$, $\alpha_1 = 1$ when $K > 2$, and additionally $\sigma_2 = 1$ when $K = 2$ (see Appendix 7.1).

5.2 Estimation of the micro-parameters

The micro log-density is given by:

$$\begin{aligned}
&\log h\left(y_{it} | y_{i,t-1}, f_t; \beta\right) \\
&= \sum_{k=1}^{K} \sum_{l=1}^{K} 1\left\{y_{i,t} = k, y_{i,t-1} = l\right\} \log\left[G\left(\frac{a_k - \alpha_l f_t - \gamma_l}{\sigma_l}\right) - G\left(\frac{a_{k-1} - \alpha_l f_t - \gamma_l}{\sigma_l}\right)\right].
\end{aligned}$$

[12] In practice, the alternative $k = K$ typically corresponds to default, which is an absorbing state. For expository purpose, we do not consider an absorbing state here.

The estimators of the factor values given β are:

$$\hat{f}_{n,t}(\beta) = \arg\max_{f_t} \sum_{k=1}^{K} \sum_{l=1}^{K} N_{lk,t} \log \left[G\left(\frac{a_k - \alpha_l f_t - \gamma_l}{\sigma_l} \right) - G\left(\frac{a_{k-1} - \alpha_l f_t - \gamma_l}{\sigma_l} \right) \right], \quad t = 1, ..., T, \tag{5.1}$$

and depend on the data through the aggregate counts $N_{lk,t}$ of transitions from rating l at time $t-1$ to rating k at time t, for $k, l = 1, ..., K$ and $t = 1, ..., T$. The (semi-)parametrically efficient estimator of the micro-parameter is:

$$\hat{\beta}^*_{nT} = \arg\max_{\beta} \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=1}^{T} N_{lk,t} \log \left[G\left(\frac{a_k - \alpha_l \hat{f}_{n,t}(\beta) - \gamma_l}{\sigma_l} \right) - G\left(\frac{a_{k-1} - \alpha_l \hat{f}_{n,t}(\beta) - \gamma_l}{\sigma_l} \right) \right]. \tag{5.2}$$

This estimator is computed from the aggregate data on rating transition counts $(N_{lk,t})$.

To compare the finite-sample distribution of estimator $\hat{\beta}^*_{nT}$ and the semi-parametric efficiency bound, we perform a Monte-Carlo study. We consider the two-state case $K = 2$ and a DGP where the transition probabilities are given by a one-factor logit specification. Under the semi-parametric identification constraints $a_1 = \gamma_1 = 0$ and $\alpha_1 = \sigma_1 = \sigma_2 = 1$, the micro-parameter to estimate is $\beta = (\alpha_2, \gamma_2)'$. The common factor f_t follows a linear Gaussian autoregressive process:

$$f_t = \mu + \rho f_{t-1} + \sigma \eta_t, \tag{5.3}$$

where (η_t) is $i.i.N(0,1)$. The parameter values used in the Monte-Carlo study are displayed in Table 1.

Table 1: Parameter values

$\alpha_1 = 1$	$\gamma_1 = 0$	$\sigma_1 = 1$	$\alpha_2 = 1$	$\gamma_2 = -0.5$	$\sigma_2 = 1$
$a_0 = -\infty$	$a_1 = 0$	$a_2 = +\infty$	$\mu = 0.1$	$\rho = 0.5$	$\sigma = 0.5$

In Figures 1 and 2, we consider the sample sizes $n = 200$, $T = 20$, and $n = 1000$, $T = 20$, respectively. In each figure, the two panels display the finite sample distributions of the estimators $\hat{\beta}^*_{nT}$ for the two micro-parameters (solid lines). We also display for each micro-parameter the Gaussian distribution (dashed lines) with mean equal to the true parameter value and variance equal to the semi-parametric efficiency bound divided by nT. The estimator $\hat{\beta}^*_{nT}$ is computed from (5.2) by numerical optimization, where for given β the estimate $\hat{f}_{n,t}(\beta)$ in (5.1) is computed by grid search. As expected from the stochastic migration literature, the α_2 parameter, which represents the sensitivity of the transition probabilities with respect to the factor, is the most difficult to estimate. Its asymptotic variance is larger and needs more granularity adjustment, that is, the convergence of the finite sample distribution to the asymptotic one is slower. By comparing Figures 1 and 2, it is seen that the standard deviations of the estimators decrease by a factor about 2 passing from $n = 200$ to $n = 1000$, as suggested by the rate of convergence $\sqrt{nT}$ of the micro-parameters. Finally, we observe that the finite sample bias is rather small for both estimators.

The semi-parametric efficiency bound for α_2 is given by [13] (see Appendix 7.2):

$$B_{\alpha_2} = E_0\left[\mu_{2,t-1}\pi_{22,t}\left(1-\pi_{22,t}\right)\left(1-\frac{\mu_{2,t-1}\pi_{22,t}\left(1-\pi_{22,t}\right)\alpha_2^2}{\mu_{1,t-1}\pi_{12,t}\left(1-\pi_{12,t}\right)+\mu_{2,t-1}\pi_{22,t}\left(1-\pi_{22,t}\right)\alpha_2^2}\right)\begin{pmatrix} f_t^2 & f_t \\ f_t & 1 \end{pmatrix}\right]^{-1}, \tag{5.4}$$

where:

$$\pi_{12,t} = \frac{1}{1+e^{f_t}} \ , \ \pi_{22,t} = \frac{1}{1+e^{\alpha_2 f_t+\gamma_2}},$$

and:

$$\mu_{1,t-1} = P\left[y_{i,t-1} = 1 | \underline{f_{t-1}}\right] = 1 - \mu_{2,t-1}.$$

The matrix B_{α_2} involves the probabilities $\mu_{1,t-1}$ and $\mu_{2,t-1}$ of the lagged states, conditional on the factor path, and the conditional variances of the indicator of state 2, that are $\pi_{21,t}(1-\pi_{21,t})$ and $\pi_{22,t}(1-\pi_{22,t})$, respectively, according to the previous state. The matrix B_{α_2} depends on macro-parameters μ, ρ, σ^2 by means of the expectation E_0. The semi-parametric efficiency bound can be approximated numerically by Monte-Carlo integration (see Appendix 7.3).

[13]Under the hypothesis $\alpha_2 = 0$ of no factor effect on the second state, the information matrix reduces to $\pi_{22}(1-\pi_{22})E_0\left[\mu_{2,t-1}\begin{pmatrix} f_t^2 & f_t \\ f_t & 1 \end{pmatrix}\right]$. Therefore, for testing the absence of factor effect, the correction for factor unobservability is not needed.

In Figure 3 we display the semi-parametric efficiency bound of parameter α_2 as a function of the autoregressive coefficient ρ and the unconditional variance $\frac{\sigma^2}{1-\rho^2}$ of the factor process (f_t). The values of the micro-parameters and μ are given in Table 1. More precisely, we display the asymptotic standard deviation $\left(\frac{1}{nT}B_{\alpha_2}\right)^{1/2}$, where $n = 1000$ and $T = 20$. The semi-parametric efficiency bound is decreasing w.r.t. the factor variance. The pattern is almost flat w.r.t. the autoregressive coefficient ρ of the factor, except for values of ρ close to 1, where the semi-parametric efficiency bound diverges to infinity.

5.3 Estimation of the macro-parameters

Let us now consider the efficient estimator of the macro-parameter $\theta = (\mu, \rho, \sigma^2)'$. This estimator is based on the cross-sectional approximations of the factor values $\hat{f}_{nT,t} = \hat{f}_{n,t}\left(\hat{\beta}^*_{nT}\right)$ from (5.1) and (5.2). The estimators $\hat{\mu}$ and $\hat{\rho}$ are obtained by OLS on the regression:

$$\hat{f}_{nT,t} = \mu + \rho \hat{f}_{nT,t-1} + u_t, \;\; t = 2, ..., T.$$

The estimator of parameter σ^2 is given by $\hat{\sigma}^2 = \dfrac{1}{T-1}\displaystyle\sum_{t=2}^{T} \hat{u}_t^2$, where $\hat{u}_t = \hat{f}_{nT,t} - \hat{\mu} - \hat{\rho}\hat{f}_{nT,t-1}$ are the OLS residuals. The estimator $\hat{\theta} = (\hat{\mu}, \hat{\rho}, \hat{\sigma}^2)'$ achieves the asymptotic efficiency bound with observable factor, that is, the Cramer-Rao bound for θ in the linear Gaussian model (5.3). Thus, the asymptotic efficiency bound is such that the estimators of $(\mu, \rho)'$ and σ^2 are asymptotically independent, root-T consistent, with asymptotic variance:

$$B^*_{(\mu,\rho)} = \sigma_0^2 E\left[\begin{pmatrix} 1 & f_t \\ f_t & f_t^2 \end{pmatrix}\right]^{-1} = \begin{pmatrix} \sigma_0^2 + \mu_0^2 \frac{1+\rho_0}{1-\rho_0} & -\mu_0(1+\rho_0) \\ -\mu_0(1+\rho_0) & 1-\rho_0^2 \end{pmatrix},$$

for $(\mu, \rho)'$, and $B^*_{\sigma^2} = 2\sigma_0^4$ for σ^2.

Figures 4 and 5 display the distributions (solid lines) of the efficient estimators $\hat{\mu}$, $\hat{\rho}$ and $\hat{\sigma}^2$ in the Monte-Carlo study for sample sizes $n = 200$, $T = 20$, and $n = 1000$, $T = 20$, respectively. The parameter values are given in Table 1. We also display Gaussian distributions (dashed lines) centered at the true values of the parameters and with variances equal to the efficiency bounds divided by T. As expected, it is more difficult to estimate the autoregressive coefficient ρ and the variance σ^2 than to estimate the intercept μ. The estimators $\hat{\rho}$ and $\hat{\sigma}^2$ feature moderate downward biases. By comparing Figure 4 and Figure

5, we notice that the standard deviations of the estimators are rather similar for the two sample sizes and do not scale with n. Moreover, by comparing Figure 2 and Figure 5, it is seen that the discrepancy between the finite-sample distribution and the asymptotic efficiency bound is more pronounced for the macro-parameters than for the micro-parameters for our sample sizes. These findings are a consequence of the different convergence rates of the two types of estimators, that are $\sqrt{T}$ and $\sqrt{nT}$, respectively.

5.4 Prediction of the factor values

Let us consider the prediction of the future factor value f_{T+L} given the information available at the last date T of the sample, where $L = 1, 2, \cdots$, is the prediction horizon. If the factor values were observable, and the macro-parameters were known, the prediction of f_{T+L} at date T is given by $\hat{f}^*_{T,T+L} = \mu \dfrac{1-\rho^L}{1-\rho} + \rho^L f_T$, for any horizon $L = 1, 2, \cdots$. The prediction error is $\varepsilon^*_{T,T+L} = \hat{f}^*_{T,T+L} - f_{T+L} = \sigma\left(\eta_{T+L} + \rho\eta_{T+L-1} + \cdots + \rho^{L-1}\eta_{T+1}\right)$, which is independent of the sample information. The prediction error has zero unconditional mean, and the unconditional variance is given by $V\left[\varepsilon^*_{T,T+L}\right] = \sigma^2 \dfrac{1-\rho^{2L}}{1-\rho^2}$. When the factor is unobservable and the macro-parameters are unknown, the factor values can be replaced by their cross-sectional approximations, and the macro-parameters μ and ρ by their efficient estimators [14]. We get the term-structure of predictions at date T:

$$\hat{f}_{T,T+L} = \hat{\mu}\frac{1-\hat{\rho}^L}{1-\hat{\rho}} + \hat{\rho}^L \hat{f}_{nT,T}, \quad L = 1, 2, \cdots . \tag{5.5}$$

The difference $\varepsilon_{T,T+L} = \hat{f}_{T,T+L} - f_{T+L}$ between the predicted and true factor values can be decomposed as:

$$\begin{aligned} \varepsilon_{T,T+L} &= \hat{\rho}^L\left(\hat{f}_{nT,T} - f_T\right) + \left[\hat{\mu}\frac{1-\hat{\rho}^L}{1-\hat{\rho}} - \mu\frac{1-\rho^L}{1-\rho} + \left(\hat{\rho}^L - \rho^L\right) f_T\right] + \varepsilon^*_{T,T+L} \\ &=: \varepsilon^{(1)}_{T,T+L} + \varepsilon^{(2)}_{T,T+L} + \varepsilon^*_{T,T+L}. \end{aligned} \tag{5.6}$$

Terms $\varepsilon^{(1)}_{T,T+L}$ and $\varepsilon^{(2)}_{T,T+L}$ are induced by the approximation of the factor values, and by the estimation of the macro-parameters μ and ρ, respectively.

[14] For given values of the macro-parameters, a different predictor is obtained by computing the conditional expectation of f_{T+L} given the available information on the observable endogenous variables $\underline{y_T}$. This predictor is equivalent to the one in (5.5) at order $1/n$ [see Gagliardini, Gouriéroux (2008)].

To assess the prediction accuracy, we compute the unconditional expectation and variance of the prediction error $\varepsilon_{T,T+L}$, and of its components $\varepsilon^{(1)}_{T,T+L}$, $\varepsilon^{(2)}_{T,T+L}$, $\varepsilon^{*}_{T,T+L}$, for prediction horizons $L = 1, 2, \cdots, 5$. The DGP parameters are given in Table 1. The results are displayed in Figure 6 for sample sizes $n = 200$, $T = 20$ (upper Panels) and $n = 1000$, $T = 20$ (lower Panels). For both sample sizes and across prediction horizons, the expectation of $\varepsilon_{T,T+L}$ is of the order $10^{-2} - 10^{-3}$, and thus the bias of the predictor $f_{T,T+L}$ is rather small. The main contribution to this bias is typically due to the estimation of the macro-parameters. The contribution of the approximation of the factor values is small. This contribution is decreasing in absolute value w.r.t. the prediction horizon, since the prediction $f_{T,T+L}$ is almost independent of the factor value for large prediction horizon. The sign of the prediction bias, and its shape as a function of the prediction horizon, are very different for the two sample sizes. For $n = 200$, $T = 20$, the prediction bias can be either positive or negative, and the expectation of $\varepsilon_{T,T+L}$ is monotonically decreasing w.r.t. the prediction horizon L. Instead, for $n = 1000$, $T = 20$, the expectation of $\varepsilon_{T,T+L}$ is a non-monotonic function of L, and the prediction bias is negative up to the investigated horizon. Let us now consider the variance of $\varepsilon_{T,T+L}$. The term structures of the prediction error variances are rather similar for the two sample sizes. At prediction horizon $L = 1$, about 90% of the variance of $\varepsilon_{T,T+1}$ is due to the variance of the prediction error with observable factor and known macro-parameters, while the remaining 10% comes from the estimation of the macro-parameters. The contribution of the approximation of the factor values is very small. The variance of $\varepsilon_{T,T+L}$ is monotonically increasing w.r.t. the prediction horizon.

6 Concluding remarks

We have considered nonlinear dynamic panel models with common unobservable factor, in which it is possible to disentangle the micro- and the macro-dynamics, the latter being captured by the factor dynamic. Such models are largely encountered in financial and insurance applications, in which structured derivative products are constructed from large homogenous pools of individual contracts such as mortgages, corporate loans, or life insurance contracts. They are also appropriate for performing macro-prediction from tendency surveys

[Gouriéroux, Monfort (2008)]. For large cross-sectional and time dimensions [$n, T \to \infty$, $T^b/n = O(1)$, $b > 1$], we have derived the semiparametric efficiency bound of the parameter β characterizing the micro-dynamics. The semi-parametric efficiency bound takes into account the factor unobservability, and coincides with the bound for known factor transition. Moreover, we have shown that the fixed effects estimator of β achieves the semi-parametric efficiency. These results require a large cross-sectional dimension to approximate the likelihood function, which involves multidimensional integrals, by a closed form expression. This expansion around $n = \infty$ is the basis for granularity adjustments.

The main results of the paper are still valid when the model is extended to include observable explanatory variables. The micro- and macro-dynamics become $h(y_{i,t}|y_{i,t-1}, x_{i,t}, z_t, f_t; \beta)$ and $g(f_t|f_{t-1}, z_t; \theta)$ respectively, where $x_{i,t}$ and z_t are observed exogenous variables. The explanatory variables $x_{i,t}$ introduce observable individual heterogeneity. The identifiability of the model requires in particular that the effects of the unobservable factor f_t and the observable macro-variables z_t in the micro-dynamics can be disentangled.

References

[1] Aigner, D., Hsiao, C., Kapteyn, A., and T., Wansbeek (1984): ”Latent Variables”, in *Handbook of Econometrics*, Vol. 3, Engle, R., and D., McFadden (Eds.), Elsevier, Amsterdam.

[2] Andrews, D. (1988): ”Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables”, Econometric Theory, 4, 458-67.

[3] Arellano, M., and S., Bonhomme (2006): ”Robust Priors in Nonlinear Panel Data Models”, forthcoming in Econometrica.

[4] Arellano, M., and J., Hahn (2006): ”A Likelihood-based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects”, Working Paper.

[5] Bai, J., and S., Ng (2002): ”Determining the Number of Factors in Approximate Factor Models”, Econometrica, 70, 191-221.

[6] Basel Committee on Banking Supervision (2001): ”The New Basel Capital Accord”, Consultative Document of the Bank for International Settlements.

[7] Bester, C., and C., Hansen (2005): ”A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects”, Working Paper.

[8] Bosq, D. (1998): *Nonparametric Statistics for Stochastic Processes*, Springer.

[9] Bougerol, P., and N., Picard (1992): ”Stationarity of GARCH Processes and of Some Nonnegative Time Series”, Journal of Econometrics, 52, 115-127.

[10] Chamberlain, G. (1987): ”Asymptotic Efficiency in Estimation with Conditional Moment Restrictions”, Journal of Econometrics, 34, 305-334.

[11] Chen, X., and X., Shen (1998): ”Sieve Extremum Estimates for Weakly Dependent Data”, Econometrica, 66, 298-314.

[12] Connor, G., Hagmann, M., and O., Linton (2007): ”Efficient Estimation of a Semiparametric Characteristic-based Factor Model of Security Returns”, London School of Economics DP.

[13] Cox, D., and N., Reid (1987): ”Parameter Orthogonality and Approximate Conditional Inference”, Journal of the Royal Statistical Society, Series B, 49, 1-39.

[14] Davidson, J. (1994): *Stochastic Limit Theory*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

[15] de Finetti, B. (1931): ”Funzione Caratteristica di un Fenomeno Aleatorio”, Atti della R. Accademia dei Lincei, 6, Memorie, Classe di Scienze Fisiche, Matematiche e Naturali, 4, 251-299.

[16] Feng, D., Gouriéroux, C., and J., Jasiak (2008): ”The Ordered Qualitative Model for Credit Rating Transitions”, Journal of Empirical Finance, 15, 111-130.

[17] Forni, M., Hallin, M., Lippi, M., and L., Reichlin (2004): ”The Generalized Dynamic Factor Model: Consistency and Rates”, Journal of Econometrics, 119, 231-255.

[18] Forni, M., and L., Reichlin (1998): ”Let’s Get Real: A Factor Analytic Approach to Disaggregated Business Cycle Dynamics”, Review of Economic Studies, 65, 453-473.

[19] Fu, J. (1982): ”Large Sample Point Estimates: A Large Deviation Approach”, Annals of Statistics, 10, 762-771.

[20] Gagliardini, P., and C., Gouriéroux (2005a): ”Migration Correlation: Definition and Efficient Estimation”, Journal of Banking and Finance, 29, 865-894.

[21] Gagliardini, P., and C., Gouriéroux (2005b): ”Stochastic Migration Models with Application to Corporate Risk”, Journal of Financial Econometrics, 3, 188-226.

[22] Gagliardini, P., and C., Gouriéroux (2008): ”Approximate Derivative Pricing in Large Homogeneous Class of Assets”, CREST and University of Lugano DP.

[23] Gordy, M., and E., Heitfield (2002): ”Estimating Default Correlation from Short Panels of Credit Rating Performance Data”, Working Paper, Federal Reserve Board.

[24] Gordy, M., and E., Lutkebohmert (2007): ”Granularity Adjustment for Basel II”, Deutsche Bank DP.

[25] Gouriéroux, C., and J., Jasiak (2008): ”Granularity Adjustment for Default Risk Factor Model with Cohorts”, York University DP.

[26] Gouriéroux, C., and A., Monfort (1995): *Statistics and Econometric Models*, Cambridge University Press, Cambridge.

[27] Gouriéroux, C., and A., Monfort (2008): “Granularity in Qualitative Factor Models”, CREST Working Paper.

[28] Granger, C. (1980): ”Long Memory Relationships and the Aggregation of Dynamic Models”, Journal of Econometrics, 14, 227-238.

[29] Granger, C., and R., Joyeux (1980): ”An Introduction to Long Memory Time Series Models and Fractional Differencing”, Journal of Time Series Analysis, 1, 15-29.

[30] Gupton, G., Finger, C., and M., Bhatia (1997): ”Creditmetrics”, Technical Report, The Risk Metrics Group.

[31] Hahn, J., and G., Kuersteiner (2004): ”Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects”, Working Paper.

[32] Hahn, J., and W., Newey (2004): ”Jackknife and Analytical Bias Reduction for Nonlinear Panel Models”, Econometrica, 72, 1295-1319.

[33] Hansen, L., Nielsen, B., and J., Nielsen (2004): ”Two Sided Analysis of Variance with a Latent Time Series”, Nuffield College, Oxford University.

[34] Hewitt, E., and L., Savage (1955): ”Symmetric Measures on Cartesian Products”, Transaction of the American Mathematical Society, 80, 470-501.

[35] Hjellvik, V., and D., Tjostheim (1999): ”Modelling Panels of Intercorrelated Autoregressive Time Series”, Biometrika, 86, 573-590.

[36] Jensen, J. (1995): *Saddlepoint Approximations,* Clarendon Press, Oxford.

[37] Lancaster, T. (2000): ”The Incidental Parameter Problem Since 1948”, Journal of Econometrics, 95, 391-413.

[38] Douc, R., Moulines, E., and T., Rydén (2004): “Asymptotic Properties of the Maximum Likelihood Estimator in Autoregressive Models with Markov Regime”, Annals of Statistics, 32, 2254-2304.

[39] Neyman, J., and E., Scott (1948): ”Consistent Estimates Based on Partially Consistent Observations”, Econometrica, 16, 1-31.

[40] Severini, T., and G., Tripathi (2001): ”A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models”, Journal of Econometrics, 102, 23-66.

[41] Stein, C. (1956): ”Efficient Nonparametric Testing and Estimation”, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, University of California Press, Berkeley, CA, 187-195.

[42] Stock, J., and M., Watson (2002): ”Forecasting Using Principal Components from a Large Number of Predictors”, Journal of the American Statistical Association, 97, 1167-1179.

[43] Sweeting, T. (1987): Discussion of Cox and Reid (1987), Journal of the Royal Statistical Society, Series B, 49, 20-21.

[44] Vasicek, O. (1987): ”Probability of Loss on Loan Portfolio”, KMV Technical Report.

[45] Vasicek, O. (1991): ”Limiting Loan Loss Probability Distribution”, KMV Technical Report.

[46] Woutersen, T. (2002): ”Robustness Against Incidental Parameters”, Working Paper.

Appendix 1

Weak LLN and Slutsky Theorem

This Appendix provides asymptotic results for nonlinear panel models with common factor to show the stochastic convergence:

$$\frac{1}{T}\sum_{t=1}^{T}\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)\right)\xrightarrow{p}E_0\left[\varphi\left(\mu_t(\beta)\right)\right], \tag{A.1}$$

uniformly in $\beta \in \mathcal{B}$, as $n, T \to \infty$, where $Y_{i,t} = (y_{i,t}, y_{i,t-1}, \cdots, y_{i,t-L})'$, $\mu_t(\beta) = E_0\left[a(Y_{i,t}, f_t(\beta), \beta)|\underline{f_t}\right]$, $\hat{f}_{n,t}(\beta)$ is a consistent estimator of $f_t(\beta)$, $\mathcal{B} \subset \mathbb{R}^q$ denotes the parameter set, and a and φ are functions. The result in Lemma A.1 is proved in Appendix B.1 on the web-site.

Lemma A.1: *Let matrix function $a(Y, f, \beta)$ admit values in $\mathbb{R}^{r\times r}$. Assume:*

(1) (i) *Parameter set $\mathcal{B} \subset \mathbb{R}^q$ is compact.*

(ii) $E_0\left[\|a(Y_{i,t}, f_t(\beta), \beta)\|^9\right] < \infty$, *for any* $\beta \in \mathcal{B}$, $E_0\left[\sup_{\beta\in\mathcal{B}}\|a(Y_{i,t}, f_t(\beta), \beta)\|^4\right] < \infty$.

(iii) $E_0\left[\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial\, vec[a(Y_{i,t}, f_t(\beta), \beta)]}{\partial\beta'}\right\|^4\right] < \infty$.

(iv) *For any* $\beta \in \mathcal{B}$: $E_0\left[\|\mu_t(\beta) - E_0[a(Y_{i,t}, f_t(\beta), \beta)|f_t, ..., f_{t-m}]\|^2\right] = O\left(m^{-\alpha}\right)$, *for some* $\alpha > 0$, *as* $m \to \infty$, *where* $\mu_t(\beta) = E_0[a(Y_{i,t}, f_t(\beta), \beta)|\underline{f_t}]$.

(v) $E_0\left[\exp\left(-u\xi_t\right)\right] \leq C_1 \exp\left(-C_2 u^\delta\right)$ *as* $u \to \infty$, *for some positive constants* $C_1, C_2, \delta > 0$, *where* $\xi_t = \left[1 + \sup_{\beta\in\mathcal{B}}\sigma_t^2(\beta)\right]^{-1}$ *and* $\sigma_t^2(\beta) = E_0\left[\|a(Y_{i,t}, f_t(\beta), \beta)\|^2|\underline{f_t}\right]$.

(vi) *Condition* (v) *holds if we replace* ξ_t *by* $\tilde{\xi}_t = \left[1 + \sup_{\beta\in\mathcal{B}}\tilde{\sigma}_t^2(\beta)\right]^{-1}$, *where* $\tilde{\sigma}_t^2(\beta) = E_0\left[b(Y_{i,t}, f_t(\beta), \beta)^2|\underline{f_t}\right]$ *and* $b(Y_{i,t}, f_t(\beta), \beta) = \sup_{f:\|f-f_t(\beta)\|\leq\eta^*}\left\|\frac{\partial\, vec[a(Y_{i,t}, f, \beta)]}{\partial f'}\right\|$, $\eta^* > 0$.

(2) *Function* $\varphi : \mathbb{R}^{r\times r} \to \mathbb{R}$ *is Lipschitz and there exists* $\tau > 2$ *such that* $E_0\left[|\varphi(\mu_t(\beta))|^\tau\right] < \infty$, *for any* $\beta \in \mathcal{B}$.

(3) $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\|\hat{f}_{n,t}(\beta) - f_t(\beta)\| = O_p(T^{-\rho})$, *for* $\rho > 0$.

(4) $n, T \to \infty$, *such that* $T/n \to 0$.

Then, under Assumptions A.1-A.5:

$$\sup_{\beta \in \mathcal{B}} \left| \frac{1}{T} \sum_{t=1}^{T} \varphi \left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) - E_0 \left[\varphi \left(\mu_t(\beta) \right) \right] \right| \xrightarrow{p} 0.$$

Lemma A.1 follows from:

(a) The convergence of estimator $\hat{f}_{n,t}(\beta)$ to $f_t(\beta)$, and the convergence of the cross-sectional average $\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, f_t(\beta), \beta)$ to $\mu_t(\beta) = E_0 \left[a(Y_{i,t}, f_t(\beta), \beta) | \underline{f_t} \right]$ by a Weak LLN (WLLN) conditional on $\underline{f_t}$, uniformly in $t = 1, \cdots, T$ and $\beta \in \mathcal{B}$,

(b) The application of the Slutsky theorem with continuous function φ,

(c) The convergence of the time series average of $\varphi(\mu_t(\beta))$ to the population expectation by the WLLN, uniformly in $\beta \in \mathcal{B}$.

Since the continuity point $\mu_t(\beta)$ for the application of the Slutsky theorem is stochastic, we need the Lipschitz condition for φ in condition (2). Condition (1) (v) in Lemma A.1 is used to apply Bernstein's inequality [e.g., Bosq (1998), Theorem 1.2] to derive a large deviation bound for $\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, f_t(\beta), \beta) - \mu_t(\beta)$ uniformly in $1 \leq t \leq T$ and $\beta \in \mathcal{B}$. Condition (1) (vi), combined with condition (3), is used to show that $\frac{1}{n}\sum_{i=1}^{n} \left[a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) - a(Y_{i,t}, f_t(\beta), \beta) \right]$ converges to zero, uniformly in $1 \leq t \leq T$ and $\beta \in \mathcal{B}$. The uniform convergence in condition (3) is proved in Appendix 4 (see Lemma A.6), when $\hat{f}_{n,t}(\beta)$ is the cross-sectional ML estimator introduced in Section 3. Finally, the uniform convergence of $\frac{1}{T}\sum_{t=1}^{T} \varphi(\mu_t(\beta))$ to $E_0 \left[\varphi(\mu_t(\beta)) \right]$ relies on a mixingale WLLN in Andrews (1988) and convergence results for Near-Epoch Dependent processes in Davidson (1994).

Lemma A.1 is also valid for multivariate functions φ whose components satisfy condition (2), in particular for the matrix identity mapping $\varphi(x) = x$, $x \in \mathbb{R}^{r \times r}$. However, the Lipschitz property in condition (2) prevents the application of Lemma A.1 when φ is the matrix inversion $\varphi(x) = x^{-1}$. Lipschitz condition (2) is relaxed in Lemma A.2, which is proved in Appendix B.2.

Lemma A.2: *Let $a(Y, f, \beta)$ admit values in the set of symmetric matrices of dimension r, and let $\mathcal{U}$ be the open subset of positive definite matrices. Assume:*

(1) *Conditions* (1) (i)-(vi) *of Lemma A.1 hold. Moreover:*

(vii) $\mu_t(\beta) = E_0\left[a(Y_{i,t}, f_t(\beta), \beta)|\underline{f_t}\right] \in \mathcal{U}$*, for any t and $\beta \in \mathcal{B}$, P-a.s., and the smallest eigenvalue $\lambda_t(\beta)$ of $\mu_t(\beta)$ is such that* $E_0\left[\left(\inf_{\beta\in\mathcal{B}}\lambda_t(\beta)\right)^{-4}\right] < \infty$.

(viii) *Conditions* (1) (v)-(vi) *of Lemma A.1 hold for* $\xi_t = \left[\inf_{\beta\in\mathcal{B}} \lambda_t(\beta)\right]\left[1 + \sup_{\beta\in\mathcal{B}} \frac{\sigma_t^2(\beta)}{\lambda_t(\beta)}\right]^{-1}$ *and* $\tilde{\xi}_t = \left[\inf_{\beta\in\mathcal{B}} \lambda_t(\beta)\right]\left[1 + \sup_{\beta\in\mathcal{B}} \frac{\tilde{\sigma}_t^2(\beta)}{\lambda_t(\beta)}\right]^{-1}$.

(2) *Function $\varphi : \mathcal{U} \to \mathbb{R}$ is such that:*

(i) *φ is Lipschitz on any compact subset of $\mathcal{U}$.*

(ii) $|\varphi(w)| \leq C\|z\|^\tau\psi(z)$*, for any $w, z \in \mathcal{U}$ such that $w = (1 + \Delta)z$, $\|\Delta\| \leq 1/2$, where constants C, τ satisfy $C > 0, \tau \leq 2$, and function ψ is such that* $E_0[\sup_{\beta\in\mathcal{B}}|\psi(\mu_t(\beta))|^4] < \infty$.

(3) $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\|\hat{f}_{n,t}(\beta) - f_t(\beta)\| = O_p(T^{-\rho})$*, for $\rho > 0$.*

(4) *$n, T \to \infty$ such that $T/n \to 0$.*

Then, under Assumptions A.1-A.5:

$$\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T}\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right) - E_0\left[\varphi\left(\mu_t(\beta)\right)\right]\right| \xrightarrow{p} 0. \tag{A.2}$$

Assumptions (1) (vii)-(viii) of Lemma A.2 involve a tail condition on the stationary distribution of the smallest eigenvalue $\lambda_t(\beta)$ of matrix $\mu_t(\beta)$ in a neighbourhood of 0. In condition (2) (i), function φ is locally Lipschitz on compact subsets of $\mathcal{U}$. The growth of $|\varphi|$ outside compact sets is bounded by condition (2) (ii). These conditions are sufficiently general to accommodate functions φ used in Appendix 6 to derive the asymptotic properties of the estimators.

Corollary A.3: *Assume that conditions (1), (3) and (4) of Lemma A.2 hold. Let function φ be either:*

(A) *The matrix inversion* $\varphi : \mathcal{U} \to \mathbb{R}^{r\times r}$, $\varphi(x) = x^{-1}$, *or*

(B) *The mapping* $\varphi : \mathcal{U} \to \mathbb{R}^{s\times s}$, $\varphi(x) = (x^{11})^{-1}$ *where* x^{11} *is the upper-left* s*-dimensional block of* x^{-1}, $s < r$.

Then, convergence (A.2) holds.

Corollary A.3 is deduced from Lemma A.2 since the inversion mapping satisfies $w^{-1} - z^{-1} = -w^{-1}(w-z)z^{-1}$, for $w, z \in \mathcal{U}$ (see Appendix B.3).

Appendix 2

Large deviation bounds for ML estimators in an i.i.d. framework

We provide two large deviation bounds for ML estimators in an i.i.d. framework. They are used in Appendix 4 to derive the rate of convergence of the factor approximations. Let us consider the ML estimator:

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta),$$

where $L_n(\theta) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} l_i(\theta)$ and $l_i(\theta) = \log h(y_i, \theta)$. Let $L(\theta) = E_0[l_i(\theta)]$, where $E_0[.]$ denotes expectation w.r.t. the true probability distribution P_0. Let us assume:

i) Parameter set $\Theta \subset \mathbb{R}^d$ is compact and convex.

ii) The observations y_i, $i = 1, ..., n$, are i.i.d. with density $h(y_i, \theta_0)$, where θ_0 is the true parameter value.

iii) Parameter $\theta_0 \in \Theta$ is globally identified, that is, $L(\theta_0) > L(\theta)$ for any $\theta \in \Theta$, $\theta \neq \theta_0$, and locally identified, that is, the matrix $J_0 = E_0\left[-\dfrac{\partial^2 \log h(y_i, \theta_0)}{\partial\theta\partial\theta'}\right]$ is non-singular.

iv) There exists $\gamma > 2$ such that:

$$R := E_0\left[\sup_{\theta\in\Theta}\left\|\frac{\partial \log h(y_i,\theta)}{\partial\theta}\right\|^{\gamma}\right] < \infty.$$

Under the compactness condition in i), condition iii) is equivalent to:

$$\mathcal{K} := 2 \inf_{\theta \in \Theta: \theta \neq \theta_0} \frac{KL(\theta, \theta_0)}{\|\theta - \theta_0\|^2} > 0, \tag{A.3}$$

where $KL(\theta, \theta_0) = L(\theta_0) - L(\theta) = E_0 \left[\log \left(\frac{h(y_i, \theta_0)}{h(y_i, \theta)} \right) \right]$ is the Kullback-Leibler discrepancy between θ and θ_0. Moreover, under condition iv):

$$\Gamma := \sup_{\theta \in \Theta} TrI(\theta, \theta_0) = \sup_{\theta \in \Theta} E_0 \left[\left\| \frac{\partial \log h(y_i, \theta)}{\partial \theta} \right\|^2 \right] < \infty,$$

where $I(\theta, \theta_0) = E_0 \left[\frac{\partial \log h(y_i, \theta)}{\partial \theta} \frac{\partial \log h(y_i, \theta)}{\partial \theta'} \right]$. Note that $I(\theta_0, \theta_0) = J_0$ by the information matrix equality, but matrix $I(\theta, \theta_0)$ differs in general from J_0 for $\theta \neq \theta_0$, even for well-specified models.

Lemma A.4: *Under conditions i)-iv), there exist constants $c_1, c_2, c_3 > 0$ (depending on d, but independent of Θ and the parametric model) such that for any n and $\varepsilon > 0$:*

$$P_0 \left[\left\| \widehat{\theta}_n - \theta_0 \right\| \geq \varepsilon \right] \leq c_1 n^d \exp \left(-c_2 n \varepsilon^2 \frac{\mathcal{K}}{1 + \Gamma / \mathcal{K}} \right) + c_3 \varepsilon^{\gamma - 2} \frac{R}{\mathcal{K}}.$$

Proof: See Appendix B.4.

Lemma A.4 differs from large deviation bounds for ML estimators known in the literature [e.g., Fu (1982), Chen, Shen (1998), Theorem 3], since Lemma A.4 makes explicit how the bound with given threshold ε and sample size n depends on the true probability distribution [15]. This dependence is summarized by statistics $\mathcal{K}$, Γ and R, and by exponent γ. In particular, the coefficient of $n\varepsilon^2$ in the exponential term involves the ratio $\frac{\mathcal{K}}{1 + \Gamma/\mathcal{K}}$. This ratio is an increasing function of the Kullback-Leibler measure $\mathcal{K}$, and and a decreasing function of the second moment of the score Γ.

The large deviation bound in Lemma A.4 can be extended to models with nuisance parameters. Let the log-density $l_i(\theta) = \log h(y_i, \theta)$ be parametrized by $\theta = (\alpha, \beta)$, where the parameter of interest is $\alpha \in \mathcal{A}$, and the nuisance parameter is $\beta \in \mathcal{B}$. We consider the concentrated ML estimator of parameter α defined by:

$$\widehat{\alpha}_n(\beta) = \arg \max_{\alpha \in \mathcal{A}} L_n(\alpha, \beta),$$

[15] Moreover, compared to the results in Fu (1982), Lemma A.4 applies for multivariate parameter θ.

for any $\beta \in \mathcal{B}$, where $L_n(\alpha, \beta) = \dfrac{1}{n}\sum_{i=1}^{n} l_i(\theta)$. Denote $L(\theta) = E_0[l_i(\theta)]$, and $\Theta = \mathcal{A} \times \mathcal{B}$.

Lemma A.5: *Assume:*

i) Set $\mathcal{A} \subset \mathbb{R}^K$ is compact and convex, and set $\mathcal{B} \subset \mathbb{R}^q$ is compact.

ii) The observations y_i are i.i.d. with density $h(y_i, \theta_0)$, where $\theta_0 = (\alpha_0, \beta_0)$ is the true parameter value.

iii) For any given $\beta \in \mathcal{B}$, the function $L(\alpha, \beta)$ is uniquely maximized w.r.t. $\alpha \in \mathcal{A}$ at $\alpha(\beta) = \arg\max_{\alpha \in \mathcal{A}} L(\alpha, \beta)$. The true values of parameters $\alpha_0 \in \mathcal{A}$ and $\beta_0 \in \mathcal{B}$ satisfy $\alpha_0 = \alpha(\beta_0)$, and the matrix $J(\beta) = E_0\left[-\dfrac{\partial^2 l_i(\alpha(\beta), \beta)}{\partial\alpha\partial\alpha'}\right]$ is non-singular, for any $\beta \in \mathcal{B}$.

iv) There exists $\gamma > 2$ such that $R := E_0\left[\sup_{\theta\in\Theta}\left\|\dfrac{\partial \log h(y_i, \theta)}{\partial\theta}\right\|^\gamma\right] < \infty$.

Then, there exist constants $c_1, c_2, c_3 > 0$ (depending on dimensions K and q, but independent of $\mathcal{A}$, $\mathcal{B}$ and the parametric model) such that for any n and $\varepsilon > 0$:

$$P\left[\sup_{\beta\in\mathcal{B}}\|\widehat{\alpha}_n(\beta) - \alpha(\beta)\| \geq \varepsilon\right] \leq c_1 Vol(\mathcal{B})\frac{n^{K+q}}{\varepsilon^q}\exp\left(-c_2 n\varepsilon^2\frac{\mathcal{K}}{1+\Gamma/\mathcal{K}}\right) + c_3\varepsilon^{\gamma-2}\frac{R}{\mathcal{K}},$$

where:

$$\mathcal{K} := \inf_{\beta\in\mathcal{B}}\inf_{\alpha\in\mathcal{A}:\alpha\neq\alpha(\beta)}\frac{2KL(\alpha, \alpha(\beta); \beta)}{\|\alpha - \alpha(\beta)\|^2} > 0,$$

and $KL(\alpha, \alpha(\beta); \beta) = L(\alpha(\beta), \beta) - L(\alpha, \beta)$ is the Kullback-Leibler discrepancy between α and $\alpha(\beta)$ for given $\beta \in \mathcal{B}$, the scalar Γ is given by:

$$\Gamma := \sup_{\theta\in\Theta} Tr I(\theta, \theta_0) = \sup_{\theta\in\Theta} E_0\left[\left\|\frac{\partial \log h(y_i, \theta)}{\partial\alpha}\right\|^2\right] < \infty,$$

with $I(\theta, \theta_0) = E_0\left[\dfrac{\partial \log h(y_i, \theta)}{\partial\alpha}\dfrac{\partial \log h(y_i, \theta)}{\partial\alpha'}\right]$, and $Vol(\mathcal{B}) = \int_{\mathcal{B}} d\lambda$ is the Lebesgue measure of $\mathcal{B}$.

Proof: See Appendix B.5.

Appendix 3

Regularity conditions

The regularity conditions used to derive the large sample properties of the estimators are given below.

H.1: *The parameter sets* $\mathcal{B} \subset \mathbb{R}^q$ *and* $\Theta \subset \mathbb{R}^p$ *are compact. The true parameter values* β_0 *and* θ_0 *are interior points of* $\mathcal{B}$ *and* Θ*, respectively.*

H.2: *The function* $h_{i,t}(\beta) := \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)$ *is such that* $E_0\left[|h_{i,t}(\beta)|^9\right] < \infty$*, for any* $\beta \in \mathcal{B}$*,* $E_0\left[\sup_{\beta\in\mathcal{B}} |h_{i,t}(\beta)|^4\right] < \infty$*, and* $E_0\left[\sup_{\beta\in\mathcal{B}} \left\| \frac{\partial h_{i,t}(\beta)}{\partial \beta} \right\|^4\right] < \infty$.

H.3: *For any* $\beta \in \mathcal{B}$*:* (i) $E_0\left[\|f_t(\beta) - E_0[f_t(\beta)|f_t, ..., f_{t-m}]\|^2\right] = O\left(m^{-\alpha}\right)$*, and* (ii) $E_0\left[\|E_0\left[h_{i,t}(\beta)|\underline{f_t}\right] - E_0[h_{i,t}(\beta)|f_t, ..., f_{t-m}]\|^2\right] = O\left(m^{-\alpha}\right)$*, for some* $\alpha > 0$*, as* $m \to \infty$.

H.4: $E_0\left[\exp\left(-u\xi_t\right)\right] \leq C_1 \exp\left(-C_2 u^\delta\right)$ *as* $u \to \infty$*, for some positive constants* $C_1, C_2, \delta > 0$*, where* $\xi_t = \left(1 + \sup_{\beta\in\mathcal{B}} E_0\left[|h_{i,t}(\beta)|^2|\underline{f_t}\right]\right)^{-1}$.

H.5: *Assumption H.4 holds for* $\tilde{\xi}_t = \left(1 + \sup_{\beta\in\mathcal{B}} E_0\left[|b_{i,t}(\beta)|^2|\underline{f_t}\right]\right)^{-1}$ *where* $b_{i,t}(\beta) := \sup_{f:\|f-f_t(\beta)\|\leq\eta^*} \left\| \frac{\partial \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial f} \right\|$, $\eta^* > 0$.

H.6: *Function* $H_{i,t}(\beta) = \left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial(f', \beta')'\partial(f', \beta')}\right]_{f=f_t(\beta)}$ *satisfies Assumptions H.2 and H.3 (ii).*

H.7: *Matrix* $I(t, \beta) := E_0\left[H_{i,t}(\beta)|\underline{f_t}\right]$ *is positive definite, for any* t *and* $\beta \in \mathcal{B}$*,* P*-a.s., and the smallest eigenvalue* $\lambda_t(\beta)$ *of* $I(t, \beta)$ *is such that* $E_0\left[\left(\inf_{\beta\in\mathcal{B}} \lambda_t(\beta)\right)^{-4}\right] < \infty$.

H.8: $E_0\left[\exp\left(-u\xi_t\right)\right] \leq C_1 \exp\left(-C_2 u^\delta\right)$ *as* $u \to \infty$*, for some positive constants* $C_1, C_2, \delta > 0$*, where* $\xi_t = \left[\inf_{\beta\in\mathcal{B}} \lambda_t(\beta)\right] \left[1 + \sup_{\beta\in\mathcal{B}} \frac{\sigma_t^2(\beta)}{\lambda_t(\beta)}\right]^{-1}$ *and* $\sigma_t^2(\beta) = E_0\left[\|H_{i,t}(\beta)\|^2|\underline{f_t}\right]$.

H.9: *Assumption H.8 holds for* $\tilde{\xi}_t = \left[\inf_{\beta\in\mathcal{B}} \lambda_t(\beta)\right]\left[1+\sup_{\beta\in\mathcal{B}}\frac{\tilde{\sigma}_t^2(\beta)}{\lambda_t(\beta)}\right]^{-1}$, *where* $\tilde{\sigma}_t^2(\beta) = E_0\left[\tilde{b}_{i,t}(\beta)^2|\underline{f_t}\right]$ *and* $\tilde{b}_{i,t}(\beta) = \sup_{f:\|f-f_t(\beta)\|\le\eta^*}\left\|\frac{\partial^3\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial(f',\beta')'\partial(f',\beta')\partial f}\right\|$, $\eta^*>0$.

H.10: *The process* $\sup_{\beta\in\mathcal{B}}\|f_t(\beta)\|$ *is such that* $P\left[\sup_{\beta\in\mathcal{B}}\|f_t(\beta)\|\ge u\right]\le C_3\exp\left(-C_4u^{1/\varrho}\right)$ *as* $u\to\infty$, *for some constants* $C_3, C_4<\infty$ *and* $\varrho>0$.

H.11: *The set* $\mathcal{F}_n\subset\mathbb{R}^d$ *is compact and convex, for any* $n\in\mathbb{N}$, *and is such that* $B_{\rho_n}(0)\subset\mathcal{F}_n$, *where* $B_{\rho_n}(0)$ *denotes a ball in* $\mathbb{R}^d$ *centered at* 0 *and with radius* $\rho_n=[(2/C_4)\log(n)]^\varrho$.

H.12: *There exists a constant* $a_1\ge 0$ *such that:*

$$\mathcal{K}_t := \inf_{n\ge1}\inf_{\beta\in\mathcal{B}}\inf_{f\in\mathcal{F}_n:f\ne f_t(\beta)}[\log(n)]^{a_1}\frac{2KL_t(f,f_t(\beta);\beta)}{\|f-f_t(\beta)\|^2}>0,$$

for any t, P*-a.s., where* $KL_t(f,f_t(\beta);\beta) = E_0\left[\log\left(\frac{h(y_{i,t}|y_{i,t-1},f_t(\beta);\beta)}{h(y_{i,t}|y_{i,t-1},f;\beta)}\right)|\underline{f_t}\right]$ *is the conditional Kullback-Leibler discrepancy between* f *and* $f_t(\beta)$ *given the factor path* $\underline{f_t}$.

H.13: *There exist constants* $\gamma\ge4$ *and* $a_2\ge0$ *such that:*

$$R_t := \sup_{n\ge1}[\log(n)]^{-a_2}E_0\left[\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}\left\|\frac{\partial\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial(f',\beta')'}\right\|^\gamma|\underline{f_t}\right]<\infty,$$

for any t, P*-a.s.*

H.14: *There exist constants* $C_5, C_6<\infty$, $a_3>0$ *such that:*

$$E_0\left[\exp\left(-u\frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right)\right]\le C_5\exp\left(-C_6u^{1/a_3}\right),\ \text{as } u\to\infty,$$

where $\Gamma_t := \sup_{n\ge1}\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}[\log(n)]^{-a_2}\,TrE_0\left[\frac{\partial\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial f}\frac{\partial\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial f'}|\underline{f_t}\right]$.

H.15: *It holds* $E_0\left[\frac{R_t}{K_t}\right]<\infty$.

H.16: *The function* $G(F_t,\theta)=\log g(f_t|f_{t-1};\theta)$, *where* $F_t=(f_t,f_{t-1})$, *is Lipschitz continuous w.r.t.* $F_t\in\mathbb{R}^{2K}$, *and such that* $E_0\left[\|G(F_t(\beta),\theta)\|^\kappa\right]<\infty$, $\kappa>2$, *for any* $\beta\in\mathcal{B}$ *and* $\theta\in\Theta$, *and* $E\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial G(F_t(\beta),\theta)}{\partial\beta}\right\|\right]<\infty$, $E\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial G(F_t(\beta),\theta)}{\partial\theta}\right\|\right]<\infty$.

H.17: $P\left[\zeta_t \geq u\right] \leq C_7 \exp\left(-C_8 u^{1/\chi}\right)$*, as* $u \to \infty$*, for some constants* $C_7, C_8, \chi > 0$*, where* $\zeta_t = \sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\sup_{F:\|F-F_t(\beta)\|\leq\eta^*}\left\|\frac{\partial G(F,\theta)}{\partial F}\right\|$*,* $\eta^* > 0$*, and* $G(F_t,\theta) = \log g(f_t|f_{t-1};\theta)$ *.*

H.18: *Assumptions H.16 and H.17 are satisfied for* $G(F_t,\theta) = \frac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial\theta'}$*,* $= \frac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial f_t'}$*, and* $= \frac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial f_{t-1}'}$*.*

H.19: $E_0\left[\left\|\frac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}\right\|^{\nu}\right] < \infty, \nu > 2.$

Assumption H.1 is a standard condition on parameter sets and true parameter values. Assumptions H.2-H.15 concern the micro log-density and the pseudo-true factor values. Specifically, Assumption H.2 requires finite higher-order moments for $\log h(y_{i,t}|y_{i,t-1}, f_t(\beta);\beta)$, and for its derivative w.r.t. β, uniformly in $\beta \in \mathcal{B}$. Under Assumption H.3, the pseudo-true factor value $f_t(\beta)$, and the conditional expectation of $\log h(y_{i,t}|y_{i,t-1}, f_t(\beta);\beta)$ given the factor path $\underline{f_t}$, can be approximated by the conditional expectation given a finite number of past factor values. Assumption H.4 concerns the decay behaviour of the stationary Laplace transform of process ξ_t, when the argument is large. This is a tail condition on the stationary distribution of process $\sup_{\beta\in\mathcal{B}} E_0\left[|h_{i,t}(\beta)|^2|\underline{f_t}\right]$. Assumption H.5 is used to control the effect of replacing $f_t(\beta)$ in $\log h(y_{i,t}|y_{i,t-1}, f_t(\beta);\beta)$ by the cross-sectional estimator $\hat{f}_{n,t}(\beta)$. Assumptions H.6-H.9 are similar to Assumptions H.2-H.5 and concern the second-order derivative matrix $\left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial(f',\beta')'\partial(f',\beta')}\right]_{f=f_t(\beta)}$. In particular, Assumption H.7 implies the concavity of the cross-sectional likelihood function $E_0\left[\log h(y_{i,t}|y_{i,t-1}, f;\beta)\right]$ w.r.t. (f,β), at $f = f_t(\beta)$ and $\beta \in \mathcal{B}$, P-a.s.. Since $I(t,\beta_0) = I(t)$, where matrix $I(t)$ is defined in (3.8), Assumption H.7 strengthens identification Assumptions A.6 and A.7 for micro-parameter β. The concavity condition in Assumption H.7 is made uniform w.r.t. β and the factor path, through the restriction on the distribution of the infimum over $\mathcal{B}$ of the smallest eigenvalue of $I(t,\beta)$. Assumptions H.8 and H.9 are tail conditions on the stationary distribution of $\inf_{\beta\in\mathcal{B}} \lambda_t(\beta)$ in a neighbourhood of zero, and for large values of the ratios $\sup_{\beta\in\mathcal{B}} \frac{\sigma_t^2(\beta)}{\lambda_t(\beta)}$ and $\sup_{\beta\in\mathcal{B}} \frac{\tilde{\sigma}_t^2(\beta)}{\lambda_t(\beta)}$. These conditions are satisfied, when the factor paths associated with very small eigenvalues of $I(t,\beta)$ are sufficiently unfrequent, and when the conditional second-order moments of $H_{i,t}(\beta)$ and $\tilde{b}_{i,t}(\beta)$ are sufficiently small compared to $\lambda_t(\beta)$, uniformly in β. Function $\tilde{b}_{i,t}(\beta)$ measures the sensitivity of the second-order derivative of the log-density w.r.t. the factor value. Assumptions H.1-H.9 are used in Appendix A.6.2 to prove the uniform convergence of the likelihood function $\mathcal{L}^*_{nT}(\beta)$ defined in (3.4), and of its second-order derivative w.r.t. β, using Lemmas A.1, A.2 and Corollary A.3 given in Appendix A.1.

Assumptions H.10-H.15 are used in Lemma A.6 to derive the uniform rate of convergence of the factor approximations (see Appendix A.4). Specifically, Assumption H.10 concerns the tail of the stationary distribution of the process $\sup_{\beta \in \mathcal{B}} \|f_t(\beta)\|$. The parameter set $\mathcal{F}_n$ is allowed to grow at a logarithmic rate as $n \to \infty$. Assumption H.11 gives a lower bound on this growth rate. Under Assumptions H.10 and H.11, the pseudo-true factor value $f_t(\beta)$ is in $\mathcal{F}_n$, for any $1 \leq t \leq T$ and $\beta \in \mathcal{B}$, with probability approaching 1 at rate $O(T/n^2)$. Assumption H.12 concerns the identifiability of the factor values. For any given n, the conditional Kullback-Leibler discrepancy between $f \in \mathcal{F}_n$ and $f_t(\beta)$ given $\underline{f_t}$ is bounded from below by a quadratic function proportional to the squared distance $\|f - f_t(\beta)\|^2$, uniformly in $\beta \in \mathcal{B}$. The scale factor converges to zero at a logarithmic rate, as parameter set $\mathcal{F}_n$ grows. Assumption H.13 introduces a uniform bound on the higher-order moments of the score of the log-density w.r.t. factor value $f \in \mathcal{F}_n$ and parameter $\beta \in \mathcal{B}$. The moment of order $\gamma \geq 4$ is allowed to diverge at a logarithmic rate as $\mathcal{F}_n$ grows. The logarithmic rates in Assumptions H.12 and H.13 imply an upper bound on the growth rate of set $\mathcal{F}_n$. Assumption H.14 is a tail condition on the stationary distribution of the process $\dfrac{\mathcal{K}_t}{1 + \Gamma_t/\mathcal{K}_t}$ in a neighbourhood of 0. The quantity $\dfrac{\mathcal{K}_t}{1 + \Gamma_t/\mathcal{K}_t}$ involves the measure of Kullback-Leibler discrepancy $\mathcal{K}_t$, and the measure Γ_t of second-order moment of the score of the log-density w.r.t. f_t, which are functions of the factor path $\underline{f_t}$. Assumption H.14 is satisfied when the probability mass of $\mathcal{K}_t$ in a neighbourhood of zero, and the probability mass for large values of the ratio $\Gamma_t/\mathcal{K}_t$, are small. Finally, Assumption H.15 is an additional condition on the left tail of the stationary distribution of process $\mathcal{K}_t$, and the right tail of R_t.

Finally, Assumptions H.16-H.19 concern the macro log-density and its derivatives w.r.t. factor values and macro-parameter θ. Specifically, Assumption H.16 requires finite moments for $\log g(f_t(\beta)|f_{t-1}(\beta);\theta)$ and its first-order derivatives w.r.t. β and θ. Assumption H.17 is a condition on the right tail of process ζ_t. This assumption is used to prove a WLLN for time series averages with true factor values replaced by cross-sectional estimators (see Lemma A.7 in Appendix 4). Assumption H.19 is a bound on the moment of the macro-score $\dfrac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}$ of order $\nu > 2$. Assumptions H.16-H.19 imply the uniform convergence of $\mathcal{L}_{1,nT}(\beta,\theta)$ [see (3.5)] and the Hessian $\dfrac{\partial^2\mathcal{L}_{1,nT}(\beta,\theta)}{\partial\theta\partial\theta'}$, uniformly in $\beta \in \mathcal{B}$, $\theta \in \Theta$, as well as the asymptotic normality of the score $\dfrac{\partial\mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\theta}$ in the proof of Proposition 5 (see Appendix 6). These assumptions are also used to prove the asymptotic efficiency of the estimator of θ in Proposition 7.

Appendix 4

Uniform rate of convergence of the cross-sectional factor approximations

Let us derive the uniform rate of convergence of the cross-sectional approximations of the factor values:

$$\widehat{f}_{n,t}(\beta) = \arg\max_{f\in\mathcal{F}_n}\sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1},f;\beta\right),$$

where set $\mathcal{F}_n \subset \mathbb{R}^K$ is compact and tends to $\mathbb{R}^K$ when $n\to\infty$ as defined in Assumption H.11.

A.4.1 Uniform rate of convergence

Lemma A.6: *Under Assumptions A.1-A.5, H.1, H.10-H.15, and if $n, T\to\infty$ such that $T^b/n = O(1)$ for a $b > 1$:*

$$\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| = O_p\left(\sqrt{\frac{(\log n)^a}{n}}\right),$$

$a = 2a_1 + a_2 + a_3 > 0$*, where* a_1, a_2, a_3 *are defined in Assumptions H.12-H.14.*

The logarithmic factor in the uniform convergence rate of $\hat{f}_{n,t}(\beta)$ depends on three parameters. Parameter a_3 controls the tail of the distribution of information measure $\dfrac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}$

in a neighbourhood of zero (see Assumption H.14). Parameters a_1 and a_2 describe the effect of the expanding parameter set $\mathcal{F}_n$ on the identifiability and higher-order moments of the score (see Assumptions H.12, H.13). The uniform rate of convergence in Lemma A.6 is valid when cross-sectional dimension n increases faster than time dimension T.

A.4.2 Proof of Lemma A.6

Let $\varepsilon_n = \sqrt{r\frac{(\log n)^a}{n}}$, where $r > 0$ is a constant. We have to show that for any $\eta > 0$, there exists a value of r such that $P\left[\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n\right] \leq \eta$, for large n and T such that $T^b/n = O(1)$, $b > 1$. We have:

$$\begin{aligned} P\left[\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n\right] &\leq TP\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n\right] \\ &= TE\left[P\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right]\right]. \end{aligned} \tag{A.4}$$

Conditional on factor path $\underline{f_t}$, the estimator $\widehat{f}_{n,t}(\beta)$ is the ML estimator of "parameter" f_t given the "nuisance" parameter β, computed on the sample $(y_{i,t}, y_{i,t-1})$, $i = 1, ..., n$. This sample is i.i.d. conditional on $\underline{f_t}$. Thus, the strategy of the proof is to first use the large deviation result in Lemma A.5 in Appendix 2 to get a bound for $P\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right]$, as a function of $\underline{f_t}$. Then, we compute the expectation of this bound w.r.t. $\underline{f_t}$.

i) Bound of $P\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right]$

Let us first consider the realizations of $\underline{f_t}$ such that $f_t(\beta) \in \mathcal{F}_n$ for any $\beta \in \mathcal{B}$. We apply Lemma A.5 with $l_i(\theta) = \log h(y_{i,t}|y_{i,t-1}, f; \beta)$, $i = 1, ..., n$, and $\theta = (f, \beta) \in \mathcal{F}_n \times \mathcal{B}$. Conditions i) and ii) are implied by Assumptions H.1 and H.11, and A.1-A.2, respectively. Condition iii) is satisfied since Assumption H.12 implies that $f_t(\beta)$ is the unique maximizer of $L_t(f, \beta) = E_0\left[\log h(y_{i,t}|y_{i,t-1}, f; \beta)|\underline{f_t}\right]$ w.r.t. $f \in \mathcal{F}_n$, and that matrix $E_0\left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f\partial f'}|\underline{f_t}\right]$ is non-singular, for any $\beta \in \mathcal{B}$. Condition iv) of Lemma A.5 is implied by Assumption H.13 and:

$$E_0\left[\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}\left\|\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial(f', \beta')'}\right\|^\gamma |\underline{f_t}\right] \leq [log(n)]^{a_2} R_t.$$

Moreover, from Assumption H.12 we know that:

$$\inf_{\beta\in\mathcal{B}} \inf_{f\in\mathcal{F}_n: f\neq f_t(\beta)} \frac{2KL_t(f, f_t(\beta);\beta)}{\|f - f_t(\beta)\|^2} \geq [\log(n)]^{-a_1}\mathcal{K}_t,$$

and:

$$\sup_{\beta\in\mathcal{B}} \sup_{f\in\mathcal{F}_n} TrE_0 \left[\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial f}\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial f'}|\underline{f_t}\right] \leq [log(n)]^{a_2}\Gamma_t,$$

from Assumptions H.13-H.14. Then, from Lemma A.5 we have:

$$\begin{aligned}
& P\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right] \\
\leq\ & c_1\frac{n^{K+q}}{\varepsilon_n^q}\exp\left(-c_2 n\varepsilon_n^2\frac{[\log(n)]^{-a_1}\mathcal{K}_t}{1+[\log(n)]^{a_1+a_2}\Gamma_t/\mathcal{K}_t}\right) + c_3\varepsilon_n^{\gamma-2}[\log(n)]^{a_1+a_2}\frac{R_t}{\mathcal{K}_t} \\
\leq\ & \frac{c_1}{r^{q/2}}n^{K+3q/2}\exp\left(-c_2 r[\log(n)]^{a_3}\frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right) + c_3\frac{r^{\gamma/2-1}}{n^{\gamma/2-1}}[\log(n)]^{a_1(\gamma-1)+(a_2+a_3)\gamma/2-a_3}\frac{R_t}{\mathcal{K}_t},
\end{aligned}$$

for any factor path such that $f_t(\beta) \in \mathcal{F}_n$ for any $\beta \in \mathcal{B}$, where c_1, c_2, c_3 are constants independent of $\underline{f_t}$ and n, T. Thus, we get:

$$\begin{aligned}
& P\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right] \\
\leq\ & \frac{c_1}{r^{q/2}}n^{K+3q/2}\exp\left(-c_2 r[\log(n)]^{a_3}\frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right) + c_3\frac{r^{\gamma/2-1}}{n^{\gamma/2-1}}[\log(n)]^{a_1(\gamma-1)+(a_2+a_3)\gamma/2-a_3}\frac{R_t}{\mathcal{K}_t} \\
& +1\left\{\bigcup_{\beta\in\mathcal{B}}[f_t(\beta)\in\mathcal{F}_n^c]\right\}, P\text{-a.s.}. \qquad (A.5)
\end{aligned}$$

ii) Integrating out the factor path

By integrating out the factor path $\underline{f_t}$, we get from (A.4) and (A.5):

$$\begin{aligned}
& P\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n\right] \\
\leq\ & \frac{c_1}{r^{q/2}}Tn^{K+3q/2}E\left[\exp\left(-c_2 r[\log(n)]^{a_3}\frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right)\right] \\
& +c_3 T\frac{r^{\gamma/2-1}}{n^{\gamma/2-1}}[\log(n)]^{a_1(\gamma-1)+(a_2+a_3)\gamma/2-a_3}E\left[\frac{R_t}{\mathcal{K}_t}\right] + TP\left[\bigcup_{\beta\in\mathcal{B}}[f_t(\beta)\in\mathcal{F}_n^c]\right] \\
=:\ & I_{1,n,T} + I_{2,n,T} + I_{3,n,T}.
\end{aligned}$$

Let us now bound these three terms.

(a) From Assumptions H.14 and using $T/n^{\gamma/2-1} = O(1)$, we have:

$$I_{1,n,T} \leq \frac{c_1}{r^{q/2}} C_5 T n^{K+3q/2} \exp\left(-C_6 (c_2 r)^{1/a_3} \log(n)\right) \leq \frac{c_1}{r^{q/2}} C_5 n^{\gamma/2-1+K+3q/2-C_6(c_2 r)^{a_3}} = o(1),$$

if $r > \frac{1}{c_2}\left(\frac{\gamma/2+K+3q/2-1}{C_6}\right)^{1/a_3}$.

(b) From Assumption H.15, $E\left[\frac{R_t}{\mathcal{K}_t}\right] < \infty$. Then, from the condition $T^b/n = O(1)$ for $b > 1$, and since $\gamma \geq 4$, we get $I_{2,n,T} = o(1)$.

(c) Finally, from Assumptions H.10 and H.11, we have:

$$P\left[\bigcup_{\beta\in\mathcal{B}} [f_t(\beta) \in \mathcal{F}_n^c]\right] \leq P\left[\sup_{\beta\in\mathcal{B}} \|f_t(\beta)\| \geq \rho_n\right] \leq C_3 \exp\left(-C_4 \rho_n^{1/\varrho}\right) \leq C_3 n^{-2}.$$

Since $T/n^2 = o(1)$, we get $I_{3,n,T} = o(1)$. This completes the proof of Lemma A.6.

A.4.3 Uniform WLLN with factor approximations

The uniform rate of convergence of cross-sectional factor approximations (Lemma A.6) can be used to derive uniform WLLN when the true factor values are replaced by their approximations.

Lemma A.7: *Let $F_t := (f_t, f_{t-1})$ and assume that function $G(F,\theta)$ is such that:*

(i) $G(F,\theta)$ *is Lipschitz continuous w.r.t. $F \in \mathbb{R}^{2K}$, for any $\theta \in \Theta$.*

(ii) *For any $\beta \in \mathcal{B}$ and $\theta \in \Theta$: $E_0[\|G(F_t(\beta),\theta)\|^\kappa] < \infty$, $\kappa > 2$, $E\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}} \left\|\frac{\partial vec[G(F_t(\beta),\theta)]}{\partial\beta'}\right\|\right] < \infty$, and $E\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}} \left\|\frac{\partial vec[G(F_t(\beta),\theta)]}{\partial\theta'}\right\|\right] < \infty$.*

(iii) $P[\zeta_t \geq u] \leq c_1 \exp\left(-c_2 u^{1/\chi}\right)$*, as $u \to \infty$, for some constants $c_1, c_2, \chi > 0$, where* $\zeta_t = \sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\sup_{F:\|F-F_t(\beta)\|\leq\eta^*} \left\|\frac{\partial vec[G(F,\theta)]}{\partial F}\right\|$, $\eta^* > 0$.

Then, under Assumptions A.1-A.5, H.1, H.3 (i), H.10-H.15, and if $n, T \to \infty$ such that $T^b/n = O(1)$ for a $b > 1$:

$$\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T} G(\hat{f}_{n,t}(\beta), \hat{f}_{n,t-1}(\beta);\theta) - E_0\left[G(f_t(\beta), f_{t-1}(\beta);\theta)\right]\right| = o_p(1).$$

Proof: See Appendix B.6.

Appendix 5

Proof of Proposition 1

We have:

$$l\left(\underline{y_T};\beta,\theta\right)=\int\cdots\int\exp\left\{\sum_{t=1}^{T}\sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1},f_t;\beta\right)+\sum_{t=1}^{T}\log g\left(f_t|f_{t-1};\theta\right)\right\}\prod_{t=1}^{T}df_t.$$

Let us now expand the integrand w.r.t. f_t around $\hat{f}_{nt}\left(\beta\right)$, $t=1,...,T$, and define:

$$\begin{aligned}\psi_{nt}\left(f_t,f_{t-1}\right) &= \sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1},f_t;\beta\right)-\sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}\left(\beta\right);\beta\right)\\ &+\frac{1}{2}\sqrt{n}\left(f_t-\hat{f}_{nt}\left(\beta\right)\right)^{\prime}I_{nt}\left(\beta\right)\sqrt{n}\left(f_t-\hat{f}_{nt}\left(\beta\right)\right)\\ &+\log g\left(f_t|f_{t-1};\theta\right)-\log g\left(\hat{f}_{nt}\left(\beta\right)|\hat{f}_{n,t-1}\left(\beta\right);\theta\right).\end{aligned}$$

Then:

$$\begin{aligned}l\left(\underline{y_T};\beta,\theta\right) &= \prod_{t=1}^{T}\prod_{i=1}^{n}h\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}\left(\beta\right);\beta\right)\prod_{t=1}^{T}g\left(\hat{f}_{nt}\left(\beta\right)|\hat{f}_{n,t-1}\left(\beta\right);\theta\right)\\ &\int\cdots\int\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\sqrt{n}\left(f_t-\hat{f}_{nt}\left(\beta\right)\right)^{\prime}I_{nt}\left(\beta\right)\sqrt{n}\left(f_t-\hat{f}_{nt}\left(\beta\right)\right)\right\}\\ &\exp\left\{\sum_{t=1}^{T}\psi_{n,t}\left(f_t,f_{t-1}\right)\right\}\prod_{t=1}^{T}df_t.\end{aligned}$$

Let us introduce the change of variable:

$$Z_t=\sqrt{n}\left[I_{nt}\left(\beta\right)\right]^{1/2}\left(f_t-\hat{f}_{nt}\left(\beta\right)\right)\Longleftrightarrow f_t=\hat{f}_{nt}\left(\beta\right)+\frac{1}{\sqrt{n}}\left[I_{nt}\left(\beta\right)\right]^{-1/2}Z_t.$$

Then:

$$\begin{aligned}&l\left(\underline{y_T};\beta,\theta\right)\\ =&\left(\frac{2\pi}{n}\right)^{TK/2}\prod_{t=1}^{T}\left[\det I_{nt}\left(\beta\right)\right]^{-1/2}\prod_{t=1}^{T}\prod_{i=1}^{n}h\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}\left(\beta\right);\beta\right)\prod_{t=1}^{T}g\left(\hat{f}_{nt}\left(\beta\right)|\hat{f}_{n,t-1}\left(\beta\right);\theta\right)\\ &\frac{1}{\left(2\pi\right)^{TK/2}}\int...\int\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}Z_t^{\prime}Z_t\right\}\\ &\exp\left\{\sum_{t=1}^{T}\psi_{n,t}\left(\hat{f}_{n,t}\left(\beta\right)+\frac{1}{\sqrt{n}}\left[I_{n,t}\left(\beta\right)\right]^{-1/2}Z_t,\hat{f}_{n,t-1}\left(\beta\right)+\frac{1}{\sqrt{n}}\left[I_{n,t-1}\left(\beta\right)\right]^{-1/2}Z_{t-1}\right)\right\}\prod_{t=1}^{T}dZ_t.\end{aligned}$$

Thus, function $\Psi_{nT}(\beta,\theta)$ is defined by the Gaussian integral:

$$
\begin{aligned}
& \exp\left[\left(\frac{T}{n}\right)\Psi_{nT}(\beta,\theta)\right] \\
= \;& \frac{1}{(2\pi)^{TK/2}}\int\ldots\int \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} Z_t' Z_t\right\} \\
& \exp\left\{\sum_{t=1}^{T}\psi_{n,t}\left(\hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2}Z_t, \hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2}Z_{t-1}\right)\right\}\prod_{t=1}^{T} dZ_t,
\end{aligned}
$$

which can be made explicit by expanding function $\exp\left\{\sum_{t=1}^{T}\psi_{n,t}\right\}$ in a power series of Z_t, $t=1,\ldots,T$.

To simplify the notation, let us consider the one-factor case, $K=1$. Then:

$$
\begin{aligned}
& \exp\left[\left(\frac{T}{n}\right)\Psi_{nT}(\beta,\theta)\right] \\
= \;& E\left[\exp\left\{\sum_{t=1}^{T}\psi_{n,t}\left(\hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2}Z_t, \hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2}Z_{t-1}\right)\right\}\right],
\end{aligned}
$$

where the expectation is taken with respect to a multivariate standard normal distribution for $Z := (Z_1,\ldots,Z_T)'$. Expanding $\psi_{n,t}$ at order $1/n$ yields:

$$
\begin{aligned}
& \psi_{n,t}\left(\hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2}Z_t, \hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2}Z_{t-1}\right) \\
= \;& \frac{1}{6}\frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-3/2}K_{3,nt}(\beta)Z_t^3+\frac{1}{24}\frac{1}{n}\left[I_{n,t}(\beta)\right]^{-2}K_{4,nt}(\beta)Z_t^4+\ldots \\
& +\frac{1}{\sqrt{n}}D_{10,nt}(\beta,\theta)\left[I_{n,t}(\beta)\right]^{-1/2}Z_t+\frac{1}{\sqrt{n}}D_{01,nt}(\beta,\theta)\left[I_{n,t-1}(\beta)\right]^{-1/2}Z_{t-1} \\
& +\frac{1}{2}\frac{1}{n}D_{20,nt}(\beta,\theta)\left[I_{n,t}(\beta)\right]^{-1}Z_t^2+\frac{1}{2}\frac{1}{n}D_{02,nt}(\beta,\theta)\left[I_{n,t-1}(\beta)\right]^{-1}Z_{t-1}^2 \\
& +\frac{1}{n}D_{11,nt}(\beta,\theta)\left[I_{n,t}(\beta)\right]^{-1/2}\left[I_{n,t-1}(\beta)\right]^{-1/2}Z_t Z_{t-1}+\ldots,
\end{aligned}
$$

where:

$$
K_{m,nt}(\beta)=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^m \log h}{\partial f_t^m}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right),\;\; m=3,4,\ldots,
$$

and:

$$
D_{pq,nt}(\beta,\theta)=\frac{\partial^{p+q}\log g}{\partial f_t^p\partial f_{t-1}^q}\left(\hat{f}_{nt}(\beta)\,|\hat{f}_{n,t-1}(\beta);\theta\right),\;\; p,q=0,1,2,\ldots\;.
$$

By expanding the exponential function $\exp\left\{\sum_{t=1}^{T}\psi_{n,t}\right\}$, and computing the expectation w.r.t. Z, it is seen that terms of orders $n^{-1/2}$, $n^{-3/2}$, ... involve odd power moments of standard normal variables, which are zero. Thus, we get:

$$\exp\left[\left(\frac{T}{n}\right)\Psi_{nT}\left(\beta,\theta\right)\right] = 1 + \frac{T}{n}A_{nT}(\beta,\theta) + o_p(T/n),$$

where:

$$\begin{aligned}
A_{nT}(\beta,\theta) \;=\;& \frac{1}{8}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-2}K_{4,n,t}(\beta) + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-1}D_{20,nt}(\beta,\theta) \\
&+\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-1}D_{02,nt}(\beta,\theta) + \frac{\mu_6}{72}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-3}K_{3,nt}^{2}(\beta) + \\
&\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}D_{10,nt}^{2}(\beta,\theta)\left[I_{n,t}\left(\beta\right)\right]^{-1} + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}D_{01,nt}^{2}(\beta,\theta)\left[I_{n,t-1}\left(\beta\right)\right]^{-1} \\
&+\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-2}D_{10,nt}(\beta,\theta)K_{3,n,t}(\beta) \\
&+\frac{1}{2}\frac{1}{T}\sum_{t=2}^{T}\left[I_{n,t-1}\left(\beta\right)\right]^{-3/2}\left[I_{n,t}\left(\beta\right)\right]^{-1/2}D_{01,nt}(\beta,\theta)K_{3,n,t-1}(\beta) \\
&+\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-1}\left[I_{n,t-1}\left(\beta\right)\right]^{-1}D_{10,n,t-1}(\beta,\theta)D_{01,nt}(\beta,\theta),
\end{aligned} \tag{A.6}$$

and μ_6 denotes the moment of order 6 of the standard normal distribution. From Lemma A.6 in Appendix 3, we know that $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| = O_p(T^{-\rho})$, for a $\rho > 0$. Then, by applying Lemmas A.1-A.2 in Appendix A.1, and Lemma A.7 in Appendix 4, we get $A_{n,T}(\beta,\theta) = O_p(1)$ uniformly in $\beta \in \mathcal{B}$ and $\theta \in \Theta$. Proposition 1 follows.

Appendix 6

Efficiency bound and efficient estimators

Let us derive the efficiency bound and prove the asymptotic efficiency of the estimators introduced in Section 4. We first give in Section A.6.1 a preliminary Lemma, used in Section A.6.2 to derive the efficiency bound (proof of Proposition 3). Then, the asymptotic properties of the estimators of the micro-parameters and the factor values are derived in Sections A.6.3 and A.6.4, respectively (proofs of Propositions 5 and 6, respectively).

A.6.1 A preliminary Lemma

Lemma A.8: *Let the estimator* $\left(\hat{\beta}, \hat{\theta}\right)$ *be defined by:*

$$\left(\hat{\beta}, \hat{\theta}\right) = \arg \max_{\beta \in \mathcal{B}, \theta \in \Theta} \mathcal{L}_{nT}\left(\beta, \theta\right),$$

where $\mathcal{B} \subset \mathbb{R}^q$ *and* $\Theta \subset \mathbb{R}^p$ *are compact sets, and:*

$$\mathcal{L}_{nT}\left(\beta, \theta\right) = \mathcal{L}^*_{nT}\left(\beta\right) + \frac{1}{n}\mathcal{L}_{1,nT}\left(\beta, \theta\right) + \frac{1}{n^2}\mathcal{L}_{2,nT}\left(\beta, \theta\right),$$

is such that:

(1) (i) $\mathcal{L}^*_{nT}(\beta)$ *converges in probability to a function* $\mathcal{L}^*(\beta)$*, uniformly in* $\beta \in \mathcal{B}$*;*
(ii) $\mathcal{L}_{1,nT}(\beta, \theta)$ *converges in probability to a function* $\mathcal{L}_1(\beta, \theta)$*, uniformly in* $\beta \in \mathcal{B}, \theta \in \Theta$.

(2) (i) *Function* $\beta \to \mathcal{L}^*(\beta)$ *is uniquely maximized at the interior point* $\beta_0 \in \mathcal{B}$*;*
(ii) *Function* $\theta \to \mathcal{L}_1(\beta_0, \theta)$ *is uniquely maximized at the interior point* $\theta_0 \in \Theta$*.*

(3) (i) *The matrix* $-\frac{\partial^2 \mathcal{L}^*_{nT}(\beta)}{\partial \beta \partial \beta'}$ *is well-defined and converges in probability to* $I^*\left(\beta\right)$*, uniformly in* $\beta \in \mathcal{B}$*, with* $I^*_0 := I^*\left(\beta_0\right)$ *positive definite;* (ii) *The matrix* $-\frac{\partial^2 \mathcal{L}_{1,nT}(\beta, \theta)}{\partial \theta \partial \theta'}$ *is well-defined and converges in probability to* $I_{1,\theta\theta}\left(\beta, \theta\right)$*, uniformly in* $\beta \in \mathcal{B}, \theta \in \Theta$*, with* $I_{1,\theta\theta} := I_{1,\theta\theta}\left(\beta_0, \theta_0\right)$ *positive definite*; (iii) $\sup_{\beta \in \mathcal{B}, \theta \in \Theta} \left\| \frac{\partial^2 \mathcal{L}_{1,nT}(\beta, \theta)}{\partial \beta \partial \beta'} \right\| = O_p(1)$ *and* $\sup_{\beta \in \mathcal{B}, \theta \in \Theta} \left\| \frac{\partial^2 \mathcal{L}_{1,nT}(\beta, \theta)}{\partial \beta \partial \theta'} \right\| = O_p(1)$.

(4) (i)

$$\left[\begin{array}{c}\sqrt{nT}\frac{\partial\mathcal{L}_{nT}^{*}(\beta_0)}{\partial\beta}\\ \sqrt{T}\frac{\partial\mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\theta}\end{array}\right]\xrightarrow{d}N\left(\left[\begin{array}{c}0\\0\end{array}\right],\left[\begin{array}{cc}I_0^{*} & 0\\ 0 & I_{1,\theta\theta}\end{array}\right]\right);$$

(ii) $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\dfrac{\partial\mathcal{L}_{1,nT}\left(\beta,\theta\right)}{\partial\beta}=O_p(1).$

(5) (i) $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\mathcal{L}_{2,nT}(\beta,\theta)=O_p(1)$; (ii) $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\dfrac{\partial\mathcal{L}_{2,nT}\left(\beta,\theta\right)}{\partial\left(\beta',\theta'\right)'}\right\|=O_p(1).$

Moreover, let:

$$\hat{\beta}_{nT}^{*}=\arg\max_{\beta\in\mathcal{B}}\mathcal{L}_{nT}^{*}\left(\beta\right).$$

Then, if $n,T\to\infty$ such that $T/n\to 0$, the estimators $\hat{\beta}$ and $\hat{\theta}$ are consistent and jointly asymptotically normal:

$$\left[\begin{array}{c}\sqrt{nT}\left(\hat{\beta}-\beta_0\right)\\ \sqrt{T}\left(\hat{\theta}-\theta_0\right)\end{array}\right]\xrightarrow{d}N\left(\left[\begin{array}{c}0\\0\end{array}\right],\left[\begin{array}{cc}(I_0^{*})^{-1} & 0\\ 0 & I_{1,\theta\theta}^{-1}\end{array}\right]\right).$$

Moreover, $\hat{\beta}$ and $\hat{\beta}_{nT}^{*}$ are asymptotically equivalent, that is, $\sqrt{nT}\left(\hat{\beta}-\hat{\beta}_{nT}^{*}\right)=o_p(1)$.

Proof: See Appendix B.7.

A.6.2 Proof of Proposition 3

The efficiency bound B^* is the asymptotic variance-covariance matrix of the ML estimator $\left(\hat{\beta},\hat{\theta}\right)=\arg\max_{\beta\in\mathcal{B},\theta\in\Theta}\mathcal{L}_{nT}\left(\beta,\theta\right)$, where $\mathcal{L}_{nT}\left(\beta,\theta\right)$ is defined in Corollary 2. This asymptotic variance-covariance matrix is derived by applying Lemma A.8. Let us verify the conditions of Lemma A.8.

Condition (1) of Lemma A.8: We have:

$$\mathcal{L}_{nT}^{*}\left(\beta\right)=\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}\left(\beta\right);\beta\right). \tag{A.7}$$

This converges to $\mathcal{L}^{*}\left(\beta\right)=E_0\left[\log h\left(y_{i,t}|y_{i,t-1},f_t\left(\beta\right);\beta\right)\right]$ in probability, uniformly in $\beta\in\mathcal{B}$, by using Lemma A.1 in Appendix 1, with $a(Y_{i,t},f_t,\beta)=\log h(y_{i,t}|y_{i,t-1},f_t;\beta)$ and φ corresponding to the identity mapping. Indeed, condition (1) of Lemma A.1 is implied by Assumptions H.1-H.5, and condition (3) of Lemma A.1 is implied by Lemma A.6. Further:

$$\mathcal{L}_{1,nT}(\beta,\theta) = -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\log\det I_{nt}(\beta) + \frac{1}{T}\sum_{t=1}^{T}\log g\left(\hat{f}_{nt}(\beta)\,|\hat{f}_{n,t-1}(\beta);\theta\right), \quad \text{(A.8)}$$

converges to:

$$\mathcal{L}_1(\beta,\theta) = -\frac{1}{2}E_0\left[\log\det I_{ff}(t;\beta)\right] + E_0\left[\log g\left(f_t(\beta)\,|f_{t-1}(\beta);\theta\right)\right],$$

uniformly in $\theta \in \Theta$, $\beta \in \mathcal{B}$, where $I_{ff}(t;\beta) = E_0\left[-\frac{\partial^2 \log h}{\partial f \partial f'}\left(y_{i,t}|y_{i,t-1}, f_t(\beta);\beta\right)|\underline{f_t}\right]$ (use Lemma A.7 in Appendix A.4.3 and Assumptions H.1, H.3 (i), H.16, H.17).

Condition (2) of Lemma A.8: Statement (i) follows from Assumptions A.6 and H.1. Statement (ii) follows from Assumptions A.8 and H.1, by using $\mathcal{L}_1(\beta_0,\theta) = E_0\left[\log g\left(f_t|f_{t-1};\theta\right)\right]$, up to a constant in θ.

Condition (3) of Lemma A.8: From (A.7), we get by differentiation:

$$\begin{aligned}
\frac{\partial \mathcal{L}^*_{nT}(\beta)}{\partial \beta} &= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial \log h}{\partial \beta}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right) \\
&\quad + \frac{1}{nT}\sum_{t=1}^{T}\frac{\partial \hat{f}_{nt}(\beta)'}{\partial \beta}\underbrace{\sum_{i=1}^{n}\frac{\partial \log h}{\partial f_t}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right)}_{=0} \\
&= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial \log h}{\partial \beta}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right),
\end{aligned}$$

and:

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}^*_{nT}(\beta)}{\partial \beta \partial \beta'} &= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial^2 \log h}{\partial \beta \partial \beta'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right) \\
&\quad + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial^2 \log h}{\partial \beta \partial f_t'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right)\frac{\partial \hat{f}_{nt}(\beta)}{\partial \beta'}.
\end{aligned}$$

By differentiating the f.o.c.

$$\sum_{i=1}^{n}\frac{\partial \log h}{\partial f_t}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right) = 0$$

w.r.t. β, we get:

$$\sum_{i=1}^{n}\frac{\partial^2 \log h}{\partial f_t \partial \beta'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right) + \sum_{i=1}^{n}\frac{\partial^2 \log h}{\partial f_t \partial f_t'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right)\frac{\partial \hat{f}_{nt}(\beta)}{\partial \beta'} = 0.$$

Let us introduce the notation:

$$\hat{I}_{\beta\beta}(t) := -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log h}{\partial\beta\partial\beta'}\left(y_{i,t}|y_{i,t-1}, \hat{f}_{nt}\left(\beta\right); \beta\right),$$

and similarly $\hat{I}_{\beta f}(t)$, $\hat{I}_{ff}(t)$. Then we get:

$$\frac{\partial \hat{f}_{nt}\left(\beta\right)}{\partial\beta'} = -\hat{I}_{ff}(t)^{-1}\hat{I}_{f\beta}(t),$$

and

$$-\frac{\partial^2 \mathcal{L}^*_{nT}\left(\beta\right)}{\partial\beta\partial\beta'} = \frac{1}{T}\sum_{t=1}^{T}\left[\hat{I}_{\beta\beta}(t) - \hat{I}_{\beta f}(t)\hat{I}_{ff}(t)^{-1}\hat{I}_{f\beta}(t)\right].$$

Thus, condition (3i) is satisfied with $I^*_0 = E\left[I_{\beta\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}I_{f\beta}(t)\right]$ by applying Corollary A.3 in Appendix 1, case (B). Indeed, condition (1) of Lemma A.2 is implied by Assumptions H.1-H.9, and condition (3) of Lemma A.2 is implied by Lemma A.6.

Moreover, from (A.8) we have:

$$\frac{\partial \mathcal{L}_{1,nT}\left(\beta,\theta\right)}{\partial\theta} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log g}{\partial\theta}\left(\hat{f}_{nt}\left(\beta\right)|\hat{f}_{n,t-1}\left(\beta\right); \theta\right),$$

and:

$$\frac{\partial^2 \mathcal{L}_{1,nT}\left(\beta,\theta\right)}{\partial\theta\partial\theta'} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log g}{\partial\theta\partial\theta'}\left(\hat{f}_{nt}\left(\beta\right)|\hat{f}_{n,t-1}\left(\beta\right); \theta\right).$$

Thus, condition (3ii) is satisfied with $I_{1,\theta\theta} = E\left[-\frac{\partial^2 \log g}{\partial\theta\partial\theta'}\left(f_t|f_{t-1};\theta_0\right)\right]$ (use Lemmas A.6-A.7 and Assumption H.18).

Condition (4) of Lemma A.8: Let us first consider the approximated score w.r.t. β. We have:

$$\sqrt{nT}\frac{\partial \mathcal{L}^*_{nT}\left(\beta_0\right)}{\partial\beta} = \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial \log h}{\partial\beta}\left(y_{i,t}|y_{i,t-1}, \hat{f}_{nt}\left(\beta_0\right); \beta_0\right).$$

By the mean-value Theorem:

$$\begin{aligned}\sqrt{nT}\frac{\partial \mathcal{L}^*_{nT}\left(\beta_0\right)}{\partial\beta} &= \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial \log h}{\partial\beta}\left(y_{i,t}|y_{i,t-1}, f_t; \beta_0\right) \\ &+\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial^2 \log h}{\partial\beta\partial f_t'}\left(y_{i,t}|y_{i,t-1}, \tilde{f}_t; \beta_0\right)\left(\hat{f}_{nt}\left(\beta_0\right) - f_t\right),\end{aligned}$$

where $\tilde{f}_t$ are mean values. Using the notation:

$$\tilde{I}_{\beta f}(t) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial \beta \partial f_t'}\left(y_{i,t}|y_{i,t-1}, \tilde{f}_t; \beta_0\right),$$

and the expansion of $\hat{f}_{nt}(\beta_0)$:

$$\sqrt{n}\left(\hat{f}_{nt}(\beta_0) - f_t\right) = -\bar{I}_{ff}(t)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log h}{\partial f_t}(y_{i,t}|y_{i,t-1}, f_t; \beta_0), \tag{A.9}$$

where $\bar{I}_{ff}(t)$ is based on a mean value $\bar{f}_t$, we get:

$$\begin{aligned}\sqrt{nT}\frac{\partial \mathcal{L}_{nT}^{*}(\beta_0)}{\partial \beta} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log h}{\partial \beta}(y_{i,t}|y_{i,t-1}, f_t; \beta_0)\right.\\ &\qquad \left. - \tilde{I}_{\beta f}(t)\bar{I}_{ff}(t)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log h}{\partial f_t}(y_{i,t}|y_{i,t-1}, f_t; \beta_0)\right].\end{aligned}$$

Then, we get:

$$\sqrt{nT}\frac{\partial \mathcal{L}_{nT}^{*}(\beta_0)}{\partial \beta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\psi_\beta(t) - I_{\beta f}(t)I_{ff}(t)^{-1}\psi_f(t)\right] + o_p(1), \tag{A.10}$$

where:

$$\psi(t) := \begin{bmatrix}\psi_\beta(t)\\ \psi_f(t)\end{bmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix}\frac{\partial \log h}{\partial \beta}(y_{i,t}|y_{i,t-1}, f_t; \beta_0)\\ \frac{\partial \log h}{\partial f_t}(y_{i,t}|y_{i,t-1}, f_t; \beta_0)\end{bmatrix}.$$

Let us now consider the approximated score w.r.t. θ. By the mean-value Theorem, we have:

$$\begin{aligned}\sqrt{T}\frac{\partial \mathcal{L}_{1,nT}(\beta_0, \theta_0)}{\partial \theta} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g}{\partial \theta}\left(\hat{f}_{nt}(\beta_0)\,|\,\hat{f}_{n,t-1}(\beta_0); \theta_0\right)\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g}{\partial \theta}(f_t|f_{t-1}; \theta_0)\\ &\quad + \sqrt{\frac{T}{n}}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log g}{\partial \theta \partial f_t'}\left(\tilde{f}_t|\tilde{f}_{t-1}; \theta_0\right)\sqrt{n}\left(\hat{f}_{nt}(\beta_0) - f_t\right)\right.\\ &\quad \left. + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log g}{\partial \theta \partial f_{t-1}'}\left(\tilde{f}_t|\tilde{f}_{t-1}; \theta_0\right)\sqrt{n}\left(\hat{f}_{n,t-1}(\beta_0) - f_{t-1}\right)\right).\end{aligned}$$

By using $T^b/n = O(1)$, $b > 1$, Assumption H.18 and Lemmas A.6 and A.7, it follows that:

$$\sqrt{T}\frac{\partial \mathcal{L}_{1,nT}(\beta_0, \theta_0)}{\partial \theta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g}{\partial \theta}(f_t|f_{t-1}; \theta_0) + o_p(1). \tag{A.11}$$

Thus, from (A.10) and (A.11) we deduce:

$$\begin{bmatrix} \sqrt{nT}\frac{\partial \mathcal{L}^*_{nT}(\beta_0)}{\partial \beta} \\ \sqrt{T}\frac{\partial \mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial \theta} \end{bmatrix} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{bmatrix} (\psi_\beta(t) - I_{\beta f}(t) I_{ff}(t)^{-1} \psi_f(t)) \\ \frac{\partial \log g}{\partial \theta} (f_t | f_{t-1}; \theta_0) \end{bmatrix} + o_p(1).$$

By using $E\left[\psi(t)|\underline{y_{t-1}}, \underline{f_t}\right] = 0$, $V\left[\psi_\beta(t) - I_{\beta f}(t) I_{ff}(t)^{-1}\psi_f(t)\right] = E\left[I_{\beta\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} I_{f\beta}(t)\right]$ and a CLT for martingale difference sequence, we get (4i).

Condition (5) of Lemma A.8: By using $\mathcal{L}_{2,nT}(\beta,\theta) = \Psi_{nT}(\beta,\theta)$, condition (i) is implied by Proposition 1.

From Lemma A.8 we deduce the efficiency bound.

A.6.3 Proof of Proposition 5

From Lemma A.8, it follows that $\sqrt{nT}\left(\hat{\beta} - \hat{\beta}^*_{nT}\right) = o_p(1)$. The conclusion follows.

A.6.4 Proof of Proposition 6

We have:

$$\sqrt{n}\left(\hat{f}_{nT,t} - f_t\right) = \sqrt{n}\left(\hat{f}_{n,t}(\beta_0) - f_t\right) + \frac{\partial \hat{f}_{n,t}\left(\dot{\beta}\right)}{\partial \beta'} \sqrt{n}\left(\hat{\beta}^*_{nT} - \beta_0\right),$$

where $\dot{\beta}$ is a mean value. The second term in the RHS is $O_p(1/\sqrt{T})$ from Proposition 5. Thus, point i) follows from expansion (A.9). Point ii) follows from Lemma A.6.

Appendix 7

Factor ordered qualitative model

A.7.1. Identification

i) Let us first consider the two-state case, $K = 2$. The transition matrix $\pi_t = [\pi_{lk,t}]$ is:

$$\pi_t = \begin{bmatrix} G\left(\frac{a_1 - \alpha_1 f_t - \gamma_1}{\sigma_1}\right) & 1 - G\left(\frac{a_1 - \alpha_1 f_t - \gamma_1}{\sigma_1}\right) \\ G\left(\frac{a_1 - \alpha_2 f_t - \gamma_2}{\sigma_2}\right) & 1 - G\left(\frac{a_1 - \alpha_2 f_t - \gamma_2}{\sigma_2}\right) \end{bmatrix}.$$

By reparametrizing coefficients γ_1 and γ_2, we can assume $a_1 = 0$. The transition matrix becomes:

$$\pi_t = \begin{bmatrix} G\left(-\frac{\alpha_1 f_t + \gamma_1}{\sigma_1}\right) & 1 - G\left(-\frac{\alpha_1 f_t + \gamma_1}{\sigma_1}\right) \\ G\left(-\frac{\alpha_2 f_t + \gamma_2}{\sigma_2}\right) & 1 - G\left(-\frac{\alpha_2 f_t + \gamma_2}{\sigma_2}\right) \end{bmatrix}.$$

We can also scale the parameters to get $\sigma_1 = \sigma_2 = 1$:

$$\pi_t = \begin{bmatrix} G\left(-\alpha_1 f_t - \gamma_1\right) & 1 - G\left(-\alpha_1 f_t - \gamma_1\right) \\ G\left(-\alpha_2 f_t - \gamma_2\right) & 1 - G\left(-\alpha_2 f_t - \gamma_2\right) \end{bmatrix}.$$

Finally, by standardizing the factor, we can set $\alpha_1 = 1$ and $\gamma_1 = 0$:

$$\pi_t = \begin{bmatrix} G\left(-f_t\right) & 1 - G\left(-f_t\right) \\ G\left(-\alpha_2 f_t - \gamma_2\right) & 1 - G\left(-\alpha_2 f_t - \gamma_2\right) \end{bmatrix}.$$

Then, the values of the factor f_t are identified by the first row of the transition matrix, $t = 1, ..., T$. The values of α_2, γ_2 are identified by the second row, when $T \geq 2$.

ii) Let us now consider the case $K > 2$. The l-th row of the transition matrix is:

$$\left[G\left(\frac{a_1 - \alpha_l f_t - \gamma_l}{\sigma_l}\right), G\left(\frac{a_2 - \alpha_l f_t - \gamma_l}{\sigma_l}\right) - G\left(\frac{a_1 - \alpha_l f_t - \gamma_l}{\sigma_l}\right), ..., 1 - G\left(\frac{a_{K-1} - \alpha_l f_t - \gamma_l}{\sigma_l}\right)\right],$$

for $l = 1, ..., K$. As above, we can first set $a_1 = 0$:

$$\left[G\left(-\frac{\alpha_l f_t + \gamma_l}{\sigma_l}\right), G\left(\frac{a_2 - \alpha_l f_t - \gamma_l}{\sigma_l}\right) - G\left(-\frac{\alpha_l f_t + \gamma_l}{\sigma_l}\right), ..., 1 - G\left(\frac{a_{K-1} - \alpha_l f_t - \gamma_l}{\sigma_l}\right)\right]. \tag{A.12}$$

Second, by normalizing the factor values and the thresholds, we can set $\alpha_1 = \sigma_1 = 1$ and $\gamma_1 = 0$ in the first row. Then, the transition matrix has a first row given by:

$$\left[G\left(-f_t\right), G\left(a_2 - f_t\right) - G\left(-f_t\right), ..., 1 - G\left(a_{K-1} - f_t\right)\right],$$

and row l is given by (A.12) for $l \geq 2$. From the first row, we can identify the factor value f_t and the $K-2$ thresholds $a_2, ..., a_K$. Then, the values of $\alpha_l, \gamma_l, \sigma_l$ are identified by the row l, for $l = 2, ..., K$, when $(K-1)T \geq 3$.

A.7.2 Semi-parametric efficiency bound [Proof of Equation (5.4)]

We have:

$$\log h\left(y_{i,t}|y_{i,t-1}, f_t; \beta\right) = \sum_{k=1}^{K} \sum_{l=1}^{K} 1\left\{y_{i,t} = k, y_{i,t-1} = l\right\} \log \pi_{lk}\left(f_t, \beta\right),$$

where $\pi_{lk}\left(f_t, \beta\right) = G\left(\frac{a_k - \alpha_l f_t - \gamma_l}{\sigma_l}\right) - G\left(\frac{a_{k-1} - \alpha_l f_t - \gamma_l}{\sigma_l}\right)$. Thus:

$$-\frac{\partial^2 \log h\left(y_{i,t}|y_{i,t-1}, f_t; \beta\right)}{\partial\left(\beta', f'\right)' \partial\left(\beta', f'\right)} = \sum_{k=1}^{K} \sum_{l=1}^{K} 1\left\{y_{i,t} = k, y_{i,t-1} = l\right\} \frac{1}{\pi_{lk}\left(f_t, \beta\right)} J_{lk}\left(f_t, \beta\right),$$

where:

$$J_{lk} = -\frac{\partial^2 \pi_{lk}}{\partial\left(\beta', f'\right)' \partial\left(\beta', f'\right)} + \frac{1}{\pi_{lk}} \frac{\partial \pi_{lk}}{\partial\left(\beta', f'\right)'} \frac{\partial \pi_{lk}}{\partial\left(\beta', f'\right)}.$$

The conditional information matrix is given by:

$$I(t) = E_0\left[-\frac{\partial^2 \log h\left(y_{i,t}|y_{i,t-1}, f_t; \beta_0\right)}{\partial\left(\beta', f'\right)' \partial\left(\beta', f'\right)}\middle| \underline{f_t}\right] = \sum_{k=1}^{K} \sum_{l=1}^{K} E_0\left[1\left\{y_{i,t} = k, y_{i,t-1} = l\right\} | \underline{f_t}\right] \frac{1}{\pi_{lk,t}} J_{lk,t},$$

where $\pi_{lk,t} = \pi_{lk}(f_t, \beta_0)$, $J_{lk,t} = J_{lk}(f_t, \beta_0)$ and all functions are evaluated at the true parameter and factor values. Under Assumption A.1:

$$\begin{aligned} E_0\left[1\left\{y_{i,t} = k, y_{i,t-1} = l\right\} | \underline{f_t}\right] &= E_0\left[E_0\left[1\left\{y_{i,t} = k\right\} | y_{i,t-1} = l, \underline{f_t}\right] 1\left\{y_{i,t-1} = l\right\} | \underline{f_t}\right] \\ &= \pi_{lk,t} P\left[y_{i,t-1} = l | \underline{f_t}\right] = \pi_{lk,t} P\left[y_{i,t-1} = l | \underline{f_{t-1}}\right] = \pi_{lk,t} \mu_{l,t-1}, \end{aligned}$$

where $\mu_{l,t-1} = P\left[y_{i,t-1} = l | \underline{f_{t-1}}\right]$. It follows that:

$$I(t) = \sum_{l=1}^{K} \mu_{l,t-1} I_{l,t},$$

where:

$$I_{l,t} = \sum_{k=1}^{K} J_{lk,t} = \sum_{k=1}^{K} \frac{1}{\pi_{lk,t}} \frac{\partial \pi_{lk,t}}{\partial\left(\beta', f_t'\right)'} \frac{\partial \pi_{lk,t}}{\partial\left(\beta', f_t'\right)}. \qquad \text{(A.13)}$$

Then, the semi-parametric efficiency bound is $(I_0^*)^{-1}$, where $I_0^* = E_0\left[I_{\beta\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} I_{f\beta}(t)\right]$.

In the two-state logit model, we have $\beta = (\alpha_2, \gamma_2)'$ and

$$\Pi_t = \begin{pmatrix} 1 - \Lambda(f_t) & \Lambda(f_t) \\ 1 - \Lambda\left(\beta' x_t\right) & \Lambda\left(\beta' x_t\right) \end{pmatrix}, \tag{A.14}$$

where $x_t = (f_t, 1)'$ and $\Lambda(x) = 1/\left(1 + e^{-x}\right)$ is the logistic function. Since $\pi_{l1,t} = -\pi_{l2,t}$ for $l = 1, 2$, we have:

$$I_{l,t} = \left(\frac{1}{\pi_{l2,t}} + \frac{1}{1 - \pi_{l2,t}}\right) \frac{\partial \pi_{l2,t}}{\partial\left(\beta', f_t\right)'} \frac{\partial \pi_{l2,t}}{\partial\left(\beta', f_t\right)} = \frac{1}{\pi_{l2,t}\left(1 - \pi_{l2,t}\right)} \frac{\partial \pi_{l2,t}}{\partial\left(\beta', f_t\right)'} \frac{\partial \pi_{l2,t}}{\partial\left(\beta', f_t\right)}, \quad l = 1, 2.$$

Since $\frac{d\Lambda(x)}{dx} = \Lambda(x)\left[1 - \Lambda(x)\right]$, we deduce:

$$I_{l,t} = \pi_{l2,t}\left(1 - \pi_{l2,t}\right) \xi_{l,t} \xi_{l,t}', \; l = 1, 2,$$

where $\xi_{1,t} = (0, 0, 1)'$ and $\xi_{2,t} = \left(f_t, 1, \alpha_2\right)'$. Thus, we have:

$$\begin{aligned} I_{\beta\beta}(t) &= \mu_{2,t-1} \pi_{22,t}\left(1 - \pi_{22,t}\right) \begin{pmatrix} f_t^2 & f_t \\ f_t & 1 \end{pmatrix}, \quad I_{\beta f}(t) = \mu_{2,t-1} \pi_{22,t}\left(1 - \pi_{22,t}\right) \alpha_2 \begin{pmatrix} f_t \\ 1 \end{pmatrix}, \\ I_{ff}(t) &= \mu_{1,t-1} \pi_{12,t}\left(1 - \pi_{12,t}\right) + \mu_{2,t-1} \pi_{22,t}\left(1 - \pi_{22,t}\right) \alpha_2^2. \end{aligned}$$

We deduce formula (5.4).

A.7.3 Numerical computation of the semi-parametric efficiency bound

The semi-parametric efficiency bound $\left(I_0^*\right)^{-1}$ can be approximated numerically by Monte-Carlo integration. Let $(f_t : t = -h, -h+1, ..., T)$ be a simulated factor path of length $S = T + h + 1$. We define $\mu_{t-1,S}$ by:

$$\mu_{t-1,S}' = e' \Pi_{-h,S} \Pi_{-h+1,S} \cdots \Pi_{t-1,S}, \quad t = 1, ..., T,$$

where $e = (1/K, ..., 1/K)'$, and

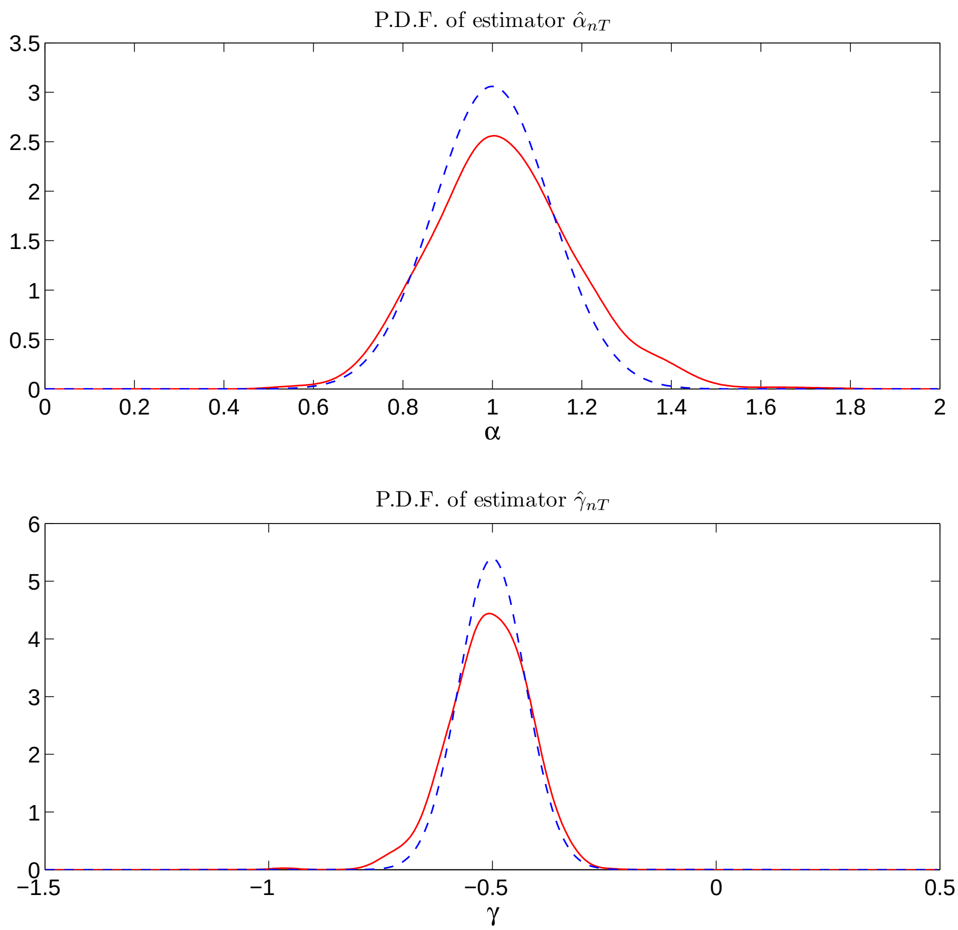
$$I_S(t) = \sum_{l=1}^{K} \mu_{l,t-1,S} I_{l,t,S}, \quad t = 1, ..., T.$$

Matrices $I_{l,t,S}$ and $\Pi_{t,S}$ correspond to the matrices in (A.13) and (A.14), respectively, based on the simulated factor values. Then we approximate matrix I_0^* by

$$I_{0,S}^* = \frac{1}{T} \sum_{t=1}^{T} \left[I_{S,\beta\beta}(t) - I_{S,\beta f}(t) I_{S,ff}(t)^{-1} I_{S,f\beta}(t)\right],$$

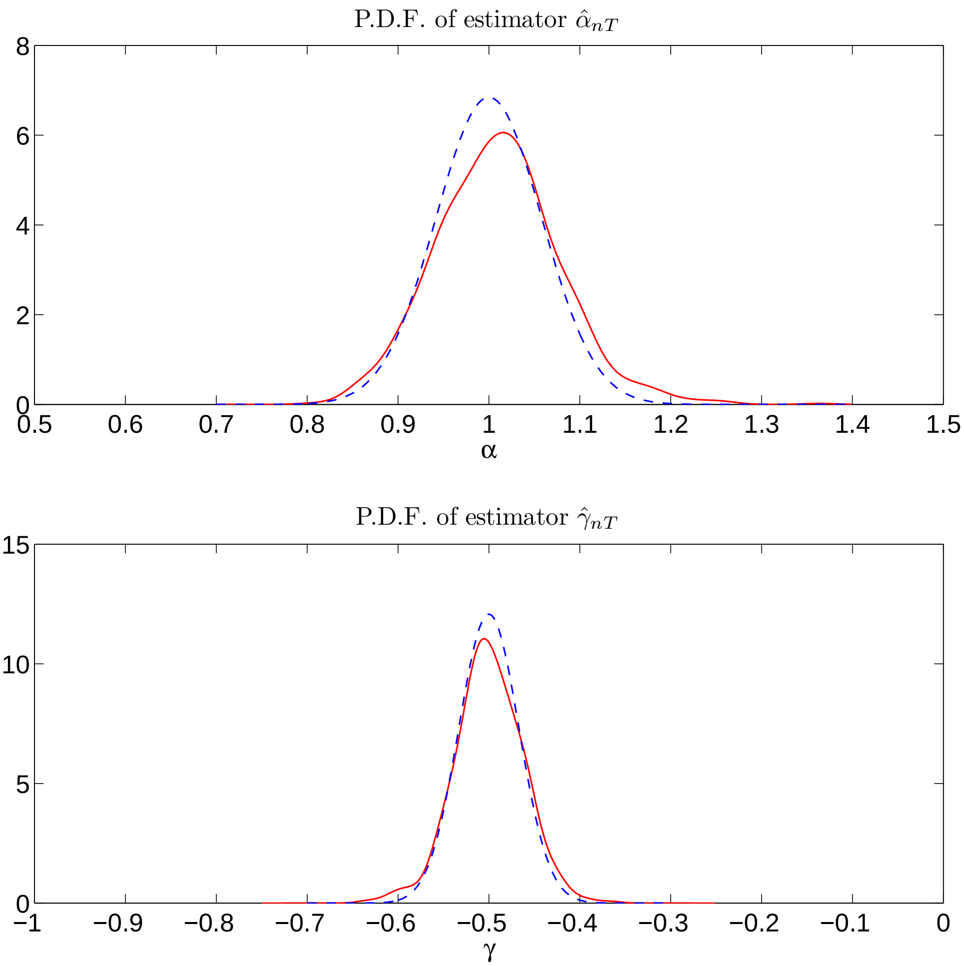
for large T and h.

Figure 1: Distribution of the semiparametrically efficient estimators of the micro-parameters, sample size $n = 200$ and $T = 20$.



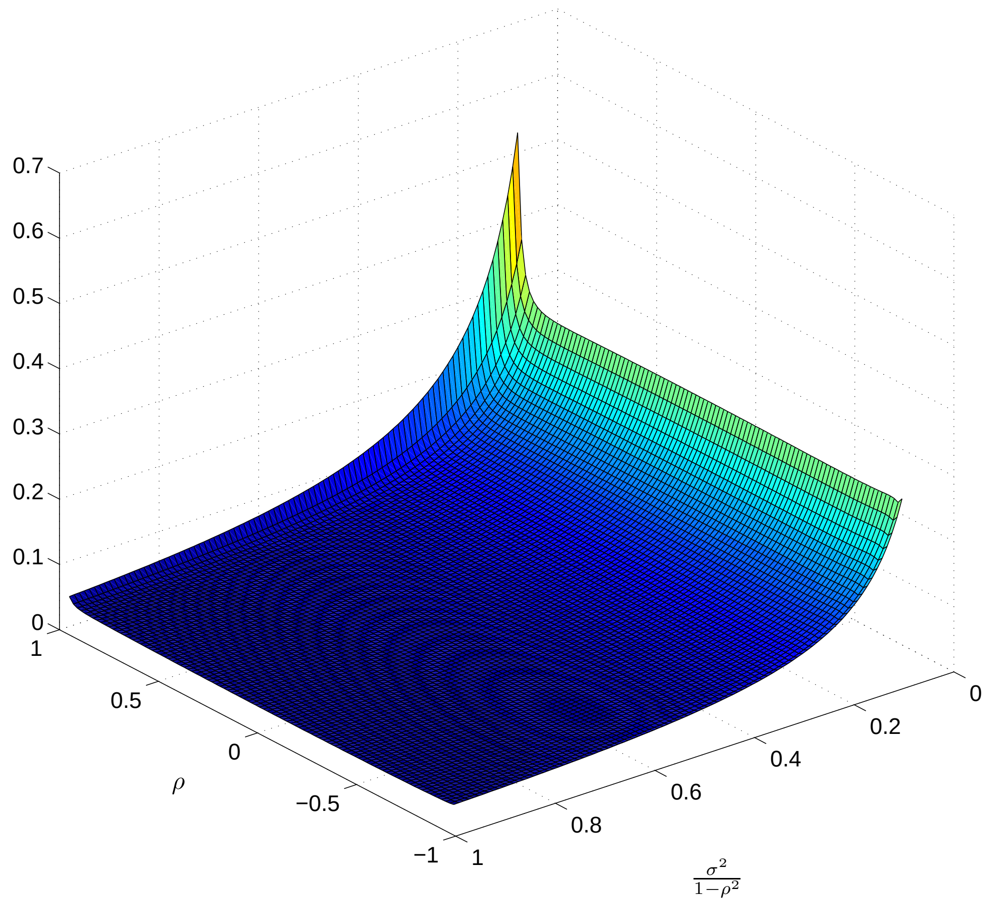
The solid lines give the pdf of the semiparametrically efficient estimators of parameter α (upper Panel, true value 1) and parameter γ (lower Panel, true value -0.5). The pdf is computed by a kernel density estimator. Sample sizes are $n = 200$ and $T = 20$. The dashed lines in the two Panels give the pdf of a normal distribution centered at the true value of the parameter and with variance equal to the semi-parametric efficiency bound divided by nT.

Figure 2: Distribution of the semiparametrically efficient estimators of the micro-parameters, sample size $n = 1000$ and $T = 20$.



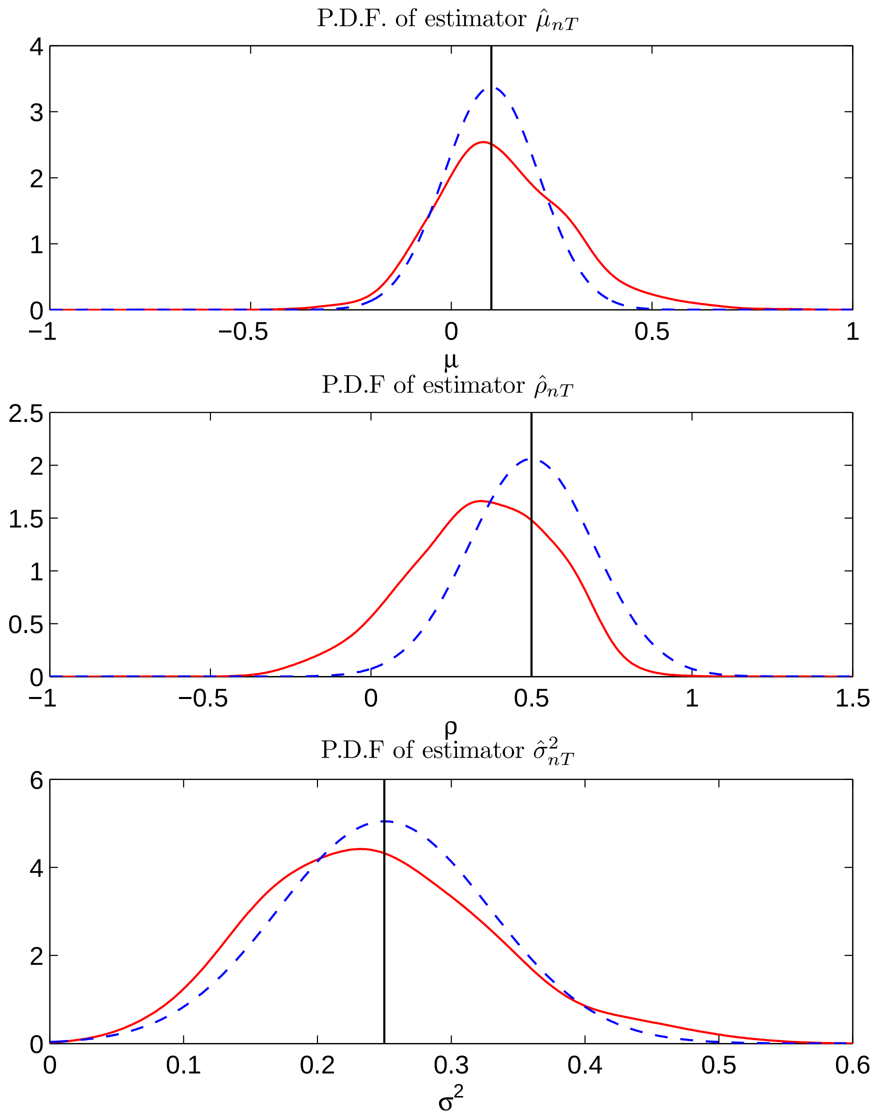
The solid lines give the pdf of the semiparametrically efficient estimators of parameter α (upper Panel, true value 1) and parameter γ (lower Panel, true value -0.5). The pdf is computed by a kernel density estimator. Sample sizes are $n = 1000$ and $T = 20$. The dashed lines in the two Panels give the pdf of a normal distribution centered at the true value of the parameter and with variance equal to the semi-parametric efficiency bound divided by nT.

Figure 3: Semiparametric efficiency bound of the micro-parameter α_2.



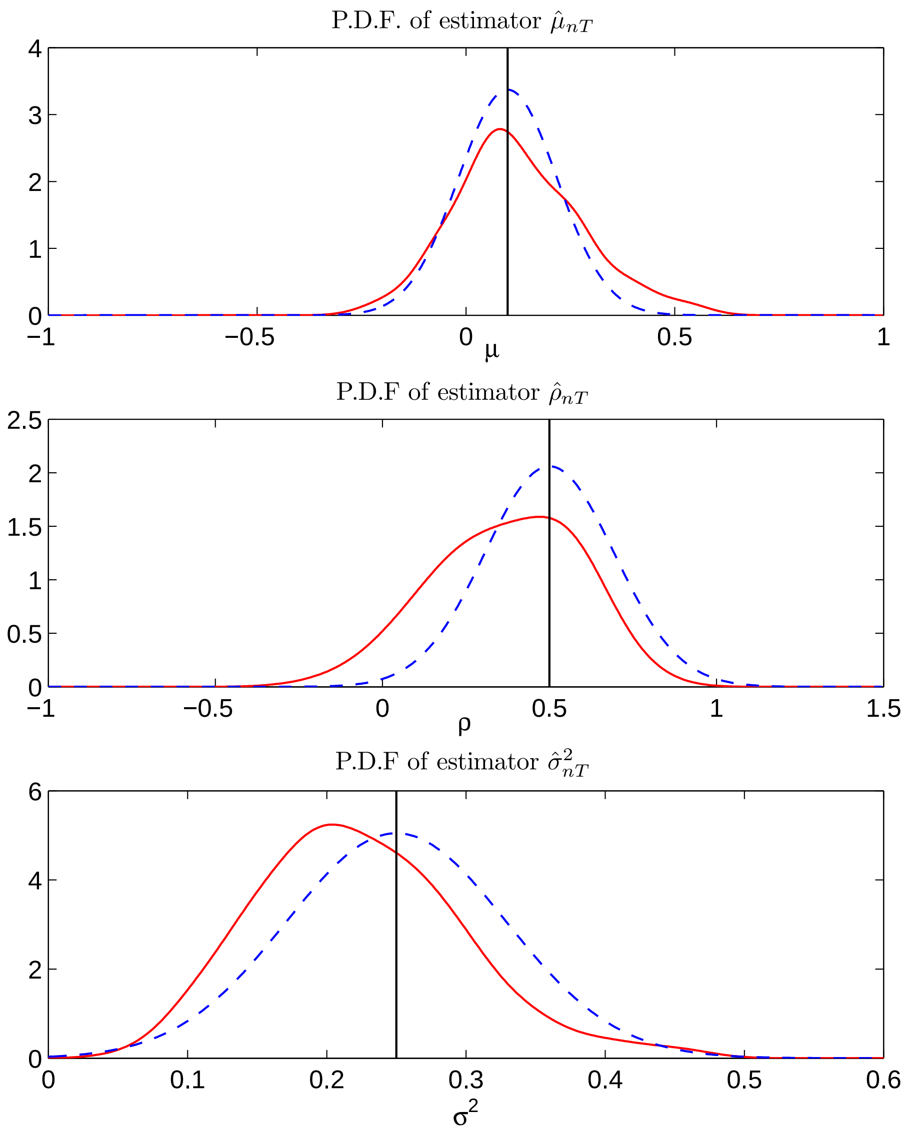
The figure displays $\left(\frac{1}{nT}B^*_{\alpha_2}\right)^{1/2}$, where $B^*_{\alpha_2}$ is the semiparametric efficiency bound for parameter α_2 and $n = 1000, T = 20$, as a function of the autoregressive coefficient ρ and the variance $\frac{\sigma^2}{1-\rho^2}$ of the factor process (f_t).

Figure 4: Distribution of the efficient estimators of the macro-parameters, sample size $n = 200$ and $T = 20$.



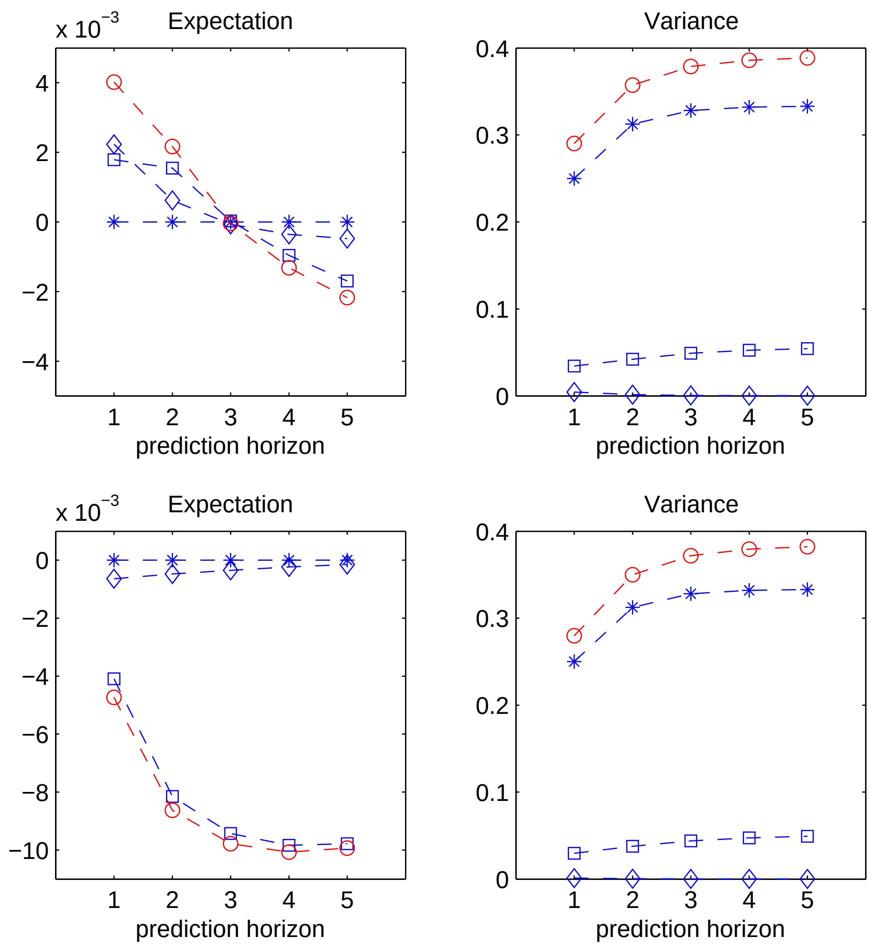
The solid lines give the pdf of the efficient estimators of parameter μ (upper Panel, true value 0.1), parameter ρ (central Panel, true value 0.5) and parameter σ^2 (lower Panel, true value 0.25). The pdf is computed by a kernel density estimator. Sample sizes are $n = 200$ and $T = 20$. The dashed lines in the three Panels give the pdf of a normal distribution centered at the true value of the parameter and with variance equal to the efficiency bound divided by T.

Figure 5: Distribution of the efficient estimators of the macro-parameters, sample size $n = 1000$ and $T = 20$.



The solid lines give the pdf of the efficient estimators of parameter μ (upper Panel, true value 0.1), parameter ρ (central Panel, true value 0.5) and parameter σ^2 (lower Panel, true value 0.25). The pdf is computed by a kernel density estimator. Sample sizes are $n = 1000$ and $T = 20$. The dashed lines in the three Panels give the pdf of a normal distribution centered at the true value of the parameter and with variance equal to the efficiency bound divided by T.

Figure 6: Term-structures of the expectations and variances of the prediction errors.



This Figure displays the term-structures of the unconditional expectations (left Panels) and variances (right Panels) of the prediction errors. The sample sizes are $n = 200$, $T = 20$ in the upper Panels, and $n = 1000$, $T = 20$ in the lower Panels. The stars, the diamonds and the squares correspond to the prediction error $\varepsilon^*_{T,T+L}$ with observable factor and known macro-parameters, the contribution $\varepsilon^{(1)}_{T,T+L}$ from the approximation of the factor value, and the contribution $\varepsilon^{(2)}_{T,T+L}$ from the estimation of the macro-parameters, respectively. The circles correspond to the term-structures of the total prediction error $\varepsilon_{T,T+L}$.