

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

Patrick Gagliardini (University of Lugano) *

Christian Gouriéroux (CREST and University of Toronto)

Preliminary version: October, 2014

Abstract

We consider (semi-)parametric models for time series and panel data including unobserved dynamic effects. When these (semi-)parametric regression models have an affine specification with respect to both lagged endogenous variables and unobserved dynamic effects, we derive moment restrictions based on appropriate Laplace transforms. We show how these moment restrictions can be used to identify not only the regression parameters of the affine model, but also the parametric or nonparametric distribution of the dynamic effects. The methodology is applied to a variety of panel models with either stochastic time effect, or with unobserved effort variables introduced to capture the moral hazard phenomena. This approach is appropriate for studying identification in the linear and nonlinear factor models encountered in macroeconomic and financial applications.

Keywords: Identification, Nonparametric Identification, Quasi-Differencing, Cross-Differencing, Panel Count Data, Dynamic Frailty, Nonlinear Factor Model.

*Corresponding author: Patrick Gagliardini, Università della Svizzera Italiana, Via Buffi 13, CH-6900 Lugano, Switzerland. Phone: +41 58 666 4660. Fax: +41 58 4734. Email: patrick.gagliardini@usi.ch. Acknowledgments: The authors thank D. Wilhelm and participants at the Econometric Society European Meeting (ESEM) 2014 in Toulouse and at the CEMMAP seminar of the University College London for very useful comments.

1 Introduction

We consider (semi-) parametric models for time series and panel data including unobserved dynamic effects. When these (semi-) parametric regression models have an affine specification with respect to both lagged endogenous variables and unobserved dynamic effects, we derive moment restrictions based on appropriate Laplace transforms. We show how these 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 effects. The methodology is applied to a variety of panel models with either stochastic time effects, or with unobserved effort variables introduced to capture the moral hazard phenomena. The approach is appropriate for studying the linear and nonlinear factor models encountered in macroeconomic and financial applications.

In Section 2, we first provide examples of models with unobserved dynamic effects. They include panel models with stochastic time effects, Poisson models with individual effort variables encountered in car insurance, linear factor models for asset returns, multivariate stochastic volatility models, models 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. In the general specification, the conditional Laplace transform of the endogenous variables is an exponential affine function of the lagged endogenous variables, the unobservable effects, and possibly exogenous observable regressors.

We derive in Section 3 moment restrictions based on the Laplace transform of a well-chosen transformation of the current and lagged observations y_t, y_{t-1} , say, of the endogenous variables. We explain how these first-order nonlinear moment restrictions can be used to estimate the (nonlinear) regression parameters and to identify nonparametrically the marginal distribution of the time effect for panel data with stochastic time effect. This is done by a nonlinear cross-sectional extension of the quasi-differencing approach introduced in Chamberlain (1992) and Mullahy (1997).

Without panel data, these first-order nonlinear moment restrictions identify the parameters and the marginal distribution of the unobserved effect, when this distribution is parametric. When the distribution of the dynamic effect is left unspecified, these moment restrictions allow to reconstitute the marginal distribution of the unobserved effects once the regression parameters of the affine models are known.

To identify the dynamic parameters of the distribution of the unobserved process, or to identify both the parameters of the affine model and the distribution of this process, when it is not parametrically specified, we need second- and third-order moment restrictions. They are obtained in Section 4 by

considering the Laplace transform of appropriate functions of y_t, y_{t-1}, y_{t-2} , and of $y_t, y_{t-1}, y_{t-2}, y_{t-3}$, respectively. The identification results are illustrated by the examples introduced in Section 2. Section 5 concludes. Proofs are gathered in the Appendix.

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 will be used later on to derive the moment restrictions in a general framework.

2.1 Examples

Example 1: The Gaussian Panel Model with Stochastic Time Effect.

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

$$y_{i,t} = f_t + x_{i,t}\alpha + \varepsilon_{i,t}, \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

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 2: Count Panel Data with Stochastic Time Effect.

The model is similar to the model in Example 1, except that the conditional regression model 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.

Example 3: Panel Model with Unobservable Effort Variables.

This type of model has been introduced in car insurance to account for moral hazard phenomena when computing the updating of the insurance premium, that is the so-called bonus-malus [see e.g. Gouieroux, Jasiak (2004)]. The variable of interest is the number of claims for the insured driver i on

a given year t . We get:

$$y_{i,t} \sim \mathcal{P}(f_{i,t} + x_{i,t}\alpha + y_{i,t-1}c), \quad i = 1, \dots, n, \quad t = 1, \dots, T.$$

The individual “effort” processes $(f_{i,t}), i = 1, \dots, n$ are positive-valued and usually assumed independent, with identical distributions. For instance, they have a log-normal autoregressive dynamics such that:

$$\log f_{i,t} = \rho \log f_{i,t-1} + u_{i,t},$$

where the innovations $u_{i,t}$ are $IIN(0, \sigma^2)$ and the autocorrelation coefficient ρ is such that $|\rho| < 1$.

Example 4: A Gaussian Linear Factor Model for Asset Returns.

The Capital Asset Pricing Model (CAPM) [see Sharpe (1964), Lintner (1965)] leads to a linear factor model with a single factor equal to the market return. However, the market return is not easy to measure in practice, and an index return used as a proxy can provide an erroneous measurement. This is the well-known Roll’s critique [see Roll (1977)]. This explains why the basic CAPM model is often replaced by a model with unobservable factors.

Let us denote by y_t the $(n, 1)$ vector of asset returns, by f_t the vector of values of the underlying latent factors. The model is written as:

$$y_t = Bf_t + \varepsilon_t,$$

where the error terms are $IIN(0, \Sigma)$ and Σ is a diagonal matrix. In this strict 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 fat tails observed in the historical distribution of returns and with volatility clustering effects, whenever these effects are created by conditionally heteroscedastic 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, Wadhvani (1994)].

Example 5: Correlation Between Markets.

Let us consider a vector 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, by means of the unobservable common factor f_t and the lagged returns y_{t-1} , respectively.

This model nests the linear factor model of Example 4 and a Vector Autoregressive (VAR) specification, which correspond to the restrictions $C = 0$, and $B = 0$, respectively.

Example 6: Correlation of Default Risks.

The decomposition of Example 5 can be adapted to analyze default risk. Such a model is introduced in Darolles, Gagliardini, Gourioux (2014). The authors consider the population of hedge funds classified by management style k , for $k = 1, \dots, K$. For each management style k , the observation $y_{k,t}$ concerns 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:

$$y_{k,t} \sim \mathcal{P}(a_k + b'_k f_t + c'_k y_{t-1}), \quad k = 1, \dots, K,$$

where the variables $y_{k,t}$, $k = 1, \dots, K$, are assumed conditionally independent given f_t and y_{t-1} . This modeling allows for disentangling the contagion effects passing through the lagged count values y_{t-1} and the effect of exogenous common shocks passing through the unobservable systematic factors f_t .

Example 7: Stochastic Volatility Model.

Such a model can be written on asset returns as:

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

where $\varepsilon_t \sim IIN(0, Id)$, the symmetric positive definite matrix Σ_t is the stochastic volatility-covolatility matrix, and $f_t = \text{vech}(\Sigma_t)$ is a vector stacking all its different components. By introducing the term $B\text{vech}(\Sigma_t)$ 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, Gourioux (2014) 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 $\underline{y}_{t-1} = (y_{t-1}, y_{t-2}, \dots)$ of these endogenous variables, and the current and past values of the unobservable dynamic effects $\underline{f}_t = (f_t, f_{t-1}, \dots)$ and the covariates $\underline{x}_t = (x_t, x_{t-1}, \dots)$. 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:*

$$E[\exp(u'y_t)|\underline{y}_{t-1}, \underline{f}_t, \underline{x}_t] = \exp \{a(u, x_t, \theta)'[Bf_t + Cy_{t-1} + d(x_t, \theta)] + b(u, x_t, \theta)\}, \quad (2.1)$$

where u is the multidimensional argument of the Laplace transform, θ , B , C are parameters, possibly constrained, and a , b , d are known functions. The model is well-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, Gouriéroux, Jasiak (2006) to the case of covariates and unobserved stochastic effect f_t . Indeed, without x_t and f_t , the condition of Assumption A.1 restricts the model to (nonlinear) dynamics of the type:

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

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 name of the associated autoregressive model.

i) Gaussian family.

We have:

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

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, 4, 5 and 7). We get for instance the Gaussian VAR process (resp., the model in Example 5) 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:

$$E[\exp(uy)] = \exp[-\lambda(1 - \exp u)].$$

This Laplace transform is defined for any real 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 Examples 2, 3 and 6).

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 sum of independent Poisson $\mathcal{P}(\lambda)$ and Binomial $Bin(y_{t-1}, p)$ distributions, 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 can be obtained by letting parameter λ be time-varying as a linear function of factor f_t .

iv) The gamma family.

Let us now assume that y follows the gamma distribution $\gamma(\nu, \lambda)$ with degree of freedom ν , $\nu > 0$, and intensity λ , $\lambda > 0$. We get:

$$E[\exp(uy)] = (1 - \lambda u)^{-\nu} = \exp[-\nu \log(1 - \lambda u)],$$

defined for any argument u such that $u < 1/\lambda$. The degree of freedom can be specified as a linear function of f_t , y_{t-1} and a possibly nonlinear function of x_t , whenever this specification satisfies the positivity condition.

v) The noncentered gamma and Wishart families.

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:

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

where $\nu, \nu > 0$, is the degree of freedom, the (K, K) symmetric positive definite matrix Σ is a variance-covariance matrix and the (K, K) matrix M a noncentrality parameter. The associated affine regression model is obtained by specifying the noncentrality parameter as an affine function of the lagged values and of the unobserved dynamic effect.

The noncentral chi-square family and noncentral gamma family are obtained from the one dimensional case. For the noncentral gamma family, we have:

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

with shape, noncentrality and scale parameters ν , m and λ , respectively. The autoregressive process corresponding to the noncentered gamma family is the AutoRegressive Gamma (ARG) process [Gourieroux, Jasiak (2006)], that is the time discretized Cox, Ingersoll, Ross process [Cox, Ingersoll, Ross (1985)]. 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 affine regression specification in Assumption A.1 is completed by additional assumptions on the unobservable dynamic effect, the regressors and their joint distribution.

Assumption A.2: *The joint process $(f'_t, x'_t)'$ of the unobservable dynamic effect and the 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.*

The notion of exogeneity in Assumption A.2 is a nonlinear notion, which requires restrictions on conditional distributions, and not on conditional expectations only. 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 $\underline{f} = (\dots, f_{t-1}, f_t, f_{t+1}, \dots)$ and $\underline{x} = (\dots, x_{t-1}, x_t, x_{t+1}, \dots)$ denote the entire histories including past, current and future values of the dynamic effects, and the regressors, respectively.

The next assumption is for expository purpose.

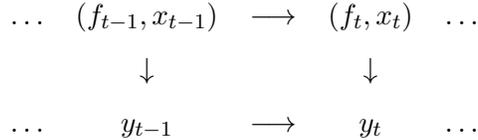
Assumption A.3: *The joint process (f_t, x_t) is a stationary Markov process of order 1.*

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 .

The joint dynamics of (f_t, x_t, y_t) is defined along the following causal scheme:

Figure 1: The Causal Scheme.



Under Assumptions A.2-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. This transition p.d.f. is either specified parametrically, or let unspecified, in the analysis of the next sections.

The (semi-) parametric framework of Assumption A.1-A.3 is more structural than the nonparametric setting considered in Hu, Shum (2012) and Hu, Shiu (2013). While the exponential affine regression in Assumption A.1, and the possibly parametric state transition in Assumptions A.2-A.3, constrain the generality of the nonlinear state space dynamics, the assumed structure helps us in deriving more informative restrictions for identification (and estimation).

3 First-order nonlinear moment restrictions

Under Assumptions A.1-A.3, it is possible to derive continuum sets of nonlinear moment restrictions which relate the observations and the unknown real or functional parameters. We first focus in this section on first-order restrictions. Then we explain how they can be used to identify parameters of interest.

3.1 The moment restrictions

Under Assumption A.1, we get: ¹

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

¹In (3.1) the qualification $\forall u$ means validity of the restriction for any argument u in a domain of \mathbb{C}^n such that the conditional Laplace transform is well-defined. For expository purpose, we do not make explicit this domain each time we write a nonlinear moment restriction.

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 of the equation and in the conditioning set in the left-hand side. However, they can be used to derive conditional moment restrictions involving observable variables only, in two different ways.

i) First, we can integrate out both sides of equation (3.1) conditional on \underline{x}_t . By the Iterated Expectation Theorem, we get the following conditions:

First-order nonlinear moment restrictions:

$$\begin{aligned} & E[\exp\{u'y_t - a(u, x_t, \theta)'[Cy_{t-1} + d(x_t, \theta)] - b(u, x_t, \theta)\}|\underline{x}_t] \\ = & E[\exp\{a(u, x_t, \theta)'Bf_t\}|\underline{x}_t], \quad \forall u, \end{aligned} \quad (3.2)$$

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 well-specified, these restrictions are valid for the true values of the parameters and the true conditional distributions. The subindex 0 denoting the true values is omitted for expository purpose. In (3.2) we have a continuum set of moment restrictions, since the argument u can vary over some domain in \mathbb{C}^n .

ii) Second, in the framework of panel data with stochastic fixed effect, conditions (3.1) can be combined to eliminate the common effect f_t in the right hand side. Then, by applying the conditional expectation given $\underline{x}_t, \underline{y}_t$, we get moment restrictions which involve parameters B, C, θ only. This extension of the quasi-differencing method to a nonlinear cross-sectional framework is discussed in the next subsection.

3.2 Nonlinear cross-differencing for panel data

Let us focus in this subsection on panel data with stochastic time effect (see Examples 1 and 2). We assume:

Assumption A.1*:

- i) Variables $y_{i,t}$ and f_t are one-dimensional.
- ii) The individual histories $(y_{i,t}, x_{i,t}, t \text{ varying})$, for $i = 1, \dots, n$, are independent conditional on \underline{f}_t .

iii) *The individual conditional Laplace transform is given by:*

$$E[\exp(uy_{i,t})|\underline{y}_{t-1}, \underline{f}_t, \underline{x}_t] = \exp\{a(u, x_{i,t}, \theta)[f_t + cy_{i,t-1} + d(x_{i,t}, \theta)] + b(u, x_{i,t}, \theta)\}.$$

Assumption A.1* is a stronger version of Assumption A.1, which is suitable for the panel data setting.

The moment restrictions (3.1) can be written for the different individuals as:

$$\begin{aligned} & E \left[\exp\{uy_{i,t} - a(u, x_{i,t}, \theta)[cy_{i,t-1} + d(x_{i,t}, \theta)] - b(u, x_{i,t}, \theta)\} | \underline{y}_{i,t-1}, \underline{f}_t, \underline{x}_t \right] \\ &= \exp[a(u, x_{i,t}, \theta)f_t], \quad \forall u, \forall f_t, \forall i = 1, \dots, n. \end{aligned} \quad (3.3)$$

Let us also assume:

Assumption A.4*: *The function $u \rightarrow a(u, x_{i,t}, \theta)$ is continuous and strictly monotonous w.r.t. the argument u , for any given $x_{i,t}$ and θ .*

Under Assumption A.4* we can define:

$$u(v, x_{i,t}, \theta) \text{ as the solution of the equation } a(u, x_{i,t}, \theta) = v,$$

and replace argument u by $u(v, x_{i,t}, \theta)$ in equation (3.3). We get:

$$\begin{aligned} & E \left[\exp\{u(v, x_{i,t}, \theta)y_{i,t} - v[cy_{i,t-1} + d(x_{i,t}, \theta)] - b[u(v, x_{i,t}, \theta), x_{i,t}, \theta]\} | \underline{y}_{i,t-1}, \underline{f}_t, \underline{x}_t \right] \\ &= \exp(vf_t), \quad \forall v, \forall f_t, \forall i = 1, \dots, n. \end{aligned}$$

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

$$\begin{aligned} & E \left[\exp\{u(v, x_{i,t}, \theta)y_{i,t} - v[cy_{i,t-1} + d(x_{i,t}, \theta)] - b[u(v, x_{i,t}, \theta), x_{i,t}, \theta]\} | \underline{y}_{i,t-1}, \underline{f}_t, \underline{x}_t \right] \\ &= E \left[\exp\left\{u(v, x_{j,t}, \theta)y_{j,t} - v[cy_{j,t-1} + d(x_{j,t}, \theta)] - b[u(v, x_{j,t}, \theta), x_{j,t}, \theta]\right\} | \underline{y}_{j,t-1}, \underline{f}_t, \underline{x}_t \right], \\ & \forall v, \forall i, j, i \neq j. \end{aligned}$$

These restrictions still depend on the unobservable component through the conditioning sets. But, by taking the conditional expectations of both sides with respect to $\underline{y}_{i,t-1}, \underline{y}_{j,t-1}, \underline{x}_t$, and by using Assumption A.1*, we get the following set of moment restrictions:

First-order nonlinear moment restrictions with cross-differencing:

$$\begin{aligned} & E \left[\exp\{u(v, x_{i,t}, \theta)y_{i,t} - v[cy_{i,t-1} + d(x_{i,t}, \theta)] - b[u(v, x_{i,t}, \theta), x_{i,t}, \theta]\} | \underline{y}_{i,t-1}, \underline{y}_{j,t-1}, \underline{x}_t \right] \\ &= E \left[\exp\{u(v, x_{j,t}, \theta)y_{j,t} - v[cy_{j,t-1} + d(x_{j,t}, \theta)] - b[u(v, x_{j,t}, \theta), x_{j,t}, \theta]\} | \underline{y}_{j,t-1}, \underline{y}_{i,t-1}, \underline{x}_t \right], \\ & \forall v, \forall i, j, i \neq j. \end{aligned} \quad (3.4)$$

We get a continuum of moment restrictions that is indexed by the pairs of individuals. The restrictions involve parameters c and θ of the affine regression model, but no parameters of the distribution of the stochastic time effect. They are generically **sufficient to identify all regression parameters**.

This approach is easily extended to multidimensional panel data $y_t = (y_{1,t}, \dots, y_{K,t})'$ with stochastic time effects $f_t = (f_{1,t}, \dots, f_{K,t})'$ specific to each type of variable. However, this approach is not applicable to time series (corresponding to a single "individual" only), where the heterogeneity of the factor sensitivities is not observable and involves unknown parameters b_i . It is also not applicable to Example 3 in which the effort variable depends also on the individual.

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 similar restrictions based on the first-order moments are simply deduced from (3.4) by equating the derivatives of both sides of equation (3.4) with respect to v at $v = 0$. Our 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 (3.4) is more informative than the ones usually considered in the panel literature and based on the first-order moment only.

By comparing the moment estimators based on the nonlinear moment restrictions with cross-differencing (3.4) with the standard quasi-differencing based on first-order moments only, we may construct new specification tests.

Example 2 (continues):

Let us illustrate the approach above with the model for count panel data. We have $y_{i,t} | \underline{y}_{i,t-1}, \underline{f}_t, \underline{x}_t \sim \mathcal{P}(f_t + x_{i,t}\alpha + y_{i,t-1}c)$. The conditional Laplace transform is:

$$E[\exp(uy_{i,t}) | \underline{y}_{i,t-1}, \underline{f}_t, \underline{x}_t] = \exp\{-(f_t + x_{i,t}\alpha + y_{i,t-1}c)(1 - \exp u)\},$$

with function a in Assumption A.1* given by $a(u) = 1 - \exp(u)$, independent of parameters and

regressors. The moment restrictions (3.4) become:

$$\begin{aligned} & E \left[\exp\{uy_{it} + (1 - \exp u)(x_{i,t}\alpha + y_{i,t-1}c)\} | \underline{y}_{i,t-1}, \underline{y}_{j,t-1}, \underline{x}_t \right] \\ = & E \left[\exp\{uy_{jt} + (1 - \exp u)(x_{j,t}\alpha + y_{j,t-1}c)\} | \underline{y}_{i,t-1}, \underline{y}_{j,t-1}, \underline{x}_t \right], \quad \forall i \neq j, \forall u. \end{aligned}$$

Since function a is independent of the regressors, we have not applied the change of argument, which is simply:

$$1 - \exp u = v \iff u = \log(1 - v).$$

By considering the first-order expansion w.r.t. u at $u = 0$, the equations above become:

$$E[y_{i,t} - x_{i,t}\alpha - y_{i,t-1}c | \underline{y}_{i,t-1}, \underline{y}_{j,t-1}, \underline{x}_t] = E[y_{j,t} - x_{j,t}\alpha - y_{j,t-1}c | \underline{y}_{i,t-1}, \underline{y}_{j,t-1}, \underline{x}_t], \quad \forall i \neq j,$$

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

3.3 Identification of the marginal distribution of f_t

Let us now come back to the general specification in Assumptions A.1-A.2. The system of moment restrictions described in (3.2) is valid for any value of argument u , whenever the Laplace transform exists. Since the moment restrictions 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.²

Assumption A.4: *There exist a change of argument from u to $v = (v'_1, v'_2)' \in \mathbb{R}^K \times \mathbb{R}^{n-K}$, such that $u(v, x_t, \theta, B)$, say, satisfies the equation:*

$$a[u(v, x_t, \theta, B), x_t, \theta]' B = v'_1, \quad \forall v, B, \theta, x_t.$$

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

Nonlinear first-order moment restrictions for the marginal distribution of (f_t) :

$$\begin{aligned} & E[\exp\{u(v, x_t, \theta, B)'y_t - a[u(v, x_t, \theta, B), x_t, \theta]'[Cy_{t-1} + d(x_t, \theta)] - b[u(v, x_t, \theta, B), x_t, \theta]\} | \underline{x}_t] \\ = & E[\exp(v'_1 f_t) | \underline{x}_t], \quad \forall v. \end{aligned} \tag{3.5}$$

²For panel data, Assumption A.4* implies Assumption A.4, since the arguments u and v are both scalar.

System (3.5) allows for identifying the true conditional distribution of f_t given \underline{x}_t , and also the true unconditional (stationary) distribution of process (f_t) , once the true regression parameters B, C, θ are known. It also allows in general to identify some functions of the regression parameters, when the number K of linearly independent components of f_t is strictly smaller than the dimension n of the endogenous variable.

3.4 Identification by the first-order nonlinear moment restrictions

The discussions above provide identification properties from the first-order nonlinear moment restrictions (3.1). The identification properties summarized below are generic, since these properties can require additional analysis of the way regression parameters B, C, θ and parameters of the distribution of (f_t) are introduced in each model. This will be illustrated in the examples of Subsection 3.5.

Proposition 1: *Under Assumptions A.1-A.4 (or A.1*-A.4*), generically, the first-order nonlinear moment restrictions identify:*

- i) *The conditional distribution of f_t given \underline{x}_t , once the parameters B, C, θ are known [by restrictions (3.5)];*
- ii) *The parameters B, C, θ by nonlinear cross-differencing; then, the conditional distribution of f_t given \underline{x}_t is nonparametrically identified, for a panel data model with stochastic time effect [by restrictions (3.4) and (3.5)].*
- iii) *The regression parameters B, C, θ and the parameters characterized by the conditional distribution of f_t given \underline{x}_t , when the distribution of the factor and covariates is specified parametrically [by restrictions (3.5)]. The parameters characterizing the dynamic of (f_t) (conditional on the covariates) are not identifiable.*

In models without covariates, the identification properties above concern the unconditional (stationary) distribution of the factor f_t , but not its dynamics in general.

3.5 Examples

As an illustration let us discuss parameter identification from the first-order nonlinear moment restrictions in the factor models encountered in Finance (see Examples 4 and 5). These models have no observable covariates. We consider *i*) a pure Gaussian model, *ii*) a conditionally Gaussian model with

one factor featuring a stochastic volatility. The Gaussian example shows how the restrictions are related to the identifiability conditions encountered in the standard factor model underlying the singular value decomposition [see e.g. Anderson (2003)] and to the identification problems encountered in Structural VAR models [see e.g. Sims (1980), Forni et al. (2000) and Alessi et al. (2008)]. The model with stochastic volatility shows that identifiability may become easier in a nonlinear dynamic framework. For the discussion below, it is important to distinguish the generic parameter values and distributions from their true values. Therefore, we introduce explicitly the subindex 0 for true values when it is needed.

Example 5: A Gaussian model with common factor and contagion (continues).

Let us consider the dynamic system:

$$\begin{cases} y_t = Bf_t + Cy_{t-1} + \Sigma^{1/2}\varepsilon_t, \\ f_t = \Phi f_{t-1} + (Id - \Phi\Phi')^{1/2}\eta_t, \end{cases} \quad (3.6)$$

where $\dim(y_t) = n$, $\dim(f_t) = K$ and the vector of error terms is a standard Gaussian white noise $(\varepsilon_t', \eta_t')' \sim IIN(0, Id)$. Without loss of generality, the multidimensional unobservable factor f_t is normalized 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. Model (3.6) is the standard dynamic factor model considered in Stock, Watson (1998). It also includes the factor augmented vector autoregression introduced by Bernanke, Boivin, Elias (2005), when vector y_t includes proxies of the unobservable factors.³

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:

$$y_t = Bf_t + \Sigma^{1/2}\varepsilon_t = B\eta_t + \Sigma^{1/2}\varepsilon_t.$$

³See e.g. Wilhelm (2013) for nonparametric identification of panel data models with measurement errors.

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 exogenous common factor f_t . This is the static linear factor model used in exploratory factor analysis [Lawley, Maxwell (1971), Anderson (2003)]. In this framework, the first-order nonlinear moment restrictions (3.2) become:

$$E_0[\exp(u'y_t)] = \exp\left\{\frac{u'(\Sigma + BB')u}{2}\right\}, \quad \forall u, \quad (3.7)$$

where E_0 denotes the expectation with respect to the Data Generating Process (DGP). Therefore, the continuum of moment restrictions (3.7) allows for identifying the value of the transformed parameter $\Sigma_0 + B_0B_0'$, where Σ_0 and B_0 denote the true values of the parameters. 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: 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 identification of the matrices Σ_0 and B_0 [see e.g. Lawley, Maxwell (1971), Section 2.3, for counterexamples, and Anderson, Rubin (1956) for either sufficient, or necessary, identification conditions]. If we assume $\Sigma_0 = \sigma_0^2 Id$, then scalar σ_0^2 is the smallest eigenvalue of matrix $\Sigma_0 + B_0'B_0$, and is identifiable.

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:

$$E_0[\exp\{u'(y_t - Cy_{t-1})\}] = \exp\left\{\frac{u'(\Sigma + BB')u}{2}\right\}, \forall u. \quad (3.8)$$

Since the joint vector $(y_t', y_{t-1}')'$ is multivariate Gaussian, the continuum of moment restrictions (3.8) is equivalent to:

$$V_0(y_t - Cy_{t-1}) = \Sigma + BB', \quad (3.9)$$

that is:

$$\Gamma_0(0) - \Gamma_0(1)C' - C\Gamma_0(1)' + C\Gamma_0(0)C' = \Sigma + BB', \quad (3.10)$$

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 autocorrelation, respectively, of process (y_t) . Equation (3.10) is a linear combination of the

first two Yule-Walker equations for VAR process (y_t) , that are:

$$\begin{aligned}\Gamma_0(0) &= C\Gamma_0(1)' + \Sigma + BB', \\ \Gamma_0(1) &= C\Gamma_0(0).\end{aligned}$$

These Yule-Walker equations jointly identify the true parameter values C_0 and $\Sigma_0 + B_0B_0'$.

Let us now study identification of the true parameters from equation (3.9). We have:

$$\begin{aligned}V_0(y_t - Cy_{t-1}) &= V_0[y_t - C_0y_{t-1} - (C - C_0)y_{t-1}] \\ &= V_0[B_0f_t + \Sigma_0^{1/2}\varepsilon_t - (C - C_0)y_{t-1}] \\ &= \Sigma_0 + B_0B_0' + (C - C_0)\Gamma_0(0)(C - C_0)'.\end{aligned}\tag{3.11}$$

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

$$\Sigma_0 + B_0B_0' + (C - C_0)\Gamma_0(0)(C - C_0)' = \Sigma + BB',\tag{3.12}$$

where B_0 and B are matrices of full rank $K = \dim(f_t)$, 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 (3.12). We deduce easily from (3.12) the following property:

Proposition 2: *We have $(\Sigma, B, C) \in \mathcal{E}_0$ if, and only if:*

$$\Sigma - \Sigma_0 + BB' - B_0B_0' \succeq 0,$$

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 $DD' = \Sigma - \Sigma_0 + BB' - B_0B_0'$.

We deduce the next two corollaries.

Corollary 1: *The parameters are not point identifiable from system (3.12).*

Proof: Parameter (Σ_0, B_0, C_0) is point identifiable if, and only if, the set \mathcal{E}_0 is a singleton, consisting of the single point (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.

However, the parameters are set identifiable, i.e., they are functions of 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 set identifiable.*

Proof: From system (3.12) we have $\Sigma + BB' \succeq \Sigma_0 + B_0B_0'$. 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. Thus, $\Sigma_0 + B_0B_0'$ is set identifiable. By assumption, the set 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 set identification**, when the true parameter value is on the boundary of the identifiable set. Moreover, some information in the Yule-Walker equations of order 0 and 1 is redundant for the purpose of set identification.

If we assume $\Sigma_0 = \sigma_0^2 Id$, Corollary 2 and the results in paragraph *i*) imply that the parameters are set identifiable. For such a case, the proof of identifiability suggests a consistent estimation method. Let us denote $G_T(\sigma^2, B, C)$ a criterion based on moment restrictions (3.8) such that G_T tends to a nondegenerate limit G_∞ , say, when the sample size T tends to infinity. Due to the lack of point identifiability, the moment estimator defined by:

$$(\hat{\sigma}_T^2, \hat{B}_T, \hat{C}_T) = \arg \min_{\sigma^2, B, C} G_T(\sigma^2, B, C),$$

where the minimization is such that matrix $B'B$ is diagonal, converges for T large towards the set \mathcal{E}_0 , but not necessarily to the true value of the parameter (σ_0^2, B_0, C_0) . Corollary 2 suggests how to modify the optimization criterion to recover consistency. Let us recall that two symmetric matrices of the same dimension A_0 and A_1 are ordered $A_0 \succeq A_1$ if, and only if, their ranked eigenvalues $\lambda_1(A) \geq \dots \geq \lambda_K(A)$ are such that: $\lambda_j(A_0) \geq \lambda_j(A_1)$, for all $j = 1, \dots, K$. Thus, to get the consistency, the estimation criterion has to be penalized:

$$\begin{aligned} (\hat{\sigma}_T^2, \hat{B}_T, \hat{C}_T) &= \arg \min_{\sigma^2, B, C} \{G_T(\sigma^2, B, C) + \alpha_T [\sigma^2 + \lambda_1(BB') + \dots + \lambda_K(BB')]\} \\ &= \arg \min_{\sigma^2, B, C} \{G_T(\sigma^2, B, C) + \alpha_T [\sigma^2 + Tr(BB')]\}, \end{aligned}$$

where the minimization is such that matrix $B'B$ is diagonal and the positive weight α_T tends to 0 at an appropriate speed when T tends to infinity. The study of the properties of such an estimator is beyond the scope of this paper.

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. This model is a state space specification, with state vector $Z_t = (y_t', f_t')'$ following a constrained VAR(1) dynamics, and measurement equation $y_t = (Id_n, 0)Z_t$. 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. This identification strategy relies on the fact that the linear combinations of the components of y_t , which are immune to the unobserved factor f_t , are uncorrelated with y_{t-2}, y_{t-3}, \dots , conditional on y_{t-1} . In the light of this result involving second-order autocovariances, we expect that the nonlinear moment restrictions at order 1 in (3.10) 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 (3.9) and (3.11), the nonlinear moment restrictions are:

$$\Sigma_0 + B_0 B_0' + (C - C_0) \Gamma_0(0) (C - C_0)' + \Lambda_0 (C - C_0)' + (C - C_0) \Lambda_0' = \Sigma + B B', \quad (3.13)$$

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 (3.13) 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 point identified as well (see Proposition 5 in the next section for a more general result). Furthermore, the extremal property of matrix $\Sigma + B B'$ 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 set identifiable.

This lack of identification was expected in the light of the discussion of identification in Structural VAR (SVAR) models. The SVAR models are in general approximations of structural dynamic models, with constraints on parameters and different types of shocks, either structural shocks with economic interpretations and effects on all variables, or measurement errors. In our framework the structural shocks correspond to the innovations on the common factors, whereas the ε_t are representing the measurement errors. The first-order moment restrictions inherit the identification problems usually encountered with the Gaussian SVAR model. In the general case the parameters are not point identifiable from conditions (3.13). However as usual they may become first-order identifiable if additional restrictions on

parameters deduced from the structural models complete appropriately the conditions (3.13). This has to be considered for each SVAR case by case. Finally the possibility to almost identify at first order the parameters in case *ii*) by a “corner solution” is mathematically similar to the structural identification by sign restrictions proposed by e.g. Uhlig (2005), Cheri, Kehoe, McGrattan (2008), but has no structural interpretation.

Example 7: A conditionally Gaussian factor model with stochastic volatility in the factor.

Let us consider the following multivariate model:

$$y_t = \beta f_{1,t}^{1/2} \eta_t + \varepsilon_t, \quad (3.14)$$

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 $\varepsilon_t \sim IIN(0, \Sigma)$, $\eta_t \sim IIN(0, 1)$. The process $(f_{1,t})$ is an Autoregressive Gamma (ARG) Markov process with conditional Laplace transform:

$$E[\exp(uf_{1,t})|f_{1,t-1}] = \exp \left\{ \frac{\gamma c u f_{1,t-1}}{1 - cu} - \nu \log(1 - cu) \right\}, \quad (3.15)$$

corresponding to a noncentral gamma transition distribution, where $\nu > 0$ is the degree of freedom parameter, $c > 0$ is a scale parameter, and $\gamma > 0$ is such that γc is the first-order autocorrelation (see Section 2.1). 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 $c = 1/(1 + \gamma)$. Then, $f_{1,t} \sim \gamma(\nu)$.

A first idea might be to compute the unconditional variance of y_t , that is, the matrix:

$$V(y_t) = \beta V[f_{1,t}^{1/2} \eta_t] \beta' + V(\varepsilon_t) = \nu \beta \beta' + \Sigma. \quad (3.16)$$

However, since the matrix Σ has not been assumed diagonal, we cannot use these moment restrictions to identify parameters ν , β and Σ separately. Moreover, even if matrix Σ were diagonal, and the variance components $\nu \beta \beta'$ and Σ were identifiable, the degree of freedom ν and the loadings vector β would not be separately identifiable from (3.16).

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

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

Thus, for identification analysis we have to consider the system :

$$\frac{u'\Sigma u}{2} - \nu \log \left(1 - \frac{(u'\beta)^2}{2} \right) = \frac{u'\Sigma_0 u}{2} - \nu_0 \log \left(1 - \frac{(u'\beta_0)^2}{2} \right), \quad \forall u. \quad (3.17)$$

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

$$u'\Sigma u = u'\Sigma_0 u, \quad \forall u \quad \Leftrightarrow \quad \Sigma = \Sigma_0,$$

$$\nu \log \left(1 - \frac{(u'\beta)^2}{2} \right) = \nu_0 \log \left(1 - \frac{(u'\beta_0)^2}{2} \right), \quad \forall u.$$

By identifying the quadratic and quartic terms in argument u in the expansion of the two sides of the second subsystem in a neighbourhood of $u = 0$, we get $\nu = \nu_0, (u'\beta)^2 = (u'\beta_0)^2, \forall u$, which implies $\beta = \beta_0$. Therefore, the true values of the parameters ν_0, β_0 and Σ_0 are all identifiable.

The parameter identifiability is simpler in the nonlinear and non Gaussian model of Example 7 compared to Example 5. Indeed, the different laws of common factor $f_{1,t}$ and innovation ε_t in Example 7 imply that the nonlinear moment restrictions lead to a continuum of equations [see (3.17)], while in the linear Gaussian model of Example 5 this continuum of restrictions boil down to a finite number of equations [see (3.10)].

4 Higher-order nonlinear moment restrictions

In general the first-order nonlinear moment restrictions (3.1) are not sufficient to identify the entire set of parameters of the models. When the distribution of (f_t) is parametrically specified, the parameters characterizing the dynamics of (f_t) are not identifiable. When the distribution of (f_t) is let unspecified, we have identified neither the distribution of (f_t) , nor the parameters B, C, θ of the affine regression model.

We derive below second- and third-order nonlinear moment restrictions which involve higher-order lags in the endogenous variables and dynamic effects. We explain how the second-order nonlinear restrictions allow for identifying the parameters characterizing the dynamics of (f_t) in the parametric case. When the distribution of Markov process (f_t) is unspecified, the complete identification of the model requires (generically) the joint use of first-, second- and third-order nonlinear moment restrictions.

4.1 Second-order nonlinear moment restrictions

These restrictions are derived 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:

$$\psi_t(u, \theta, C) = a(u, x_t, \theta)'[Cy_{t-1} + d(x_t, \theta)] + b(u, x_t, \theta). \quad (4.1)$$

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

$$\begin{aligned} & E[\exp(u'y_t - \psi_t(u, \theta, C) + \tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u}, \theta, C)) | \underline{y_{t-2}}, \underline{f_t}, \underline{x_t}] \\ &= E\{E[\exp(u'y_t - \psi_t(u, \theta, C) + \tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u}, \theta, C)) | \underline{y_{t-1}}, \underline{f_t}, \underline{x_t}] | \underline{y_{t-2}}, \underline{f_t}, \underline{x_t}\} \\ &= E\{E[\exp(u'y_t - \psi_t(u, \theta, C)) | \underline{y_{t-1}}, \underline{f_t}, \underline{x_t}] \exp(\tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u}, \theta, C)) | \underline{y_{t-2}}, \underline{f_t}, \underline{x_t}\} \\ &= \exp[a(u, x_t, \theta)'Bf_t] E[\exp(\tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u}, \theta, C)) | \underline{y_{t-2}}, \underline{f_t}, \underline{x_t}] \\ &= \exp[a(u, x_t, \theta)'Bf_t + a(\tilde{u}, x_{t-1}, \theta)'Bf_{t-1}], \quad \forall u, \tilde{u}. \end{aligned} \quad (4.2)$$

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:

$$\begin{aligned} & E\left[\exp\{u'y_t - \psi_t(u, \theta, C) + \tilde{u}'y_{t-1} - \psi_{t-1}(\tilde{u}, \theta, C)\} | \underline{x_t}\right] \\ &= E[\exp\{a(u, x_t, \theta)'Bf_t + a(\tilde{u}, x_{t-1}, \theta)'Bf_{t-1}\} | \underline{x_t}], \quad \forall u, \tilde{u}. \end{aligned} \quad (4.3)$$

We get another continuum of moment restrictions, which can be used to complete identification. Since $a(0, x_t, \theta) = 0$ and $b(0, x_t, \theta) = 0$, for argument $\tilde{u} = 0$ the second-order nonlinear moment restrictions (4.3) boil down to the first-order restrictions (3.2).

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 $v = (v'_1, v'_2)' \in \mathbb{R}^K \times \mathbb{R}^{n-K}$, and from argument \tilde{u} to argument $\tilde{v} = (\tilde{v}'_1, \tilde{v}'_2)' \in \mathbb{R}^K \times \mathbb{R}^{n-K}$, such that $u(v, x_t, \theta, B)$ and $u(\tilde{v}, x_{t-1}, \theta, B)$ satisfy:

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

From equation (4.3) we get:

$$\begin{aligned} & E\left[\exp\{u(v, x_t, \theta, B)'y_t - \psi_t[u(v, x_t, \theta, B), \theta, C] \right. \\ & \quad \left. + u(\tilde{v}, x_{t-1}, \theta, B)'y_{t-1} - \psi_{t-1}[u(\tilde{v}, x_{t-1}, \theta, B), \theta, C]\} | \underline{x_t}\right] \\ &= E[\exp\{v'_1f_t + \tilde{v}'_1f_{t-1}\} | \underline{x_t}], \quad \forall v, \tilde{v}. \end{aligned} \quad (4.4)$$

We have the following (generic) identification property.

Proposition 3: *Under Assumptions A.1-A.4, the first- and second-order nonlinear moment restrictions [i.e. restrictions (4.4)] allow for identifying (generically):*

- i) *The complete model, if the conditional distribution of process (f_t) given (x_t) is specified parametrically.*
- ii) *The joint distribution of (f_t, f_{t-1}) conditional on \underline{x}_t , if the regression parameters are known and the (conditional) distribution of process (f_t) is left unspecified.*

4.2 Third-order nonlinear moment restrictions

The technique developed in Subsection 4.1 can be applied at any lag. The third-order nonlinear moment restrictions are derived by applying the Iterated Expectation Theorem to the joint Laplace transform of y_t, y_{t-1}, y_{t-2} given $\underline{y}_{t-3}, \underline{f}_t, \underline{x}_t$ (see Appendix 1). These moment restrictions relate the regression parameters and the joint distribution of f_t, f_{t-1}, f_{t-2} .

For expository purpose, let us focus on the case without observable exogenous covariates x_t . Then, from the third-order nonlinear moment restrictions we identify in particular the conditional transition of the unobserved component at horizon 1 : $g_1(f_t|f_{t-1}; B, C, \theta)$, and the conditional transition at horizon 2: $g_2(f_t|f_{t-2}; B, C, \theta)$, say, once the regression parameters B, C, θ are known. However, by the Markov Assumption A.3, we get the Kolmogorov relationship:

$$g_2(f_t|f_{t-2}; B, C, \theta) = \int g_1(f_t|f_{t-1}; B, C, \theta)g_1(f_{t-1}|f_{t-2}, B, C, \theta)df_{t-1}, \quad (4.5)$$

$$\forall f_t, f_{t-2}, B, C, \theta,$$

that is an infinite number of moment restrictions, which can be used to identify the true values of the regression parameters.

Proposition 4: *Under Assumptions A.1-A.4, the first-, second- and third-order nonlinear moment restrictions (generically) identify the regression parameters and the unspecified transition of process (f_t) .*

4.3 Examples

Example 5: A Gaussian model with common factor and contagion (continues).

The nonlinear moment restrictions at order 1, 2 and 3 are derived in Appendix 2.1. They are:

$$\begin{aligned}
\Gamma_0(0) + C\Gamma_0(0)C' - \Gamma_0(1)C' - C\Gamma_0(1)' &= BB' + \Sigma, \text{ for order 1,} \\
\Gamma_0(1) + C\Gamma_0(1)C' - \Gamma_0(2)C' - C\Gamma_0(0) &= B\Phi B', \text{ to be added for order 2,} \\
\Gamma_0(2) + C\Gamma_0(2)C' - \Gamma_0(3)C' - C\Gamma_0(1) &= B\Phi^2 B', \text{ to be added for order 3,} \quad (4.6)
\end{aligned}$$

where the autocovariance function $\Gamma_0(h) = Cov_0(y_t, y_{t-h})$ is identifiable. We explain in Appendix 2.2 that these nonlinear restrictions are well-chosen (parameter-dependent) linear combinations of a subset of the Yule-Walker equations for the VAR(1) process $(y_t', f_t')'$. Such linear combinations of Yule-Walker equations are such that they 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 that matrix $B'B$ is 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 3.4, Example 5, the parameters cannot be identified from moment restrictions at order 1 only. It is easily verified that the order condition for identification at order 2 is also not satisfied. In fact, the next proposition shows that the second-order moment restrictions do not identify the true model parameters.

Proposition 5: *In the Gaussian linear model with common factor and contagion, for any n and K the true parameter value is not identifiable from the second-order nonlinear moment restrictions.*

The proof of Proposition 5 in Appendix 2.3 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 . However, Proposition 5 shows that the order condition is not sufficient for identification. The proof of Proposition 5 also shows that the lack of identification applies both locally and globally.

Finally, the order condition for identification at order 3 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.

Example 7: 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:

$$\begin{aligned} E[\exp(u'y_t + \tilde{u}'y_{t-1})] &= E\{E[\exp(u'y_t + \tilde{u}'y_{t-1})|f_{1,t}]\} \\ &= \exp\left(\frac{1}{2}u'\Sigma u + \frac{1}{2}\tilde{u}'\Sigma\tilde{u}\right) E\left[\exp\left(\frac{(u'\beta)^2}{2}f_{1,t} + \frac{(\tilde{u}'\beta)^2}{2}f_{1,t-1}\right)\right], \quad \forall u, \tilde{u}. \end{aligned} \quad (4.7)$$

Parameters β_0 and Σ_0 are identified by the first-order nonlinear moment restrictions since the factor $f_{1,t}$ has a parametric distribution (see Section 3.5). 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 $u = (\sqrt{2v}/\beta_{0,1}, 0, \dots, 0)'$ and $\tilde{u} = (\sqrt{2\tilde{v}}/\beta_{0,1}, 0, \dots, 0)'$, where v, \tilde{v} are positive. We get:

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

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 $f_{1,t}$.

Example 8 : Multivariate Poisson model with common stochastic intensity

The model is defined by:

$$y_{k,t} \sim \mathcal{P}(\beta_k f_t), \quad k = 1, \dots, K,$$

where the count variables are independent conditional on f_t , and the factor dynamics is let unspecified. This model corresponds to a special case of the specification in Example 6, when the intercepts a_k and the contagion coefficients $c_{k,k'}$ are zero. Without loss of generality, we set $\beta_1 = 1$. Thus, one of the regression parameters is assumed known.

Let us first write the second-order nonlinear moment restrictions on $y_{1,t}, y_{1,t-1}$. We get:

$$E(\exp[\log(1 - v)y_{1,t} + \log(1 - \tilde{v})y_{1,t-1}]) = E[\exp(vf_t + \tilde{v}f_{t-1})]. \quad (4.8)$$

Thus the dynamics of the common intensity f_t is nonparametrically identifiable.

Let us now apply the first-order nonlinear moment restrictions on $y_{k,t}$, for $k \geq 2$. We get:

$$\begin{aligned} E[\exp(uy_{k,t})] &= E[\exp\{-\beta_k f_t(1 - \exp u)\}] \\ &= \psi[\beta_k(\exp u - 1)], \forall u, \end{aligned} \quad (4.9)$$

where ψ denotes the Laplace transform of the stationary distribution of f_t . Function ψ is identifiable by (4.8), and then parameters $\beta_k, k = 2, \dots, K$ are identifiable by (4.9).

Alternatively the identifiability of β parameters might have been obtained by nonlinear cross-differencing. Indeed we deduce from (4.9) :

$$E\{\exp[\log(1 - v/\beta_k)y_{k,t}]\} = E[\exp(-vf_t)], \quad \forall k.$$

Thus we get for any pair (k, l) , with $k \neq l$, the moments in cross-difference :

$$E\{\exp[\log(1 - v/\beta_k)y_{k,t}]\} = E\{\exp[\log(1 - v/\beta_l)y_{l,t}]\}, \quad \forall k \neq l. \quad (4.10)$$

4.4 Semi-nonparametric identification of the linear model with contagion and common factor

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

$$y_t = Cy_{t-1} + Bf_t + \varepsilon_t, \quad (4.11)$$

where (f_t) is a Markov process, and (ε_t) is a strong white noise process independent of (f_t) . Contrary to Example 5, 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 (4.11) 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 number K of factors is strictly smaller than the number of endogenous variables n , and matrix B_0 has full column-rank.

i) Identification of the number of factors, factor loadings and contagion matrix

The nonlinear moment restrictions at order 2 are:

$$\begin{aligned} E_0 [\exp (u'(y_t - C_0y_{t-1}) + \tilde{u}'(y_{t-1} - C_0y_{t-2}))] &= E_0 [\exp (u'\varepsilon_t + \tilde{u}'\varepsilon_{t-1} + u'B_0f_t + \tilde{u}'B_0f_{t-1})] \\ &= \exp (\psi_\varepsilon^0(u) + \psi_\varepsilon^0(\tilde{u}) + \psi_f^0(B_0'u, B_0'\tilde{u})), \quad \forall u, \tilde{u}, \end{aligned}$$

where E_0 denotes the expectation w.r.t. the DGP. Then, the function h_0 defined by:

$$h_0(u, \tilde{u}, C) = \log E_0 \left[\exp \left(u'(y_t - Cy_{t-1}) + \tilde{u}'(y_{t-1} - Cy_{t-2}) \right) \right], \quad (4.12)$$

is such that:

$$h_0(u, \tilde{u}, C_0) = \psi_\varepsilon^0(u) + \psi_\varepsilon^0(\tilde{u}) + \psi_f^0(B_0'u, B_0'\tilde{u}), \quad \forall u, \tilde{u}. \quad (4.13)$$

The quantity $h_0(u, \tilde{u}, C)$ depends on the DGP through the expectation E_0 . Since this expectation involves the observable variables only, the function $h_0 : (u, \tilde{u}, C) \rightarrow h_0(u, \tilde{u}, C)$ is identifiable.

The cross-terms of function $h_0(\cdot, \cdot, C_0)$ with respect to arguments u and \tilde{u} , which capture the dependence between variables $y_t - C_0y_{t-1}$ and $y_{t-1} - C_0y_{t-2}$, are due to the serial persistence of the unobservable factor process. These cross-terms have specific patterns. Indeed, from equation (4.13), function $h_0(\cdot, \cdot, C_0)$ is additively separable w.r.t. arguments u, \tilde{u} on the linear subspace defined by the condition $B_0'u = 0$ (that is, the subspace in which the argument u is orthogonal to the column space of matrix B_0) and on the linear subspace defined by $B_0'\tilde{u} = 0$. The arguments u and \tilde{u} on such linear subspaces define linear combinations of the components of y_t that are immune to the common factor f_t , and thus variables $u'(y_t - C_0y_{t-1})$ and $\tilde{u}'(y_{t-1} - C_0y_{t-2})$ are independent. We exploit these cross-terms patterns for identification. The matrix of second-order mixed partial derivatives of function $h_0(\cdot, \cdot, C_0)$ is:

$$\frac{\partial^2 h_0(u, \tilde{u}, C_0)}{\partial u \partial \tilde{u}'} = B_0 \frac{\partial^2 \psi_f^0(B_0'u, B_0'\tilde{u})}{\partial v \partial w'} B_0', \quad \forall u, \tilde{u}. \quad (4.14)$$

This matrix has reduced rank smaller or equal to K , for all u, \tilde{u} . Moreover, if (u, \tilde{u}) is such that $\partial^2 \psi_f^0(B_0'u, B_0'\tilde{u}) / \partial v \partial w'$ is full-rank, the null space of matrix $\partial h_0^2(u, \tilde{u}, C_0) / \partial u \partial \tilde{u}'$ is equal to the orthogonal complement of the column space of matrix B_0 .

Assumption A.5: i) The matrix $\partial^2 h_0(u, \tilde{u}, C) / \partial u \partial \tilde{u}'$ has reduced rank for all (u, \tilde{u}) if, and only if, $C = C_0$. ii) 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, there exists (v^*, w^*) such that matrix $\partial^2 \psi_f^0(v^*, w^*) / \partial v \partial w'$ has full rank.

In Assumption A.5 ii) we explicitly exclude the case of Gaussian factor and innovation processes. Indeed, in such a case, the log Laplace transform functions $h_0(\cdot, \cdot, C)$ and ψ_0^f are quadratic, and their the second-order partial derivative matrices are constant. Then, from (4.14) we get a system of n^2 equations independent of the arguments u, \tilde{u} , that correspond to the additional moment restrictions for order 2 given in (4.6). We have verified in Section 4.3 that these restrictions (even when considered

together with the first-order moment restrictions) are not sufficient for identification of the parameters. Moreover, another necessary condition for the full-rank property in Assumption A.5 ii) is that process (f_t) is serially dependent, which explains why we explicitly exclude the case of a white noise factor process.

Proposition 6: *Under Assumption A.5, the dimension K of the factor space, the contagion matrix C_0 , and the column space of matrix B_0 in model (4.11) are identifiable.*

Proof: The matrix $\partial^2 h_0(u, \tilde{u}, C)/\partial u \partial \tilde{u}'$ is identifiable, for any given u , \tilde{u} and C . Thus, matrix C_0 is identifiable from the reduced-rank property in Assumption A.5 i). From equation (4.14), Assumption A.5 ii), and the full-rank property of B_0 , it follows that the largest rank among matrices $\partial^2 h_0(u, \tilde{u}, C_0)/\partial u \partial \tilde{u}'$, with (u, \tilde{u}) varying, is equal to K . Thus, the factor dimension is identifiable. Finally, the range $\mathcal{R}(B_0)$ of matrix B_0 is identifiable, because:

$$\mathcal{R}(B_0) = \bigcup_{(u, \tilde{u})} \mathcal{R} \left(\frac{\partial^2 h_0(u, \tilde{u}, C_0)}{\partial u \partial \tilde{u}'} \right), \quad (4.15)$$

and

$$\mathcal{R}(B_0)^\perp = \bigcap_{(u, \tilde{u})} \mathcal{N} \left(\frac{\partial^2 h_0(u, \tilde{u}, C_0)}{\partial u \partial \tilde{u}'} \right), \quad (4.16)$$

where \perp denotes the orthogonal complement of a vector space, and $\mathcal{N}(\cdot)$ and $\mathcal{R}(\cdot)$ denote the null space and the range (i.e., the column space), respectively, of a matrix.⁴

Let us first show equation (4.15). The inclusion of the set in the RHS in the space in the LHS is a consequence of equation (4.14). To show the inclusion in the opposite direction, we use that matrix B_0 has row rank K and thus there exist arguments u^* and \tilde{u}^* such that $B_0' u^* = v^*$ and $B_0' \tilde{u}^* = w^*$, where v^* and w^* are as in Assumption A.5 ii). Then from (4.14) we get:

$$\frac{\partial^2 h_0(u^*, \tilde{u}^*, C_0)}{\partial u \partial \tilde{u}'} = B_0 \frac{\partial^2 \psi_f^0(v^*, w^*)}{\partial v \partial w'} B_0',$$

and matrix $\frac{\partial^2 \psi_f^0(v^*, w^*)}{\partial v \partial w'} B_0'$ has column rank K . It follows that $\mathcal{R} \left(\frac{\partial^2 h_0(u^*, \tilde{u}^*, C_0)}{\partial u \partial \tilde{u}'} \right) = \mathcal{R}(B_0)$, which implies the inclusion of the space in the LHS of (4.15) in the set in the RHS.

Let us now show equation (4.16). The inclusion of the space in the LHS in the space in the RHS is a consequence of equation (4.14). To show the inclusion in the opposite direction, let us consider

⁴The union of two linear spaces is not necessarily a linear space. However, equation (4.15) shows that the set $\bigcup_{(u, \tilde{u})} \mathcal{R} \left(\frac{\partial^2 h_0(u, \tilde{u}, C_0)}{\partial u \partial \tilde{u}'} \right)$ is a linear space.

a n -dimensional vector ξ , which belongs to the null space of matrix $\partial^2 h_0(u, \tilde{u}, C_0)/\partial u \partial \tilde{u}'$, for any (u, \tilde{u}) . Then, from equation (4.14) written for arguments u^* and \tilde{u}^* defined above, it follows:

$$B_0 \frac{\partial^2 \psi_f^0(v^*, w^*)}{\partial v \partial w'} B_0' \xi = 0.$$

Since matrices B_0 and $\partial^2 \psi_f^0(v^*, w^*)/\partial v \partial w'$ are full-rank, it follows that $B_0' \xi = 0$, that is $\xi \in \mathcal{R}(B_0)^\perp$. Q.E.D.

ii) 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) . 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:

$$E[\exp(u' f_t) | f_{t-1}] = \exp[a(u)' f_{t-1} + b(u)], \quad (4.17)$$

where functions a and b are left 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:

$$\tilde{y}_t = f_t + \tilde{\varepsilon}_t, \quad (4.18)$$

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 3.1 we show that:

$$\tau_0(u, v; h) = \exp \left(\varphi_f^0[a_0^{oh}(u) + v] - \varphi_f^0[a_0^{oh}(u)] - \varphi_f^0(v) \right), \quad \forall u, v, \forall h \geq 1, \quad (4.19)$$

where a_0^{oh} 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:

$$\tau_0(u, v; 2) = \tau_0[a_0(u), v; 1], \quad \forall u, v. \quad (4.20)$$

Proposition 7: *Consider the semi-nonparametric model with common factor and contagion defined by equations (4.11) and (4.17). Suppose that Assumption A.5 holds and function $a_0 : \mathcal{D}_0 \rightarrow \mathcal{R}_0$ is one-to-one, where domain \mathcal{D}_0 is a neighbourhood of 0 in \mathbb{C}^K , and $\mathcal{R}_0 \subset \mathcal{D}_0$. Then: i) function $a_0(u)$ is identifiable for $u \in \mathcal{D}_0$, ii) function $\varphi_f^0(u)$ is identifiable for $u \in \mathcal{R}_0$, and iii) function $\psi_\varepsilon(u)$ is 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 3.2 we show that function $\tau_0(\cdot, \cdot; 1)$ is such that:

$$\tau_0(u_1, v; 1) = \tau_0(u_2, v; 1), \quad \forall v \quad \Rightarrow \quad u_1 = u_2. \quad (4.21)$$

Then, for any given $u \in \mathcal{D}_0$, the value $a_0(u)$ is identifiable from (4.20).

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, 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 zero, we have, $\partial \varphi_f^0(0)/\partial u = E_0(f_t) = 0$. It follows:

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

Since function a_0 is one-to-one and identifiable by part i), the above equation implies that:

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

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 . We have:

$$E_0[\exp(u'(y_t - C_0 y_{t-1}))] = \exp(\varphi_f^0(B'_0 u) + \psi_\varepsilon^0(u)).$$

The LHS is identifiable, and $\varphi_f^0(B'_0 u)$ is identifiable if $B'_0 u \in \mathcal{R}_0$ from part ii). The conclusion follows. Q.E.D.

5 Concluding remarks

The aim of this paper is to highlight how moment restrictions based on appropriate Laplace transforms can be used to identify nonlinear panel data or time series models with unobservable dynamic effects. The identification approach requires exponential affine regression models in lagged endogenous variables and unobservable dynamic effects, but can be applied both to a parametric specification of the dynamics of the unobserved component as well as when the distribution of the unobservable process is left unspecified. We have provided several examples to show the variety of models for which this identification approach can be used.

In nonlinear models with unobservable dynamic components, it is difficult to estimate the model by applying the maximum likelihood approach (resp. an empirical maximum likelihood approach), when the distribution of the unobserved component is specified parametrically (resp. is left unspecified). Indeed, the likelihood function (resp. the empirical likelihood function) involves multidimensional integrals of a large dimension increasing with the number T of observations, since the unobserved factor path has to be integrated out. 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. The continuum of nonlinear conditional moment restrictions at first-, second-, third- ... orders can be the basis of moment-based estimation methods [see Carrasco, Florens (2000), Singleton (2001), Jiang, Knight (2002), Chacko, Viceira (2003), Carrasco et al. (2007)], avoiding the numerical computation of such multidimensional integrals. The analysis of the associated moment methods depends on the type of data (panel or time series), on the asymptotics (n and/or T tending to infinity), on the assumptions on the distribution of the unobserved component (parametric or nonparametric), on selected moment restrictions, etc. Such an analysis is clearly out of the scope of this paper and is left for future research.

⁵See however Darolles, Gagliardini, Gouriéroux (2014) for an application of such a moment method for the analysis of liquidation risk (Example 6) with a parametric specification of the distribution of the latent common factor.

References

- Al-Osh, M., and A., Alzaid (1987): "First-Order Integer Valued Autoregressive INAR(1) Process", *Journal of Time Series Analysis*, 8, 261-275.
- Alessi, L., Barigozzi, M., and M., Carpasso (2008): "A Review of Nonfundamentalness and Identification in Structural VAR Models", DP European Central Bank, 922.
- 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., Elias (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach", *Quarterly Journal of Economics*, 120, 387-422.
- Bockenholt, U. (1999): "Mixed INAR(1) Poisson Regression Models Analyzing Heterogeneity and Serial Dependencies in Longitudinal Count Data", *Journal of Econometrics*, 89, 317-338.
- Blundell, R., Griffith, R., and F., Windmeijer (2002): "Individual Effects and Dynamics in Count Data Model", *Journal of Econometrics*, 108, 113-131.
- Brannas, K. (1994): "Estimation and Testing in Integer Valued AR(1) Models", DP 335, University of Umea.
- 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.
- 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.

Chari, V., Kehoe, P., and E., McGrattan (2008): "Are Structural VAR's with Long Run Restrictions Useful in Developing Business Cycle Theory", *Journal of Monetary Economics*, 55, 1337-1352.

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., Gouriéroux (2014): "Contagion Analysis in the Banking Sector", working paper.

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

Darolles, S., Gouriéroux, 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.

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", *Kybernetika*, 33, 583-606.

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., Gouriéroux (2014): "Wishart Stochastic Volatility Model", DP CREST.

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 (2004): "Heterogenous INAR(1) Model with Application to Car Insurance", *Insurance : Mathematics and Economics*, 34, 177-192.

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.

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., Wadhvani (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.

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.

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

Stock, J., and M., Watson (1998): "Diffusion Indexes", NBER Working Paper, 6701.

Uhlig, H. (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure", *Journal of Monetary Economics*, 52, 381-419.

D. Wilhelm (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: Third-order nonlinear moment restrictions

These 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 “correction terms”. By using equation (4.2) in Section 4.1 and applying the Iterated Expectation Theorem, we get:

$$\begin{aligned}
& E[\exp(u'_0 y_t - \psi_t(u_0, \theta, C) + u'_1 y_{t-1} - \psi_{t-1}(u_1, \theta, C) + u'_2 y_{t-2} - \psi_{t-2}(u_2, \theta, C)) | \underline{y_{t-3}}, \underline{f_t}, \underline{x_t}] \\
= & E \left\{ E[\exp(u'_0 y_t - \psi_t(u_0, \theta, C) + u'_1 y_{t-1} - \psi_{t-1}(u_1, \theta, C)) | \underline{y_{t-2}}, \underline{f_t}, \underline{x_t}] \right. \\
& \left. \exp(u'_2 y_{t-2} - \psi_{t-2}(u_2, \theta, C)) | \underline{y_{t-3}}, \underline{f_t}, \underline{x_t} \right\} \\
= & \exp[a(u_0, x_t, \theta)' B f_t + a(u_1, x_{t-1}, \theta)' B f_{t-1}] E[\exp(u'_2 y_{t-2} - \psi_{t-2}(u_2, \theta, C)) | \underline{y_{t-3}}, \underline{f_t}, \underline{x_t}] \\
= & \exp[a(u_0, x_t, \theta)' B f_t + a(u_1, x_{t-1}, \theta)' B f_{t-1} + a(u_2, x_{t-2}, \theta)' B f_{t-2}], \quad \forall u_0, u_1, u_2.
\end{aligned}$$

Then, as in Section 4.1, we apply the conditional expectation given $\underline{x_t}$ only on both sides to get:

$$\begin{aligned}
& E[\exp(u'_0 y_t - \psi_t(u_0, \theta, C) + u'_1 y_{t-1} - \psi_{t-1}(u_1, \theta, C) + u'_2 y_{t-2} - \psi_{t-2}(u_2, \theta, C)) | \underline{x_t}] \\
= & E \left[\exp[a(u_0, x_t, \theta)' B f_t + a(u_1, x_{t-1}, \theta)' B f_{t-1} + a(u_2, x_{t-2}, \theta)' B f_{t-2}] | \underline{x_t} \right]. \quad (\text{a.1})
\end{aligned}$$

Finally, under Assumption A.4 we apply the change of variables from u_0, u_1, u_2 to v_0, v_1, v_2 such that:

$$\begin{aligned}
a(u_0, x_t, \theta)' B &= v'_{0,1}, \\
a(u_1, x_{t-1}, \theta)' B &= v'_{1,1}, \\
a(u_2, x_{t-2}, \theta)' B &= v'_{2,1},
\end{aligned}$$

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.

Appendix 2: Identification in the Gaussian linear model with common factor and contagion

In this Appendix we study the identification of model (3.6) in Example 5:

$$y_t = Bf_t + Cy_{t-1} + \Sigma^{1/2}\varepsilon_t, \quad (\text{a.2})$$

$$f_t = \Phi f_{t-1} + (Id - \Phi\Phi')^{1/2}\eta_t, \quad (\text{a.3})$$

where the shocks $\varepsilon_t \sim IIN(0, Id)$ and $\eta_t \sim N(0, Id)$ are mutually independent Gaussian white noise processes.

A.2.1 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:

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

$\forall u, \tilde{u}$. 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 restrictions are:

$$\Gamma_0(0) + C\Gamma_0(0)C' - \Gamma_0(1)C' - C\Gamma_0(1)' = BB' + \Sigma, \quad (\text{a.4})$$

$$[\text{corresponding to } V_0(y_t - Cy_{t-1}) = V_0(Bf_t + \Sigma^{1/2}\varepsilon_t)]$$

and the additional restrictions at second-order are:

$$\Gamma_0(1) + C\Gamma_0(1)C' - \Gamma_0(2)C' - C\Gamma_0(0) = B\Phi B', \quad (\text{a.5})$$

$$[\text{corresponding to } Cov_0(y_t - Cy_{t-1}, y_{t-1} - Cy_{t-2}) = Cov_0(Bf_t + \Sigma^{1/2}\varepsilon_t, Bf_{t-1} + \Sigma^{1/2}\varepsilon_{t-1})]$$

where $\Gamma_0(h) = Cov_0(y_t, y_{t-h})$.

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:

$$\Gamma_0(2) + C\Gamma_0(2)C' - \Gamma_0(3)C' - C\Gamma_0(1) = B\Phi^2 B'. \quad (\text{a.6})$$

A.2.2 Link with the Yule-Walker equations

In this section we show that the nonlinear moment restrictions (a.4)-(a.6) are well-chosen (parameter-dependent) linear combinations of the Yule-Walker equations for the structural VAR model (a.2)-(a.3). The Yule-Walker equations involving the covariance function $\Gamma_0(h) = Cov_0(y_t, y_{t-h})$, for $h = 0, 1, 2, \dots$ of the observable process are derived by multiplying equation (a.2) by $y'_t, y'_{t-1}, y'_{t-2}, \dots$, respectively, and computing the expectation on both sides. We get:

$$\Gamma_0(0) = B\tilde{\Gamma}_0(0) + C\Gamma_0(1)' + \Sigma, \quad (\text{a.7})$$

$$\Gamma_0(1) = B\tilde{\Gamma}_0(1) + C\Gamma_0(0), \quad (\text{a.8})$$

$$\Gamma_0(h) = B\tilde{\Gamma}_0(h) + C\Gamma_0(h-1), \quad h \geq 2, \quad (\text{a.9})$$

where $\tilde{\Gamma}_0(h) = Cov_0(f_t, y_{t-h})$ is the cross-covariance function of processes (f_t) and (y_t) . These covariances are not nonparametrically identifiable from the data. However, some linear combinations of the Yule-Walker equations (a.7)-(a.9) do not involve the unidentifiable covariances $\tilde{\Gamma}_0(h)$. To see this, let us multiply equation (a.2) by f'_{t+h} , for $h = 0, 1, 2, \dots$ and compute the expectation, to get:

$$\tilde{\Gamma}_0(h)' = B(\Phi')^h + C\tilde{\Gamma}_0(h+1)', \quad h = 0, 1, 2, \dots \quad (\text{a.10})$$

Then, by considering equation (a.7), subtracting equation (a.8) right-multiplied by C' , and using equation (a.10) with $h = 0$, we get:

$$\Gamma_0(0) - \Gamma_0(1)C' = C\Gamma_0(1)' - C\Gamma_0(0)C' + BB' + \Sigma,$$

i.e., the first-order nonlinear moment restriction (a.4). Similarly, by considering equation (a.8), subtracting equation (a.9) for $h = 2$ right-multiplied by C' , and using equation (a.10) with $h = 1$, we get:

$$\Gamma_0(1) - \Gamma_0(2)C' = C\Gamma_0(0) - C\Gamma_0(1)C' + B\Phi B',$$

i.e., the second-order nonlinear moment restriction (a.5). The third-order nonlinear moment restrictions are obtained in an analogous manner as a linear combination of the equations in (a.9) for $h = 2$ and $h = 3$.

A.2.3 Lack of identification from the second-order nonlinear moment restrictions (proof of Proposition 5)

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 [see (a.4)-(a.5)]:

$$\Gamma_0(0) + C\Gamma_0(0)C' - \Gamma_0(1)C' - C\Gamma_0(1)' = \Sigma + BB', \quad (\text{a.11})$$

$$\Gamma_0(1) + C\Gamma_0(1)C' - \Gamma_0(2)C' - C\Gamma_0(0) = B\Phi B', \quad (\text{a.12})$$

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 that the existence of solutions (B, C, Σ, Φ) different from the true parameter value $(B_0, C_0, \Sigma_0, \Phi_0)$ and with $\Sigma = \Sigma_0$.

Let us first write the autocovariance matrices $\Gamma_0(1)$ and $\Gamma_0(2)$ in terms of the true model parameters. By replacing equation (a.3) into equation (a.2), 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:

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

We have:

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

and:

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

For the true parameter values, we deduce:

$$\Gamma_0(1) = C_0\Gamma_0(0) + B_0\Phi_0\Delta_0(0), \quad (\text{a.13})$$

$$\Gamma_0(2) = C_0^2\Gamma_0(0) + (C_0B_0\Phi_0 + B_0\Phi_0^2)\Delta_0(0), \quad (\text{a.14})$$

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

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.11)-(a.12), these equations can be written as:

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

By using formulas (a.13)-(a.14), we get

$$\begin{aligned}\Sigma_0 + B_0B_0' - (C - C_0)\Delta_0(0)'\Phi_0' B_0' - B_0\Phi_0\Delta_0(0)(C - C_0)' + (C - C_0)\Gamma_0(0)(C - C_0)' \\ = \Sigma + BB', \\ B_0\Phi_0B_0' - (C - C_0)[\Gamma_0(0) - C_0\Gamma_0(0)C_0' - B_0\Phi_0\Delta_0(0)C_0'] - B_0\Phi_0^2\Delta_0(0)(C - C_0)' \\ + (C - C_0)\Gamma_0(1)(C - C_0)' = B\Phi B'.\end{aligned}\tag{a.15}$$

Let us rewrite the term in the square brackets in (a.15). By computing the variance of both sides of equation (a.2) written for the true parameter values, 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 (a.3) 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 are:

$$\begin{aligned}\Sigma_0 + B_0B_0' - (C - C_0)\Delta_0(0)'\Phi_0' B_0' - B_0\Phi_0\Delta_0(0)(C - C_0)' + (C - C_0)\Gamma_0(0)(C - C_0)' \\ = \Sigma + BB',\end{aligned}\tag{a.16}$$

$$\begin{aligned}B_0\Phi_0B_0' - (C - C_0)[B_0B_0' + \Sigma_0 + C_0\Delta_0(0)'\Phi_0' B_0'] - B_0\Phi_0^2\Delta_0(0)(C - C_0)' \\ + (C - C_0)\Gamma_0(1)(C - C_0)' = B\Phi B'.\end{aligned}\tag{a.17}$$

Let us now consider solutions (B, C, Σ, Φ) of the matrix equations system (a.16)-(a.17) such that:

$$C = C_0 + B_0AB_0'\Sigma_0^{-1},\tag{a.18}$$

$$\Sigma = \Sigma_0,\tag{a.19}$$

where A is a (K, K) matrix. By replacing (a.18)-(a.19) into (a.16), we get:

$$B_0 [Id - AB'_0 \Sigma_0^{-1} \Delta_0(0)' \Phi'_0 - \Phi_0 \Delta_0(0) \Sigma_0^{-1} B_0 A' + AB'_0 \Sigma_0^{-1} \Gamma_0(0) \Sigma_0^{-1} B_0 A'] B_0 = BB'. \quad (\text{a.20})$$

The symmetric matrix:

$$\begin{aligned} R_0 &= Id - AB'_0 \Sigma_0^{-1} \Delta_0(0)' \Phi'_0 - \Phi_0 \Delta_0(0) \Sigma_0^{-1} B_0 A' + AB'_0 \Sigma_0^{-1} \Gamma_0(0) \Sigma_0^{-1} B_0 A' \\ &= Id - \Phi'_0 \Phi_0 + V_0 [\Phi_0 f_t - AB'_0 \Sigma_0^{-1} y_t], \end{aligned}$$

is positive definite. Let us define:

$$B = B_0 R_0^{1/2} Q_0, \quad (\text{a.21})$$

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.21) solves equation (a.20) and satisfies the restriction: $B'B$ is diagonal.

Finally, let us replace (a.18), (a.19), (a.21) into equation (a.17). We get:

$$\begin{aligned} B_0 [\Phi_0 - AB'_0 \Sigma_0^{-1} B_0 - A - AB'_0 \Sigma_0^{-1} C_0 \Delta_0(0)' \Phi'_0 - \Phi_0^2 \Delta_0(0) \Sigma_0^{-1} B_0 A' \\ + AB'_0 \Sigma_0^{-1} \Gamma_0(1) \Sigma_0^{-1} B_0 A'] B'_0 = B_0 R_0^{1/2} Q_0 \Phi Q'_0 R_0^{1/2} B'_0, \end{aligned}$$

which is satisfied if:

$$\begin{aligned} \Phi &= Q'_0 R_0^{-1/2} [\Phi_0 - AB'_0 \Sigma_0^{-1} B_0 - A - AB'_0 \Sigma_0^{-1} C_0 \Delta_0(0)' \Phi'_0 - \Phi_0^2 \Delta_0(0) \Sigma_0^{-1} B_0 A' \\ &\quad + AB'_0 \Sigma_0^{-1} \Gamma_0(1) \Sigma_0^{-1} B_0 A'] R_0^{-1/2} Q_0. \end{aligned} \quad (\text{a.22})$$

Thus, (B, C, Σ, Φ) defined in (a.18), (a.19), (a.21), (a.22) 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 3: Semi-nonparametric identification in the linear model with common factor and contagion

In this Appendix we show equation (4.19) and property (4.21) of Section 4.4, which are used to prove Proposition 7 on the identification of function a_0 in the semi-nonparametric linear model with contagion and CaR common factor defined by equations (4.11) and (4.17).

A.3.1 Proof of equation (4.19)

By equation (4.18), 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:

$$E_0[\exp(u' \tilde{y}_t + v' \tilde{y}_{t-h})] = \exp(\psi_{\tilde{\varepsilon}}(u) + \psi_{\tilde{\varepsilon}}(v)) E_0[\exp(u' f_t + v' f_{t-h})],$$

where $\psi_{\tilde{\varepsilon}}$ is the log-Laplace transform of $\tilde{\varepsilon}_t$. It follows:

$$\tau_0(u, v, h) = \frac{E_0[\exp(u' \tilde{y}_t + v' \tilde{y}_{t-h})]}{E_0[\exp(u' \tilde{y}_t)] E_0[\exp(v' \tilde{y}_{t-h})]} = \frac{E_0[\exp(u' f_t + v' f_{t-h})]}{E_0[\exp(u' f_t)] E_0[\exp(v' f_{t-h})]}. \quad (\text{a.23})$$

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

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

By using the Law of Iterated Expectation, we can show:

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

for any $h \geq 1$. Then, we have:

$$\begin{aligned} E_0[\exp(u' f_t + v' f_{t-h})] &= \exp\left(\varphi_f^0(u) - \varphi_f^0[a_0^{\circ h}(u)]\right) E_0\left[\exp\left([a_0^{\circ h}(u) + v]' f_{t-h}\right)\right] \\ &= \exp\left(\varphi_f^0[a_0^{\circ h}(u) + v] + \varphi_f^0(u) - \varphi_f^0[a_0^{\circ h}(u)]\right). \end{aligned} \quad (\text{a.24})$$

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.23) and (a.24) we get equation (4.19).

A.3.2 Proof of property (4.21)

Function $\tau_0(\cdot, \cdot; 1)$ is defined by $\tau_0(u, v; 1) = \exp\left(\varphi_f^0[a_0(u) + v] - \varphi_f^0[a_0(u)] - \varphi_f^0(v)\right)$. Thus, property (4.21) is equivalent to:

$$\varphi_f^0[a_0(u_1) + v] - \varphi_f^0[a_0(u_1)] = \varphi_f^0[a_0(u_2) + v] - \varphi_f^0[a_0(u_2)], \quad \forall v \quad \Rightarrow \quad u_1 = u_2. \quad (\text{a.25})$$

To prove property (a.25), 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:

$$\frac{\partial \varphi_f^0}{\partial u}[a_0(u_1)] = \frac{\partial \varphi_f^0}{\partial u}[a_0(u_2)].$$

Then, we get:

$$\begin{aligned} 0 &= \frac{\partial \varphi_f^0}{\partial u}[a_0(u_2)] - \frac{\partial \varphi_f^0}{\partial u}[a_0(u_1)] \\ &= \int_0^1 \frac{\partial^2 \varphi_f^0[a_0(u_1) + t(a_0(u_2) - a_0(u_1))]}{\partial u \partial u'} dt (a_0(u_2) - a_0(u_1)). \end{aligned} \quad (\text{a.26})$$

Now, function φ_f^0 is strictly convex. Indeed, we have:

$$\frac{\partial^2 \varphi_f^0(u)}{\partial u \partial u'} = V_0^u(f_t), \quad (\text{a.27})$$

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:

$$dP_0^u/dP_0 = \frac{\exp(u' f_t)}{E_0[\exp(u' f_t)]}.$$

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.26) 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 $u_1 = u_2$.