EFFICIENCY IN LARGE DYNAMIC PANEL MODELS WITH COMMON FACTORS

Patrick, GAGLIARDINI[(1)] and Christian, GOURIEROUX[(2)]

First version: August 2008
This version: October 2012

[1]Università della Svizzera Italiana. Corresponding author: Patrick Gagliardini, Università della Svizzera Italiana, Faculty of Economics, Via Buffi 13, CH-6900 Lugano, Switzerland. Email: patrick.gagliardini@usi.ch.

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

Acknowledgements: We thank a co-editor, anonymous referees, A. Monfort, F. Peracchi, E. Renault, M. Rockinger, O. Scaillet and participants at the CEMMAP 2008 Workshop on Unobserved Factor Models in London, the ESEM 2009 in Milan, the 2009 Financial Econometrics Conference in Toulouse, the SoFiE 2009 Conference in Geneva, the Journées de Statistique 2009 in Bordeaux, the 2009 Panel Data Conference in Bonn, the FINRISK 2009 Research Workshop in Gerzensee, the Triangle Conference 2010 and seminars at Toronto University, CEMFI, Einaudi Institute for Economics and Finance and Carlos III University 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 Factors

Abstract

This paper deals with asymptotically efficient estimation in exchangeable nonlinear dynamic panel models with common unobservable factors. These models are especially relevant for applications to large portfolios of credits, corporate bonds, or life insurance contracts. For instance, the special case of an Asymptotic Risk Factor (ARF) model is recommended in the current regulation in Finance (Basel II and Basel III) and Insurance (Solvency II) for risk prediction and computation of the required capital. The specification accounts for both micro- and macro-dynamics, induced by the lagged individual observations and the common stochastic factors, 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. The results are based on an asymptotic expansion of the log-likelihood function in powers of $1/n$, and are linked to granularity theory. The results are illustrated with the stochastic migration model for credit risk analysis.

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

JEL classification: C23, C13, G12.

1 Introduction

This paper considers the asymptotically efficient estimation of nonlinear dynamic panel models with common unobservable factors. 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 often 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 factors represent latent macro-variables, such as the sector and country specific business cycle, that introduce dependence across the nonlinear 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. The panel data may also correspond to other risk characteristics in a pool of corporate loans, household mortgages or life insurance contracts, such as prepayment, lapse, or mortality.

The model considered in this paper involves both micro- and macro-dynamics. Conditional on a given factor path, the individuals are assumed independent and identically distributed, with the histories of observations $y_{i,t}$, t varying, following the same time-inhomogeneous Markov process for any individual i. The transition density $h(y_{i,t}|y_{i,t-1}, f_t; \beta)$ between dates $t-1$ and t depends on the (multivariate) factor value f_t and the unknown parameter β. The micro-dynamics is captured by the lagged individual observation $y_{i,t-1}$ and unknown parameter β. The macro-dynamics is driven by the time-varying stochastic common factor f_t. The latter follows a Markov process with transition density $g(f_t|f_{t-1}; \theta)$, which depends on the unknown parameter θ. In credit risk applications, the common factor f_t has to be considered unobservable in order to account for systematic risk. 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 real-valued or discrete (as for default and rating histories in the credit risk application), while the components of the vector f_t are real valued (corresponding to a continuum of latent states). The model is potentially nonlinear in both micro- and macro-dynamics.

When the cross-sectional dimension n is fixed and the time dimension T tends to infinity, the Maximum Likelihood (ML) estimators of micro-parameter β and macro-parameter θ are asymptotically normal and efficient [1]. 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. In applications to mortgage or life insurance, we typically have $n \simeq 100,000 - 1,000,000$ contracts and $T \simeq 200$ months. Moreover, the numerical computation of the ML estimate 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 β and the macro-parameter θ, and to introduce asymptotically efficient estimators of β and θ that are easier to compute than the ML estimator. We consider the double asymptotics $n, T \to \infty$, such that $T^\nu/n = O(1)$, with either $\nu > 1$, for estimators maximizing a first-order expansion of the log-likelihood function w.r.t. $1/n$, or $\nu > 3/2$, for estimators maximizing a more accurate second-order expansion. We summarize our theoretical contributions as follows. First, we show that the efficiency bound for the micro-parameter β does not depend on the parametric model defining the macro-dynamics. In particular, this bound coincides with the parametric 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, the efficiency bound for the macro-parameter θ is the same as if the factor values were observable. These findings correspond to oracle properties w.r.t. the factor dynamics for the micro-parameter, and w.r.t. the factor values for the macro-parameter. Third, the asymptotic efficiency bound can be reached by optimizing approximated likelihood functions which do not involve integrals w.r.t. the factor path.

In Section 2 we introduce the nonlinear dynamic panel model with common factors. To provide motivation and grounding on potential applications, we first describe the Asymptotic Single Risk Factor (ASRF) model, which is the simplest benchmark model suggested for the regulation of credit risk in Basel II [BCBS (2001), (2003)]. Then, we present the general specification and discuss the stationarity and ergodicity assumptions needed for the asymptotic analysis. Our theoretical results are mainly based on a second-order asymptotic expansion of the log-likelihood function in powers of $1/n$ given in Section 3. The basic idea behind this expansion is that the integration of the latent factor path is performed along the lines of the Laplace approximation. In Section 4 we

[1] See e.g. Douc, Moulines, Rydèn (2004) for the asymptotic properties of the ML estimator in autoregressive models with Markov regimes.

introduce estimators of both micro- and macro-parameters, that do not involve numerical integration w.r.t. the unobservable factor. These estimators are obtained by maximizing approximations of the log-likelihood function at order $1/n$, and $1/n^2$, respectively. They are called Cross-Sectional Asymptotic (CSA) and Granularity Adjusted (GA) maximum likelihood estimators, respectively. We study the asymptotic properties of these estimators under suitable identification conditions and prove their asymptotic efficiency. In Section 5 we introduce an asymptotically efficient estimation approach, in which the estimators of the micro- and macro-parameters can be computed in two steps. The estimator of the micro-component is a fixed effects estimator, which considers the factor values as nuisance parameters. The estimator of the macro-parameter is obtained by maximizing the likelihood function of the macro-dynamics, in which the unobservable factor values are replaced by suitable cross-sectional factor approximations. 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 factors account for migration correlation. The patterns of the efficiency bound and the computation of the efficient estimators are illustrated for this example. We also investigate the finite-sample properties of the estimators in a Monte-Carlo experiment. Section 7 concludes. Appendix A.1 provides the regularity conditions for the large sample properties of the estimators. The proofs of the results are gathered in Appendices A.2 and A.3. The proofs rely on some Limit Theorems for uniform stochastic convergence with panel data and technical Lemmas. The details of these Theorems and Lemmas are provided online at Cambridge Journals Online in supplementary material to this article. Readers may refer to the supplementary material associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

2 Exchangeable nonlinear panel model with common factors

Exchangeable nonlinear panel models with common factors are the basis for risk analysis of homogenous retail portfolios encountered in Finance and Insurance. Before describing the general specification, we review as an illustration the Asymptotic Single Risk Factor (ASRF) model introduced for default risk analysis by Vasicek (1987), (1991).

2.1 The Asymptotic (Single) Risk Factor (ASRF) model for default

The general specification considered in Section 2.2 is motivated by the ASRF model introduced by Vasicek (1987), (1991) and based on the Value of the Firm model [Merton (1974)]. This model, possibly extended to include more factors, 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 may contain several millions of individual contracts (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 ASRF model is applied to these homogeneous subportfolios separately (or jointly), with parameters and factors which can depend on the segment. The sizes of these subportfolios may still be rather large including some ten 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. Let us consider a given subpopulation and a single-factor model. 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 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 latter component is introduced to represent the risk dependence. It is especially important for financial stability analysis. Indeed, the standard stress testing methodology corresponds to assessing the impact of extreme shocks on some components of the systematic risk factor. The coefficients α, γ, σ are independent of the individuals, according to the definition of an homogeneous portfolio. The parameters and factors depend on the segment, but the index of the segment is omitted for expository purpose [2]. The

[2] When the subpopulation index k, with $k = 1, ..., K$, is introduced explicitly, the variables are triply indexed by k, i, t, and the latent model becomes $y^*_{k,i,t} = \alpha_k + \gamma_k F_{k,t} + \sigma_k u_{k,i,t}$, where $k = 1, ..., K$, $i \in PaR_{k,t}$ and $t = 1, ..., T$. The subpopulations fixed effects are α_k, γ_k, σ_k and the model allows for a crossing of fixed effects γ_k

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:

$$PD_t = \mathbb{P}\left[y_{i,t} = 1 | y_{i,t-1} = 0, F_t\right] = \Phi\left[-\left(\alpha/\sigma\right) - \left(\gamma/\sigma\right) F_t\right], \tag{2.1}$$

where Φ 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. To summarize, the qualitative observations $y_{i,t}$ are independent conditional on the factor path with Bernoulli distribution:

$$y_{i,t} | F_t \sim \mathcal{B}(1, PD_t). \tag{2.2}$$

We get a probit model in which the explanatory variable F_t is unobservable and captures the systematic default risk. This basic static model can be extended by allowing for several factors in the given subpopulation, for a dynamics of the common factors [e.g., Duffie, Singleton (1998), Loeffler (2003), Dembo, Deuschel, Duffie (2004), McNeil, Wendin (2007), Duffie et al. (2009)], 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 and references therein).

The unconditional probability of default is $PD = \mathbb{P}[y_{i,t} = 1] = \Phi\left(-\alpha/\sqrt{\gamma^2 + \sigma^2}\right)$, whereas the unconditional default correlation between any two firms i and j is:

$$\rho = Corr\left(y_{i,t}, y_{j,t}\right) = \frac{\Psi\left(-\alpha/\sqrt{\gamma^2+\sigma^2}, -\alpha/\sqrt{\gamma^2+\sigma^2}; \rho^*\right) - PD^2}{PD(1-PD)}, \tag{2.3}$$

where $\rho^* = \gamma^2/\left(\gamma^2 + \sigma^2\right)$ is the asset correlation, that is the correlation between the log asset/liability ratios of any two firms, and $\Psi(.,.;\rho^*)$ denotes the joint cdf of the bivariate standard Gaussian distribution with correlation coefficient ρ^*. In the new regulation for credit risk, the required capital depends on the values of PD and ρ^*, that is, indirectly on the values α/σ and γ/σ, and is especially sensitive to the asset correlation parameter ρ^*. In the standard implementation of the above risk factor model, the unknown parameters PD and ρ^* are replaced by their empirical counterparts, which are close to the true values when the subpopulation sizes are large. This

with time stochastic effects $F_{k,t}$. Moreover, we get a joint multi-factor model, whenever the factors $F_{k,t}$ are different among classes.

explains the term "asymptotic" appearing in the usual methodology. However, it is important to check if not only the consistency, but also the efficiency can be reached by computationally simple estimators of the structural parameters [3].

The above ASRF model assumes that the individual fixed effects depend on the segment only, that is, the individual fixed effects α_i, γ_i, σ_i, say, are identical for two individuals in a same segment. This model assumption is compatible with the two-step approach considered in credit risk applications. First, models with individual fixed effects are used to get the homogeneous subportfolios; then the ASRF model is written for each homogeneous subportfolio to derive the distribution of the future portfolio value and the corresponding 1% quantile, called CreditVaR. Such a two-step procedure has been preferred in the current regulation for at least the following reasons: First, in the standard regulation approach that applies for the banks with the least advanced risk management systems, a common segmentation can be proposed by the regulator itself. Thus, the risk analysis is performed by the banks with a same segmentation, which facilitates the aggregation of bank portfolios when analyzing the global risk of the system. Second, and more importantly, the introduction of several millions of individual fixed effects beyond segment effects would diminish the estimated magnitude of idiosyncratic risks. In a regulatory perspective, this would yield a significantly lower level of required capital. Indeed, the reserves for credit risk are typically computed with unknown parameters directly replaced by their estimates. [4] Finally, models without individual fixed effects are common in the credit risk literature on bankruptcy prediction [e.g., Shumway (2001), Chava, Jarrow (2004), Campbell, Hilscher, Szilagyi (2008)], where individual heterogeneity is accounted for by observable characteristics. Duffie et al. (2009) estimate their model on US corporate default data and find that the inclusion of individual fixed effects does not lead to a significant improvement of the results.

[3] The underestimation of the asset correlation parameter in 2007-2008 played a key role in the underpricing of Collateralized Debt Obligations (CDO) contracts and lead to severe losses during the recent subprime crisis.

[4] In Basel II regulation, the lack of accuracy on estimated model parameters might be taken into account by means of reserves for estimation risk. However, in the current implementation, these reserves are usually set to zero. Moreover, the updating of the estimated individual fixed effects would induce a large volatility of the required capital for credit risk, with undesirable effects on financial market stability.

2.2 The general specification

The basic ASRF model can be extended to include any number of factors and any type of parametric nonlinear dynamics. This extended model is introduced in this section. Let us consider panel data $y_{i,t}$ for a large homogeneous population of individuals $i = 1, ..., n$ observed at dates $t = 1, ..., T$. We assume that there exists a common (multidimensional) factor such that [5]:

A.1: *Conditional on the factor path* (f_t)*, the individual histories* $(y_{i,t}, t = 1, 2, \cdots)$*, for* i *varying, 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 an exogenous Markov process of order 1 in* $\mathbb{R}^m$*, that is, the conditional distribution of* f_t *given the past of the factor* $\underline{f_{t-1}} = (f_{t-1}, f_{t-2}, ...)$ *and of the individual histories* $\underline{y_{i,t-1}} = (y_{i,t-1}, y_{i,t-2}, ...)$*,* $i = 1, ..., n$*, depends on* f_{t-1} *only, with transition pdf* $g(f_t|f_{t-1}; \theta)$ *and unknown parameter* $\theta \in \Theta$*, where* $\Theta \subset \mathbb{R}^p$*.*

We denote by β_0 and θ_0 the true values of parameters β and θ, respectively. Factor f_t is assumed unobservable. [6] Thus, it has to be integrated out to derive the joint density of observations $y_{i,t}$. The latent factor introduces both non-Markovian individual dynamics and dependence across individuals. The exogeneity assumption means that: (i) there is no feedback from one specific individual history on the future factor values; and (ii) the lagged factor value includes all informative macro-summaries of the past. The distribution of the individual histories $(y_{i,t})$ 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)] [7]. Such exchangeability

[5] In an unobservable factor model, the factor process is usually defined up to some nonlinear dynamic transformation. Assumptions A.1-A.2 have to be satisfied for an appropriate choice of factor f_t. As a consequence the Markov assumption on factor f_t is rather mild. For instance, let us consider a dynamic model with a factor f_t satisfying Assumption A.1 and admitting a nonlinear moving average representation $f_t = a(\varepsilon_t, \varepsilon_{t-1}; \theta)$, say, with $\varepsilon_t \sim IIN(0, 1)$. Then Assumptions A.1-A.2 are satisfied with f_t replaced by $f_t^* = (\varepsilon_t, \varepsilon_{t-1})'$ and $h(y_{i,t}|y_{i,t-1}, f_t; \beta)$ replaced by $h^*(y_{i,t}|y_{i,t-1}, f_t^*; \beta^*) = h(y_{i,t}|y_{i,t-1}, a(\varepsilon_t, \varepsilon_{t-1}; \theta); \beta)$, where $\beta^* = (\beta', \theta')'$.

[6] As in the ASRF model for default, we can introduce explicitly the fixed effects of the segments, that is, the factors $f_{k,t}$ can differ among the segments and parameters β_k, θ_k can depend on k, with $k = 1, ..., K$.

[7] 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)]. Here, we assume that the sigma-field $\mathcal{F}$ is generated by the

assumptions have been introduced in the literature on linear dynamics [see e.g. Andrews (2005) and Hjellwig, Tjostheim (1999)]. The focus of our paper is on the efficient estimation of both micro-parameter β and macro-parameter θ in the nonlinear exchangeable panel model A.1-A.2.

Without Assumption A.2 on the parametric factor dynamics, the model introduced in Assumption A.1 might be seen as a model with time fixed effects instead of individual fixed effects. Thus, we might expect to derive the asymptotic results from the nonlinear panel literature with individual fixed effects by simply interchanging the roles of individual and time indices i and t, and the sizes n and T [see e.g. Hahn, Newey (2004) for estimation of nonlinear panel models with fixed individual effects]. However, this intuition is not correct since there are important differences between our setting and the ones considered by the individual fixed effects panel literature:

i) In applications to credit risk the size n of the segment is much larger than the number T of dates, and therefore the incidental parameter problem [see Neyman, Scott (1948) for the pioneering paper and Lancaster (2000) for a review] is much less pronounced with time fixed effects than with individual fixed effects. In particular, bias corrections in the first-order asymptotic distributions are not required in our setting since we assume $T/n \to 0$.

ii) Assumption A.2 shows that the nonlinear panel model with common factor is a time series model introduced for prediction purpose. This fact is illustrated in Section 2.1 on default risk analysis, in which the final aim is the computation of reserves by means of a quantile of the conditional distribution of the future portfolio value, that is, the CreditVaR. Therefore, we are interested not only in the micro-parameter β, but also in the macro-parameter θ.

iii) The parametric Assumption A.2 on the factor dynamics provides additional information, which might allow for a more efficient estimation of the micro-parameter β.

To establish the large sample properties of the estimators, we introduce the next Assumptions A.3, A.4 and A.5. Assumptions A.3 and A.4 concern the stationarity and mixing properties of the factor process, and of the individual histories conditional on the factor process, respectively.

A.3: *The process (f_t) is strictly stationary and geometrically strong mixing, that is, $\alpha(s) = O(\rho^s)$ as $s \to \infty$, for some $\rho \in (0,1)$, where $\alpha(s) = \sup_{A \in \mathcal{H}_{-\infty}^t, B \in \mathcal{H}_{t+s}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ denotes the alpha mixing coefficient at lag $s \in \mathbb{N}$, and $\mathcal{H}_{-\infty}^t = \sigma(f_t, f_{t-1}, ...)$ and $\mathcal{H}_{t+s}^\infty = \sigma(f_{t+s}, f_{t+s+1}, ...)$ denote the sigma-fields generated by process (f_t) up to time t, and from time $t+s$ onward, respec-*

finite-dimensional Markov process (f_t).

tively.

A.4: *Conditional on the factor path (f_t), the individual process $(y_{i,t})$ is beta mixing, such that the conditional beta mixing coefficients:*

$$\beta_t(s) \equiv \sup_{A \in \mathscr{B}(\mathbb{R})} \int \left| \mathbb{P}\left[y_{i,t} \in A | y_{i,t-s} = \eta, f_t, f_{t-1}, ..., f_{t-s+1}\right] - \mathbb{P}\left[y_{i,t} \in A | \underline{f_t}\right] \right| \lambda(\eta) d\eta, \quad s \in \mathbb{N},$$

are measurable functions of $\underline{f_t}$ and satisfy $\beta_t(s) \to 0$ as $s \to \infty$, for any t and $\mathbb{P}$-a.s., where $\mathscr{B}(\mathbb{R})$ denotes the Borel sigma-field on $\mathbb{R}$, λ is a strictly positive p.d.f. on $\mathbb{R}$, and $\underline{f_t} = (f_t, f_{t-1}, \cdots)$.

Assumption A.4 requires that the Markov transition distribution of $y_{i,t}$ conditional on $y_{i,t-s}$ and the factor path converges to the long run conditional distribution of $y_{i,t}$, denoted $\mathbb{P}[\cdot|\underline{f_t}]$, as the lag s tends to ∞. The conditional long run distribution $\mathbb{P}[\cdot|\underline{f_t}]$ and the conditional beta mixing coefficients $\beta_t(s)$ at date t depend on the factor path $\underline{f_t}$, and thus are stochastic. The beta mixing coefficients $\beta_t(s)$ are assumed to converge to zero as lag s increases, for any factor path, implying the irrelevance of the initial values of the $y_{i,t}$'s in the long run conditional on the 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, by the Lebesgue Theorem, under Assumption A.4 the integrated mixing coefficients $E_0\left[\beta_t(s)\right]$ are such that $E_0\left[\beta_t(s)\right] \to 0$ as $s \to \infty$. 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. As usual, it is convenient for expository purpose to disregard the short run effect of the initial observations by introducing a suitable assumption on their distribution.

A.5: *The initial observations $y_{i,0}$, with $i = 1, ..., n$, are i.i.d. conditional on the factor path (f_t), with distribution corresponding to the long run distribution $\mathbb{P}[\cdot|\underline{f_0}]$ at time $t = 0$.*

Assumption A.5 implies that at each date t the distribution of $y_{i,t}$ conditional on the factor path is the long run distribution $\mathbb{P}[\cdot|\underline{f_t}]$. This property is the analog of stationarity for the individual histories conditional on the factor path.

3 The likelihood expansion

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:

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

If the factors were observable, the terms $l_{micro}\left(\underline{y_T}|\underline{f_T}; \beta\right)$ and $l_{macro}\left(\underline{f_T}; \theta\right)$ 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 observations $\underline{y_T}$ is obtained by integrating out the factor path $\underline{f_T}$:

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

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 computation of the ML estimates [8]. However, for large n, this integral can be approximated along the lines of the Laplace approximation [Laplace (1774)]. The use of the Laplace approximation in

[8] 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 approximate likelihood function given in Proposition 1 avoids the numerically cumbersome expectation step, while controlling the approximation error.

the econometric literature is as early as by Holly, Phillips (1979) and Phillips (1983) for the derivation of the marginal distribution of instrumental variable estimators. Tierney, Kadane (1986) used this device to derive the posterior distribution in Bayesian statistics. More recently, the Laplace approximation has been used in Arellano, Bonhomme (2009) to derive the bias of the integrated likelihood in nonlinear panel models with individual fixed effects. Huber, Scaillet, Victoria-Feser (2009) use the Laplace approximation to develop a tractable estimator for a multivariate logit model in a latent factor framework in finance. In our setting with serially dependent factors, the Laplace approximation is applied to an integral w.r.t. the full path of time effects. Specifically, we start by defining for any parameter value $\beta \in \mathcal{B}$ and date $t = 1, ..., T$ the cross-sectional ML estimator of the factor value:

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

where the compact set $\mathcal{F}_n \subset \mathbb{R}^m$ grows when $n \to \infty$ as described by Assumption H.7 in Appendix A.1. Then, by a Taylor expansion of the integrand in the RHS of equation (3.2) around $\left(\hat{f}_{n,1}(\beta)', ..., \hat{f}_{n,T}(\beta)'\right)'$, that is the maximizer of $\sum_{t=1}^{T}\sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f_t; \beta)$ w.r.t. the factor path, we get:

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

where:

$$I_{n,t}(\beta) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f_t \partial f_t'}\left(y_{i,t}|y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta\right), \tag{3.4}$$

and the remainder term $\psi_{n,t}$ is defined in (a.3) in Appendix A.2.1. By introducing the change of variables $z_t = \sqrt{n}\left[I_{n,t}(\beta)\right]^{1/2}\left(f_t - \hat{f}_{n,t}(\beta)\right) \iff f_t = \hat{f}_{n,t}(\beta) + \frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2} z_t$, for $t = 1, ..., T$, and expanding function $\exp\left(\sum_t \psi_{n,t}\right)$ in a power series of the $n^{-1/2} z_t$, the multivariate integral in the expression of the likelihood can be written as a linear combination of power moments of the standard Gaussian distribution, with coefficients depending on the observations. The next proposition gives the expansion for the (nT-standardized) log-likelihood function of the sample:

$$\mathcal{L}_{nT}(\beta, \theta) = \frac{1}{nT} \log l\left(\underline{y_T}; \beta, \theta\right), \tag{3.5}$$

as a power series of $1/n$, and controls the stochastic order of the remainder term.

PROPOSITION 1. *Let Assumptions A.1-A.5 and H.1-H.13 in Appendix A.1 be satisfied.*
(i) If $n, T \to \infty$ such that $T^{\nu}/n = O(1)$, for a value $\nu > 1$, we have:

$$\mathcal{L}_{nT}(\beta, \theta) = \mathcal{L}_{nT}^{*}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta, \theta) + \Psi_{nT}(\beta, \theta), \tag{3.6}$$

where:

$$\mathcal{L}_{nT}^{*}(\beta) = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\log h\left(y_{i,t}|y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta\right), \tag{3.7}$$

$$\mathcal{L}_{1,nT}(\beta, \theta) = -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\log\det I_{n,t}(\beta) + \frac{1}{T}\sum_{t=1}^{T}\log g\left(\hat{f}_{n,t}(\beta)|\hat{f}_{n,t-1}(\beta); \theta\right), \tag{3.8}$$

with $I_{n,t}(\beta)$ defined as in (3.4), and the remainder term $\Psi_{nT}(\beta,\theta)$ is such that $\sup_{\beta\in\mathcal{B},\theta\in\Theta}|\Psi_{nT}(\beta,\theta)| = o_p(1/n)$ w.r.t. the true distribution.
(ii) If $n, T \to \infty$ such that $T^{\nu}/n = O(1)$, for a value $\nu > 3/2$, we have:

$$\mathcal{L}_{nT}(\beta, \theta) = \mathcal{L}_{nT}^{*}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta, \theta) + \frac{1}{n^2}\mathcal{L}_{2,nT}(\beta, \theta) + \tilde{\Psi}_{nT}(\beta, \theta), \tag{3.9}$$

where the reminder term $\tilde{\Psi}_{nT}(\beta,\theta)$ is such that $\sup_{\beta\in\mathcal{B},\theta\in\Theta}|\tilde{\Psi}_{nT}(\beta,\theta)| = o_p(1/n^2)$. When the factor is one-dimensional, i.e. $m = 1$, the expression of term $\mathcal{L}_{2,nT}(\beta,\theta)$ is given by:

$$\begin{aligned}
\mathcal{L}_{2,nT}(\beta,\theta) \quad = \quad & \frac{1}{8}\frac{1}{T}\sum_{t=1}^{T}J_{4,n,t}(\beta) + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}D_{20,nt}(\beta,\theta) + \frac{1}{2}\frac{1}{T}\sum_{t=2}^{T}D_{02,nt}(\beta,\theta) \\
& + \frac{5}{24}\frac{1}{T}\sum_{t=1}^{T}[J_{3,nt}(\beta)]^2 + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}[D_{10,nt}(\beta,\theta)]^2 + \frac{1}{2}\frac{1}{T}\sum_{t=2}^{T}[D_{01,nt}(\beta,\theta)]^2 \\
& + \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}J_{3,n,t}(\beta)D_{10,nt}(\beta,\theta) + \frac{1}{2}\frac{1}{T}\sum_{t=2}^{T}J_{3,n,t-1}(\beta)D_{01,nt}(\beta,\theta) \\
& + \frac{1}{T}\sum_{t=2}^{T}D_{10,n,t-1}(\beta,\theta)D_{01,nt}(\beta,\theta),
\end{aligned} \tag{3.10}$$

with $J_{p,nt}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^p \log h}{\partial f_t^p}\left(y_{i,t}|y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta\right)[I_{n,t}(\beta)]^{-p/2}$, *for* $p = 3, 4$, *and* $D_{pq,nt}(\beta,\theta) = \frac{\partial^{p+q}\log g}{\partial f_t^p \partial f_{t-1}^q}\left(\hat{f}_{nt}(\beta)|\hat{f}_{n,t-1}(\beta); \theta\right)[I_{n,t}(\beta)]^{-p/2}[I_{n,t-1}(\beta)]^{-q/2}$, *for* $p, q = 0, 1, 2$.

Proof. See Appendix A.2.1. □

Function $\mathcal{L}_{nT}^{*}(\beta)$, called profile log-likelihood function, is the micro log-likelihood of β concentrated w.r.t. the factor values, as if the latter ones were nuisance parameters. It contains the information on β which is independent of the factor dynamics. Proposition 1 shows that the leading term in the asymptotic expansion of the log-likelihood function $\mathcal{L}_{nT}(\beta, \theta)$ in powers of $1/n$ involves parameter β only and is equal to $\mathcal{L}_{nT}^{*}(\beta)$. The next term $\mathcal{L}_{1,nT}(\beta, \theta)$ at order $1/n$ is the first to provide information on parameter θ characterizing the factor dynamics. It corresponds to the macro log-likelihood after replacing the unobservable factor values with cross-sectional approximations depending on β. The log-det component comes from the Jacobian in the change of variable for Laplace approximation. The term of order $1/n^2$ involves first- and second-order derivatives of the macro log-density function, and third- and fourth-order derivatives of the micro log-density w.r.t. the factor value. Its specific expression seems difficult to interpret in the general framework. It is possible to derive $\mathcal{L}_{2,nT}(\beta, \theta)$ also in the multiple factor case ($m \geq 2$), but its expression is notationally cumbersome and is not provided here. Functions $\mathcal{L}_{nT}^{*}(\beta)$, $\mathcal{L}_{1,nT}(\beta, \theta)$ and $\mathcal{L}_{2,nT}(\beta, \theta)$ do not involve integrals w.r.t. the factor path, but only nonlinear aggregates of sample observations. In fact, all multidimensional integrals are included in the residual terms $o_p(1/n)$, or $o_p(1/n^2)$. Thus, Propositions 1 (i) and (ii) provide closed-form approximations of the log-likelihood function at order $o_p(1/n)$, and $o_p(1/n^2)$, respectively. The condition $T^{\nu}/n = O(1)$, $\nu > 1$, is used in Appendix A.2.1 to control the stochastic remainder term in the Laplace approximation at order $o_p(1/n)$. This condition constrains the growth rate of the dimension Tm of the integral in equation (3.2) relatively to the cross-sectional size n, which plays the role of the parameter tending to infinity in our application of the Laplace approximation method. The more restrictive condition $T^{\nu}/n = O(1)$, $\nu > 3/2$, is used to derive the more accurate log-likelihood approximation at order $o_p(1/n^2)$.

The true log-likelihood function $\mathcal{L}_{nT}(\beta, \theta)$ is invariant to one-to-one transformations of the factor vector $f \to \phi(f)$, say, where ϕ is any invertible mapping in $\mathbb{R}^m$. The leading term $\mathcal{L}_{nT}^{*}(\beta)$ in the log-likelihood expansion is invariant to such transformations, since it corresponds to the concentrated micro log-likelihood. As a consequence, also the terms $\mathcal{L}_{1,nT}(\beta, \theta)$ and $\mathcal{L}_{2,nT}(\beta, \theta)$ at order $1/n$ and $1/n^2$ are invariant to one-to-one factor transformations, as can be directly verified from their expressions in (3.8) and (3.10) (for $m = 1$). In particular, the invariance of $\mathcal{L}_{1,nT}(\beta, \theta)$ explains the log-det component $-\dfrac{1}{2}\dfrac{1}{T}\sum_{t=1}^{T} \log\det I_{n,t}(\beta)$. This component corresponds to the term introduced by Cox, Reid (1987) in their modified profile likelihood [see also Sweeting (1987)]. [9]

We can interpret the leading term in the expansions given in Proposition 1 as an example of the asymptotic equivalence of frequentist and Bayesian methods in large sample [see e.g. Bickel, Yahav (1969), Ibragimov, Has'minskii (1981)]. To get the intuition, let time dimension T be fixed and parameter θ be given for a moment. Then, our specification with stochastic common factor can be seen as a Bayesian approach w.r.t. parameter β and time effects $\underline{f_T}$. The prior distribution is such that the density of $\underline{f_T}$ given β is $\prod_{t=1}^{T} g\left(f_t|f_{t-1};\theta\right)$, independent of β, and the prior distribution of β is diffuse. Then, the posterior density of $(\beta, \underline{f_T})$ corresponds to the RHS of equation (3.1), while the posterior density of β corresponds to the RHS of equation (3.2), up to multiplicative constants. Thus, as $n \to \infty$, 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 β concentrated w.r.t. parameters f_t, $t = 1, ..., T$. The asymptotic irrelevance of the second term in the RHS of (3.6), or (3.9), involving the transition density of the factor corresponds to the irrelevance of the prior distribution in large samples. Our results show that this asymptotic equivalence is still valid when the number of time effects parameters tends to infinity: $T \to \infty$, such that $T^\nu/n \to 0$, $\nu > 1$. [10]

4 Maximum Likelihood and Maximum Approximated Likelihood estimators

4.1 The estimators of micro- and macro-parameters

The ML estimator of (β, θ) is derived by maximizing the log-likelihood function $\mathcal{L}_{nT}(\beta,\theta)$ defined in equation (3.5). Alternative estimators can be defined by maximizing jointly w.r.t. β and θ approximations of the log-likelihood function at probability order $1/n$, and $1/n^2$, respectively.

[9] When the micro-parameter β and the time effect f_t are information orthogonal, that is, $E_0\left[-\dfrac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t;\beta_0)}{\partial\beta\partial f_t'}|\underline{f_t}\right] = 0$, $\mathbb{P}$-a.s., the score w.r.t. β of the approximated log-likelihood in Proposition 1 (i) corresponds to the score of the profile log-likelihood in Cox, Reid (1987), and to the score of the penalized log-likelihood in Bester, Hansen (2009), up to order $o_p(1/n)$. When information orthogonality does not apply, the scores of the three log-likelihoods differ at order $O_p(1/n)$.

[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 different regularity conditions.

From Proposition 1 (i), an approximation at order $o_p(1/n)$ is given by:

$$\mathcal{L}_{nT}^{\mathrm{CSA}}(\beta,\theta) = \mathcal{L}_{nT}^{*}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta,\theta). \tag{4.1}$$

This approximation defines the cross-sectional asymptotic (CSA) log-likelihood function. Similarly, from Proposition 1 (ii) an approximation valid up to order $o_p(1/n^2)$ is:

$$\mathcal{L}_{nT}^{\mathrm{GA}}(\beta,\theta) = \mathcal{L}_{nT}^{*}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta,\theta) + \frac{1}{n^2}\mathcal{L}_{2,nT}(\beta,\theta). \tag{4.2}$$

This approximated log-likelihood function defines the granularity adjusted (GA) log-likelihood function. Then, we define the maximum likelihood and maximum approximated likelihood estimators as follows:

DEFINITION 1. *(i) The maximum likelihood estimator is* $\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right) = \underset{\beta\in\mathcal{B},\theta\in\Theta}{\arg\max}\, \mathcal{L}_{nT}\left(\beta,\theta\right)$.
(ii) The CSA maximum likelihood estimator is $\left(\tilde{\beta}_{nT}^{CSA}, \tilde{\theta}_{nT}^{CSA}\right) = \underset{\beta\in\mathcal{B},\theta\in\Theta}{\arg\max}\, \mathcal{L}_{nT}^{CSA}(\beta,\theta)$.
(iii) The GA maximum likelihood estimator is $\left(\tilde{\beta}_{nT}^{GA}, \tilde{\theta}_{nT}^{GA}\right) = \underset{\beta\in\mathcal{B},\theta\in\Theta}{\arg\max}\, \mathcal{L}_{nT}^{GA}(\beta,\theta)$.

The CSA and GA maximum likelihood estimators are computationally more convenient than the standard ML estimator, since the CSA and GA log-likelihood functions do not involve integrals w.r.t. the factor path. 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 asymptotic limit of an infinitely fine grained portfolio ($n = \infty$), and compute the adjustment at order $1/n$ [see Gagliardini, Gouriéroux, Monfort (2012), Section 5, for a general presentation of granularity for risk measures]. A similar approach is applied here on the likelihood function and ML estimators instead of being applied on the future portfolio value distribution and its quantiles.

4.2 Identification

To analyze the asymptotic properties of the estimators in Definition 1, we introduce suitable identification assumptions for the micro- and macro-parameters. Identification is ensured by the global and local behaviour of the large sample limit of the likelihood function around the true parameter value. We exploit the asymptotic expansion of the log-likelihood function in Proposition 1 and consider the case in which the next two conditions are satisfied: (i) the micro-parameter β is identifiable from the leading term $\mathcal{L}^*_{nT}(\beta)$, and (ii) the full parameter vector (β, θ) is identifiable from the log-likelihood approximation at first-order in $1/n$, that is, the CSA log-likelihood $\mathcal{L}^{CSA}_{nT}(\beta, \theta)$. The cases in which the identification of the micro-parameter requires the first-order term $n^{-1}\mathcal{L}_{1,nT}(\beta, \theta)$, or the identification of some parameters requires the second-order term $n^{-2}\mathcal{L}_{2,nT}(\beta, \theta)$, lead to different asymptotic behaviours of the estimators and are not considered in this paper. Let us now derive the identification assumptions, starting from the micro-parameter.

(i) Let us first define the population counterpart of the cross-sectional estimate of the factor value:

$$f_t(\beta) = \arg\max_{f \in \mathbb{R}^m} E_0 \left[\log h\left(y_{i,t} | y_{i,t-1}, f; \beta\right) | \underline{f_t} \right], \tag{4.3}$$

where $E_0\left[.|\underline{f_t}\right]$ denotes the expectation w.r.t. the true conditional distribution of $(y_{i,t}, y_{i,t-1})$ given $\underline{f_t} = (f_t, f_{t-1}, ...)$. The pseudo-true factor value $f_t(\beta)$ maximizes the limiting cross-sectional log-likelihood at date t for given parameter value β. It is a function of both parameter β and factor path $\underline{f_t}$. Thus, $f_t(\beta)$ is a stochastic process, for any $\beta \in \mathcal{B}$. We assume that the pseudo-true factor value is globally and locally identified (see Assumption H.2 in Appendix A.1). By the properties of the Kullback-Leibler discrepancy, at true parameter value β_0 the pseudo-true factor value $f_t(\beta_0)$ coincides with the true factor value f_t, $\mathbb{P}$-a.s., for any t.

Let us now define the function:

$$\begin{aligned} \mathcal{L}^*(\beta) &= \operatorname*{plim}_{n,T\to\infty} \mathcal{L}^*_{nT}(\beta) = \operatorname*{plim}_{n,T\to\infty} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \log h\left(y_{i,t} | y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta \right) \\ &= E_0 \left[\log h\left(y_{i,t} | y_{i,t-1}, f_t(\beta); \beta\right) \right], \end{aligned} \tag{4.4}$$

where the convergence is uniform w.r.t. $\beta \in \mathcal{B}$, and is proved in Lemma 1 (i) (see supplementary material). Intuitively, function $\mathcal{L}^*(\beta)$ is the asymptotic micro log-likelihood concentrated w.r.t. the stochastic process (f_t). The assumptions below concern the identification of parameter β.

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

A.7 (Local identification assumption for β): *The matrix $I^*_0 = -\dfrac{\partial^2 \mathcal{L}^*(\beta_0)}{\partial\beta\partial\beta'}$ is positive definite.*

The matrix I_0^* is given by:

$$I_0^* = E_0\left[I_{\beta\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} I_{f\beta}(t)\right] = E_0\left[U_{it} U_{it}'\right], \tag{4.5}$$

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:

$$I(t) = E_0\left[-\frac{\partial^2 \log h\left(y_{i,t}|y_{i,t-1}, f_t; \beta_0\right)}{\partial\left(\beta', f_t'\right)' \partial\left(\beta', f_t'\right)} |\underline{f_t}\right], \tag{4.6}$$

and $U_{it} = \dfrac{\partial \log h(y_{i,t}|y_{i,t-1}; f_t; \beta_0)}{\partial \beta} - I_{\beta f}(t) I_{ff}(t)^{-1} \dfrac{\partial \log h(y_{i,t}|y_{i,t-1}; f_t; \beta_0)}{\partial f_t}$. Thus, I_0^* is the variance-covariance matrix of the residual U_{it} in the orthogonal conditional projection of the score w.r.t. the micro-parameter on the score w.r.t the factor value given $\underline{f_t}$.

(ii) Let us now consider the macro-component of the log-likelihood. Under Assumptions A.6-A.7, parameter β can be estimated at a rate faster than the rate for parameter θ. Hence, the relevant criterion for identification of θ is the mapping $\theta \to \mathcal{L}_1(\beta_0, \theta)$, where $\mathcal{L}_1(\beta, \theta)$ is the large sample limit of $\mathcal{L}_{1,nT}(\beta, \theta)$ in equation (3.8). We have $\mathcal{L}_1(\beta_0, \theta) = E_0\left[\log g(f_t|f_{t-1}; \theta)\right]$, up to a term constant in θ [see Lemma 1 (ii) in the supplementary material]. Thus, the identification assumptions for the macro-parameter are the following:

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

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

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

4.3 Asymptotic properties of the estimators

We consider the asymptotic properties of the CSA, GA and true ML estimators in Definition 1 under Assumptions A.1-A.9 and H.1-H.15 in Appendix A.1. Assumptions A.1-A.9 are invariant to one-to-one transformations of the factor vector (if the transformation is independent of the parameters β, θ), whereas some of the Assumptions H.1-H.15 are not. Moreover, the CSA, GA and true ML estimators are numerically invariant to one-to-one transformations of the factor. Thus, in order to establish their asymptotic properties it is enough that the regularity conditions H.1-H.15 in Appendix A.1 are satisfied for a suitable choice of the factor representation.

Let us first study the probability order of the difference between the CSA and GA ML estimators on the one hand, and the true ML estimators on the other hand.

PROPOSITION 2. *Under Assumptions A.1-A.9 and H.1-H.15, the CSA, GA and true ML estimators in Definition 1 are such that:*

$$\tilde{\beta}^{CSA}_{nT} - \tilde{\beta}_{nT} = o_p(1/n), \qquad \tilde{\theta}^{CSA}_{nT} - \tilde{\theta}_{nT} = O_p\left(\frac{(\log n)^{\delta_1}}{\sqrt{n}}\right), \tag{4.7}$$

$$\tilde{\beta}^{GA}_{nT} - \tilde{\beta}_{nT} = o_p(1/n), \qquad \tilde{\theta}^{GA}_{nT} - \tilde{\theta}_{nT} = O_p\left(\frac{(\log n)^{\delta_1}}{\sqrt{n}}\right), \tag{4.8}$$

for a constant $\delta_1 > 0$, if $n, T \to \infty$ such that $T^\nu/n = O(1)$, $\nu > 1$, and:

$$\tilde{\beta}^{CSA}_{nT} - \tilde{\beta}_{nT} = O_p(1/n^2), \qquad \tilde{\theta}^{CSA}_{nT} - \tilde{\theta}_{nT} = O_p(1/n), \tag{4.9}$$

$$\tilde{\beta}^{GA}_{nT} - \tilde{\beta}_{nT} = o_p(1/n^2), \qquad \tilde{\theta}^{GA}_{nT} - \tilde{\theta}_{nT} = o_p(1/n), \tag{4.10}$$

if $n, T \to \infty$ such that $T^\nu/n = O(1)$, $\nu > 3/2$.

Proof. See Appendix A.2.2. □

Proposition 2 states that the CSA, GA and true ML estimators are asymptotically equivalent, and provides the probability orders of this equivalence. If $T^\nu/n = O(1)$, $\nu > 3/2$, the GA maximum likelihood estimator provides a more accurate approximation of the true ML estimator compared to the CSA maximum likelihood estimator. The accuracy of the approximation is superior for the micro- than for the macro-parameters. Under the less restrictive condition $T^\nu/n = O(1)$, $\nu > 1$, the CSA and GA ML estimators have the same order of accuracy in approximating the true ML estimator, and this accuracy is again superior for the micro-parameters.

The joint asymptotic distribution of the estimators of the micro- and macro-parameters are given in the next proposition.

PROPOSITION 3. *Let Assumptions A.1-A.9 and H.1-H.15 be satisfied, and let $(\hat{\beta}_{nT}, \hat{\theta}_{nT})$ be either the CSA, GA, or true ML estimator in Definition 1. Then, if $n, T \to \infty$ such that $T^\nu/n = O(1)$, $\nu > 1$, estimator $(\hat{\beta}_{nT}, \hat{\theta}_{nT})$ is consistent and asymptotically normal:*

$$\begin{bmatrix} \sqrt{nT}\left(\hat{\beta}_{nT} - \beta_0\right) \\ \sqrt{T}\left(\hat{\theta}_{nT} - \theta_0\right) \end{bmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} B^*_{\beta\beta} & B^*_{\beta\theta} \\ B^*_{\theta\beta} & B^*_{\theta\theta} \end{pmatrix}\right), \tag{4.11}$$

with asymptotic variance-covariance matrix

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

where $I_0^* = E_0\left[I_{\beta\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} I_{f\beta}(t)\right]$ *and* $I_{1,\theta\theta} = E_0\left[-\dfrac{\partial^2 \log g\left(f_t | f_{t-1}; \theta_0\right)}{\partial\theta\partial\theta'}\right]$.

Proof. See Appendix A.2.3. □

Proposition 3 states that the CSA, GA and ML estimators are asymptotically normal with different rates of convergence for the micro- and macro-component, that are root-nT and root-T, respectively, if $T^\nu/n = O(1)$, $\nu > 1$. The asymptotic variance-covariance matrix B^* defines the joint efficiency bound for estimating both micro- and macro-parameters (β, θ). Matrix B^* is block-diagonal for the micro- and macro-components, with the diagonal blocks corresponding to the Hessian matrices $I_0^* = -\dfrac{\partial^2 \mathcal{L}^*\left(\beta_0\right)}{\partial\beta\partial\beta'}$ and $I_{1,\theta\theta} = -\dfrac{\partial^2 \mathcal{L}_1\left(\beta_0, \theta_0\right)}{\partial\theta\partial\theta'}$. The zero out-of-diagonal blocks in the efficiency bound imply that parameters β and θ can be considered independently for estimation purpose. This justifies ex-post their interpretation as micro- and macro-parameters, respectively, since parameter β (resp. θ) contains no macro-information (resp. no micro-information) under identification Assumptions A.6-A.9. The condition $T^\nu/n = O(1)$, $\nu > 1$, implies that the asymptotic distributions of the estimators are centered. Thus, in our framework there is no incidental parameter bias [see e.g. Neyman, Scott (1948) and Lancaster (2000)].

The result in Proposition 3 is a consequence of the expansion of the likelihood function in Proposition 1. Indeed, under identification Assumptions A.6-A.7 and the regularity conditions in Appendix A.1, for large n and T the relevant term for estimation of parameter β is $\mathcal{L}^*_{nT}(\beta)$. The corresponding limit log-likelihood function is $\mathcal{L}^*(\beta)$, and the efficiency bound $B^*_{\beta\beta}$ for β is the inverse of the Hessian I_0^*. Similarly, the efficiency bound $B^*_{\theta\theta}$ for θ is the inverse of the Hessian $I_{1,\theta\theta}$. Moreover, the (standardized) ML estimators of β and θ are asymptotically independent. Therefore, the efficiency bound $B^*_{\beta\beta}$ for β given in Proposition 3 is the same as the efficiency bound for β with known transition density of the factor. Finally, the information matrix I_0^* is smaller than the information matrix $I_0^{**} = E_0\left[I_{\beta\beta}(t)\right]$ corresponding to the case of observable factor, while matrix $I_{1,\theta\theta}$ is equal to the information for θ with observable factor. 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. Therefore, the unobservability of the factor has no efficiency impact asymptotically for estimating θ, but has an impact for estimating β. Indeed, the factor values can be estimated at a rate close to $1/\sqrt{n}$ (see Proposition 5 below), a rate which is faster than the rate $1/\sqrt{T}$ for estimating θ, if $T^{\nu}/n = O(1)$, $\nu > 1$, and slower than the rate $1/\sqrt{nT}$ for estimating β.

Proposition 3 shows that the computationally convenient CSA and GA ML estimators are asymptotically efficient estimators of parameters β and θ (see also Section 5 for other asymptotically efficient estimators). This result concerns first-order asymptotics only. It is out of the scope of the present paper to get the higher-order expansion of the asymptotic distribution of the standardized estimators $\left[\sqrt{nT}\left(\hat{\beta}_{nT} - \beta_0\right)', \sqrt{T}\left(\hat{\theta}_{nT} - \theta_0\right)'\right]'$ in the sense of Ghosh, Subramanyam (1974) and Pfanzagl, Wefelmeyer (1978), for instance to correct for the higher-order bias in n and/or T. It is likely difficult to derive the higher-order expansions due to the double asymptotics and the different rates of convergence of the estimators of micro- and macro-parameters. The GA ML estimator is closer to the unfeasible ML estimator than the CSA ML estimator is, if $T^{\nu}/n = O(1)$, $\nu > 3/2$, and likely inherits its finite-sample properties. In some applications to credit risk, the ML and GA ML estimators can feature worse finite-sample properties than the CSA ML estimator [see Gourieroux, Jasiak (2012)]. Therefore, we may expect different higher-order expansions for the CSA and GA ML estimators.

4.4 Semi-parametric efficiency

The efficiency bound $B^*_{\beta\beta}$ for parameter β in Proposition 3 is independent of the parametric model $g(f_t|f_{t-1};\theta)$, $\theta \in \mathbb{R}^p$, for the transition 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 β is the efficiency bound for estimating β in the semi-parametric model in which the transition $g(f_t|f_{t-1})$ of the factor is a functional parameter. The semi-parametric efficiency bound B can be computed by using Stein's heuristic [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.12-H.15 in Appendix A.1, and let $B^*_{\beta\beta}(g_\theta)$ be the corresponding parametric efficiency bound for estimating β.

DEFINITION 2. *The semi-parametric efficiency bound B is defined by:*

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

where the maximization is performed w.r.t. the well-specified parametric models g_θ for the transition of f_t that satisfy Assumptions A.8-A.9 and H.12-H.15.

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

COROLLARY 4. *Under Assumptions A.1-A.7 and H.1-H.11, and if $n, T \to \infty$ such that $T^\nu/n = O(1)$, $\nu > 1$, the semi-parametric efficiency bound for β is equal to the parametric efficiency bound: $B = B^*_{\beta\beta} = E_0\left[I_{\beta\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}I_{f\beta}(t)\right]^{-1}$.*

Thus, any well-specified parametric model g_θ is the least-favorable one in the sense of Chamberlain (1987). 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 β. [11]

4.5 Approximation of the factor values

Given a consistent estimator of the micro-parameter β, we can use the cross-sectional aggregate $\hat{f}_{n,t}(\beta)$ to get consistent approximations of the factor value f_t. [12]

DEFINITION 3. *Let $\hat{\beta}_{nT}$ denote either the CSA, GA, or true ML estimator of the micro-parameter β in Definition 1. Then a cross-sectional approximation of the factor value at date t is:*

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

for $t = 1, ..., T$, where $\hat{f}_{n,t}(\beta)$ is defined in equation (3.3).

For any given date t, the factor approximation $\hat{f}_{nT,t}$ depends on the whole individual histories and is a kind of smoothed factor value. Its asymptotic properties are given in the next proposition.

PROPOSITION 5. *Suppose Assumptions A.1-A.9 and H.1-H.15 hold, and let $n, T \to \infty$ such that $T^\nu/n = O(1)$, $\nu > 1$. Then:*
i) For any date t, conditional on $\underline{f_t}$ we have: $\sqrt{n}\left(\hat{f}_{nT,t} - f_t\right) \xrightarrow{d} N\left(0, I_{ff}(t)^{-1}\right)$.

[11] The proof of Proposition 3 shows that the CSA and GA ML estimators of parameter β based on a misspecified factor model remain consistent and first-order asymptotically efficient (but not the CSA and GA ML estimators of parameter θ).

[12] Approximations of factor values in panel data with large cross-sectional and time dimensions have been proposed in, e.g., Forni, Reichlin (1998), Forni, Hallin, Lippi, Reichlin (2000), Bai, Ng (2002), Stock, Watson (2002), Connor, Hagmann, Linton (2012). All these papers consider linear factor models for the micro-dynamics.

ii) $\sup_{1\leq t\leq T}\left\|\hat{f}_{nT,t}-f_t\right\| = O_p\left(\frac{(\log n)^{\delta_2}}{\sqrt{n}}\right)$, *where* $\delta_2 = \gamma_2+\gamma_3/2+2/d_3+1/2$ *and constants* $\gamma_2,\gamma_3\geq 0$, $d_3>0$ *are defined in Assumptions H.8-H.10 in Appendix A.1.*

Proof. See Appendix A.2.4. □

Conditionally on the factor path, the factor approximation converges to the true factor value f_t at rate $1/\sqrt{n}$. Since $\hat{\beta}_{nT}$ is root-nT consistent, estimator $\hat{f}_{nT,t}$ is asymptotically equivalent to the unfeasible ML estimator $\hat{f}_{n,t}(\beta_0)$ for known micro-parameter β_0. The asymptotic variance $I_{ff}(t)^{-1}$ of $\hat{f}_{nT,t}$ is the inverse of the Fisher information for estimating f_t in the cross-section at date t with known β_0. The uniform convergence in Proposition 5 ii) follows from the convergence of $\hat{f}_{n,t}(\beta)$ to $f_t(\beta)$ uniformly in $\beta\in\mathcal{B}$ and $t=1,...,T$ (see Limit Theorem 1 in the supplementary materials) and the root-nT consistency of estimator $\hat{\beta}_{nT}$ (see Proposition 3). Proposition 5 ii) is not invariant to one-to-one transformations of the factor, since the regularity assumptions include tail conditions on the factor distribution (see Assumptions H.8-H.10).

5 Two-step efficient estimators

In this section we introduce another asymptotically efficient estimation approach, in which the estimators of the micro- and macro-parameters can be computed in two steps and are easy to interpret.

DEFINITION 4. *The two-step estimator is defined by:*

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

and:

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

where $\hat{f}_{n,t}(\beta)$ *is defined in equation (3.3) and* $\hat{f}^*_{nT,t}=\hat{f}_{n,t}\left(\hat{\beta}^*_{nT}\right)$ *for* $t=1,...,T$.

In the first step, the estimator $\hat{\beta}^*_{nT}$ of the micro-parameter is obtained by maximizing the profile likelihood function $\mathcal{L}^*_{nT}(\beta)$ defined in equation (3.7). Thus, $\hat{\beta}^*_{nT}$ is the time fixed effects estimator of β which considers the f_t values as additional unknown parameters. Since the function $\mathcal{L}^*_{nT}(\beta)$

does not involve the transition pdf of the factor, the estimator $\hat{\beta}^*_{nT}$ does not depend on the specification of the factor dynamics. In this sense, $\hat{\beta}^*_{nT}$ is a semi-parametric estimator, which is not the case for the CSA and GA ML estimators. Estimator $\hat{\beta}^*_{nT}$ is used to derive cross-sectional approximations $\hat{f}^*_{nT,t}$ of the factor values. These cross-sectional factor approximations correspond to the ML estimates of the time fixed effects. In the second step, the approximations of the factor values are used to derive the approximation of the macro-likelihood function $\sum_{t=1}^{T} \log g\left(\hat{f}^*_{nT,t}|\hat{f}^*_{nT,t-1};\theta\right)$. By maximizing this approximate likelihood w.r.t. θ, we get an estimator of the macro-parameter.

The asymptotic distribution of the two-step estimator is given in the next proposition.

PROPOSITION 6. *Suppose Assumptions A.1-A.9 and H.1-H.15 hold, and let $n, T \to \infty$ such that $T^\nu/n = O(1)$, $\nu > 1$. Then the estimators in Definition 4 are such that:*
i) $\hat{\beta}^*_{nT} - \tilde{\beta}_{nT} = O_p(1/n)$, $\hat{\theta}^*_{nT} - \tilde{\theta}_{nT} = O_p\left(\frac{(\log n)^{\delta_1}}{\sqrt{n}}\right)$, for $\delta_1 > 0$ as in Proposition 2.
ii) The estimator $(\hat{\beta}^{*\prime}_{nT}, \hat{\theta}^{*\prime}_{nT})'$ is consistent and asymptotically normal such that:

$$\left[\begin{array}{c} \sqrt{nT}\left(\hat{\beta}^*_{nT} - \beta_0\right) \\ \sqrt{T}\left(\hat{\theta}^*_{nT} - \theta_0\right) \end{array}\right] \xrightarrow{d} N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} (I^*_0)^{-1} & 0 \\ 0 & I^{-1}_{1,\theta\theta} \end{array}\right)\right),$$

where matrices I^*_0 and $I_{1,\theta\theta}$ are given in Proposition 3.

Proof. See Appendix A.2.5. □

From Propositions 2 and 6 i), the two-step estimator of the micro-parameter provides a less accurate approximation of the true ML estimator compared with the CSA and GA ML estimators. However, the semi-parametric estimator $\hat{\beta}^*_{nT}$ still achieves asymptotically the (semi-) parametric efficiency bound. In other words, the conditional likelihood estimator of β (based on concentrating out the f_t) is first-order asymptotically equivalent to the full likelihood estimator of β.

The first-order asymptotic distribution of the fixed effects estimator $\hat{\beta}^*_{nT}$ in Proposition 6 (ii) is not surprising in view of Theorem 1 in Hahn, Newey (2004), who consider a nonlinear setting with micro-density $h(y_{i,t}|\alpha_i;\beta)$ and individual fixed effects α_i. In particular, the interpretation of the asymptotic variance I^*_0 in equation (4.5) as the outer product of the residual in the orthogonal projection of the score w.r.t. the micro-parameter on the score w.r.t. the fixed effect, is the same as in Theorem 1 in Hahn, Newey (2004). However, Proposition 6 cannot be obtained by interchanging the individual and time indices, and also the sizes n and T, and by letting $\rho \to 0$ in Hahn, Newey

(2004), where their parameter $\rho > 0$ is such that $n/T \to \rho$. Indeed, in our paper the micro-density $h(h_{i,t}|y_{i,t-1}, f_t; \beta)$ depends on the lagged variable $y_{i,t-1}$, and our asymptotic results are under a probability measure such that the time effects f_t define a stochastic process with parametric dynamics and are not a sequence of fixed constants. Hahn, Kuersteiner (2002) consider a linear dynamic panel model with individual fixed effects and prove that the (bias-corrected) fixed effects estimator is asymptotically efficient in the sense of Hayek's convolution theorem. Proposition 6 differs from Hahn, Kuersteiner (2002) since we define the efficiency bound as the asymptotic variance of the ML estimator under a parametric dynamics of the random time effects.

6 Stochastic migration model

In this section we illustrate the finite sample properties of the two-step estimators in Definition 4 with a stochastic migration model.

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 (2005a, b), Feng, Gouriéroux, Jasiak (2008), Koopman, Lucas, Monteiro (2008)]. A basic stochastic migration model is the ordered qualitative model with one factor, which extends the ASRF model of Section 2.1 to more than two alternatives. Let us denote by $y_{i,t}$, with t varying, the sequence of ratings for corporation i. The possible ratings are $k = 1, 2, ..., K$, say [13]. The micro-dynamic

[13] In practice, the alternative $k = K$ corresponds typically to default, which is an absorbing state. Then, the stationarity and mixing conditions in Assumptions A.3-A.4 are not satisfied and the estimators might be inconsistent. A stationary and mixing framework can be recovered if we assume that the number n of operating firms in the portfolio is kept constant in time by replacing each defaulted firm by a new one, whose initial rating is randomly distributed across classes $k = 1, ..., K-1$ according to some distribution. This mechanism reflects the "static pool" definition of Standard & Poor's [see Brady and Bos (2002)]. Then, the methodology can be applied considering the model for the transitions between rating classes $k = 1, ..., K-1$ [see Gagliardini, Gouriéroux (2005b)]. For expository purpose, we do not consider an absorbing state here and refer to Gagliardini, Gouriéroux (2005b), Section 4.2, for more details.

model specifies the transition matrices with elements depending on the factor value:

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

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, when $G(x) = \Phi(x)$, and to the logistic distribution for the logit model, when $G(x) = 1/\left(1 + e^{-x}\right)$. Thus, we get a set of ordered probit or logit models with latent factors and common parameters, since the thresholds c_k appear in each row of the transition matrix. The ratios $a_{l,k,t} = \left(c_k - \gamma_l f_t - \alpha_l\right)/\sigma_l$ in the above transition probabilities identify semiparametrically the micro-parameters and the factor values up to location and scale transformations. Assumptions A.6-A.7 for semi-parametric identification are satisfied if 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 Appendix A.3). For instance, the vector of free micro-parameters is $\beta = (\alpha_l, \gamma_l, \sigma_l, l = 2, ..., K, c_k, k = 2, ..., K)$ when $K > 2$. Finally, we assume for illustration a single common factor f_t, which follows a linear Gaussian autoregressive process:

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

where (η_t) is $IIN(0, 1)$, and μ, ρ and σ are unknown macro-parameters.

6.2 Estimation of the micro-parameters

The micro log-density is given by:

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

The estimators of the factor values given β are:

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

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 two-step (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 assume a logistic function G. 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 $\beta = (\gamma_2, \alpha_2)'$. 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 of the two micro-parameters (solid lines), that are the components of $\hat{\beta}^*_{nT}$. 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 equation (6.3) by numerical optimization. To evaluate the profile micro-loglikelihood function for any given β, the estimate $\hat{f}_{n,t}(\beta)$ in equation (6.2) is computed by grid search. As expected from the stochastic migration literature, the γ_2 parameter, which represents the sensitivity of the transition probabilities with respect to the systematic 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 estimators. Finally, the latter estimators feature a rather small finite sample bias.

The semi-parametric efficiency bound for $\beta = (\gamma_2, \alpha_2)'$ is easily derived from Proposition 3 and is given by:

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

where $\pi_{12,t} = 1/(1+e^{-f_t})$, $\pi_{22,t} = 1/(1+e^{-\gamma_2 f_t - \alpha_2})$ and $\mu_{1,t-1} = P\left[y_{i,t-1} = 1|\underline{f_{t-1}}\right] = 1-\mu_{2,t-1}$. 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 μ, ρ, σ^2 by means of the expectation E_0.

Let us now study the pattern of the semi-parametric efficiency bound of parameter γ_2 as a function of the autoregressive coefficient ρ and the unconditional variance $\frac{\sigma^2}{1-\rho^2}$ of the factor process (f_t). Figure 3 displays the asymptotic standard deviation $\left(\frac{1}{nT}B^*_{\gamma_2\gamma_2}\right)^{1/2}$ as a function of these two macro-parameters, where $n = 1000$ and $T = 20$, and the semi-parametric efficiency bound $B^*_{\gamma_2\gamma_2}$ is approximated numerically by Monte-Carlo integration. The values of the micro-parameters and of μ are given in Table 1. The semi-parametric efficiency bound is decreasing w.r.t. the factor variance. The pattern is almost flat w.r.t. the autoregressive coefficient ρ of the factor, except for values of ρ close to 1, where the semi-parametric efficiency bound diverges to infinity.

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 equations (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 σ^2 is given by $\hat{\sigma}^2 = \dfrac{1}{T-1}\sum_{t=2}^{T} \hat{u}_t^2$, where $\hat{u}_t = \hat{f}^*_{nT,t} - \hat{\mu} - \hat{\rho}\hat{f}^*_{nT,t-1}$ are the OLS residuals. The estimator $\hat{\theta}^* = (\hat{\mu}, \hat{\rho}, \hat{\sigma}^2)'$ achieves the asymptotic efficiency bound with observable factor, that is, the Cramer-Rao bound for θ in the linear Gaussian model (6.1). Thus, the asymptotic efficiency bound is such that the estimators of $(\mu, \rho)'$ and σ^2 are asymptotically independent, root-T consistent, with asymptotic variance:

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

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

Figures 4 and 5 display the distributions (solid lines) of the efficient estimators $\hat{\mu}$, $\hat{\rho}$ and $\hat{\sigma}^2$ in the Monte-Carlo study for sample sizes $n = 200$, $T = 20$, and $n = 1000$, $T = 20$, respectively.

The parameter values are given in Table 1. We also display Gaussian distributions (dashed lines) centered at the true values of the parameters, with variances equal to the efficiency bounds divided by T. As expected, it is more difficult to estimate the autoregressive coefficient ρ and the variance σ^2 than to estimate the intercept μ. The estimators $\hat{\rho}$ and $\hat{\sigma}^2$ feature moderate downward biases. By comparing Figure 4 and Figure 5, we notice that the standard deviations of the estimators are rather similar for the two sample sizes and do not scale with n. Moreover, by comparing Figure 2 and Figure 5, it is seen that the discrepancy between the finite-sample distribution and the asymptotic efficiency bound is more pronounced for the macro-parameters than for the micro-parameters for our sample sizes. These findings are a consequence of the different convergence rates of the two types of estimators, that are $\sqrt{T}$ and $\sqrt{nT}$, respectively.

7 Concluding remarks

We have considered nonlinear dynamic panel models with common unobservable factors, in which it is possible to disentangle the micro- and the macro-dynamics, the latter ones being captured by the factor dynamics. Such models are often encountered in finance and insurance when the joint individual risks 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 extracting the business cycle from tendency surveys [Gouriéroux, Monfort (2009)], to disentangle inequality and mobility features in the dynamic analysis of income distributions, or to analyze longevity risk [e.g. Lee, Carter (1972), Schrager (2006), Gouriéroux, Monfort (2008)]. The considered specifications include both segment fixed effects and dynamic factors, but no individual fixed effects. For large cross-sectional and time dimensions $n, T \to \infty$, such that $T^\nu/n = O(1)$, $\nu > 1$, we have derived the semi-parametric efficiency bound of the parameter β characterizing the micro-dynamics. This semi-parametric efficiency bound takes into account the factor unobservability, and coincides with the bound for known factor transition. The efficiency bound for parameter θ characterizing the macro-dynamics is the same as if the factor were observable. Moreover, we have shown that the efficiency bound for (β, θ) can be reached by estimators that do not involve numerical integration w.r.t. the factor path and thus are easy to implement. These results require a large cross-sectional dimension to approximate the likelihood function by a closed form expression. When $T^\nu/n = O(1)$, $\nu > 3/2$, 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. For prediction purposes, it could be useful to include time-invariant observable individual characteristics x_i in the micro-density $h(y_{i,t}|y_{i,t-1}, x_i, f_t; \beta)$. The results in the paper can be easily extended to this case.

The condition $T^\nu/n = O(1)$, $\nu > 1$, implies that in our framework the incidental parameters problem does not induce a bias in the first-order asymptotic distribution of the estimators. An interesting venue for future research is to investigate the properties of the CSA, GA and true ML estimators, as well as of the two-step estimators, when T/n converges to a non-zero constant. This asymptotic scheme is common in the panel literature with individual fixed effects, which focuses on bias correction of the fixed effects estimator [see e.g. Woutersen (2002), Hahn, Newey (2004), Arellano, Hahn (2006), Bester, Hansen (2009), Hahn, Kuersteiner (2011) for analytical bias correction, and Hahn, Newey (2004), Dhaene, Jochmans, Thuysbaert (2006), Gourieroux, Phillips, Yu (2010) for bias correction by jackknife and indirect inference]. When $n, T \to \infty$ such that $T/n \to c$ (say), $c > 0$, it is possible to prove that the fixed effects estimator $\hat{\beta}^*_{nT}$, as well as the CSA and GA ML estimators of β are asymptotically normal, with variance-covariance matrix $(I_0^*)^{-1}$, and feature an asymptotic bias. Since the true ML estimator of β admits an interpretation as a random effects estimator (see Section 3), the results in Hahn, Kuersteiner, Cho (2005) and Arellano, Bonhomme (2009) suggest that the true ML estimator of parameter β could be first-order asymptotically unbiased when $T/n \to c$, $c > 0$. The proof of this conjecture is beyond the scope of the present paper.

References

[1] Amemiya, T. (1985): *Advanced Econometrics*, Harvard University Press.

[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.

[8] 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 Economic 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] Brady, B., and R., Bos (2004): "Record Default in 2001: The Result of Poor Credit Quality and a Weak Economy", Technical Report, Standard & Poor's.

[14] Campbell, J., Hilscher, J., and J., Szilagyi (2008): "Search for Distress Risk", Journal of Finance, 63, 2899-2939.

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

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

[17] Chava, S., and R., Jarrow (2004): "Bankruptcy Prediction with Industry Effects", Review of Finance 8, 537-569.

[18] Connor, G., Hagmann, M., and O., Linton (2012): "Efficient Estimation of the Fama-French Model and Extensions", Econometrica, 80, 713-754.

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

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

[21] 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.

[22] Dembo, A., Deuschel, T., and D., Duffie (2004): "Large Portfolio Losses", Finance and Stochastics, 8, 3-16.

[23] 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.

[24] Dhaene, G., Jochmans, K., and B., Thuysbaert (2006): "Jackknife Bias Reduction for Nonlinear Dynamic Panel Data Models with Fixed Effects", Working Paper, Leuven University.

[25] Dixit, A. (1990): *Optimization in Economic Theory*, Oxford University Press.

[26] 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.

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

[28] Duffie, D., and K., Singleton (1998): "Simulating Correlated Defaults", Working Paper, Stanford University.

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

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

[31] Forni, M., Hallin, M., Lippi, M., and L., Reichlin (2000): "The Generalized Dynamic Factor Model: Identification and Estimation", Review of Economics and Statistics, 82, 540-554.

[32] 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.

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

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

[35] Gagliardini, P., Gouriéroux, C., and A., Monfort (2012): "Microinformation, Nonlinear Filtering and Granularity", Journal of Financial Econometrics, 10, 1-53.

[36] Ghosh, J., and K., Subramanyam (1974): "Second-Order Efficiency of Maximum Likelihood Estimators", Sankya, Series A, 36, 325-358.

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

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

[39] Gouriéroux, C., and J., Jasiak (2012): "Granularity Adjustment for Default Risk Factor Model with Cohorts", Journal of Banking and Finance, 36, 1464-1477.

[40] Gouriéroux, C., and A., Monfort (2008): "Quadratic Stochastic Intensity and Prospective Mortality Tables", Insurance Mathematics and Economics, 43, 174-184.

[41] Gouriéroux, C., and A., Monfort (2009): "Granularity in Qualitative Factor Models", Journal of Credit Risk, 5, 29-61.

[42] Gouriéroux, C., Phillips, P., and J., Yu (2010): "Indirect Inference for Dynamic Panel Models", Journal of Econometrics, 157, 68-77.

[43] 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.

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

[45] Hahn, J., and G., Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large", Econometrica, 70, 1639-1657.

[46] Hahn, J., and G., Kuersteiner (2011): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects", Econometric Theory, 27, 1152-1191.

[47] Hahn, J., Kuersteiner, G., and M., Cho (2005): "Asymptotic Distribution of Misspecified Random Effects Estimators for a Dynamic Panel Model with Fixed Effects When Both n and T are Large", Economic Letters, 84, 117-125.

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

[49] Hall, P., and C., Heyde (1980): *Martingale Limit Theory and Its Application*, Academic Press, New York.

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

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

[52] Holly, A., and P., Phillips (1979): “A Saddlepoint Approximation of the Distribution of the k-Class Estimator of a Coefficient in a Simultaneous System”, Econometrica, 47, 1527-1547.

[53] Huber, P., Scaillet, O., and M.-P., Victoria-Feser (2009): “Assessing Multivariate Predictors of Financial Market Movements: A Latent Factor Framework for Ordinal Data”, Annals of Applied Statistics, 3, 249-271.

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

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

[56] 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.

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

[58] 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.

[59] Lee, R., and L., Carter (1972): “Modeling and Forecasting US Mortality”, Journal of the American Statistical Association, 87, 659-675.

[60] Loeffler, G. (2003): “The Effects of Estimation Error on Measures of Portfolio Credit Risk”, Journal of Banking and Finance, 27, 1427-1453.

[61] McNeil, A., and J., Wendin (2007): “Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk”, Journal of Empirical Finance, 14, 131-149.

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

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

[64] Pfanzagl, J., and W., Wefelmeyer (1978): “A Third-Order Optimum Property of the Maximum Likelihood Estimator”, Journal of Multivariate Analysis, 8, 1-29.

[65] Phillips, P. (1983): “Marginal Densities of Instrumental Variable Estimators in the General Single Equation Case”, Advances in Econometrics, 2, 1-24.

[66] Robinson, P. (1988): “The Stochastic Difference Between Econometric Estimators”, Econometrica, 56, 531-548.

[67] Schrager, D. (2006): “Affine Stochastic Mortality”, Insurance: Mathematics and Economics, 38, 87-97.

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

[69] Shumway, T. (2001): "Forecasting Bankruptcy More Accurately: A Simple Hazard Model", Journal of Business 74, 101-124.

[70] 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.

[71] 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.

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

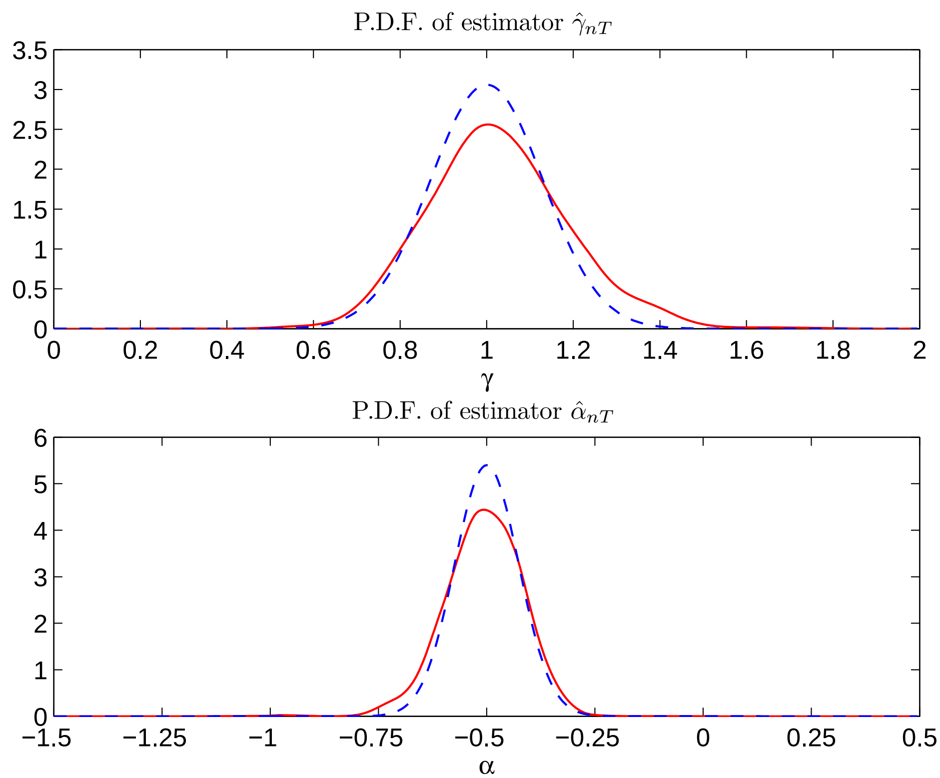
[73] Tierney, L., and Kadane, J. (1986): "Accurate Approximations for Posterior Moments and Marginal Densities", Journal of the American Statistical Association, 81, 82-86.

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

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

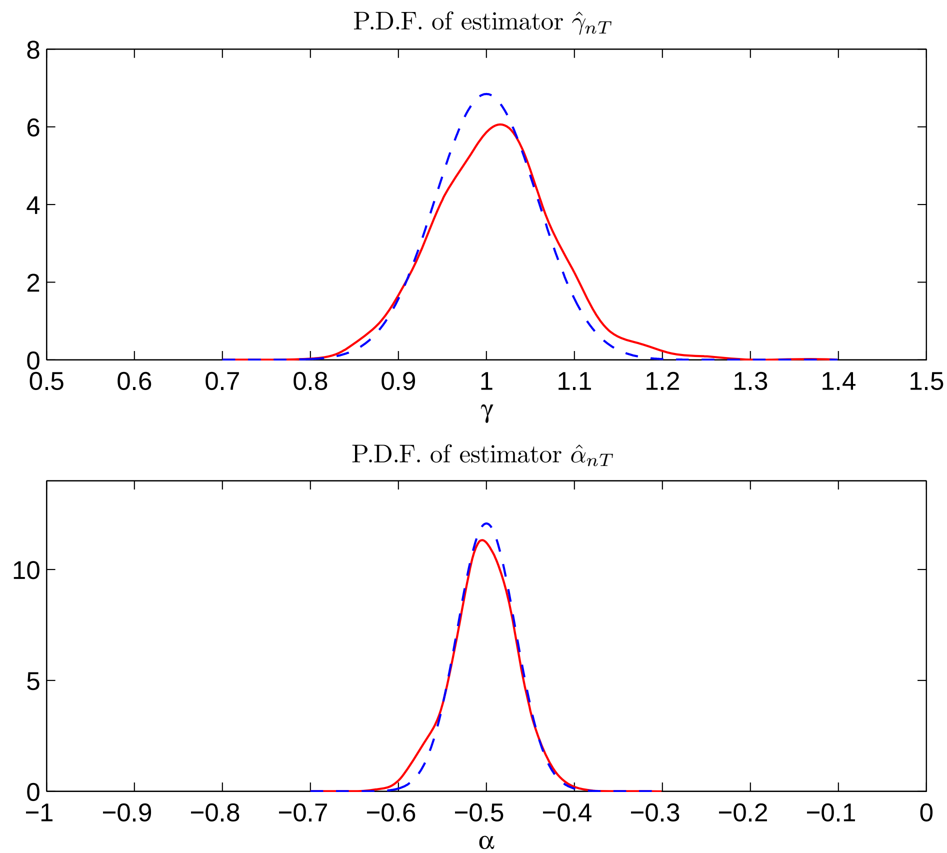
[76] Woutersen, T. (2002): "Robustness Against Incidental Parameters", Working Paper.

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



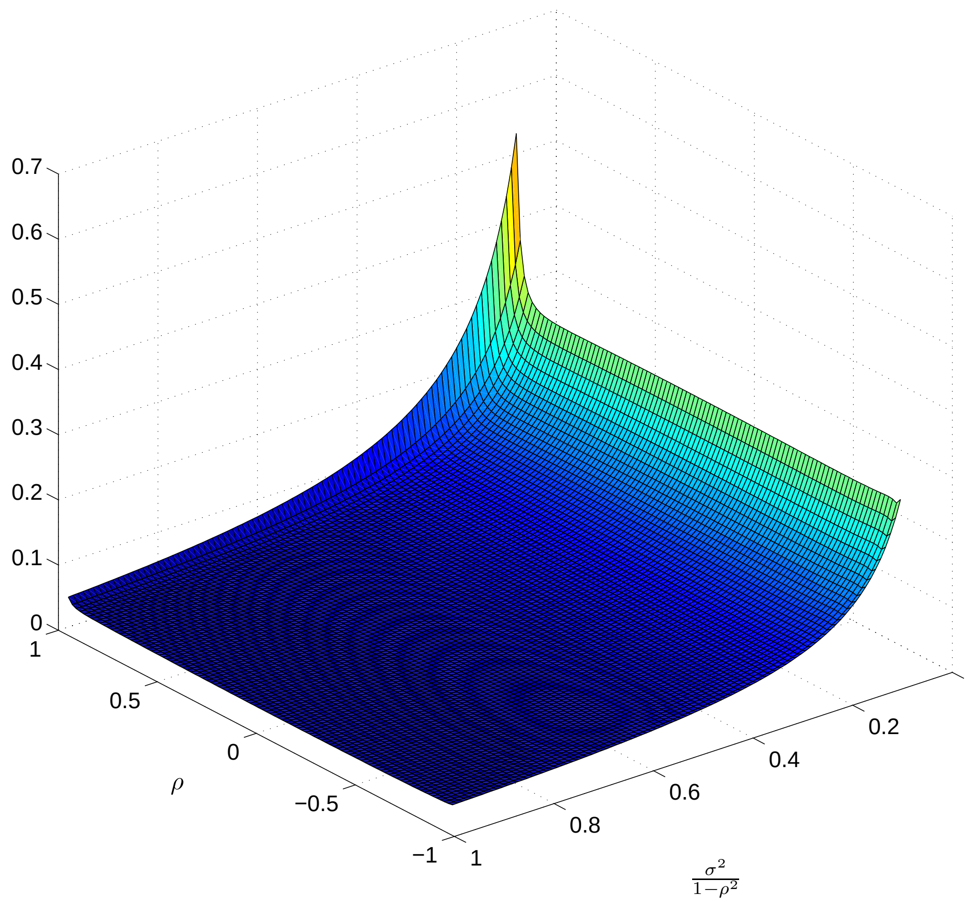
The solid lines give the pdf of the two-step semiparametrically efficient estimators of parameter γ (upper Panel, true value 1) and parameter α (lower Panel, true value -0.5). The pdf is computed by a kernel density estimator. Sample sizes are $n = 200$ and $T = 20$. The dashed lines in the two Panels 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 two-step semiparametrically efficient estimators of the micro-parameters, sample size $n = 1000$ and $T = 20$.



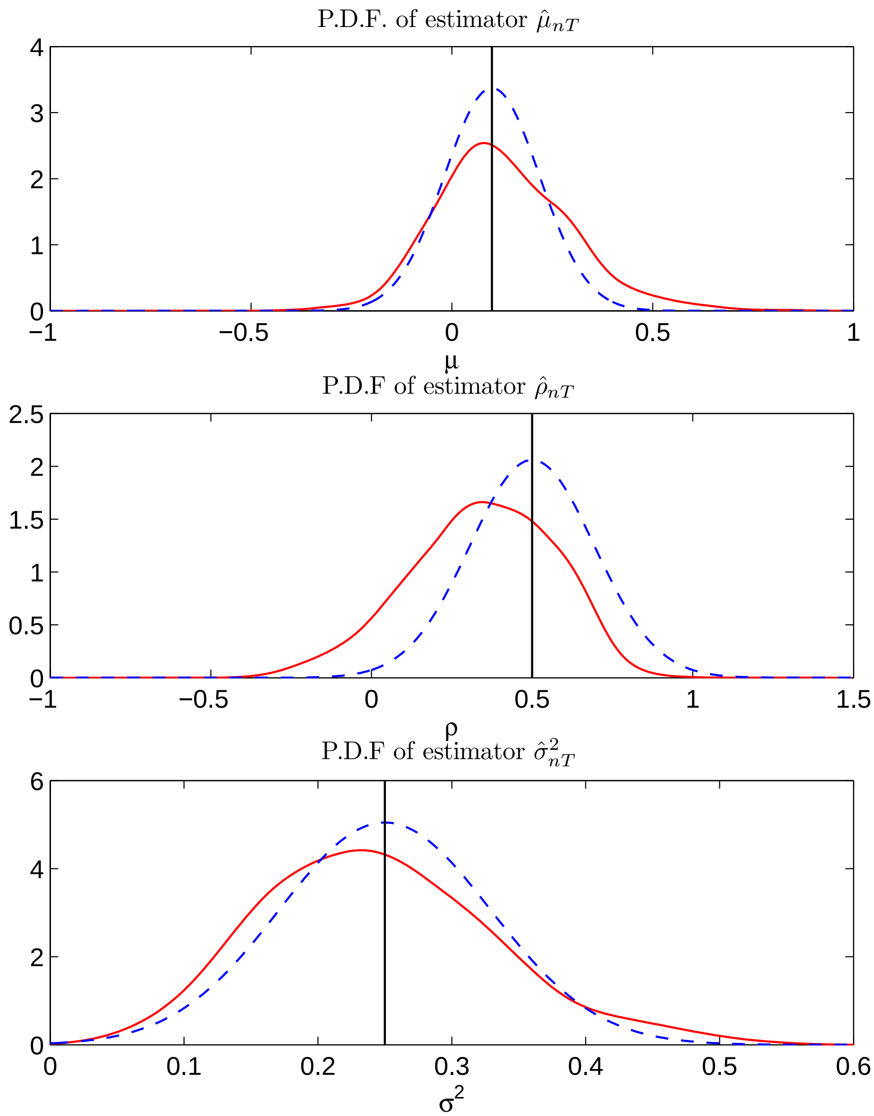
The solid lines give the pdf of the two-step semiparametrically efficient estimators of parameter γ (upper Panel, true value 1) and parameter α (lower Panel, true value -0.5). The pdf is computed by a kernel density estimator. Sample sizes are $n = 1000$ and $T = 20$. The dashed lines in the two Panels 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 γ_2.



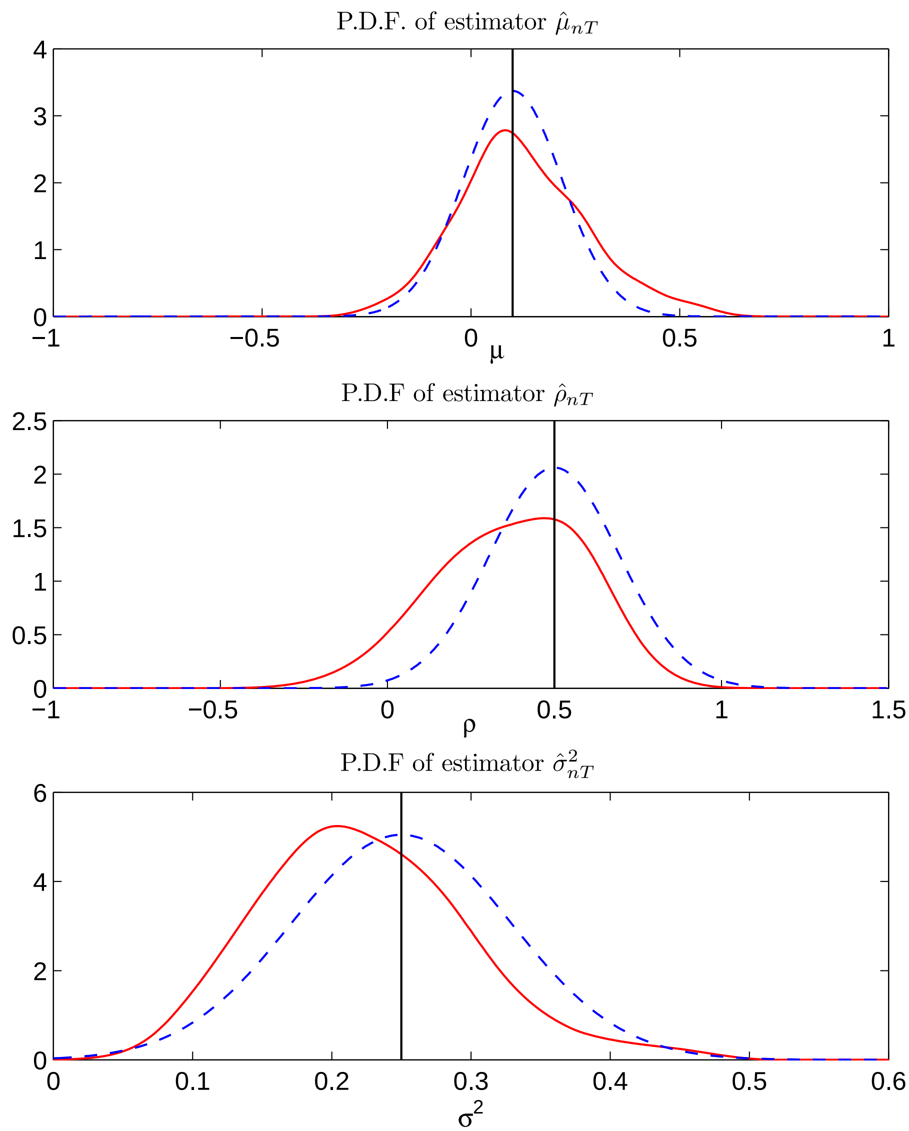
The figure displays $\left(\frac{1}{nT}B^*_{\gamma_2\gamma_2}\right)^{1/2}$, where $B^*_{\gamma_2\gamma_2}$ is the semiparametric efficiency bound for parameter γ_2 and $n = 1000, T = 20$, as a function of the autoregressive coefficient ρ and the variance $\frac{\sigma^2}{1-\rho^2}$ of the factor process (f_t).

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



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

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



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

APPENDIX A

In Appendix A.1 we provide the list of regularity conditions for the asymptotic analysis. The proofs of Propositions 1, 2, 3, 5 and 6 are given in Appendix A.2. They rely on Limit Theorems 1-3 and Lemmas 1-8, which are provided in the supplementary material. Appendix A.3 presents the proof of identification of the micro-parameters in the stochastic migration model. We denote by $\|A\|$ the Frobenius norm of matrix A. Moreover, b_i, c_i, d_i and γ_i, for $i = 1, 2, ...$, denote constants in the regularity conditions, while $C_1, C_2, ...$ denote generic constants used in the proofs.

A.1 Regularity conditions

In addition to Assumptions A.1-A.9, we use the regularity conditions given below to derive the large sample properties of the estimators. Due to the invariance of the true and approximate log-likelihood functions under one-to-one factor transformations $f \to \phi(f)$, the validity of Propositions 1, 2, 3 and 6 only requires that the regularity conditions are satisfied for a suitable choice of the factor process.

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

H.2: *The mapping $f \to E_0\left[\log h\left(y_{i,t}|y_{i,t-1}, f; \beta\right)|\underline{f_t}\right]$ defined on $\mathbb{R}^m$ admits a unique maximum, denoted by $f_t(\beta)$, for any parameter value $\beta \in \mathcal{B}$ and any factor path $\underline{f_t}$, $\mathbb{P}$-a.s. Moreover, $E_0\left[\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f_t}|\underline{f_t}\right] = 0$, and the matrix $E_0\left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f_t \partial f_t'}|\underline{f_t}\right]$ is positive definite, for any $\beta \in \mathcal{B}$ and any factor path $\underline{f_t}$, $\mathbb{P}$-a.s.*

H.3: *The micro-density is such that* (i) $\sup\{h(y_{i,t}|y_{i,t-1}, f_t; \beta) : y_{i,t}, y_{i,t-1} \in \mathbb{R}, f_t \in \mathbb{R}^m, \beta \in \mathcal{B}\} < \infty$, *and* (ii) $E_0\left[\sup_{\beta\in\mathcal{B}}\left|\frac{\partial^{|\alpha|} \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial^\alpha(\beta', f_t')'}\right|^8\right] < \infty$, *for any multi-index $\alpha \in \mathbb{N}^{q+m}$ with $|\alpha| \leq 3$.*

H.4: *For any $\beta \in \mathcal{B}$:* (i) *The pseudo-true factor value $f_t(\beta)$ is a measurable mapping of the factor path $\underline{f_t} \in \mathbb{R}^\infty$, where $\mathbb{R}^\infty$ denotes the set of real sequences, and measurability is defined w.r.t. the Borel field $\mathscr{B}(\mathbb{R}^\infty)$, i.e., the smallest sigma-field of subsets of $\mathbb{R}^\infty$ containing all finite-dimensional rectangles;*
(ii) *The cross-sectional log-likelihood $\mathcal{L}_t(\beta) = E_0[\log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)|\underline{f_t}]$ is measurable w.r.t. $\underline{f_t}$;*
(iii) *The Hessian matrix $I_t(\beta) = E_0\left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial(\beta', f_t')'\partial(\beta', f_t')}|\underline{f_t}\right]$ is measurable w.r.t. $\underline{f_t}$.*

H.5: $\mathbb{P}\left[\xi_{t,1} \geq u\right] \leq b_1 \exp\left(-c_1 u^{d_1}\right)$ *as $u \to \infty$, for some constants $b_1, c_1, d_1 > 0$, where $\xi_{t,1} = \max\{\xi_{t,1}^*, \xi_{t,1}^{**}\}$, with $\xi_{t,1}^* = \left(\inf_{\beta\in\mathcal{B}} \inf_{f\in\mathbb{R}^m:\|f-f_t(\beta)\|\leq\eta^*} \lambda_t(\beta, f)\right)^{-1}$, $\lambda_t(\beta, f) > 0$ denotes the smallest* *eigenvalue of the positive definite matrix* $I_t(\beta, f) \equiv E_0\left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial f \partial f'}|\underline{f_t}\right]$, $\eta^* > 0$, *and* $\xi^{**}_{t,1} = \sup_{\alpha \in \mathbb{N}^{q+m}:|\alpha|\leq 5} \sup_{\beta \in \mathcal{B}} E_0\left[\sup_{f \in \mathbb{R}^m:\|f - f_t(\beta)\| \leq \eta^*}\left|\frac{\partial^{|\alpha|}\log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial^{\alpha}(\beta', f')'}\right|^2 |\underline{f_t}\right]$.

H.6: *The process* $\xi_{t,2} = \sup_{\beta \in \mathcal{B}} \|f_t(\beta)\|$ *is such that* $\mathbb{P}\left[\xi_{t,2} \geq u\right] \leq b_2 \exp\left(-c_2 u^{d_2}\right)$ *as* $u \to \infty$, *for some constants* $b_2, c_2, d_2 > 0$.

H.7: *The set* $\mathcal{F}_n \subset \mathbb{R}^m$ *is* (i) *compact and convex, for any* $n \in \mathbb{N}$, *and such that* (ii) $B_{r_n}(0) \subset \mathcal{F}_n$, *where* $B_{r_n}(0)$ *denotes the open ball in* $\mathbb{R}^m$ *centered at* 0 *and with radius* $r_n = [(2/c_2)\log(n)]^{1/d_2}$ *and* (iii) $\mathcal{F}_n \subset B_{R_n}(0)$, *where* $R_n = O([\log(n)]^{\gamma_1})$ *for a constant* γ_1 *with* $\gamma_1 \geq 1/d_2$.

H.8: *There exists a constant* $\gamma_2 \geq 0$ *such that:*

$$\mathcal{K}_t \equiv \inf_{n\geq 1} \inf_{\beta \in \mathcal{B}} \inf_{f \in \mathcal{F}_n: f \neq f_t(\beta)} [\log(n)]^{\gamma_2} \frac{2KL_t(f, f_t(\beta);\beta)}{\|f - f_t(\beta)\|^2} > 0,$$

for any t, $\mathbb{P}$*-a.s., where* $KL_t(f, f_t(\beta);\beta) \equiv E_0\left[\log\left(\frac{h(y_{i,t}|y_{i,t-1}, f_t(\beta);\beta)}{h(y_{i,t}|y_{i,t-1}, f;\beta)}\right)|\underline{f_t}\right]$.

H.9: *There exists a constant* $\gamma_3 \geq 0$ *such that:*

$$\mathcal{R}_t \equiv \sup_{n\geq 1} [\log(n)]^{-\gamma_3} E_0\left[\sup_{\beta \in \mathcal{B}} \sup_{f \in \mathcal{F}_n}\left\|\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial(\beta', f')'}\right\|^4 |\underline{f_t}\right] < \infty,$$

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

H.10: $\mathbb{P}\left[\xi_{t,3} \geq u\right] \leq b_3 \exp\left[-c_3 u^{d_3}\right]$ *as* $u \to \infty$, *for some constants* $b_3, c_3, d_3 > 0$, *where* $\xi_{t,3} = \max\{\mathcal{K}_t^{-1}, \Gamma_t\}$, *with* $\Gamma_t \equiv \sup_{n\geq 1} \sup_{\beta \in \mathcal{B}} \sup_{f \in \mathcal{F}_n} [\log(n)]^{-\gamma_3} E_0\left[\left\|\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial f}\right\|^2 |\underline{f_t}\right]$.

H.11: (i) *There exists a constant* $\gamma_4 \geq 0$ *such that* $\mathbb{P}[\xi_{t,4} \geq u] \leq b_4 \exp\left(-c_4 u^{d_4}\right)$, *as* $u \to \infty$, *for some constants* $b_4, c_4, d_4 > 0$, *where* $\xi_{t,4} = \sup_{n\geq 1} \sup_{f \in \mathcal{F}_n} \sup_{\beta \in \mathcal{B}} [\log(n)]^{-\gamma_4} E_0\left[|\log h(y_{i,t}|y_{i,t-1}, f;\beta)|\,|\underline{f_t}\right]$. (ii) *There exists a constant* $\gamma_5 \geq 0$ *such that* $E_0\left[\sup_{\beta \in \mathcal{B}} \sup_{f \in \mathcal{F}_n} |\log h(y_{i,t}|y_{i,t-1}, f;\beta)|^4\right] = O\left([\log(n)]^{\gamma_5}\right)$ *and* $E_0\left[\sup_{\beta \in \mathcal{B}} \sup_{f \in \mathcal{F}_n}\left\|\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial(\beta', f')'}\right\|\right] = O\left([\log(n)]^{\gamma_5}\right)$. (iii) *Conditions* (i) *and* (ii) *are satisfied when replacing* $\log h(y_{i,t}|y_{i,t-1}, f;\beta)$ *by* $\frac{\partial^{|\alpha|}\log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial^{\alpha}(\beta', f')'}$, *for any multi-index* $\alpha \in \mathbb{N}^{q+m}$ *with* $|\alpha| \leq 5$.

H.12: *The stationary distribution* $\mathbb{P}_\theta$ *of Markov process* (f_t) *associated with the transition density* $g(f_t|f_{t-1};\theta)$ *is such that* $\sup_{\theta \in \Theta} \mathbb{P}_\theta[f_t \in \mathcal{F}_n^c] = O\left(e^{-\gamma_6 n^2}\right)$, *for a constant* $\gamma_6 > 0$.

H.13: *The function* $G(F_t;\theta) = \log g(f_t|f_{t-1};\theta)$*, where* $F_t = (f_t', f_{t-1}')'$*, is:* (i) *differentiable w.r.t.* $F_t \in \mathbb{R}^{2m}$ *and* $\theta \in \Theta$*, and such that* (ii) $E\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial G(F_t(\beta);\theta)}{\partial\theta}\right\|\right] < \infty$ *and* (iii) $\mathbb{P}\left[\xi_{t,5} \geq u\right] \leq b_5 \exp\left(-c_5 u^{d_5}\right)$*, as* $u \to \infty$*, for some constants* $b_5, c_5, d_5 > 0$*, where* $\xi_{t,5} = \sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\sup_{F\in\mathbb{R}^{2m}:\|F-F_t(\beta)\|\leq\eta^*}\left\|\frac{\partial^{|\alpha|}G(F;\theta)}{\partial F^\alpha}\right\|$, $\eta^* > 0$*, for any multi-index* $\alpha \in \mathbb{N}^{2m}$ *such that* $|\alpha| \leq 3$.

H.14: *Assumption H.13 is 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.15: *The macro-score is such that* $E_0\left[\left\|\frac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}\right\|^4\right] < \infty$.

Assumption H.1 is a standard condition on parameter sets and true parameter values. Assumptions H.2-H.5 concern the micro log-density and the pseudo-true factor values. Specifically, Assumption H.2 corresponds to the global and local identification conditions for the pseudo-true factor value $f_t(\beta)$ as maximizer of the asymptotic cross-sectional likelihood function. In Assumption H.3 (i) the micro-density is upper bounded, uniformly w.r.t. the factor value and micro-parameter. Assumption H.3 (ii) requires finite higher-order moments for $\log h(y_{i,t}|y_{i,t-1}, f;\beta)$ and its derivatives w.r.t. β and f, evaluated at $f = f_t(\beta)$, uniformly in $\beta \in \mathcal{B}$. The measurability conditions in Assumption H.4, together with Assumption A.3, are used to prove that the pseudo-true factor value $f_t(\beta)$, the cross-sectional log-likelihood $\mathcal{L}_t(\beta)$ and the Hessian matrix $I_t(\beta)$ are strictly stationary and ergodic processes, for any given value of the micro-parameter $\beta \in \mathcal{B}$. Assumption H.5 strengthens the local identification condition of the pseudo-true factor value in Assumption H.2. It requires that matrix $I_t(\beta, f)$ is positive definite for any factor value f in a neighborhood of $f_t(\beta)$, uniformly w.r.t. the micro-parameter $\beta \in \mathcal{B}$, and for any factor path $\underline{f_t}$, $\mathbb{P}$-a.s. Moreover, Assumption H.5 implies a tail condition on the stationary distribution of the positive process $\xi_{1,t}^*$. This condition is satisfied, when the factor paths associated with very small eigenvalues $\lambda_t(\beta, f)$, for some parameter value $\beta \in \mathcal{B}$ and factor value f close to $f_t(\beta)$, are sufficiently unfrequent. Assumptions H.5 also implies a tail condition for the stationary distribution of process $\xi_{t,1}^{**}$ involving higher-order derivatives of the micro log-density function. Assumptions H.1-H.5 are used to show that the Regularity Conditions RC.2 and RC.3 in Limit Theorem 3 are satisfied when proving the uniform convergence of the profile log-likelihood function $\mathcal{L}_{nT}^*(\beta)$, and of its second-order derivative matrix w.r.t. β [see Lemmas 1 (i) and 6 (1i) in the supplementary material].

Assumptions H.6, H.7 (i)-(ii), H.8-H.10 are used in Limit Theorem 1 to derive the uniform rate of convergence of the factor approximations. Specifically, Assumption H.6 concerns the tail of the stationary distribution of process $\sup_{\beta \in \mathcal{B}} \|f_t(\beta)\|$. Assumptions H.7 (ii) and (iii) introduce lower and upper bounds on the growth rate of set $\mathcal{F}_n$ as $n \to \infty$. These bounds are given in terms of expanding balls with radii of the order of powers of $\log(n)$. Under Assumptions H.6 and H.7 (ii), 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 (w.p.a.) 1 at rate $O(T/n^2)$. Assumption H.8 concerns the identifiability of the pseudo-true factor values from the asymptotic cross-sectional log-likelihood function. For any given micro-parameter value β and date t, the mapping $f \to KL_t(f, f_t(\beta); \beta)$ is a Kullback-Leibler divergence of the conditional p.d.f. $h(\cdot|\cdot, f; \beta)$ parametrized by $f \in \mathcal{F}_n$ from the pseudo-true conditional p.d.f. $h(\cdot|\cdot, f_t(\beta); \beta)$ given $\underline{f_t}$ under misspecification. From the global identification Assumption H.2, we have $KL_t(f, f_t(\beta); \beta) > 0$, for any factor value $f \neq f_t(\beta)$, parameter value β and date t, $\mathbb{P}$-a.s. Assumption H.8 strenghtens this condition by requiring that mapping $f \to KL_t(f, f_t(\beta); \beta)$ is lower bounded by a quadratic function proportional to the squared distance $\|f - f_t(\beta)\|^2$, uniformly in $\beta \in \mathcal{B}$, $f \in \mathcal{F}_n$ and $n \in \mathbb{N}$. The scale factor is allowed to converge to zero at most at a logarithmic rate, as set $\mathcal{F}_n$ increases. 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 conditional moment of order 4 is allowed to diverge at a logarithmic rate as $\mathcal{F}_n$ increases. Assumption H.10 is a tail condition on the stationary distribution of the processes $\mathcal{K}_t^{-1}$ and Γ_t. These processes correspond to the inverse of the measure $\mathcal{K}_t$ related to the conditional Kullback-Leibler discrepancy for cross-sectional factor approximation, and the measure Γ_t of second-order conditional moment of the score of the log-density w.r.t. f_t: they are both 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 Γ_t, are small.

Assumption H.11 introduces tail conditions and uniform bounds on conditional moments of the log micro-density, and of its derivatives w.r.t. factor f_t and parameter β. This assumption is used in Lemma 2 (see the supplementary material) to show the convergence in probability of the cross-sectional log-likelihood function, and of its derivatives w.r.t. the factor values, uniformly over the parameter value $\beta \in \mathcal{B}$, factor value $f \in \mathcal{F}_n$, and dates $1 \leq t \leq T$.

Assumptions H.12-H.15 concern the macro log-density and its derivatives w.r.t. factor values and macro-parameter θ. Specifically, Assumption H.12 requires that the tail of the stationary distribution of the factor process are sufficiently thin, uniformly w.r.t. the macro-parameter θ. This condition is used in Proposition A.2 (see Appendix A.2.1) to show that the contribution to the log-likelihood function coming from factor paths admitting some values outside set $\mathcal{F}_n$ is asymptotically negligible. Assumptions H.13 (i) and (ii) require that function $\log g(f_t|f_{t-1}; \theta)$ is differentiable w.r.t. the factor values and the macro-parameter θ, with uniformly finite expectation of the first-order derivative w.r.t. θ. Assumption H.13 (iii) is a condition on the tail of process $\xi_{t,5}$ involving the derivatives of $\log g(f_t|f_{t-1};\theta)$ w.r.t. the factor values. Assumption H.15 is a bound on the fourth-order moment of the macro-score $\dfrac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}$. Assumptions H.13-H.15 are used to show that Regularity Condition RC.1 in Limit Theorem 2 is satisfied when proving the convergence of $\mathcal{L}_{1,nT}(\beta,\theta)$, and of the Hessian $\dfrac{\partial^2\mathcal{L}_{1,nT}(\beta,\theta)}{\partial\theta\partial\theta'}$, uniformly in $\beta\in\mathcal{B}$, $\theta\in\Theta$ [see Lemmas 1 (ii) and 6 (1ii) in the supplementary material].

A.2 Proofs of the asymptotic results

A.2.1 Proof of Proposition 1

i) Preliminary expansions

Let us write the joint density in equation (3.2) as $l\left(\underline{y_T};\beta,\theta\right) = \int \exp\left[n\phi_{nT}(\boldsymbol{f}_T;\beta)\right] g_T(\boldsymbol{f}_T;\theta)d\boldsymbol{f}_T$, where $\boldsymbol{f}_T = (f_1', f_2', ..., f_T')' \in \mathbb{R}^{Tm}$, function ϕ_{nT} is defined by $\phi_{nT}(\boldsymbol{f}_T;\beta) = \sum_{t=1}^{T}\mathcal{L}_{n,t}(f_t;\beta)$, with $\mathcal{L}_{n,t}(f_t;\beta) = \dfrac{1}{n}\sum_{i=1}^{n}\log h(y_{i,t}|y_{i,t-1},f_t;\beta)$, and $g_T(\boldsymbol{f}_T;\theta) = \prod_{t=1}^{T} g(f_t|f_{t-1};\theta)$. Let $\varepsilon_n \downarrow 0$ be a sequence indexed by n, and let $B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))$ denote the open ball in $\mathbb{R}^{Tm}$ with radius ε_n centered in $\hat{\boldsymbol{f}}_{nT}(\beta) = (\hat{f}_{n,1}(\beta)', ..., \hat{f}_{n,T}(\beta)')'$. The integral in $l\left(\underline{y_T};\beta,\theta\right)$ can be decomposed as:

$$
\begin{aligned}
l\left(\underline{y_T};\beta,\theta\right) &= \int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))} \exp\left[n\phi_{nT}(\boldsymbol{f}_T;\beta)\right] g_T(\boldsymbol{f}_T;\theta)d\boldsymbol{f}_T \\
&\quad + \int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))^c\cap\mathcal{F}_{nT}} \exp\left[n\phi_{nT}(\boldsymbol{f}_T;\beta)\right] g_T(\boldsymbol{f}_T;\theta)d\boldsymbol{f}_T \\
&\quad + \int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))^c\cap\mathcal{F}_{nT}^c} \exp\left[n\phi_{nT}(\boldsymbol{f}_T;\beta)\right] g_T(\boldsymbol{f}_T;\theta)d\boldsymbol{f}_T, \qquad \text{(a.1)}
\end{aligned}
$$

where $\mathcal{F}_{nT} = \mathcal{F}_n \times \cdots \times \mathcal{F}_n \subset \mathbb{R}^{Tm}$ and $\mathcal{F}_n \subset \mathbb{R}^m$ is the sequence of sets involved in the definition of estimator $\hat{f}_{n,t}(\beta)$ [see equation (3.3)].

Let us consider the first integral in the RHS of equation (a.1). We apply the Laplace approximation method with an explicit expression for the remainder term.

PROPOSITION A.1. *We have:*

$$
\int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))} \exp\left[n\phi_{nT}(\boldsymbol{f}_T;\beta)\right] g_T(\boldsymbol{f}_T;\theta)d\boldsymbol{f}_T = \left(\frac{2\pi}{n}\right)^{Tm/2} \exp\left[nT\mathcal{L}^*_{nT}(\beta) + T\mathcal{L}_{1,nT}(\beta,\theta)\right]\Lambda_{nT}(\beta,\theta),
$$

where:

$$
\begin{aligned}
\Lambda_{nT}(\beta,\theta) = \frac{1}{(2\pi)^{Tm/2}} \int_{\mathcal{Z}_{nT}(\beta)} &\exp\left(-\frac{1}{2}\sum_{t=1}^{T} z_t' z_t\right) \\
\cdot \exp&\left[\sum_{t=1}^{T}\psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}}z_t, \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}}z_{t-1};\beta,\theta\right)\right] dz, \qquad \text{(a.2)}
\end{aligned}
$$

the function $\psi_{n,t}$ is defined by:

$$\begin{aligned}\psi_{n,t}(f_t, f_{t-1}; \beta, \theta) &= n[\mathcal{L}_{n,t}(f_t; \beta) - \mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta); \beta)] + \frac{n}{2}[f_t - \hat{f}_{n,t}(\beta)]'[I_{n,t}(\beta)][f_t - \hat{f}_{n,t}(\beta)] \\ &+ \log g(f_t | f_{t-1}; \theta) - \log g(\hat{f}_{n,t}(\beta) | \hat{f}_{n,t-1}(\beta); \theta), \qquad \text{(a.3)}\end{aligned}$$

and the integration domain is $\mathcal{Z}_{nT}(\beta) = \left\{ z = (z_1', ..., z_T')' \in \mathbb{R}^{Tm} : \sum_{t=1}^{T} z_t' I_{n,t}(\beta)^{-1} z_t \leq n\varepsilon_n^2 \right\}$.

Proof of Proposition A.1: By the definition of function $\psi_{n,t}$ in equation (a.3), we have:

$$\begin{aligned}\int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))} \exp\left[n\phi_{nT}(\boldsymbol{f}_T; \beta)\right] g_T(\boldsymbol{f}_T; \theta) d\boldsymbol{f}_T &= \prod_{t=1}^{T} \prod_{i=1}^{n} h\left(y_{i,t} | y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta\right) \prod_{t=1}^{T} g\left(\hat{f}_{n,t}(\beta) | \hat{f}_{n,t-1}(\beta); \theta\right) \\ &\cdot \int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))} \exp\left\{ \sum_{t=1}^{T} \left(\psi_{n,t}(f_t, f_{t-1}; \beta, \theta) - \frac{n}{2}[f_t - \hat{f}_{n,t}(\beta)]'[I_{n,t}(\beta)][f_t - \hat{f}_{n,t}(\beta)] \right) \right\} d\boldsymbol{f}_T.\end{aligned}$$

Let us introduce the change of variable from f_t to $z_t = \sqrt{n}[I_{n,t}(\beta)]^{1/2}[f_t - \hat{f}_{n,t}(\beta)]$, for $t = 1, ..., T$. Then, we get:

$$\begin{aligned}\int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))} \exp\left[n\phi_{nT}(\boldsymbol{f}_T; \beta)\right] g_T(\boldsymbol{f}_T; \theta) d\boldsymbol{f}_T &= \left(\frac{2\pi}{n}\right)^{Tm/2} \prod_{t=1}^{T} \left[\det I_{n,t}(\beta)\right]^{-1/2} \\ &\cdot \prod_{t=1}^{T} \prod_{i=1}^{n} h\left(y_{i,t} | y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta\right) \prod_{t=1}^{T} g\left(\hat{f}_{n,t}(\beta) | \hat{f}_{n,t-1}(\beta); \theta\right) \Lambda_{nT}(\beta, \theta). \qquad \text{(a.4)}\end{aligned}$$

By the definition of functions $\mathcal{L}_{nT}^*(\beta)$ and $\mathcal{L}_{1,nT}(\beta, \theta)$ in equations (3.7)-(3.8), the conclusion follows. ■

Let us now consider the next two terms in the RHS of equation (a.1). We bound these two terms at the beginning of the proof of Proposition A.2. The second integral in the RHS of equation (a.1) is asymptotically negligible for the expansion of the log-likelihood function in powers of $1/n$, if the sequence ε_n converges to zero slowly enough, namely if $\frac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$ for some $\mu_1 > 0$. This condition on sequence $\varepsilon_n = o(1)$ can be satisfied if $T^\nu/n = O(1)$, with $\nu > 1$. The third integral in the RHS of equation (a.1) is asymptotically negligible if the set $\mathcal{F}_n$ expands fastly enough as $n \to \infty$, whereas the tails of the factor distribution are not too heavy (see Assumption H.12).

PROPOSITION A.2. *Under Assumptions A.1-A.5 and H.1-H.13, if $T^\nu/n = O(1)$, for $\nu > 1$, and $\frac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, for $\mu_1 > 0$, then:*

$$\mathcal{L}_{nT}(\beta, \theta) = \mathcal{L}_{nT}^*(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta, \theta) + \frac{1}{nT} \log\left[\Lambda_{nT}(\beta, \theta) + o_p\left(n^{-\mu_2}\right)\right],$$

for any $\mu_2 > 0$, where the term $o_p\left(n^{-\mu_2}\right)$ is uniform w.r.t. $\beta \in \mathcal{B}$, $\theta \in \Theta$, and function $\Lambda_{nT}(\beta, \theta)$ is defined in equation (a.2).

Proof of Proposition A.2: (*) The second integral in the RHS of equation (a.1) is such that:

$$
\begin{aligned}
\int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))^c\cap\mathcal{F}_{nT}} \exp\left[n\phi_{nT}(\boldsymbol{f}_T;\beta)\right] g_T(\boldsymbol{f}_T;\theta)d\boldsymbol{f}_T &= \prod_{t=1}^{T}\prod_{i=1}^{n} h\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right) \\
&\quad\cdot\int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))^c\cap\mathcal{F}_{nT}} \exp\left(-n[\phi_{nT}(\hat{\boldsymbol{f}}_{nT}(\beta);\beta)-\phi_{nT}(\boldsymbol{f}_T;\beta)]\right) g_T(\boldsymbol{f}_T;\theta)d\boldsymbol{f}_T \\
\leq \prod_{t=1}^{T}\prod_{i=1}^{n} h\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right)\exp\left(-n\tau_{nT}(\beta)\right) &= \exp\left[nT\mathcal{L}^*_{nT}(\beta)-n\tau_{nT}(\beta)\right], \qquad \text{(a.5)}
\end{aligned}
$$

where:

$$
\tau_{nT}(\beta) = \inf_{\boldsymbol{f}_T\in B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))^c\cap\mathcal{F}_{nT}} [\phi_{nT}(\hat{\boldsymbol{f}}_{nT}(\beta);\beta)-\phi_{nT}(\boldsymbol{f}_T;\beta)]. \qquad \text{(a.6)}
$$

(**) The third integral in the RHS of equation (a.1) is such that:

$$
\begin{aligned}
\int_{B_{\varepsilon_n}(\hat{\boldsymbol{f}}_{nT}(\beta))^c\cap\mathcal{F}^c_{nT}} \exp\left[n\phi_{nT}(\boldsymbol{f}_T;\beta)\right] g_T(\boldsymbol{f}_T;\theta)d\boldsymbol{f}_T &\leq \bar{H}^{nT}\mathbb{P}_\theta[\boldsymbol{f}_T\in\mathcal{F}^c_{nT}] \leq \bar{H}^{nT}T\mathbb{P}_\theta\left[f_t\in\mathcal{F}^c_n\right] \\
&= O\left(T\bar{H}^{nT}e^{-\gamma_6 n^2}\right), \qquad \text{(a.7)}
\end{aligned}
$$

uniformly in $\beta\in\mathcal{B},\theta\in\Theta$, where $\bar{H}=\sup\{h(y_{i,t}|y_{i,t-1},f_t;\beta): y_{i,t},y_{i,t-1}\in\mathbb{R}, f_t\in\mathbb{R}^m,\beta\in\mathcal{B}\}<\infty$ [Assumption H.3 (i)] and $\gamma_6>0$ (Assumption H.12).

(***) Then, from equation (a.1), inequality (a.5), the bound in (a.7), and Proposition A.1, we get:

$$
l\left(\underline{y_T};\beta,\theta\right) = \left(\frac{2\pi}{n}\right)^{Tm/2}\exp\left[nT\mathcal{L}^*_{nT}(\beta)+T\mathcal{L}_{1,nT}(\beta,\theta)\right]\left[\Lambda_{nT}(\beta,\theta)+\Delta_{nT}(\beta,\theta)\right], \qquad \text{(a.8)}
$$

where:

$$
\begin{aligned}
0\leq\Delta_{nT}(\beta,\theta)\leq&\left(\frac{2\pi}{n}\right)^{-Tm/2}\exp\left[T|\mathcal{L}_{1,nT}(\beta,\theta)|\right] \\
&\cdot\left\{\exp\left[-n\tau_{nT}(\beta)\right]+\exp\left[nT(|\mathcal{L}^*_{nT}(\beta)|+C_1)-\gamma_6 n^2\right]\right\}, \qquad \text{(a.9)}
\end{aligned}
$$

for a constant $C_1>0$. To bound the RHS of inequality (a.9) we need the uniform asymptotic behaviour of functions $\mathcal{L}^*_{nT}(\beta)$ and $\mathcal{L}_{1,nT}(\beta,\theta)$. These functions involve mixtures of cross-sectional and time series aggregates. We prove in Lemma 1 in the supplementary material that $\mathcal{L}^*_{nT}(\beta)$ and $\mathcal{L}_{1,nT}(\beta,\theta)$ converge in probability to the corresponding population quantities $\mathcal{L}^*(\beta)=E_0[\log h(y_{i,t}|y_{i,t-1},f_t(\beta);\beta)]$ and:

$$
\mathcal{L}_1(\beta,\theta) = -\frac{1}{2}E_0\left[\log\det I_{t,ff}(\beta)\right]+E_0[\log g(f_t(\beta)|f_{t-1}(\beta);\theta)], \qquad \text{(a.10)}
$$

where $I_{t,ff}(\beta) = E_0\left[-\frac{\partial^2\log h\left(y_{i,t}|y_{i,t-1},f_t(\beta);\beta\right)}{\partial f\partial f'}|\underline{f_t}\right]$, uniformly in $\beta\in\mathcal{B}$, $\theta\in\Theta$. Moreover, $\sup_{\beta\in\mathcal{B}}|\mathcal{L}^*(\beta)|<\infty$ and $\sup_{\beta\in\mathcal{B},\theta\in\Theta}|\mathcal{L}_1(\beta,\theta)|<\infty$ from Assumptions H.1, H.3, H.5 and H.13. We deduce that:

$$
\sup_{\beta\in\mathcal{B}}\mathcal{L}^*_{nT}(\beta)=O_p(1), \qquad \sup_{\beta\in\mathcal{B},\theta\in\Theta}\mathcal{L}_{1,nT}(\beta,\theta)=O_p(1). \qquad \text{(a.11)}
$$

Let us now prove that:

$$\inf_{\beta\in\mathcal{B}} \tau_{nT}(\beta) \geq C_2 \frac{\varepsilon_n^2}{[\log(n)]^{C_3}}, \tag{a.12}$$

w.p.a. 1, for some constants $C_2, C_3 > 0$, where $\tau_{nT}(\beta)$ is defined in equation (a.6). We have:

$$\begin{aligned}
\inf_{\beta\in\mathcal{B}} \tau_{nT}(\beta) &= \inf_{\beta\in\mathcal{B}} \inf_{f_T\in B_{\varepsilon_n}(\hat{f}_{nT}(\beta))^c\cap\mathcal{F}_{nT}} \sum_{t=1}^{T} [\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta) - \mathcal{L}_{n,t}(f_t;\beta)] \\
&= \inf_{\beta\in\mathcal{B}} \inf_{f_T\in B_{\varepsilon_n}(\hat{f}_{nT}(\beta))^c\cap\mathcal{F}_{nT}} \sum_{t=1}^{T} \frac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta) - \mathcal{L}_{n,t}(f_t;\beta)}{\|\hat{f}_{n,t}(\beta) - f_t\|^2} \|\hat{f}_{n,t}(\beta) - f_t\|^2 \\
&\geq \left(\inf_{1\leq t\leq T} \inf_{\beta\in\mathcal{B}} \inf_{f_t\in\mathcal{F}_n} \frac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta) - \mathcal{L}_{n,t}(f_t;\beta)}{\|\hat{f}_{n,t}(\beta) - f_t\|^2} \right) \varepsilon_n^2.
\end{aligned}$$

In Lemma 2 in the supplementary material we prove that the term in the round brackets is lower bounded by $C_2[\log(n)]^{-C_3}$, w.p.a. 1, for some constants $C_2, C_3 > 0$. Then, the lower bound (a.12) follows.

From inequalities (a.9) and (a.12), the bounds in (a.11), and condition $\frac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, $\mu_1 > 0$, we get:

$$\begin{aligned}
\sup_{\beta\in\mathcal{B}} \sup_{\theta\in\Theta} \Delta_{nT}(\beta,\theta) &\leq \exp\left\{ -\frac{C_2 n\varepsilon_n^2}{[\log(n)]^{C_3}} \left[1 + O_p\left(\frac{T[\log(n)]^{C_3}}{n\varepsilon_n^2} \right) + O\left(\frac{T[\log(n)]^{C_3+1}}{n\varepsilon_n^2} \right) \right] \right\} \\
&\quad + \exp\left\{ -\gamma_6 n^2 \left[1 + O_p(T/n) \right] \right\} = o_p(n^{-\mu_2}),
\end{aligned}$$

for any $\mu_2 > 0$. By taking the log on equation (a.8), Proposition A.2 follows. ■

ii) CSA log-likelihood expansion [proof of Proposition 1 (i)]

In order to derive an expansion of the log-likelihood function at order $o_p(1/n)$ from Proposition A.2, we have to control the term $\Lambda_{nT}(\beta,\theta)$ uniformly in $\beta \in \mathcal{B}$, $\theta \in \Theta$. Since $\Lambda_{nT}(\beta,\theta)$ can take values both above, or below, 1, we need a uniform upper bound on the absolute value of $\log \Lambda_{nT}(\beta,\theta)$. Such a bound is provided next.

PROPOSITION A.3. *Under Assumptions A.1-A.5 and H.1-H.13, if $T^\nu/n = O(1)$, for $\nu > 1$, and $\frac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, for $\mu_1 > 0$, then:*

$$\mathcal{L}_{nT}(\beta,\theta) = \mathcal{L}_{nT}^*(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta,\theta) + \frac{1}{nT}\log\left[\Lambda_{nT}(\beta,\theta) + o_p\left(n^{-\mu_2}\right) \right],$$

and:

$$|\log(\Lambda_{nT}(\beta,\theta))| \leq C_4 T \varepsilon_n [\log(n)]^{C_5}, \tag{a.13}$$

uniformly in $\beta \in \mathcal{B}, \theta \in \Theta$, w.p.a. 1, for any $\mu_2 > 0$ and some constants $C_4, C_5 > 0$.

Proof of Proposition A.3: (*) Let us perform a Taylor expansion of function $\psi_{n,t}$ defined in equation (a.3) around $(f_t, f_{t-1}) = (\hat{f}_{n,t}(\beta), \hat{f}_{n,t-1}(\beta))$, and then use this expansion to derive an upper bound for term

$\psi_{n,t}\left(\hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2}z_t,\hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2}z_{t-1};\beta,\theta\right)$ in the RHS of equation (a.2). To simplify the notation, we consider the case $m=1$. We get for $z\in\mathcal{Z}_{nT}(\beta)$:

$$\begin{aligned}&\left|\psi_{n,t}\left(\hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2}z_t,\hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2}z_{t-1};\beta,\theta\right)\right|\\&\quad\leq\frac{1}{3!\sqrt{n}}\tilde{J}_{3,nt}(\beta)|z_t|^3+\frac{1}{\sqrt{n}}\tilde{D}_{10,nt}(\beta,\theta)|z_t|+\frac{1}{\sqrt{n}}\tilde{D}_{01,nt}(\beta,\theta)|z_{t-1}|,\end{aligned}\tag{a.14}$$

where $\tilde{J}_{3,nt}(\beta)=\sup_{f_t:|f_t-\hat{f}_{n,t}(\beta)|\leq\varepsilon_n}\left|\frac{\partial^3\mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^3}\right|\left|I_{n,t}(\beta)\right|^{-3/2}$ and $\tilde{D}_{pq,nt}(\beta,\theta)=\left|I_{n,t}(\beta)\right|^{-p/2}\left|I_{n,t-1}(\beta)\right|^{-q/2}$ $\cdot\sup_{f_t,f_{t-1}}\left\{\left|\frac{\partial^{p+q}\log g}{\partial f_t^p\partial f_{t-1}^q}(f_t|f_{t-1};\theta)\right|:|f_t-\hat{f}_{n,t}(\beta)|\leq\varepsilon_n,|f_{t-1}-\hat{f}_{n,t-1}(\beta)|\leq\varepsilon_n\right\}$, for $p+q=1$. We use Lemma 3 in the supplementary material to get upper bounds for the coefficients $\tilde{J}_{3,nt}(\beta)$, $\tilde{D}_{10,nt}(\beta,\theta)$, and $\tilde{D}_{01,nt}(\beta,\theta)$ in the RHS of inequality (a.14), uniformly in $\beta\in\mathcal{B}$, $\theta\in\Theta$ and $1\leq t\leq T$. By exploiting the tail conditions in Assumptions H.5 and H.13 (iii), and $T^\nu/n=O(1)$, $\nu>1$, the bounds diverge slowly with sample sizes n,T, namely as powers of $\log(n)$. More precisely, let us define the sequence:

$$\kappa_n=2[\log(n)/C_6]^{C_7},\quad n\in\mathbb{N},\tag{a.15}$$

where constants $C_6,C_7>0$ are such that $C_6\leq\min\{c_1,c_5\}$ and $C_7\geq\max\{3/d_1,2/d_5\}$, for c_1,d_1 and c_5,d_5 defined in Assumptions H.5 and H.13 (iii), respectively. If $z\in\mathcal{Z}_{nT}(\beta)$, we have $\|z\|^2\leq[\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}I_{n,t}(\beta)]n\varepsilon_n^2$. This implies $|z_t|\leq\|z\|\leq\sqrt{n}\varepsilon_n\kappa_n^{1/2}$ for any t, w.p.a. 1, since $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}I_{n,t}(\beta)\leq\kappa_n$ w.p.a. 1 from Lemma 3 (ii). Then, by Lemma 3 (iii-iv) and inequality (a.14), we get:

$$\begin{aligned}&\left|\sum_{t=1}^T\psi_{n,t}\left(\hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2}z_t,\hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2}z_{t-1};\beta,\theta\right)\right|\\&\leq\frac{1}{3!}\kappa_n^{3/2}\varepsilon_n\|z\|^2+2T\kappa_n^{3/2}\varepsilon_n,\end{aligned}\tag{a.16}$$

uniformly in $\beta\in\mathcal{B}$, $\theta\in\Theta$, w.p.a. 1.

(**) Let us now use inequality (a.16) to derive uniform upper and lower bounds for $\Lambda_{nT}(\beta,\theta)$, whose expression is given in equation (a.2).

a) Uniform upper bound. From inequality (a.16) we have (for $m=1$):

$$\begin{aligned}\Lambda_{nT}(\beta,\theta)&\leq\frac{e^{2T\kappa_n^{3/2}\varepsilon_n}}{(2\pi)^{T/2}}\int_{\mathbb{R}^T}\exp\left(-\frac{1}{2}\left(1-\frac{1}{3}\kappa_n^{3/2}\varepsilon_n\right)\|z\|^2\right)dz=\frac{e^{2T\kappa_n^{3/2}\varepsilon_n}}{\left(1-\frac{1}{3}\kappa_n^{3/2}\varepsilon_n\right)^{T/2}}\\&\sim\exp\left(\frac{13}{6}T\kappa_n^{3/2}\varepsilon_n\right),\end{aligned}\tag{a.17}$$

uniformly in $\beta\in\mathcal{B}$, $\theta\in\Theta$.

b) Uniform lower bound. If $\|z\|^2 \leq n\varepsilon_n^2 \inf_{1\leq t\leq T} \inf_{\beta\in\mathcal{B}} I_{n,t}(\beta)$, then $z \in \mathcal{Z}_{nT}(\beta)$. Moreover, from Lemma 3 (i) we have $\inf_{1\leq t\leq T} \inf_{\beta\in\mathcal{B}} I_{n,t}(\beta) \geq \kappa_n^{-1}$, w.p.a. 1. Thus, from (a.2) and (a.16) we get:

$$\begin{aligned}\Lambda_{nT}(\beta,\theta) &\geq \frac{e^{-2T\kappa_n^{3/2}\varepsilon_n}}{(2\pi)^{T/2}} \int_{\|z\|^2\leq n\varepsilon_n^2/\kappa_n} \exp\left(-\frac{1}{2}\left(1+\frac{1}{3}\kappa_n^{3/2}\varepsilon_n\right)\|z\|^2\right) dz \\ &= \frac{e^{-2T\kappa_n^{3/2}\varepsilon_n}}{(2\pi)^{T/2}} \int_0^{\sqrt{n\varepsilon_n^2/\kappa_n}} \int_{S^{T-1}} \exp\left(-\frac{1}{2}\left(1+\frac{1}{3}\kappa_n^{3/2}\varepsilon_n\right)r^2\right) r^{T-1} dz' dr,\end{aligned}$$

w.p.a. 1, where $r^{T-1}dz'dr$ is the integration element in spherical coordinates in dimension T and S^{T-1} denotes the unit sphere in dimension T. By using $\int_{S^{T-1}} dz' = \frac{2\pi^{T/2}}{\Gamma(T/2)}$ and the change of variable from r to $u = \frac{1}{2}\left(1+\frac{1}{3}\kappa_n^{3/2}\varepsilon_n\right)r^2$, we get:

$$\Lambda_{nT}(\beta,\theta) \geq \frac{e^{-2T\kappa_n^{3/2}\varepsilon_n}}{\left(1+\frac{1}{3}\kappa_n^{3/2}\varepsilon_n\right)^{T/2}} \frac{1}{\Gamma(T/2)} \int_0^{a_n} u^{T/2-1} \exp(-u)\, du,$$

where $a_n = \frac{1}{2}n\varepsilon_n^2\kappa_n^{-1}\left(1+\frac{1}{3}\kappa_n^{3/2}\varepsilon_n\right)$. The quantity $q_{nT} = \frac{1}{\Gamma(T/2)}\int_0^{a_n} u^{T/2-1}\exp(-u)\,du$ is the value at a_n of the cumulative distribution function (cdf) of a Gamma distribution $\gamma(T/2)$ with parameter $T/2$. Equivalently, $q_{nT} = \mathbb{P}[X_{nT} \leq 1]$, where the random variable X_{nT} is such that $a_n X_{nT} \sim \gamma(T/2)$. The moment generating function of X_{nT} is $M_{nT}(s) = E\left[\exp(-sX_{nT})\right] = \left(1+\frac{s}{a_n}\right)^{-T/2}$, for $s \in \mathbb{R}_+$. Thus, $M_{nT}(s) \sim \exp\left(-\frac{Ts}{2a_n}\right) \to 1$, as $n, T \to \infty$, for any $s \in \mathbb{R}_+$, since $T/a_n = o(1)$ from the condition $\frac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, $\mu_1 > 0$. Thus, X_{nT} converges in distribution to the constant 1, as $n, T \to \infty$. This implies $q_{nT} = 1 + o(1)$. Thus, we get:

$$\Lambda_{nT}(\beta,\theta) \geq \frac{e^{-2T\kappa_n^{3/2}\varepsilon_n}}{\left(1+\frac{1}{3}\kappa_n^{3/2}\varepsilon_n\right)^{T/2}}(1+o(1)) \sim \exp\left(-\frac{13}{6}T\kappa_n^{3/2}\varepsilon_n\right), \tag{a.18}$$

uniformly in $\beta \in \mathcal{B}$, $\theta \in \Theta$, w.p.a. 1. From bounds (a.17)-(a.18), and the expression of κ_n in (a.15), the upper bound (a.13) in Proposition A.3 follows. ■

To prove the CSA expansion in Proposition 1 (i), we use Proposition A.3. If $n, T \to \infty$ such that $T^\nu/n = O(1)$, for $\nu > 1$, then there exists a sequence $\varepsilon_n \downarrow 0$ such that $\frac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, for some $\mu_1 > 0$, and $\varepsilon_n[\log(n)]^{C_5} = o(1)$, for constant C_5 of Proposition A.3. Thus, from Proposition A.3 and equation (a.8), we deduce that equation (3.6) holds with $\Psi_{nT}(\beta,\theta) = \frac{1}{nT}\log\left[\Lambda_{nT}(\beta,\theta) + \Delta_{nT}(\beta,\theta)\right]$, where $\Delta_{nT}(\beta,\theta) \geq 0$, $\Delta_{nT}(\beta,\theta) = o_p(n^{-\mu_2})$, for any $\mu_2 > 0$, and $|\log\Lambda_{nT}(\beta,\theta)| = o_p(T)$, uniformly in $\beta \in \mathcal{B}$, $\theta \in \Theta$. Then, from the monotonicity of the logarithm, we have w.p.a. 1:

$$\begin{aligned}\Psi_{nT}(\beta,\theta) &\leq \frac{1}{nT}\max\left\{\log[2\Lambda_{nT}(\beta,\theta)], \log[2\Delta_{nT}(\beta,\theta)]\right\} \\ &\leq O\left(\frac{1}{nT}\right) + \frac{1}{nT}\max\left\{\log[\Lambda_{nT}(\beta,\theta)], 0\right\} = o_p(1/n),\end{aligned}$$

and $\Psi_{nT}(\beta,\theta) \geq \dfrac{1}{nT}\log[\Lambda_{nT}(\beta,\theta)] = o_p(1/n)$, uniformly in $\beta \in \mathcal{B}$ and $\theta \in \Theta$. Proposition 1 (i) follows.

iii) GA log-likelihood expansion [proof of Proposition 1 (ii)]

In order to derive an expansion of the log-likelihood function at order $o_p(1/n^2)$, we need a more accurate analysis of the term $\Lambda_{nT}(\beta,\theta)$ compared to Proposition A.3. A uniform asymptotic expansion for $\Lambda_{nT}(\beta,\theta)$ at order $o_p(T/n)$ is provided in Proposition A.4 below under an additional condition on the convergence rate of sequence ε_n, namely $\sqrt{T}\varepsilon_n^2 = O(n^{-\mu_3})$, with $\mu_3 > 0$. This condition is compatible with condition $\dfrac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, with $\mu_1 > 0$, if $n, T \to \infty$ such that $T^\nu/n = O(1)$, with $\nu > 3/2$.

PROPOSITION A.4. *Under Assumptions A.1-A.5 and H.1-H.13, if* $T^\nu/n = O(1)$*, with* $\nu > 3/2$*, and if* ε_n *is such that* $\dfrac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$ *and* $\sqrt{T}\varepsilon_n^2 = O(n^{-\mu_3})$*, for some* $\mu_1, \mu_3 > 0$*, then:*

$$\mathcal{L}_{nT}(\beta,\theta) \quad = \quad \mathcal{L}^*_{nT}(\beta) + \frac{1}{n}\mathcal{L}_{1,nT}(\beta,\theta) + \frac{1}{nT}\log\left[\Lambda_{nT}(\beta,\theta) + o_p\left(n^{-\mu_2}\right)\right],$$

for any $\mu_2 > 0$*, and:*

$$\Lambda_{nT}(\beta,\theta) = 1 + \frac{T}{n}\mathcal{L}_{2,nT}(\beta,\theta) + o_p(T/n), \tag{a.19}$$

uniformly in $\beta \in \mathcal{B}$*,* $\theta \in \Theta$*.*

Proof of Proposition A.4: We perform a Taylor expansion of function $\psi_{n,t}$ in (a.3) around $(f_t, f_{t-1}) = (\hat{f}_{n,t}(\beta), \hat{f}_{n,t-1}(\beta))$. The expansion is of fifth-order for the part of the function in the RHS of the first line in equation (a.3), and of third order for the part of the function in the second line in equation (a.3), so that the remainder term involves a power $n^{-3/2}$. To simplify the notation, we consider the case $m = 1$. We get:

$$\begin{aligned}
&\psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2} z_t, \hat{f}_{n,t-1}(\beta) + \frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2} z_{t-1}; \beta, \theta\right) \\
&\quad = \frac{1}{3!\sqrt{n}}J_{3,nt}(\beta)z_t^3 + \frac{1}{4!n}J_{4,nt}(\beta)z_t^4 + \frac{1}{\sqrt{n}}D_{10,nt}(\beta,\theta)z_t + \frac{1}{\sqrt{n}}D_{01,nt}(\beta,\theta)z_{t-1} \\
&\quad + \frac{1}{2n}D_{20,nt}(\beta,\theta)z_t^2 + \frac{1}{2n}D_{02,nt}(\beta,\theta)z_{t-1}^2 + \frac{1}{n}D_{11,nt}(\beta,\theta)z_t z_{t-1} + R_{n,t}(z_t, z_{t-1}; \beta, \theta),
\end{aligned} \tag{a.20}$$

where the remainder term is such that:

$$|R_{n,t}(z_t, z_{t-1}; \beta, \theta)| \leq \frac{1}{5!n^{3/2}}\tilde{J}_{5,nt}(\beta)|z_t|^5 + \frac{1}{3!n^{3/2}}\sum_{j=0}^{3}\binom{3}{j}\tilde{D}_{3-j,j,nt}(\beta,\theta)|z_t|^{3-j}|z_{t-1}|^j, \tag{a.21}$$

with $\tilde{J}_{5,nt}(\beta) = \sup\limits_{f_t : |f_t - \hat{f}_{n,t}(\beta)| \leq \varepsilon_n}\left|\dfrac{\partial^5 \mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^5}\right| |I_{n,t}(\beta)|^{-5/2}$ and $\tilde{D}_{pq,nt}(\beta,\theta) = |I_{n,t}(\beta)|^{-p/2}|I_{n,t-1}(\beta)|^{-q/2}$ $\cdot \sup\limits_{f_t, f_{t-1}}\left\{\left|\dfrac{\partial^{p+q}\log g}{\partial f_t^p \partial f_{t-1}^q}(f_t|f_{t-1};\theta)\right| : |f_t - \hat{f}_{n,t}(\beta)| \leq \varepsilon_n, |f_{t-1} - \hat{f}_{n,t-1}(\beta)| \leq \varepsilon_n\right\}$, for $p + q = 3$. Let us write the exponential $\exp\left(\sum\limits_{t=1}^{T}\psi_{n,t}\right)$ in equation (a.2) as a series, and interchange the series and the integral by applying the Lebesgue theorem on the bounded domain $\mathcal{Z}_{nT}(\beta)$. We get $\Lambda_{nT}(\beta,\theta) = \sum_{j=0}^{\infty} \frac{1}{j!}\Lambda_{j,nT}(\beta,\theta)$, where:

$$\Lambda_{j,nT}(\beta,\theta) = \frac{1}{(2\pi)^{T/2}} \int_{\mathcal{Z}_{nT}(\beta)} \exp\left(-\frac{1}{2}\|z\|^2\right)$$
$$\cdot \left[\sum_{t=1}^{T} \psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}} z_t, \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}} z_{t-1}; \beta,\theta\right)\right]^j dz. \quad \text{(a.22)}$$

We analyze the terms $\Lambda_{j,nT}(\beta,\theta)$, for $j = 0, 1, ...$, separately. By replacing expansion (a.20) into equation (a.22), we show below that:

$$\Lambda_{0,nT}(\beta,\theta) = 1 + o_p(T/n), \quad \text{(a.23)}$$

$$\Lambda_{1,nT}(\beta,\theta) = \frac{1}{8n}\sum_{t=1}^{T} J_{4,nt}(\beta) + \frac{1}{2n}\sum_{t=1}^{T} D_{20,nt}(\beta,\theta) + \frac{1}{2n}\sum_{t=2}^{T} D_{02,nt}(\beta,\theta) + o_p(T/n), \quad \text{(a.24)}$$

$$\Lambda_{2,nT}(\beta,\theta) = \frac{5}{12n}\sum_{t=1}^{T} J_{3,nt}(\beta)^2 + \frac{1}{n}\sum_{t=1}^{T} D_{10,nt}(\beta,\theta)^2 + \frac{1}{n}\sum_{t=2}^{T} D_{01,nt}(\beta,\theta)^2 + \frac{1}{n}\sum_{t=1}^{T} J_{3,nt}(\beta)D_{10,nt}(\beta,\theta)$$
$$+\frac{1}{n}\sum_{t=2}^{T} J_{3,n,t-1}(\beta)D_{01,nt}(\beta,\theta) + \frac{2}{n}\sum_{t=2}^{T} D_{10,n,t-1}(\beta,\theta)D_{01,nt}(\beta,\theta) + o_p\left(T/n\right), \quad \text{(a.25)}$$

and:

$$\sum_{j=3}^{\infty} \frac{1}{j!}|\Lambda_{j,nT}(\beta,\theta)| = o_p(T/n), \quad \text{(a.26)}$$

uniformly in $\beta \in \mathcal{B}, \theta \in \Theta$. By combining equations (a.23)-(a.26), equation (a.19) follows.

a) Proof of equivalence (a.23). We have $\Lambda_{0,nT}(\beta,\theta) = 1 - \frac{1}{(2\pi)^{T/2}} \int_{\mathcal{Z}_{nT}(\beta)^c} \exp\left(-\frac{1}{2}\|z\|^2\right) dz$. Let us derive an upper bound for the integral $\frac{1}{(2\pi)^{T/2}} \int_{\mathcal{Z}_{nT}(\beta)^c} \exp\left(-\frac{1}{2}\|z\|^2\right) z_t^{2k} dz$, with $k \in \mathbb{N}$ and $t = 1, ..., T$. If $z \in \mathcal{Z}_{nT}(\beta)^c$, we have $\|z\|^2 \geq [\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}} I_{n,t}(\beta)]n\varepsilon_n^2 \geq n\varepsilon_n^2\kappa_n^{-1}$, w.p.a. 1, from Lemma 3 (i). We get:

$$\frac{1}{(2\pi)^{T/2}} \int_{\mathcal{Z}_{nT}(\beta)^c} \exp\left(-\frac{1}{2}\|z\|^2\right) z_t^{2k} dz \leq \frac{1}{(2\pi)^{T/2}} \int_{\|z\|^2 \geq n\varepsilon_n^2\kappa_n^{-1}} \exp\left(-\frac{1}{2}\|z\|^2\right) z_t^{2k} dz$$
$$\leq \frac{1}{(2\pi)^{T/2}T} \int_{\sqrt{n\varepsilon_n^2\kappa_n^{-1}}}^{\infty} \int_{S^{T-1}} \exp(-r^2/2) r^{T+2k-1} dz' dr$$
$$= \frac{1}{T2^{T/2-1}\Gamma(T/2)} \int_{\sqrt{n\varepsilon_n^2\kappa_n^{-1}}}^{\infty} \exp(-r^2/2) r^{T+2k-1} dr, \quad \text{(a.27)}$$

uniformly in $1 \leq t \leq T$ and $\beta \in \mathcal{B}$, where we have used spherical coordinates as in the proof of Proposition A.3. By the change of variable from r to $u = \frac{1}{2}r^2$, we have:

$$\frac{1}{2^{T/2-1}\Gamma(T/2)} \int_{\sqrt{n\varepsilon_n^2\kappa_n^{-1}}}^{\infty} exp(-r^2/2) r^{T+2k-1} dr = \frac{2^k\Gamma(T/2+k)}{\Gamma(T/2)} \frac{1}{\Gamma(T/2+k)} \int_{\bar{a}_n}^{\infty} e^{-u} u^{T/2+k-1} du, \quad \text{(a.28)}$$

where $\bar{a}_n = \frac{1}{2}n\varepsilon_n^2\kappa_n^{-1}$. The RHS involves the survivor function of the Gamma distribution $\gamma(T/2+k)$ evaluated at $\bar{a}_n$. Since $\bar{a}_n \to \infty$ as $n \to \infty$, to upper bound the RHS of equation (a.28) it is enough to upper bound the cdf of the Gamma distribution in the right tail. By repeated partial integration, we get for any $s \geq 1$ and $\delta \geq 1$:

$$\begin{aligned}
&\frac{1}{\Gamma(\delta)}\int_s^\infty e^{-u}u^{\delta-1}du = \frac{e^{-s}s^{\delta-1}}{\Gamma(\delta)} + \frac{e^{-s}s^{\delta-2}}{\Gamma(\delta-1)} + \cdots + \frac{e^{-s}s^{l+1}}{\Gamma(l+2)} + \frac{1}{\Gamma(l+1)}\int_s^\infty e^{-u}u^l du \\
&\leq e^{-s}(s^{\delta-1} + s^{\delta-2} + \cdots + s^{l+1}) + \int_s^\infty e^{-u}udu \leq (\lfloor\delta\rfloor + 1)e^{-s}s^{\delta-1}, \qquad \text{(a.29)}
\end{aligned}$$

where $\lfloor\delta\rfloor$ denotes the integer part of δ and $l = \delta - \lfloor\delta\rfloor$ is the decimal part of δ. From inequality (a.27) and equation (a.28), and by using bound (a.29) with $s = \bar{a}_n$ and $\delta = T/2 + k$, we get:

$$\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\frac{1}{(2\pi)^{T/2}}\int_{\mathcal{Z}_{nT}(\beta)^c}\exp\left(-\frac{1}{2}\|z\|^2\right)z_t^{2k}dz \leq \frac{2^k\Gamma(T/2+k)}{\Gamma(T/2)}\frac{T/2+k+1}{T}e^{-\bar{a}_n}\bar{a}_n^{T/2+k-1}.$$

By the Stirling's formula, we have $\frac{\Gamma(T/2+k)}{\Gamma(T/2)} = O(T^k)$ for large T. Moreover, from condition $\frac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, $\mu_1 > 0$, we have:

$$e^{-\bar{a}_n}\bar{a}_n^{T/2+k-1} = \exp\left\{-\frac{n\varepsilon_n^2}{2\kappa_n}\left[1 + o\left(\frac{T\kappa_n\log(n)}{n\varepsilon_n^2}\right)\right]\right\} \leq \exp\left(-\frac{n\varepsilon_n^2}{4\kappa_n}\right) = o(n^{-\mu_4}),$$

for any $\mu_4 > 0$. Thus, we get for any $k \in \mathbb{N}$:

$$\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\frac{1}{(2\pi)^{T/2}}\int_{\mathcal{Z}_{nT}(\beta)^c}\exp\left(-\frac{1}{2}\|z\|^2\right)z_t^{2k}dz = o_p\left(n^{-\mu_4}\right), \qquad \text{(a.30)}$$

for any $\mu_4 > 0$. In particular, equivalence (a.23) follows.

b) Proof of equivalence (a.24). By the symmetry of the domain of integration $\mathcal{Z}_{nT}(\beta)$ we have:

$$\begin{aligned}
\Lambda_{1,nT}(\beta,\theta) &= \frac{1}{4!n}\sum_{t=1}^T J_{4,nt}(\beta)a_{2,nT,t}(\beta) + \frac{1}{2n}\sum_{t=1}^T D_{20,nt}(\beta,\theta)a_{1,nT,t}(\beta) \\
&+\frac{1}{2n}\sum_{t=2}^T D_{02,nt}(\beta,\theta)a_{1,nT,t-1}(\beta) + \sum_{t=1}^T\frac{1}{(2\pi)^{T/2}}\int_{\mathcal{Z}_{nT}(\beta)}\exp\left(-\frac{1}{2}\|z\|^2\right)R_{n,t}(z_t,z_{t-1};\beta,\theta)dz,
\end{aligned} \qquad \text{(a.31)}$$

where we use the notation $a_{k,nT,t}(\beta) = \frac{1}{(2\pi)^{T/2}}\int_{\mathcal{Z}_{nT}(\beta)}\exp\left(-\frac{1}{2}\|z\|^2\right)z_t^{2k}dz$. To control the RHS of equation (a.31), we use Lemma 4 in the supplementary material, which provides uniform upper bounds for terms $J_{p,nt}(\beta)$ and $D_{pq,nt}(\beta,\theta)$ involving higher-order partial derivatives w.r.t. the factor values. From inequality (a.21) and Lemma 4, the last term in the RHS of equation (a.31) is $O_p\left(\frac{T\kappa_n}{n^{3/2}}\right) = o_p(T/n)$, uniformly in $\beta \in \mathcal{B}, \theta \in \Theta$, where sequence κ_n is defined in (a.15). By using the bound in (a.30), we have $a_{2,nT,t} = 3 + o_p(n^{-\mu_5})$ and $a_{1,nT,t} = 1 + o_p(n^{-\mu_5})$, uniformly in $t = 1, ..., T$ and $\beta \in \mathcal{B}$, for any $\mu_5 > 0$. Then, from equation (a.31) and Lemma 4 in the supplementary material, we get equivalence (a.24).

c) Proof of equivalence (a.25). By the symmetry of domain $\mathcal{Z}_{nT}(\beta)$, we have:

$$\Lambda_{2,nT}(\beta,\theta) = \frac{1}{(3!)^2 n}\sum_{t=1}^{T} J_{3,nt}(\beta)^2 a_{3,nT,t}(\beta) + \frac{1}{n}\sum_{t=1}^{T} D_{10,nt}(\beta,\theta)^2 a_{1,nT,t}(\beta)$$
$$+\frac{1}{n}\sum_{t=2}^{T} D_{01,nt}(\beta,\theta)^2 a_{1,nT,t-1}(\beta) + \frac{2}{3!n}\sum_{t=1}^{T} J_{3,nt}(\beta) D_{10,nt}(\beta,\theta) a_{2,nT,t}(\beta)$$
$$+\frac{2}{3!n}\sum_{t=2}^{T} J_{3,n,t-1}(\beta) D_{01,nt}(\beta,\theta) a_{2,nT,t-1}(\beta) + \frac{2}{n}\sum_{t=2}^{T} D_{10,n,t-1}(\beta,\theta) D_{01,nt}(\beta,\theta) a_{1,nT,t-1}(\beta) + O_p\left(\frac{T^2\kappa_n^2}{n^2}\right).$$

From equation (a.30) we get $a_{3,nT,t}(\beta) = 15 + o_p(n^{-\mu_6})$ uniformly, for any $\mu_6 > 0$. Then, from Lemmas 3 and 4 in the supplementary material, equivalence (a.25) follows.

d) Proof of equivalence (a.26). We use Lemma 5 in the supplementary material, which provides the following uniform upper bounds for $\Lambda_{j,nT}(\beta,\theta)$, for any integer $j \geq 3$:

$$\Lambda_{j,nT}(\beta,\theta) \leq C_j^* \left(\frac{T^2\kappa_n^j}{n^2}\right), \tag{a.32}$$

and:

$$\Lambda_{j,nT}(\beta,\theta) \leq C_8 \kappa_n^{2j} j! \left(\frac{T}{n} + \sqrt{T}\varepsilon_n^2\right)^j, \tag{a.33}$$

uniformly in $\beta \in \mathcal{B}, \theta \in \Theta$, w.p.a. 1, for some constants $C_j^* > 0$, $j = 3, 4, ...$, and $C_8 > 0$, and where sequence κ_n is defined in (a.15). The sequence of constants C_j^* in bound (a.32) diverges rapidly as j increases, and the sequence $C_j^*\kappa_n^j/j!$ might not be summable. This explains why, for any given $J \geq 3$ independent of n and T, we use the bound in (a.32) for $j \leq J$ and the bound in (a.33) for $j > J$, to get w.p.a. 1:

$$\sum_{j=3}^{\infty}\frac{1}{j!}|\Lambda_{j,nT}(\beta,\theta)| \leq \sum_{j=3}^{J} C_j^* \frac{T^2\kappa_n^j}{j! n^2} + \sum_{j=J+1}^{\infty} C_8\kappa_n^{2j}\left(\frac{T}{n} + \sqrt{T}\varepsilon_n^2\right)^j$$
$$= \sum_{j=3}^{J} C_j^* \frac{T^2\kappa_n^j}{j! n^2} + C_8\frac{\rho_{nT}^{J+1}}{1-\rho_{nT}} = o_p(T/n) + O_p\left(\rho_{nT}^{J+1}\right),$$

uniformly in $\beta \in \mathcal{B}$, $\theta \in \Theta$, where $\rho_{nT} = \kappa_n^2\left(\frac{T}{n} + \sqrt{T}\varepsilon_n^2\right) = o(n^{-\mu_7})$, for any μ_7 such that $0 < \mu_7 < \min\{\mu_3, 1 - 1/\nu\}$, if $T^\nu/n = O(1)$, for $\nu > 3/2$, and $\sqrt{T}\varepsilon_n^2 = O(n^{-\mu_3})$, $\mu_3 > 0$. If we choose $J \geq \max\{3, 1/\mu_7 - 1\}$, we get $\rho_{nT}^{J+1} = o(n^{-1})$, which implies equation (a.26). ■

From Lemmas 3 and 4 in the supplementary material, we have that $\frac{T}{n}\mathcal{L}_{2,nT}(\beta,\theta) = o_p(1)$, uniformly in $\beta \in \mathcal{B}$, $\theta \in \Theta$. Then, from Proposition A.4 and the expansion of the logarithm in a neighbourhood of 1, Proposition 1 (ii) follows.

A.2.2 Proof of Proposition 2

The proof is in two steps. We first show the consistency of the estimators, which is then used to derive the stochastic difference between the estimators.

i) Consistency of the estimators

Let us prove the consistency of the estimators when $n, T \to \infty$ such that $T^\nu/n = O(1)$, $\nu > 1$. We start with the ML estimator $(\tilde{\beta}_{nT}, \tilde{\theta}_{nT})$. Let us first prove that $\tilde{\beta}_{nT}$ is consistent. For any $\varepsilon > 0$ we have:

$$\begin{aligned}\mathbb{P}\left[\|\tilde{\beta}_{nT} - \beta_0\| \geq \varepsilon\right] &\leq \mathbb{P}\left[\sup_{\beta \in \mathcal{B}: \|\beta - \beta_0\| \geq \varepsilon} \mathcal{L}_{nT}(\beta, \tilde{\theta}_{nT}) \geq \mathcal{L}_{nT}(\tilde{\beta}_{nT}, \tilde{\theta}_{nT})\right] \\ &\leq \mathbb{P}\left[\sup_{\beta \in \mathcal{B}: \|\beta - \beta_0\| \geq \varepsilon} \mathcal{L}_{nT}(\beta, \tilde{\theta}_{nT}) \geq \mathcal{L}_{nT}(\beta_0, \theta_0)\right].\end{aligned}$$

By using Proposition 1 (i), Lemma 1 (i) in the supplementary material, and the second bound in (a.11), we get:

$$\mathbb{P}\left[\|\tilde{\beta}_{nT} - \beta_0\| \geq \varepsilon\right] \leq \mathbb{P}\left[\sup_{\beta \in \mathcal{B}: \|\beta - \beta_0\| \geq \varepsilon} \mathcal{L}^*(\beta) - \mathcal{L}^*(\beta_0) \geq o_p(1)\right], \tag{a.34}$$

where $\mathcal{L}^*(\beta)$ is the probability limit of $\mathcal{L}^*_{nT}(\beta)$ defined in equation (4.4). The probability in the RHS of inequality (a.34) is $o(1)$, since $\sup_{\beta \in \mathcal{B}: \|\beta - \beta_0\| \geq \varepsilon} \mathcal{L}^*(\beta) - \mathcal{L}^*(\beta_0) < 0$ by global identification Assumption A.6, continuity of function $\mathcal{L}^*(\beta)$ and compactness of set $\mathcal{B}$.

Let us now show that $\tilde{\theta}_{nT}$ is consistent. For any $\varepsilon > 0$ we have:

$$\begin{aligned}\mathbb{P}\left[\|\tilde{\theta}_{nT} - \theta_0\| \geq \varepsilon\right] &\leq \mathbb{P}\left[\sup_{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon} \mathcal{L}_{nT}(\tilde{\beta}_{nT}, \theta) \geq \mathcal{L}_{nT}(\tilde{\beta}_{nT}, \tilde{\theta}_{nT})\right] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon} \mathcal{L}_{nT}(\tilde{\beta}_{nT}, \theta) \geq \mathcal{L}_{nT}(\tilde{\beta}_{nT}, \theta_0)\right].\end{aligned}$$

Using Proposition 1 (i), Lemma 1 (ii), and the consistency of $\tilde{\beta}_{nT}$, the RHS probability is such that:

$$\begin{aligned}\mathbb{P}\left[\sup_{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon} \mathcal{L}_{nT}(\tilde{\beta}_{nT}, \theta) \geq \mathcal{L}_{nT}(\tilde{\beta}_{nT}, \theta_0)\right] &= \mathbb{P}\left[\sup_{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon} \frac{1}{n}[\mathcal{L}_{1,nT}(\tilde{\beta}_{nT}, \theta) - \mathcal{L}_{1,nT}(\tilde{\beta}_{nT}, \theta_0)] \geq o_p\left(1/n\right)\right] \\ &= \mathbb{P}\left[\sup_{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon} \mathcal{L}_1(\beta_0, \theta) - \mathcal{L}_1(\beta_0, \theta_0) \geq o_p\left(1\right)\right],\end{aligned}$$

where $\mathcal{L}_1(\beta, \theta)$ is the probability limit of $\mathcal{L}_{1,nT}(\beta, \theta)$ defined in equation (a.10). Therefore we get:

$$\mathbb{P}\left[\|\tilde{\theta}_{nT} - \theta_0\| \geq \varepsilon\right] \leq \mathbb{P}\left[\sup_{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon} \mathcal{L}_1(\beta_0, \theta) - \mathcal{L}_1(\beta_0, \theta_0) \geq o_p\left(1\right)\right].$$

The RHS probability is $o(1)$, since $\sup_{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon} \mathcal{L}_1(\beta_0, \theta) - \mathcal{L}_1(\beta_0, \theta_0) < 0$ from global identification Assumption A.8, continuity of mapping $\theta \to \mathcal{L}_1(\beta_0, \theta)$ and the compactness of set Θ. The consistency of $\tilde{\theta}_{nT}$ follows.

The proof of the consistency of $(\tilde{\beta}_{nT}^{CSA}, \tilde{\theta}_{nT}^{CSA})$ and $(\tilde{\beta}_{nT}^{GA}, \tilde{\theta}_{nT}^{GA})$ is similar, by replacing criterion $\mathcal{L}_{nT}(\beta, \theta)$ with $\mathcal{L}_{nT}^{CSA}(\beta, \theta)$, and $\mathcal{L}_{nT}^{GA}(\beta, \theta)$, respectively, in the above arguments.

ii) Stochastic difference between estimators (proof of Proposition 2)

Since the CSA, GA and true ML estimators are consistent, the stochastic difference between these estimators can be derived along the lines of Robinson (1988), Theorem 1. However, we have to carefully take into account the double asymptotics in n and T. We provide the proof for $n, T \to \infty$ such that $T^\nu/n = O(1)$ with $\nu > 1$ (the proof for $\nu > 3/2$ is similar).

Let us first prove the stochastic difference between $(\tilde{\beta}_{nT}^{CSA}, \tilde{\theta}_{nT}^{CSA})$ and $(\tilde{\beta}_{nT}, \tilde{\theta}_{nT})$ [equivalence (4.7) in Proposition 2]. From the first-order conditions of the true and CSA ML estimators, Proposition 1 (i) and the mean value Theorem, we have:

$$
\begin{aligned}
0 &= \frac{\partial \mathcal{L}_{nT}\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right)}{\partial \left(\beta', \theta'\right)'} = \frac{\partial \mathcal{L}_{nT}^{CSA}\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right)}{\partial \left(\beta', \theta'\right)'} + \frac{\partial \Psi_{nT}\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right)}{\partial (\beta', \theta')'} \\
&= \frac{\partial^2 \mathcal{L}_{nT}^{CSA}\left(\bar{\beta}_{nT}, \bar{\theta}_{nT}\right)}{\partial \left(\beta', \theta'\right)' \partial \left(\beta', \theta'\right)} \begin{pmatrix} \tilde{\beta}_{nT} - \tilde{\beta}_{nT}^{CSA} \\ \tilde{\theta}_{nT} - \tilde{\theta}_{nT}^{CSA} \end{pmatrix} + \frac{\partial \Psi_{nT}\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right)}{\partial (\beta', \theta')'},
\end{aligned} \tag{a.35}
$$

where $\bar{\beta}_{nT}$ is between $\tilde{\beta}_{nT}$ and $\tilde{\beta}_{nT}^{CSA}$, and similarly for $\bar{\theta}_{nT}$. [14] From section i) above, $(\bar{\beta}_{nT}, \bar{\theta}_{nT})$ converges to (β_0, θ_0) in probability. Let us now use Lemma 6 in the supplementary material, which provides the uniform convergence of functions $\mathcal{L}_{nT}^*$, $\mathcal{L}_{1,nT}$, $\mathcal{L}_{2,nT}$, Ψ_{nT}, $\tilde{\Psi}_{nT}$ in the asymptotic expansion of the log-likelihood function, and of their partial derivatives. From Lemmas 6 (1), (2iii-iv) we get $\frac{\partial^2 \mathcal{L}_{nT}^{CSA}\left(\bar{\beta}_{nT}, \bar{\theta}_{nT}\right)}{\partial \beta \partial \beta'} = -I_0^* + o_p(1)$, $\frac{\partial^2 \mathcal{L}_{nT}^{CSA}\left(\bar{\beta}_{nT}, \bar{\theta}_{nT}\right)}{\partial \theta \partial \theta'} = -\frac{1}{n} I_{1,\theta\theta} + o_p(1/n)$ and $\frac{\partial^2 \mathcal{L}_{nT}^{CSA}\left(\bar{\beta}_{nT}, \bar{\theta}_{nT}\right)}{\partial \beta \partial \theta'} = O_p(1/n)$, where matrices I_0^* and $I_{1,\theta\theta}$ are defined in Assumptions A.7 and A.9. Moreover, from Lemma 6 (3), we have $\frac{\partial \Psi_{nT}\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right)}{\partial (\beta', \theta')'} = \left[o_p(1/n), O_p\left(\frac{[\log(n)]^{C_9}}{n^{3/2}}\right)\right]'$, for a constant $C_9 > 0$. From equation (a.35) we deduce:

$$
-I_0^*\left(\tilde{\beta}_{nT} - \tilde{\beta}_{nT}^{CSA}\right) + o_p\left(\tilde{\beta}_{nT} - \tilde{\beta}_{nT}^{CSA}\right) + O_p\left(\frac{1}{n}\left(\tilde{\theta}_{nT} - \tilde{\theta}_{nT}^{CSA}\right)\right) = o_p(1/n), \tag{a.36}
$$

$$
-I_{1,\theta\theta}\left(\tilde{\theta}_{nT} - \tilde{\theta}_{nT}^{CSA}\right) + o_p\left(\tilde{\theta}_{nT} - \tilde{\theta}_{nT}^{CSA}\right) + O_p\left(\tilde{\beta}_{nT} - \tilde{\beta}_{nT}^{CSA}\right) = O_p\left(\frac{[\log(n)]^{C_9}}{n^{1/2}}\right). \tag{a.37}
$$

Since matrix I_0^* is positive definite, and $\tilde{\theta}_{nT} - \tilde{\theta}_{nT}^{CSA} = o_p(1)$ by consistency of the estimators, equation (a.36) implies $\tilde{\beta}_{nT} - \tilde{\beta}_{nT}^{CSA} = o_p(1/n)$, that is the first equation in the equivalence (4.7) in Proposition 2.

[14] More precisely, the mean value Theorem is applied separately to each component of the vector $\frac{\partial \mathcal{L}_{nT}^{CSA}\left(\tilde{\beta}_{nT}, \tilde{\theta}_{nT}\right)}{\partial \left(\beta', \theta'\right)'}$, and the mean values can be different across components. For expository purpose, we do not make this explicit in equation (a.35) and in the rest of the proofs.

Then, since $I_{1,\theta\theta}$ is a positive definite matrix, equation (a.37) implies the second equation in the equivalence (4.7) in Proposition 2 (with $\delta_1 = C_9$).

To derive the stochastic difference between the true and GA ML estimators [equivalence (4.8) in Proposition 2], we use that $\mathcal{L}_{nT}(\beta,\theta) = \mathcal{L}_{nT}^{GA}(\beta,\theta) + \tilde{\Psi}_{nT}(\beta,\theta)$, where $\tilde{\Psi}_{nT}(\beta,\theta) = \Psi_{nT}(\beta,\theta) - \frac{1}{n^2}\mathcal{L}_{2,nT}(\beta,\theta)$. From Lemma 6 (2ii), (3), we get $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial\tilde{\Psi}_{nT}(\beta,\theta)}{\partial\beta}\right\| = o_p(1/n)$ and $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial\tilde{\Psi}_{nT}(\beta,\theta)}{\partial\theta}\right\| = O_p\left(\frac{[\log(n)]^{C_9}}{n^{3/2}}\right)$ when $T^\nu/n = O(1)$, $\nu > 1$. Then, by similar arguments as above, the equivalence (4.8) follows.

A.2.3 Proof of Proposition 3

The proof is in three steps. We first derive the asymptotic expansion of the standardized CSA ML estimator in terms of the standardized score. Then, we prove the asymptotic normality of the standardized score. Finally, this asymptotic normality and the asymptotic equivalences (Proposition 2) are used to deduce the asymptotic normality of the different estimators.

i) Asymptotic expansion of the CSA ML estimator

The first-order conditions for $(\hat{\beta}_{nT}, \hat{\theta}_{nT}) = (\tilde{\beta}_{nT}^{CSA}, \tilde{\theta}_{nT}^{CSA})$ are:

$$\begin{aligned} 0 &= \frac{\partial\mathcal{L}_{nT}^{CSA}}{\partial\beta}\left(\hat{\beta}_{nT},\hat{\theta}_{nT}\right) = \frac{\partial\mathcal{L}_{nT}^{*}}{\partial\beta}\left(\hat{\beta}_{nT}\right) + \frac{1}{n}\frac{\partial\mathcal{L}_{1,nT}}{\partial\beta}\left(\hat{\beta}_{nT},\hat{\theta}_{nT}\right), \\ 0 &= \frac{\partial\mathcal{L}_{nT}^{CSA}}{\partial\theta}\left(\hat{\beta}_{nT},\hat{\theta}_{nT}\right) \quad\Leftrightarrow\quad 0 = \frac{\partial\mathcal{L}_{1,nT}}{\partial\theta}\left(\hat{\beta}_{nT},\hat{\theta}_{nT}\right), \end{aligned}$$

where the factor $1/n$ in the second equation cancels. Let us multiply the first equation by $\sqrt{nT}$, the second equation by $\sqrt{T}$, and use the mean value Theorem to get:

$$\begin{aligned} 0 &= \sqrt{nT}\frac{\partial\mathcal{L}_{nT}^{*}(\beta_0)}{\partial\beta} + \frac{\partial^2\mathcal{L}_{nT}^{*}(\bar{\beta}_{nT})}{\partial\beta\partial\beta'}\sqrt{nT}\left(\hat{\beta}_{nT}-\beta_0\right) + \sqrt{\frac{T}{n}}\frac{\partial\mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\beta} \\ &\quad + \frac{1}{n}\frac{\partial^2\mathcal{L}_{1,nT}(\bar{\beta}_{nT},\bar{\theta}_{nT})}{\partial\beta\partial\beta'}\sqrt{nT}\left(\hat{\beta}_{nT}-\beta_0\right) + \frac{1}{\sqrt{n}}\frac{\partial^2\mathcal{L}_{1,nT}(\bar{\beta}_{nT},\bar{\theta}_{nT})}{\partial\beta\partial\theta'}\sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right), \end{aligned}$$

and:

$$0 = \sqrt{T}\frac{\partial\mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\theta} + \frac{1}{\sqrt{n}}\frac{\partial^2\mathcal{L}_{1,nT}(\bar{\beta}_{nT},\bar{\theta}_{nT})}{\partial\theta\partial\beta'}\sqrt{nT}\left(\hat{\beta}_{nT}-\beta_0\right) + \frac{\partial^2\mathcal{L}_{1,nT}(\bar{\beta}_{nT},\bar{\theta}_{nT})}{\partial\theta\partial\theta'}\sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right),$$

where $\bar{\beta}_{nT}$ and $\bar{\theta}_{nT}$ are mean values. In matrix form we have:

$$\begin{aligned} &-\begin{bmatrix} \frac{\partial^2\mathcal{L}_{nT}^{*}(\bar{\beta}_{nT})}{\partial\beta\partial\beta'} + \frac{1}{n}\frac{\partial^2\mathcal{L}_{1,nT}(\bar{\beta}_{nT},\bar{\theta}_{nT})}{\partial\beta\partial\beta'} & \frac{1}{\sqrt{n}}\frac{\partial^2\mathcal{L}_{1,nT}(\bar{\beta}_{nT},\bar{\theta}_{nT})}{\partial\beta\partial\theta'} \\ \frac{1}{\sqrt{n}}\frac{\partial^2\mathcal{L}_{1,nT}(\bar{\beta}_{nT},\bar{\theta}_{nT})}{\partial\theta\partial\beta'} & \frac{\partial^2\mathcal{L}_{1,nT}(\bar{\beta}_{nT},\bar{\theta}_{nT})}{\partial\theta\partial\theta'} \end{bmatrix}\begin{bmatrix} \sqrt{nT}\left(\hat{\beta}_{nT}-\beta_0\right) \\ \sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right) \end{bmatrix} \\ &\qquad = \begin{bmatrix} \sqrt{nT}\frac{\partial\mathcal{L}_{nT}^{*}(\beta_0)}{\partial\beta} \\ \sqrt{T}\frac{\partial\mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\theta} \end{bmatrix} + \begin{bmatrix} \sqrt{\frac{T}{n}}\frac{\partial\mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\beta} \\ 0 \end{bmatrix} + o_p(1). \quad \text{(a.38)} \end{aligned}$$

The second term in the RHS of (a.38) contributes to the asymptotic bias. From Lemma 6 (2i) in the supplementary material, and since $T/n \to 0$, this term is $o_p(1)$. From Lemma 6 (1), (2iii-iv) , we get:

$$\begin{bmatrix} \sqrt{nT}\left(\hat{\beta}_{nT}-\beta_0\right) \\ \sqrt{T}\left(\hat{\theta}_{nT}-\theta_0\right) \end{bmatrix} = \left[\begin{pmatrix} (I_0^*)^{-1} & 0 \\ 0 & I_{1,\theta\theta}^{-1} \end{pmatrix} + o_p(1)\right] \begin{bmatrix} \sqrt{nT}\frac{\partial \mathcal{L}_{nT}^*(\beta_0)}{\partial\beta} \\ \sqrt{T}\frac{\partial \mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\theta} \end{bmatrix} + o_p(1). \quad \text{(a.39)}$$

ii) Asymptotic normality of the standardized score vector

PROPOSITION A.5. *Let Assumptions A.1-A.5 and H.1-H.15 be satisfied. If* $n, T \to \infty$ *such that* $T^\nu/n = O(1)$, $\nu > 1$, *the standardized approximate score vector of the partial derivatives of functions* $\mathcal{L}_{nT}^*(\beta)$ *and* $\mathcal{L}_{1,nT}(\beta,\theta)$ *w.r.t.* β *and* θ, *respectively, is such that:*

$$\begin{bmatrix} \sqrt{nT}\dfrac{\partial \mathcal{L}_{nT}^*(\beta_0)}{\partial\beta} \\ \sqrt{T}\dfrac{\partial \mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial\theta} \end{bmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_0^* & 0 \\ 0 & I_{1,\theta\theta} \end{pmatrix}\right),$$

where $I_0^* = E_0\left[I_{\beta\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}I_{f\beta}(t)\right]$ *and* $I_{1,\theta\theta} = E_0\left[-\dfrac{\partial^2 \log g\left(f_t|f_{t-1};\theta_0\right)}{\partial\theta\partial\theta'}\right]$.

Proof of Proposition A.5: Let us first consider the approximate score w.r.t. β. By the envelope Theorem [e.g., Dixit (1990)] we have $\sqrt{nT}\dfrac{\partial \mathcal{L}_{nT}^*(\beta_0)}{\partial\beta} = \dfrac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\dfrac{\partial \log h}{\partial\beta}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta_0);\beta_0\right)$. By the mean value Theorem we get:

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

where $\tilde{f}_t$ are mean values. By Assumption H.11, Limit Theorem 1 in the supplementary material and condition $T^\nu/n = O(1)$, $\nu > 1$, we can show that $\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{\partial^2 \log h}{\partial\beta\partial f_t'}\left(y_{i,t}|y_{i,t-1},\tilde{f}_t;\beta_0\right) = -I_{\beta f}(t) + O_p\left(\dfrac{(\log n)^{C_{10}}}{\sqrt{n}}\right)$, uniformly in $1 \le t \le T$, for some constant $C_{10} > 0$, where $I_{\beta f}(t)$ is the (β, f) block of the matrix $I(t)$ defined in equation (4.6). Then, by Limit Theorem 1 and the condition $T^\nu/n = O(1)$, $\nu > 1$, we have:

$$\sqrt{nT}\frac{\partial \mathcal{L}_{nT}^*(\beta_0)}{\partial\beta} = \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial \log h}{\partial\beta}\left(y_{i,t}|y_{i,t-1},f_t;\beta_0\right) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}I_{\beta f}(t)\sqrt{n}\left(\hat{f}_{n,t}(\beta_0)-f_t\right) + o_p(1). \quad \text{(a.40)}$$

Let us now derive an asymptotic expansion for $\sqrt{n}\left(\hat{f}_{n,t}(\beta_0)-f_t\right)$. Since f_t is in the interior of set $\mathcal{F}_n$ w.p.a. 1 from Assumptions H.6 and H.7 (i)-(ii), and $\hat{f}_{n,t}(\beta_0)$ converges in probability to f_t by Limit Theorem

1, the first-order condition $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log h(y_{i,t}|y_{i,t-1}, \hat{f}_{n,t}(\beta_0); \beta_0)}{\partial f_t} = 0$ holds w.p.a. 1. Then, by the mean value Theorem, we have:

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log h}{\partial f_t}(y_{i,t}|y_{i,t-1}, f_t; \beta_0) + \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f_t \partial f_t'}(y_{i,t}|y_{i,t-1}, \bar{f}_t; \beta_0) \right) \sqrt{n} \left(\hat{f}_{n,t}(\beta_0) - f_t \right),$$

where $\bar{f}_t$ is a mean value. Similarly to above, by Assumption H.11, Limit Theorem 1 and condition $T^\nu / n = O(1)$, $\nu > 1$, we have $\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f_t \partial f_t'} \left(y_{i,t}|y_{i,t-1}, \bar{f}_t; \beta_0 \right) = -I_{ff}(t) + O_p \left(\frac{(\log n)^{C_{11}}}{\sqrt{n}} \right)$, uniformly in $1 \leq t \leq T$, for some constant $C_{11} > 0$, where $I_{ff}(t)$ is the (f, f) block of the matrix $I(t)$ defined in equation (4.6). Then, by Limit Theorem 1 and Assumption H.5 we get:

$$\sqrt{n} \left(\hat{f}_{n,t}(\beta_0) - f_t \right) = I_{ff}(t)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log h}{\partial f_t}(y_{i,t}|y_{i,t-1}, f_t; \beta_0) + O_p \left(\frac{(\log n)^{C_{12}}}{\sqrt{n}} \right), \tag{a.41}$$

uniformly in $1 \leq t \leq T$, for some constant $C_{12} > 0$. By replacing expansion (a.41) into expansion (a.40), and by using the condition $T^\nu / n = O(1)$, $\nu > 1$, and Assumption H.5, we get:

$$\sqrt{nT} \frac{\partial \mathcal{L}^*_{nT}(\beta_0)}{\partial \beta} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\psi_{n,\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} \psi_{n,f}(t) \right] + o_p(1), \tag{a.42}$$

where:

$$\psi_{n,\beta}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log h}{\partial \beta}(y_{i,t}|y_{i,t-1}, f_t; \beta_0), \quad \psi_{n,f}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log h}{\partial f_t}(y_{i,t}|y_{i,t-1}, f_t; \beta_0). \tag{a.43}$$

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

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

By using $T^\nu / n = O(1)$, $\nu > 1$, Assumption H.14 and Limit Theorem 1, it follows that:

$$\sqrt{T} \frac{\partial \mathcal{L}_{1,nT}(\beta_0, \theta_0)}{\partial \theta} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \log g}{\partial \theta}(f_t|f_{t-1}; \theta_0) + o_p(1). \tag{a.44}$$

Thus, from equations (a.42) and (a.44) we deduce:

$$\begin{bmatrix} \sqrt{nT} \frac{\partial \mathcal{L}^*_{nT}(\beta_0)}{\partial \beta} \\ \sqrt{T} \frac{\partial \mathcal{L}_{1,nT}(\beta_0, \theta_0)}{\partial \theta} \end{bmatrix} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \zeta_{n,t} + o_p(1), \quad \zeta_{n,t} \equiv \begin{bmatrix} \psi_{n,\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} \psi_{n,f}(t) \\ \frac{\partial \log g}{\partial \theta}(f_t|f_{t-1}; \theta_0) \end{bmatrix}, \tag{a.45}$$

where $\psi_{n,\beta}(t)$ and $\psi_{n,f}(t)$ are defined in (a.43). Proposition A.5 follows if we prove that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\zeta_{n,t} \stackrel{d}{\to}$ $N(0,\Omega)$ as $n,T\to\infty$, where $\Omega = \begin{pmatrix} I_0^* & 0 \\ 0 & I_{1,\theta\theta} \end{pmatrix}$. Since $\{\zeta_{n,t},\mathcal{G}_{n,t},\ 1\le t\le T;\ n\in\mathbb{N}\}$ is a martingale difference array w.r.t. the filtration $\mathcal{G}_{n,t} = \left(\underline{y_{i,t}},\ 1\le i\le n,\ \underline{f_{t+1}}\right)$, t varying, namely $\zeta_{n,t}$ is measurable w.r.t. $\mathcal{G}_{n,t}$ and $E[\zeta_{n,t}|\mathcal{G}_{n,t-1}] = 0$ for any $t\le T$ and $n\in\mathbb{N}$, we can apply Theorem 3.2 in Hall and Heyde (1980). [15] Thus, Proposition A.5 follows if we prove the next three conditions:

$$\text{(a)}\quad \frac{1}{\sqrt{T}}\max_{1\le t\le T}\|\zeta_{n,t}\| \stackrel{p}{\to} 0;\quad \text{(b)}\quad \frac{1}{T}\sum_{t=1}^{T}\zeta_{n,t}\zeta_{n,t}' \stackrel{p}{\to} E[\zeta_{n,t}\zeta_{n,t}'] = \Omega;\quad \text{(c)}\quad \frac{1}{T}E\left(\max_{1\le t\le T}\|\zeta_{n,t}\|^2\right) = O(1).$$

These conditions are checked in Lemma 7 in the supplementary material when $n,T\to\infty$ such that $T^\nu/n = O(1)$ with $\nu > 0$. In particular, the variance-covariance matrix Ω of the random vector $\zeta_{n,t}$ in (a.45) is block-diagonal, since the micro-component $\psi_{n,\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}\psi_{n,f}(t)$ is zero-mean conditional on the factor path, while the macro-component $\partial\log g\left(f_t|f_{t-1};\theta_0\right)/\partial\theta$ depends on the factor path only. ■

iii) Asymptotic normality of the estimators (proof of Proposition 3)

The joint asymptotic normality of the CSA ML estimator $(\hat{\beta}_{nT},\hat{\theta}_{nT})$ follows from the asymptotic expansion (a.39) and Proposition A.5. The asymptotic normality of the GA and true ML estimators is implied by the asymptotic normality of the CSA ML estimator and the asymptotic equivalences (4.7)-(4.8) in Proposition 2 when $T^\nu/n = O(1)$, $\nu > 1$.

A.2.4 Proof of Proposition 5

i) Proof of Proposition 5 i)

By the mean value Theorem we have:

$$\sqrt{n}\left(\hat{f}_{nT,t} - f_t\right) = \sqrt{n}\left(\hat{f}_{n,t}(\beta_0) - f_t\right) + \frac{\partial\hat{f}_{n,t}\left(\dot{\beta}_{nT}\right)}{\partial\beta'}\sqrt{n}\left(\hat{\beta}_{nT} - \beta_0\right), \tag{a.46}$$

where $\dot{\beta}_{nT}$ is a mean value. Let us consider the first term in the RHS. By the proof of Limit Theorem 1 in the supplementary material, we get that $\hat{f}_{n,t}(\beta_0)$ converges in probability to f_t, conditional on $\underline{f_t}$, for $\mathbb{P}$-almost every (a.e.) $\underline{f_t}$. Thus, $\hat{f}_{n,t}(\beta_0)$ coincides with the maximizer of the cross-sectional log-likelihood function

[15] To apply Theorem 3.2 in Hall and Heyde (1980), replace t for i, $\zeta_{n,t}/\sqrt{T}$ for $X_{i,n}$, and T_n for k_n, where T_n denotes the time dimension of the panel written as a function of the cross-sectional dimension n in the double asymptotic scheme. Theorem 3.2 in Hall and Heyde (1980) holds as long as $T_n\uparrow\infty$ when $n\to\infty$. Condition (3.21) in Hall and Heyde (1980) is satisfied in our setting, but is not needed, since the variance in the limit distribution is non-stochastic, see the remark on page 58 in Hall and Heyde (1980). Finally, while Theorem 3.2 in Hall and Heyde (1980) is stated for univariate processes, the multivariate extension is immediate by using the Cramer-Wold device.

$\sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f; \beta_0)$ w.r.t. f in set $\{f \in \mathbb{R}^m : \|f - f_t\| \leq r\}$, w.p.a. 1, conditional on $\underline{f_t}$, for any $r > 0$. From Assumptions A.1 and H.2, we get $\sqrt{n}\left(\hat{f}_{n,t}(\beta_0) - f_t\right) \xrightarrow{d} N(0, I_{ff}(t)^{-1})$, conditionally on $\underline{f_t}$, by applying Theorem 4.2.4 of Amemiya (1985) on the asymptotic normality of ML estimators. In checking the conditions of Theorem 4.2.4 of Amemiya (1985), we use that observations $(y_{i,t}, y_{i,t-1})$, for $i = 1, ..., n$, are i.i.d. conditional on the factor path $\underline{f_t}$ from Assumption A.1, and that Assumption H.2 implies the global and local identification conditions of f_t. Moreover, the score $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log h(y_{i,t}|y_{i,t-1}, f_t; \beta_0)}{\partial f_t}$ is asymptotically $N(0, I_{ff}(t))$ distributed, conditional on $\underline{f_t}$, by applying a standard CLT and using Assumption H.2.

Let us now consider the second term in the RHS of equation (a.46). We use Lemma 8 in the supplementary material, which provides a probability bound for $\partial \hat{f}_{n,t}(\beta)/\partial \beta'$, uniformly in $\beta \in \mathcal{B}$, conditionally on $\underline{f_t}$. Then, from Lemma 8 and Proposition 3, the second term in the RHS of equation (a.46) is $o_p(1)$, conditionally on $\underline{f_t}$. The asymptotic normality in Proposition 5 (i) follows.

ii) Proof of Proposition 5 ii)

We have $\|\hat{f}_{nT,t} - f_t\| \leq \|\hat{f}_{n,t}(\hat{\beta}_{nT}) - f_t(\hat{\beta}_{nT})\| + \|f_t(\hat{\beta}_{nT}) - f_t(\beta_0)\|$ and thus:

$$\sup_{1\leq t\leq T} \left\|\hat{f}_{nT,t} - f_t\right\| \leq \sup_{1\leq t\leq T} \sup_{\beta \in \mathcal{B}} \left\|\hat{f}_{n,t}(\beta) - f_t(\beta)\right\| + \sup_{1\leq t\leq T} \sup_{\beta \in \mathcal{B}} \left\|\frac{\partial f_t(\beta)}{\partial \beta'}\right\| \left\|\hat{\beta}_{nT} - \beta_0\right\|. \tag{a.47}$$

From Limit Theorem 1, the first term in the RHS of inequality (a.47) is $O_p\left(n^{-1/2}[\log(n)]^{\delta_2}\right)$. Let us consider the second term. By differentiating the first-order condition $E_0\left[\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f_t}|\underline{f_t}\right] = 0$ w.r.t. β, we deduce $\frac{\partial f_t(\beta)}{\partial \beta'} = -I_{t,ff}(\beta)^{-1} I_{t,f\beta}(\beta)$, where $I_{t,ff}(\beta)$ and $I_{t,f\beta}(\beta)$ are the blocks of the Hessian matrix $I_t(\beta)$ defined in Assumption H.4 (iii). From Assumptions H.5, we get $\sup_{1\leq t\leq T} \sup_{\beta \in \mathcal{B}} \left\|\frac{\partial f_t(\beta)}{\partial \beta'}\right\| = O_p\left([\log(n)]^{C_{13}}\right)$, for some $C_{13} > 0$. Then, from Proposition 3, the second term in RHS of inequality (a.47) is $O_p\left((nT)^{-1/2}[\log(n)]^{C_{13}}\right)$. The uniform convergence rate in Proposition 5 (ii) follows.

A.2.5 Proof of Proposition 6

i) Consistency

Let us first show that the estimator $(\hat{\beta}^*_{nT}, \hat{\theta}^*_{nT})$ is consistent. The consistency of $\hat{\beta}^*_{nT}$ follows by similar arguments as in Section A.2.2 i), by setting functions $\mathcal{L}_{1,nT}(\beta, \theta)$ and $\Psi_{nT}(\beta, \theta)$ equal to zero. To prove the consistency of $\hat{\theta}^*_{nT}$, we use that $\hat{\theta}^*_{nT}$ is the maximizer of criterion $Q_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \log g[\hat{f}_{n,t}(\hat{\beta}^*_{nT})|\hat{f}_{n,t-1}(\hat{\beta}^*_{nT}); \theta]$ over the set Θ. We have $Q_T(\theta) = \mathcal{L}_{1,nT}(\hat{\beta}^*_{nT}, \theta)$, up to a constant independent of θ. By a slight modification of Lemma 1 (ii) and the consistency of $\hat{\beta}^*_{nT}$, criterion $Q_T(\theta)$ converges in probability to $Q_\infty(\theta) =$

$E_0\left[\log g(f_t|f_{t-1};\theta)\right]$ uniformly in $\theta \in \Theta$. Since function Q_T is continuous, set Θ is compact, and θ_0 is the unique maximizer of function Q_∞ by the global identification Assumption A.8, we can apply the standard consistency theorem for extremum estimators [e.g., Amemiya (1985), Theorem 4.1.1]; it follows that $\hat{\theta}^*_{nT}$ converges to θ_0 in probability.

ii) Stochastic difference between estimators [proof of Proposition 6 (i)]

The first-order conditions of estimators $(\tilde{\beta}^{CSA}_{nT}, \tilde{\theta}^{CSA}_{nT})$ and $(\hat{\beta}^*_{nT}, \hat{\theta}^*_{nT})$ are given by:

$$\left\{\begin{array}{c} \dfrac{\partial \mathcal{L}^*_{nT}(\tilde{\beta}^{CSA}_{nT})}{\partial \beta} + \dfrac{1}{n}\dfrac{\partial \mathcal{L}_{1,nT}(\tilde{\beta}^{CSA}_{nT}, \tilde{\theta}^{CSA}_{nT})}{\partial \beta} = 0, \\ \dfrac{\mathcal{L}_{1,nT}(\tilde{\beta}^{CSA}_{nT}, \tilde{\theta}^{CSA}_{nT})}{\partial \theta} = 0, \end{array}\right. \qquad \left\{\begin{array}{c} \dfrac{\partial \mathcal{L}^*_{nT}(\hat{\beta}^*_{nT})}{\partial \beta} = 0, \\ \dfrac{\partial \mathcal{L}_{1,nT}(\hat{\beta}^*_{nT}, \hat{\theta}^*_{nT})}{\partial \theta} = 0, \end{array}\right.$$

respectively. Let us expand the first-order conditions of $(\tilde{\beta}^{CSA}_{nT}, \tilde{\theta}^{CSA}_{nT})$ around $(\hat{\beta}^*_{nT}, \hat{\theta}^*_{nT})$. By the mean value Theorem, and the first-order conditions of $(\hat{\beta}^*_{nT}, \hat{\theta}^*_{nT})$, we get:

$$0 = \frac{\partial^2 \mathcal{L}^*_{nT}(\bar{\beta}_{nT})}{\partial \beta \partial \beta'}\left(\tilde{\beta}^{CSA}_{nT} - \hat{\beta}^*_{nT}\right) + \frac{1}{n}\frac{\partial \mathcal{L}_{1,nT}(\tilde{\beta}^{CSA}_{nT}, \tilde{\theta}^{CSA}_{nT})}{\partial \beta}, \tag{a.48}$$

and:

$$0 = \frac{\partial^2 \mathcal{L}_{1,nT}(\bar{\beta}_{nT}, \bar{\theta}_{nT})}{\partial \theta \partial \beta'}\left(\tilde{\beta}^{CSA}_{nT} - \hat{\beta}^*_{nT}\right) + \frac{\partial^2 \mathcal{L}_{1,nT}(\bar{\beta}_{nT}, \bar{\theta}_{nT})}{\partial \theta \partial \theta'}\left(\tilde{\theta}^{CSA}_{nT} - \hat{\theta}^*_{nT}\right), \tag{a.49}$$

where $(\bar{\beta}_{nT}, \bar{\theta}_{nT})$ are mean values. Since the CSA and two-step estimators are consistent by Proposition 3 and section i) above, the mean values $(\bar{\beta}_{nT}, \bar{\theta}_{nT})$ are consistent as well. From Lemmas 6 (1), (2i) and (2iii) we get $\dfrac{\partial^2 \mathcal{L}^*_{nT}(\bar{\beta}_{nT})}{\partial \beta \partial \beta'} = -I^*_0 + o_p(1)$, $\dfrac{\partial \mathcal{L}_{1,nT}(\tilde{\beta}^{CSA}_{nT}, \tilde{\theta}^{CSA}_{nT})}{\partial \beta} = O_p(1)$, $\dfrac{\partial^2 \mathcal{L}_{1,nT}(\bar{\beta}_{nT}, \bar{\theta}_{nT})}{\partial \theta \partial \beta'} = O_p(1)$ and $\dfrac{\partial^2 \mathcal{L}_{1,nT}(\bar{\beta}_{nT}, \bar{\theta}_{nT})}{\partial \theta \partial \theta'} = -I_{1,\theta\theta} + o_p(1)$. Then, equation (a.48) implies $\tilde{\beta}^{CSA}_{nT} - \hat{\beta}^*_{nT} = O_p(1/n)$, and equation (a.49) implies $\tilde{\theta}^{CSA}_{nT} - \hat{\theta}^*_{nT} = O_p\left(\tilde{\beta}^{CSA}_{nT} - \hat{\beta}^*_{nT}\right) = O_p(1/n)$. Then, from equivalence (4.7) in Proposition 2, we get $\hat{\beta}^*_{nT} - \tilde{\beta}_{nT} = O_p(1/n)$ and $\hat{\theta}^*_{nT} - \tilde{\theta}_{nT} = O_p\left(\dfrac{[\log(n)]^{\delta_1}}{\sqrt{n}}\right)$.

iii) Asymptotic normality [proof of Proposition 6 (ii)]

From the condition $T^\nu/n = O(1)$, $\nu > 1$, and Proposition 6 (i), we get $\left(\sqrt{nT}(\hat{\beta}^*_{nT} - \beta_0)', \sqrt{T}(\hat{\theta}^*_{nT} - \theta_0)'\right)' = \left(\sqrt{nT}(\tilde{\beta}_{nT} - \beta_0)', \sqrt{T}(\tilde{\theta}_{nT} - \theta_0)'\right)' + o_p(1)$. Then, Proposition 6 (ii) follows from Proposition 3.

A.3 Identification in the stochastic migration model

The stochastic migration model is a set of ordered qualitative models, with an unobservable stochastic factor and a common vector of threshold parameters c_k, $k = 1, .., K-1$. This explains why the identification conditions have to be derived carefully.

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

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

By reparametrizing coefficients α_1 and α_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 γ_2, α_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 $c_1 = 0$:

$$\left[G\left(-\frac{\gamma_l f_t+\alpha_l}{\sigma_l}\right), G\left(\frac{c_2-\gamma_l f_t-\alpha_l}{\sigma_l}\right) - G\left(-\frac{\gamma_l f_t+\alpha_l}{\sigma_l}\right), ..., 1-G\left(\frac{c_{K-1}-\gamma_l f_t-\alpha_l}{\sigma_l}\right)\right]. \tag{a.50}$$

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 equation (a.50) for $l \geq 2$. From the first row, we 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 $(K-1)T \geq 3$.