# EFFICIENCY IN LARGE DYNAMIC PANEL MODELS WITH COMMON FACTOR

Patrick GAGLIARDINI[1] and Christian GOURIEROUX[2]

First version: August 2008

This version: April 2010

[1] University of Lugano and Swiss Finance Institute.

[2] CREST, CEPREMAP (Paris), and University of Toronto.

Acknowledgements: We thank A. Monfort, E. Renault, M. Rockinger and O. Scaillet 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 asymptotically efficient estimation in exchangeable nonlinear dynamic panel models with common unobservable factor. These models are especially relevant for applications to large portfolios of credits, corporate bonds, or life insurance contracts, and are recommended in the current regulation in Finance (Basel II) and Insurance (Solvency II). 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 $n$ and $T$, respectively, we derive the efficiency bound and introduce computationally simple efficient estimators for both the micro- and macro-parameters. In particular, we show that the fixed effects estimator of the micro-parameter is asymptotically efficient. The results are based on an asymptotic expansion of the log-likelihood function in powers of $1/n$. This expansion is used to investigate the second-order bias properties of the estimators. The results are illustrated with the stochastic migration model for credit risk analysis.



**Keywords:** Nonlinear Panel Model, Factor Model, Exchangeability, Semi-parametric Efficiency, Fixed Effects Estimator, Bayesian Statistics, Credit Risk, Stochastic Migration, Basel II, Granularity Adjustment, State Space Model.

**JEL classification:** C23, C13, G12.

# 1 Introduction

This paper considers the asymptotically efficient estimation of nonlinear dynamic panel models with common unobservable factor. We focus on exchangeable specifications that are appropriate to analyze the set of histories of a large homogeneous population of individuals featuring serial and cross-sectional dependence. Such a framework is largely encountered in credit risk applications. For instance, for the risk analysis in portfolios of corporate debt, the panel data are the default, loss given default and rating migration histories of a large pool of firms in a given industrial sector and country. The common factor represents a latent macro-variable, such as the sector and country specific business cycle, that introduces dependence across the individual risks, such as default, loss given default, or migration correlations. The purpose of the analysis is to predict the future risk in a large portfolio of corporate bonds or credit derivatives issued by the firms in the pool. [1]

The model considered in this paper involves both a micro- and a macro-dynamic. Conditional on a given factor path, the individuals are assumed independent and identically distributed, with the histories of 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 $\beta$. The micro-dynamic is captured by the lagged individual observation $y_{i,t-1}$ and unknown parameter $\beta$. 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 $\theta$. When this common factor is integrated out, it introduces both non-Markovian serial dependence within the individual histories, and cross-sectional dependence between individuals. The variables $y_{i,t}$ are either quantitative, or qualitative (as for default and rating histories in the credit risk application), while the components of vector $f_t$ are real valued (corresponding to a continuum of latent states). The model is potentially nonlinear in both the micro- and the macro-dynamic.

When the cross-sectional dimension $n$ is fixed and the time dimension $T$ tends to infinity,

[1] The panel data may also correspond to other risk characteristics of a pool of corporate loans, household mortgages or life insurance contracts, such as prepayment, lapse, mortality.

the Maximum Likelihood (ML) estimators of micro-parameter $\beta$ and macro-parameter $\theta$ 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 [2]. Moreover, the numerical computation of the ML estimate [3] is complicated since the likelihood function involves a large dimensional integral w.r.t. the unobservable factor path.

The aim of this paper is to derive the asymptotic efficiency bound for estimating both the micro-parameter $\beta$ and the macro-parameter $\theta$, and to introduce asymptotically efficient estimators of $\beta$ and $\theta$ that are easier to compute than the ML estimator. We consider the double asymptotics $n, T \to \infty$, such that $T^b/n = O(1)$, for $b > 1$. We summarize our contributions as follows. First, we show that the efficiency bound for micro-parameter $\beta$ 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 $\beta$ 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 [4]. 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 [Bickel, Yahav (1969), Ibragimov, Has'minskii (1981)]. Third, an efficient estimator of the macro-parameter $\theta$ 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

[2] In applications to mortgage or life insurance data, we typically have $n \simeq 100,000 - 1,000,000$ contracts and $T \simeq 200$ months.

[3] For instance, by means of an Expectation-Maximization (EM) algorithm [Dempster, Laird, Rubin (1977)], where the Expectation step is performed via a Gibbs sampler.

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

in Basel II [BCBS (2001), (2003)]. 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 in powers of $1/n$. For this purpose, the integration of the latent factor is performed along the lines of the Laplace approximation [see Tierney, Kadane (1986) for the use of Laplace approximation to compute posterior moments in Bayesian statistics]. If the micro-parameter is semi-parametrically identified, we show that the efficiency bound for micro-parameter $\beta$ 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 asymptotically 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 the macro-likelihood after substitution of the factor values $f_t$ by their approximations $\hat{f}_t$ is asymptotically efficient. Finally, we discuss the link between our likelihood expansion and the granularity adjustment introduced in Pillar 2 of the Basel II regulation. The higher-order terms in the likelihood expansion can be used to get more accurate approximations of the ML estimators. We compare the bias at order $1/n$ of different asymptotically efficient estimators, that are the Cross-Sectional Asymptotic (CSA) and Granularity Adjusted (GA) maximum likelihood estimators, respectively. Section 5 describes a class of models with macro-parameters only, where the impact of a vector of Gaussian autoregressive macro-factors on the micro-dynamics is summarized through some noisy linear transformations of the macro-factors. Asymptotically efficient estimators can be easily computed by applying a linear Kalman filter to appropriate approximate state space models. The class of models in Section 5 include specifications with heterogeneity between different substrata of the populations. In Section 6, 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 7 concludes. The proofs of the results are gathered in Appendices A.1-A.9. 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.people.usi.ch/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 the factor path* $(f_t)$*, the individual histories* $(y_{it}, t = 1, \cdots, T)$*,* $i = 1, ..., n$*, are i.i.d. time-inhomogeneous Markov processes of order 1, with transition pdf* $h(y_{i,t}|y_{i,t-1}, f_t; \beta)$ *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 $\beta_0$ and $\theta_0$ the true values of parameters $\beta$ and $\theta$, 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. invariant by permutation of the individuals. The exchangeability property is equivalent to the existence of a factor representation [de Finetti (1931), Hewitt, Savage (1955)] [5]. Andrews (2005) considers linear regression models for cross-sectional data with very general exchangeable dependence structure. Exchangeable linear models for panel data are considered in Hjellwig, Tjostheim

[5]More precisely, by the de Finetti-Hewitt-Savage theorem, the infinite sequence of histories $y_i = (y_{i,t}, t = 1, \cdots, T)$, $i = 1, 2, \cdots$, is exchangeable if and only if there exists a sigma-field $\mathcal{F}$ such that $y_i$, $i = 1, 2, \cdots$, are i.i.d. conditional on $\mathcal{F}$ [see also Kingman (1978)]. In our model, the sigma-field $\mathcal{F}$ is generated by the Markov process $(f_t)$.

(1999). The focus of our paper is on the efficient estimation of both micro-parameter $\beta$ and macro-parameter $\theta$ in the **nonlinear** exchangeable panel model A.1-A.2.

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 geometrically strong mixing.*

**A.5:** *Conditional on the factor path* $(f_t)$*, the individual process* $(y_{i,t})$ *is ergodic and beta mixing, such that the conditional beta mixing coefficients:*

$$\beta_t(h) = \int \left\{ \sup_{A \in \mathcal{B}(\mathbb{R})} \left| P\left[y_{i,t} \in A | y_{i,t-h} = \eta, \underline{f_t}\right] - P\left[y_{i,t} \in A | \underline{f_t}\right] \right| \right\} \lambda(\eta) d\eta, \quad h \in \mathbb{N},$$

*satisfy* $E\left[\beta_t(h)\right] = O(h^{-\alpha})$ *as* $h \to \infty$*, for some* $\alpha > 0$*, where* $\mathcal{B}(\mathbb{R})$ *denotes the Borel sigma field on* $\mathbb{R}$*,* $\lambda$ *is a strictly positive p.d.f. on* $\mathbb{R}$*, and* $\underline{f_t} = (f_t, f_{t-1}, \cdots)$*. Moreover:*

$$\sup_{A \in \mathcal{B}(\mathbb{R})} E\left[\left|P\left[y_{i,t} \in A | \underline{f_t}\right] - P\left[y_{i,t} \in A | f_t, \cdots, f_{t-m}\right]\right|\right] = O(m^{-\alpha}),$$

*as* $m \to \infty$.

Assumption A.5 requires that the individual processes $(y_{i,t})$ are ergodic and beta mixing, conditional on the factor path. The conditional mixing coefficients $\beta_t(h)$ can depend on the factor path, and converge to zero as lag $h$ increases, for any factor path. The convergence rate can be geometric, for instance. The integration w.r.t. the factor path is expected to decrease the decay rate of the mixing coefficients [Granger, Joyeux (1980)]. However, under Assumption A.5 the integrated mixing coefficients converge to zero at least as a negative power of the lag. The 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. Moreover, under Assumption A.5 the stationary distribution of $y_{i,t}$ conditional on the factor path can be well approximated by a finite number of lags of the factor. Assumptions A.3-A.5 are used to give a Weak Law of Large Numbers (WLLN) for 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 $\beta$, and $\varphi$ 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 [6].

## 2.2 The Single Risk Factor (SRF) model for default

The specification considered in Section 2.1 is motivated by the SRF model introduced by Vasicek (1987), (1991) and based on the Value of the Firm model [Merton (1974)]. The SRF model is recommended for the analysis of credit risk in Pillar 1 of Basel II regulation, concerning the minimum required capital, and in Pillar 2, concerning internal risk models [BCBS (2001), (2003)]. 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 (or jointly, see Section 5). 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 household borrowers. 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:

$$y^*_{i,t} = \alpha + \gamma F_t + \sigma u_{i,t}, \quad i \in PaR_t, \;\; t = 1, ..., T,$$

where $PaR_t$ denotes the Population-at-Risk, that is the set of firms in the portfolio which are still alive at time $t$, and where the common factor $(F_t)$ and the errors $(u_{i,t})$ are independent

[6] The stationarity and ergodicity Assumptions A.3-A.5 for asymptotic analysis with $n, T \to \infty$, $T/n \to 0$, 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 in Section 4 depend on cross-sectional aggregates of the type $\frac{1}{n}\sum_{i=1}^{n} a(y_{i,t}, f_t, \beta)$, which become 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).

standard Gaussian white noise processes. This specification distinguishes the idiosyncratic risks $u_{i,t}$, which can be diversified, and the undiversifiable systematic risk $F_t$. The sensitivity coefficients $\alpha, \gamma, \sigma$ are independent of the individuals, according to the definition of an homogeneous portfolio. The observed endogenous variable is the indicator for the default event, that occurs when the asset is below the liability:

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

We deduce the Probability of Default (PD) at date $t$ conditional on the common factor:

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

where $\Phi$ denotes the cumulative distribution function (cdf) of the standard normal distribution. Thus, the conditional probability of default is time-varying and driven by the common stochastic factor $F_t$. This basic model can be extended by allowing for a dynamics of the common factor, and for a joint analysis of more than two rating levels by means of stochastic migration models describing the transitions between rating classes AAA, AA, ..., C, D, say (see Section 6).

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

$$\rho = Corr\left(y_{i,t}, y_{j,t}\right) = \frac{\Psi\left(-\alpha/\sqrt{\gamma^2+\sigma^2}, -\alpha/\sqrt{\gamma^2+\sigma^2}; \rho^*\right) - PD^2}{PD(1-PD)}, \tag{2.1}$$

where $\rho^* = \gamma^2/\left(\gamma^2+\sigma^2\right)$ is the asset correlation, that is the correlation between the log asset-to-liability ratios of any two firms, and $\Psi(.,.;\rho^*)$ denotes the joint cdf of the bivariate standard Gaussian distribution with correlation coefficient $\rho^*$. In the new regulation for credit risk, the required capital depends on the values of $PD$ and $\rho^*$, that is, indirectly on the values $\alpha/\sigma$ and $\gamma/\sigma$, and is especially sensitive to the asset correlation parameter $\rho^*$. This explains the importance of a computationally simple and statistically efficient estimation of the structural parameters.

## 2.3 The panel model with fixed effects

The econometric literature on nonlinear panel models with fixed effects considers specifications such that the variables $y_{i,t}$, for $i = 1, 2, ..., n$, and $t = 1, ..., T$, are independent with pdf $f(y_{i,t};\alpha_i,\beta)$, where $\alpha_i$ is the fixed effect of individual $i$ [Hahn, Newey (2004); see e.g. Hahn, Kuersteiner (2004) and Arellano, Bonhomme (2009) for extensions to a dynamic setting]. The focus of this literature is on the correction of the bias of the ML estimator of $\beta$ caused by the incidental parameters problem [Neyman, Scott (1948); see Lancaster (2000) for a review]. Without Assumption A.2 on the parametric factor dynamic, the model introduced in Section 2.1 could 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 applications to credit risk $n$ 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 particular, the bias corrections are less important in our setting and even not required if $T/n \to 0$.

ii) Assumption A.2 shows that the nonlinear panel model with common factor in Section 2.1 is clearly a time series model introduced for prediction purpose. At the opposite, a model with fixed individual effects is used to get a segmentation of the population into homogeneous classes, i.e. with similar $\alpha_i$ values. For instance, in the credit risk application, 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 derive the distribution of the future portfolio value and the corresponding 1% quantile, called CreditVaR.

iii) As a consequence, we are also interested in the filtering of the factor values, in their dynamics, that is, in macro-parameter $\theta$, 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 $\underline{f_T} = (f_t, t = 1,...,T)$ (conditionally on the initial values) is given by:

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

If the factors were observable, the terms $l\left(\underline{y_T}|\underline{f_T};\beta\right)$ and $l\left(\underline{f_T};\theta\right)$ would correspond to the conditional micro-density of the endogeneous variables, and the macro-density of the factors, respectively. Since the factors are unobservable, the density of $\underline{y_T}$ is obtained by integrating out factors $\underline{f_T}$:

$$
\begin{aligned}
l\left(\underline{y_T};\beta,\theta\right) &= \int\cdots\int\prod_{t=1}^{T}\prod_{i=1}^{n}h\left(y_{i,t}|y_{i,t-1},f_t;\beta\right)\prod_{t=1}^{T}g(f_t|f_{t-1};\theta)\prod_{t=1}^{T}df_t \qquad (3.2)\\
&= \int\cdots\int\exp\left\{\sum_{t=1}^{T}\sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1},f_t;\beta\right)\right\}\prod_{t=1}^{T}g(f_t|f_{t-1};\theta)\prod_{t=1}^{T}df_t.
\end{aligned}
$$

This likelihood function involves an integral with a large dimension increasing with $T$, which complicates the analytical study of the Maximum Likelihood (ML) estimators and the numerical derivations of the ML estimates [7]. However, for large $n$, this integral can be approximated by expanding the integrand around its maximum w.r.t. the factor values, along the lines of the Laplace approximation [Laplace (1774); see Jensen (1995) for the general setting and Tierney, Kadane (1986) for application to Bayesian statistics]. The Laplace approximation has been used in Arellano, Bonhomme (2009) to derive the bias of the integrated likelihood in nonlinear panel models with fixed individual effects. In our setting with serially dependent factor, the Laplace approximation is applied to the integral w.r.t. the full path of time effects. Specifically, let us define for any parameter value $\beta$ and date $t$ the cross-sectional ML

[7] In such a model with unobservable factors, the ML estimate could be computed numerically by means of an Expectation-Maximization (EM) algorithm [Dempster, Laird, Rubin (1977)]. The EM algorithm applies recursively the Expectation step, which computes the function:

$$Q\left[(\beta,\theta)|(\beta^{(p)},\theta^{(p)})\right] = \underset{(\beta^{(p)},\theta^{(p)})}{E}\left[\log l\left(\underline{y_T},\underline{f_T};\beta,\theta\right)|\underline{y_T}\right],$$

and the maximization step, providing the next value of the parameter as:

$$(\beta^{(p+1)},\theta^{(p+1)}) = \underset{(\beta,\theta)}{\arg\max}\, Q\left[(\beta,\theta)|(\beta^{(p)},\theta^{(p)})\right].$$

In our nonlinear dynamic framework, the Expectation step requires the numerical approximation of function $Q$ by means of a Gibbs sampler [see e.g. Cappé, Moulines, Rydén (2005) for general properties, and Fiorentini, Sentana, Shephard (2004), Duffie et al. (2009) for applications to credit and finance]. The closed form expression of the likelihood function given in Proposition 1 allows to avoid the numerically cumbersome expectation step, while controlling the approximation error.

estimator of the factor value [8]:

$$\hat{f}_{n,t}(\beta) = \arg\max_{f_t} \sum_{i=1}^{n} \log h\left(y_{i,t}|y_{i,t-1}, f_t; \beta\right). \tag{3.3}$$

**Proposition 1.** *Under the regularity Assumptions H.1-H.14 in Appendix A.3, the joint density of* $\left(\underline{y_T}\right)$ *is such that:*

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

*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:*

$$\mathcal{L}_{nT}\left(\beta,\theta\right) = \mathcal{L}^*_{nT}\left(\beta\right) + \frac{1}{n}\mathcal{L}_{1,nT}\left(\beta,\theta\right) + \frac{1}{n^2}\Psi_{nT}\left(\beta,\theta\right), \tag{3.4}$$

*where:*

$$\mathcal{L}^*_{nT}\left(\beta\right) = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} \log h\left(y_{i,t}|y_{i,t-1}, \hat{f}_{nt}\left(\beta\right); \beta\right), \tag{3.5}$$

$$\mathcal{L}_{1,nT}\left(\beta,\theta\right) = -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\log\det I_{nt}\left(\beta\right) + \frac{1}{T}\sum_{t=1}^{T}\log g\left(\hat{f}_{nt}\left(\beta\right)|\hat{f}_{n,t-1}\left(\beta\right); \theta\right), \tag{3.6}$$

*and* $\sup_{\beta\in\mathcal{B},\theta\in\Theta} \Psi_{nT}\left(\beta,\theta\right) = O_p(1)$.

[8] 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 $\mathcal{F}_n$, that converges to the set $\mathbb{R}^K$ when $n \to \infty$.

Function $\mathcal{L}^*_{nT}(\beta)$, called profile log-likelihood function, is the log-likelihood of $\beta$ concentrated w.r.t. the factor values, as if the latter were 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 $\beta$ when $n \to \infty$ (see Section 3.3 for a precise statement).

## 3.2 Bayesian interpretation

The results in Proposition 1 and Corollary 2 are an example of the asymptotic equivalence of frequentist and Bayesian methods in large sample. To get the intuition, let time dimension $T$ be fixed and parameter $\theta$ be given for a moment. Then, our specification with stochastic common factor can be seen as a Bayesian approach w.r.t. parameter $\beta$ and time effects $\underline{f_T}$. The prior distribution is such that the density of $\underline{f_T}$ given $\beta$ is $\prod_{t=1}^{T} g(f_t|f_{t-1};\theta)$ [9], independent of $\beta$, and the prior distribution of $\beta$ is diffuse. Then, the posterior density of $(\beta, \underline{f_T})$ corresponds to the RHS of equation (3.1), while the posterior density of $\beta$ corresponds to the RHS of equation (3.2), up to multiplicative constants. As the cross-sectional dimension $n$ tends to infinity, it is expected from results in 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 $\beta$ [see e.g. Bickel, Yahav (1969), Ibragimov, Has'minskii (1981)]. This is why the integral in the RHS of (3.2) approaches asymptotically the density of $\left(\underline{y_T}\right)$ given $\left(\underline{f_T}\right)$ with $f_t$ replaced by $\hat{f}_{nt}(\beta)$, $t = 1, ..., T$, for large $n$, up to higher order terms. Thus, the "Bayesian" log posterior density $\mathcal{L}_{nT}(\beta, \theta)$ approaches the log-likelihood $\mathcal{L}^*_{nT}(\beta)$, which is the "frequentist" log-likelihood for $\beta$ concentrated w.r.t. parameters $f_t$, $t = 1, ..., T$. The asymptotic irrelevance of the second term in the RHS of (3.4) involving the transition density of the factor corresponds to the irrelevance of the prior distribution in large sample. Our results show that this asymptotic equivalence is still valid when the number of time effects parameters tends to infinity [10]: $T \to \infty$, such that $T^b/n \to 0$, $b > 1$. Function $\mathcal{L}_{1,nT}(\beta, \theta)$

[9] This prior depends on "hyperparameter" $\theta$.

[10] See Belloni, Chernozhukov (2009) for another extension of the asymptotic normality of the (quasi-) posterior distribution when the number of parameters increases with the sample size. This extension is derived under

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 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 [11].

## 3.3 Efficiency bound

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

$$\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right) = \arg\max_{\beta,\theta} \mathcal{L}_{nT}(\beta, \theta). \tag{3.7}$$

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

$$\left[\begin{array}{c} \sqrt{nT}\left(\tilde{\beta}_{nT} - \beta_0\right) \\ \sqrt{T}\left(\tilde{\theta}_{nT} - \theta_0\right) \end{array}\right] \xrightarrow{d} N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} B^*_{\beta\beta} & B^*_{\beta\theta} \\ B^*_{\theta\beta} & B^*_{\theta\theta} \end{array}\right)\right), \tag{3.8}$$

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

To compute the efficiency bound, let us introduce the large sample counterparts of the likelihood terms in the RHS of (3.4).

(i) Let us first consider $\mathcal{L}^*_{nT}(\beta)$. We can define at each date $t$ the cross-sectional 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 $\underline{f_t}$. This function yields the factor value $f_t(\beta)$ that maximizes the limiting

a different set of regularity conditions.

[11] 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.

cross-sectional log-likelihood at date $t$, for any given parameter value $\beta$. It corresponds to the population counterpart of $\hat{f}_{n,t}(\beta)$ in (3.3) when $n \to \infty$. The pseudo-true factor value $f_t(\beta)$ is a function of both parameter $\beta$ and information $\underline{f_t}$. The pseudo-true factor value is stochastic due to its conditional interpretation. Moreover, by the properties of the Kullback-Leibler discrepancy, at true parameter value $\beta_0$ 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 $\beta$.

**A.6 (Global semi-parametric identification assumption for** $\beta$**):** *The mapping* $\beta \to \mathcal{L}^*(\beta)$ *is uniquely maximized at the true parameter value* $\beta_0$.

**A.7 (Local semi-parametric identification assumption for** $\beta$**):** *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):

$$I_0^* = E_0\left[I_{\beta\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}I_{f\beta}(t)\right], \tag{3.9}$$

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 $t$:

$$I(t) = E_0\left[-\frac{\partial^2\log h\left(y_{it}|y_{i,t-1}, f_t;\beta_0\right)}{\partial\left(\beta', f'\right)'\partial\left(\beta', f'\right)}|\underline{f_t}\right]. \tag{3.10}$$

Assumptions A.6 and A.7 correspond to identification conditions for parameter $\beta$ 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 (3.5), that is, the likelihood of $\beta$ concentrated w.r.t. "parameters" $f_t$, $t = 1, ..., T$. [12]

(ii) Let us now consider the macro component $\mathcal{L}_{1,nT}(\beta,\theta)$ of the log-likelihood. Under Assumptions A.6-A.7 parameter $\beta$ can be estimated at a rate infinitely faster than the rate for

[12] When Assumptions A.6 and A.7 are not satisfied, the identification of parameter $\beta$ relies on the parametric

parameter $\theta$ and the relevant criterion for identification of $\theta$ 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 (3.6) for $\beta = \beta_0$. We have $\mathcal{L}_1(\beta_0, \theta) = E_0 [\log g(f_t | f_{t-1}; \theta)]$, up to a term constant in $\theta$. Thus, the identification assumptions for the macro-parameter are the following:

**A.8 (Global identification assumption for** $\theta$**):** *The mapping* $\theta \to E_0 [\log g(f_t | f_{t-1}; \theta)]$ *is uniquely maximized at the true parameter value* $\theta_0$*.*

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

Assumptions A.8 and A.9 are the standard global and local identification conditions for estimating $\theta$ in a model with observable factor values.

**Proposition 3.** *Under Assumptions A.1-A.9 and H.1-H.14, and if* $n, T \to \infty$ *such that* $T^b/n = O(1)$*,* $b > 1$*, the efficiency bound for* $(\beta, \theta)$ *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*

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

*Proof.* See Appendix 6. □

The zero out-of-diagonal blocks in the efficiency bound imply that parameters $\beta$ and $\theta$ can be considered independently for estimation purpose. This justifies ex-post their interpretation as micro- and macro-parameters, respectively, since parameter $\beta$ (resp. $\theta$) contains no

model $g(f_t | f_{t-1}; \theta)$ for the transition of the factor. Intuitively, we would have to distinguish the transformations of vector $\beta$ that are identified by criterion $\mathcal{L}^*(\beta)$, and the transformations of $\beta$ 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 analysis of this general setting is beyond the scope of this paper.

macro-information (resp. no micro-information) under Assumptions A.6-A.9. 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 the regularity conditions in Appendix 3, for large $n$ and $T$ the relevant term for estimation of parameter $\beta$ is $\mathcal{L}^*_{nT}(\beta)$. The corresponding limit log-likelihood function is $\mathcal{L}^*(\beta)$ , and the efficiency bound $B^*_{\beta\beta}$ for $\beta$ 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 $\theta$ 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 $\beta$ and $\theta$ are asymptotically independent. Therefore, the efficiency bound $B^*_{\beta\beta}$ for $\beta$ given in Proposition 3 is the same as the efficiency bound for $\beta$ with known transition density of the factor. Finally, matrix $I^*_0$ in (3.9) 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 $\theta$ with observable factor. Therefore, the unobservability of the factor has no efficiency impact asymptotically for estimating $\theta$, but has an impact for estimating $\beta$. 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 $\theta$, if $T^b/n = O(1)$, $b > 1$, and infinitely slower than the rate $1/\sqrt{nT}$ for estimating $\beta$.

The efficiency bound $B^*_{\beta\beta}$ for parameter $\beta$ in Proposition 3 is independent of the parametric model $g(f_t|f_{t-1};\theta)$, $\theta \in \mathbb{R}^p$, for the transition density 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 $\beta$ is the efficiency bound for estimating $\beta$ 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 [Stein (1956), Severini, Tripathi (2001)]. 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.11-H.14 in Appendix 3, and let $B^*_{\beta\beta}(g_\theta)$ be the corresponding parametric efficiency bound for estimating $\beta$.

**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 $g_\theta$ for the transition of $f_t$ that satisfy Assumptions A.8-A.9 and H.11-H.14.*

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

**Corollary 4.** *Under Assumptions A.1-A.7 and H.1-H.10, and if $n, T \to \infty$ such that $T^b/n = O(1)$, $b > 1$, the semi-parametric efficiency bound for $\beta$ 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 $g_\theta$ 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 asymptotically efficient estimation of micro-parameter $\beta$.

## 3.4 Identification in the SRF model for default

The SRF model of Section 2.2 is such that $y_{i,t}$ admits the Bernoulli distribution $\mathcal{B}\left(1, \Phi\left[-(\alpha/\sigma) - (\gamma/\sigma) F_t\right]\right)$ conditional on the factor $F_t$, and the observations can be summarized by the sufficient statistics $\widehat{PD}_t = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$, that are the cross-sectional default frequencies. In a semi-parametric framework, in which the transition of the factor is left unspecified, parameters $\alpha/\sigma$ and $\gamma/\sigma$ are not identified. Indeed, the initial factor can be replaced by $f_t = \Phi\left[-(\alpha/\sigma) - (\gamma/\sigma) F_t\right] = PD_t$, and the model becomes:

$$y_{i,t} \sim \mathcal{B}\left(1, f_t\right), \tag{3.11}$$

conditionally on $f_t$. The factor values are approximated by $\hat{f}_{n,t} = \widehat{PD}_t$. Parameters $\alpha/\sigma$ and $\gamma/\sigma$ can be identified when a parametric specification for the factor dynamics is introduced. For instance, the SRF model considered by Basel II is identifiable due to the assumption that the factor values $F_t$ are independent standard normal. Then, the transformed factors $f_t$ are such that:

$$\Phi^{-1}(f_t) \sim IIN\left(-\alpha/\sigma, \gamma^2/\sigma^2\right). \tag{3.12}$$

The model defined by equations (3.11) and (3.12) satisfies Assumptions A.1-A.9, with no micro-parameter and macro-parameter $\theta = (\alpha/\sigma, \gamma/\sigma)'$. From equation (2.1), the default correlation is a function of the macro-parameters only. We see in Section 6 that micro-parameters, and their semi-parametric identification, are 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 estimator. These estimators rely on the asymptotic expansion of the log-likelihood function and do not involve the numerical integration w.r.t. the unobservable factor. We also compare the bias at order $1/n$ of the efficient estimators, and deduce the higher-order expansion of the true ML estimator.

## 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 $\beta$ 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.10, 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 $\beta$ 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), estimator $\hat{\beta}^*_{nT}$ is consistent, but its asymptotic distribution is not centered at the true parameter value $\beta_0$ [13]. The fixed effects estimator $\hat{\beta}^*_{nT}$ becomes efficient if $n, T \to \infty$ such that $T^b/n = O(1), b > 1$.

The numerical computation of an asymptotically efficient estimator of the micro-parameter $\beta$ can be simplified when the micro-dynamics is such that:

$$h\left(y_{i,t}|y_{i,t-1}, f_t; \beta\right) = \tilde{h}\left(y_{i,t}|y_{i,t-1}, a_t\right), \quad \text{say}, \tag{4.1}$$

where the impact of factor $f_t$ and parameter $\beta$ is summarized through the canonical factor vector $a_t = c\left(f_t, \beta\right)$, with $\dim(a_t) \geq \dim(f_t) + 1$. Let us assume that the canonical factor is cross-sectionally identifiable, that is, the cross-sectional log-likelihood function $E_0\left[\log \tilde{h}(y_{i,t}|y_{i,t-1}, a)|\underline{a_t}\right]$ is uniquely maximized at the true value $a = a_t$ of the canonical factor at date $t$, $P$-a.s., for any $t$. Another asymptotically efficient estimator of $\beta$ can be derived by first computing the cross-sectional fixed effects estimators of the canonical factor values:

$$\hat{a}_{n,t} = \arg\max_{a_t} \sum_{i=1}^{n} \log \tilde{h}\left(y_{i,t}|y_{i,t-1}, a_t\right),$$

and their estimated asymptotic variances:

$$\hat{\Sigma}_{n,t} = \left(-\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log \tilde{h}\left(y_{i,t}|y_{i,t-1}, \hat{a}_{n,t}\right)}{\partial a_t \partial a_t'}\right)^{-1}.$$

Then, the estimator $\hat{\beta}^{**}_{nT}$ obtained by solving the optimization:

$$\min_{\beta, f_1, \ldots, f_T} \sum_{t=1}^{T} \left[\hat{a}_{n,t} - c\left(f_t, \beta\right)\right]' \hat{\Sigma}^{-1}_{n,t} \left[\hat{a}_{n,t} - c\left(f_t, \beta\right)\right], \tag{4.2}$$

is asymptotically equivalent to $\hat{\beta}^*_{nT}$ by applying general results on Asymptotic Least Squares. This shows that for the computation of the asymptotically efficient estimators the data can be aggregated through the fixed effects estimators of the canonical factors. The computation of the estimator $\hat{\beta}^{**}_{nT}$ is greatly simplified when the canonical factors $a_t$ are linear w.r.t. the factors $f_t$, that is:

$$a_t = \alpha(\beta) + \gamma(\beta) f_t, \tag{4.3}$$

[13] It is beyond the scope of this paper to investigate whether a bias reduction approach in the spirit of Woutersen (2002), Hahn, Kuersteiner (2004), Arellano, Hahn (2006), Bester, Hansen (2009) can be applied in our framework.

where $\alpha(.)$ and $\gamma(.)$ are known vector and matrix functions of the micro-parameter, respectively. Indeed, in this case the factor values are easily concentrated in the optimization problem (4.2), and the concentrated criterion for $\beta$ can then be optimized to get $\hat{\beta}^{**}_{nT}$. The stochastic transition model for rating migration considered in Section 6 is a specification satisfying equations (4.1) and (4.3).

## 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 [14]. A consistent approximation of the factor value at date $t$ is:

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

**Proposition 6.** *Suppose Assumptions A.1-A.7 and H.1-H.10 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)* $\sup_{1\leq t\leq T}\left\|\hat{f}_{nT,t} - f_t\right\| = O_p\left(\sqrt{\frac{\log(n)^a}{n}}\right)$, *where* $a = 2a_1 + a_2 + a_3$, *and* $a_1, a_2, a_3 > 0$ *are defined in Assumptions H.8-H.10 in Appendix A.3.*

*Proof.* See Appendix 6. □

For any given date $t$, the factor approximation $\hat{f}_{nT,t}$ depends on the whole individual histories and can be considered as a smoothed factor value. It 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 $\beta_0$. 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

[14]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 $\beta_0$. 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 applied 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 $\hat{f}_{n,t}(\beta)$ w.r.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. $\theta$, we get an efficient estimator of the macro-parameter.

**Proposition 7.** *Under Assumptions A.1-A.9 and H.1-H.14, 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.11-H.14, 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]. □

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 $\theta$ 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 $\beta$ and $\theta$, respectively, then the joint estimator $\left(\hat{\beta}^*_{nT}, \hat{\theta}_{nT}\right)$ is also asymptotically efficient [see Gouriéroux, Monfort (1995)].

## 4.4 Cross-sectional Asymptotic and Granularity Adjusted Maximum Likelihood estimators

The asymptotic analysis can be refined by considering expansions of the log-likelihood function at probability order $1/n$, and $1/n^2$. From Corollary 2, an approximation at order $1/n$ is given by:

$$\mathcal{L}_{nT}^{\text{CSA}}(\beta,\theta) = \mathcal{L}_{nT}^{*}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta,\theta). \tag{4.4}$$

This approximation defines the cross-sectional asymptotic (CSA) log-likelihood function. Similarly, we can derive an approximation valid up to order $1/n^2$:

$$\mathcal{L}_{nT}^{\text{GA}}(\beta,\theta) = \mathcal{L}_{nT}^{*}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta,\theta) + \frac{1}{n^2}\mathcal{L}_{2,nT}(\beta,\theta), \tag{4.5}$$

where $\mathcal{L}_{2,nT}(\beta,\theta)$ is given in (A.5) in Appendix 5. This approximated log-likelihood function defines the granularity adjusted (GA) log-likelihood function.

Both CSA and GA log-likelihood functions have closed form expressions, that is, without integrals w.r.t. the factor values. Then, we can define the CSA and GA maximum likelihood estimators as follows.

**Definition 2.** *(i) The CSA maximum likelihood estimator is:*

$$\left(\tilde{\beta}_{nT}^{CSA}, \tilde{\theta}_{nT}^{CSA}\right) = \arg\max_{\beta,\theta} \mathcal{L}_{nT}^{CSA}(\beta,\theta).$$

*(ii) The GA maximum likelihood estimator is:*

$$\left(\tilde{\beta}_{nT}^{GA}, \tilde{\theta}_{nT}^{GA}\right) = \arg\max_{\beta,\theta} \mathcal{L}_{nT}^{GA}(\beta,\theta).$$

The difference between the GA and CSA maximum likelihood estimators is called the granularity adjustment. This terminology is explained by the link with the recent literature on granularity adjustment in credit risk [see e.g. BCBS (2001), Gordy (2003)]. This literature focuses on the computation of risk measures, such as the Value-at-Risk, for large homogeneous portfolios of $n$ assets, whose values are affected by systematic risk factors. The basic idea is to expand the risk measure around the cross-sectional limit of an infinitely fine grained portfolio ($n = \infty$), and compute the adjustment at order $1/n$ [see Gagliardini, Gouriéroux (2010) for a general presentation of granularity for risk measures].

The CSA and GA estimates can be computed directly from Definition 2. They can also be approximated at the appropriate order by other estimates with simpler expressions. For instance, we have the following result.

**Proposition 8.** *(i) Let us denote:*

$$\hat{\beta}_{nT}^{CSA} = \hat{\beta}_{nT}^{*} + \left(-\frac{\partial \mathcal{L}_{nT}^{*}(\hat{\beta}_{nT}^{*})}{\partial\beta\partial\beta'}\right)^{-1} \frac{1}{n}\frac{\partial \mathcal{L}_{1,nT}}{\partial\beta}(\hat{\beta}_{nT}^{*}, \hat{\theta}_{nT}), \qquad \hat{\theta}_{nT}^{CSA} = \hat{\theta}_{nT},$$

*where* $\left(\hat{\beta}_{nT}^{*}, \hat{\theta}_{nT}\right)$ *are defined in Propositions 5 and 7. Then:*

$$\hat{\beta}_{nT}^{CSA} - \tilde{\beta}_{nT}^{CSA} = o_p(1/n), \qquad \hat{\theta}_{nT}^{CSA} - \tilde{\theta}_{nT}^{CSA} = O_p(1/n).$$

*(ii) The estimator defined by:*

$$\begin{pmatrix} \hat{\beta}_{nT}^{GA} \\ \hat{\theta}_{nT}^{GA} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{nT}^{CSA} \\ \hat{\theta}_{nT}^{CSA} \end{pmatrix} + \left(-\frac{\partial^2 \mathcal{L}_{nT}^{CSA}\left(\hat{\beta}_{nT}^{CSA}, \hat{\theta}_{nT}^{CSA}\right)}{\partial(\beta',\theta')'\partial(\beta',\theta')}\right)^{-1} \frac{\partial \mathcal{L}_{nT}^{GA}\left(\hat{\beta}_{nT}^{CSA}, \hat{\theta}_{nT}^{CSA}\right)}{\partial(\beta',\theta')'},$$

*is such that:*

$$\hat{\beta}_{nT}^{GA} - \tilde{\beta}_{nT}^{GA} = o_p(1/n^2), \qquad \hat{\theta}_{nT}^{GA} - \tilde{\theta}_{nT}^{GA} = o_p(1/n).$$

*Proof.* See Appendix B.9. □

Thus, computationally tractable CSA and GA approximations are easily derived by applying an iteration step in modified Newton-Raphson algorithms with appropriate starting values [see also Gouriéoux, Jasiak (2008) in a static framework].

The interest in introducing several CSA and GA type estimators is threefold. First, it provides approximations of the true ML estimator $(\tilde{\beta}_{nT}, \tilde{\theta}_{nT})$ with different accuracies, as seen in Corollary 9 below.

**Corollary 9.** *The CSA, GA and true ML estimators are such that:*

$$\hat{\beta}_{nT}^{CSA} - \tilde{\beta}_{nT} = O_p(1/n^2), \qquad \hat{\theta}_{nT}^{CSA} - \tilde{\theta}_{nT} = O_p(1/n),$$

*and:*

$$\hat{\beta}_{nT}^{GA} - \tilde{\beta}_{nT} = o_p(1/n^2), \qquad \hat{\theta}_{nT}^{GA} - \tilde{\theta}_{nT} = o_p(1/n).$$

*Proof.* See Appendix B.10. □

Second, the different CSA and GA estimators introduced in Definition 2 and Proposition 8 are all asymptotically efficient, but have different finite sample properties. Therefore, we get a set of consistent and asymptotically efficient estimators in which we can select the preferred one in finite sample, according to the specific model and sample size. Third, as seen in Section 4.5, the higher order expansion of the GA estimator can be used to derive the second-order bias of the true ML estimator in the nonstandard framework considered in this paper.

### 4.5 Higher-order bias of CSA, GA and ML estimators

It is possible to derive higher-order stochastic expansions for the CSA and GA estimators introduced in Definition 2 and Proposition 8 by exploiting the closed form of the CSA and GA log-likelihood functions given in Section 4.4. Then, Corollary 9 allows us to deduce the stochastic expansion of the true ML estimator at the appropriate order, in particular its second-order bias. For expository purpose, we focus on a model with a scalar macro-parameter only, and a single static factor with unconditional pdf $g(f_t;\theta)$. Moreover, let us assume that the cross-sectional and time dimensions are such that $T^b/n$ is bounded and bounded away from 0, for $1 < b < 3/2$. Then, the asymptotic expansion of the bias $E[\hat{\theta}_{nT}] - \theta_0$ at order $o(1/n)$, where $\hat{\theta}_{nT}$ denotes either the CSA, or the GA, or the true ML estimator, involves terms of order [15] $1/T$, $1/\sqrt{nT}$ and $1/n$, that is:

$$E\left[\hat{\theta}_{nT}\right] - \theta_0 = \frac{b_1(\theta_0)}{T} + \frac{b_2(\theta_0)}{\sqrt{nT}} + \frac{b_3(\theta_0)}{n} + o(1/n). \tag{4.6}$$

The coefficients $b_1(\theta_0), b_2(\theta_0), b_3(\theta_0)$ for the second-order bias expansion of the CSA and GA estimators are given in the next proposition.

**Proposition 10.** *(i) The second-order bias expansion of the CSA maximum likelihood estima-*

[15] The bias expansion does not involve a term at order $1/\sqrt{T}$ since the efficient score $I_{1,\theta\theta}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t;\theta_0\right)}{\partial f_t}$ with observable factors has zero expectation. Moreover, when the condition $n = o(T^{3/2})$ is not satisfied, the development at order $1/n$ involves additional terms, of order $1/T^{3/2}$, $1/\left(\sqrt{n}T\right)$, etc. The extension of the results to the general case is theoretically straightforward, but notationally cumbersome.

*tor of the macro-parameter* $\theta_0$ *is such that:*

$$b_1^{CSA}(\theta_0) = I_{1,\theta\theta}^{-2} Cov\left(\frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial \theta^2}, \frac{\partial \log g\left(f_t, \theta_0\right)}{\partial \theta}\right) + \frac{1}{2} I_{1,\theta\theta}^{-2} E\left[\frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta^3}\right],$$

$b_2^{CSA}(\theta_0) = 0$, *and:*

$$b_3^{CSA}(\theta_0) = I_{1,\theta\theta}^{-1} E\left[I_{ff}(t)^{-2}\left(K_{1,2}(t) + \frac{1}{2}K_3(t)\right)\frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f} + \frac{1}{2} I_{ff}(t)^{-1} \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f^2}\right],$$

*where* $K_{1,2}(t) = Cov\left(\frac{\partial^2 \log h\left(y_{i,t}|y_{i,t-1}, f_t\right)}{\partial f_t^2}, \frac{\partial \log h\left(y_{i,t}|y_{i,t-1}, f_t\right)}{\partial f_t}|\underline{f_t}\right)$ *and* $K_3(t) = E\left[\frac{\partial^3 \log h\left(y_{i,t}|y_{i,t-1}, f_t\right)}{\partial f_t^3}|\underline{f_t}\right]$.

*(ii) The second-order bias expansion of the GA maximum likelihood estimator is such that* $b_1^{GA}(\theta_0) = b_1^{CSA}(\theta_0)$, $b_2^{GA}(\theta_0) = b_2^{CSA}(\theta_0) = 0$ *and:*

$$\begin{aligned} b_3^{GA}(\theta_0) &= I_{1,\theta\theta}^{-1} E\left[I_{ff}(t)^{-2}\left(K_{1,2}(t) + K_3(t)\right)\frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f}\right] \\ &\quad + I_{1,\theta\theta}^{-1} E\left[I_{ff}(t)^{-1}\left(\frac{\partial \log g\left(f_t, \theta_0\right)}{\partial f}\frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial f \partial \theta} + \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f^2}\right)\right]. \end{aligned}$$

*Proof.* See Appendix 7. □

The bias term at order $1/T$ for the CSA estimator coincides with the second-order bias of the ML estimator with observable factor values [see e.g. Gouriéroux, Monfort (1995), Section 23.2.1 for the second-order bias of the ML estimator]. The term at order $1/\sqrt{nT}$ does not contribute to the bias. The bias term at order $1/n$ involves third-order (cross-) conditional moments of the micro-density given the factor path, as well as third-order derivatives of the macro-density w.r.t. the factor value and the parameter. From Corollary 9, it follows that the CSA and GA estimators of the macro-parameter differ at order $O_p(1/n)$, which is reflected in Proposition 10 in the different bias coefficients $b_3^{CSA}(\theta_0)$ and $b_3^{GA}(\theta_0)$. Finally, from Corollary 9, we deduce that the bias expansion of the true ML estimator coincides up to order $1/n$ with that of the GA estimator given in Proposition 10 (ii). The bias expansion in Proposition 10 can be used to eliminate the bias up to order $1/n$ by considering the estimator:

$$\bar{\theta}_{nT}^{GA} = \hat{\theta}_{nT}^{GA} - \frac{b_1^{GA}(\hat{\theta}_{nT}^{GA})}{T} - \frac{b_3^{GA}(\hat{\theta}_{nT}^{GA})}{n}.$$

This estimator is asymptotically efficient at first-order and such that $E\left[\bar{\theta}_{nT}^{GA}\right] - \theta_0 = o(1/n)$.

# 5 Approximate linear state space model

In this Section, we consider an exchangeable panel factor model, where the impact of the macro-factors on the micro-dynamics is summarized by a vector of canonical factors. The vector of canonical factors is a noisy affine transformation of a reduced set of factors admitting a Gaussian autoregressive dynamics. For this panel factor model we show that the CSAML and GAML estimators derived in Section 4 are asymptotically equivalent to estimators derived by a linear Kalman filter at appropriate orders in $1/n$. The measurement variables are suitable cross-sectional aggregates that approximate the canonical factors. The methodology is illustrated by the Single Risk Factor model with heterogeneity.

## 5.1 Panel model with canonical factors

The model is defined in three steps. First, the micro-dynamics conditional on the canonical factors is given by the transition pdf:

$$\tilde{h}(y_{i,t}|y_{i,t-1}; a_t), \tag{5.1}$$

where $a_t$ is a $(m, 1)$ vector of latent canonical factors. Second, the canonical factors are such that:

$$a_t = \alpha + \gamma F_t + u_t, \tag{5.2}$$

where $F_t$ is a $(J, 1)$ vector, $J \leq m-1$, and corresponds to a reduced set of latent factors, $\alpha$ is a $(m, 1)$ vector of intercept parameters and $\gamma$ is a $(m, J)$ full rank matrix of factor loadings. The time specific noise is such that $u_t \sim IIN(0, \Delta)$ with $(m, m)$ diagonal variance-covariance matrix $\Delta = diag(\eta_1^2, \cdots, \eta_m^2)$. Third, the reduced factors satisfy a Gaussian Vector Autoregressive (VAR) model of order 1:

$$F_t = \mu + \Phi F_{t-1} + v_t, \tag{5.3}$$

where $(v_t)$ is $IIN(0, \Omega)$ and independent of the time specific noise $(u_t)$. In this panel model the micro-dynamics (5.1) is potentially nonlinear, while the macro-dynamics is linear and given by the Gaussian state-space model (5.2)-(5.3) for the canonical factors.

Let us discuss factor and parameter identification. First, we assume that the canonical factors $a_t$ are cross-sectionally identified. Second, the model is invariant under invertible affine transformations of the latent factor $F_t$ and the associated transformations of the parameters $\gamma$, $\mu$, $\Phi$. To get the identification of factors and parameters, we may assume that $\mu = 0$, $\Omega = Id_J$ and $\gamma'\gamma = Id_J$. The model (5.1)-(5.3) differs from the specification (4.1), (4.3) considered in Section 4.1. Indeed, by writing $f_t = (F_t', u_t')'$ the order condition dim$(a_t) \geq$ dim$(f_t) + 1$ is not satisfied. In fact, the parameters $\alpha$, $\gamma$ and $\Delta$ have to be considered as macro-parameters, and no micro-parameter is included in the model. By setting $\Delta = 0$ we recover the model in equations (4.1) and (4.3), but then the interpretation of $\alpha$ and $\gamma$ is as micro-parameters.

**Example 1: Default Risk Factor Model with stratified heterogeneity**

Let us consider a population of $n$ credits partitioned at each date into $K$ homogeneous strata with size $n_k$, for $k = 1, \cdots, K$, with $K \geq 2$. The individuals are doubly indexed by $(i, k)$, $i = 1, \cdots, n_k$ and $k = 1, \cdots, K$. An extension of the basic SRF model of Section 2.2 accounts for the heterogeneity between strata. The default indicators $y_{i,k,t}$, $i = 1, \cdots, n_k$, $k = 1, \cdots, K$, $t = 1, \cdots, T$, are independent conditionally on the underlying factors, with Bernoulli distribution:

$$y_{i,k,t}|(F_t), (u_{k,t}) \sim \mathcal{B}(1, PD_{k,t}),$$

and conditional default probability:

$$PD_{k,t} = \Phi\left(\alpha_k + \gamma_k' F_t + u_{k,t}\right),$$

where $(F_t)$ and $(u_{k,t})$ are independent, such that $F_t \sim IIN(0, Id_J)$ and $u_{k,t} \sim IIN(0, \eta_k^2)$, and $u_{k,t}$ is a stratum-specific effect common to all individuals in stratum $k$ at date $t$.

The model can be rewritten as:

$$y_{i,k,t}|(F_t), (u_{k,t}) \sim \mathcal{B}\left[1, \Phi(a_{k,t})\right],$$

where the canonical factors $a_t = (a_{1,t}, \cdots, a_{K,t})'$ satisfy equation (5.2) with $\alpha = (\alpha_1, \cdots, \alpha_K)'$, $\gamma = (\gamma_1, \cdots, \gamma_K)'$ and $u_t = (u_{1,t}, \cdots, u_{K,t})'$. In this example, we have $m = K$. The canonical factor satisfies the cross-sectional identification condition and the number $J$ of reduced factors has to be strictly smaller than the number of strata, i.e. $J \leq K - 1$.

## 5.2 The CSA Maximum Likelihood estimator

Under the identification conditions on the canonical factors and parameters, the model includes macro-parameters only, gathered in vector $\theta$. The asymptotic expansion of the ($nT$-standardized) log-likelihood function becomes (see Appendix 8):

$$\mathcal{L}_{nT}(\theta) = \mathcal{L}_{nT}^{*} + \frac{1}{n}\mathcal{L}_{1,nT}(\theta) + O_p\left(\frac{1}{n^2}\right), \tag{5.4}$$

where $\mathcal{L}_{nT}^{*}$ is constant in $\theta$ and function $\mathcal{L}_{1,nT}(\theta)$ is given by:

$$\begin{aligned}
\mathcal{L}_{1,nT}(\theta) \quad = \quad & \frac{1}{T}\log\left(\frac{1}{[(2\pi)^m \det\Delta]^{T/2}[(2\pi)^J \det\Omega]^{T/2}}\right. \\
& \int\cdots\int \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)'\Delta^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)\right. \\
& \left.\left.-\frac{1}{2}\sum_{t=1}^{T}\left(F_t-\mu-\Phi F_{t-1}\right)'\Omega^{-1}\left(F_t-\mu-\Phi F_{t-1}\right)\right\}\prod_{t=1}^{T}dF_t\right),
\end{aligned} \tag{5.5}$$

where vector $\hat{a}_{n,t}$ denotes the cross-sectional ML estimator of the canonical factor at date $t$:

$$\hat{a}_{n,t} = \arg\max_{a_t}\sum_{i=1}^{n}\log\tilde{h}(y_{i,t}|y_{i,t-1};a_t).$$

Function $\mathcal{L}_{1,nT}(\theta)$ is the ($T$-standardized) log-density of the canonical factor path $\underline{a_T}$ evaluated at $a_t = \hat{a}_{n,t}$, $t = 1,\cdots,T$, for parameter value $\theta$. We deduce the next result.

**Proposition 11.** *An asymptotically efficient estimator of parameter $\theta$ can be obtained by applying the linear Kalman filter to the linear state-space model:*

$$\begin{cases} \hat{a}_{n,t} = \alpha + \gamma F_t + u_t, \;\; u_t \sim IIN\left(0,\Delta\right), \\ \;\; F_t = \mu + \Phi F_{t-1} + v_t, \;\; v_t \sim IIN(0,\Omega). \end{cases}$$

*Proof.* See Appendix 8. □

The approximate log-likelihood (5.5) associated with the state space model of Proposition 11 still includes multiple integrals of large dimension since the macro-factors $F_t$ are not semi-parametrically identified. However, due to the interpretation in terms of linear state space model, an asymptotically efficient estimator is easily derived numerically by means of a linear Kalman filter. The linear state space model in Proposition 11 corresponds to equations (5.2)-(5.3) after replacing the unobservable canonical factors $a_t$ by their consistent ML cross-sectional approximations $\hat{a}_{n,t}$. Thus, we have reduced the estimation of a nonlinear latent factor model written for a large number $n$ of individuals to the estimation of a linear state space model with $m$ cross-sectional aggregates.

**Example 1: Default Risk Factor Model with heterogeneity (cont.)**

The cross-sectional estimator of $a_{k,t}$ is simply the transformed default frequency in the category $k$ for date $t$, that is,

$$\hat{a}_{n,k,t} = \Phi^{-1}(\widehat{PD}_{n,k,t}), \tag{5.6}$$

where $\widehat{PD}_{n,k,t} = \frac{1}{n_k}\sum_{i=1}^{n_k} y_{i,k,t}$. The approximate CSA state space model becomes:

$$\begin{cases} \left(\Phi^{-1}(\widehat{PD}_{n,1,t}), \cdots, \Phi^{-1}(\widehat{PD}_{n,K,t})\right)' = \alpha + \gamma F_t + u_t, \quad u_t \sim IIN(0, \Delta), \\ F_t = \mu + \Phi F_{t-1} + v_t, \quad v_t \sim IIN(0, \Omega). \end{cases}$$

## 5.3 The GA Maximum Likelihood estimator

Let us now derive the GAML estimator. For this purpose, the term $\mathcal{L}_{2,nT}(\theta)$ of order $O_p(1/n^2)$ in the log-likelihood expansion has to be taken explicitly into account. In Appendix 8, we show that:

$$\begin{aligned} \mathcal{L}_{nT}(\theta) &= \frac{1}{nT}\log\Bigg[(2\pi)^{-(m+J)T/2}(\det\Omega)^{-T/2}\left(\prod_{t=1}^{T}\det\Psi_{n,t}\right)^{-1/2} \\ &\int\cdots\int\exp\Bigg\{-\frac{1}{2}\sum_{t=1}^{T}\left(\hat{a}_t - \alpha - \gamma F_t + \frac{1}{n}\hat{\xi}_{n,t}\right)'\Psi_{n,t}^{-1}\left(\hat{a}_t - \alpha - \gamma F_t + \frac{1}{n}\hat{\xi}_{n,t}\right) \\ &-\frac{1}{2}\sum_{t=1}^{T}\left(F_t - \mu - \Phi F_{t-1}\right)'\Omega^{-1}\left(F_t - \mu - \Phi F_{t-1}\right)\Bigg\}\prod_{t=1}^{T}dF_t\Bigg] + o_p(1/n^2), \end{aligned} \tag{5.7}$$

where:

$$\hat{\Psi}_{n,t} = \Delta + \frac{1}{n}\hat{\Sigma}_{n,t}, \quad \hat{\Sigma}_{n,t} = \left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log\tilde{h}}{\partial a_t \partial a_t'}(y_{i,t}|y_{i,t-1};\hat{a}_{n,t})\right)^{-1}, \tag{5.8}$$

and $\hat{\xi}_{n,t}$ is a $(m,1)$ vector with components:

$$\hat{\xi}_{n,t,r} = \frac{1}{2}\sum_{l,p,q=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3\log\tilde{h}}{\partial a_{l,t}\partial a_{p,t}\partial a_{q,t}}(y_{i,t}|y_{i,t-1};\hat{a}_{n,t})\right)\hat{\Sigma}_{n,t,lp}\hat{\Sigma}_{n,t,qr}, \quad r = 1, ..., m. \tag{5.9}$$

Thus, we deduce the following result.

**Proposition 12.** *An estimator asymptotically equivalent at order $1/n$ to the GAML estimator of parameter $\theta$ can be obtained by applying the linear Kalman filter to the linear state-space model:*

$$\begin{cases} \hat{a}_{n,t} + \frac{1}{n}\hat{\xi}_{n,t} = \alpha + \gamma F_t + u_t, \ u_t \sim IIN\left(0, \Delta + \frac{1}{n}\hat{\Sigma}_{n,t}\right), \\ F_t = \mu + \Phi F_{t-1} + v_t, \ v_t \sim IIN(0, \Omega), \end{cases}$$

*where $\hat{\Sigma}_{n,t}$ and $\hat{\xi}_{n,t}$ are defined in (5.8) and (5.9), respectively.*

*Proof.* See Appendix 8. □

The state-space model to compute the GAML estimator involves corrections at order $1/n$ in the measurement variable and in the variance of the measurement error. As seen in Corollary 9, corrections of such an order are sufficient for GA estimation of macro-parameters. The matrix $\hat{\Sigma}_{n,t}$ in the variance adjustment is the inverse of the Fisher information matrix for cross-sectional estimation of the canonical factors. The vector $\hat{\xi}_{n,t}$ involves the third-order derivatives of the log micro-density and adjusts the bias of the cross-sectional ML estimator of $a_t$ at order $1/n$. This bias adjustment is not complete, since the GAML (and thus the true ML) estimator has a non-zero second-order bias as seen in Proposition 10.

**Example 1: Default Risk Factor Model with heterogeneity (cont.)**

From Proposition 12, the GA approximate state space model corresponds to approximate measurement equations for the different strata:

$$\hat{a}_{n,k,t} + \frac{1}{n_k}\hat{\xi}_{n,k,t} = \alpha_k + \gamma_k F_t + u^*_{k,t}, \tag{5.10}$$

where the $u^*_{k,t}$ are independent across strata and time, with time-inhomogeneous Gaussian distribution:

$$u^*_{k,t} \sim N\left(0, \eta_k^2 + \frac{\widehat{PD}_{n,k,t}(1-\widehat{PD}_{n,k,t})}{n_k}\left[\frac{d\Phi^{-1}(p)}{dp}\right]^2_{p=\widehat{PD}_{n,k,t}}\right),$$

and:

$$\hat{\xi}_{n,k,t} = (1-2\widehat{PD}_{n,k,t})\left[\frac{d\Phi^{-1}(p)}{dp}\right]_{p=\widehat{PD}_{n,k,t}} + \frac{3}{2}\widehat{PD}_{n,k,t}(1-\widehat{PD}_{n,k,t})\left[\frac{d^2\Phi^{-1}(p)}{dp^2}\right]_{p=\widehat{PD}_{n,k,t}}.$$

The measurement equations (5.10) involve two granularity adjustments. The first one is a variance adjustment and corresponds to the delta-method applied to the transformation of the default frequencies (5.6). The second one is a partial bias adjustment of the estimated transformed default frequencies.

# 6 Stochastic migration model

In this Section we illustrate the finite sample properties of the efficient estimators in Propositions 5 and 7 for a dynamic panel model with factor structure as in equations (4.1) and (4.3).

## 6.1 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 an homogeneous pool of credits [e.g., Gupton et al (1997), Gordy, Heitfield (2002), Gagliardini, Gouriéroux (2005b), Feng, Gouriéroux, Jasiak (2008), Koopman, Lucas, Monteiro (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 [16]. The micro-dynamic model specifies the transition matrices with elements depending on the factor value:

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

where $c_1 < c_2 < ... < c_{K-1}$ and $\alpha_l, \gamma_l, \sigma_l$, $l = 1, ..., K$ are unknown micro-parameters, and $c_0 = -\infty$, $c_K = +\infty$. 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 $a_{l,k,t} = \left(c_k - \gamma_l f_t - \alpha_l\right)/\sigma_l$ 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 (Assumptions A.6-A.7), we impose the constraints $c_1 = 0$, $\sigma_1 = 1$, $\alpha_1 = 0$, $\gamma_1 = 1$ when $K > 2$, and additionally $\sigma_2 = 1$ when $K = 2$ (see

[16] 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.

Appendix 9.1). The model can be written as in equations (4.1) and (4.3), where the vector of canonical factors is $a_t = vec[a_{l,k,t}]$. Finally, the common factor $f_t$ follows a linear Gaussian autoregressive process:

$$f_t = \mu + \rho f_{t-1} + \sigma \eta_t, \tag{6.1}$$

where $(\eta_t)$ is $IIN(0,1)$, and $\mu$, $\rho$ and $\sigma$ are unknown macro-parameters.

## 6.2 Estimation of the micro-parameters

The micro log-density is given by:

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

The estimators of the factor values given $\beta$ are:

$$\hat{f}_{n,t}(\beta) = \arg\max_{f_t} \sum_{k=1}^{K} \sum_{l=1}^{K} N_{lk,t} \log \left[G\left(\frac{c_k - \gamma_l f_t - \alpha_l}{\sigma_l}\right) - G\left(\frac{c_{k-1} - \gamma_l f_t - \alpha_l}{\sigma_l}\right)\right], \quad t = 1, ..., T, \tag{6.2}$$

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:

$$\hat{\beta}^*_{nT} = \arg\max_{\beta} \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=1}^{T} N_{lk,t} \log \left[G\left(\frac{c_k - \gamma_l \hat{f}_{n,t}(\beta) - \alpha_l}{\sigma_l}\right) - G\left(\frac{c_{k-1} - \gamma_l \hat{f}_{n,t}(\beta) - \alpha_l}{\sigma_l}\right)\right]. \tag{6.3}$$

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 $c_1 = \alpha_1 = 0$ and $\gamma_1 = \sigma_1 = \sigma_2 = 1$, the micro-parameter to estimate is $\beta = (\alpha_2, \gamma_2)'$. The parameter values used in the Monte-Carlo study are displayed in Table 1.

**Table 1: Parameter values**

| | | | | | |
|---|---|---|---|---|---|
| $\alpha_1 = 0$ | $\gamma_1 = 1$ | $\sigma_1 = 1$ | $\alpha_2 = -0.5$ | $\gamma_2 = 1$ | $\sigma_2 = 1$ |
| $c_0 = -\infty$ | $c_1 = 0$ | $c_2 = +\infty$ | $\mu = 0.1$ | $\rho = 0.5$ | $\sigma = 0.5$ |

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 (6.3) by numerical optimization, where for given $\beta$ the estimate $\hat{f}_{n,t}(\beta)$ in (6.2) is computed by grid search [17]. As expected from the stochastic migration literature, the $\gamma_2$ 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 the convergence of the finite sample distribution to the asymptotic one is slower. A comparison of Figures 1 and 2 shows that the standard deviations of the estimators decrease by a factor of about 2 when passing from $n = 200$ to $n = 1000$, as suggested by the rate of convergence $\sqrt{nT}$ of the micro-parameters. Finally, we observe a rather small finite sample bias for both estimators.

The semi-parametric efficiency bound for $(\alpha_2, \gamma_2)'$ is given by (see Appendix 9.2):

$$B_{\beta\beta} = E_0 \left[ \mu_{2,t-1}\pi_{22,t}(1-\pi_{22,t}) \left( 1 - \frac{\mu_{2,t-1}\pi_{22,t}(1-\pi_{22,t})\gamma_2^2}{\mu_{1,t-1}\pi_{12,t}(1-\pi_{12,t}) + \mu_{2,t-1}\pi_{22,t}(1-\pi_{22,t})\gamma_2^2} \right) \begin{pmatrix} f_t^2 & f_t \\ f_t & 1 \end{pmatrix} \right]^{-1}, \tag{6.4}$$

where $\pi_{12,t} = 1/(1+e^{f_t})$, $\pi_{22,t} = 1/(1+e^{\gamma_2 f_t + \alpha_2})$ and:

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

[17] The stochastic migration model can be easily extended to include multiple factor in the transition probabilities. In such a model, the estimation procedure described in Section 4.1 based on optimization problem (4.2) is numerically convenient. In particular, the cross-sectional estimators of the canonical factors $\hat{a}_{l,k,t}$ are given in closed form. This estimation procedure has been applied in Gagliardini and Gouriéroux (2005b) without proving its asymptotic efficiency.

The matrix $B_{\beta\beta}$ 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 $B_{\beta\beta}$ depends on macro-parameters $\mu, \rho, \sigma^2$ by means of the expectation $E_0$. The semi-parametric efficiency bound can be approximated numerically by Monte-Carlo integration (see Appendix 9.3).

Figure 3 displays the semi-parametric efficiency bound of parameter $\gamma_2$ as a function of the autoregressive coefficient $\rho$ and the unconditional variance $\frac{\sigma^2}{1-\rho^2}$ of the factor process $(f_t)$. The values of the micro-parameters and $\mu$ are given in Table 1. More precisely, we display the asymptotic standard deviation $\left(\frac{1}{nT}B_{\gamma_2\gamma_2}\right)^{1/2}$, 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 $\rho$ of the factor, except for values of $\rho$ close to 1, where the semi-parametric efficiency bound diverges to infinity.

## 6.3 Estimation of the macro-parameters

Let us now consider the efficient estimation of the macro-parameter $\theta = (\mu, \rho, \sigma^2)'$. The 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 (6.2) and (6.3). 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 $\sigma^2$ 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 $\theta$ in the linear Gaussian model (6.1). Thus, the asymptotic efficiency bound is such that the estimators of $(\mu, \rho)'$ and $\sigma^2$ 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 $\sigma^2$.

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, with variances equal to the efficiency bounds divided by $T$. As expected, it is more difficult to estimate the autoregressive coefficient $\rho$ and the variance $\sigma^2$ than to estimate the intercept $\mu$. 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.

# 7 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 finance and insurance when the joint risk dynamics are followed in large homogenous pools of individual contracts such as corporate loans, household mortgages or life insurance contracts. In such applications the model allows to disentangle the dynamics of systematic and unsystematic risks. These models are also appropriate for performing macro-prediction from tendency surveys [Gouriéroux, Monfort (2009)]. 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 $\beta$ 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 $\beta$ achieves the (semi-) parametric efficiency. We have also shown that an asymptotically efficient estimator of the macro-parameter $\theta$ is obtained by replacing the unobservable factor values by consistent cross-sectional approximations in the likelihood function. These results require a large cross-sectional dimension to approximate the likelihood function, which involves multidimensional integrals, by a closed form expression. The higher-order terms in this expansion around $n = \infty$ are the basis for granularity adjustments, which yield asymptotically efficient estimators, that are more accurate approximations of the true ML estimator.

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 Models in Econometrics”, in *Handbook of Econometrics*, Vol. 2, 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] Andrews, D. (2005): “Cross-Section Regression with Common Shocks”, Econometrica, 73, 1551-1585.

[4] Arellano, M., and S., Bonhomme (2009): ”Robust Priors in Nonlinear Panel Data Models”, Econometrica, 77, 489-536.

[5] 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.

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

[7] Basel Committee on Banking Supervision (2001): ”The New Basel Capital Accord”, Consultative Document of the Bank for International Settlements, April 2001, Part 2: Pillar 1, Section “Calculation of IRB Granularity Adjustment to Capital”.

[8] Basel Committee on Banking Supervision (2003): ”The New Basel Capital Accord”, Consultative Document of the Bank for International Settlements, April 2003, Part 3: The Second Pillar.

[9] Belloni, A., and V., Chernozhukov (2009): “On the Computational Complexity of MCMC-based Estimators in Large Sample”, Annals of Statistics, 37, 2011-2055.

[10] Bester, C., and C., Hansen (2009): ”A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects”, Journal of Business and Economics Statistics, 27, 131-148.

[11] Bickel, P., and J., Yahav (1969): “Some Contributions to the Asymptotic Theory of Bayes Solutions”, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 11, 257-276.

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

[13] Cappé, O., Moulines, E., and T., Rydén (2005): *Inference in Hidden Markov Models*, Springer Verlag, New York.

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

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

[16] 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.

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

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

[19] 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.

[20] Dempster, A., Laird, N., and D., Rubin (1977): ”Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm (with discussion)”, Journal of the Royal Statistical Society, B, 39, 1-38.

[21] 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.

[22] Duffie, D., Eckner, A., Horel, G., and L., Saita (2009): ”Frailty Correlated Defaults”, Journal of Finance, 64, 2089-2123.

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

[24] Fiorentini, G., Sentana, E., and N., Shephard (2004): “Likelihood-Based Estimation of Latent Generalized ARCH Structures”, Econometrica, 72, 1481-1517.

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

[26] 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.

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

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

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

[30] Gagliardini, P., and C., Gouriéroux (2010): “Granularity Adjustment in Dynamic Multiple Factor Models: Systematic vs Unsystematic Risk”, CREST DP.

[31] Gordy, M. (2003): ”A Risk-Factor Model Foundation for Rating-Based Bank Capital Rules”, Journal of Financial Intermediation, 12, 199-232.

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

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

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

[35] Gouriéroux, C., and A., Monfort (2009): “Granularity in Qualitative Factor Models”, forthcoming in the Journal of Credit Risk.

[36] 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.

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

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

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

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

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

[42] Ibragimov, I., and R., Has’minskii (1981): *Statistical Estimation: Asymptotic Theory*, Springer, Berlin.

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

[44] Kingman, J. (1978): “Uses of Exchangeability”, The Annals of Probability, 6, 183-197.

[45] Koopman, S., Lucas, A., and A., Monteiro (2008): “The Multi-State Latent Factor Intensity Model for Credit Rating Transitions”, Journal of Econometrics, 142, 399-424.

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

[47] Laplace, P. (1774): “Mémoire sur la probabilité des causes par les événements”, Mémoires de mathématique et de physique présentés à l’Academie Royale des Sciences par divers savants, dans ses assemblées, 6, 621-656.

[48] Merton, R. (1974): ”On the Pricing of Corporate Debt: The Risk Structure of Interest Rates”, Journal of Finance, 29, 449-470.

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

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

[51] 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.

[52] 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.

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

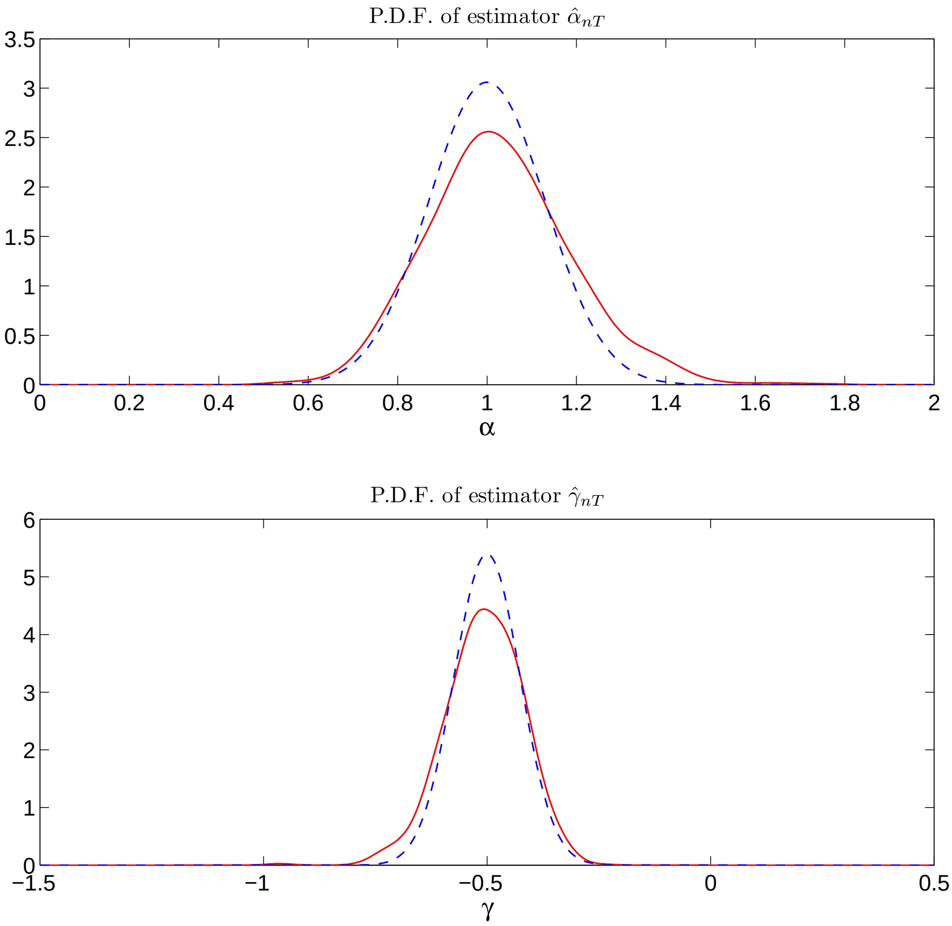
[54] Tierney, L., and Kadane, J. (1986): “Accurate Approximations for Posterior Moments and Marginal Densities”, Journal of the American Statistical Association, 81, 82-86.

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

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

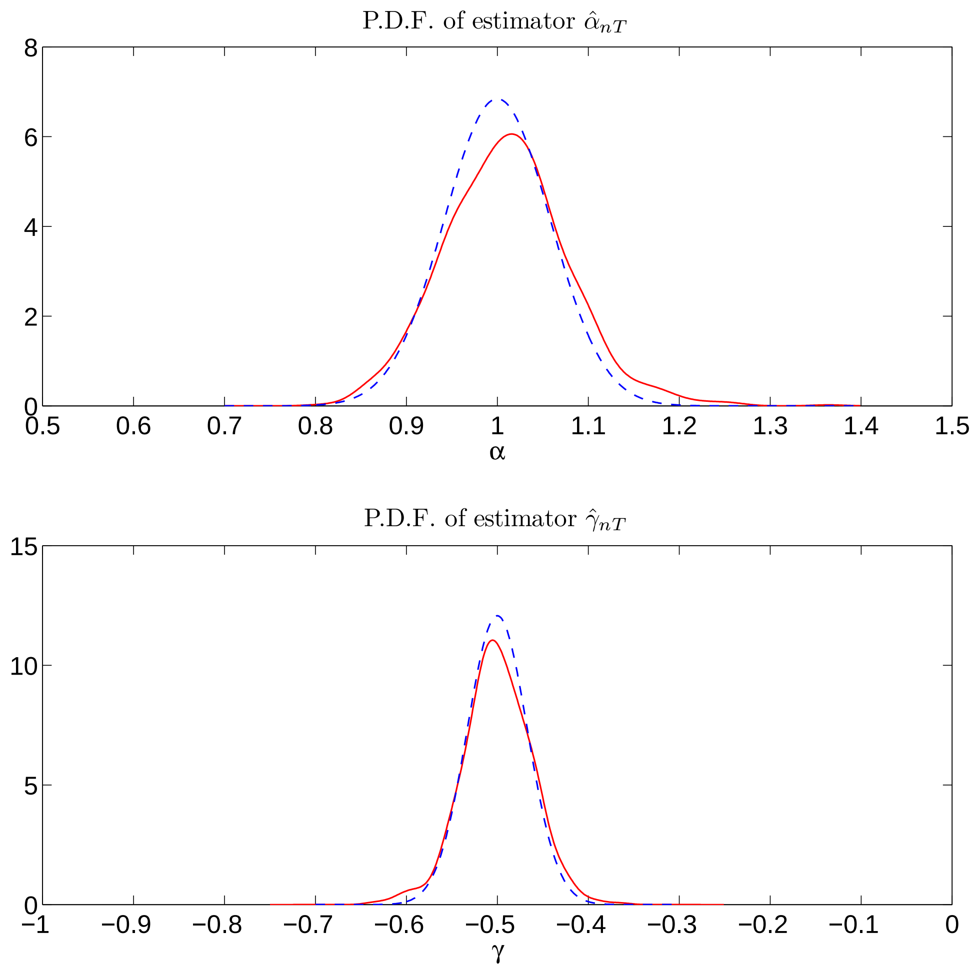
[57] Woutersen, T. (2002): ”Robustness Against Incidental Parameters”, Working Paper.

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



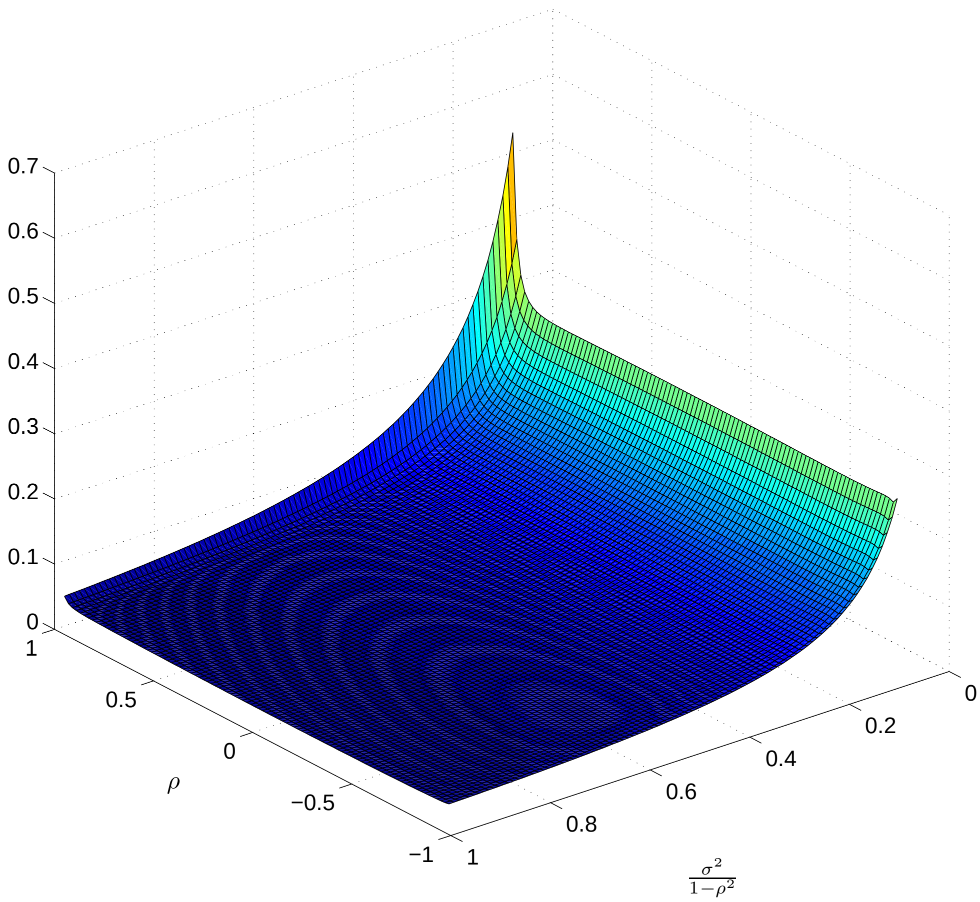

The solid lines give the pdf of the semiparametrically efficient estimators of parameter $\gamma$ (upper Panel, true value 1) and parameter $\alpha$ (lower Panel, true value $-0.5$). The pdf is computed by a kernel density estimator. Sample sizes are $n = 200$ and $T = 20$. The dashed lines in the two 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 $T = 20$.



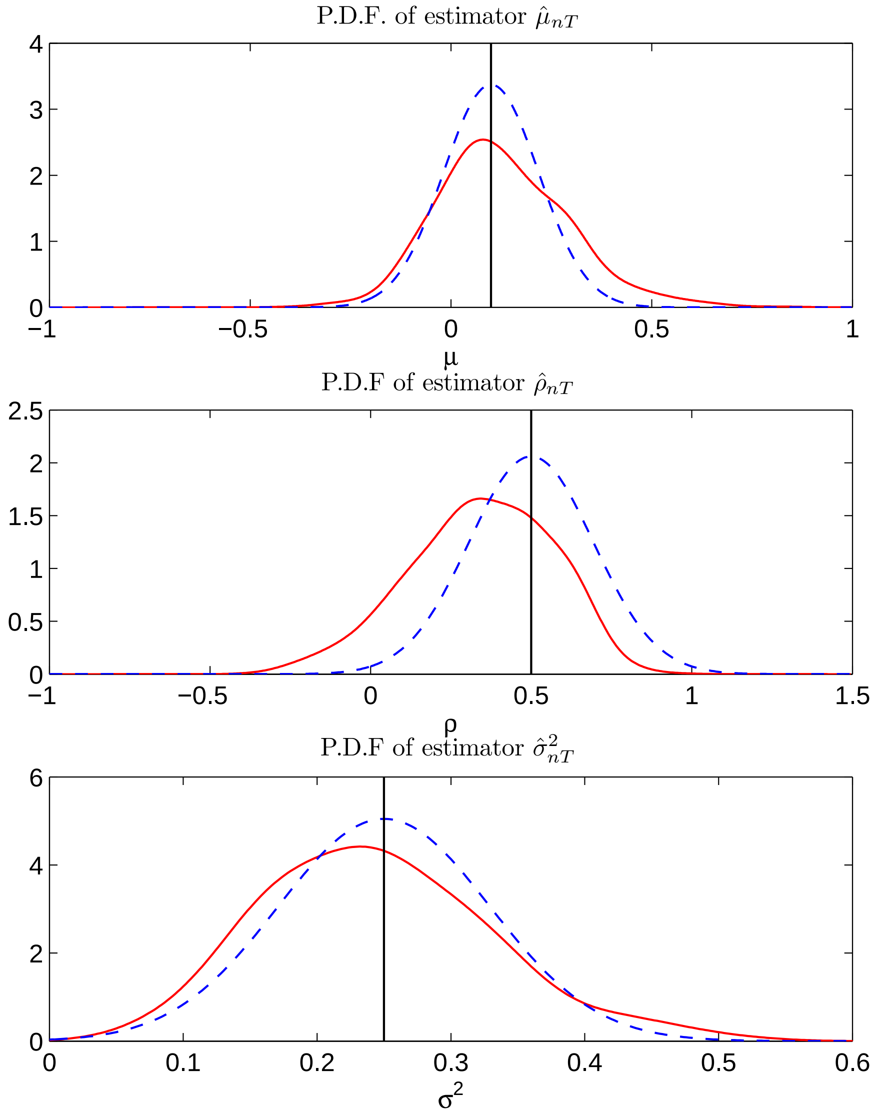

The solid lines give the pdf of the semiparametrically efficient estimators of parameter $\gamma$ (upper Panel, true value 1) and parameter $\alpha$ (lower Panel, true value $-0.5$). The pdf is computed by a kernel density estimator. Sample sizes are $n = 1000$ and $T = 20$. The dashed lines in the two 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 $\gamma_2$.



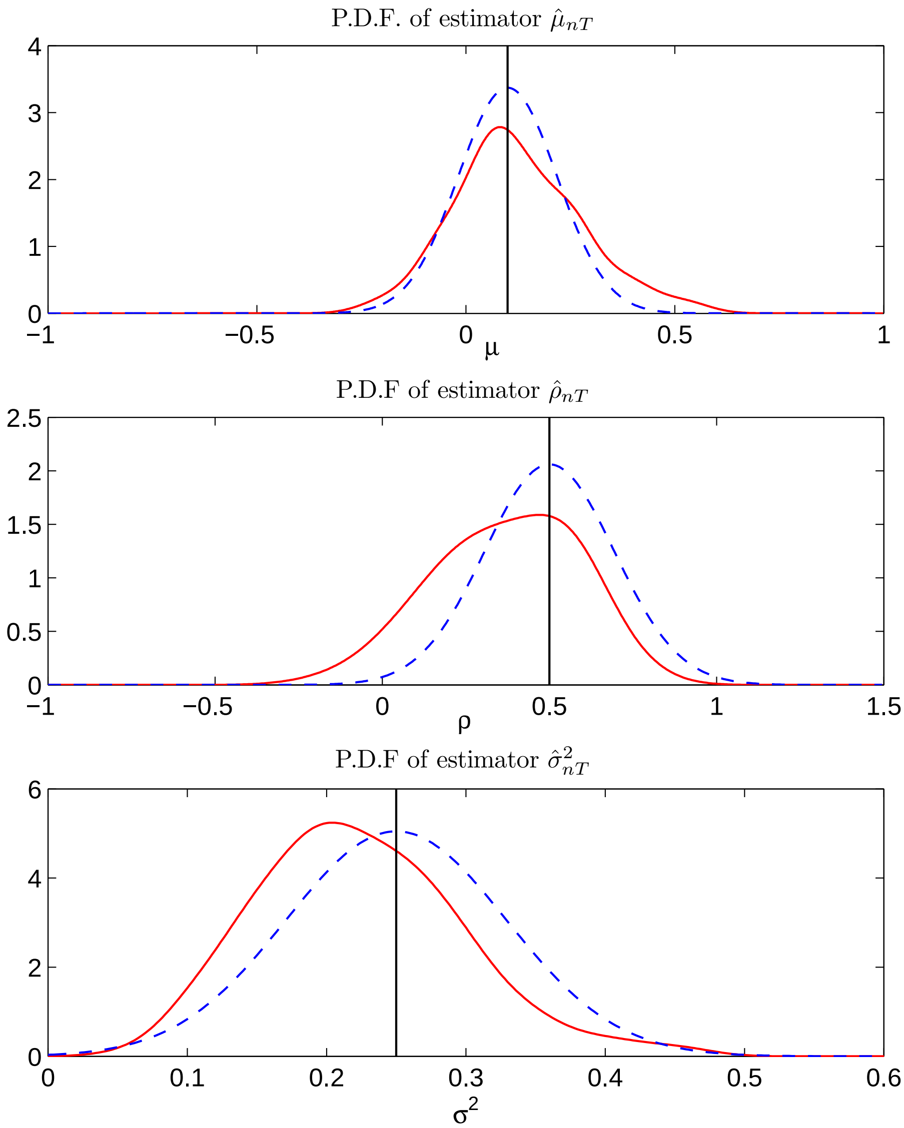

The figure displays $\left(\frac{1}{nT}B_{\gamma_2\gamma_2}\right)^{1/2}$, where $B_{\gamma_2\gamma_2}$ is the semiparametric efficiency bound for parameter $\gamma_2$ and $n = 1000, T = 20$, as a function of the autoregressive coefficient $\rho$ and the variance $\frac{\sigma^2}{1-\rho^2}$ of the factor 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 $\mu$ (upper Panel, true value $0.1$), parameter $\rho$ (central Panel, true value $0.5$) and parameter $\sigma^2$ (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 $\mu$ (upper Panel, true value $0.1$), parameter $\rho$ (central Panel, true value $0.5$) and parameter $\sigma^2$ (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 $T$.

# APPENDIX 1

## Weak LLN and Slutsky Theorem

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

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

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 $\varphi$ 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) $P\left[\xi_t\geq u\right]\leq C_1\exp\left(-C_2u^{\delta}\right)$ *as $u\to\infty$, for some constants $C_1,C_2,\delta>0$, where $\xi_t=\sup_{\beta\in\mathcal{B}}\sigma_t^2(\beta)$ 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 for $\xi_t=\sup_{\beta\in\mathcal{B}}\tilde{\sigma}_t^2(\beta)$, 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, the uniform stochastic convergence (A.1) holds.*

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 $\varphi$,

(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 $\varphi$ 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 $\varphi$ 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 $\varphi$ 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.*

(viii) *Condition* (1) (v) *of Lemma A.1 holds for* $\xi_t = \left(\inf_{\beta\in\mathcal{B}} \lambda_t(\beta)\right)^{-1}$, *where $\lambda_t(\beta)$ is* *the smallest eigenvalue of matrix* $\mu_t(\beta)$*, for* $\xi_t = \sup_{\beta\in\mathcal{B}} \dfrac{\sigma_t(\beta)}{\lambda_t(\beta)}$*, and for* $\xi_t = \sup_{\beta\in\mathcal{B}} \dfrac{\tilde{\sigma}_t(\beta)}{\lambda_t(\beta)}$*, where* $\sigma_t(\beta)$ *and* $\tilde{\sigma}_t(\beta)$ *are defined in Lemma A.1.*

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

(i) $\varphi$ *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* $C, \tau$ *satisfy* $C > 0$*,* $\tau \leq 2$*, and function* $\psi$ *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, the uniform stochastic convergence (A.1) holds.*

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 $\varphi$ 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 $\varphi$ 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* $\varphi$ *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, the uniform stochastic convergence (A.1) 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 through a conditioning argument. Let us consider the ML estimator:

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

where $L_n(\theta) = \dfrac{1}{n}\sum_{i=1}^{n} l_i(\theta)$ 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 $\theta_0$ 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:

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

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

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

where $KL(\theta,\theta_0) = L(\theta_0) - L(\theta) = E_0\left[\log\left(\dfrac{h(y_i,\theta_0)}{h(y_i,\theta)}\right)\right]$ is the Kullback-Leibler discrepancy between $\theta$ and $\theta_0$. 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 $I(\theta,\theta_0) = E_0\left[\dfrac{\partial \log h(y_i,\theta)}{\partial\theta}\dfrac{\partial \log h(y_i,\theta)}{\partial\theta'}\right]$. 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 $\Theta$ and the parametric model) such that for any $n$ and $\varepsilon > 0$:*

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

**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 $\varepsilon$ and sample size $n$ depends on the true probability distribution [18]. This dependence is summarized by statistics $\mathcal{K}$, $\Gamma$ and $\mathcal{R}$, and by exponent $\gamma$. In particular, the coefficient of $n\varepsilon^2$ in the exponential term involves the ratio $\dfrac{\mathcal{K}}{1+\Gamma/\mathcal{K}}$. 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 $\Gamma$.

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 $\alpha$ defined by:

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

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

[18] Moreover, compared to the results in Fu (1982), Lemma A.4 applies for multivariate parameter $\theta$.

*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 $\varepsilon > 0$:*

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

*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 $\alpha$ and $\alpha(\beta)$ for given $\beta \in \mathcal{B}$, the scalar $\Gamma$ is given by:*

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

*with $I(\theta, \theta_0) = E_0\left[\frac{\partial \log h(y_i, \theta)}{\partial\alpha}\frac{\partial \log h(y_i, \theta)}{\partial\alpha'}\right]$, 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 $\beta_0$ and $\theta_0$ are interior points of $\mathcal{B}$ and $\Theta$, respectively.*

**H.2:** *For any multi-index $\alpha \in \mathbb{N}^{q+K}$ such that $|\alpha| \leq 3$:*

$$E_0 \left[ \sup_{\beta \in \mathcal{B}} \left| \frac{\partial^{|\alpha|} \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial^{\alpha}(\beta', f')} \right|^9_{f=f_t(\beta)} \right] < \infty,$$

*where $K = dim(f)$. Moreover $E_0 \left[ \sup_{\beta \in \mathcal{B}} \left\| \frac{\partial f_t(\beta)}{\partial \beta'} \right\|^8 \right] < \infty$.*

**H.3:** *For any $\beta \in \mathcal{B}$:* (i) $E_0 \left[ \| f_t(\beta) - E_0[f_t(\beta)|f_t, ..., f_{t-m}] \|^2 \right] = O\left(m^{-\alpha}\right)$*,*
(ii) $E_0 \left[ |E_0 \left[ l_{i,t}(\beta)|\underline{f_t} \right] - E_0[l_{i,t}(\beta)|f_t, ..., f_{t-m}]|^2 \right] = O\left(m^{-\alpha}\right)$*,*
(iii) $E_0 \left[ \|E_0 \left[ H_{i,t}(\beta)|\underline{f_t} \right] - E_0[H_{i,t}(\beta)|f_t, ..., f_{t-m}]\|^2 \right] = O\left(m^{-\alpha}\right)$*, as $m \to \infty$, for some $\alpha > 0$, where $l_{i,t}(\beta) = \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)$ and $H_{i,t}(\beta) = \left[ -\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial(f', \beta')' \partial(f', \beta')} \right]_{f=f_t(\beta)}$.*

**H.4:** *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.*

**H.5:** *$P\left[\xi_t \geq u\right] \leq C_1 \exp\left(-C_2 u^{\delta}\right)$ as $u \to \infty$, for some constants $C_1, C_2, \delta > 0$ and:*
(i) $\xi_t = \left( \inf_{\beta \in \mathcal{B}} \lambda_t(\beta) \right)^{-1}$*, where $\lambda_t(\beta)$ is the smallest eigenvalue of matrix $I(t, \beta)$;*
(ii) $\xi_t = \sup_{\beta \in \mathcal{B}} E_0 \left[ |l_{i,t}(\beta)|^2 |\underline{f_t} \right]$;
(iii) $\xi_t = \sup_{\beta \in \mathcal{B}} E_0 \left[ \sup_{f: \|f - f_t(\beta)\| \leq \eta^*} \left\| \frac{\partial \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial f} \right\|^2 |\underline{f_t} \right]$*, where $\eta^* > 0$;*
(iv) $\xi_t = \sup_{\beta \in \mathcal{B}} \frac{\sigma_t(\beta)}{\lambda_t(\beta)}$*, where $\sigma_t^2(\beta) = E_0 \left[ \|H_{i,t}(\beta)\|^2 |\underline{f_t} \right]$;*
(v) $\xi_t = \sup_{\beta \in \mathcal{B}} \frac{\tilde{\sigma}_t(\beta)}{\lambda_t(\beta)}$*, where $\tilde{\sigma}_t^2(\beta) = E_0 \left[ \sup_{f: \|f - f_t(\beta)\| \leq \eta^*} \left\| \frac{\partial^3 \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial(f', \beta')' \partial(f', \beta') \partial f} \right\|^2 |\underline{f_t} \right]$ and $\eta^* > 0$.*

**H.6:** *The process* $\sup_{\beta\in\mathcal{B}}\|f_t(\beta)\|$ *is such that* $P\left[\sup_{\beta\in\mathcal{B}}\|f_t(\beta)\|\geq u\right]\leq C_3\exp\left(-C_4u^{\varrho}\right)$ *as* $u\to\infty$, *for some constants* $C_3, C_4, \varrho>0$.

**H.7:** *The set* $\mathcal{F}_n\subset\mathbb{R}^K$ *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* $\mathbb{R}^K$ *centered at* $0$ *and with radius* $\rho_n=\left[(2/C_4)\log(n)\right]^{1/\varrho}$.

**H.8:** *There exists a constant* $a_1\geq 0$ *such that:*

$$\mathcal{K}_t:=\inf_{n\geq 1}\inf_{\beta\in\mathcal{B}}\inf_{f\in\mathcal{F}_n:f\neq f_t(\beta)}[\log(n)]^{a_1}\frac{2KL_t(f,f_t(\beta);\beta)}{\|f-f_t(\beta)\|^2}>0,$$

*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.9:** *There exist constants* $\gamma\geq 4$ *and* $a_2\geq 0$ *such that:*

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

*for any* $t$, $P$*-a.s. Moreover* $E_0\left[\mathcal{R}_t^2\right]<\infty$.

**H.10:** *There exist constants* $C_5, C_6>0$ *and* $a_3>1$ *such that* $P\left[\xi_t\geq u\right]\leq C_5\exp\left[-C_6u^{1/(a_3-1)}\right]$ *as* $u\to\infty$, *where* $\xi_t=\left(\frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right)^{-1}$ *and:*

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

**H.11:** *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.12:** $P\left[\zeta_t\geq u\right]\leq C_7\exp\left(-C_8u^{1/\chi}\right)$, *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.13:** *Assumptions H.11 and H.12 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.14:** $E_0\left[\left\|\dfrac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}\right\|^{\nu}\right] < \infty$, $\nu > 2$.

Assumption H.1 is a standard condition on parameter sets and true parameter values. Assumptions H.2-H.6 concern the micro log-density and the pseudo-true factor values. Specifically, Assumption H.2 requires finite higher-order moments for $\log h(y_{i,t}|y_{i,t-1}, f; \beta)$ and its derivatives w.r.t. $\beta$ and $f$, evaluated at $f = f_t(\beta)$, uniformly in $\beta \in \mathcal{B}$. Similarly, finite higher moments for $\partial f_t(\beta)/\partial\beta'$ uniformly in $\beta$ are required. Under Assumption H.3, the pseudo-true factor value $f_t(\beta)$, and the conditional expectation of $l_{i,t}(\beta) = \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)$ and $H_{i,t}(\beta) = \left[-\dfrac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial(f', \beta')'\partial(f', \beta')}\right]_{f=f_t(\beta)}$ 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 implies the concavity of the cross-sectional likelihood function $E_0\left[\log h(y_{i,t}|y_{i,t-1}, f; \beta)|\underline{f_t}\right]$ w.r.t. $(f, \beta)$, 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 (3.10), Assumption H.4 strengthens identification Assumptions A.6 and A.7 for micro-parameter $\beta$. Assumption H.4 is equivalent to the condition $\lambda_t(\beta) > 0$, for any $t$ and $\beta \in \mathcal{B}$, P-a.s., on the smallest eigenvalue of matrix $I(t, \beta)$. Assumption H.5 (i) is a tail condition on the stationary distribution of process $\xi_t = \left(\inf_{\beta\in\mathcal{B}} \lambda_t(\beta)\right)^{-1}$. This condition is satisfied, when the factor paths associated with very small $\lambda_t(\beta)$ for some $\beta$ are sufficiently unfrequent. Assumptions H.5 (ii)-(v) are similar tail conditions for the stationary distributions of process $\xi_t = \sup_{\beta\in\mathcal{B}} E_0\left[|l_{i,t}(\beta)|^2|\underline{f_t}\right]$ as well as processes involving the derivatives of the micro log-density function. Assumptions H.1-H.5 are used in Appendix A.6.2 to prove the uniform convergence of the likelihood function $\mathcal{L}^*_{nT}(\beta)$ defined in (3.5), and of its second-order derivative w.r.t. $\beta$, using Lemmas A.1, A.2 and Corollary A.3 given in Appendix A.1.

Assumptions H.6-H.10 are used in Lemma A.6 to derive the uniform rate of convergence of the factor approximations (see Appendix A.4). Specifically, Assumption H.6 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.7 gives a lower bound on this growth rate. Under Assumptions H.6 and H.7, 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.8 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.9 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.8 and H.9 imply an upper bound on the growth rate of set $\mathcal{F}_n$. Assumption H.10 is a tail condition on the stationary distribution of the process $\frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}$ in a neighbourhood of 0. The quantity $\frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}$ involves the measure of Kullback-Leibler discrepancy $\mathcal{K}_t$, and the measure $\Gamma_t$ 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.10 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, Assumptions H.11-H.14 concern the macro log-density and its derivatives w.r.t. factor values and macro-parameter $\theta$. Specifically, Assumption H.11 requires finite moments for $\log g(f_t(\beta)|f_{t-1}(\beta);\theta)$ and its first-order derivatives w.r.t. $\beta$ and $\theta$. Assumption H.12 is a condition on the right tail of process $\zeta_t$. This assumption is used to prove a WLLN for time series averages with true factor values replaced by cross-sectional estimators (see Lemma A.8 in Appendix 4). Assumption H.14 is a bound on the moment of the macro-score $\frac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}$ of order $\nu > 2$. Assumptions H.11-H.14 imply the uniform convergence of $\mathcal{L}_{1,nT}(\beta,\theta)$ [see (3.6)] and the Hessian $\frac{\partial^2 \mathcal{L}_{1,nT}(\beta,\theta)}{\partial\theta\partial\theta'}$, uniformly in $\beta \in \mathcal{B}$, $\theta \in \Theta$, as well as the asymptotic normality of the score $\frac{\partial \mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\theta}$ in the proof of Proposition 5 (see Appendix 6). These assumptions are also used to prove the asymptotic efficiency of the estimator of $\theta$ 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.7.

### A.4.1 Uniform rate of convergence

**Lemma A.6:** *Under Assumptions A.1-A.5, H.1, H.6-H.10, 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.8-H.10.*

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.10). 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.8 and H.9). 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 $\varepsilon_n = \sqrt{r \dfrac{(\log n)^a}{n}}$, where $r > 0$ is a constant. We have to show that for any $\eta > 0$, there exists a value of $r$ such that $P\left[ \sup_{1 \le t \le T} \sup_{\beta \in \mathcal{B}} \left\| \widehat{f}_{n,t}(\beta) - f_t(\beta) \right\| \ge \varepsilon_n \right] \le \eta$, for large $n$ and $T$ such that $T^b/n = O(1)$, $b > 1$. We have:

$$\begin{aligned} P\left[ \sup_{1 \le t \le T} \sup_{\beta \in \mathcal{B}} \left\| \widehat{f}_{n,t}(\beta) - f_t(\beta) \right\| \ge \varepsilon_n \right] &\le TP\left[ \sup_{\beta \in \mathcal{B}} \left\| \widehat{f}_{n,t}(\beta) - f_t(\beta) \right\| \ge \varepsilon_n \right] \\ &= TE\left[ P\left[ \sup_{\beta \in \mathcal{B}} \left\| \widehat{f}_{n,t}(\beta) - f_t(\beta) \right\| \ge \varepsilon_n \mid \underline{f_t} \right] \right]. \end{aligned} \tag{A.3}$$

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 $\beta$, 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.7, and A.1-A.2, respectively. Condition iii) is satisfied since Assumption H.8 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.9 and:

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

Moreover, from Assumption H.8 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.9-H.10. Then, from Lemma A.5 we have:

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

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:

$$\begin{aligned}
& P\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta)-f_t(\beta)\right\|\geq\varepsilon_n \mid \underline{f_t}\right] \\
\leq\ & \frac{c_1}{r^{q/2}}n^{K+3q/2}\exp\left(-c_2 r[\log(n)]^{a_3}\frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right)+c_3\frac{r^{\gamma/2-1}}{n^{\gamma/2-1}}[\log(n)]^{a_1(\gamma-1)+(a_2+a_3)\gamma/2-a_3}\frac{\mathcal{R}_t}{\mathcal{K}_t} \\
& +1\left\{\bigcup_{\beta\in\mathcal{B}}[f_t(\beta)\in\mathcal{F}_n^c]\right\}, P\text{-a.s..} \qquad \text{(A.4)}
\end{aligned}$$

**ii) Integrating out the factor path**

By integrating out the factor path $\underline{f_t}$, we get from (A.3) and (A.4):

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

Let us now bound these three terms.

(a) To bound $I_{1,n,T}$ we use the next Lemma A.7.

**Lemma A.7:** *Let $\xi$ be a positive random variable such that $P[\xi\geq u]\leq C_1\exp\left(-C_2u^{\varrho}\right)$ as $u\to\infty$, for some constants $C_1, C_2, \varrho>0$. Then $E[\exp\left(-u\xi^{-1}\right)]\leq\tilde{C}_1\exp\left(-\tilde{C}_2u^{\varrho/(1+\varrho)}\right)$ as $u\to\infty$, for some constants $\tilde{C}_1, \tilde{C}_2>0$.*

**Proof:** See Appendix B.6.

From Assumption H.10 and Lemma A.7, and using $T/n^{\gamma/2-1}=O(1)$, we have for some constants $c_4, c_5>0$:

$$I_{1,n,T} \leq \frac{c_1}{r^{q/2}}c_4Tn^{K+3q/2}\exp\left(-c_5(c_2r)^{1/a_3}\log(n)\right)\leq\frac{c_1}{r^{q/2}}c_4n^{\gamma/2-1+K+3q/2-c_5(c_2r)^{a_3}}=o(1),$$

if $r>\dfrac{1}{c_2}\left(\dfrac{\gamma/2+K+3q/2-1}{c_5}\right)^{1/a_3}$.

(b) From Assumptions H.9 and H.10, $E\left[\dfrac{\mathcal{R}_t}{\mathcal{K}_t}\right]\leq E\left[\mathcal{R}_t^2\right]^{1/2}E\left[\mathcal{K}_t^{-2}\right]^{1/2}<\infty$. 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.6 and H.7, we have:

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

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.8:** *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 $\theta\in\Theta$: $E_0\left[\|G(F_t(\beta),\theta)\|^{\kappa}\right]<\infty$, $\kappa>2$, $E\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial vec[G(F_t(\beta),\theta)]}{\partial\beta'}\right\|\right]<\infty$, 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) *$P\left[\zeta_t\geq u\right]\leq c_1\exp\left(-c_2u^{1/\chi}\right)$, 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.6-H.10, 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.7.

Lemma A.8 is used in the proofs of Proposition 1 (Appendix 5), Proposition 3 (Appendix 6) and Proposition 7.

## 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 $\hat{f}_{nt}(\beta)$, $t=1,...,T$, and define:

$$\begin{aligned}\psi_{nt}\left(f_t,f_{t-1}\right) &= \sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1},f_t;\beta\right)-\sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right)\\ &+\frac{1}{2}\sqrt{n}\left(f_t-\hat{f}_{nt}(\beta)\right)' I_{nt}(\beta)\sqrt{n}\left(f_t-\hat{f}_{nt}(\beta)\right)\\ &+\log g\left(f_t|f_{t-1};\theta\right)-\log g\left(\hat{f}_{nt}(\beta)\,|\hat{f}_{n,t-1}(\beta);\theta\right).\end{aligned}$$

Then:

$$\begin{aligned}l\left(\underline{y_T};\beta,\theta\right) &= \prod_{t=1}^{T}\prod_{i=1}^{n}h\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right)\prod_{t=1}^{T}g\left(\hat{f}_{nt}(\beta)\,|\hat{f}_{n,t-1}(\beta);\theta\right)\\ &\int\cdots\int\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\sqrt{n}\left(f_t-\hat{f}_{nt}(\beta)\right)' I_{nt}(\beta)\sqrt{n}\left(f_t-\hat{f}_{nt}(\beta)\right)\right\}\\ &\exp\left\{\sum_{t=1}^{T}\psi_{n,t}\left(f_t,f_{t-1}\right)\right\}\prod_{t=1}^{T}df_t.\end{aligned}$$

Let us introduce the change of variable:

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

Then:

$$\begin{aligned}&l\left(\underline{y_T};\beta,\theta\right)\\ =\;&\left(\frac{2\pi}{n}\right)^{TK/2}\prod_{t=1}^{T}\left[\det I_{nt}(\beta)\right]^{-1/2}\prod_{t=1}^{T}\prod_{i=1}^{n}h\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta\right)\prod_{t=1}^{T}g\left(\hat{f}_{nt}(\beta)\,|\hat{f}_{n,t-1}(\beta);\theta\right)\\ &\frac{1}{(2\pi)^{TK/2}}\int\ldots\int\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}Z_t'Z_t\right\}\\ &\exp\left\{\sum_{t=1}^{T}\psi_{n,t}\left(\hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2}Z_t,\hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2}Z_{t-1}\right)\right\}\prod_{t=1}^{T}dZ_t.\end{aligned}$$

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

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

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

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 $Z := (Z_1,...,Z_T)'$. Expanding $\psi_{n,t}$ at order $1/n$ yields:

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

where:

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

and:

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

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:

$$\exp\left[\left(\frac{T}{n}\right)\Psi_{nT}\left(\beta,\theta\right)\right] = 1 + \frac{T}{n}\mathcal{L}_{2,nT}(\beta,\theta) + o_p(T/n),$$

where:

$$\begin{aligned}
\mathcal{L}_{2,nT}(\beta,\theta) \;\; = \;\; & \frac{1}{8}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-2}K_{4,n,t}(\beta) + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-1}D_{20,nt}(\beta,\theta) \\
& + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t-1}\left(\beta\right)\right]^{-1}D_{02,nt}(\beta,\theta) + \frac{5}{24}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-3}K_{3,nt}^{2}(\beta) \\
& + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}D_{10,nt}^{2}(\beta,\theta)\left[I_{n,t}\left(\beta\right)\right]^{-1} + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}D_{01,nt}^{2}(\beta,\theta)\left[I_{n,t-1}\left(\beta\right)\right]^{-1} \\
& + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t}\left(\beta\right)\right]^{-2}D_{10,nt}(\beta,\theta)K_{3,n,t}(\beta) \\
& + \frac{1}{2}\frac{1}{T}\sum_{t=2}^{T}\left[I_{n,t-1}\left(\beta\right)\right]^{-2}D_{01,nt}(\beta,\theta)K_{3,n,t-1}(\beta) \\
& + \frac{1}{T}\sum_{t=1}^{T}\left[I_{n,t-1}\left(\beta\right)\right]^{-1}D_{10,n,t-1}(\beta,\theta)D_{01,nt}(\beta,\theta). \qquad \text{(A.5)}
\end{aligned}$$

From Lemma A.6 in Appendix 4, 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.8 in Appendix 4, we get $\mathcal{L}_{2,nT}(\beta,\theta) = O_p(1)$ 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.9:** *Let the estimator* $\left(\tilde{\beta}_{nT},\tilde{\theta}_{nT}\right)$ *be defined by:*

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

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

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

*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* $\beta\rightarrow\mathcal{L}^{*}(\beta)$ *is uniquely maximized at the interior point* $\beta_{0}\in\mathcal{B}$*;*
(ii) *Function* $\theta\rightarrow\mathcal{L}_{1}(\beta_{0},\theta)$ *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) $\displaystyle\sup_{\beta\in\mathcal{B},\theta\in\Theta}\frac{\partial\mathcal{L}_{1,nT}\left(\beta,\theta\right)}{\partial\beta}=O_{p}(1).$

(5) (i) $\displaystyle\sup_{\beta\in\mathcal{B},\theta\in\Theta}\Psi_{nT}(\beta,\theta)=O_{p}(1)$; (ii) $\displaystyle\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial\Psi_{nT}\left(\beta,\theta\right)}{\partial\left(\beta',\theta'\right)'}\right\|=O_{p}(1).$

*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 $\tilde{\beta}_{nT}$ and $\tilde{\theta}_{nT}$ are consistent and jointly asymptotically normal:*

$$\begin{bmatrix} \sqrt{nT}\left(\tilde{\beta}_{nT} - \beta_0\right) \\ \sqrt{T}\left(\tilde{\theta}_{nT} - \theta_0\right) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} (I_0^*)^{-1} & 0 \\ 0 & I_{1,\theta\theta}^{-1} \end{bmatrix}\right).$$

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

**Proof:** See Appendix B.8.

### A.6.2 Proof of Proposition 3

The efficiency bound $B^*$ is the asymptotic variance-covariance matrix of the ML estimator $\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right) = \arg\max_{\beta \in \mathcal{B}, \theta \in \Theta} \mathcal{L}_{nT}(\beta, \theta)$, where $\mathcal{L}_{nT}(\beta, \theta)$ is defined in Corollary 2. This asymptotic variance-covariance matrix is derived by applying Lemma A.9. Let us verify the conditions of Lemma A.9.

**Condition (1) of Lemma A.9:** We have:

$$\mathcal{L}^*_{nT}(\beta) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \log h\left(y_{i,t} | y_{i,t-1}, \hat{f}_{nt}(\beta); \beta\right). \tag{A.6}$$

This converges to $\mathcal{L}^*(\beta) = E_0\left[\log h\left(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta\right)\right]$ 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 $\varphi$ corresponding to the identity mapping. Indeed, condition (1) of Lemma A.1 is implied by Assumptions H.1, H.2, H.3 (ii), H.5 (ii)-(iii), and condition (3) of Lemma A.1 is implied by Lemma A.6. Further:

$$\mathcal{L}_{1,nT}(\beta, \theta) = -\frac{1}{2}\frac{1}{T} \sum_{t=1}^{T} \log\det I_{nt}(\beta) + \frac{1}{T}\sum_{t=1}^{T} \log g\left(\hat{f}_{nt}(\beta) | \hat{f}_{n,t-1}(\beta); \theta\right), \tag{A.7}$$

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.8 in Appendix A.4.3 and Assumptions H.1, H.3 (i), H.6-H.12].

**Condition (2) of Lemma A.9:** Statement (i) follows from Assumptions A.6 and H.1. Statement (ii) follows from Assumptions A.8 and H.1, by using $\mathcal{L}_1(\beta_0, \theta) = E_0 [\log g(f_t | f_{t-1}; \theta)]$, up to a constant in $\theta$.

**Condition (3) of Lemma A.9:** From (A.6), 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. $\beta$, 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:

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

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

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

and

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

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.5, and condition (3) of Lemma A.2 is implied by Lemma A.6.

Moreover, from (A.7) we have:

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

and:

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

Thus, condition (3ii) is satisfied with $I_{1,\theta\theta} = E\left[-\frac{\partial^2 \log g}{\partial \theta \partial \theta'}(f_t|f_{t-1};\theta_0)\right]$ (use Lemmas A.6-A.7 and Assumption H.13).

**Condition (4) of Lemma A.9:** Let us first consider the approximated score w.r.t. $\beta$. We have:

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

By the mean-value Theorem:

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

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 $\hat{f}_{nt}(\beta_0)$:

$$\sqrt{n}\left(\hat{f}_{nt}(\beta_0) - f_t\right) = -\bar{I}_{ff}(t)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log h}{\partial f_t}(y_{i,t}|y_{i,t-1},f_t;\beta_0), \qquad \text{(A.8)}$$

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

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

Then, we get:

$$\sqrt{nT}\frac{\partial \mathcal{L}_{nT}^{*}\left(\beta_0\right)}{\partial \beta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\psi_\beta(t) - I_{\beta f}(t) I_{ff}(t)^{-1}\psi_f(t)\right] + o_p(1), \tag{A.9}$$

where:

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

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

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

By using $T^b/n = O(1)$, $b > 1$, Assumption H.13 and Lemmas A.6 and A.8, it follows that:

$$\sqrt{T}\frac{\partial \mathcal{L}_{1,nT}\left(\beta_0,\theta_0\right)}{\partial \theta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g}{\partial \theta}\left(f_t|f_{t-1};\theta_0\right) + o_p(1). \tag{A.10}$$

Thus, from (A.9) and (A.10) we deduce:

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

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.9:** Condition (i) is implied by Proposition 1.

From Lemma A.9 we deduce the efficiency bound.

### A.6.3 Proof of Proposition 5

From Lemma A.9, it follows that $\sqrt{nT}\left(\tilde{\beta}_{nT} - \hat{\beta}_{nT}^{*}\right) = o_p(1)$. The conclusion follows.

### A.6.4 Proof of Proposition 6

We have:

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

where $\dot{\beta}_{nT}$ 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.8). Point ii) follows from Lemma A.6.

## APPENDIX 7

## Proof of Proposition 10

### A.7.1 Higher-order asymptotic expansion of the cross-sectional factor approximations

By conditioning on the factor path $(f_t)$, we can apply results on higher-order expansion of the ML estimator in the iid case. From Gouriéroux, Monfort (1995), Section 23.1.2, we get:

$$\hat{f}_{n,t}=f_t+\frac{1}{\sqrt{n}}A_{n,t}+\frac{1}{n}B_{n,t}+o_p(1/n), \tag{A.11}$$

where:

$$A_{n,t}=I_{ff}(t)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log h\left(y_{i,t}|y_{i,t-1},f_t\right)}{\partial f_t},$$

and:

$$\begin{aligned}B_{n,t} \quad = \quad & I_{ff}(t)^{-2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial^2 \log h\left(y_{i,t}|y_{i,t-1},f_t\right)}{\partial f_t^2}+I_{ff}(t)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log h\left(y_{i,t}|y_{i,t-1},f_t\right)}{\partial f_t}\right)\\ & +\frac{1}{2}I_{ff}(t)^{-3}K_3(t)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log h\left(y_{i,t}|y_{i,t-1},f_t\right)}{\partial f_t}\right)^2.\end{aligned}$$

Conditionally on the factor path, the statistics $A_{n,t}$, $t$ varying, are a martingale difference sequence with $E\left[A_{n,t}^2|\underline{f_t}\right]=I_{ff}(t)^{-1}$. Moreover:

$$E\left[B_{n,t}|\underline{f_t}\right]=I_{ff}(t)^{-2}\left[K_{1,2}(t)+\frac{1}{2}K_3(t)\right]=B(t), \text{ say.}$$

### A.7.2 Asymptotic expansion of the CSA estimator

The CSA log-likelihood function is $\mathcal{L}_{nT}^{CSA}(\theta) = \mathcal{L}_{nT}^* + \frac{1}{n}\mathcal{L}_{1,nT}(\theta)$, where $\mathcal{L}_{nT}^*$ is independent of $\theta$ and

$$\mathcal{L}_{1,nT}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log g\left(\hat{f}_{n,t}, \theta\right).$$

Thus, the first-order condition for the CSA estimator $\hat{\theta}_{nT} = \hat{\theta}_{nT}^{CSA}$ is:

$$\frac{1}{T}\sum_{t=1}^{T} \frac{\partial \log g}{\partial \theta}\left(\hat{f}_{n,t}, \hat{\theta}_{nT}\right) = 0.$$

Let us expand this first-order condition w.r.t. both $\hat{f}_{n,t}$ and $\hat{\theta}_{nT}$ up to $o_p(1/n)$, and replace expansion (A.11). Since $n \ll T^{3/2}$, it is enough to keep terms which are at most of second-order in $\left(\hat{f}_{n,t} - f_t\right)$ and $\left(\hat{\theta}_{nT} - \theta_0\right)$. We get:

$$\begin{aligned}
0 \;=\;& \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \log g\left(f_t, \theta_0\right)}{\partial \theta} + \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial \theta^2}\left(\hat{\theta}_{nT} - \theta_0\right) \\
&+ \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f}\left(\frac{1}{\sqrt{n}}A_{n,t} + \frac{1}{n}B_{n,t}\right) + \frac{1}{2T}\sum_{t=1}^{T} \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta^3}\left(\hat{\theta}_{nT} - \theta_0\right)^2 \\
&+ \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta^2 \partial f}\left(\hat{\theta}_{nT} - \theta_0\right)\left(\frac{1}{\sqrt{n}}A_{n,t} + \frac{1}{n}B_{n,t}\right) \\
&+ \frac{1}{2nT}\sum_{t=1}^{T} \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f^2}A_{n,t}^2 + o_p(1/n). \qquad \text{(A.12)}
\end{aligned}$$

Now, we use that:

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f}B_{n,t} + \frac{1}{2T}\sum_{t=1}^{T} \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f^2}A_{n,t}^2 \\
=\;& E\left[\frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f}E\left[B_{n,t}|\underline{f_t}\right]\right] + \frac{1}{2}E\left[\frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f^2}E\left[A_{n,t}^2|\underline{f_t}\right]\right] + o_p(1) \\
=\;& E\left[B(t)\frac{\partial^2 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f}\right] + \frac{1}{2}E\left[I_{ff}(t)^{-1}\frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta \partial f^2}\right] + o_p(1),
\end{aligned}$$

and:

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta^2 \partial f}\left(\hat{\theta}_{nT} - \theta_0\right)\left(\frac{1}{\sqrt{n}}A_{n,t} + \frac{1}{n}B_{n,t}\right) \\
=\;& \left[\frac{1}{T\sqrt{n}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta^2 \partial f}A_{n,t}\right) + \frac{1}{n\sqrt{T}}\left(\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^3 \log g\left(f_t, \theta_0\right)}{\partial \theta^2 \partial f}B_{n,t}\right)\right]\sqrt{T}\left(\hat{\theta}_{nT} - \theta_0\right) \\
=\;& O_p\left(\frac{1}{T\sqrt{n}}\right) + O_p\left(\frac{1}{n\sqrt{T}}\right) = o_p(1/n).
\end{aligned}$$

Thus, the expansion in (A.12) simplifies to:

$$
\begin{aligned}
0 &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta} + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta^2}\left(\hat{\theta}_{nT}-\theta_0\right) + \frac{1}{\sqrt{n}}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}A_{n,t} \\
&\quad + \frac{1}{2T}\sum_{t=1}^{T}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta^3}\left(\hat{\theta}_{nT}-\theta_0\right)^2 \\
&\quad + \frac{1}{n}E\left[B(t)\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}\right] + \frac{1}{2n}E\left[I_{ff}(t)^{-1}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f^2}\right] + o_p(1/n). \qquad \text{(A.13)}
\end{aligned}
$$

By multiplying (A.13) by $\sqrt{T}$ and solving for $\sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right)$, we get:

$$
\begin{aligned}
\sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right) &= \left(-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta^2}\right)^{-1}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right. \\
&\quad + \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}A_{n,t}\right) \\
&\quad + \frac{1}{2\sqrt{T}}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta^3}\right)\left[\sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right)\right]^2 \\
&\quad + \frac{\sqrt{T}}{n}E\left[B(t)\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}\right] + \frac{\sqrt{T}}{2n}E\left[I_{ff}(t)^{-1}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f^2}\right] \\
&\quad \left. + o_p(\sqrt{T}/n)\right\}. \qquad \text{(A.14)}
\end{aligned}
$$

Let us now expand the inverse matrix in the RHS:

$$
\begin{aligned}
\left(-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta^2}\right)^{-1} &= \left(I_{1,\theta\theta} + \frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[-\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta^2} - I_{1,\theta\theta}\right]\right)^{-1} \\
&= I_{1,\theta\theta}^{-1} - I_{1,\theta\theta}^{-2}\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[-\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta^2} - I_{1,\theta\theta}\right]\right) \\
&\quad + o_p(\sqrt{T}/n), \qquad \text{(A.15)}
\end{aligned}
$$

where we used $1/T = o(\sqrt{T}/n)$. By plugging (A.15) into (A.14), we get:

$$
\begin{aligned}
\sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right) &= I_{1,\theta\theta}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right) \\
&\quad - \frac{1}{\sqrt{T}}I_{1,\theta\theta}^{-2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[-\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta^2} - I_{1,\theta\theta}\right]\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right) \\
&\quad + \frac{1}{\sqrt{n}}I_{1,\theta\theta}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}A_{n,t}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{2\sqrt{T}} I_{1,\theta\theta}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta^3}\right)\left[\sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right)\right]^2 \\
&+\frac{\sqrt{T}}{n} I_{1,\theta\theta}^{-1}\left(E\left[B(t)\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}\right]+\frac{1}{2}E\left[I_{ff}(t)^{-1}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f^2}\right]\right) \\
&+o_p(\sqrt{T}/n).
\end{aligned}
$$

By iterating this expansion, it is seen that:

$$
\begin{aligned}
& \frac{1}{2\sqrt{T}} I_{1,\theta\theta}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta^3}\right)\left[\sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right)\right]^2 \\
= \; & \frac{1}{2\sqrt{T}} I_{1,\theta\theta}^{-3} E\left[\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta^3}\right]\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right)^2+o_p(\sqrt{T}/n).
\end{aligned}
$$

Thus, by replacing the expression of $B(t)$, we deduce the asymptotic expansion of the CSA estimator:

$$
\begin{aligned}
\sqrt{T}\left(\hat{\theta}_{nT}^{CSA}-\theta_0\right) \; = \; & I_{1,\theta\theta}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right) \\
&-\frac{1}{\sqrt{T}} I_{1,\theta\theta}^{-2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[-\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta^2}-I_{1,\theta\theta}\right]\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right) \\
&+\frac{1}{\sqrt{n}} I_{1,\theta\theta}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}A_{n,t}\right) \\
&+\frac{1}{2\sqrt{T}} I_{1,\theta\theta}^{-3} E\left[\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta^3}\right]\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right)^2 \\
&+\frac{\sqrt{T}}{n} I_{1,\theta\theta}^{-1}\left(E\left[I_{ff}(t)^{-2}\left(K_{1,2}(t)+\frac{1}{2}K_3(t)\right)\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}\right]\right. \\
&\left.+\frac{1}{2}E\left[I_{ff}(t)^{-1}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f^2}\right]\right)+o_p(\sqrt{T}/n). \qquad \text{(A.16)}
\end{aligned}
$$

### A.7.3 Asymptotic expansion of the GA estimator

The GA log-likelihood function is $\mathcal{L}_{nT}^{GA}(\theta)=\mathcal{L}_{nT}^{*}+\frac{1}{n}\mathcal{L}_{1,nT}(\theta)+\frac{1}{n^2}\mathcal{L}_{2,nT}(\theta)$, where [see Appendix 5, equation (A.5)]:

$$
\begin{aligned}
\mathcal{L}_{2,nT}(\theta) \; = \; & \frac{1}{2T}\sum_{t=1}^{T} I_{nt}^{-1}\frac{\partial^2 \log g}{\partial f^2}\left(\hat{f}_{n,t},\theta\right)+\frac{1}{2T}\sum_{t=1}^{T} I_{nt}^{-1}\left[\frac{\partial \log g}{\partial f}\left(\hat{f}_{n,t},\theta\right)\right]^2 \\
&+\frac{1}{2T}\sum_{t=1}^{T} I_{nt}^{-2}\frac{\partial \log g}{\partial f}\left(\hat{f}_{n,t},\theta\right)K_{3,nt}.
\end{aligned}
$$

From the first-order condition for the GA estimator, we deduce the asymptotic expansion:

$$
\begin{aligned}
\sqrt{T}\left(\hat{\theta}_{nT}^{GA}-\theta_0\right) &= I_{1,\theta\theta}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right) \\
&-\frac{1}{\sqrt{T}}I_{1,\theta\theta}^{-2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[-\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta^2}-I_{1,\theta\theta}\right]\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right) \\
&+\frac{1}{\sqrt{n}}I_{1,\theta\theta}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}A_{n,t}\right) \\
&+\frac{1}{2\sqrt{T}}I_{1,\theta\theta}^{-3}E\left[\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta^3}\right]\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial\theta}\right)^2 \\
&+\frac{\sqrt{T}}{n}I_{1,\theta\theta}^{-1}\left(E\left[I_{ff}(t)^{-2}\left(K_{1,2}(t)+K_3(t)\right)\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f}\right]\right. \\
&\left.+E\left[I_{ff}(t)^{-1}\frac{\partial \log g\left(f_t,\theta_0\right)}{\partial f}\frac{\partial^2 \log g\left(f_t,\theta_0\right)}{\partial f\partial\theta}\right]+E\left[I_{ff}(t)^{-1}\frac{\partial^3 \log g\left(f_t,\theta_0\right)}{\partial\theta\partial f^2}\right]\right) \\
&+o_p(\sqrt{T}/n). \qquad \text{(A.17)}
\end{aligned}
$$

Proposition 10 follows from (A.16) and (A.17).

## APPENDIX 8

## Proofs of Propositions 11 and 12

### A.8.1 Asymptotic expansion of the log-likelihood function (Proof of Proposition 11)

Let us write the density of the sample of observations $\underline{y_T}$ (given an initial condition) as:

$$
l(\underline{y_T};\theta) = \int\cdots\int l(\underline{y_T}|\underline{F_T};\theta)\prod_t g(F_t|F_{t-1};\theta)\prod_t dF_t, \qquad \text{(A.18)}
$$

where:

$$
\begin{aligned}
l(\underline{y_T}|\underline{F_T};\theta) &= \int\cdots\int \exp\left\{\sum_i\sum_t \log\tilde{h}(y_{i,t}|y_{i,t-1};a_t)+\sum_t \log l(a_t|F_t;\theta)\right\}\prod_t da_t \\
&\propto \frac{1}{(\det\Delta)^{T/2}}\int\cdots\int \exp\left\{\sum_i\sum_t \log\tilde{h}(y_{i,t}|y_{i,t-1};a_t)\right. \\
&\left.-\frac{1}{2}\sum_t\left(a_t-\alpha-\gamma F_t\right)'\Delta^{-1}\left(a_t-\alpha-\gamma F_t\right)\right\}\prod_t da_t.
\end{aligned}
$$

By applying the argument in the proof of Proposition 1 (see Appendix 5) to the conditional density $l(\underline{y_T}|\underline{F_T};\theta)$ and replacing it into (A.18), we get:

$$\begin{aligned} l(\underline{y_T};\theta) &\propto \prod_t \prod_i \tilde{h}(y_{i,t}|y_{i,t-1},\hat{a}_{n,t}) \\ &\quad \cdot \int \cdots \int \prod_t g(\hat{a}_{n,t}|F_t;\theta) \exp\left[\frac{T}{n}\Psi_{nT}(\theta)\right] \prod_t g(F_t|F_{t-1};\theta) \prod_t dF_t \\ &\propto \int \cdots \int \prod_t g(\hat{a}_{n,t}|F_t;\theta) \exp\left[\frac{T}{n}\Psi_{nT}(\theta)\right] \prod_t g(F_t|F_{t-1};\theta) \prod_t dF_t, \end{aligned} \tag{A.19}$$

where $\Psi_{nT}(\theta)$ is such that:

$$\exp\left[\frac{T}{n}\Psi_{nT}(\theta)\right] = E\left[\exp\left\{\sum_{t=1}^{T}\psi_{n,t}\left(\hat{a}_{n,t}+\frac{1}{\sqrt{n}}\hat{\Sigma}_{n,t}^{1/2}Z_t\right)\right\}\right], \tag{A.20}$$

the expectation is w.r.t. the standard Gaussian vector $(Z_1',\cdots,Z_T')'$, the matrix $\hat{\Sigma}_{n,t}$ is defined by:

$$\hat{\Sigma}_{n,t} = \left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log\tilde{h}}{\partial a_t \partial a_t'}(y_{i,t}|y_{i,t-1};\hat{a}_{n,t})\right)^{-1},$$

and the function $\psi_{n,t}(a_t)$ is given by:

$$\begin{aligned} \psi_{n,t}(a_t) &= \sum_i \log\tilde{h}(y_{i,t}|y_{i,t-1};a_t) - \sum_i \log\tilde{h}(y_{i,t}|y_{i,t-1};\hat{a}_{n,t}) + \frac{n}{2}(a_t-\hat{a}_{n,t})'\hat{\Sigma}_{n,t}^{-1}(a_t-\hat{a}_{n,t}) \\ &\quad -\frac{1}{2}\left[(a_t-\alpha-\gamma F_t)'\Delta^{-1}(a_t-\alpha-\gamma F_t) - (\hat{a}_{n,t}-\alpha-\gamma F_t)'\Delta^{-1}(\hat{a}_{n,t}-\alpha-\gamma F_t)\right]. \end{aligned}$$

Let us now compute the term $\Psi_{nT}(\theta)$. By expanding $\psi_{n,t}\left(\hat{a}_{n,t}+\frac{1}{\sqrt{n}}\hat{\Sigma}_{n,t}^{1/2}Z_t\right)$ at order $1/n$ we have:

$$\begin{aligned} \psi_{n,t}\left(\hat{a}_{n,t}+\frac{1}{\sqrt{n}}\hat{\Sigma}_{n,t}^{1/2}Z_t\right) &= \frac{1}{6\sqrt{n}}\sum_{l,p,q=1}^{m}\hat{K}_{n,t}(l,p,q)\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_l\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_p\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_q \\ &\quad +\frac{1}{24n}\sum_{l,p,q,r=1}^{m}\hat{K}_{n,t}(l,p,q,r)\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_l\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_p\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_q\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_r \\ &\quad -\frac{1}{\sqrt{n}}(\hat{a}_{n,t}-\alpha-\gamma F_t)'\Delta^{-1}\hat{\Sigma}_{n,t}^{1/2}Z_t - \frac{1}{2n}Z_t'\hat{\Sigma}_{n,t}^{1/2}\Delta^{-1}\hat{\Sigma}_{n,t}^{1/2}Z_t + o_p(1/n), \end{aligned}$$

where:

$$\begin{aligned} \hat{K}_{n,t}(l,p,q) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3\log\tilde{h}}{\partial a_{l,t}\partial a_{p,t}\partial a_{q,t}}(y_{i,t}|y_{i,t-1};\hat{a}_{n,t}), \\ \hat{K}_{n,t}(l,p,q,r) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^4\log\tilde{h}}{\partial a_{l,t}\partial a_{p,t}\partial a_{q,t}\partial a_{r,t}}(y_{i,t}|y_{i,t-1};\hat{a}_{n,t}). \end{aligned}$$

By expanding the exponential function and computing the Gaussian integral in (A.20), we get:

$$\exp\left[\frac{T}{n}\Psi_{nT}(\theta)\right] = 1 + \frac{1}{2n}\sum_{t=1}^{T}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)'\Delta^{-1}\hat{\Sigma}_{n,t}\Delta^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)$$

$$-\frac{1}{2n}\sum_{t=1}^{T}Tr\left(\hat{\Sigma}_{n,t}\Delta^{-1}\right) - \frac{1}{6n}\sum_{t=1}^{T}\sum_{l,p,q,r=1}^{m}\hat{K}_{n,t}\left(l,p,q\right)$$

$$\cdot E\left[\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_l\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_p\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_q\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_r\right]\left[\Delta^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)\right]_r + \frac{T}{n}B_{nT} + o_p(T/n),$$

where $B_{nT}$ is independent of both the factor values $F_t$ and the parameter $\theta$, and:

$$\begin{aligned} E\left[\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_l\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_p\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_q\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_r\right] &= \frac{\partial^4}{\partial\lambda_l\partial\lambda_p\partial\lambda_q\partial\lambda_r}\exp\left(\frac{1}{2}\lambda'\hat{\Sigma}_{n,t}\lambda\right)\Big|_{\lambda=0} \\ &= \hat{\Sigma}_{n,t,lp}\hat{\Sigma}_{n,t,qr} + \hat{\Sigma}_{n,t,lq}\hat{\Sigma}_{n,t,pr} + \hat{\Sigma}_{n,t,lr}\hat{\Sigma}_{n,t,pq}. \end{aligned}$$

Then, by using the symmetry of $\hat{K}_{n,t}\left(l,p,q\right)$ w.r.t. $l,p,q$, we get:

$$\begin{aligned} &\sum_{l,p,q,r=1}^{m}\hat{K}_{n,t}\left(l,p,q\right)E\left[\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_l\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_p\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_q\left(\hat{\Sigma}_{n,t}^{1/2}Z_t\right)_r\right]\left[\Delta^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)\right]_r \\ = \; &3\hat{\Lambda}_{n,t}'\Delta^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right), \end{aligned}$$

where $\hat{\Lambda}_{n,t}$ is a $(m,1)$ vector with elements:

$$\hat{\Lambda}_{n,t,r} = \sum_{l,p,q=1}^{m}\hat{K}_{n,t}\left(l,p,q\right)\hat{\Sigma}_{n,t,lp}\hat{\Sigma}_{n,t,qr}, \quad r=1,...,m.$$

Thus, we get:

$$\begin{aligned} \exp\left[\frac{T}{n}\Psi_{nT}(\theta)\right] &= 1 + \frac{T}{n}A_{nT}(\theta) + \frac{T}{n}B_{nT} + o_p(T/n), && \text{(A.21)} \\ &= \exp\left(\frac{T}{n}A_{nT}(\theta) + \frac{T}{n}B_{nT} + o_p(T/n)\right), && \text{(A.22)} \end{aligned}$$

where:

$$\begin{aligned} A_{nT}(\theta) = \; &\frac{1}{2T}\sum_{t=1}^{T}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)'\Delta^{-1}\hat{\Sigma}_{n,t}\Delta^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right) \\ &-\frac{1}{2T}\sum_{t=1}^{T}\hat{\Lambda}_{n,t}'\Delta^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right) - \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}Tr\left(\hat{\Sigma}_{n,t}\Delta^{-1}\right). \end{aligned}$$

By replacing (A.22) into (A.19), it is seen that the term $B_{nT}$ is irrelevant for maximization w.r.t. $\theta$. By replacing (A.21) into (A.19), we get:

$$
\begin{aligned}
l(\underline{y_T};\theta) \quad \propto \quad & \int \cdots \int \prod_t g(\hat{a}_{n,t}|F_t;\theta) \prod_t g(F_t|F_{t-1};\theta) \prod_t dF_t \\
& + \frac{T}{n} \int \cdots \int \prod_t g(\hat{a}_{n,t}|F_t;\theta) A_{nT}(\theta) \prod_t g(F_t|F_{t-1};\theta) \prod_t dF_t + o_p(T/n).
\end{aligned}
$$

Thus, the ($nT$-standardized) log-likelihood function $\mathcal{L}_{nT}(\theta)$ is such that:

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

where $\mathcal{L}^*_{nT}$ is constant in $\theta$,

$$
\begin{aligned}
& \mathcal{L}_{1,nT}(\theta) = \frac{1}{T}\log\left(\int \cdots \int \prod_t g(\hat{a}_{n,t}|F_t;\theta) \prod_t g(F_t|F_{t-1};\theta) \prod_t dF_t\right) \\
& = \frac{1}{T}\log\left(\frac{1}{[(2\pi)^{m+J}(\det\Delta)(\det\Omega)]^{T/2}} \int \cdots \int \exp\left\{-\frac{1}{2}\sum_t \left(\hat{a}_{n,t} - \alpha - \gamma F_t\right)' \Delta^{-1}\right.\right. \\
& \left.\left. \cdot \left(\hat{a}_{n,t} - \alpha - \gamma F_t\right) - \frac{1}{2}\sum_t \left(F_t - \mu - \Phi F_{t-1}\right)' \Omega^{-1} \left(F_t - \mu - \Phi F_{t-1}\right)\right\} \prod_t dF_t\right),
\end{aligned}
$$

and:

$$
\mathcal{L}_{2,nT}(\theta) = \frac{\displaystyle\int \cdots \int \prod_t g(\hat{a}_{n,t}|F_t;\theta) A_{nT}(\theta) \prod_t g(F_t|F_{t-1};\theta) \prod_t dF_t}{\displaystyle\int \cdots \int \prod_t g(\hat{a}_{n,t}|F_t;\theta) \prod_t g(F_t|F_{t-1};\theta) \prod_t dF_t}.
$$

By applying Lemma A.9 (Appendix 6), parameter $\theta$ is estimated at rate $1/\sqrt{T}$ and an efficient estimator is obtained by maximizing the criterion function $\mathcal{L}_{1,nT}(\theta)$. Proposition 11 follows.

### A.8.2 Granularity adjustment (Proof of Proposition 12)

By replacing (A.22) into (A.19) we get:

$$
\begin{aligned}
& l(\underline{y_T};\beta,\theta) \\
\propto \quad & \frac{1}{(\det\Delta)^{T/2}} \int \cdots \int \exp\left\{-\frac{1}{2}\sum_t \left(\hat{a}_{n,t} - \alpha - \gamma F_t\right)' \Delta^{-1} \left(I_m - \frac{1}{n}\hat{\Sigma}_{n,t}\Delta^{-1}\right) \left(\hat{a}_{n,t} - \alpha - \gamma F_t\right)\right. \\
& \left. -\frac{1}{2n}\sum_t \hat{\Lambda}'_{n,t}\Delta^{-1}\left(\hat{a}_{n,t} - \alpha - \gamma F_t\right) - \frac{1}{2n}\sum_t Tr\left(\hat{\Sigma}_{n,t}\Delta^{-1}\right)\right\} \\
& \cdot \prod_t g(F_t|F_{t-1};\theta) \prod_t dF_t \cdot (1 + o_p(T/n)).
\end{aligned}
$$

Now, since:

$$-\frac{1}{2}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)'\Delta^{-1}\left(I_m-\frac{1}{n}\hat{\Sigma}_{n,t}\Delta^{-1}\right)\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)-\frac{1}{2n}\hat{\Lambda}'_{n,t}\Delta^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t\right)$$
$$=-\frac{1}{2}\left(\hat{a}_{n,t}-\alpha-\gamma F_t+\frac{1}{n}\hat{\xi}_{n,t}\right)'\Psi_{n,t}^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t+\frac{1}{n}\hat{\xi}_{n,t}\right)+o_p(1/n),$$

where:

$$\hat{\Psi}_{n,t}=\Delta+\frac{1}{n}\hat{\Sigma}_{n,t}\ \text{ and }\hat{\xi}_{n,t}=\frac{1}{2}\hat{\Lambda}_{n,t},$$

and:

$$\begin{aligned}\det\Psi_{n,t} &= \det(\Delta)\det\left(I_m+\frac{1}{n}\hat{\Sigma}_{n,t}\Delta^{-1}\right)=\det(\Delta)\left(1+\frac{1}{n}Tr\left(\hat{\Sigma}_{n,t}\Delta^{-1}\right)+O_p(1/n^2)\right)\\ &= \det(\Delta)\exp\left(\frac{1}{n}Tr\left(\hat{\Sigma}_{n,t}\Delta^{-1}\right)+o_p(1/n)\right),\end{aligned}$$

we get:

$$\begin{aligned}l(\underline{y_T};\beta,\theta) \;\propto\; & \frac{1}{\prod_t\left(\det\Psi_{n,t}\right)^{1/2}}\\ & \cdot\int\cdots\int\exp\left\{-\frac{1}{2}\sum_t\left(\hat{a}_{n,t}-\alpha-\gamma F_t+\frac{1}{n}\hat{\xi}_{n,t}\right)'\Psi_{n,t}^{-1}\left(\hat{a}_{n,t}-\alpha-\gamma F_t+\frac{1}{n}\hat{\xi}_{n,t}\right)\right\}\\ & \prod_t g(F_t|F_{t-1};\theta)\prod_t dF_t\cdot(1+o_p(T/n)).\end{aligned}$$

Thus, we get equation (5.7), and the conclusion follows.

# APPENDIX 9

## Factor ordered qualitative model

### A.9.1. Identification

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

$$\pi_t=\left[\begin{array}{cc} G\left(\frac{c_1-\gamma_1 f_t-\alpha_1}{\sigma_1}\right) & 1-G\left(\frac{c_1-\gamma_1 f_t-\alpha_1}{\sigma_1}\right)\\ G\left(\frac{c_1-\gamma_2 f_t-\alpha_2}{\sigma_2}\right) & 1-G\left(\frac{c_1-\gamma_2 f_t-\alpha_2}{\sigma_2}\right)\end{array}\right].$$

By reparametrizing coefficients $\alpha_1$ and $\alpha_2$, we can assume $c_1 = 0$. The transition matrix becomes:

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

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

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

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

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

Then, the values of the factor $f_t$ are identified by the first row of the transition matrix, $t = 1, ..., T$. The values of $\gamma_2, \alpha_2$ 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:

$$\left[G\left(\frac{c_1 - \gamma_l f_t - \alpha_l}{\sigma_l}\right), G\left(\frac{c_2 - \gamma_l f_t - \alpha_l}{\sigma_l}\right) - G\left(\frac{c_1 - \gamma_l f_t - \alpha_l}{\sigma_l}\right), ..., 1 - G\left(\frac{c_{K-1} - \gamma_l f_t - \alpha_l}{\sigma_l}\right)\right],$$

for $l = 1, ..., K$. As above, we can first set $c_1 = 0$:

$$\left[G\left(-\frac{\gamma_l f_t + \alpha_l}{\sigma_l}\right), G\left(\frac{c_2 - \gamma_l f_t - \alpha_l}{\sigma_l}\right) - G\left(-\frac{\gamma_l f_t + \alpha_l}{\sigma_l}\right), ..., 1 - G\left(\frac{c_{K-1} - \gamma_l f_t - \alpha_l}{\sigma_l}\right)\right]. \tag{A.23}$$

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

$$\left[G\left(-f_t\right), G\left(c_2 - f_t\right) - G\left(-f_t\right), ..., 1 - G\left(c_{K-1} - f_t\right)\right],$$

and row $l$ is given by (A.23) for $l \geq 2$. From the first row, we can identify the factor value $f_t$ and the $K-2$ thresholds $c_2, ..., c_K$. Then, the values of $\gamma_l, \alpha_l, \sigma_l$ are identified by the row $l$, for $l = 2, ..., K$, when $(K-1)T \geq 3$.

### A.9.2 Semi-parametric efficiency bound [Proof of Equation (6.4)]

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

$$I_{l,t}=\sum_{k=1}^{K}J_{lk,t}=\sum_{k=1}^{K}\frac{1}{\pi_{lk,t}}\frac{\partial\pi_{lk,t}}{\partial\left(\beta',f_t'\right)'}\frac{\partial\pi_{lk,t}}{\partial\left(\beta',f_t'\right)}. \tag{A.24}$$

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 $\beta=(\gamma_2,\alpha_2)'$ and

$$\Pi_t=\begin{pmatrix}1-\Lambda\left(f_t\right)&\Lambda\left(f_t\right)\\1-\Lambda\left(\beta'x_t\right)&\Lambda\left(\beta'x_t\right)\end{pmatrix}, \tag{A.25}$$

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 $\xi_{2,t} = (f_t, 1, \gamma_2)^{'}$. Thus, we have:

$$\begin{aligned} I_{\beta\beta}(t) &= \mu_{2,t-1}\pi_{22,t}\left(1-\pi_{22,t}\right)\begin{pmatrix} f_t^2 & f_t \\ f_t & 1 \end{pmatrix} \quad , \quad I_{\beta f}(t) = \mu_{2,t-1}\pi_{22,t}\left(1-\pi_{22,t}\right)\gamma_2 \begin{pmatrix} f_t \\ 1 \end{pmatrix}, \\ I_{ff}(t) &= \mu_{1,t-1}\pi_{12,t}\left(1-\pi_{12,t}\right) + \mu_{2,t-1}\pi_{22,t}\left(1-\pi_{22,t}\right)\gamma_2^2. \end{aligned}$$

We deduce formula (6.4).

### A.9.3 Numerical computation of the semi-parametric efficiency bound

The semi-parametric efficiency bound $(I_0^*)^{-1}$ 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.24) and (A.25), 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 $h$.