Identification by Laplace Transforms in Nonlinear Time Series and Panel Models with Unobserved Stochastic Dynamic Effects*

Patrick Gagliardini[†], Christian Gouriéroux [‡]

First version: August, 2014. Revised version: August, 2017.

Abstract

We consider nonlinear parametric and semi-parametric models for time series and panel data including unobserved dynamic effects. These regression models have an affine specification with respect to lagged endogenous variables and unobserved dynamic effects. We derive conditional moment restrictions based on suitable Laplace transforms. We show how to deploy these nonlinear moment restrictions to identify the parameters of the affine regression model, and the parametric or nonparametric distribution of the unobserved effects. This approach is appropriate for studying identification in (nonlinear) latent factor models encountered in macroeconomic and financial applications as well as in panel models with stochastic time effects.

Keywords: Semi-Parametric Identification, Nonlinear Factor Model, Conditional Moment Restrictions, Cross-Differencing, Count Panel Data.

JEL classification: G12, C23.

---

*The authors thank a co-Editor, two anonymous referees, A. de Paula, J.-M. Dufour, D. Wilhelm and participants at the Econometric Society European Meeting (ESEM) 2014 in Toulouse, the CEMMAP seminar of the University College London, and at the Econometrics Seminar of Concordia University for very useful comments.

[†]Università della Svizzera Italiana (USI), Lugano, and Swiss Finance Institute. Corresponding author: Patrick Gagliardini, Università della Svizzera Italiana (USI), Faculty of Economics, Via Buffi 13, CH-6900 Lugano, Switzerland. Phone: +41 58 6664660. Email: patrick.gagliardini@usi.ch.

[‡]CREST and University of Toronto (christian.gourieroux@ensae.fr).

1 Introduction

This paper deals with identification in nonlinear models for time series and panel data including unobserved dynamic effects (or factors). Our results apply to possibly nonlinear regression models which have a parametric affine specification with respect to both lagged endogenous variables and unobserved dynamic effects. The dynamics of the unobservable effects is either parametrically specified, or let unconstrained, leading to parametric and semi-parametric frameworks, respectively. By exploiting the affine regression property we derive conditional moment restrictions based on appropriate Laplace transforms of the endogeneous observable variables and their lags. The main contribution of the paper is to show how these continuum sets of first-, second-, and third-order moment restrictions can be used to identify the regression parameters of the affine model as well as the parametric, or nonparametric, distribution of the dynamic unobservable effects. Our findings are particularly innovative in the semi-parametric setting, where identification results for dynamic nonlinear models with unobservable components are relatively scarce. Compared to other identification approaches considered in the literature (see Sections 3.4 and 5.1 for reviews), we treat the distribution of the unobservable time effects as a parameter of main interest, and not as a nuisance parameter. Moreover, our identification strategy based on conditional moment restrictions is constructive as it naturally leads to estimators of model parameters within the class of the Generalized Method of Moments (GMM). This paper focuses on deriving nonlinear moment restrictions and studying identification, and we leave estimation and inference for future work. We postpone to the concluding remarks a brief discussion on the implementation of the GMM estimators implied by our moment restrictions.

In Section 2, to motivate our paper we first provide examples of linear and nonlinear time series models with unobserved dynamic effects encountered in macroeconomic and financial applications. They include linear factor models for asset returns, multivariate stochastic volatility models, and models which are appropriate to disentangle dynamic frailty from contagion phenomena applied to either market, or credit risk. Then, we introduce a general specification which encompasses all these models. The conditional Laplace transform of the endogenous variables is an exponential affine function of the lagged endogenous variables and the unobservable effects, possibly given exogenous observable regressors.

In Section 3, we derive a continuum of conditional moment restrictions based on the Laplace transform of a well-chosen linear combination of the current and lagged observations of the endogenous variables.

These moment restrictions depend on the finite-dimensional parameter of interest, but also involve the Laplace transform of the distribution of the unobserved factor. We distinguish these moment restrictions according to the lag order of the variables in the joint Laplace transform, which yields first-, second-, and third-order conditional moment restrictions.

In Section 4, we explain how to identify the model parameters when the factor dynamics is specified parametrically. The first-order nonlinear moment restrictions identify the nonlinear regression parameters and the marginal distribution of the unobserved effect. To identify the dynamic parameters of the distribution of the unobserved process, we need second-order nonlinear moment restrictions based on the current value, and the two most recent lags, of the endogenous observable variable. In nonlinear fully parametric models with unobservable factors, the continuum set of conditional moment restrictions derived in this paper is the basis for defining GMM estimators, which can be a computationally more convenient alternative to Maximum Likelihood (ML) estimators implemented via simulation (see references in Section 7).

In Section 5, we consider a model with an unspecified dynamics for the factor. In this semi-parametric framework, we first show how to eliminate the infinite-dimensional parameter, i.e. the transition distribution of the unobservable factor, from the second-order nonlinear moment restrictions by a suitable transformation. We investigate the identifiability of the nonlinear regression parameters from these transformations of the second-order nonlinear moment restrictions. Then, once the regression parameters are identified, the transition distribution of the unobservable factor is identifiable from the second-order nonlinear moment restrictions themselves. The identification results are illustrated by studying semi-parametric identification in the multivariate Poisson model with common stochastic intensity and in a non-Gaussian linear dynamic model with latent factor and contagion.

Section 6 considers nonlinear panel data models with stochastic time effects and exponentially affine conditional Laplace transform, such as autoregressive count data models with unobservable common factors. We explain how the first-order nonlinear moment restrictions obtained by a cross-differencing approach can be used to identify the regression parameters and the nonparametric stationary distribution of the time effect.

Section 7 concludes. We establish our results under a set of regularity assumptions. We distinguish between Assumptions A.1, A.2, ... used for the general time series setting, Assumptions B.1, B.2, ... adopted in the general panel data setting, and Assumptions 1, 2, ... concerning the examples. Proofs of the results are gathered in the Appendices. In Supplementary Materials available online, we derive the GMM semi-parametric efficiency bounds for estimating the regression parameters from a continuum of nonlinear conditional moment restrictions. The patterns of these bounds are illustrated for the Poisson count panel data model with stochastic time effect estimated by cross-differencing. In the Supplementary Materials we also provide the proofs of some technical lemmas and additional identification results.

2 The model

We first provide examples of models with unobserved dynamic effects proposed in the literature. Next we introduce a specification which includes all these examples and is used later on to derive the moment restrictions in a general framework.

2.1 Examples

Example 1: A Gaussian Linear Factor Model for Asset Returns.

In Finance, the most popular model to explain the expected excess returns of risky securities is the Capital Asset Pricing Model (CAPM) [see Sharpe (1964), Lintner (1965)]. It relies on a linear factor model for asset returns with a single factor corresponding to the market return. In empirical work, the market return is often treated as an observable factor. However, the market return is not easy to measure, and an index return used as a proxy can provide an erroneous measurement. This is the well-known Roll's critique [see Roll (1977)], which explains why the basic CAPM model is sometimes replaced by a model with unobservable factors.

Let us denote by y_t the vector of returns for n assets, and by f_t the vector of values of the underlying latent factors. The model is written as:

$$y_t = Bf_t + \varepsilon_t,$$

where B is the matrix of the factor sensitivities, the error terms are $IIN(0, \Sigma)$ and Σ is a diagonal matrix. In this strict (or exact) factor model, the dependencies across the asset returns are generated by the multiple factor f_t, called systematic risk factor [Sharpe (1964), p. 436]. Such a model is compatible with the heavy tails observed in the historical distribution of returns and with volatility clustering effects, whenever these effects are created by conditionally heteroskedastic latent factors. In this respect, this type of modeling includes the standard linear factor model [see e.g. Geweke (1977), Sargent, Sims (1977), Engle, Watson (1981), Fernandez-Macho (1997)] as well as factor ARCH models [Diebold, Nerlove (1989), King, Sentana, Wadhwani (1994)]. We consider a specification with a conditionally heteroskedastic factor in Example 5 in Section 4.

Example 2: Correlation Between Markets.

Let us consider the return vector y_t of market indexes corresponding to different stock exchanges. The dependence between these markets can be analyzed by a dynamic model of the type:

$$y_t = Bf_t + Cy_{t-1} + \varepsilon_t,$$

with $\varepsilon_t \sim IIN(0, \Sigma)$. This model accounts for both systematic risk factors and contagion effects across markets, through the unobservable common factor f_t and the lagged returns y_{t-1}, respectively. This model nests the linear factor model of Example 1 and a Vector Autoregressive (VAR) specification, which correspond to the restrictions $C = 0$, and $B = 0$, respectively [see Darolles, Gourieroux (2015) and the references therein for a general presentation of contagion modeling in Finance]. Such a specification may be completed by a noisy measurement of f_t in the Factor Augmented Vector Autoregressive (FAVAR) model introduced by Bernanke, Boivin, Eliasz (2005).

Example 3: Correlation of Default Risks.

The decomposition between exogenous common factors and lagged endogenous effects put forward in Example 2 can be adapted to analyze default risk. Such a model is introduced in Darolles, Gagliardini, Gourieroux (2014). The authors consider a population of hedge funds classified by management style k, for $k = 1, \ldots, K$. For each management style k, the observation $y_{k,t}$ is the number of liquidated funds at a given month t, and vector y_t stacks these count data for the different management styles. The dynamic Poisson model is:

$$y_{k,t} \sim \mathcal{P}(a_k + b_k' f_t + c_k' y_{t-1}), \quad k = 1, \ldots, K,$$

where the variables $y_{k,t}$, $k = 1, \ldots, K$, are assumed conditionally independent given f_t and y_{t-1}. This modeling disentangles the contagion effects passing through the lagged liquidation count values y_{t-1} and the effect of exogenous common shocks passing through the unobservable systematic factor f_t. The latter is interpreted in terms of aggregate funding liquidity shocks in Darolles, Gagliardini, Gourieroux (2014).

Example 4: Stochastic Volatility Model.

A multivariate stochastic volatility model for asset returns can be defined by the equation:

$$y_t = a + B\ vech(\Sigma_t) + \Sigma_t^{1/2}\varepsilon_t,$$

where $\varepsilon_t \sim IIN(0, Id)$, the symmetric positive definite matrix Σ_t is the stochastic volatility-covolatility matrix, and $f_t = vech(\Sigma_t)$ is a vector stacking all its different components. By introducing the term $Bvech(\Sigma_t)$ in the vector of conditional expected returns, we allow for risk premia corresponding not only to the volatility of the asset of interest, but also to the volatilities of the other assets, and to the covolatilities as well. Gagliardini, Gourieroux (2014b) discusses the estimation of such multivariate stochastic volatility models based on conditional moment restrictions.

2.2 General specification

In the general framework, the (nonlinear) regression model is written by means of the conditional Laplace transform. The latter provides the conditional expectations of exponential transformations of the endogenous variable y_t given the lagged values $\underline{y_{t-1}} = (y_{t-1}, y_{t-2}, ...)$ of these endogenous variables, and the current and past values of the unobservable dynamic effects $\underline{f_t} = (f_t, f_{t-1}, ...)$ and the covariates $\underline{x_t} = (x_t, x_{t-1}, ...)$. We denote by $n = \dim(y_t)$ and $K = \dim(f_t)$ the dimensions of the endogenous variable and the unobservable factor, respectively, and we assume $K \leq n$.

Assumption A.1. ***(Affine Regression Model)*** *We have:*

$$E[\exp(u'y_t)|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}] \quad = \quad \exp\left\{a(u, x_t, \theta)'[Bf_t + Cy_{t-1} + d(x_t, \theta)] + b(u, x_t, \theta)\right\}, \tag{2.1}$$

where $u \in \mathcal{U}$ is the multidimensional argument of the Laplace transform, $\mathcal{U} \subset \mathbb{C}^n$, vector θ and matrices B, C are parameters, possibly constrained, and a, b, d are known functions. The model is correctly specified with true values of the parameters in the conditional distribution denoted by θ_0, B_0, C_0.

The model is called affine regression model, since the log-Laplace transform is an affine function of both f_t and y_{t-1}. The specification above extends the Compound Autoregressive (CaR) processes introduced in Darolles, Gourieroux, Jasiak (2006) to the case of covariates and unobserved stochastic effects. Indeed, without processes x_t and f_t, the condition of Assumption $A.1$ restricts the model to (nonlinear) dynamics of the type:

$$E[\exp(u'y_t)|\underline{y_{t-1}}] = \exp[a^*(u, \theta)'y_{t-1} + b^*(u, \theta)], \quad \text{say}, \tag{2.2}$$

for some parameterized functions $a^*(\cdot,\theta)$ and $b^*(\cdot,\theta)$. The dynamics in Assumption A.1 correspond to special families of distributions, such that the Laplace transform (moment generating function) is exponentially affine w.r.t. a subset of the parameters. Their associated autoregressive models and regression models are obtained by letting these parameters be a linear function of y_{t-1}, f_t and a (possibly nonlinear) function of x_t. Some of these families are given below with the associated autoregressive model.

i) Gaussian family. We have the moment generating function:

$$E[\exp(u'y)] = \exp\left(u'm + \frac{u'\Sigma u}{2}\right),$$

where m and Σ are the mean vector and the variance-covariance matrix of the multivariate Gaussian distribution. This family is used to construct autoregressive models and regression models by considering m and/or Σ as function of the conditioning variables (see Examples 1, 2 and 4). We get the Gaussian VAR process (resp., the model in Example 2) when the conditional mean m_t is affine in the lagged observations (resp., in the current factor value as well) and the variance-covariance matrix is constant.

ii) Poisson family. If variable y follows a Poisson distribution with parameter $\lambda > 0$, we get:

$$E[\exp(uy)] = \exp[-\lambda(1-\exp u)]. \tag{2.3}$$

This Laplace transform is defined for any real (or complex) argument u. The associated autoregressive/regression model is obtained by specifying the time-varying stochastic intensity λ_t as a positive linear function of f_t, y_{t-1} and a possibly nonlinear function of x_t (see Example 3).

iii) Binomial family. If variable y follows a Binomial distribution $Bin(n,p)$ with integer parameter n and probability p, we have:

$$E[\exp(uy)] = (1-p+p\exp u)^n = \exp[n\log(1-p+p\exp u)],$$

which is exponential affine in parameter n. This property is used in the definition of autoregressive models for integer-valued data. For instance, the INteger AutoRegressive (INAR) process is a Markov process such that the transition distribution of y_t given y_{t-1} is the distribution of the sum of independent Poisson $\mathcal{P}(\lambda)$ and Binomial $Bin(y_{t-1},p)$ variables, with parameters $\lambda > 0$ and $p \in (0,1)$ [see e.g. Al-Osh, Alzaid (1987), Bockenholt (1994), Brannas (1994)]. Then:

$$E[\exp(uy_t)|y_{t-1}] = \exp[-\lambda(1-\exp u) + y_{t-1}\log(1-p+p\exp u)].$$

An extension of this model including unobservable effects is obtained by letting parameter λ be time-varying as a linear positive function of latent factor f_t.

iv) The noncentered gamma and Wishart families. For the noncentral gamma family we have:

$$E[\exp(uy)] = \frac{1}{(1-\lambda u)^{\nu}} \exp\left(-\frac{\lambda^2 um}{1-\lambda u}\right),$$

where ν, $\nu > 0$, is the degree of freedom, λ with $\lambda > 0$ is the scale parameter, and m is the noncentrality parameter. [1] The above Laplace transform is defined for any real argument u such that $u < 1/\lambda$. The corresponding autoregressive process is the AutoRegressive Gamma (ARG) process [Gourieroux, Jasiak (2006)], that is obtained by letting m be a linear positive function of the lagged value y_{t-1}, and is the time discretized Cox, Ingersoll, Ross process [Cox, Ingersoll, Ross (1985)]. We get a specification with unobservable effects by letting m be a linear positive function of y_{t-1} and f_t.

The multivariate extension of the noncentral gamma family is provided by the Wishart family. The Wishart distribution is a distribution for a stochastic symmetric positive semi-definite matrix Y, with dimension (K, K). Its Laplace transform can be written in matrix notation by considering a (K, K) symmetric matrix of arguments Γ and by writing a linear combination of the elements of Y as $Tr(\Gamma Y)$, where Tr denotes the trace operator providing the sum of the diagonal elements of a square matrix. The Laplace transform of the Wishart distribution is:

$$E[\exp Tr(\Gamma Y)] = \frac{\exp Tr[\Gamma(Id - 2\Sigma\Gamma)^{-1}M]}{[det(Id - 2\Sigma\Gamma)]^{\nu/2}},$$

where $\nu, \nu > 0$, is the degree of freedom parameter, the (K, K) symmetric positive definite matrix Σ is a variance-covariance matrix and the (K, K) matrix M a noncentrality parameter. The autoregressive process corresponding to the Wishart family is the Wishart AutoRegressive (WAR) process [see e.g. Gourieroux, Jasiak, Sufana (2009), Chiriac, Voev (2011)]. The associated affine regression model is obtained by specifying the noncentrality parameter matrix M as an affine function of the lagged values and of the unobserved dynamic effect.

The affine regression specification in Assumption A.1 is completed by additional assumptions on the unobservable dynamic effect, the regressors and their joint distribution.

[1] When $m = 0$ we get the (centered) gamma distribution $\gamma(\nu, \lambda)$ with degree of freedom ν and intensity λ.

Assumption A. 2. *The joint process $(f_t', x_t')'$ of the unobservable dynamic effects and the observable regressors is strongly exogenous, that is, the conditional distribution of $(f_t', x_t')'$ given $\underline{y_{t-1}}$, $\underline{f_{t-1}}$ and $\underline{x_{t-1}}$, is equal to the conditional distribution of $(f_t', x_t')'$ given $\underline{f_{t-1}}$ and $\underline{x_{t-1}}$ only.*

Assumption A.2 excludes feedback effects from lagged values of the endogeneous variable. Moreover, Assumptions A.1 and A.2 imply that the conditional distribution of y_t given $\underline{y_{t-1}}$, $\underline{f}$, $\underline{x}$ is equal to the conditional distribution of y_t given $\underline{y_{t-1}}$, $\underline{f_t}$, $\underline{x_t}$, where $\underline{f} = (..., f_{t-1}, f_t, f_{t+1}, ...)$ and $\underline{x} = (..., x_{t-1}, x_t, x_{t+1}, ...)$ denote the entire histories including past, current and future values of the dynamic effects, and the regressors, respectively. Exogeneity of the unobservable component is typical in state space models. As far as the observable regressor x_t is concerned, we could accommodate feedback effects from lagged values of y by including (some components of) x_t among the endogenous observable variables. The developments in Sections 3-6 show that we can dispense of the exogeneity condition on x_t (and on f_t) without compromising completely the validity of our identification strategy.

The next assumption is convenient for expository purpose.

Assumption A.3. *The joint process $(f_t', x_t')'$ is a stationary Markov process of order 1.*

Assumption A.3 is rather weak. Indeed, a Markov process of order p larger than 1 can easily be transformed into a Markov process of order 1 by changing the definition of f_t, x_t and considering the new factor obtained by stacking in a vector the current and lagged values of f and x. Moreover, the stationarity assumption is empirically relevant for the unobservable factors in the examples presented in Section 2.1, e.g. the market portfolio return in Example 1 with $K = 1$ and the funding liquidity factor in Example 3.

The joint dynamics of $(f_t', x_t', y_t')'$ is defined along the causal scheme displayed below, where each arrow represents a direct causality effect.

The causality scheme.

$$\begin{array}{cccccc} \dots & (f_{t-1}, x_{t-1}) & \longrightarrow & (f_t, x_t) & \dots \\ & \downarrow & & \downarrow & \\ \dots & y_{t-1} & \longrightarrow & y_t & \dots \end{array}$$

Under Assumptions A.2 and A.3, the dynamics of the exogenous process $(f_t', x_t')'$ is characterized by its transition p.d.f. $g(f_t, x_t | f_{t-1}, x_{t-1})$, say. In the analysis of the next sections, this transition p.d.f. is either specified parametrically, or let unspecified, which yields - together with Assumption A.1 - a fully parametric model, or a semi-parametric framework, respectively. In the former case, the parameters are the regression coefficients B, C, θ and the parameters of the transition p.d.f. g. In the latter case the parameter set includes both the finite-dimensional component B, C, θ and the infinite-dimensional parameter g. [2]

3 Moment restrictions

In this section we define nonlinear moment restrictions by exploiting the exponential affine structure of the model. These nonlinear moment restrictions differ by the number of lags of the endogenous variable. More precisely, the first-order (resp., second-order, third-order) restrictions involve the first (resp., the second, third) lag of the endogenous variable.

3.1 First-order nonlinear moment restrictions

Under Assumption A.1, we get:

$$E[\exp\{u'y_t - a(u, x_t, \theta)'[Cy_{t-1} + d(x_t, \theta)] - b(u, x_t, \theta)\}|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}] = \exp\{a(u, x_t, \theta)'Bf_t\}, \ \forall u \in \mathcal{U}. \quad (3.1)$$

The restrictions in (3.1) cannot be used directly for parameter identification, since they involve the unobserved factor f_t, both in the function in the right-hand side (r.h.s.) of the equation and in the conditioning set in the left-hand side (l.h.s.). However, they can be used to derive conditional moment restrictions involving observable variables only by integrating out both sides of equation (3.1) conditional on $\underline{x_t}$. Then, by the Iterated Expectation Theorem, we get the following conditions:

First-order nonlinear moment restrictions:

$$E[\exp\{u'y_t - a(u, x_t, \theta)'[Cy_{t-1} + d(x_t, \theta)] - b(u, x_t, \theta)\}|\underline{x_t}] = E\left[\exp\{a(u, x_t, \theta)'Bf_t\}|\underline{x_t}\right], \ \forall u \in \mathcal{U}, \quad (3.2)$$

where the conditional expectation in the right hand side is w.r.t. the (parametric, or nonparametric) model distribution of f_t given $\underline{x_t}$. Since the model is assumed correctly specified, these restrictions are valid for the true values of the parameters and the true conditional distributions. The subscript $_0$ denoting the true

[2]Concerning the transition p.d.f. g, our focus is on the identification of (non-)parametric features of the distribution of latent factors (f_t) given observable covariates (x_t), since the distribution of the latter can be identified by standard methods.

values is omitted for expository purpose. In (3.2) we have a continuum set of moment restrictions, because the argument u can vary over some domain $\mathcal{U} \subset \mathbb{C}^n$ for which the Laplace transform exists. This domain depends on the specific model.

Since the moment restrictions in (3.2) are conditional on $\underline{x_t}$, these restrictions are also valid for values u_t of the argument that are functions of $\underline{x_t}$ and possibly of parameters B, C and θ. Assumption A.4 below introduces a new argument v, say.

Assumption A.4 (Normalization): *There exists a change of argument from u to $v = (v_1', v_2')' \in \mathbb{C}^K \times \mathbb{C}^{n-K}$, and a domain $\mathcal{V} \subset \mathbb{C}^n$ such that $u(v, x_t, \theta, B) \in \mathcal{U}$, say, satisfies the equation:*

$$a[u(v, x_t, \theta, B), x_t, \theta]'B = v_1', \quad \forall v \in \mathcal{V}, B, \theta, x_t.$$

Under Assumptions A.1-A.4 we deduce from (3.2) the following moment conditions:

First-order nonlinear moment restrictions for the marginal distribution of (f_t):

$$\begin{aligned} & E[\exp\{u(v, x_t, \theta, B)'y_t - a[u(v, x_t, \theta, B), x_t, \theta]'[Cy_{t-1} + d(x_t, \theta)] - b[u(v, x_t, \theta, B), x_t, \theta]\}|\underline{x_t}] \\ = \; & E[\exp(v_1' f_t)|\underline{x_t}], \quad \forall v \in \mathcal{V}. \end{aligned} \tag{3.3}$$

The r.h.s. of equation (3.3) involves the Laplace transform of f_t conditional on $\underline{x_t}$.

3.2 Second-order nonlinear moment restrictions

In general, the first-order nonlinear moment restrictions (3.2) [or equivalently (3.3)] are not sufficient to identify the entire set of model parameters. For instance, any (parametric or nonparametric) feature, which characterizes the dynamics of the unobservable factor process (f_t), is unidentifiable from restrictions (3.2), because the latter ones only involve the distribution of f_t (conditional on $\underline{x_t}$). We derive below second-order nonlinear moment restrictions which involve second-order lags in the endogenous variables and dynamic features of the unobservable effects.

The second-order nonlinear moment restrictions are obtained by considering joint exponential transforms of y_t, y_{t-1} given $\underline{y_{t-2}}, \underline{f_t}, \underline{x_t}$, and accounting for an adjustment factor. Let us define:

$$\psi_t(u, \theta, C) = a(u, x_t, \theta)'[Cy_{t-1} + d(x_t, \theta)] + b(u, x_t, \theta). \tag{3.4}$$

We have by equation (3.1) and the Iterated Expectation Theorem:

$$
\begin{aligned}
& E[\exp(u'y_t - \psi_t(u,\theta,C) + \tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u},\theta,C))|\underline{y_{t-2}}, \underline{f_t}, \underline{x_t}] \\
= \;& E\{E[\exp(u'y_t - \psi_t(u,\theta,C) + \tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u},\theta,C))|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}]|\underline{y_{t-2}}, \underline{f_t}, \underline{x_t}\} \\
= \;& E\{E[\exp(u'y_t - \psi_t(u,\theta,C))|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}]\exp(\tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u},\theta,C))|\underline{y_{t-2}}, \underline{f_t}, \underline{x_t}\} \\
= \;& \exp[a(u,x_t,\theta)'Bf_t]E[\exp(\tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u},\theta,C))|\underline{y_{t-2}}, \underline{f_t}, \underline{x_t}] \\
= \;& \exp[a(u,x_t,\theta)'Bf_t + a(\tilde{u},x_{t-1},\theta)'Bf_{t-1}], \quad \forall u,\tilde{u} \in \mathcal{U}. \qquad (3.5)
\end{aligned}
$$

Then, we apply the conditional expectation given $\underline{x_t}$ only on both sides of the above equation. We get:

Second-order nonlinear moment restrictions:

$$
\begin{aligned}
& E\Big[\exp\left\{u'y_t - \psi_t(u,\theta,C) + \tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u},\theta,C)\right\}|\underline{x_t}\Big] \\
& \qquad = E[\exp\{a(u,x_t,\theta)'Bf_t + a(\tilde{u},x_{t-1},\theta)'Bf_{t-1}\}|\underline{x_t}], \quad \forall u,\tilde{u} \in \mathcal{U}. \qquad (3.6)
\end{aligned}
$$

We get another continuum of moment restrictions, which can be used to complete identification.

To recover the joint Laplace transform of (f_t, f_{t-1}) conditional on $\underline{x_t}$, we need a change of arguments. Under Assumption A.4, there exists a change of variables from argument u to argument $v = (v_1', v_2')' \in \mathbb{C}^K \times \mathbb{C}^{n-K}$, and from argument $\tilde{u}$ to argument $\tilde{v} = (\tilde{v}_1', \tilde{v}_2')' \in \mathbb{C}^K \times \mathbb{C}^{n-K}$, such that $u(v, x_t, \theta, B)$ and $u(\tilde{v}, x_{t-1}, \theta, B)$ satisfy:

$$
a[u(v,x_t,\theta,B),x_t,\theta]'B = v_1', \quad a[u(\tilde{v},x_{t-1},\theta,B),x_{t-1},\theta]'B = \tilde{v}_1', \quad \forall v,\tilde{v} \in \mathcal{V}, x_t, x_{t-1}, \theta, B.
$$

From equation (3.6) we get:

$$
\begin{aligned}
& E\Big[\exp\left\{u(v,x_t,\theta,B)'y_t - \psi_t[u(v,x_t,\theta,B),\theta,C] + u(\tilde{v},x_{t-1},\theta,B)'y_{t-1}\right. \\
& \qquad \left. -\psi_{t-1}[u(\tilde{v},x_{t-1},\theta,B),\theta,C]\right\}|\underline{x_t}\Big] = E[\exp\{v_1'f_t + \tilde{v}_1'f_{t-1}\}|\underline{x_t}], \quad \forall v,\tilde{v} \in \mathcal{V}. \qquad (3.7)
\end{aligned}
$$

For $\tilde{v} = 0$ the equations in (3.7) reduce to the equations in (3.3). Thus, the second-order nonlinear restrictions include the first-order nonlinear restrictions as a special case.

3.3 Third-order nonlinear moment restrictions

The technique developed in Subsection 3.2 can be applied at any lag. In particular, the third-order nonlinear moment restrictions are derived by considering the joint Laplace transform of y_t, y_{t-1}, y_{t-2} given

$\underline{y_{t-3}}, \underline{f_t}, \underline{x_t}$, and accounting for adjustment terms. By using equation (3.5) and applying the Iterated Expectation Theorem, we get:

$$
\begin{aligned}
& E[\exp(u_0' y_t - \psi_t(u_0,\theta,C) + u_1' y_{t-1} - \psi_{t-1}(u_1,\theta,C) + u_2' y_{t-2} - \psi_{t-2}(u_2,\theta,C))|\underline{y_{t-3}}, \underline{f_t}, \underline{x_t}] \\
= \ & E\Big\{E[\exp(u_0' y_t - \psi_t(u_0,\theta,C) + u_1' y_t - \psi_{t-1}(u_1,\theta,C))|\underline{y_{t-2}}, \underline{f_t}, \underline{x_t}] \\
& \exp(u_2' y_{t-2} - \psi_{t-2}(u_2,\theta,C))|\underline{y_{t-3}}, \underline{f_t}, \underline{x_t}\Big\} \\
= \ & \exp[a(u_0,x_t,\theta)' B f_t + a(u_1,x_{t-1},\theta)' B f_{t-1}] E[\exp(u_2' y_{t-2} - \psi_{t-2}(u_2,\theta,C))|\underline{y_{t-3}}, \underline{f_t}, \underline{x_t}] \\
= \ & \exp[a(u_0,x_t,\theta)' B f_t + a(u_1,x_{t-1},\theta)' B f_{t-1} + a(u_2,x_{t-2},\theta)' B f_{t-2}], \quad \forall u_0,u_1,u_2 \in \mathcal{U}.
\end{aligned}
$$

Then, as in Section 3.2, we apply the conditional expectation given $\underline{x_t}$ only on both sides, to get:

Third-order nonlinear moment restrictions:

$$
\begin{aligned}
& E[\exp(u_0' y_t - \psi_t(u_0,\theta,C) + u_1' y_{t-1} - \psi_{t-1}(u_1,\theta,C) + u_2' y_{t-2} - \psi_{t-2}(u_2,\theta,C))|\underline{x_t}] \\
= \ & E\left[\exp[a(u_0,x_t,\theta)' B f_t + a(u_1,x_{t-1},\theta)' B f_{t-1} + a(u_2,x_{t-2},\theta)' B f_{t-2}]|\underline{x_t}\right], \quad \forall u_0,u_1,u_2 \in \mathcal{U}. \qquad (3.8)
\end{aligned}
$$

Similarly to Section 3.2, under Assumption A.4 we can apply the change of variables from u_0, u_1, u_2 to v_0, v_1, v_2 such that $a(u_0,x_t,\theta)'B = v_{0,1}'$, $a(u_1,x_{t-1},\theta)'B = v_{1,1}'$, $a(u_2,x_{t-2},\theta)'B = v_{2,1}'$, where $v_{0,1}$, $v_{1,1}$ and $v_{2,1}$ denote the $(K,1)$ upper blocks of vectors v_0, v_1, v_2, respectively.

3.4 Link to the literature on moment restrictions with latent variables

The moment restrictions derived in Sections 3.1-3.3 are of the form:

$$
E[G(\underline{y_t}, \underline{x_t}, \underline{f_t}; \beta)|\underline{x_t}] = 0, \quad \text{say,}
$$

where β is the parameter vector consisting of the elements of B, C and θ. [3] Unlike the conventional Generalized Method of Moments (GMM) [Hansen (1982)], some of the variables entering the moment function, namely $\underline{f_t}$, are not observable. Therefore, these moment restrictions cannot be used for identification and estimation in a standard GMM setting. Parametrized moment restrictions with unobservable variables have been considered in Schennach (2014). Schennach (2014) introduces a consistent estimation method for the identified set of parameter β. In this method, the distribution of the unobservable variable is

[3] The moment function G has a continuum of components in our setting.

replaced by a least favorable distribution parameterized by an auxiliary parameter γ, say, whose dimension is equal to the number of unconditional moment restrictions. The methodology of Schennach (2014) can be applied to our framework at the cost of two drawbacks. First, in our framework the number of moment restrictions, that is the dimension of vector function G, is typically very large - in principle infinite, due to the continuum of arguments for the Laplace transform - and thus the dimension of the auxiliary parameter is also very large. Second, we are not only interested in the estimation of parameter β, but also in the estimation of the transition p.d.f. of the latent factor f_t. The estimated least favorable distribution does not estimate consistently this p.d.f. In the following sections, we explain how to identify both parameter β and the factor dynamics in parametric and semi-parametric settings. [4]

In a panel framework, Chamberlain (1992) considers models defined by conditional moment restrictions involving unobservable individual effects. After interchanging the role of individuals and time dates, the analysis of Chamberlain (1992) applies to models of the type $E[y_t - d(x_t, \beta) - B(x_t, \beta)f_t|x_t, f_t] = 0$, say, and provides an identification strategy for parameter β based on quasi-differencing. We stress two important differences compared to our framework. First, in our setting the moment function is exponential affine w.r.t. the latent factor f_t, as opposed to linear. Secondly, we consider dynamic latent factors that are possibly serially dependent, while the individual effects are typically assumed cross-sectionally i.i.d. in the panel literature. Therefore, the results in Chamberlain (1992) cannot be directly applied to our framework by simply interchanging the role of indices i and t (even if we were considering a linear model instead of an exponentially linear one).

4 Parametric identification

In this section, we assume that the dynamics of the unobservable effects is specified parametrically and investigate the identifiability of the model parameters from the nonlinear moment restrictions introduced in Section 3. We start with the first-order moment restrictions (Section 4.1) and then consider the higher-order moment restrictions (Section 4.2).

[4] Conditional moment restrictions with latent variables have also been considered in Gallant, Giacomini, Ragusa (2017). They consider a semi-parametric modeling in which the transition of the latent factor is parametrized, but the conditional distribution of the observation given the factor is partly nonparametric.

4.1 First-order nonlinear identification

Let us assume a parametric specification for the dynamics of the latent factor (f_t) conditional on the covariate process (x_t), which yields a fully parametric framework. The discussion in Section 3.1 suggests that identification of model parameters is possible from the first-order nonlinear moment restrictions (3.3), where the conditional expectation in the r.h.s. is taken w.r.t. the model distribution of the unobserved effects given the covariates. More precisely, let us denote by γ the parameter characterizing the conditional distribution of f_t given $\underline{x_t}$ and consider the next definition.

Definition 1. *The regression parameters B, C, θ and the parameter γ are first-order nonlinearly identifiable iff the continuum set of moment restrictions (3.3) admits the true values of the parameters: $B = B_0$, $C = C_0$, $\theta = \theta_0$, $\gamma = \gamma_0$, as the unique solution.*

In models without covariates, first-order nonlinear identification concerns the parameters characterizing the unconditional (stationary) distribution of the factor f_t. In any case, first-order nonlinear identification is not possible for parameters that control the serial dependence in process (f_t) (conditional on the covariates), because restrictions (3.3) do not involve the dynamics of the unobservable effect.

As an illustration of Definition 1, let us discuss first-order nonlinear identification in the factor models encountered in Finance (see Examples 1 and 2). These models have no observable covariates. We consider *i)* a pure Gaussian model, and *ii)* a conditionally Gaussian model with one factor featuring stochastic volatility. For the discussion below, it is important to distinguish the generic parameter values and distributions from their true values. Therefore, we introduce explicitly the subscript $_0$ for true parameter values and expectations under the Data Generating Process (DGP) when it is needed.

Example 2: A Gaussian model with common factor and contagion (continues).

Let us consider the dynamic system:

$$\left\{ \begin{array}{lcl} y_t & = & B f_t + C y_{t-1} + \Sigma^{1/2} \varepsilon_t, \\ \\ f_t & = & \Phi f_{t-1} + (Id - \Phi\Phi')^{1/2} \eta_t, \end{array} \right. \tag{4.1}$$

where the vector y_t of endogenous variables is n-dimensional, the vector f_t of unobservable factors is K-dimensional, with $K \leq n$, the matrix of loadings B has full rank K, and the vector of error terms is a standard Gaussian white noise $(\varepsilon_t', \eta_t')' \sim IIN(0, Id)$. We normalize the unobservable factor f_t by setting its unconditional variance equal to the identity matrix. The serial dependence between the current and lagged values of the endogenous variable is due to the dynamics of the common factor, but also to the effect of lagged observable value y_{t-1}, which represents the contagion phenomena. The joint process $(y_t', f_t')'$ admits a Gaussian structural VAR(1) representation, and the stationarity condition requires that the eigenvalues of matrices C and Φ are smaller than one in modulus.

Let us distinguish three cases, depending on the restrictions that are imposed on the matrix parameters C, Φ and Σ in the model specification.

i) Let us first consider the special case where $C = 0$, $\Phi = 0$, and the variance-covariance matrix Σ is diagonal. We get $y_t = B\eta_t + \Sigma^{1/2}\varepsilon_t$. Thus, the observations y_t, for t varying, are $IIN(0, \Sigma + BB')$, and the cross-sectional dependence is entirely captured by the effect of the unobservable common factor f_t. This is the static linear factor model used in exploratory factor analysis [Lawley, Maxwell (1971), Anderson (2003)]. The first-order nonlinear moment restrictions (3.2) satisfied by the true parameter values Σ_0, B_0 become:

$$E_0[\exp(u'y_t)] = \exp\left\{\frac{u'(\Sigma + BB')u}{2}\right\}, \quad \forall u \in \mathbb{R}^n, \tag{4.2}$$

where E_0 denotes the expectation with respect to the DGP. Therefore, the continuum of moment restrictions (4.2) allows for identifying the value of the transformed parameter $\Sigma_0 + B_0B_0'$. However, the identification of matrix $\Sigma_0 + B_0B_0'$ is in general not sufficient for identifying matrices Σ_0 and B_0 themselves. For instance, we have $\Sigma_0 + B_0B_0' = \Sigma_0 + B_0Q(B_0Q)'$ for any orthogonal (K, K) matrix Q. Therefore, identification requires additional constraints, such as the assumption that matrix $B_0'B_0$ is diagonal. Then, the order condition is satisfied if the number of linearly independent identifiable parameters $n(n+1)/2$ is larger than, or equal to, the number of free structural parameters $n(1+K) - K(K-1)/2$, that is, if $\frac{1}{2}[(n-K)^2 - (n+K)] \geq 0$. This order condition is not sufficient for identifying the matrices Σ_0 and B_0 [see e.g. Lawley, Maxwell (1971), Section 2.3, for counterexamples]. Anderson, Rubin (1956) provides either sufficient, or necessary, conditions for identification of Σ_0, B_0 from $\Sigma_0 + B_0B_0'$.

ii) Let us now consider the case with $\Phi = 0$, i.e. a static factor, but allowing for nonzero contagion matrix C. Moreover, let us still assume that the variance-covariance matrix Σ is diagonal. Then, process (y_t) follows a VAR(1) model with constrained variance-covariance matrix $\Sigma + BB'$ of the innovations.

The first-order nonlinear moment restrictions satisfied by the true parameter values C_0, B_0, Σ_0 are:

$$E_0\left[\exp\{u'(y_t - Cy_{t-1})\}\right] = \exp\left\{\frac{u'(\Sigma + BB')u}{2}\right\}, \quad \forall u \in \mathbb{R}^n. \tag{4.3}$$

Since the joint vector $(y_t', y_{t-1}')'$ is multivariate Gaussian, the continuum of moment restrictions (4.3) is equivalent to the $n(n+1)/2$ equations:

$$V_0(y_t - Cy_{t-1}) = \Sigma + BB', \tag{4.4}$$

that is:

$$\Gamma_0(0) - \Gamma_0(1)C' - C\Gamma_0(1)' + C\Gamma_0(0)C' = \Sigma + BB', \tag{4.5}$$

where $\Gamma_0(0) = V_0(y_t)$ and $\Gamma_0(1) = Cov_0(y_t, y_{t-1})$ are the true unconditional variance, and the true first-order autocovariance, respectively, of process (y_t).

Let us now study identification of the true parameter values from equation (4.4). We have:

$$\begin{aligned} V_0(y_t - Cy_{t-1}) &= V_0[y_t - C_0 y_{t-1} - (C - C_0)y_{t-1}] \qquad (4.6)\\ &= V_0[B_0 f_t + \Sigma_0^{1/2}\varepsilon_t - (C - C_0)y_{t-1}] = \Sigma_0 + B_0 B_0' + (C - C_0)\Gamma_0(0)(C - C_0)'. \end{aligned}$$

Thus, we have to look for the solutions (Σ, B, C) of the matrix equation:

$$\Sigma_0 + B_0 B_0' + (C - C_0)\Gamma_0(0)(C - C_0)' = \Sigma + BB', \tag{4.7}$$

where B_0 and B are matrices of full rank K, such that $B_0'B_0$ and $B'B$ are diagonal. This matrix equation corresponds to a system of $n(n+1)/2$ nonlinear equations that involve $n(1+K+n) - K(K-1)/2$ unknown free parameters in matrices Σ, B and C. For any $n \geq K \geq 1$, we have $n(1 + K + n) - K(K-1)/2 > n(n+1)/2$, i.e., more unknown parameters than equations, and the order condition is not satisfied. Let us denote $\mathcal{E}_0$ the set of solutions (Σ, B, C) of system (4.7). This set is identifiable by construction. We deduce easily from (4.7) the following property:

Proposition 1. *We have* $(\Sigma, B, C) \in \mathcal{E}_0$ *if, and only if:*

$$\Sigma - \Sigma_0 + BB' - B_0 B_0' \succeq 0,$$

where $\succeq$ *denotes the ordering on symmetric matrices. Then,* $C = C_0 - D\Gamma_0(0)^{-1/2}$*, where* D *is a* (n, n) *matrix such that* $DD' = \Sigma - \Sigma_0 + BB' - B_0 B_0'$*.*

We deduce the next two corollaries.

Corollary 1. *The parameters are not first-order nonlinearly identifiable.*

Proof: Parameter (Σ_0, B_0, C_0) is first-order identifiable if, and only if, the set $\mathcal{E}_0$ is a singleton, consisting of the single element (Σ_0, B_0, C_0). This condition is not satisfied. For instance, the condition $\Sigma - \Sigma_0 + BB' - B_0B_0' \succeq 0$ is satisfied when the smallest eigenvalue of matrix $\Sigma - \Sigma_0$ is larger than, or equal to, the largest eigenvalue of matrix B_0B_0'. Then, there exist several ways to derive matrix D, for instance by considering the square root of matrix $\Sigma - \Sigma_0 + BB' - B_0B_0'$, or its Cholevski decomposition. Q.E.D.

Even if the identifiable set $\mathcal{E}_0$ is not a singleton, the next Corollary shows that the parameters are known functions of the identifiable set $\mathcal{E}_0$ - under the identification condition for a static factor model (see paragraph *i)*).

Corollary 2. *Suppose that parameters B_0 and Σ_0 are identifiable from matrix $\Sigma_0 + B_0B_0'$. Then, the parameters are identifiable from the identifiable set $\mathcal{E}_0$.*

Proof: From Proposition 1 we have $\Sigma + BB' \succeq \Sigma_0 + B_0B_0'$ for any $(\Sigma, B, C) \in \mathcal{E}_0$. We deduce:

$$\Sigma_0 + B_0B_0' = \min\{\Sigma + BB' : (\Sigma, B, C) \in \mathcal{E}_0\},$$

where the minimum is with respect to the ordering on symmetric matrices and is unique. Thus, $\Sigma_0 + B_0B_0'$ is identifiable. By assumption, the identifiability of Σ_0 and B_0, and hence of C_0, follows. Q.E.D.

Corollary 2 shows that **the standard order condition is not necessary for identification**, when the true parameter value is characterized by some extremality property in the identifiable set. Further, if we assume $\Sigma_0 = \sigma_0^2 Id_n$ (cross-sectionally homoskedastic errors), Corollary 2 together with the remark that scalar σ_0^2 is the smallest eigenvalue of matrix $\Sigma_0 + B_0B_0'$, imply that the parameters are identifiable from set $\mathcal{E}_0$ (see the Supplementary Materials for a consistent estimation method based on this idea).

iii) Finally, let us consider the general case in which the autoregressive matrices Φ and C, and the variance-covariance matrix Σ, are unconstrained, that is, a model with both contagion and dynamic factor, and conditional contemporaneous correlations. Darolles, Dubecq, Gourieroux (2014) show that the factor loading matrix B is identifiable under a full-rank condition for a multivariate partial autocovariance of order 2 of the observable process. In the light of this result involving second-order autocovariances, we expect that the nonlinear moment restrictions at order 1 in (4.5) are not sufficient for identification. Let us check this. When the unobservable factor is dynamic, $y_t - C_0 y_{t-1} = B_0 f_t + \Sigma_0^{1/2}\varepsilon_t$ is not necessarily uncorrelated with the lagged value y_{t-1}. From equations (4.4) and (4.6), the first-order nonlinear moment restrictions are:

$$\Sigma_0 + B_0 B_0' + (C - C_0)\Gamma_0(0)(C - C_0)' + \Lambda_0 (C - C_0)' + (C - C_0)\Lambda_0' = \Sigma + BB', \tag{4.8}$$

where matrix $\Lambda_0 = Cov_0(y_t - C_0 y_{t-1}, y_{t-1}) = \Gamma_0(1) - C_0\Gamma_0(0)$ does not vanish in general. The matrix equation (4.8) does not involve the factor dynamic parameter Φ, which is therefore unidentifiable from the first-order restrictions. Moreover, this matrix equation does not admit a unique solution for (Σ, B, C), and parameters (Σ_0, B_0, C_0) are not first-order nonlinearly identifiable as well (see Proposition 2 in the next subsection, which shows parameter unidentifiability even from second-order moment restrictions). Furthermore, the extremal property of matrix $\Sigma + BB'$ with respect to the solution set $\mathcal{E}_0$ does not apply when matrix Φ is not restricted to be zero in the model, and the parameters (Σ_0, B_0, C_0) are not uniquely defined from the identifiable set of solutions of equation (4.8).

The first-order moment restrictions inherit the identification problems usually encountered with the Gaussian structural VAR model [see e.g. Sims (1980), Forni et al. (2000)]. In the general case the parameters are not identifiable from conditions (4.8). However they may become first-order identifiable if additional structural restrictions on parameter values complete appropriately the conditions (4.8).

Example 5: A conditionally Gaussian factor model with stochastic volatility in the factor.

Let us consider the multivariate model:

$$y_t = \beta f_{1,t}^{1/2}\eta_t + \varepsilon_t, \tag{4.9}$$

where $f_t = (f_{1,t}, \eta_t)'$ is a bivariate factor, the scalar processes $(f_{1,t})$, (η_t) and the n-dimensional process (ε_t) are mutually independent, with $\varepsilon_t \sim IIN(0, \Sigma)$ and $\eta_t \sim IIN(0, 1)$. The process $(f_{1,t})$ is an Autoregressive Gamma (ARG) Markov process with conditional Laplace transform:

$$E[\exp(u f_{1,t})|f_{1,t-1}] = \exp\left\{\frac{\rho u f_{1,t-1}}{1 - \delta u} - \nu\log(1 - \delta u)\right\}, \tag{4.10}$$

corresponding to a noncentral gamma transition distribution, where $\nu > 0$ is the degree of freedom parameter, $\delta > 0$ is a scale parameter, and ρ, with $\rho \in [0, 1)$, is the first-order autocorrelation (see Section 2.2). The unconditional distribution of $f_{1,t}$ is a central gamma distribution and, without loss of generality, it is possible to choose the parameters of the ARG process such that the unconditional scale parameter is equal to 1. This condition applies when $\delta = 1 - \rho$. Then, the stationary distribution is $f_{1,t}$ is the gamma distribution $\gamma(\nu)$ with parameter ν.

Let us consider the first-order nonlinear moment restrictions. They are given by:

$$
\begin{aligned}
E[\exp(u'y_t)] &= \exp\left(\frac{u'\Sigma u}{2}\right) E\left[\exp\left(u'\beta f_{1,t}^{1/2}\eta_t\right)\right] \\
&= \exp\left(\frac{u'\Sigma u}{2}\right) E\left[\exp\left\{\frac{(u'\beta)^2 f_{1,t}}{2}\right\}\right] \quad \text{(by the Iterated Expectation Theorem)} \\
&= \exp\left\{\frac{u'\Sigma u}{2} - \nu \log\left(1 - \frac{(u'\beta)^2}{2}\right)\right\}, \quad \forall u \in \mathcal{U},
\end{aligned}
$$

where argument u varies in a suitable neighbourhood $\mathcal{U}$ of the origin. Thus, for first-order nonlinear identification we have to consider the system:

$$
\frac{u'\Sigma u}{2} - \nu \log\left(1 - \frac{(u'\beta)^2}{2}\right) = \frac{u'\Sigma_0 u}{2} - \nu_0 \log\left(1 - \frac{(u'\beta_0)^2}{2}\right), \quad \forall u \in \mathcal{U}. \tag{4.11}
$$

By using the linear independence between the quadratic and logarithmic functions, we deduce:

$$
\begin{aligned}
&u'\Sigma u = u'\Sigma_0 u, \ \forall u \in \mathcal{U} \quad \Leftrightarrow \quad \Sigma = \Sigma_0, \\
&\nu \log\left(1 - \frac{(u'\beta)^2}{2}\right) = \nu_0 \log\left(1 - \frac{(u'\beta_0)^2}{2}\right), \forall u \in \mathcal{U}.
\end{aligned}
$$

By identifying the quadratic and quartic terms in argument u in the expansion of the two sides of the second equation in a neighbourhood of $u = 0$, we get: $\nu = \nu_0, (u'\beta)^2 = (u'\beta_0)^2$, $\forall u \in \mathcal{U}$, which implies $\beta = \beta_0$. Therefore, the true values of the parameters ν_0, β_0 and Σ_0 are first-order nonlinearly identifiable. As expected, the correlation parameter ρ_0 of the ARG factor dynamics is unidentifiable from first-order moment restrictions.

Parameter first-order identifiability applies in the nonlinear and non Gaussian model of Example 5, whereas it does not in Example 2. Indeed, the different laws of common factor $f_{1,t}$ and innovation ε_t in Example 5 imply that the nonlinear moment restrictions lead to a continuum of equations [see (4.11)], while in the linear Gaussian model of Example 2 this continuum of restrictions boil down to a finite number of equations [see equation (4.5)].

4.2 Second- and third-order nonlinear identification

When the parameters are not first-order identifiable, we can increase the set of nonlinear moment restrictions to second- and third-order.

Definition 2. *The parameters B, C, θ and γ are second-order (resp., third-order) nonlinearly identifiable iff the true parameter values B_0, C_0, θ_0, γ_0 yield the unique solution to the set of second-order (resp. third-order) moment restrictions.*

As noted before, the second- (resp., third-) order nonlinear moment restrictions include as special case the first- (resp., the first- and second-) order nonlinear moment restrictions. Therefore, the parameters can be more often second- (resp., third-) order identifiable than first- (resp., second-) order identifiable. In particular, we expect that parameters characterizing the serial dependence of the unobservable factors are second-order identifiable, whereas they are not first-order identifiable as already remarked in Section 4.1.

Example 2: A Gaussian model with common factor and contagion (continues).

Let us consider the unrestricted specification in (4.1). The nonlinear moment restrictions at order 1, 2 and 3 are derived in Appendix 1.1. They are:

$$\begin{aligned}
\Gamma_0(0) + C\Gamma_0(0)C' - \Gamma_0(1)C' - C\Gamma_0(1)' &= BB' + \Sigma, \text{ for order } 1, \\
\Gamma_0(1) + C\Gamma_0(1)C' - \Gamma_0(2)C' - C\Gamma_0(0) &= B\Phi B', \text{ to be added for order } 2, \\
\Gamma_0(2) + C\Gamma_0(2)C' - \Gamma_0(3)C' - C\Gamma_0(1) &= B\Phi^2 B', \text{ to be added for order } 3,
\end{aligned} \tag{4.12}$$

where the autocovariance function $\Gamma_0(h) = Cov_0(y_t, y_{t-h})$ is identifiable. We explain in the Supplementary Materials that these nonlinear restrictions are well-chosen (parameter-dependent) linear combinations of a subset of the Yule-Walker equations for the VAR(1) process $(y_t', f_t')'$, which involve the autocovariance function of the observable component (y_t) only.

In the unrestricted specification, the number of unknown parameters is $n^2+nK+K^2+n(n+1)/2$ for matrices C, B, Φ and Σ, respectively, subject to $K(K-1)/2$ identification conditions from the constraint on matrix $B'B$ to be diagonal. The numbers of independent nonlinear moment restrictions are: $n(n+1)/2$ at order 1, $n(n+1)/2+n^2$ at order 2, and $n(n+1)/2+2n^2$ at order 3, respectively. As already remarked in Section 4.1, Example 2, the parameters cannot be identified from moment restrictions at order 1 only. It is easily verified that the order condition for second-order nonlinear identification is also not satisfied. In fact, the next proposition shows that the second-order moment restrictions do not identify the true model parameters.

Proposition 2. *In the Gaussian linear model with common factor and contagion, for any n and K the true parameter value $(B_0, C_0, \Phi_0, \Sigma_0)$ is not second-order nonlinearly identifiable.*

The proof of Proposition 2 in Appendix 1.2 shows that the lack of identification from second-order restrictions holds true even when the restriction of a diagonal variance-covariance matrix Σ is imposed. Under this restriction, the order condition for identification at order 2 is $n^2 \geq n(2K+1) + K^2 + K$ and is satisfied if the number of factors K is sufficiently small compared to the number of endogenous variables n. Proposition 2 shows that the order condition is not sufficient for second-order identification. The proof of Proposition 2 also shows that the lack of identification applies both locally and globally.

Finally, the order condition for third-order identification in the unrestricted model is $2n^2 \geq K(2n + K + 1)$. This inequality is satisfied when the number of factors K is sufficiently small with respect to the dimension n of the endogenous variable. Due to the nonlinearity of the system of matrix equations (4.12), the analysis of global identification from third-order nonlinear moment restrictions is challenging and is left for future research.

Example 5: A conditionally Gaussian factor model with stochastic volatility in the factor (continues).

The second-order nonlinear moment restrictions are based on the unconditional Laplace transform of the joint process $(y_t', y_{t-1}')'$. We get:

$$\begin{aligned} & E[\exp(u'y_t + \tilde{u}'y_{t-1})] = E\{E[\exp(u'y_t + \tilde{u}'y_{t-1})|\underline{f_{1,t}}]\} \\ = \ & \exp\left(\frac{1}{2}u'\Sigma u + \frac{1}{2}\tilde{u}'\Sigma\tilde{u}\right) E\left[\exp\left(\frac{(u'\beta)^2}{2}f_{1,t} + \frac{(\tilde{u}'\beta)^2}{2}f_{1,t-1}\right)\right], \quad \forall u, \tilde{u} \in \mathcal{U}. \end{aligned} \tag{4.13}$$

Parameters β_0 and Σ_0 are first-order nonlinearly identified (see Section 4.1). Let us now check how to identify the factor dynamics from the second-order nonlinear moment restrictions. Let us assume $\beta_{0,1} \neq 0$, and apply the conditions above to argument vectors $u = (\sqrt{2v}/\beta_{0,1}, 0, \ldots, 0)'$ and $\tilde{u} = (\sqrt{2\tilde{v}}/\beta_{0,1}, 0, \ldots, 0)'$, where the real arguments $v, \tilde{v} \in \mathcal{V}$ are positive. We get:

$$E_0\left[\exp\left(\frac{\sqrt{2v}}{\beta_{0,1}}y_{1,t} + \frac{\sqrt{2\tilde{v}}}{\beta_{0,1}}y_{1,t-1}\right)\right] = E\left[\exp\left(v(f_{1,t} + \lambda_0) + \tilde{v}(f_{1,t-1} + \lambda_0)\right)\right], \quad \forall v, \tilde{v} \in \mathcal{V},$$

where the shift factor $\lambda_0 = \sigma_{0,1}^2/\beta_{0,1}^2$ is identified from the first-order restrictions. Thus, we can use these second-order nonlinear moment restrictions to identify the dynamics of common factor $f_{1,t}$. In particular, parameter ρ_0 corresponding to factor autocorrelation is identified from $\rho_0 = \dfrac{\partial^2\psi_0(0,0)}{\partial v \partial \tilde{v}}\left[\dfrac{\partial^2\psi_0(0,0)}{\partial v^2}\right]^{-1}$, where $\psi_0(v, \tilde{v}) := \log E_0[\exp(vf_{1,t} + \tilde{v}f_{1,t-1})] = \log E_0\left[\exp\left(\dfrac{\sqrt{2v}}{\beta_{0,1}}y_{1,t} + \dfrac{\sqrt{2\tilde{v}}}{\beta_{0,1}}y_{1,t-1}\right)\right] - \lambda_0(v + \tilde{v})$. The second-order partial derivatives of function ψ_0 in $(0,0)$ are identifiable from the function values for positive arguments $v, \tilde{v} \in \mathcal{V}$, because ψ_0 is twice differentiable in a neighbourhood of the origin.

5 Semi-parametric identification

In this section, the dynamics of the unobservable effects is let unspecified, which leads to a semi-parametric framework. We investigate the identifiability of the nonlinear regression parameters and of the distribution function of the unobservable effects (conditionally on the covariates) based on second-order nonlinear moment restrictions. Semiparametric identification from third-order nonlinear restrictions is discussed in the Supplementary Materials.

5.1 Second-order nonlinear identification

The second-order nonlinear moment restrictions (3.7) involve both the finite-dimensional regression parameters B, C, θ and the functional parameter, that is the joint distribution of f_t, f_{t-1}, or equivalently the transition distribution of the unobservable effect (conditionally on the covariates). As noted in Section 3.2, they contain as special case the first-order nonlinear moment restrictions. Let us assume that $K < n$, that is, the number of latent factors is strictly smaller than the number of endogenous observable variables. We first assume that K is known, and leave the discussion on the identification of K to the examples. The r.h.s. of equation (3.7) is independent of the arguments v_2, $\tilde{v}_2$. Thus, the l.h.s. of (3.7), which is identifiable up to the regression parameters B, C, θ, is independent of v_2, $\tilde{v}_2$ as well, when evaluated at the true parameters. By differentiating the l.h.s. of (3.7) w.r.t. v_2, and $\tilde{v}_2$, and equalizing the gradients to zero, we get a continuum set of conditional moment restrictions that only involves the finite-dimensional regression parameters:

$$\begin{aligned}
& E\Big[\frac{\partial u(v,x_t,\theta,B)'}{\partial v_2}\left(y_t - \frac{\partial a[u(v,x_t,\theta,B),x_t,B]'}{\partial u}Cy_{t-1} - \frac{\partial b[u(v,x_t,\theta,B),x_t,B]}{\partial u}\right) \\
& \times \exp\left\{u(v,x_t,\theta,B)'y_t - \psi_t[u(v,x_t,\theta,B),\theta,C] + u(\tilde{v},x_{t-1},\theta,B)'y_{t-1} - \psi_{t-1}[u(\tilde{v},x_{t-1},\theta,B),\theta,C]\right\}|\underline{x_t}\Big] \\
& = 0, \quad \forall v,\tilde{v} \in \mathcal{V},
\end{aligned} \tag{5.1}$$

and:

$$\begin{aligned}
& E\Big[\frac{\partial u(\tilde{v},x_{t-1},\theta,B)'}{\partial \tilde{v}_2}\left(y_{t-1} - \frac{\partial a[u(\tilde{v},x_{t-1},\theta,B),x_{t-1},B]'}{\partial u}Cy_{t-2} - \frac{\partial b[u(\tilde{v},x_{t-1},\theta,B),x_{t-1},B]}{\partial u}\right) \\
& \times \exp\left\{u(v,x_t,\theta,B)'y_t - \psi_t[u(v,x_t,\theta,B),\theta,C] + u(\tilde{v},x_{t-1},\theta,B)'y_{t-1} - \psi_{t-1}[u(\tilde{v},x_{t-1},\theta,B),\theta,C]\right\}|\underline{x_t}\Big] \\
& = 0, \quad \forall v,\tilde{v} \in \mathcal{V},
\end{aligned} \tag{5.2}$$

respectively.

Definition 3. *The regression parameters B, C, θ are semi-parametrically second-order nonlinearly identifiable if the continuum set of conditional nonlinear moment restrictions in (5.1) and (5.2) admits the true parameter values as unique solution:* $B = B_0, C = C_0, \theta = \theta_0$.

Our identification strategy for the regression parameters is similar in spirit to a differencing approach to eliminate the distribution of the unobservable time effects from the conditional moment restrictions. Differencing methods are popular for eliminating unobservable individual effects in panel data, see for instance Chamberlain (1992) for linear models, Bonhomme (2012) for a recent contribution in a nonlinear setting relying on the conditional p.d.f. of the observable variables, and the analysis in Section 6.

From equation (3.7) we get immediately the following proposition:

Proposition 3. *If the regression parameters B, C, θ are semi-parametrically second-order nonlinearly identifiable, the joint Laplace transform of (f_t, f_{t-1}) conditional on $\underline{x_t}$ is identifiable (in a suitable neighbourhood of the origin).*

Before investigating the conditions for second-order nonlinear identifiability in some examples, let us briefly compare this semi-parametric identifiability concept with the conditions for nonparametric identification in Markov processes with latent components provided in Hu, Shum (2012). Under the latter conditions, the model is nonparametrically identifiable from the joint density of y_t, y_{t-1}, y_{t-2}, y_{t-3}, or equivalently from all the cross-moments based on y_t, y_{t-1}, y_{t-2}, y_{t-3}. The nonparametric identification in Hu, Shum (2012) is largely based on the assumption of a unique spectral decomposition of the conditional expectation operator of y_t, y_{t-3} given y_{t-1}, y_{t-2}. [5] For our identification strategy based on the structure of the conditional Laplace transform of the affine regression model (Assumption A.1), a triplet of variables y_t, y_{t-1}, y_{t-2} and a specific subset of cross-moment restrictions suffice.

5.2 Multivariate Poisson model with common stochastic intensity

The model is defined by:

$$y_{i,t} \sim \mathcal{P}(\beta_i' f_t), \quad i = 1, \ldots, n, \tag{5.3}$$

[5] This assumption is used to apply the identification strategy of Carroll, Chen, Hu (2010) introduced for models with nonclassical measurement errors. Models with measurement errors can be written as specifications with observable and unobservable components featuring specific restrictions. See e.g. Wilhelm (2013) for nonparametric identification of panel data models with measurement errors in the regressors.

where the count variables are independent conditional on the unobservable factor process (f_t), and the factor dynamics is let unspecified. The factor process and the regression coefficients are strictly positive. This model corresponds to a special case of the specification in Example 3, when the intercepts and the contagion coefficients are zero. Since the factor is unobservable and its dynamic unspecified, the identifiable nonlinear regression parameters are at best the elements of the range $\mathcal{R}(B)$ of the (n, K) loadings matrix B with rows β_i'. We normalize the matrix B such that it has full column rank K and $B'B = Id_K$.

Let us derive the second-order nonlinear moment restrictions. From the Laplace transform of the Poisson distribution we have:

$$E\left[\exp\left(u'y_t + \tilde{u}'y_{t-1}\right)\right] = E\left[\exp\left\{\xi(u)'Bf_t + \xi(\tilde{u})'Bf_{t-1}\right\}\right], \quad \forall u, \tilde{u} \in \mathcal{U}, \tag{5.4}$$

where the components of the $(n, 1)$ vector $\xi(u)$ are $\xi_i(u) = e^{u_i} - 1$. Let $B_\perp$ be a $(n, n-K)$ matrix whose columns define an orthonormal basis of $\mathcal{R}(B)^\perp$, i.e. the linear space orthogonal to the range of B, and define the (n, n) orthogonal matrix $\tilde{B} = (B \vdots B_\perp)$ with rows $\tilde{\beta}_i'$, say. This matrix is unique, as a function of the column space of B, up to a block-diagonal rotation. We use the change of variables:

$$v' = \xi(u)'\tilde{B} \qquad \Leftrightarrow \qquad \xi(u) = \tilde{B}v,$$

which yields function $u(v, B)$ with components $u_i(v, B) = \log(\tilde{\beta}_i'v + 1)$. Thus, we get:

$$E\left[\exp\left\{u(v, B)'y_t + u(\tilde{v}, B)'y_{t-1}\right\}\right] = E\left[\exp\left\{v_1'f_t + \tilde{v}_1'f_{t-1}\right\}\right], \quad \forall v, \tilde{v} \in \mathcal{V}. \tag{5.5}$$

By using:

$$\frac{\partial u(v, B)'}{\partial v_2} = B_\perp' D(v, B)^{-1}, \qquad \text{with } D(v, B) = diag(\tilde{\beta}_i'v + 1),$$

the continuum set of moment restrictions (5.1) becomes:

$$E\left[B_\perp' D(v, B)^{-1} y_t \exp\left\{u(v, B)'y_t + u(\tilde{v}, B)'y_{t-1}\right\}\right] = 0, \quad \forall v, \tilde{v} \in \mathcal{V}. \tag{5.6}$$

Let us now introduce an assumption on the factor dynamics.

Assumption 1. *The factor process (f_t) is **not** i.i.d., and the log joint Laplace transform $\psi_f(w, \tilde{w}) = \log E[\exp(w'f_t + \tilde{w}'f_{t-1})]$ is such that:*

$$\eta' \frac{\partial^2 \psi_f(w, \tilde{w})}{\partial w \partial \tilde{w}'} \eta = 0, \ \eta \in \mathbb{R}^K, \ \forall w, \tilde{w} \in \mathcal{W} \ \Rightarrow \eta = 0, \tag{5.7}$$

for any neighbourhood $\mathcal{W}$ of 0.

We have $\dfrac{\partial^2 \psi_f(w,\tilde{w})}{\partial w \partial \tilde{w}'} = \widetilde{Cov}(f_t, f_{t-1})$, where $\widetilde{Cov}$ denotes covariance w.r.t. the modified measure with expectation $\tilde{E}(Z) = E[\exp(w'f_t + \tilde{w}'f_{t-1})Z]/E[\exp(w'f_t + \tilde{w}'f_{t-1})]$, which depends on w, $\tilde{w}$. Hence, Assumption 1 requires that, if $\eta' f_t$ is a linear combination of the components of f_t which has a zero first-order autocorrelation under the modified measure for any $w, \tilde{w} \in \mathcal{W}$, then this linear combination is the zero process. This condition is not satisfied if process (f_t) is a strong white noise. This explains why we exclude this possibility in Assumption 1. Under the additional condition that function $\dfrac{\partial^2 \psi_f}{\partial w \partial \tilde{w}'}$ is analytic in a neighbourhood of 0 in the complex domain, we can establish a closer link between Assumption 1 and the white noise properties of the factor process (see the Supplementary Materials for the proofs). Namely, when the factor is one-dimensional, Assumption 1 is equivalent to: Process (f_t) is *not* i.i.d. In the general framework with multidimensional factor, Assumption 1 *is implied by* the following condition:

$$\text{If } \eta' f_t \text{ and } \eta' f_{t-1} \text{ are independent, for a } \eta \in \mathbb{R}^K, \text{ then } \eta = 0. \tag{5.8}$$

The latter condition is related to the literature on codependence [Gourieroux, Peaucelle (1992)] and serial correlation common features [Engle, Kosicki (1993), Vahid, Engle (1993)].

Proposition 4. *Under Assumption 1, the range of loadings matrix B in the Poisson model (5.3) is semi-parametrically second-order nonlinearly identifiable.*

Proof: Let us show that (5.6) admits the unique solution $B = B_0$ up to a rotation. For this purpose, let us first rewrite (5.6) in terms of the true factor dynamics. From (5.4) evaluated at the true parameter values, we have:

$$\begin{aligned} E\left[y_t \exp\left(u'y_t + \tilde{u}'y_{t-1}\right)\right] &= \frac{\partial}{\partial u} E\left[\exp\left(u'y_t + \tilde{u}'y_{t-1}\right)\right] = \frac{\partial}{\partial u} \Psi_f[B_0'\xi(u), B_0'\xi(\tilde{u})] \\ &= \frac{\partial \xi(u)'}{\partial u} B_0 \frac{\partial \Psi_f}{\partial w}[B_0'\xi(u), B_0'\xi(\tilde{u})] = diag(e^{u_i}) B_0 \frac{\partial \Psi_f}{\partial w}[B_0'\xi(u), B_0'\xi(\tilde{u})], \end{aligned}$$

where $\Psi_f(w, \tilde{w}) = \exp[\psi_f(w, \tilde{w})]$ denotes the joint Laplace transform of f_t, f_{t-1}. We get:

$$E\left[y_t \exp\left\{u(v,B)'y_t + u(\tilde{v},B)'y_{t-1}\right\}\right] = D(v,B) B_0 \frac{\partial \Psi_f}{\partial w}[B_0'\tilde{B}v, B_0'\tilde{B}\tilde{v}].$$

Thus, the conditional moment restrictions in (5.6) are equivalent to:

$$B_\perp' B_0 \frac{\partial \Psi_f}{\partial w}[B_0'\tilde{B}v, B_0'\tilde{B}\tilde{v}] = 0, \quad \forall v, \tilde{v} \in \mathcal{V}. \tag{5.9}$$

Since $\Psi_f > 0$, we can replace Ψ_f by ψ_f in (5.9). Moreover, we can differentiate the resulting equation w.r.t. $\tilde{v}_2$ to get:

$$B'_\perp B_0 \frac{\partial^2 \psi_f}{\partial w \partial \tilde{w}'}[B'_0\tilde{B}v, B'_0\tilde{B}\tilde{v}]B'_0 B_\perp = 0, \quad \forall v, \tilde{v} \in \mathcal{V}. \tag{5.10}$$

For any B, the matrix $B'_0\tilde{B}$ has rank K. Hence, vectors $B'_0\tilde{B}v$, for v varying, span $\mathbb{R}^K$. Therefore, (5.10) is equivalent to:

$$B'_\perp B_0 \frac{\partial^2 \psi_f(w, \tilde{w})}{\partial w \partial \tilde{w}'} B'_0 B_\perp = 0, \quad \forall w, \tilde{w} \in \mathcal{W}^*, \tag{5.11}$$

in a suitable neighbourhood $\mathcal{W}^*$ of the origin. From condition (5.7) in Assumption 1, we get $B'_0 B_\perp = 0$. Thus, the column space of $B_\perp$ belongs to the orthogonal complement of the column space of B_0, i.e. $\mathcal{R}(B)^\perp \subset \mathcal{R}(B_0)^\perp$. Since these linear spaces have the same dimension $n - K$, we get $\mathcal{R}(B) = \mathcal{R}(B_0)$. Q.E.D.

The proof of Proposition 4 shows that semi-parametric second-order identification can be achieved by a subset of the nonlinear moment restrictions (5.1)-(5.2), and that the number of unobservable factors is identifiable as well from the continuum of moment restrictions (5.6).

Proposition 5. *Under Assumption 1, the true number of factors K_0, say, is equal to the smallest dimension K of the column space of matrix B such that the continuum of moment restrictions (5.6) admits a solution.*

Proof: From the proof of Proposition 4 it follows that, if the continuum of moment restrictions (5.6) admits a solution for a (n, K) matrix B, then $B'_0 B_\perp = 0$, where B_0 is a full-rank (n, K_0) matrix and $B_\perp$ is $(n, n-K)$. If $K < K_0$, we cannot find $n - K$ linearly independent vectors that are orthogonal to the columns of B_0. In that case the continuum of moment restrictions (5.6) does not admit a solution. Q.E.D.

Finally, from Proposition 3 the joint Laplace transform of f_t, f_{t-1} is identifiable from equation (5.5) evaluated at the true value B_0.

5.3 Semi-parametric identification of the non Gaussian linear model with contagion and common factor

In this section we focus on the semi-parametric identification in the non Gaussian linear model with contagion and common factor:

$$y_t = C y_{t-1} + B f_t + \varepsilon_t, \tag{5.12}$$

where (f_t) is a Markov process and (ε_t) is a strong white noise process independent of (f_t). Contrary to Example 2 studied in Section 4, the distributions of processes (ε_t) and (f_t) are let unspecified. We assume that the unconditional distribution of ε_t and the joint distribution of (f_t, f_{t-1}) admit Laplace transforms in a neighbourhood of the zero arguments, that we denote by $E[\exp(u'\varepsilon_t)] = \exp[\psi_\varepsilon(u)]$ and $E[\exp(v'f_t + w'f_{t-1})] = \exp[\psi_f(v,w)]$, respectively. Model (5.12) satisfies the exponential affine property in Assumption A.1, with an infinite-dimensional parameter θ corresponding to the log Laplace transform ψ_ε of the error distribution. The full parameter vector $(B, C, \psi_\varepsilon, \psi_f)$ contains both finite-dimensional and infinite-dimensional components, and we denote by $(B_0, C_0, \psi_\varepsilon^0, \psi_f^0)$ the true parameter values. We assume that the (unknown) number K of factors is strictly smaller than the number of endogenous variables n, and matrix B_0 has full column-rank.

i) Second-order identification of the number of factors, factor loadings and contagion matrix

The second-order nonlinear moment restrictions are:

$$\begin{aligned} E\left[\exp\left\{u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2})\right\}\right] &= E\left[\exp\left(u'\varepsilon_t + \tilde{u}'\varepsilon_{t-1} + u'Bf_t + \tilde{u}'Bf_{t-1}\right)\right] \\ &= \exp\left\{\psi_\varepsilon(u) + \psi_\varepsilon(\tilde{u}) + \psi_f(B'u, B'\tilde{u})\right\}, \ \forall u, \tilde{u} \in \mathcal{U}, \end{aligned}$$

which hold for the true parameter values. Let us consider the change of variables:

$$u'\tilde{B} = v' \quad \Leftrightarrow \quad u(v, B) = \tilde{B}v,$$

where $\tilde{B} = (B \vdots B_\perp)$ is defined as in the previous subsection. Thus, we get:

$$E\left[\exp\left\{u(v,B)'(y_t - Cy_{t-1}) + u(\tilde{v},B)'(y_{t-1} - Cy_{t-2})\right\}\right] = \exp\left\{\psi_\varepsilon(Bv) + \psi_\varepsilon(\tilde{B}\tilde{v}) + \psi_f(v_1, \tilde{v}_1)\right\}, \tag{5.13}$$

$\forall v, \tilde{v} \in \mathcal{V}$. The identification strategy is slightly different in this example compared to the general setting in Section 5.1, because we have a second infinite-dimensional parameter ψ_ε in addition to the factor dynamics. The r.h.s. of equation (5.13) is the product of a function of v, a function of $\tilde{v}$, and a function of $(v_1, \tilde{v}_1)$. Thus, the cross-derivative w.r.t. v_2 and $\tilde{v}$ of the log of this product function, as well as the cross-derivative w.r.t. v and $\tilde{v}_2$, vanish. By computing the log of the l.h.s. of (5.13), taking the cross-derivative w.r.t. v_2 and $\tilde{v}$, and equalizing the result to zero we get $B_\perp' H(\tilde{B}v, \tilde{B}\tilde{v}, C)\tilde{B}' = 0$, $\forall v, \tilde{v} \in \mathcal{V}$, where:

$$H(u, \tilde{u}, C) = \frac{\partial^2 \log E\left[\exp\left\{u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2})\right\}\right]}{\partial u \partial \tilde{u}'}. \tag{5.14}$$

Since the matrix $\tilde{B}$ is non-singular, we get equivalently the continuum of nonlinear restrictions:

$$B_{\perp}' H(u, \tilde{u}, C) = 0, \quad \forall u, \tilde{u} \in \mathcal{U}. \tag{5.15}$$

Similarly, by taking the cross-derivative w.r.t. v and $\tilde{v}_2$ of the log l.h.s. of (5.13), we get:

$$H(u, \tilde{u}, C) B_{\perp} = 0, \quad \forall u, \tilde{u} \in \mathcal{U}. \tag{5.16}$$

Function H in (5.14) is identifiable. Matrices B and C are semi-parametrically second-order nonlinearly identifiable if the unique solution of (5.15) and (5.16) is $B = B_0$, $C = C_0$. Function H is not an expectation, and (5.15) and (5.16) are not moment restrictions. Moreover, equations (5.15) and (5.16) generally contain only a subset of the information in the second-order nonlinear moment restrictions (5.13). [6]

Let us now show that parameters B and C are semi-parametrically second-order nonlinearly identifiable. In Appendix 2.1 we show that:

$$\begin{aligned} H(u, \tilde{u}, C) &= \widetilde{Cov}\left(y_t - C y_{t-1}, y_{t-1} - C y_{t-2}\right) \\ &= [B_0 \quad -\Delta]\, \widetilde{Cov}\left[\begin{pmatrix} f_t \\ y_{t-1} \end{pmatrix}, \begin{pmatrix} f_{t-1} \\ y_{t-2} \end{pmatrix}\right] \begin{pmatrix} B_0' \\ -\Delta' \end{pmatrix} - \Delta \tilde{V}(\varepsilon_t), \end{aligned} \tag{5.17}$$

where $\Delta = C - C_0$, and $\tilde{V}$ and $\widetilde{Cov}$ denote variance and covariance w.r.t. the modified probability measure with expectation operator:

$$\tilde{E}(Z) = \frac{E[\exp\{u'(y_t - C y_{t-1}) + \tilde{u}'(y_{t-1} - C y_{t-2})\} Z]}{E[\exp\{u'(y_t - C y_{t-1}) + \tilde{u}'(y_{t-1} - C y_{t-2})\}]}. \tag{5.18}$$

This modified probability measure depends on parameter matrix C, arguments u, $\tilde{u}$ and date t.

Assumption 2. *The factor process (f_t) and the innovation process (ε_t) are **not** both Gaussian, and the factor process (f_t) is **not** i.i.d. Moreover: i) The $(n+K) \times n$ matrix:*

$$M(u, \tilde{u}, C) = \widetilde{Cov}\left[\begin{pmatrix} f_t \\ y_{t-1} \end{pmatrix}, \begin{pmatrix} f_{t-1} \\ y_{t-2} \end{pmatrix}\right] \begin{pmatrix} B_0' \\ -\Delta' \end{pmatrix} - \begin{pmatrix} 0 \\ \tilde{V}(\varepsilon_t) \end{pmatrix},$$

[6] To see better the link with Section 5.1, note that by considering the moment restriction (5.1) for the semi-parametric linear model (5.12), concentrating out the functional parameter ψ_{ε}, and taking the derivative w.r.t. $\tilde{v}$ we get equation (5.15). A similar remark applies to the link between equations (5.2) and (5.16).

is such that lin $\left(\bigcup_{(u,\tilde{u})}\mathcal{R}\left(M(u,\tilde{u},C)\right)\right) = \mathbb{R}^{n+K}$*, for any* C*, where* $\mathcal{R}(\cdot)$ *denotes the range of a matrix, and lin*(S) *denotes the linear space spanned by the elements of set* S*; ii) We have lin* $\left(\bigcup_{(u,\tilde{u})}\mathcal{R}\left(\tilde{V}(\varepsilon_t)B_0^{\perp}\right)\right) = \mathbb{R}^n$*, for any* C*, where* $B_0^{\perp}$ *is a* $n \times (n-K)$ *matrix whose columns span the space* $\mathcal{R}(B_0)^{\perp}$*, that is the orthogonal complement of* $\mathcal{R}(B_0)$.

In Assumption 2 we explicitly exclude the case of Gaussian factor and innovation processes. Indeed, in such a case the log Laplace transform of $(y_t - Cy_{t-1}, y_{t-1} - Cy_{t-2})$ is quadratic, its second-order partial derivative matrix is constant, and thus equations (5.15)-(5.16) are independent of arguments u and $\tilde{u}$. We get a system of equations which correspond to the additional moment restrictions for order 2 given in (4.12). We have verified in Section 4.1 that these restrictions (even when considered together with the first-order moment restrictions) are not sufficient for identifying the parameters. In fact, in the Gaussian case, Assumptions 2 i) and ii) do not hold, since matrices $M(u,\tilde{u},C)$ and $\tilde{V}(\varepsilon_t)$ are independent of u, $\tilde{u}$. When the processes are not Gaussian, equations (5.15) and (5.16) in general imply a continuum set of restrictions depending on the values of arguments u, $\tilde{u}$. Assumptions 2 i) and ii) require that some variance and covariance matrices of the variables under the modified expectation operator $\tilde{E}$ have sufficient variation w.r.t. the arguments $(u,\tilde{u})$, so that the associated ranges span the entire Euclidean space, for any C. A necessary condition for Assumption 2 i) is that process (f_t) is serially dependent, otherwise the upper (K,n) block of matrix $M(u,\tilde{u},C_0)$ vanishes (see Lemma A.1 in Appendix 2). This fact explains why we explicitly exclude the case of a white noise factor process similarly as in Assumption 1. When the model is restricted to have $C = 0$, i.e. no contagion, we can replace Assumption 2 i) with the weaker Assumption 1, and we can dispense of Assumption 2 ii). This parallels the analysis in Section 5.2 on the Poisson model with common stochastic intensity.

Proposition 6. *Under Assumption 2, the dimension* K *of the factor space, the contagion matrix* C*, and the column space of matrix* B *in model (5.12) are second-order nonlinearly identifiable.*

Proof: Let us first prove that equation (5.15) identifies the factor dimension K_0, the range of matrix B_0 and some features of matrix parameter C_0. Using (5.17) and the definition of matrix $M(u,\tilde{u},C)$ in Assumption 2, we can write (5.15) as $B_{\perp}'[B_0 \quad -\Delta]M(u,\tilde{u},C) = 0$, $\forall u,\tilde{u} \in \mathcal{U}$. From Assumption 2 i) we get:

$$B_{\perp}'[B_0 \quad -\Delta] = 0. \tag{5.19}$$

For $K = K_0$, equation (5.19) is satisfied iff i) the columns of matrix Δ are linear combinations of the columns of matrix B_0, i.e., parameter matrix C is in set $\mathcal{C} = \{C \in \mathbb{R}^{n\times n} \ : \ C = C_0 + B_0\alpha', \ \alpha \in \mathbb{R}^{n\times K}\}$, and ii) $\mathcal{R}(B_\perp) = \mathcal{R}(B_0)^\perp$, i.e. $\mathcal{R}(B) = \mathcal{R}(B_0)$. For $K < K_0$, equation (5.19) has no solution, because the column space of matrix $[B_0 \quad -\Delta]$ has dimension larger or equal to K_0, while the number of columns of $B_\perp$ is strictly larger than $n - K_0$. Therefore, K_0 is identifiable as the smallest factor dimension K for which (5.15) has a solution. Moreover, the column space of matrix B_0 and set $\mathcal{C}$ are identifiable as well.

Let us now show that matrix C_0 is identifiable from equation (5.16). Suppose that $C \in \mathcal{C}$, that is $\Delta = C - C_0 = B_0\alpha'$ for some (n, K) matrix α. Then, from equation (5.17) we get:

$$H(u,\tilde{u},C) = B_0[Id \quad -\alpha']\,\widetilde{Cov}\left[\begin{pmatrix} f_t \\ y_{t-1}\end{pmatrix}, \begin{pmatrix} f_{t-1} \\ y_{t-2}\end{pmatrix}\right]\begin{pmatrix} Id \\ -\alpha\end{pmatrix} B_0' - B_0\alpha'\tilde{V}(\varepsilon_t). \tag{5.20}$$

Let $v \in \mathcal{R}(B_0)^\perp$. Then, from equation (5.16) with $\mathcal{R}(B_\perp) = \mathcal{R}(B_0)^\perp$, we get $B_0\alpha'\tilde{V}(\varepsilon_t)v = 0, \forall u, \tilde{u}$. Since matrix B_0 has full column-rank, we deduce $\alpha'\tilde{V}(\varepsilon_t)v = 0, \quad \forall u, \tilde{u}, \quad \forall v \in \mathcal{R}(B_0)^\perp$. From Assumption 2 ii) it follows $\alpha = 0$, that is, $C = C_0$. Q.E.D.

ii) Second-order nonlinear identification of the distributions of the factor process and innovations

Let us now consider the identification of the distributions of the innovation process (ε_t) and unobservable common factor (f_t). Proposition 3 does not apply here, because the r.h.s. of (5.13) involves two functional parameters, namely ψ_ε and ψ_f. We focus on the case in which the common factor is an exponentially affine (i.e. CaR) Markov process (see Section 2.2). Then, the Laplace transform of the transition is such that:

$$E[\exp(u'f_t)|f_{t-1}] = \exp[a(u)'f_{t-1} + b(u)], \tag{5.21}$$

where functions a and b are let unspecified. Under the stationarity condition [see Darolles, Gourieroux, Jasiak (2006)] and by the Law of Iterated Expectation, the Laplace transform of the stationary distribution $E[\exp(u'f_t)] = \exp(\varphi_f(u))$ is such that $\varphi_f(u) = b(u) + \varphi_f[a(u)]$. Thus, we have $b(u) = \varphi_f(u) - \varphi_f[a(u)]$, and the distributions of processes (ε_t) and (f_t) can be parameterized by the functional parameters a, φ_f and ψ_ε. We denote by a_0, φ_f^0 and ψ_ε^0 the true parameters.

The (n, K) factor loading matrix B_0 has row rank K. Let us assume that the first K rows of B_0 are linearly independent (possibly after renumbering of the individuals). Then, as identification constraints for the unobservable factor, it is convenient to assume that the upper (K, K) block of matrix B_0 is the identity matrix Id_K and that the factor process has zero unconditional mean. We have:

$$\tilde{y}_t = f_t + \tilde{\varepsilon}_t, \tag{5.22}$$

where $\tilde{y}_t$ and $\tilde{\varepsilon}_t$ denote $(K, 1)$ vector processes which contain the first K elements of $y_t - C_0 y_{t-1}$, and of ε_t, respectively. Let us define the function:

$$\tau_0(u, v; h) = \frac{E_0[\exp(u'\tilde{y}_t + v'\tilde{y}_{t-h})]}{E_0[\exp(u'\tilde{y}_t)]E_0[\exp(v'\tilde{y}_{t-h})]},$$

that is, the ratio between the joint Laplace transform of $\tilde{y}_t$ and $\tilde{y}_{t-h}$ for horizon $h \geq 1$, and the product of the two marginal Laplace transforms. Since matrix C_0 is identifiable from Proposition 6, process $\tilde{y}_t$ involves observable data and identifiable parameters only. Thus, function τ_0 is identifiable. Moreover, in Appendix 2.2 we show that:

$$\tau_0(u, v; h) = \exp\left(\varphi_f^0[a_0^{\circ h}(u) + v] - \varphi_f^0[a_0^{\circ h}(u)] - \varphi_f^0(v)\right), \quad \forall u, v, \ \forall h \geq 1, \tag{5.23}$$

where $a_0^{\circ h}$ denotes function a_0 compounded h times with itself. Function τ_0 depends on the DGP by means of the functional parameters a_0 and φ_f^0 of the factor process only, and we have:

$$\tau_0(u, v; 2) = \tau_0[a_0(u), v; 1], \quad \forall u, v. \tag{5.24}$$

Proposition 7. *Consider the semi-parametric model with common factor and contagion defined by equations (5.12) and (5.21). Suppose that Assumption 2 holds and function $a_0 : \mathcal{D}_0 \to \mathcal{R}_0$ is one-to-one, where domains $\mathcal{D}_0$ and $\mathcal{R}_0$ are neighbourhoods of 0 in $\mathbb{C}^K$, and $\mathcal{R}_0 \subset \mathcal{D}_0$. Then: i) function $a_0(u)$ is second-order identifiable for $u \in \mathcal{D}_0$, ii) function $\varphi_f^0(u)$ is second-order identifiable for $u \in \mathcal{R}_0$, and iii) function $\psi_\varepsilon(u)$ is second-order identifiable for $B_0'u \in \mathcal{R}_0$.*

Proof: i) Let us first show that function a_0 is identifiable on domain $\mathcal{D}_0$. By using the one-to-one property of function a_0 and the convexity of function φ_f^0, in Appendix 2.3 we show that function $\tau_0(\cdot, \cdot; 1)$ is such that:

$$\tau_0(u_1, v; 1) = \tau_0(u_2, v; 1), \ \forall v \quad \Rightarrow \quad u_1 = u_2. \tag{5.25}$$

Then, for any given $u \in \mathcal{D}_0$, the value $a_0(u)$ is identifiable from (5.24) because function $\tau(\cdot, \cdot; 2)$ is identifiable.

ii) Let us now prove that function φ_f^0 is identifiable on $\mathcal{R}_0$. Since function $\tau_0(\cdot, \cdot; 1)$ is identifiable, the function:

$$\frac{\partial \log \tau_0}{\partial v}(u, 0; 1) = \frac{\partial \varphi_f^0}{\partial u}[a_0(u)] - \frac{\partial \varphi_f^0}{\partial u}(0),$$

is identifiable as well. Since the unconditional expectation of the latent factor is normalized to zero, we have $\partial\varphi_f^0(0)/\partial u = E_0(f_t) = 0$. It follows that:

$$\frac{\partial \log \tau_0}{\partial v}(u, 0; 1) = \frac{\partial \varphi_f^0}{\partial u}[a_0(u)].$$

Since function a_0 is one-to-one and identifiable by part i), the above equation implies that $\partial\varphi_f^0/\partial u(u) = \partial \log \tau_0/\partial v[a_0^{-1}(u), 0; 1]$ is identifiable for $u \in \mathcal{R}_0$. Then:

$$\varphi_f^0(u) = \varphi_f^0(u) - \varphi_f^0(0) = \int_0^1 \frac{\partial \varphi_f^0}{\partial u}(tu)' u dt = \int_0^1 \frac{\partial \log \tau_0}{\partial v}[a_0^{-1}(tu), 0; 1]' u dt,$$

is identifiable.

iii) Finally, let us show the identification of function ψ_ε^0. This follows from:

$$E_0[\exp(u'(y_t - C_0 y_{t-1}))] = \exp(\varphi_f^0(B_0' u) + \psi_\varepsilon^0(u)),$$

where the l.h.s. is identifiable, and $\varphi_f^0(B_0' u)$ is identifiable if $B_0' u \in \mathcal{R}_0$ from part ii). Q.E.D.

6 Nonlinear cross-differencing in panel data

In this section we study semi-parametric identification in (nonlinear) panel data models with stochastic time effects. In this framework, moment restrictions based on cross-differencing can be used to achieve parameter identification. Identification is studied for a number of time periods T tending to infinity, and a finite cross-sectional dimension n.

6.1 Linear and nonlinear panel data models with unobservable effects

We introduce the modeling framework by means of some examples.

Example 6: The Gaussian Panel Model with Stochastic Time Effect.

The one-dimensional observed endogenous variable is defined by the regression model:

$$y_{i,t} = f_t + x_{i,t}' \alpha + \varepsilon_{i,t}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T,$$

where $x_{i,t}$ denotes the observed explanatory variables and the error terms $\varepsilon_{i,t}$ are $IIN(0, \sigma^2)$ conditional on the x variables. The time effect f_t is assumed stochastic, independent of the error terms, and is unobservable for the econometrician.

Example 7: Count Panel Data with Stochastic Time Effect.

This model is similar to the model in Example 6, except that the conditional distribution is Poisson to account for the interpretation of the endogenous variable as a count variable [see e.g. Blundell, Griffith, Windmeijer (2002)]. We have:

$$y_{i,t} \sim \mathcal{P}(f_t + x'_{i,t}\alpha + y_{i,t-1}c), \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

with possibly effects of the lagged count corresponding to the same individual. This model is used for the dynamic analysis of corporate defaults by segment i of corporates [see e.g. Agosto et al. (2016)]. In this application f_t represents the systematic (systemic) factor, and the model is the basis for the stress test exercises conducted for financial stability. The model is appropriate to disentangle the direct effect of an exogenous systematic shock (f_t) from the indirect contagion effect passing through the lagged counts. Such models are also used for a joint analysis of trading activity and the measure of systematic illiquidity risk, when $y_{i,t}$ is the number of trades on stock i on day t [see e.g. Quoreshi (2008), Fokianos (2012), Bourguignon, Vasconcellos (2016)].

The general panel data specification, that we consider in this section, is defined in the next assumptions.

Assumption B.1. ***(Affine Panel Data Regression)***

i) Variables $y_{i,t}$ and f_t are one-dimensional.

ii) The individual histories $(y_{i,t}, x_{i,t},\ t$ varying), for $i = 1, \dots, n$, are independent conditional on (f_t).

iii) The individual conditional Laplace transform is given by:

$$E[\exp(uy_{i,t})|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}] = \exp\{a(u, x_{i,t}, \theta)[f_t + cy_{i,t-1} + d(x_{i,t}, \theta)] + b(u, x_{i,t}, \theta)\},$$

where argument $u \in \mathcal{U}$ is scalar.

Assumption B.2. *Process (f_t) is strongly exogenous, strictly stationary and Markov of order* 1.

Assumption B.1 is a stronger version of Assumption A.1, which is suitable for the panel data setting. Examples 6 and 7 are compatible with Assumption B.1. The dynamics of the unobservable effects (f_t) is let unspecified except for the nonparametric restrictions in Assumption B.2, which yields a semi-parametric framework.

By Assumption B.1 iii), the panel model has a particular structure defined in terms of Laplace transform. This structure is appropriate for models encountered in Finance and Insurance, when measuring systematic and idiosyncratic risks. It is not appropriate for dynamic discrete choice panel models or dynamic censored panel models used for instance for the analysis of labor force participation. First, qualitative or censoring features are not easily written in terms of Laplace transform. Second, these latter models focus more on individual unobserved heterogeneity than on stochastic time dependent factors [see e.g. Hu, Shiu (2013) for a study of nonparametric identification in these other models].

6.2 First-order nonlinear moment restrictions by cross-differencing

In the framework of panel data with stochastic time effect defined by Assumptions B.1 and B.2, the analogue of the first-order moment restrictions (3.1) can be written for the different individuals as:

$$\begin{aligned} &E\left[\exp\{uy_{i,t} - a(u, x_{i,t}, \theta)[cy_{i,t-1} + d(x_{i,t}, \theta)] - b(u, x_{i,t}, \theta)\}|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}\right] \\ &\qquad = \exp[a(u, x_{i,t}, \theta) f_t], \quad \forall u \in \mathcal{U},\ \forall i = 1, \dots, n. \end{aligned} \tag{6.1}$$

Conditions (6.1) can be combined across individuals to eliminate the common effect f_t in the r.h.s. Then, by applying the conditional expectation given $\underline{x_t}$, $\underline{y_{t-1}}$, we get moment restrictions which involve parameters B, C, θ only. This provides an extension of the quasi-differencing method to a nonlinear cross-sectional framework. Specifically, let us assume: [7]

Assumption B.3. *The function $u \to a(u, x_{i,t}, \theta)$ is continuous and strictly monotonous w.r.t. the argument u, for any given $x_{i,t}$ and θ.*

Under Assumption B.3 we can define $u(v, x_{i,t}, \theta)$ as the solution of the equation $a(u, x_{i,t}, \theta) = v$, and replace argument u by $u(v, x_{i,t}, \theta)$ in equation (6.1). We get:

$$\begin{aligned} &E\left[\exp\{u(v, x_{i,t}, \theta)y_{i,t} - v[cy_{i,t-1} + d(x_{i,t}, \theta)] - b[u(v, x_{i,t}, \theta), x_{i,t}, \theta]\}|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}\right] = \exp(vf_t), \\ &\forall v \in \mathcal{V},\ \forall i = 1, \dots, n. \end{aligned} \tag{6.2}$$

Then, let us consider a pair of individuals i and j, with $i \neq j$. We deduce:

$$\begin{aligned} &E\left[\exp\{u(v, x_{i,t}, \theta)y_{i,t} - v[cy_{i,t-1} + d(x_{i,t}, \theta)] - b[u(v, x_{i,t}, \theta), x_{i,t}, \theta]\}|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}\right] \\ = \ &E\left[\exp\left\{u(v, x_{j,t}, \theta)y_{j,t} - v[cy_{j,t-1} + d(x_{j,t}, \theta)] - b[u(v, x_{j,t}, \theta), x_{j,t}, \theta]\right\}|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}\right], \end{aligned}$$

$\forall v \in \mathcal{V},\ \forall i, j,\ i \neq j$. These restrictions still depend on the unobservable component through the conditioning sets. But, by taking the conditional expectations of both sides with respect to $\underline{y_{t-1}}, \underline{x_t}$, we get the following set of moment restrictions:

[7]For panel data, Assumption B.3 implies Assumption A.4, since the arguments u and v are both scalar.

First-order nonlinear moment restrictions with cross-differencing:

$$
\begin{aligned}
& E\left[\exp\{u(v,x_{i,t},\theta)y_{i,t}-v[cy_{i,t-1}+d(x_{i,t},\theta)]-b[u(v,x_{i,t},\theta),x_{i,t},\theta]\}|\underline{y_{t-1}},\underline{x_t}\right] \\
= \; & E\left[\exp\{u(v,x_{j,t},\theta)y_{j,t}-v[cy_{j,t-1}+d(x_{j,t},\theta)]-b[u(v,x_{j,t},\theta),x_{j,t},\theta]\}|\underline{y_{t-1}},\underline{x_t}\right], \\
& \forall v\in\mathcal{V},\ \forall i,j,\ i\neq j. \qquad (6.3)
\end{aligned}
$$

We get a continuum of moment restrictions that is indexed by the pairs of individuals and scalar argument v. The restrictions involve parameters c and θ of the affine regression model, but no parameters of the distribution of the stochastic time effect. This leads to the following definition.

Definition 4. *The regression parameters c and θ are first-order nonlinearly identifiable by cross-differencing if the true values $c=c_0$, $\theta=\theta_0$ yield the unique solution to system (6.3).*

We have the following proposition:

Proposition 8. *If parameters c and θ are first-order nonlinearly identifiable by cross-differencing, then the conditional distribution of f_t given $\underline{x_t}$ is nonparametrically identifiable by equation (6.2).*

Thus, in panel models defined by Assumptions B.1-B.3, we obtain semi-parametric identification of the nonlinear regression parameters and stationary distribution of the unobservable effects (given the covariates) without using higher-order moment restrictions, but cross-differencing instead. Higher-order nonlinear moment restrictions can be used to identify the transition distribution of the unobservable effects as in Section 5. The cross-differencing approach is easily extended to multidimensional panel data $y_{i,t}=(y_{1,it},\dots,y_{K,it})'$ with stochastic time effects $f_t=(f_{1,t},\dots,f_{K,t})'$ specific to each type of variable.

The approach above is a nonlinear cross-sectional extension of the standard quasi-differencing approach usually proposed for panel data with fixed individual effect and based on the first-order moments only [see Mullahy (1997), Wooldridge (1997), (1999)]. Indeed, the derivative of a Laplace transform $E[\exp(uy_{i,t})]$ w.r.t. argument u at $u=0$ is equal to the expectation $E(y_{i,t})$, if this expectation exists. By construction, we have $a(0,x_{i,t},\theta)=0$, $\forall x_{i,t},\forall\theta$. Thus, an analysis of the restrictions in a neighbourhood of $u=0$ is equivalent to an analysis in a neighbourhood of $v=0$. The analogue restrictions based on the first-order moments are simply deduced from (6.3) by equating the derivatives of both sides of equation (6.3) with respect to v at $v=0$. Our nonlinear cross-differencing approach takes into account not only the first-order moment of appropriate transformations of the observations, but also higher-order moments by means of the exponential transform with varying argument v. Thus, the set of moment restrictions (6.3) is more informative than the ones usually considered in the panel literature and based on the first-order moment only. In the Supplementary Materials we derive the GMM asymptotic efficiency bound for estimating parameters c, θ from (6.3), and investigate the efficiency gain resulting from the nonlinear moment restrictions in Example 7. [8]

Example 7: Count Panel Data with Stochastic Time Effect (continues)

Let us illustrate the nonlinear cross-differencing approach above with the model for count panel data. We have $y_{i,t}|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t} \sim \mathcal{P}(f_t + x_{i,t}'\alpha + y_{i,t-1}c)$. The conditional Laplace transform is:

$$E[\exp(uy_{i,t})|\underline{y_{t-1}}, \underline{f_t}, \underline{x_t}] = \exp\{-(f_t + x_{i,t}'\alpha + y_{i,t-1}c)(1 - \exp u)\}, \tag{6.4}$$

with function a in Assumption B.1 given by $a(u) = 1 - \exp(u)$, independent of parameters and regressors. The moment restrictions (6.3) become:

$$\begin{aligned} & E\left[\exp\{uy_{it} + (1 - \exp u)(x_{i,t}'\alpha + y_{i,t-1}c)\}|\underline{y_{t-1}}, \underline{x_t}\right] \\ = \; & E\left[\exp\{uy_{j,t} + (1 - \exp u)(x_{j,t}'\alpha + y_{j,t-1}c)\}|\underline{y_{t-1}}, \underline{x_t}\right], \; \forall i \neq j, \; \forall u \in \mathcal{U}. \end{aligned} \tag{6.5}$$

Since function a is independent of the regressors, we have not applied the change of argument, which is simply $1 - \exp u = v$, or equivalently $u = \log(1 - v)$. By considering the first-order expansion of the moment restrictions for arguments close to $u = 0$, the equality of the first-order derivatives of the l.h.s. and r.h.s. of (6.5) w.r.t. u in $u = 0$ yields:

$$E[y_{i,t} - x_{i,t}'\alpha - y_{i,t-1}c|\underline{y_{t-1}}, \underline{x_t}] = E[y_{j,t} - x_{j,t}'\alpha - y_{j,t-1}c|\underline{y_{t-1}}, \underline{x_t}], \quad \forall i \neq j, \tag{6.6}$$

which are the moment restrictions considered in Windmeijer (2000), Blundell et al. (2002).

The conditional moment restrictions (6.6) identify the true parameter values α_0, c_0, because these restrictions are equivalent to $x_{i,t}'(\alpha - \alpha_0) + y_{i,t-1}(c - c_0) = x_{j,t}'(\alpha - \alpha_0) + y_{j,t-1}(c - c_0)$, $\forall i \neq j$, which implies $\alpha = \alpha_0$ and $c = c_0$. A fortiori, the true parameter values are first-order nonlinearly identifiable from the more informative set of conditional moment restrictions (6.5). In particular, this holds true even if the arguments' set $\mathcal{U}$ is restricted to be a subset of the real line.

[8] The GMM efficiency bound derived in the Supplementary Materials is for large T and finite n. The semi-parametric efficiency bound in the asymptotics with T and n tending jointly to infinity at an appropriate relative rate is derived in Gagliardini, Gourieroux (2014a).

7 Concluding remarks

The aim of this paper is to highlight how continuum sets of conditional moment restrictions based on appropriate Laplace transforms can be used to identify nonlinear time series or panel data models with unobservable dynamic effects. The identification approach requires regression models which are exponential affine in lagged endogenous variables and unobservable dynamic effects. It can be applied both with a parametric specification for the dynamics of the unobserved component as well as when the distribution of the unobservable process is let unspecified. We provide several examples inspired by macroeconomic and financial time series models, as well as microeconomic panel models, to show the variety of specifications for which this identification approach can be used.

Our identification strategy is constructive and leads to moment-based estimation methods which can be computationally less demanding compared to standard parametric approaches such as the Maximum Likelihood (ML) approach [9], or other recently proposed semi/non-parametric estimation methods [10]. Indeed, the continuum of nonlinear conditional moment restrictions can be the basis to define convenient moment-based estimation methods avoiding the numerical computation of multidimensional integrals [see e.g. Carrasco, Florens (2000), Singleton (2001), Jiang, Knight (2002), Chacko, Viceira (2003), Carrasco et al. (2007), Carrasco, Florens (2014); see also Darolles, Gagliardini, Gourieroux (2014) for an application to the estimation of the default risk model with frailty and contagion in Example 3]. When the model involves observable explanatory variables, the conditional moment restrictions have to be transformed into unconditional moment restrictions via the choice of instruments. Moreover, when a continuum of moment restrictions are used, the implementation of GMM estimators with optimal weighting involves the adoption of a regularization approach to deal with the non-invertibility of the variance-covariance operator.

[9] The likelihood function involves multidimensional integrals, whose dimension increases with the number T of observations, since the unobserved factor path has to be integrated out. ML estimation can be implemented by simulation-based methods [see e.g. Cappé, Moulines, Rydén (2005) and Doucet, Johansen (2010) for methods relying on particle filtering and similar ideas, and Gourieroux and Monfort (1997) for indirect inference and similar methods]. In a semi-affine, parametric nonlinear state space framework, Bates (2006) develops a nonlinear filtering and estimation method that requires the computation of numerical integrals with dimension equal to the number of endogenous observable variables.

[10] In specific semiparametric frameworks with unobservable components which differ from our, Bonhomme (2012) and Schennach (2014) introduce consistent estimation methods relying on functional differencing, and entropic latent variable integration via simulation, respectively. The spectral operator decomposition deployed by Hu, Shum (2012) for nonparametric identification could be the basis for introducing consistent estimation methods, which however seem not easy to implement.

The detailed discussion of the implementation of such GMM estimators is out of the scope of this paper. Further, since the $(l + 1)$-th order nonlinear moment restrictions include the l-th order nonlinear moment restrictions as a special case, for $l = 1, 2$, the use of higher-order restrictions - even when not necessary for parameter identification - leads to GMM estimators with equal or better asymptotic efficiency.

This paper focuses on nonlinear models with unobservable time effects. We stress that semi-parametric identification strategies developed for panel data with individual unobservable effects generally cannot be directly applied to our setting by interchanging the role of individuals and time dates, because we allow for dynamics in the unobservable time effects. However, it would be interesting to combine the identification strategies developed in this paper and in the panel data literature to study models with both individual and time unobservable effects. Moreover, in a panel data framework with a conditional moment function that is linear in the unobservable effects, Chamberlain (1992) shows that the estimator of the regression parameter obtained by quasi-differencing reaches the semi-parametric efficiency bound. It would be interesting to address this question in our exponentially affine framework with a continuum of conditional moment restrictions and investigate, e.g., if the GMM efficiency bound for cross-differencing derived in the Supplementary Materials corresponds to the semi-parametric efficiency bound for the regression parameter. We leave these challenging questions for future research.

References

Agosto, A., Cavaliere, G., Kristensen, D., and A., Rabhek (2016): "Modeling Corporate Defaults: Poisson Autoregressions with Exogenous Covariates (PARX)", Journal of Empirical Finance, 38, 640-663.

Al-Osh, M., and A., Alzaid (1987): "First-Order Integer Valued Autoregressive INAR(1) Process", Journal of Time Series Analysis, 8, 261-275.

Anderson, T. (2003): *An Introduction to Multivariate Statistical Analysis*, Wiley, 3rd edition.

Anderson, T., and H., Rubin (1956): "Statistical Inference in Factor Analysis", in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman ed., Volume 5, 111-150.

Bates, D. (2006): "Maximum Likelihood Estimation of Latent Affine Processes", Review of Financial Studies, 19, 909-965.

Bernanke, B., Boivin, J., and P., Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach", Quarterly Journal of Economics, 120, 387-422.

Blundell, R., Griffith, R., and F., Windmeijer (2002): "Individual Effects and Dynamics in Count Data Model", Journal of Econometrics, 108, 113-131.

Bockenholt, U. (1999): "Mixed INAR(1) Poisson Regression Models Analyzing Heterogeneity and Serial Dependencies in Longitudinal Count Data", Journal of Econometrics, 89, 317-338.

Bonhomme, S. (2012): "Functional Differencing", Econometrica, 80, 1337-1385.

Bourguignon, M., and K., Vasconcellos (2016): "A New Skew Integer Valued Time Series Process", Statistical Methodology, 31, 8-19.

Brannas, K. (1994): "Estimation and Testing in Integer Valued AR(1) Models", DP 335, University of Umea.

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

York.

Carrasco, M., Chernov, M., Florens, J.P., and E., Ghysels (2007): “Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions”, Journal of Econometrics, 140, 529-573.

Carrasco, M., and J.-P., Florens (2000): “Generalization of GMM to a Continuum of Moment Conditions”, Econometric Theory, 16, 797-834.

Carrasco, M., and J.-P., Florens (2014): “On the Asymptotic Efficiency of GMM”, Econometric Theory, 30, 372-406.

Carroll, R., Chen, X., and Y., Hu (2010): “Identification and Estimation of Nonlinear Models Using Two Samples with Nonclassical Measurement Errors”, Journal of Nonparametric Statistics, 22, 379-399.

Chamberlain, G. (1992): “Efficiency Bounds for Semi-Parametric Regression”, Econometrica, 60, 567-596.

Chacko, G., and L., Viceira (2003): “Spectral GMM Estimation of Continuous-Time Processes”, Journal of Econometrics, 116, 259-292.

Chiriac, R., and V., Voev (2011): “Modelling and Forecasting Multivariate Realized Volatility”, Journal of Applied Econometrics, 26, 922-947.

Cox, J., Ingersoll, J., and S., Ross (1985): “A Theory of the Term Structure of Interest Rates”, Econometrica, 53, 385-407.

Darolles, S., Dubecq, S., and C., Gourieroux (2014): “Contagion Analysis in the Banking Sector”, working paper.

Darolles, S., Gagliardini, P., and C., Gourieroux (2014): “Survival of Hedge Funds : Frailty vs Contagion”, DP CREST.

Darolles, S., and C., Gourieroux (2015): *Contagion Phenomena with Applications in Finance*, Elsevier.

Darolles, S., Gourieroux, C., and J., Jasiak (2006): “Structural Laplace Transform and Compound Autoregressive Models”, Journal of Time Series Analysis, 27, 477-450.

Diebold, F., and M., Nerlove (1989): “The Dynamics of Exchange Rate Volatility : A Multivariate Latent-Factor ARCH Model”, Journal of Applied Econometrics, 4, 1-22.

Doucet, A., and A. Johansen (2010): “A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later”, in D. Crisan, and B. Rozovsky, ed.: Handbook of Nonlinear Filtering (Oxford University Press: Oxford).

Engle, R., and S., Kozicki (1993): “Testing for Common Features”, Journal of Business and Economic Statistics, 11(4), 369-380.

Engle, R., and M., Watson (1981): “A One-Factor Multivariate Time Series Models of Metropolitan Wage Rates”, Journal of the American Statistical Society, 76, 774-781.

Fernandez-Macho, F. (1997): “A Dynamic Factor Model for Economic Time Series”, Kybernetica, 33, 583-606.

Fokianos, K. (2012): “Count Time Series Models”, in *Handbook of Statistics: Time Series, Methods and Applications*, 30, 315-347.

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.

Gagliardini, P., and C., Gourieroux (2014a): “Efficiency in Large Dynamic Panel Models with Common Factors”, Econometric Theory, 30, 961-1020.

Gagliardini, P., and C., Gourieroux (2014b): “Wishart Stochastic Volatility Model”, DP CREST.

Gallant, R., Giacomini, R., and G., Ragusa (2017): “Bayesian Estimation of State Space Models Using Moment Conditions”, Journal of Econometrics, forthcoming.

Geweke, J. (1977): “The Dynamic Factor Analysis of Economic Time-Series Models”, in D., Aigner and A., Goldberger, eds, ”Latent Variables in Socio-Economic Models”, North-Holland, New-York.

Gourieroux, C., and J., Jasiak (2006): “Autoregressive Gamma Processes”, Journal of Forecasting, 25, 129-152.

Gourieroux, C., Jasiak, J., and R., Sufana (2009): “The Wishart Autoregressive Process of Multivariate Stochastic Volatility”, Journal of Econometrics, 150, 167-181.

Gourieroux, C., and A., Monfort (1997): *Simulation-based Econometric Methods*, Oxford University Press.

Gourieroux, C., and I., Peaucelle (1992): “Séries Codépendantes: Applications à l’hypothèse de Parité du Pouvoir d’Achat”, Revue d’Analyse Economique, 68, 283-304.

Hansen, L. (1982): “Large Sample Properties of Generalized Method of Moments Estimators”, Econometrica, 50, 1029-1054.

Hu, Y., and J.-L., Shiu (2013): “Identification and Estimation of Nonlinear Dynamic Panel Data Models with Unobserved Covariates”, Journal of Econometrics, 175, 116-131.

Hu, Y., and M., Shum (2012): “Nonparametric Identification of Dynamic Models with Unobserved State Variables”, Journal of Econometrics, 171, 32-44.

Jiang, G., and J., Knight (2002): “Estimation of Continuous-Time Processes via the Empirical Characteristic Function”, Journal of Business and Economic Statistics, 20, 198-212.

King, M., Sentana, E., and S., Wadhwani (1994): “Volatility and Links Between National Stock Markets”, Econometrica, 62, 901-933.

Lawley, D. N., and A. E., Maxwell (1971): *Factor Analysis as a Statistical Method*, Elsevier, New York.

Lintner, J. (1965): “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets”, Review of Economics and Statistics, 47, 13-37.

Mullahy, J. (1997): “Instrumental Variable Estimation of Poisson Regression Models : Application to Models of Cigarette Smoking Behaviour”, Review of Economics and Statistics, 79, 586-593.

Quoreshi, A. (2008): "A Vector Integer Valued Moving Average Model for High Frequency Financial Count Data", Economic Letters, 101, 258-261.

Roll, R. (1977): "A Critique of the Asset Pricing Theory's Tests Part I : On Past and Potential Testability of the Theory", Journal of Financial Economics, 4, 129-176.

Sargent, T. J., and C. A., Sims (1977): "Business Cycle Modeling Without Pretending to Have Too Much a-priori Economic Theory", in *New Methods in Business Cycle Research*, ed. by C. Sims, Federal Reserve Bank of Minneapolis, Minneapolis.

Sharpe, W. (1964): "Capital Asset Prices : A Theory of Market Equilibrium Under Conditions of Risk", Journal of Finance, 19, 425-442.

Sims, C. (1980): "Macroeconomics and Reality", Econometrica, 48, 1-48.

Schennach, S. (2014): "Entropic Latent Variable Integration via Simulation", Econometrica, 82, 345-385.

Singleton, K. (2001): "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function", Journal of Econometrics, 102, 111-141.

Vahid, F., and R., Engle (1993): "Common Trends and Common Cycles", Journal of Applied Econometrics, 8, 341-360.

Wilhelm, D. (2013): "Identification and Estimation of Nonparametric Data Regressions with Measurement Errors", working paper.

Windmeijer, F. (2000): "Moment Conditions for Fixed Effects Count Data Model with Endogenous Regressors", Economics Letters, 68, 21-24.

Wooldridge, J. (1997): "Multiplicative Panel Data Models Without the Strict Exogeneity Assumption", Econometric Theory, 13, 667-678.

Wooldridge, J. (1999): "Distribution-Free Estimation of Some Nonlinear Panel Data Models", Journal of Econometrics, 90, 77-97.

Appendix 1: Identification in the Gaussian linear model with common factor and contagion

In this Appendix we study the identification of the Gaussian linear model (4.1) with common factor and contagion.

A.1.1 Higher-order nonlinear moment restrictions

Since $y_t - Cy_{t-1} = Bf_t + \Sigma^{1/2}\varepsilon_t$, the second-order nonlinear moment restrictions are:

$$E_0[\exp\{u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2})\}] = E_0[\exp(u'Bf_t + \tilde{u}'B'f_{t-1} + u'\Sigma^{1/2}\varepsilon_t + \tilde{u}'\Sigma^{1/2}\varepsilon_{t-1})], \quad \forall u, \tilde{u}.$$

Since the variables are Gaussian, the restrictions concern the expression of $V_0(y_t - Cy_{t-1})$, which corresponds to the first-order moment restrictions, and the expression of $Cov_0(y_t - Cy_{t-1}, y_{t-1} - Cy_{t-2})$. The first-order nonlinear restrictions are:

$$\Gamma_0(0) + C\Gamma_0(0)C' - \Gamma_0(1)C' - C\Gamma_0(1)' = BB' + \Sigma, \tag{a.1}$$

corresponding to $V_0(y_t - Cy_{t-1}) = V_0(Bf_t + \Sigma^{1/2}\varepsilon_t)$, and the additional second-order nonlinear restrictions are:

$$\Gamma_0(1) + C\Gamma_0(1)C' - \Gamma_0(2)C' - C\Gamma_0(0) = B\Phi B', \tag{a.2}$$

corresponding to $Cov_0(y_t - Cy_{t-1}, y_{t-1} - Cy_{t-2}) = Cov_0(Bf_t + \Sigma^{1/2}\varepsilon_t, Bf_{t-1} + \Sigma^{1/2}\varepsilon_{t-1})$, where $\Gamma_0(h) = Cov_0(y_t, y_{t-h})$. Finally, the additional constraints from the third-order nonlinear moment restrictions correspond to the equality $Cov_0(y_t - Cy_{t-1}, y_{t-2} - Cy_{t-3}) = Cov_0(Bf_t + \Sigma^{1/2}\varepsilon_t, Bf_{t-2} + \Sigma^{1/2}\varepsilon_{t-2})$. They are:

$$\Gamma_0(2) + C\Gamma_0(2)C' - \Gamma_0(3)C' - C\Gamma_0(1) \quad = \quad B\Phi^2 B'. \tag{a.3}$$

A.1.2 Lack of identification from the second-order nonlinear moment restrictions (proof of Proposition 2)

In this section we show that the true parameter values B_0, C_0, Σ_0 and Φ_0 are not identifiable from the second-order nonlinear moment restrictions. For this purpose, we prove that the set of solutions (B, C, Σ, Φ) of the nonlinear matrix equations system in (a.1)-(a.2) is not a singleton. In order to show that the lack of identification holds even when the restriction of a diagonal matrix Σ is imposed, we establish the existence of solutions (B, C, Σ_0, Φ) different from the true parameter value $(B_0, C_0, \Sigma_0, \Phi_0)$.

Let us first write the autocovariance matrices $\Gamma_0(1)$ and $\Gamma_0(2)$ in terms of the true model parameters. By replacing the second equation of system (4.1) into the first equation, we have:

$$\begin{cases} y_t &= B\Phi f_{t-1} + Cy_{t-1} + \Sigma^{1/2}\varepsilon_t + B(Id - \Phi\Phi')^{1/2}\eta_t, \\ f_t &= \Phi f_{t-1} + (Id - \Phi\Phi')^{1/2}\eta_t. \end{cases}$$

Thus, the joint process $(y_t', f_t')'$ follows the restricted VAR(1) model:

$$\begin{pmatrix} y_t \\ f_t \end{pmatrix} = \begin{pmatrix} C & B\Phi \\ 0 & \Phi \end{pmatrix} \begin{pmatrix} y_{t-1} \\ f_{t-1} \end{pmatrix} + u_t,$$

where u_t is a Gaussian white noise process with variance-covariance matrix:

$$\Omega = V(u_t) = \left(\begin{array}{cc} \Sigma + B(Id - \Phi\Phi')B' & B(Id - \Phi\Phi') \\ (Id - \Phi\Phi')B' & Id - \Phi\Phi' \end{array} \right).$$

We have:

$$Cov\left[\left(\begin{array}{c} y_t \\ f_t \end{array} \right), \left(\begin{array}{c} y_{t-1} \\ f_{t-1} \end{array} \right) \right] = \left(\begin{array}{cc} C & B\Phi \\ 0 & \Phi \end{array} \right) V\left[\left(\begin{array}{c} y_t \\ f_t \end{array} \right) \right],$$

and:

$$Cov\left[\left(\begin{array}{c} y_t \\ f_t \end{array} \right), \left(\begin{array}{c} y_{t-2} \\ f_{t-2} \end{array} \right) \right] = \left(\begin{array}{cc} C & B\Phi \\ 0 & \Phi \end{array} \right)^2 V\left[\left(\begin{array}{c} y_t \\ f_t \end{array} \right) \right] = \left(\begin{array}{cc} C^2 & CB\Phi + B\Phi^2 \\ 0 & \Phi^2 \end{array} \right) V\left[\left(\begin{array}{c} y_t \\ f_t \end{array} \right) \right].$$

For the true parameter values, we deduce:

$$\Gamma_0(1) = C_0\Gamma_0(0) + B_0\Phi_0\Delta_0(0), \tag{a.4}$$

$$\Gamma_0(2) = C_0^2\Gamma_0(0) + (C_0B_0\Phi_0 + B_0\Phi_0^2)\Delta_0(0), \tag{a.5}$$

where $\Delta_0(0) = Cov_0(f_t, y_t)$ is the true contemporaneous cross-covariance between f_t and y_t.

Let us now rewrite the second-order nonlinear moment restrictions using the above formulas for the true autocovariances. Since the true parameter values solve equations (a.1)-(a.2), these equations can be written as:

$$\begin{gathered} \Sigma_0 + B_0B_0' + (C - C_0)\Gamma_0(0)C_0' + C_0\Gamma_0(0)(C - C_0)' + (C - C_0)\Gamma_0(0)(C - C_0)' \\ -\Gamma_0(1)(C - C_0)' - (C - C_0)\Gamma_0(1)' = \Sigma + BB', \\ B_0\Phi_0B_0' + (C - C_0)\Gamma_0(1)C_0' + C_0\Gamma_0(1)(C - C_0)' + (C - C_0)\Gamma_0(1)(C - C_0)' \\ -\Gamma_0(2)(C - C_0)' - (C - C_0)\Gamma_0(0) = B\Phi B'. \end{gathered}$$

By using formulas (a.4)-(a.5), we get

$$\begin{aligned} &\Sigma_0 + B_0B_0' - (C - C_0)\Delta_0(0)'\Phi_0'B_0' - B_0\Phi_0\Delta_0(0)(C - C_0)' + (C - C_0)\Gamma_0(0)(C - C_0)' = \Sigma + BB', \\ &B_0\Phi_0B_0' - (C - C_0)[\Gamma_0(0) - C_0\Gamma_0(0)C_0' - B_0\Phi_0\Delta_0(0)C_0'] - B_0\Phi_0^2\Delta_0(0)(C - C_0)' \\ &\qquad +(C - C_0)\Gamma_0(1)(C - C_0)' = B\Phi B'. \end{aligned} \tag{a.6}$$

Let us rewrite the term in the square brackets in (a.6). By computing the variance of both sides of equation $y_t = B_0f_t + C_0y_{t-1} + \Sigma_0^{1/2}\varepsilon_t$, we have:

$$\Gamma_0(0) = B_0B_0' + C_0\Gamma_0(0)C_0' + \Sigma_0 + B_0Cov_0(f_t, y_{t-1})C_0' + C_0Cov_0(y_{t-1}, f_t)B_0'.$$

By using that $Cov_0(f_t, y_{t-1}) = \Phi_0Cov_0(f_{t-1}, y_{t-1}) = \Phi_0\Delta_0(0)$ from (4.1) and stationarity, we get:

$$\Gamma_0(0) - C_0\Gamma_0(0)C_0' - B_0\Phi_0\Delta_0(0)C_0' = B_0B_0' + \Sigma_0 + C_0\Delta_0(0)'\Phi_0'B_0'.$$

Thus, the second-order nonlinear moment restrictions become:

$$\Sigma_0 + B_0 B_0' - (C - C_0)\Delta_0(0)'\Phi_0' B_0' - B_0 \Phi_0 \Delta_0(0)(C - C_0)' + (C - C_0)\Gamma_0(0)(C - C_0)' = \Sigma + BB', \quad \text{(a.7)}$$

$$B_0 \Phi_0 B_0' - (C - C_0)[B_0 B_0' + \Sigma_0 + C_0 \Delta_0(0)'\Phi_0' B_0'] - B_0 \Phi_0^2 \Delta_0(0)(C - C_0)'$$
$$+ (C - C_0)\Gamma_0(1)(C - C_0)' = B\Phi B'. \quad \text{(a.8)}$$

Let us now consider solutions (B, C, Σ, Φ) of the matrix equations system (a.7)-(a.8) such that:

$$C = C_0 + B_0 A B_0' \Sigma_0^{-1}, \qquad \Sigma = \Sigma_0, \quad \text{(a.9)}$$

where A is a (K, K) matrix. By replacing (a.9) into (a.7), we get:

$$B_0 \left[Id - A B_0' \Sigma_0^{-1} \Delta_0(0)' \Phi_0' - \Phi_0 \Delta_0(0) \Sigma_0^{-1} B_0 A' + A B_0' \Sigma_0^{-1} \Gamma_0(0) \Sigma_0^{-1} B_0 A' \right] B_0 = BB'. \quad \text{(a.10)}$$

The symmetric matrix:

$$R_0 = Id - A B_0' \Sigma_0^{-1} \Delta_0(0)' \Phi_0' - \Phi_0 \Delta_0(0) \Sigma_0^{-1} B_0 A' + A B_0' \Sigma_0^{-1} \Gamma_0(0) \Sigma_0^{-1} B_0 A' = Id - \Phi_0 \Phi_0' + V_0[\Phi_0 f_t - A B_0' \Sigma_0^{-1} y_t],$$

is positive definite. Let us define:

$$B = B_0 R_0^{1/2} Q_0, \quad \text{(a.11)}$$

where the orthogonal (K, K) matrix Q_0 is such that $Q_0' R_0^{1/2} B_0' B_0 R_0^{1/2} Q_0$ is diagonal. Then, matrix B defined in (a.11) solves equation (a.10) and satisfies the restriction that $B'B$ is diagonal.

Finally, let us replace (a.9), (a.11) into equation (a.8). We get:

$$B_0 \left[\Phi_0 - A B_0' \Sigma_0^{-1} B_0 - A - A B_0' \Sigma_0^{-1} C_0 \Delta_0(0)' \Phi_0' - \Phi_0^2 \Delta_0(0) \Sigma_0^{-1} B_0 A' \right.$$
$$\left. + A B_0' \Sigma_0^{-1} \Gamma_0(1) \Sigma_0^{-1} B_0 A' \right] B_0' = B_0 R_0^{1/2} Q_0 \Phi Q_0' R_0^{1/2} B_0',$$

which is satisfied if:

$$\Phi = Q_0' R_0^{-1/2} \left[\Phi_0 - A B_0' \Sigma_0^{-1} B_0 - A - A B_0' \Sigma_0^{-1} C_0 \Delta_0(0)' \Phi_0' - \Phi_0^2 \Delta_0(0) \Sigma_0^{-1} B_0 A' \right.$$
$$\left. + A B_0' \Sigma_0^{-1} \Gamma_0(1) \Sigma_0^{-1} B_0 A' \right] R_0^{-1/2} Q_0. \quad \text{(a.12)}$$

Thus, (B, C, Σ, Φ) defined in (a.9), (a.11), (a.12) is a solution of the second-order nonlinear moment restrictions.

The above set of solutions of the second-order nonlinear moment restrictions is indexed by matrix A. When the elements of this matrix are selected sufficiently close to zero, the solution (B, C, Σ, Φ) gets arbitrarily close to the true parameter value $(B_0, C_0, \Sigma_0, \Phi_0)$. This shows that unidentifiability applies both locally and globally.

Appendix 2: Semi-parametric identification in the non-Gaussian linear model with common factor and contagion

In this Appendix we first show equation (5.17), which is used to prove Proposition 6 on the identification in the semi-parametric linear model with contagion and CaR common factor considered in Section 5.3. Then, we show equation (5.23) and property (5.25), which are used to prove Proposition 7.

A.2.1 Proof of equation (5.17)

Let us define $h(u,\tilde{u},C) = \log E[\exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]$. We have:

$$\frac{\partial h(u,\tilde{u},C)}{\partial u} = \frac{E[(y_t - Cy_{t-1})\exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]}{E[\exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]},$$

and:

$$\begin{aligned} H(u,\tilde{u},C) = \frac{\partial^2 h(u,\tilde{u}',C)}{\partial u \partial \tilde{u}'} &= \frac{E[(y_t - Cy_{t-1})(y_{t-1} - Cy_{t-2})' \exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]}{E[\exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]} \\ &\quad - \frac{E[(y_t - Cy_{t-1})\exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]}{E[\exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]} \\ &\quad \times \frac{E[(y_{t-1} - Cy_{t-2})' \exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]}{E[\exp(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}))]} \\ &= \widetilde{Cov}(y_t - Cy_{t-1}, y_{t-1} - Cy_{t-2}), \end{aligned}$$

where $\widetilde{Cov}$ denotes the covariance w.r.t. the modified probability measure $\tilde{P}$ defined by (5.18). By using $y_t - Cy_{t-1} = B_0 f_t + \varepsilon_t - \Delta y_{t-1}$, where $\Delta = C - C_0$, we get:

$$H(u,\tilde{u},C) = \widetilde{Cov}(B_0 f_t + \varepsilon_t - \Delta y_{t-1}, B_0 f_{t-1} + \varepsilon_{t-1} - \Delta y_{t-2}). \tag{a.13}$$

Let us now compute the covariance in the r.h.s. Using the model $y_t = C_0 y_{t-1} + B_0 f_t + \varepsilon_t$, we can write y_t as a linear filter of $\underline{\varepsilon_t}$ and $\underline{f_t}$. Thus, the change of measure from P to $\tilde{P}$ is such that:

$$\frac{d\tilde{P}}{dP} = \frac{\exp(\sum_{j=0}^{\infty} a_j' \varepsilon_{t-j} + \sum_{j=0}^{\infty} b_j' f_{t-j})}{E[\exp(\sum_{j=0}^{\infty} a_j' \varepsilon_{t-j} + \sum_{j=0}^{\infty} b_j' f_{t-j})]}, \tag{a.14}$$

for some coefficients a_j and b_j that depend on u, $\tilde{u}$ and C. This change of measure is multiplicative w.r.t. variables ε_{t-j} and f_{t-j}, for j varying. We use the next Lemma, which is proved at the end of this subsection.

Lemma A.1: *Let X and Y be independent random variables under the probability measure P. Define the modified probability measure $\tilde{P}$, absolutely continuous w.r.t. P with derivative $d\tilde{P}/dP = a(X)b(Y)$, for some functions a and b with $E[a(X)] = E[b(Y)] = 1$. Then, variables X and Y are independent also under the modified probability measure $\tilde{P}$.*

From Lemma A.1, under $\tilde{P}$ variables ε_t, for t varying, are independent, and independent of the factor process (f_t). Then, from (a.13) we get:

$$\begin{aligned} H(u,\tilde{u},C) &= B_0 \widetilde{Cov}(f_t, f_{t-1}) B_0' - B_0 \widetilde{Cov}(f_t, y_{t-2}) \Delta' \\ &\quad - \Delta \widetilde{Cov}(y_{t-1}, f_{t-1}) B_0' - \Delta \widetilde{Cov}(y_{t-1}, \varepsilon_{t-1}) + \Delta \widetilde{Cov}(y_{t-1}, y_{t-2}) \Delta'. \end{aligned}$$

By using $\widetilde{Cov}(y_{t-1}, \varepsilon_{t-1}) = \tilde{V}(\varepsilon_t)$, and rearranging terms, equation (5.17) follows.

Proof of Lemma A.1: Let g and h be functions, that are integrable w.r.t. the distributions of random variables X and Y, respectively. We have:

$$\begin{aligned}\widetilde{Cov}[g(X),h(Y)] &= \tilde{E}[g(X)h(Y)]-\tilde{E}[g(X)]\tilde{E}[h(Y)] = E[a(X)b(Y)g(X)h(Y)]-E[a(X)b(Y)g(X)]E[a(X)b(Y)h(Y)] \\ &= E[a(X)g(X)]E[b(Y)h(Y)]-E[a(X)g(X)]E[b(Y)]E[a(X)]E[b(Y)h(Y)]=0,\end{aligned}$$

since X and Y are independent under P, and $E[a(X)]=E[b(Y)]=1$. Since the covariance $\widetilde{Cov}[g(X),h(Y)]$ vanishes for any integrable functions g and h, variables X and Y are independent under $\tilde{P}$. Q.E.D.

A.2.2 Proof of equation (5.23)

By equation (5.22), the serial independence of process $(\tilde{\varepsilon}_t)$, and the independence between processes (f_t) and $(\tilde{\varepsilon}_t)$, for any $h\geq 1$ we have $E_0[\exp(u'\tilde{y}_t+v'\tilde{y}_{t-h})]=\exp\left(\psi_{\tilde{\varepsilon}}(u)+\psi_{\tilde{\varepsilon}}(v)\right)E_0[\exp(u'f_t+v'f_{t-h})]$, where $\psi_{\tilde{\varepsilon}}$ is the log-Laplace transform of $\tilde{\varepsilon}_t$. It follows:

$$\tau_0(u,v,h)=\frac{E_0[\exp(u'\tilde{y}_t+v'\tilde{y}_{t-h})]}{E_0[\exp(u'\tilde{y}_t)]E_0[\exp(v'\tilde{y}_{t-h})]}=\frac{E_0[\exp(u'f_t+v'f_{t-h})]}{E_0[\exp(u'f_t)]E_0[\exp(v'f_{t-h})]}. \tag{a.15}$$

Let us now compute the ratio in the r.h.s. By using $b_0(u)=\varphi_f^0(u)-\varphi_f^0[a_0(u)]$, the transition of the factor is such that:

$$E_0[\exp(u'f_t)|f_{t-1}]=\exp(a_0(u)'f_{t-1}+b_0(u))=\exp\left(\varphi_f^0(u)+a_0(u)'f_{t-1}-\varphi_f^0[a_0(u)]\right).$$

By using the Law of Iterated Expectation, we can show $E_0[\exp(u'f_t)|f_{t-h}]=\exp\Big(\varphi_f^0(u)+a_0^{\circ h}(u)'f_{t-h}-\varphi_f^0[a_0^{\circ h}(u)]\Big)$, for any $h\geq 1$. Then, we have:

$$\begin{aligned}E_0[\exp(u'f_t+v'f_{t-h})] &= \exp\left(\varphi_f^0(u)-\varphi_f^0[a_0^{\circ h}(u)]\right)E_0\left[\exp\left([a_0^{\circ h}(u)+v]'f_{t-h}\right)\right] \\ &= \exp\left(\varphi_f^0[a_0^{\circ h}(u)+v]+\varphi_f^0(u)-\varphi_f^0[a_0^{\circ h}(u)]\right). \end{aligned} \tag{a.16}$$

By using $E_0[\exp(u'f_t)]=\exp[\varphi_f^0(u)]$ and $E_0[\exp(v'f_{t-h})]=\exp[\varphi_f^0(v)]$, from equations (a.15) and (a.16) we get (5.23).

A.2.3 Proof of property (5.25)

From (5.23) we have $\tau_0(u,v;1)=\exp\Big(\varphi_f^0[a_0(u)+v]-\varphi_f^0[a_0(u)]-\varphi_f^0(v)\Big)$. Thus, (5.25) is equivalent to:

$$\varphi_f^0[a_0(u_1)+v]-\varphi_f^0[a_0(u_1)]=\varphi_f^0[a_0(u_2)+v]-\varphi_f^0[a_0(u_2)],\ \forall v \quad \Rightarrow \quad u_1=u_2. \tag{a.17}$$

To prove property (a.17), let us assume that $u_1,u_2\in\mathcal{D}_0$ are such that $\varphi_f^0[a_0(u_1)+v]-\varphi_f^0[a_0(u_1)]=\varphi_f^0[a_0(u_2)+v]-\varphi_f^0[a_0(u_2)]$, $\forall v$. By considering infinitesimal vectors v, it follows $\partial\varphi_f^0/\partial u[a_0(u_1)]=\partial\varphi_f^0/\partial u[a_0(u_2)]$. Then, we get:

$$0=\frac{\partial\varphi_f^0}{\partial u}[a_0(u_2)]-\frac{\partial\varphi_f^0}{\partial u}[a_0(u_1)]=\int_0^1\frac{\partial^2\varphi_f^0[a_0(u_1)+t(a_0(u_2)-a_0(u_1))]}{\partial u\partial u'}dt(a_0(u_2)-a_0(u_1)). \tag{a.18}$$

Now, function φ_f^0 is strictly convex. Indeed, we have $\partial^2\varphi_f^0(u)/\partial u\partial u'=V_0^u(f_t)$, where $V_0^u(\cdot)$ denotes the variance-covariance operator w.r.t. the probability measure P_0^u defined by the change of measure $dP_0^u/dP_0=\exp(u'f_t)/E_0[\exp(u'f_t)]$. Thus, matrix $\int_0^1\frac{\partial^2\varphi_f^0[a_0(u_1)+t(a_0(u_2)-a_0(u_1))]}{\partial u\partial u'}dt$ is positive definite, and in particular non-singular. Then, from (a.18) it follows $a_0(u_2)-a_0(u_1)=0$, i.e. $a_0(u_1)=a_0(u_2)$. By the one-to-one property of function a_0, we conclude $u_1=u_2$.