UNIVERSITÀ DELLA SVIZZERA ITALIANA

Faculty of Economics - Institute of Finance

and

SWISS FINANCE INSTITUTE

Latent Factor Models for Large and Mixed-frequency Data in Finance and Macroeconomics

A thesis submitted by

MIRCO RUBIN

for the degree of

DOCTOR IN ECONOMICS

March 2016

Members of the jury:

Prof. Fabio Trojani, Università della Svizzera Italiana, President of the jury

Prof. Patrick Gagliardini, Università della Svizzera Italiana, Advisor

Prof. Monica Billio, Università Ca' Foscari di Venezia, External Member

Prof. Eric Ghysels, University of North Carolina - Chapel Hill, External Member

Prof. Christian Gouriéroux, CREST e University of Toronto, External Member

Contents

Chapter 1

Introduction

In the last century, the increasing amount of quantitative data available due to the improvements of information technology has been accompanied by the development of dimensionality reduction techniques allowing the researchers in all disciplines to easily interpret the data. Latent - or unobserved - factor models have been among the earliest tools proposed by the statistical literature since the beginning of the twentieth century to summarize the vast amount of quantitative information, and capture its relevant dependence structure. In this type of models, the repeated observations over time of a numerical variable for different units of measurement - or individuals - are assumed to be functions of (a) few unobservable factors common to all the units in the panel examined, and (b) observational errors with negligible dependence to the factors. Undoubtedly, linear factor models are the most widely used and studied among the general class of unobserved factor models. For this type of models, individual observations are a linear combination of a small number of common factors and idiosyncratic errors. This statistical tool has been used to examine a wide variety of phenomena in economics and finance, as it can be exploited to efficiently describe the common features of multiple observable variables, and it often involves limited computational complexity even if applied on large panels of data.

Among the first and most prominent applications of factor models in economics and finance are empirical models of macroeconomic fluctuations and models of financial asset pricing. In the latter case, the stochastic return of an asset is a linear combination of few systematic risk factors, plus an asset-specific random shock. This idea is at the core of the equilibrium arbitrage pricing theory explaining returns of assets in large economies initiated by Ross (1976). The typical dataset used in empirical finance consists of a large number of observations, usually returns, for many different assets, i.e. they feature a large cross-sectional dimension. In these settings, the natural estimators of the common factors for a linear factor model consist of computing cross-sectional averages of the observations at each date, mainly using extensions of the principal component methods as in Chamberlain and Rothschild (1983), Connor and Korajczyk (1986), Connor and Korajczyk (1988), Connor and Korajczyk (1993), Bai and Ng (2002), and Bai (2003). Also Macroeconomics has contributed - both from the theoretical and empirical sides - to the development of factor models for large number of economic indicators, especially with the seminal contributions of Forni and Reichlin (1996), Forni and Reichlin (1998), Stock and Watson (1998), Forni, Hallin, Lippi, and Reichlin (2000), Stock and Watson (2002a). The main findings in this strand of literature is that business cycle fluctuations are determined by few factors driving a large cross-section of macroeconomic time series. These factors are particularly useful for the construction of leading and coincident indicators of the economic activity, or when used in forecasting, nowcasting, and other policy exercises.

This thesis considers new latent factor models, and their estimation methodologies, suitable for settings relatively unexplored in the econometric literature as (i) a nonlinear model for the joint dynamics of a large cross-sectional distribution of asset returns, and the persistence of the ranks of the individuals inside it; (ii) approximate linear latent factor models for large panels of mixed-frequency data; and (iii) small scale state space models featuring multiple time series with stochastic volatility and observed at different frequencies.

Regarding point (i), the literature on asset allocation has found convenient to use the linear models developed in asset pricing theory to obtain a parsimonious representation of the expected excess returns of the different assets to be included in a portfolio, and of their dependence structure as measured by the associated variance-covariance matrix. These models, depending on the dynamic specification of the factors and the errors, allow to predict the evolution of the entire cross-sectional distribution of returns, the expected return and the riskiness of each asset, and those of the portfolio itself. They proved to be central tools to solve, both theoretically and empirically, asset allocation problems consisting on the maximization of the risk-adjusted performance of the portfolio, where the risk adjustment is due to the manager's utility function which is written on the portfolio value. More recently the literature on "fund tournaments", initiated by Brown, Harlow, and Starks (1996), has noted that money managers might also be interested in the relative performance of their portfolios with respect to competitors, justifying new types of utility functions defined on the rank, or relative position, of the portfolio return within a certain cross-sectional distribution of competing returns (see, for instance, Guéant (2013)). In order to solve portfolio allocation problems based on the maximization of these new rank-based utility functions, standard linear model of asset returns might not be flexible enough as they imply a specific dynamic for the shape of the entire cross-sectional distribution of returns, and also for the relative positions of asset returns inside the cross-sectional distribution. In the literature there seem to be no available model for the disjoint characterization of the dynamic of the shape of the cross-sectional distribution of returns, and the dynamic of the relative positions of the assets inside it, which is valuable to solve rank-utility maximization problems. Inspired by - but differently from - the copula model of Bonhomme and Robin (2009), we propose a model with two type of latent factors: the first one driving the functional form of the cross-sectional distribution of returns, while a second type of factors driving the autoregressive coefficient of the ranks of the single individuals. The latter factors allow to capture commonalities over time in the persistence of the relative positions. This model implies a wider range of possible rank dynamics than the standard linear factor model, due to the various nonlinearity features, and at the same time can be estimated by maximum likelihood on a large panel of assets, as the estimators involve only the alternation of time-series and cross-sectional averages of appropriate functions of the data.

Turning to point (ii), the vast majority of the models mentioned up to this point assume that all the available data are (time) sampled at the same frequency. Nevertheless, the data used in empirical research in economics and finance are often observed at different frequencies, for example stock returns are available at daily or higher frequency, while macroeconomic indicators are usually published at monthly or quarterly frequency. On one side, the direct use of mixed frequency data generates new opportunities to better understand the underlying structure of the economy, but on the other it poses some challenges from an econometric point of view. In fact, empirical research generally avoids the direct use of mixed frequency data by either first aggregating higher frequency series and then performing estimation and testing at the low frequency common across the series, or neglecting the low frequency data and working only with the high frequency series. The literature on large scale factor models is no exception to this practice, see e.g. Forni and Reichlin (1998), Stock and Watson (2002a,b) and Stock and Watson (2010), although recently a number of mixed frequency factor models have been proposed. Usually these models rely on small cross-sections, as in the seminal papers of Mariano and Murasawa (2003), Nunes (2005), Aruoba, Diebold, and Scotti (2009). Stock and Watson (2002b) and Jungbacker, Koopman, and Van der Wel (2011) are among the rare recent attempts to extend these models to large panels of mixed-frequency observables. Moreover, often the data sampled at different frequencies are of a slightly different in nature, therefore a natural question to ask is whether groups of data corresponding to different frequencies share any common factor, or their dynamics are decoupled. Therefore, there appears to be need of a generic class of linear models which can be easily estimated using extensions of principal component methods (i.e. that do not require computationally intensive recursive estimation methods), and can handle large panels of mixed frequency observables, allowing for common factors pervasive to all data, and also for factors loading only on the group of data observed at a specific frequency.

Concluding with point (iii), the first type of factor models developed for panels with small cross-sectional dimensions, and actually very much used also for modeling single time series, are known under the name of "state space models". In order to be able to identify and estimate the latent factors from the data, specific parametric assumptions are explicitly made on the dynamic specification for the latent factors and for the idiosyncratic errors, and not only on the dependence structure between the factor and the observables. When the specifications of the first two dynamics are linear and Gaussian, and the observables load linearly on the factors, then maximum likelihood estimation is possible using a special kind of filtering algorithm for the latent factor, called Kalman filter. When either one of the linearity and Gaussianity assumptions on the idiosyncratic errors are relaxed, like in the case of idiosyncratic errors featuring stochastic volatility (which can be seen as an additional latent factor loading on its variance), then ad-hoc nonlinear filters can be applied for maximum likelihood estimation. An advantage of these models is that they can be easily adapted to handle mixed frequency observables. On the other hand, any departure from the linearity and Gaussianity assumptions, or the increase of the number of observables, often imply an higher computational burden of the filtering technique, with the associated risk of numerical instability of the estimation procedure. In these situations the use of the indirect inference methods (see Gouriéroux, Monfort, and Renault (1993), Smith (1993) and Gallant and Tauchen (1996)) becomes a fast and easy to implement alternative, providing an appropriate auxiliary model - i.e. a misspecified model matching the main features of the true one, and easily estimable - can be found. In a recent paper, Bai, Ghysels, and Wright (2013) find a one-to-one mapping among the optimal forecast of a low frequency variable attainable from a state space model with mixed frequency observables, and a specific type of Mixed Data Sampling (MIDAS) forecasting regressions, i.e. regressions involving as explanatory variables the same high frequency variables included in the state space model (MIDAS regressions were introduced in the seminal paper of Ghysels, Santa-Clara, and Valkanov (2006)). This insight, together with the older tradition in finance of using ARCH-type auxiliary models to estimate stochastic volatility models (see, for instance, Engle and Lee (1999)), allow to have the proper auxiliary models to perform indirect inference estimation of small scale state space models featuring stochastic volatility in the idiosyncratic term of the observables.

The objectives of this thesis are the development of new econometric methodologies for the estimation of latent factor models on large and mixed frequency datasets, and their application to asset allocation, the study of comovement in the output growth of different sectors of the US economy, and the forecasting of European GDP. The rest of the thesis is articulated in three additional chapters. Chapter 2, corresponds to an article coauthored with Patrick Gagliardini and Christian Gourieroux, and presents a new type of asset allocation strategies based on a novel dynamic model of the cross-sectional distribution of returns and the ranks of assets inside the same cross-sectional distribution. These strategies are implemented on a large panel of US stocks, and are shown to perform well compared to traditional asset allocation strategies.

Chapter 3, corresponds to a paper coauthored with Elena Andreou, Patrick Gagliardini and Eric Ghysels, and proposes a new class of approximate latent factor models suitable for large panels of data observed at different frequencies. An empirical application uncovers the common components of monthly data output growth rates of the US Industrial Production sectors, and the yearly output growth rates of all the remaining sectors of the US economy, mainly services. Chapter 4, corresponds to a paper coauthored with Patrick Gagliardini and Eric Ghysels. It introduces indirect inference estimators for state space models featuring mixed frequency observables and stochastic volatility, and considers an application to forecasting quarterly European GDP using monthly macroeconomic indicators. The content of each chapter is outlined more extensively below.

Detailed outline of the chapters

Chapter 2: Positional portfolio management

The second chapter, corresponds to the paper *Positional Portfolio Management*. It introduces a new type of asset allocation strategies named "positional portfolio" strategies. Positional portfolio management is based on the maximization of the expected utility of the future rank, or position, of the portfolio value, as opposed to the traditional portfolio management which focuses on the expected utility of the future portfolio value itself. Specifically, we study positional portfolio management in which the manager maximizes an expected utility function written on the cross-sectional rank, or position, of the portfolio return. This objective function is introduced to reflect the portfolio manager's goal to be well-ranked among competitors. This goal, which does not necessarily coincide with the standard objective of maximizing the risk-adjusted performance of the portfolio, is compatible with the desire of the manager to increase his/her asset under management, and therefore his/her fees, when faced with potential new investors who value only the relative performance with respect to competing funds. In this respect, also the manager might be more interested in relative performance than in absolute performance, as documented by the literature on "fund tournaments" (see e.g. Brown, Harlow, and Starks (1996) and Guéant (2013), among others).

To implement positional allocation strategies, we specify a nonlinear factor model for the asset returns featuring unobservable individual heterogeneities and latent dynamic factors. The model disentangles the dynamics of the cross-sectional distribution of the returns and the dynamics of the ranks, or positions, of the individual assets within the cross-sectional distribution. We estimate our model on a large set of stocks traded in the NYSE, AMEX and NASDAQ markets between 1990/1 and 2009/12, and implement the positional strategies for different investment universes. The factor model that we specify allows enough reduction in the dimension of the parameter space to be able to implement the strategies on the same large dimensional set of assets considered for the estimation. We compare the performance of positional strategies with traditional momentum, reversal, mean-variance, minimum-variance and equally-weighted portfolio allocation strategies. We find that the positional strategies implemented out-of-sample outperform momentum and reversal strategies, as well as mean-variance and minimum-variance strategies in terms of average positional utility and Sharpe ratio. The performance of the positional strategies is similar to that of the equally-weighted portfolio according to these criteria, but the former outperform the latter in terms of probability to be well-ranked. The fact that the strategies based on our model perform well when implemented both in-sample, and out-of-sample, is an indication that the model is sophisticated enough to capture relevant features of the dynamics of the cross-section of US stocks returns.

Chapter 3: Is Industrial Production Still the Dominant Factor for the US Economy?

The third chapter corresponds to the paper *Is Industrial Production Still the Dominant Factor for the US Economy?*. This article introduces a new class of approximate latent factor models in the spirit of Bai and Ng (2002) suitable for large panels of data observed at different frequencies. For expository purposes and compatibility with the empirical application, we consider the simplified setting in which a first panel of measurements is observed at high frequency and a second panel of measurements is observed at low frequency. The model allows for the presence of latent factors common to both high and low frequency data panels, and factors specific to the high and low frequency data. It can be written as a group factor model, and in the proposed identification strategy, the groups correspond to panels observed at different sampling frequencies. While there is a literature on how to estimate factors in a grouped model (see e.g. Krzanowski (1979), Flury (1984), Kose, Otrok, and Whiteman (2008), Goyal, Pérignon, and Villa (2008), Bekaert, Hodrick, and Zhang (2009) and Moench, Ng, and Potter (2013)), there does not exist a general unifying asymptotic theory for large panel data. We propose estimators for the common and group specific factors, and an inference procedure for the number of common and group specific factors based on canonical analysis of the principal components estimators on each subgroup. The procedure is general in scope, can be extended straightforwardly to the case of more than two groups of observables, and has many applications in economics and other fields.

The empirical application revisits the analysis of Foerster, Sarte, and Watson (2011) who use factor analytic methods to decompose industrial production (IP) into components arising from aggregate and sector-specific shocks. They focus exclusively on the industrial production sectors of the US economy. Yet, IP has featured steady decline as a share of US output over the past 30 years, as the US economy has become more a service sector economy. Contrary to IP, we do not have monthly or quarterly data about the cross-section of US output across non-IP sectors, but we do on an annual basis. We identify three factors with our mixed frequency approximate factor model, where the first is a high frequency factor common to all sectors, the second is a high frequency factor specific to IP-sectors and a third is a low frequency factor pertaining only to non-IP sectors. Despite the growth of service sectors, we find that a single common factor explaining 90% of the variability in IP output growth index also explains 60% of total GDP output growth fluctuations. A single low frequency factor unrelated to manufacturing explains 14% of GDP growth. Moreover, we re-examine whether these common factors reflect sectoral shocks that have propagated by way of input-output linkages between service sectors and manufacturing. Hence, our analysis completes an important part missing in the original study of Foerster, Sarte, and Watson (2011), which omitted, among others, the services sector: a major ingredient of US economic activity. A structural factor analysis indicates that one common and two frequency-specific factors to the productivity innovations, or technological shocks, continue to be the dominant source of variation in the output growth of the US economy, but IP sectors technology shocks do not play a dominant role.

Chapter 4: Indirect inference estimation of mixed frequency state space models with stochastic volatility

The fourth and concluding chapter corresponds to the paper *Indirect Inference Estimation of Mixed Frequency Stochastic Volatility State Space Models using MIDAS Regressions and ARCH Models*, joint work with Patrick Gagliardini and Eric Ghysels. This article examines the relationship between MIDAS regressions and the estimation of state space models applied to mixed frequency data featuring stochastic volatility.

To estimate the parameters of the model, we introduce indirect inference estimation procedures proposed by Gouriéroux, Monfort, and Renault (1993), Smith (1993) and Gallant and Tauchen (1996). More specifically, we use MIDAS regressions introduced by Ghysels, Santa-Clara, and Valkanov (2006) combined with ARCH specifications for the errors, as well as mixed frequency Vector Autoregressive (VAR) models (see e.g. Ghysels (2014)) as auxiliary models. It is worth noting that in some specific cases we know the binding function between the state space model and the implied MIDAS regression, as discussed in Bai, Ghysels, and Wright (2013). However, these cases are rather too simple to be practical, so that the use of indirect inference is a natural way to tackle the unknown binding function. In addition to the estimation of the parameters of the model, we filter the latent factors given observables, using reprojection methods proposed by Gallant and Tauchen (1998). Also these filters involve simple regression-based methods combined with ARCH. Hence, we provide a simple estimation methodology alternative to computationally demanding approximate nonlinear filters.

We assess the efficiency of our indirect inference estimator for the stochastic volatility model by comparing it with the Maximum Likelihood estimator (MLE) in Monte Carlo simulation experiments. The ML estimate is computed with a simulation-based Expectation-Maximization algorithm, in which the smoothing distribution required in the Expectation step is obtained via a particle forward-filtering/backward-smoothing algorithm. The Monte Carlo simulations show that the proposed Indirect Inference procedure is very appealing, as its statistical accuracy is close to that of MLE but the former procedure has clear advantages in terms of computational efficiency. Even in the linear Gaussian case, we find our indirect inference methods remarkably accurate, when compared to the standard MLE based on the Kalman filter. An application to forecasting quarterly GDP growth in the Euro area with monthly macroeconomic indicators illustrates the usefulness of our procedure in empirical analysis. Last, it is worth noting that same frequency data settings are a special case of mixed frequency ones. The methodology proposed by this paper is therefore also applicable to standard state space models.

Bibliograpghy of Chapter 1

ARUOBA, S. B., F. X. DIEBOLD, AND C. SCOTTI (2009): "Real-time Measurement of Business Conditions," *Journal of Business and Economic Statistics*, 27(4), 417–427.

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

BAI, J., E. GHYSELS, AND J. H. WRIGHT (2013): "State Space Models and MIDAS Regressions," *Econometric Reviews*, 32, 779–813.

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

BEKAERT, G., R. J. HODRICK, AND X. ZHANG (2009): "International Stock Return Comovements," *Journal of Finance*, 64, 2591–2626.

BONHOMME, S., AND J.-M. ROBIN (2009): "Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000," *Review of Economic Studies*, 76(1), 63–92.

BROWN, K. C., W. V. HARLOW, AND L. STARKS (1996): "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry," *Journal of Finance*, 51, 85–110.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure and Mean-variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

CONNOR, G., AND R. KORAJCZYK (1986): "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394.

——— (1988): "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics*, 21, 255–289.

——— (1993): "A Test of the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263–1291.

ENGLE, R. F., AND G. G. LEE (1999): "A Permanent and Transitory Component Model of Stock Return Volatility," in *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, ed. by R. Engle, and H. White, pp. 475–497. Oxford University Press, Oxford.

FLURY, B. N. (1984): "Common Principal Components in k Groups," *Journal of the American Statistical Association*, 79, 892–898.

FOERSTER, A. T., P.-D. G. SARTE, AND M. W. WATSON (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119(1), 1–38.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The generalized dynamic-factor model: Identification and estimation," *Review of Economics and Statistics*, 82, 540–554.

FORNI, M., AND L. REICHLIN (1996): “Dynamic Common Factors in Large Cross-Sections,” *Empirical Economics*, 21, 27–42.

——— (1998): “Let’s Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics,” *The Review of Economic Studies*, 65, 453–473.

GALLANT, A. R., AND G. TAUCHEN (1996): “Which Moments to Match?,” *Econometric Theory*, 12, 657–681.

——— (1998): “Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions,” *Journal of the American Statistical Association*, 93, 10–24.

GHYSELS, E. (2014): “Macroeconomics and the Reality of Mixed Frequency Data,” *Journal of Econometrics* (forthcoming).

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2006): “Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies,” *Journal of Econometrics*, 131, 59–95.

GOURIÉROUX, C., A. MONFORT, AND E. RENAULT (1993): “Indirect Inference,” *Journal of Applied Econometrics*, 8, 85–118.

GOYAL, A., C. PÉRIGNON, AND C. VILLA (2008): “How Common are Common Return Factors across the NYSE and Nasdaq?,” *Journal of Financial Economics*, 90, 252–271.

GUÉANT, O. (2013): “Tournament-induced Risk-shifting: A Mean Field Games Approach,” *Risk and Decision Analysis*.

JUNGBACKER, B., S. J. KOOPMAN, AND M. VAN DER WEL (2011): “Maximum Likelihood Estimation for Dynamic Factor Models with Missing Data,” *Journal of Economic Dynamics and Control*, 35(8), 1358 – 1368.

KOSE, A. M., C. OTROK, AND C. H. WHITEMAN (2008): “Understanding the Evolution of World Business Cycles,” *Journal of International Economics*, 75, 110–130.

KRZANOWSKI, W. (1979): “Between-groups Comparison of Principal Components,” *Journal of the American Statistical Association*, 74, 703–707.

MARIANO, R. S., AND Y. MURASAWA (2003): “A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series,” *Journal of Applied Econometrics*, 18(4), 427–443.

MOENCH, E., S. NG, AND S. POTTER (2013): “Dynamic Hierarchical Factor Models,” *Review of Economics and Statistics*, 95, 1811–1817.

NUNES, L. C. (2005): “Nowcasting Quarterly GDP Growth in a Monthly Coincident Indicator Model,” *Journal of Forecasting*, 24(8), 575–592.

ROSS, S. A. (1976): “The Arbitrage Theory of Capital Asset Pricing,” *Journal of Economic Theory*, 13, 341–360.

SMITH, A. A. (1993): "Estimating Nonlinear Time-series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, 63–84.

STOCK, J., AND M. WATSON (1998): "Diffusion Indexes," *NBER Workin Paper 6702, August*.

STOCK, J. H., AND M. W. WATSON (2002a): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167 – 1179.

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–162.

——— (2010): "Dynamic Factor Models," in *Oxford Handbook of Economic Forecasting*, ed. by E. Graham, C. Granger, and A. Timmerman, pp. 87–115. Michael P. Clements and David F. Hendry (eds), Oxford University Press, Amsterdam.

Chapter 2

Positional Portfolio Management

Abstract

We study positional portfolio management in which the manager maximizes an expected utility function written on the cross-sectional rank (position) of the portfolio return. The objective function reflects the manager's goal to be well-ranked among competitors. To implement positional allocation strategies, we specify a nonlinear unobservable factor model for the asset returns which disentangles the dynamics of the cross-sectional distribution and the dynamics of the ranks of the individual assets. Using a large dataset of stocks returns we find that positional strategies outperform standard momentum, reversal and mean-variance allocation strategies, as well as equally weighted portfolio for criteria based on position.

JEL Codes: C38, C55, G11.

Keywords: Positional Good, Robust Portfolio Management, Rank, Fund Tournament, Factor Model, Big Data, Equally Weighted Portfolio, Momentum, Positional Risk Aversion.

2.1 Introduction

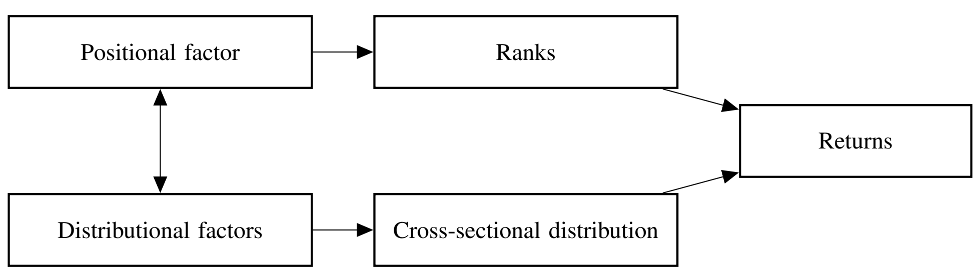
The management fees of portfolio managers should be designed to reconcile the objectives of these managers with the objectives of the investors. They depend on the asset under management for mutual funds, and also on the returns of the portfolio above some benchmark threshold, the so-called high-water mark, for hedge funds [Brown, Harlow, and Starks (1996), Aragon and Nanda (2012), Darolles and Gourieroux (2014)]. These designs might be not entirely satisfactory and induce spurious portfolio management. For instance, the effect of high-water mark can lead managers to take too risky short term positions and use a high leverage. Similarly, to increase his/her market share, that is the asset under management, the manager has to get better performance than his/her competitors. In this respect, the manager might be more interested in relative performance than in absolute performance, especially when the journals for investors write lead articles or even make their cover page on the ranking of funds. For instance, "in the Managed and Personal Investing section of the *Wall Street Journal Europe*, the Fund Scorecard provides the return of the top fifteen performers in a category" [Goriaev, Palomino, and Prat (2001)].

The traditional Finance theory assesses the quality of a portfolio management strategy by considering the expected (indirect) utility of the portfolio value, or of the portfolio return. A portfolio with 10% expected return is preferred to a portfolio with 8% expected return for a given level of risk. However, this preference ordering can be questioned if we account for the *context*, that is, for competing portfolio managements. Do we prefer a 10% return when the competing portfolio return is 20%, or a 8% return when the competing portfolio return is 5% ? Indeed, with 8% return the portfolio manager is number one, whereas he/she is not with 10% return. Economic theory uses the term *positional good* to "denote the good for which the link between context", i.e., the behaviour of other economic agents, "and evaluation is the strongest", and the term *nonpositional good* to denote that for which the link is the weakest [Hirsch (1976), Frank (1991)]. Positional theory has proved useful to explain the escalation of expenditures in armaments, the race for technology in electronic financial markets [Biais, Foucault, and Moinas (2013)], the negative association between happiness measures and average neighbourhood income [Easterlin (1995), Frey and Stutzer (2002)], the sharp increase in the surface of newly constructed houses in the United States, the labour force participation of married women [Neumark and Postlewaite (1998)], and the demand for luxury goods [Frank (1999)]. The application of positional theory in Finance, which is the closest to the topic of this paper, is the competition for talented agents, especially for CEOs, fund managers, or traders in the finance sector [see e.g. Gabaix and Landier (2008), Thanassoulis (2012)]. Indeed, the fact that investors look for talented fund managers might explain the incentive for positioning introduced in the contracts for management fees, as well as the race of fund managers to be well ranked, i.e. the so-called fund tournament [Goriaev, Palomino, and Prat (2001), Goriaev, Nijman, and Werker (2005), Chen and Pennacchi (2009), Schwarz (2012)].

The aim of this paper is to introduce the positional concern in portfolio management. The positional portfolio management is based on the maximization of the expected utility of the future rank (or position) of the portfolio value, as opposed to the traditional portfolio management which focuses on the expected utility of the future portfolio value itself. The positional portfolio management leads to new types of allocations strategies, which we compare theoretically and empirically with traditional allocation strategies, such as mean-variance, momentum and contrarian (or reversal) strategies, as well as the naive $1/n$ portfolio. We measure the ability of positional strategies to yield portfolio returns that rank well cross-sectionally. A positional strategy diverts resources to be well ranked in the race among portfolio managers and might diminish the absolute performance compared to nonpositional strategies. In this respect, such a management does not necessarily act in the interest of investors. Therefore, one goal of our analysis is to measure the loss (or gain) of absolute performance due to a positional strategy.[1]

In Section 2.2, we introduce the notion of cross-sectional rank (position). This notion is used to define a positional portfolio management, and is at the core of the distinction of this management from the standard management based on the expected utility of future portfolio returns. A positional strategy can be interpreted as a standard strategy in which the utility function is replaced by a stochastic utility, which is function of the stochastic future cross-sectional distribution of returns. To implement the positional portfolio strategy we need an appropriate specification which disentangles the dynamics of the ranks from those of the cross-sectional distribution of returns. The model for the dynamics of ranks is introduced in Section 2.3. The Gaussian ranks follow a conditionally Gaussian autoregressive process, with the autoregressive coefficient accounting for positional persistence. The latter can depend on unobservable individual heterogeneities and stochastic dynamic factors. The dynamic model for the ranks is used in Section 4 to construct a first type of positional portfolio allocation strategies, which are compared with standard momentum and reversal strategies on a large panel of returns for stocks traded in the NYSE, AMEX and NASDAQ markets. The investment universe for these positional strategies consists of about 1000 stocks, which illustrates the big data aspect of our analysis. In Section 2.5 we complete the model by introducing an appropriate specification for the dynamics of the cross-sectional distribution of individual stock returns. The distribution is chosen in the Variance-Gamma family, with stochastic mean, variance, skewness and kurtosis driven by unobservable common factors, in order to accommodate time-varying higher-order moments of the cross-sectional returns distribution. The full vector of macro-factors driving positional persistence and the moments of the cross-sectional distribution follows a vector autoregressive (VAR) process. The specifications for the dynamics of positions, cross-sectional distribution and underlying factors define the joint dynamics of returns. The complete dynamic model is summarized in Scheme 1.

Scheme 1: The model structure



This complete dynamic model is used in Section 6 to construct efficient positional portfolio allocation strategies. We compare the performance of the momentum and efficient positional strategies with the performance of traditional mean-variance, minimum-variance and $1/n$ strategies. We find that the positional strategies implemented out-of-sample outperform momentum and reversal strategies, as well as mean-variance and minimum-variance strategies in terms of average positional utility and Sharpe ratio. The performance of the positional strategies is similar to that of the equally-weighted portfolio according to these criteria, but the former outperform the latter in terms of probability to be well-ranked. Section 2.7 concludes. Technical proofs and a discussion of the Nash equilibrium of positional strategies are gathered in Appendices.

[1]We do not discuss in this paper the positive or negative effects of such tournaments on the global social welfare, as possible misuse of resources on the negative side, and acceleration of innovation on the positive side.

2.2 Positional portfolio management

2.2.1 Returns and positions

Let us consider a set of n risky assets $i = 1, \ldots, n$, which can be either stocks, or fund portfolios, and a riskfree asset with riskfree rate $r_{f,t}$. We denote by $y_{i,t}$ the return of risky asset i in period t, for $t = 1, ..., T$. At any given date, the observed returns can be used to define the ranks (or positions) of the assets. For this purpose, it is necessary to distinguish the ex-ante and ex-post notions of rank (or position). The ex-post ranks are simply obtained by ranking at any given date t the asset returns from the smallest one to the largest one, and then taking their positions in this ranking (divided by n). Formally, the ex-post ranks are defined as $\hat{u}^*_{i,t} = \hat{H}^*_t(y_{i,t})$, where $\hat{H}^*_t(\cdot)$ is the empirical cross-sectional (CS) cumulative distribution function (c.d.f.) of the returns at date t. In the ex-ante analysis, the empirical cross-sectional c.d.f. at the current date t is replaced by its theoretical analogue, denoted by $H^*_t(\cdot)$ (see Appendix 1). Then, the ex-ante ranks are given by $u^*_{i,t} = H^*_t(y_{i,t})$. The rank $u^*_{i,t}$ corresponds to the position of return $y_{i,t}$ ex-ante with respect to the observation of the other asset returns. The ex-ante ranks have a cross-sectional uniform distribution on the interval $[0, 1]$, whereas the ex-post ranks have the discrete empirical uniform distribution on $\{1/n, 2/n, ..., 1\}$.

Since the ranks are defined up to an increasing transformation, we can also introduce the ex-ante and ex-post Gaussian ranks. They are obtained from the corresponding uniform ranks by applying the quantile function of the standard normal distribution:

$$u_{i,t} = \Phi^{-1}(u^*_{i,t}) \qquad \text{and} \qquad \hat{u}_{i,t} = \Phi^{-1}(\hat{u}^*_{i,t}), \tag{2.2.1}$$

where Φ is the c.d.f. of the standard normal distribution. The ex-ante Gaussian ranks $u_{i,t}$ (resp. the ex-post Gaussian ranks $\hat{u}_{i,t}$) are standardized to ensure a cross-sectional standard normal distribution (resp. a cross-sectional distribution close to the standard normal one for large n). For instance, if asset i has ex-post rank $\hat{u}^*_{i,t} = 0.95$, there are 95% of assets in the sample with a smaller or equal return on time t, and 5% of assets with a larger return. The corresponding ex-post Gaussian rank is $\hat{u}_{i,t} = 1.64$, that is the 95% quantile of the standard normal distribution. If an asset i has ex-ante rank $u_{i,t} = 0.95$, there is a probability equal to 0.95 that the return at time t of any other asset is smaller or equal to the return of asset i. The ex-ante Gaussian ranks are related to the returns by the equation $u_{i,t} = H_t(y_{i,t})$, where H_t is the compound function $H_t = \Phi^{-1} \circ H^*_t$.

To illustrate the notions of ex-ante and ex-post cross-sectional distributions, we consider the subsample of all Center for Research in Security Prices (CRSP) common stocks [2] traded on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX) and the NASDAQ, for which the monthly holding-period returns are available for the period ranging from January 1990 to December 2009. We exclude from the dataset the stocks for which monthly volume data are either missing, or equal to 0, at some months. We get a balanced panel for the returns of $n = 939$ companies, with $T = 240$ monthly observations. We compute the empirical cross-sectional distribution of returns $\hat{H}^*_t$ at the end of each month of the sample. The associated smoothed probability density functions are displayed in Figure 2.1.

[2] Common stocks are stocks with CRSP End of Period Share Code 10 and 11. Therefore, our sample does not include Certificates, American Depositary Receipts (ADR), Shares of Beneficial Interest (SBI), Units, Exchange-Traded Funds (ETF), Companies incorporated outside the U.S., Close-ended funds, and Real Estate Investment Trusts (REIT). Stock prices are denominated in US dollars. The CRSP dataset includes 15044 different common stocks listed in the NYSE, the AMEX and the NASDAQ, in the period from January 1990 to December 2009.

[FIGURE 2.1: Time series of cross-sectional distributions of monthly CRSP stock returns.]

We deduce from these distributions the associated 5%, 25%, 50%, 75%, 95% empirical cross-sectional quantiles, which are time varying. The time series of these quantiles are displayed in Figure 2.2.

[FIGURE 2.2: Time series of quantiles of the CS distributions of monthly CRSP stock returns.]

The empirical cross-sectional distributions are generally unimodal, with a mode close to the zero return. They vary over time, mainly in their concentration and tails. As expected, we observe in Figure 2.2 an endogenous clustering of these effects: the individual returns are more cross-sectionally concentrated at some periods of time, and less concentrated at some other ones.

In Figure 2.3 we consider an hypothetical riskfree asset with a constant monthly return 0.05 and provide the time series of its ex-post Gaussian ranks.

[FIGURE 2.3: Time series of ex-post Gaussian ranks associated with a constant monthly return of 0.05.]

This constant return is below the CS median in some months, and above the 95% CS quantile in other months. In Figure 2.3 these effects are reflected by the fact that the ex-post Gaussian rank is smaller than 0, or larger than 1.64, respectively, at some months.

2.2.2 Positional management strategies

Let us assume that the investor's information at date t, denoted by I_t, includes the current and past realizations of all asset returns: $I_t = (\underline{r_{f,t}}, \underline{y_t})$, where $\underline{y_t} = (y_t, y_{t-1}, \dots)$ and $y_t = (y_{1,t}, \dots, y_{n,t})'$. The standard (myopic) portfolio management summarizes the preferences of the investor by means of an increasing concave indirect utility function U written on the future portfolio value. The investor selects at time t the portfolio allocation which maximizes the expected utility of the future portfolio value. Let us consider a portfolio invested in both risky and riskfree assets and denote by γ the vector of dollar allocations in the risky assets, $w_r = \gamma' e$ the budget invested in the risky assets, and e the n-dimensional unit vector. Then, $\alpha = \gamma / w_r$ is the vector of relative allocations in the risky assets. By taking into account the budget constraint, the future portfolio value is equal to:

$$W_{t+1} = W_t(1 + r_{f,t}) + \gamma' \tilde{y}_{t+1} = W_t(1 + r_{f,t}) + w_r \alpha' \tilde{y}_{t+1},$$

where W_t is the portfolio value at date t and $\tilde{y}_{t+1} = y_{t+1} - r_{f,t} e$ is the vector of excess returns. The optimization problem provides the optimal allocations $\hat{\gamma}_t$ by:

$$\hat{\gamma}_t = \arg\max_{\gamma} E_t \left(U[W_t(1 + r_{f,t}) + \gamma' \tilde{y}_{t+1}] \right), \tag{2.2.2}$$

where $E_t(\cdot) = E(\cdot | I_t)$ is the conditional expectation given the available information at time t, and the allocation $\hat{\gamma}_t$ can depend on this information. The optimal values $\hat{\gamma}_t$, $\hat{w}_{r,t} = \hat{\gamma}_t' e$ and $\hat{\alpha}_t = \hat{\gamma}_t / \hat{w}_{r,t}$ are also solutions of the two equivalent constrained optimization problems:

$$
\begin{aligned}
\hat{\gamma}_t &= \arg\max_{\gamma} E_t\left(U[W_t(1+r_{f,t}) + \gamma' \tilde{y}_{t+1}]\right), \\
&\quad \text{s.t. } \gamma' e = \hat{w}_{r,t},
\end{aligned}
$$

and:

$$
\begin{aligned}
\hat{\alpha}_t &= \arg\max_{\alpha} E_t\left(U[W_t(1+r_{f,t}) + \hat{w}_{r,t}\alpha' \tilde{y}_{t+1}]\right), \\
&\quad \text{s.t. } \alpha' e = 1.
\end{aligned} \tag{2.2.3}
$$

Thus, the optimization can be splitted into two parts. In a first step we consider the optimal allocation of the total budget between the riskfree asset and the set of risky assets, that is $W_t - \hat{w}_{r,t}$ and $\hat{w}_{r,t}$. Then, the budget $\hat{w}_{r,t}$ is allocated between risky assets. For a CARA indirect utility function and conditionally Gaussian returns, we get the standard mean-variance efficient allocation [see e.g. Ingersoll (1987), p. 98]. In this case the quantity $\hat{w}_{r,t}$ depends on the risk aversion and on the conditional distribution of excess returns, but not on the initial portfolio value W_t. The relative allocations vector $\hat{\alpha}_t$ depends on the conditional distribution of excess returns only:

$$
\hat{\alpha}_t = \frac{1}{e' \left[V_t(\tilde{y}_{t+1})\right]^{-1} E_t(\tilde{y}_{t+1})} \cdot \left[V_t(\tilde{y}_{t+1})\right]^{-1} E_t(\tilde{y}_{t+1}).
$$

The objective of a fund manager could be, for instance, to provide a high portfolio (excess) return, or perhaps to provide a better (excess) return than his competitors. In the latter case, he can prefer to be in the "top ten", whatever the return levels are. Such a positional strategy can be developed for the whole portfolio including both riskfree and risky assets, or only for the risky part of the portfolio once the budgets for the riskfree and risky parts of the portfolio have been fixed. We follow the second approach, that is, we derive the optimal positional allocations vector γ subject to the constraint $\gamma' e = w_r$, for w_r given. We keep the same definition of the ranks as in Section 2.1, that is, we compare the position of portfolios with the positions of each individual stock. Thus, in a first step we consider as exogenous competitors portfolios invested in single assets. We show in Appendix 2 *i)* that the optimal positional strategy is $\gamma_t^* = w_r \alpha_t^*$, where:

$$
\begin{aligned}
\alpha_t^* &= \arg\max_{\alpha:\, \alpha' e = 1} E_t\left[\mathcal{U}\left(H_{t+1}(\alpha' y_{t+1})\right)\right] & (2.2.4)\\
&= \arg\max_{\alpha:\, \alpha' e = 1} E_t\left[\mathcal{U}\left(H_{t+1}\left(\sum_{i=1}^{n} \alpha_i H_{t+1}^{-1}(u_{i,t+1})\right)\right)\right], & (2.2.5)
\end{aligned}
$$

where $\mathcal{U}(\cdot)$ is a utility function written on the Gaussian rank $H_{t+1}(\alpha' y_{t+1})$ of the future return $\alpha' y_{t+1}$ of the risky part of the portfolio. Equation (2.2.5) leads to three remarks. First, the optimal positional relative allocations vector α_t^* is independent of w_r, i.e., it can be computed for a risky portfolio of unitary value 1. The reason is that a positional strategy is not interested in the levels of the portfolio values, but only on their comparison. Second, the ranks are computed on the returns. Indeed, the ranks computed on the returns, or on the excess returns, are the same. Third, in equation (2.2.5) the future portfolio rank $H_{t+1}\left(\sum_{i=1}^{n} \alpha_i H_{t+1}^{-1}(u_{i,t+1})\right)$ is a nonlinear aggregate of the individual future ranks [see Appendix 2 *ii)*].

The nonlinear aggregation scheme involves the stochastic future cross-sectional distribution of returns H^*_{t+1} via function $H_{t+1} = \Phi^{-1} \circ H^*_{t+1}$.

By comparing equations (2.2.3) and (2.2.4), we note that the positional utility function $\mathcal{U}$ is different from the rescaled indirect utility function U_t, with $U_t(r) = U[W_t(1+r_{f,t}) + \hat{w}_{r,t} r]$, written on the portfolio excess return $r = \alpha' \tilde{y}_{t+1}$ of the risky part of the portfolio. In particular, the argument of the rescaled indirect utility function U_t admits the unit: \$ at time $t+1$ over \$ at time t, while the argument of the positional utility function $\mathcal{U}$ is dimensionless. A positional strategy replaces the increasing and concave rescaled utility function U_t by an endogenous stochastic utility function $U_{t+1} = \mathcal{U} \circ H_{t+1}$, which is strictly increasing, but non-concave in general. The positional portfolio management depends on the choice of the positional utility function $\mathcal{U}$, but also on the selected definition of ranks, that can be uniform or Gaussian, and on the universe of stocks used to compute these ranks. Moreover, the optimal allocation α^*_t of the fund manager is defined by considering in a first step the function H_{t+1} as exogenous. In particular, we have chosen the return distribution of portfolios invested in single assets as such exogenous benchmark. When the rank is computed with respect to the performance of other managed portfolios, the future "portfolio returns" distribution has also to account for the possible reactions of the other fund managers, who also want to be in the "top ten". In this case, all portfolios associated with the funds would be optimized jointly. Thus, the performance of the fund is seen as a public good [see e.g. Hirsch (1976), Frank (1991)]. In other words, if we considered the positional equilibrium condition as the analogue of the standard CAPM, the equilibrium would be with respect to the prices, information set and also to the cross-sectional distribution H_{t+1}. In Appendix 2 *iii)* we describe the Nash equilibrium for the positional allocation problem of fund managers in such a complex framework. The objective function in equation (2.2.4) would involve the behaviours of the other fund managers. Therefore, our analysis is related to the literature on social interactions [see e.g. Davezies, D'Haultfoeuille, and Fougère (2009) and Blume, Brock, Durlauf, and Jayaraman (2013)], especially the part of this literature interested in strategic complementarities in production [Calvo-Armengol, Patacchini, and Zenou (2009)]. However, it differs from this literature because of the more sophisticated objective function which is considered [3]. In particular, at the equilibrium we do not get bilateral effects only, i.e. peer effects only. In our framework the individual decision involves in a complicated way the complete distribution of other managers' decisions. The extension to an endogenous benchmark H_{t+1} is not considered further in the main body of the paper.

The preferences based on expected positional utility satisfy some axioms of the expected utility theory introduced by von Neumann and Morgenstern (1944), but not all of them. For instance, the expected positional utility is a linear function of the probability of the future state including in our case the returns of all assets. However, the compatibility with the second-order stochastic dominance for the portfolio returns is clearly not satisfied, since the preferences also involve the distributions of the other stock returns.[4]

In order to implement the positional strategies defined in (2.2.5) and to compare them with the standard allocation strategies based on the expected utility of future portfolio values, we need an appropriate dynamic model for both the rank processes and the transformed cross-sectional distribution H_t linking the returns and

[3]In the standard social interaction models, the individual utility functions are quadratic functions of the individual actions (i.e. the portfolio allocations). They depend on these individual actions by summaries interpretable as cross-sectional portfolio values up to a change of probability. This leads to linear models for the Bayesian-Nash equilibrium strategies.

[4]In this respect our approach differs from the theory of anticipated utility, also called rank dependent expected utility theory. This theory introduced by Quiggin (1982) [see also Kahneman and Tversky (1992) for the extension to prospect theory] is using some ranks to overweigh rare extreme events. In this approach, the rank would be computed as the rank of the portfolio value across states of nature, that is, with respect to the (conditional) distribution of the portfolio return. In our framework the rank is computed with respect to the cross-sectional distribution of the stock returns. Thus, the benchmark distribution used to compute the rank varies with the state of nature.

the ranks. An illustrative example is discussed below and is extended in the next sections to accommodate the empirical features of the return processes.

2.2.3 A toy-model of cross-sectional Gaussian returns

Let us consider a simple joint dynamic model for returns and Gaussian ranks, in which the returns are *cross-sectionally* Gaussian and the ranks are serially persistent. We use this toy-model to provide a simple intuition for positional portfolio allocation, and will extend it for empirical analysis.

The model is defined in two steps, by specifying first the dynamics of the Gaussian ranks and then the link between the individual asset returns and their ranks. The ex-ante Gaussian ranks $u_{i,t}$ are assumed such that:

$$u_{i,t} = \rho u_{i,t-1} + \sqrt{1-\rho^2}\varepsilon_{i,t}, \tag{2.2.6}$$

where the idiosyncratic disturbance terms $\varepsilon_{i,t}$ are independent and identically distributed (i.i.d.) standard normal variables. The autoregressive coefficient ρ has a modulus smaller than 1 in order to ensure the stationarity of the process of Gaussian ranks. The unconditional distribution of $u_{i,t}$ coincides with the theoretical cross-sectional distribution and is standard normal. When coefficient ρ increases, the position of any asset features more serial persistence.

Suppose that the returns are defined from the Gaussian ranks by an affine stochastic transformation:

$$y_{i,t} = \sigma_t u_{i,t} + \mu_t, \tag{2.2.7}$$

where the scale and drift coefficients define the macro-dynamic factor $F_t = (\mu_t, \sigma_t)'$. The scale σ_t, that is the cross-sectional standard deviation, is a strictly positive process. Then, the cross-sectional return distribution at date t is Gaussian $N(\mu_t, \sigma_t^2)$. The function H_t mapping returns into Gaussian ranks is given by $H_t(y) = (y - \mu_t)/\sigma_t$. It simply consists in cross-sectionally demeaning and standardizing the returns. The individual return processes are not Gaussian, since they feature stochastic mean and variance due to factors μ_t and σ_t.

Let us now consider a portfolio invested in both risky and riskfree assets, with relative risky allocation vector α. The future return of the risky part of the portfolio is given by:

$$\alpha' y_{t+1} = \sigma_{t+1}\alpha' u_{t+1} + \mu_{t+1},$$

since $\alpha' e = 1$, and the corresponding excess return is:

$$\alpha' \tilde{y}_{t+1} = \sigma_{t+1}\alpha' u_{t+1} + \mu_{t+1} - r_{f,t}.$$

The future position of return $\alpha' y_{t+1}$ is:

$$H_{t+1}(\alpha' y_{t+1}) = \frac{(\sigma_{t+1}\alpha' u_{t+1} + \mu_{t+1}) - \mu_{t+1}}{\sigma_{t+1}} = \alpha' u_{t+1}. \tag{2.2.8}$$

Thus, the position of the future return of the risky part of the portfolio is a linear combination of the Gaussian ranks of the individual risky assets, with weights equal to the relative risky allocations α. This property is a consequence of the linearity of the (transformed) quantile function $H_{t+1}(\cdot)$, that is, of the Gaussian assumption for the CS distribution, and holds for any dynamics of the ranks. By taking into account the dynamics (2.2.6) of the Gaussian ranks, we get:

$$\alpha' \tilde{y}_{t+1} = \sigma_{t+1}\rho\alpha' u_t + \sigma_{t+1}\sqrt{1-\rho^2}\alpha'\varepsilon_{t+1} + \mu_{t+1} - r_{f,t}, \tag{2.2.9}$$

and:

$$H_{t+1}(\alpha' y_{t+1}) = \rho\alpha' u_t + \sqrt{1-\rho^2}\alpha'\varepsilon_{t+1}. \tag{2.2.10}$$

In the standard approach to portfolio management, we assume a CARA indirect utility function $U(W;A) = -\exp(-AW)$ written on the portfolio value, where $A > 0$ is the absolute risk aversion of the investor. From (2.2.3) and (2.2.9) the expected utility is:

$$\begin{aligned}
& -E[\exp(-AW_t(1+r_{f,t}) - Aw_r\alpha'\tilde{y}_{t+1})|\underline{F_t}, \underline{y_t}] \\
= \;& -\exp(-AW_t(1+r_{f,t}))E\left\{E[\exp(-Aw_r\alpha'\tilde{y}_{t+1})|\underline{F_{t+1}}, \underline{y_t}]|\underline{F_t}, \underline{y_t}\right\} \\
= \;& -\exp(-A(W_t + (W_t - w_r)r_{f,t})) \\
& \times E\left\{\exp(-Aw_r\sigma_{t+1}\rho\alpha' u_t - Aw_r\mu_{t+1} + \frac{A^2}{2}w_r^2\sigma_{t+1}^2(1-\rho^2)\alpha'\alpha)|\underline{F_t}, \underline{y_t}\right\}.
\end{aligned}$$

The optimal portfolio is obtained by maximizing the above expected utility with respect to w_r and α subject to $\alpha' e = 1$. The optimal allocation depends on the joint dynamics of the cross-sectional mean and cross-sectional variance. If these dynamics are Markovian and exogenous with respect to the ranks, the optimal allocation depends on the current factor values (μ_t, σ_t) and ranks vector u_t. The allocations $\hat{\gamma}_t$ and $\hat{\alpha}_t$ in the risky assets are independent of the initial portfolio value W_t.

In the positional approach, we assume a CARA utility function $\mathcal{U}(v;\mathcal{A}) = -\exp(-\mathcal{A}v)$ written on the Gaussian rank of the future return of the risky part of the portfolio, with a positional risk aversion parameter $\mathcal{A} > 0$. By using equation (2.2.10), the expected positional utility is:

$$-E[\exp(-\mathcal{A}H_{t+1}(\alpha' y_{t+1}))|\underline{F_t}, \underline{y_t}] = -\exp\left(-\mathcal{A}\rho\alpha' u_t + \frac{\mathcal{A}^2}{2}(1-\rho^2)\alpha'\alpha\right).$$

The expected positional utility is independent of the factor values at time t and depends on the returns histories by means of the current positions vector u_t only. The optimal positional portfolio allocation is derived by maximizing $\mathcal{A}\rho\alpha' u_t - \frac{\mathcal{A}^2}{2}(1-\rho^2)\alpha'\alpha$ with respect to vector α subject to the budget constraint $\alpha' e = 1$. We get the optimal relative positional allocation in the risky assets:

$$\alpha_t^* = \frac{1}{n}e + \frac{1}{\mathcal{A}}\frac{\rho}{1-\rho^2}(u_t - \bar{u}_t e), \tag{2.2.11}$$

where $\bar{u}_t = u_t' e/n$ denotes the cross-sectional average of the Gaussian ranks at date t. This cross-sectional average tends to 0, which is the mean of the standard normal distribution, when the number of assets n tends to infinity. The optimal relative positional allocation α_t^* is a linear combination of two popular portfolios. The first one is the equally weighted portfolio, with weight $1/n$ in each asset [see e.g. DeMiguel, Garlappi, and Uppal (2009) and Beleznay, Markov, and Panchekha (2012)]. We see from (2.2.10) that this portfolio minimizes the conditional variance of future portfolio rank. For large n, the $1/n$ portfolio ensures the risk free median rank, but its return is still risky due to the effect of macro-factors. The second portfolio is an arbitrage portfolio (zero-cost portfolio) with dynamic allocations proportional to the current ranks of the assets in deviation from their cross-sectional average. The weight of the arbitrage portfolio in the relative risky allocation α_t^* is increasing with respect to the persistence ρ of the ranks, and decreasing with respect to the positional risk aversion coefficient $\mathcal{A}$ of the investor. The optimal positional allocation α_t^* deviates from the $1/n$ portfolio by overweighting the assets with larger (resp. smaller) current ranks, when the persistence parameter is positive (resp. negative)[5]. Thus, in this example the optimal positional allocation strategy combines the $1/n$ portfolio with momentum (resp. reversal) kind of strategies. The term $\rho(u_t - \bar{u}_t e)$ in equation (2.2.11) is equal to the vector of expected future ranks in deviation from their cross-sectional average. Thus, we can also interpret the arbitrage portfolio in (2.2.11) as a portfolio investing long in assets with large expected future rank and short in assets with small expected future rank, irrespective of the sign of the persistence parameter. This interpretation applies to more general specifications of the individual ranks dynamics, as shown in the next section.

2.3 The dynamics of positions

This section extends the toy dynamic model of positions in Section 2.3 to accommodate relevant empirical features. The main issue is that in our sample positional persistence varies across stocks and time [see Appendix 3 for evidence based on an Analysis of Variance (ANOVA)]. Therefore, we let the positional persistence depend on both stock-specific random effects and stochastic common dynamic factors.

2.3.1 Model specification

The joint dynamics of the individual Gaussian rank processes $(u_{i,t})$ is now specified as:

$$u_{i,t} = \rho_{i,t} u_{i,t-1} + \sqrt{1-\rho_{i,t}^2}\varepsilon_{i,t}, \tag{2.3.1}$$

$$\rho_{i,t} = \Psi(\beta_i + \gamma_i F_{p,t}), \tag{2.3.2}$$

where *i)* the idiosyncratic shocks $(\varepsilon_{i,t})$, the individual random effects $\delta_i = (\beta_i, \gamma_i)'$, and the macro-factor $F_{p,t}$ are mutually independent, *ii)* the shocks $(\varepsilon_{i,t})$ are standard Gaussian white noise processes independent across assets, and *iii)* the individual random effects δ_i are i.i.d. across assets. In equation (2.3.1) we assume that the Gaussian rank process $(u_{i,t})$ of any stock follows a conditionally Gaussian first-order Auto-Regressive [$AR(1)$] model. The autoregressive coefficient $\rho_{i,t}$ characterizes the positional persistence of stock i between months $t-1$ and t. The dependence of the autoregressive coefficient $\rho_{i,t}$ on the macro-factor and the individual effects is specified in equation (2.3.2). The single stochastic factor $F_{p,t}$ drives the positional persistence over time, that is, it is a positional macro-factor, whose interpretation has to be discussed jointly with the interpretation of the distributional macro-factors driving the cross-sectional return distribution (see Section 5). The individual effects β_i and γ_i introduce heterogeneity across stocks in the long run average positional persistence and in the sensitivity to the positional persistence factor, respectively. We select function $\Psi(s) = (e^{2s}-1)/(e^{2s}+1)$, for $s \in \mathbb{R}$, to guarantee an autoregressive coefficient $\rho_{i,t}$ between -1 and 1 and to get a one-to-one increasing relationship between $\rho_{i,t}$ and the positional persistence score $\beta_i + \gamma_i F_{p,t}$. Since $\Psi(s) \approx s$ for an argument s close to 0, the positional persistence $\rho_{i,t}$ is approximately equal to the score $\beta_i + \gamma_i F_{p,t}$, when the latter is small in absolute value. The model in equations (2.3.1)-(2.3.2) extends specification (2.2.6) to individual and time dependent positional persistence. The joint process of individual Gaussian ranks defined in equations (2.3.1)-(2.3.2) satisfies the constraint of a standard Gaussian CS distribution [see Appendix 4, Subsections *i)* and *ii)*].

[5]Even if the model of the example is symmetric in the individual assets, the portfolio allocation is not symmetric in the assets, since they have different returns, and then ranks, at date t.

The dynamic model (2.3.1)-(2.3.2) introduced for describing the evolution of the ranks is appropriate to represent the idea of market dislocations, that are "circumstances in which financial markets operating under stressful conditions cease to price assets correctly on a ... relative basis" [see Pasquariello (2014)]. Such a situation of dislocation arises for instance at the beginning of the 2007 financial crisis, more precisely on August 8 and 9 as shown in Khandani and Lo (2011) by analyzing the properties of contrarian portfolio management strategies. To illustrate this point, let us consider a homogeneous set of assets (for expository purpose), with individual effects equal to $\bar{\beta}$ and $\bar{\gamma}$. The serial dependence of the individual ranks in this set of assets is generally positive, if the probability for the positional factor to take values larger than $-\bar{\beta}/\bar{\gamma}$ is large enough. Then, the ranking of these assets is rather stable in time. Let us now consider a stressed situation where the factor takes a very negative value; then, the serial correlations between the individual ranks become negative, which implies that a lot of well ranked assets will become bad ranked, and vice-versa.

As usual in latent factor models, the factor values and the factor loadings are identifiable up to a one-to-one linear (affine) transformation. Indeed, systems $(F_{p,t}, \beta_i, \gamma_i)$ and $(cF_{p,t} + d, \beta_i - d/c, \gamma_i/c)$ are observationally equivalent, for any values of constants c and d, with $c \neq 0$. Therefore, without loss of generality, we assume:

$$E(F_{p,t}) = 0, \qquad E(F_{p,t}^2) = 1, \tag{2.3.3}$$

for identification purpose. Thus, for an asset i with small β_i and γ_i, the historical mean and variance of the positional persitence are approximately β_i and γ_i^2, respectively.

2.3.2 Model estimation

Let us now estimate the model of ranks dynamics on the dataset of $n = 939$ CRSP stocks described in Section 2.2.1.

i) Estimation procedure

We estimate the values of the positional persistence factor $F_{p,t}$ at all months t, and heterogeneities β_i and γ_i for all stocks i, by maximizing the Gaussian conditional log-likelihood function of rank processes $(u_{i,t})$ after replacing the unobservable ex-ante Gaussian rank $u_{i,t}$ with the empirical ex-post Gaussian rank $\hat{u}_{i,t}$ defined in Section 2.2. Indeed, the ex-post and ex-ante ranks are close, when the cross-sectional size n is large[6]. We treat factor values and individual heterogeneities as unknown parameters. The fixed effects estimators $\hat{F}_{p,t}$ of the factor values, for $t = 1, ..., T$, and $\hat{\beta}_i$, $\hat{\gamma}_i$ of the heterogeneities, for $i = 1, ..., n$, are obtained from the maximization problem:

$$\max_{\substack{F_{p,t},\, t=1,...,T \\ \beta_i, \gamma_i,\, i=1,...,n}} \sum_{t=1}^{T} \sum_{i=1}^{n} \left\{ -\frac{1}{2} \log(1 - \rho_{i,t}^2) - \frac{(\hat{u}_{i,t} - \rho_{i,t}\hat{u}_{i,t-1})^2}{2(1 - \rho_{i,t}^2)} \right\}, \tag{2.3.4}$$

where $\rho_{i,t} = \Psi(\beta_i + \gamma_i F_{p,t})$, subject to the constraints:

$$\frac{1}{T}\sum_{t=1}^{T} F_{p,t} = 0, \qquad \frac{1}{T}\sum_{t=1}^{T} F_{p,t}^2 = 1. \tag{2.3.5}$$

[6]This is a consequence of the Law of Large Numbers (LLN) applied cross-sectionally conditionally on the path of macro-factors.

The constraints (2.3.5) are the empirical analogues of the identification conditions (2.3.3). In Appendix 4 *iii)*, we provide a sequential updating algorithm for the iterative computation of the estimates that are solutions of the constrained maximization problem (2.3.4)-(2.3.5). This sequential updating algorithm avoids the inversion of matrices of dimensions (n, n) or (T, T) corresponding to the parameter dimension. Thus, it has a small degree of numerical complexity appropriate in our big data framework.

ii) Empirical results

We provide in Figure 2.4 the time series of factor estimates $\hat{F}_{p,t}$. The estimated serial autocorrelations are not significant. Thus, we will assume that the factor values $F_{p,t}$ are independent and identically distributed over time. This assumption implies the independence across time of common shocks to positional persistence, but not the absence of positional persistence itself.

[FIGURE 2.4 : Time series of positional factor estimates $\hat{F}_{p,t}$.]

[FIGURE 2.5 : Positional factor vs. CRSP EW index returns.]

Figure 2.5 shows a negative association between the estimated positional factor values and the monthly returns of the equally weighted (EW) CRSP index returns, at least for the months with negative EW CRSP index returns. This finding is similar to Figure 4 in Moskowitz, Ooi, and Pedersen (2012) who report a U-shape relationship between their momentum strategy and the S&P 500 index returns.

Let us now consider the estimated heterogeneity parameters $\hat{\beta}_i$ and $\hat{\gamma}_i$. Their marginal distributions are displayed in Figure 2.6, and some insight on their joint distribution is given by the scatterplot in Figure 2.7. The marginal distributions are unimodal and the values of $\hat{\beta}_i$ and $\hat{\gamma}_i$ in the support have the same order of magnitude. The marginal distribution of the $\hat{\beta}_i$ is close to a Gaussian distribution, while the marginal distribution of $\hat{\gamma}_i$ features right skewness. Thus, for large positive (resp. negative) values of positional factor $F_{p,t}$, we expect a large proportion of stocks with large positive (resp. negative) positional persistence. Figure 2.7 shows that the nonparametric regression of $\hat{\gamma}_i$ on $\hat{\beta}_i$ is almost linear. We observe a significant positive slope in this regression. Hence, the stocks which feature more positional persistence on average, also feature more time variation in this positional persistence.[7]

[FIGURE 2.6 : Histograms of estimated individual effects.]

[FIGURE 2.7 : Scatterplot of $\hat{\gamma}_i$ vs. $\hat{\beta}_i$.]

[7] We find some association between the individual effects (β_i, γ_i) and the industrial sectors. For instance the sector with the largest average value of sensitivity γ_i to the positional factor is Energy, and the one with the smallest average value is Healthcare, Medical Equipment, and Drugs. A standard t-test rejects the null hypothesis that the mean values of the distributions of the γ_i in the two sectors are the same.

The effect of the heterogeneity parameters on positional persistence is rather complex, since it involves the distribution of individual effects β_i and γ_i including their dependence, the level of the factor $F_{p,t}$, and passes through the nonlinear transformation Ψ. Figure 2.8 displays the distribution of the positional persistence for different factor levels.

[FIGURE 2.8 : Histograms of positional persistence $\hat{\rho}_{i,t}$ as function of $F_{p,t}$.]

When the positional factor value is negative (resp. positive), we observe a negative (resp. positive) average value of the positional persistence. When the factor $F_{p,t}$ gets larger in absolute value, the main effect comes from the γ_i distribution and the dispersion of the distribution of positional persistence increases. When the factor is close to 0, corresponding to the median of its distribution, we observe mainly the distribution of sensitivities β_i. Overall Figures 2.6 - 2.8 show that the estimated model accommodates for some stocks featuring momentum and others featuring reversal at a given date. Moreover, a given stock might feature momentum at some dates, and reversal at other dates, depending on the value of the positional factor $F_{p,t}$.

2.3.3 Efficient positional allocation with Gaussian CS distribution

In this section we derive the optimal positional allocation when the individual positions follow the autoregressive model (2.3.1)-(2.3.2) with stochastic persistence, and the CS distributions of returns are in the Gaussian family with stochastic mean and variance as in (2.2.7). Let us consider the CARA positional utility function $\mathcal{U}(v) = -\exp(-\mathcal{A}v)$, with $\mathcal{A} > 0$. The optimal positional allocation α_t^* defined in (2.2.4) is such that [see Appendix 4 *iv)*]:

$$\alpha_{i,t}^* = w_{i,t} + \frac{1}{\mathcal{A}}\left(\xi_{i,t} - w_{i,t}\sum_{i=1}^{n}\xi_{i,t}\right), \tag{2.3.6}$$

where:

$$w_{i,t} = [E_t^{\alpha}(1-\rho_{i,t+1}^2)]^{-1} \Big/ \sum_{i=1}^{n}[E_t^{\alpha}(1-\rho_{i,t+1}^2)]^{-1}, \tag{2.3.7}$$

$$\xi_{i,t} = \frac{E_t^{\alpha}(\rho_{i,t+1})u_{i,t}}{E_t^{\alpha}(1-\rho_{i,t+1}^2)}, \tag{2.3.8}$$

and $E_t^{\alpha}(\cdot)$ denotes the conditional expectation under a modified probability distribution such that $E_t^{\alpha}(\rho_{i,t+1}) = E\left[\rho_{i,t+1}\exp\left(-\mathcal{A}H_{t+1}(\alpha_t^{*\prime}y_{t+1})\right)|\underline{F_t}, \underline{y_t}\right] / E\left[\exp\left(-\mathcal{A}H_{t+1}(\alpha_t^{*\prime}y_{t+1})\right)|\underline{F_t}, \underline{y_t}\right]$. The optimal positional allocation (2.3.6) extends the allocation derived in (2.2.11) to the case of stochastic positional persistence. This allocation is a linear combination of two portfolios. The first one has positive weights $w_{i,t}$, that vary across assets as an increasing function of the modified conditional expectation $E_t^{\alpha}(\rho_{i,t+1}^2)$ of the squared positional persistence. If the positional persistence were time invariant, i.e. $\rho_{i,t+1} = \rho_i$, this portfolio would be the portfolio with the least risky future rank, conditional on the current values of the ranks. The second portfolio is an arbitrage portfolio (zero-cost portfolio), with weights involving the modified conditional expected ranks $E_t^{\alpha}(\rho_{i,t+1})u_{i,t}$. Equation (2.3.6) defines the optimal positional allocation α_t^* in an implicit way, since the RHS of this equation depends on vector α_t^* through the modified conditional expectation of the positional persistence and its square.

Since the positional persistence values for most assets and dates are rather small (see Figure 2.8), in order to get more intuition on equation (2.3.6), we can consider its first-order expansion w.r.t. $\rho_{i,t+1}$. In this approximation we have $w_{i,t} \simeq 1/n$ and $\xi_{i,t} \simeq E_t(\rho_{i,t+1})u_{i,t}$ [see Appendix 4 *vi)*], where $E_t(\rho_{i,t+1}) = E(\rho_{i,t+1}|F_t, \delta_i)$. This yields an explicit formula for the approximate optimal positional allocation:

$$\alpha^*_{i,t} = \frac{1}{n} + \frac{1}{\mathcal{A}}\left(E_t(\rho_{i,t+1})u_{i,t} - \frac{1}{n}\sum_{i=1}^{n} E_t(\rho_{i,t+1})u_{i,t}\right). \tag{2.3.9}$$

Thus, the optimal positional allocation is the linear combination of the equally weighted portfolio and an arbitrage portfolio, whose weights are given by the conditional expected ranks of the assets in deviation from their cross-sectional average. Equation (2.3.9) is the generalization of (2.2.11), for small stock-specific and time varying positional persistence.

2.4 Momentum strategies based on ranks

In this section we compare simple (suboptimal) positional allocation strategies which are based on the ranks of the assets and their dynamics, only.

2.4.1 Investment universe versus positioning universe

When analyzing a positional strategy, it is important to precisely define the investment universe, that is the set of assets potentially introduced in the portfolio, and the positional universe, that is the set of assets and portfolios used to define the rankings. For instance, for a fund (resp. fund of funds) manager, the investment universe may be a fraction of the stocks (resp. funds), whereas the positioning universe can be the set of all stocks (resp. all funds, or all funds including the funds of funds). The dynamic model for positions developed in Section 2.3 is appropriate for an investment universe nested in the positioning universe. In this section, we consider simple positional strategies for which both the positioning universe and the investment universe are the set of 939 stocks in our balanced panel from CRSP.

2.4.2 Positional momentum strategies

As a first illustration of positional strategies, let us consider momentum and contrarian (reversal) approaches [see e.g. Lehmann (1990), Jegadeesh and Titman (1993), Chan, Jegadeesh, and Lakonishok (1996)]. These strategies will be applied on the complete universe of stocks. We consider below the nine following strategies:

i) The (positional) momentum strategies denoted by PMS1 (resp. PMS2), which select an equally weighted portfolio including all stocks whose current return is in the upper 5% quantile of the CS distribution (resp., between the upper 10% and 5% quantiles), i.e. the past winners. The current (ex-post) Gaussian ranks of these stocks are such that $\hat{u}_{i,t} \geq 1.64$ (resp., $1.64 \geq \hat{u}_{i,t} \geq 1.28$). These strategies are similar to standard momentum strategies, but are based on the rank of the return on the current month, instead of the rank of the return over a longer period in the past. It is commonly believed that many stocks feature reversal in returns at a short monthly horizon [see e.g. Jegadeesh (1990) and Avramov, Chordia, and Goyal (2006)], likely due to overreaction of some investors to news[8] [De Bondt and Thaler (1985)]. Therefore, we also consider (positional) reversal strategies PRS1 (resp. PRS2), which select an equally weighted portfolio including all stocks with current rank in the lowest 5% quantile (resp., between the lower 10% and 5% quantiles), i.e. the past losers.

ii) The expected positional momentum strategies EPMS1 IN and EPMS1 OUT (resp. EPMS2 IN and EPMS2 OUT) based on the information on the rank histories. These strategies select equally weighted portfolios including the stocks with the 5% largest *expected* future ranks at each month (resp., the stocks with expected future ranks between the 5% and 10% upper quantiles). The estimated model in Section 3 is used to compute the conditional expectation of the future ranks given the current information. As the positional factor ($F_{p,t}$) is assumed i.i.d. over time, the expected future rank of asset i is given by $E_t(u_{i,t+1}) = \bar{\rho}_i u_{i,t}$, where the expected positional persistence $\bar{\rho}_i = E[\Psi(\beta_i + \gamma_i F_{p,t+1})|\beta_i, \gamma_i]$ involves the expectation with respect to the historical distribution of $F_{p,t+1}$. In our numerical implementation, the expectation is replaced by a sample average over the factor estimates $\hat{F}_{p,t+1}$, the stock-specific effects are replaced by the estimates $\hat{\beta}_i$ and $\hat{\gamma}_i$, and the ex-ante current rank is replaced by the ex-post rank $\hat{u}_{i,t}$. For the in-sample strategies EPMS1 IN and EPMS2 IN the entire available sample of returns from January 1990 to December 2010 is used to estimate the factor model. On the other hand, in order to assess the out-of-sample performance, for the out-of-sample strategies EPMS1 OUT and EPMS2 OUT the model is re-estimated at each month using a rolling window of 10 years of data. The expected future ranks determining the EPMS allocation are computed using these rolling estimates.

iii) As a benchmark, we also consider the market portfolio defined as the equally weighted portfolio computed on all stocks.

We provide in Figure 2.9 the ex-post properties of these portfolios over the period from January 2000 to December 2009.

[FIGURE 2.9 : Ex-post properties of the portfolio strategies, 2000-2009.]

Panel (a) provides the evolution of the Gaussian ranks for the management strategies PMS2, PRS1, EPMS1 OUT, EPMS2 OUT, and the equally weighted portfolio, and panels (b) and (c) the evolution of their excess returns and cumulated returns over the period. The series of Gaussian ranks of the equally weighted portfolio is less disperse than the others. For ease of comparison, we provide historical summary statistics of Gaussian ranks and returns in Table 2.1.

[TABLE 2.1 : Ex-post properties of the portfolio strategies, 2000-2009.]

Even if the standard financial theory suggests that the market portfolio has some efficiency properties, we observe that it is not systematically well ranked, or with the highest return. The historical average of the Gaussian ranks of the equally weighted portfolio and momentum strategies PMS1 and PMS2 are slightly larger than 0. Thus, on average, the return of these strategies is slightly above the CS median, while the historical averages of the ranks of the reversal and expected positional momentum strategies are larger.

[8] More precisely, Chan, Jegadeesh, and Lakonishok (1996) report reversal in stock returns at short horizons smaller than 6 months, and for periods between 3 and 5 years, but momentum in returns between 6 months and 3 years.

The largest average Gaussian rank is featured by the positional strategy EPMS1 IN, and is equal to 0.22, followed by PRS1, EPMS2 IN and EPMS2 OUT with 0.20, 0.17 and 0.14, respectively. As expected from the discussion in Section 3.3, the equally weighted portfolio is close to the median rank. The Sharpe ratios of the expected positional momentum strategies in sample are the largest ones (both higher than 1.10), while PRS1 has a Sharpe ratio (0.96) only slightly larger than EMPS2 OUT (0.94). All positional strategies based on the 5%-10% quantile range are less volatile than the corresponding positional strategies based on the first 5% quantile since they avoid extreme effects. However, while this fact results in a larger Sharpe ratio for EPMS2 OUT compared to EPMS1 OUT, for strategies based on expected future ranks in-sample and reversal strategies, considering the upper 5%-10% quantile range yields a smaller Sharpe ratio than considering the upper 5% quantile. The ranks of the reversal strategy PRS1 are the most volatile ones. Moreover, the series of excess returns of the reversal strategies are negatively skewed and feature the largest negative values, as can be deduced by the historical 5% quantile. To summarize, the out-of-sample expected positional momentum strategy based on the 5%-10% quantile range performs almost as good as the reversal strategy PRS1 in terms of Sharpe ratio, but has substantially smaller downside risk. This is made possible by the combination of momentum and reversal type of allocations across stocks, that is implicit in this strategy. Moreover, the in- and out-of-sample expected positional momentum strategies outperform the equally weighted portfolio, both in terms of average Gaussian rank and Sharpe ratio.

Finally, panel (d) in Figure 2.9 and the bottom part of Table 2.1 give some insight on the asset turnover of the portfolio strategies. The asset turnover is measured by the unweighted proportion of selected stocks which are not kept in the portfolio between two consecutive dates. This unweighted measure of asset turnover provides an information on the potential transaction costs of the portfolio updating. However, this information is rather crude, since it does not account for the variation in the quantities of assets included in the portfolio. For instance, there is no unweighted asset turnover in the equally weighted portfolio, but the transaction costs are not zero for this portfolio, as rebalancing is required to keep the relative allocations in value constant at $1/n$. Among the eight reversal and momentum strategies, the ones based on expected future ranks in the first 5% quantile, both in-sample and out-of-sample, have the smallest average asset turnover.

2.5 The full-fledged model

The positional portfolio strategies implemented in Section 4 rely on the dynamics of the individual ranks only. However, the optimal positional allocation defined in Section 2 generally involves the entire distribution of the return histories. In fact, the portfolio rank is a nonlinear aggregate of the individual ranks depending on the cross-sectional return distribution. Two features have to be considered in order to pass from the dynamics of the ranks to the dynamics of the returns (see Scheme 1 in the Introduction). First, we have to specify the cross-sectional distribution in a flexible way. This cross-sectional distribution varies in time as a function of macro factors. Second, we have to explain how these macro factors, that impact the cross-sectional distribution, are linked to the positional factor, that drives the persistence of ranks dynamics. To get a tractable model, we assume in Section 2.5.1 that the cross-sectional distributions belong to the Variance-Gamma (VG) family (see Appendix 5 for a review on the VG family). The macro factors are time varying parameters characterizing the distributions in this family. Next, in Section 2.5.2 we specify the joint dynamics of the positional and distributional macro-factors by a Gaussian Vector Autoregressive (VAR) model.

2.5.1 Specification of the cross-sectional distributions

Let us first complete the analysis of Section 2 by investigating if the empirical CS distributions are close to Gaussian distributions, and studying how empirical CS summary statistics, such as mean, standard deviation, skewness and kurtosis, vary over time. In Figures 2.10 and 2.11, we provide the empirical CS distributions and their Gaussian approximations at some months. In particular, in Figure 2.11 we focus on the period around the 2008 Lehman Brothers bankruptcy.

[FIGURE 2.10: Cross-sectional distributions of monthly CRSP stock returns.]

[FIGURE 2.11: Cross-sectional distributions around the 2008 Lehman Brothers bankruptcy.]

The comparison between the panels in Figure 2.10 shows that the empirical CS distribution may be close to a Gaussian in some months (e.g. August 1998), may feature rather fat tails (e.g. July 1995, December 2006), or be asymmetric (e.g. November 2000). In Figure 2.11, we see that in July and August 2008, before the Lehman Brothers crisis, the CS distribution is non-normal, with a peak close to 0 and is slightly right-skewed. Instead, in October 2008, the month after Lehman Brothers filed for Chapter 11 bankruptcy protection (September 15, 2008), the CS distribution is close to Gaussian with a large negative mean of about -18%.

The above empirical evidence shows that it is necessary to choose the cross-sectional distributions in an extended family including the Gaussian family as a special case, and to introduce additional macro-factors accounting for time-varying higher-order moments. We consider in our analysis the Variance-Gamma (VG) family. The distributions in this family are indexed by four parameters, that are in a one-to-one relationship with the mean μ_t, the log-volatility $\log \sigma_t$, the skewness s_t and the log excess kurtosis $\log k_t^*$, where $k_t^* = k_t - 3(1 + s_t^2/2)$ and k_t denotes the kurtosis (see Appendix 5). The excess kurtosis k_t^* is a measure of the fatness of the tails of the CS distribution of returns at month t, in excess of $3(1 + s_t^2/2)$. The latter value is the minimum admissible kurtosis for a VG distribution with skewness parameter s_t. Since the above four transformed parameters can vary independently on the entire real line, they are chosen to define the vector of distributional macro-factors:

$$F_{d,t} = (\mu_t, \log \sigma_t, s_t, \log k_t^*)'. \tag{2.5.1}$$

We provide in Figure 2.12 the time series of estimated distributional macro-factor values $\hat{F}_{d,t}$, obtained from the empirical CS moments, along with their asymptotic (large n) pointwise 95 % confidence bands [9].

[FIGURE 2.12: Time series of estimated distributional macro-factors.]

The series of the cross-sectional mean (Panel (a)) is rather close to the return series of the CRSP Equally Weighted Index (not shown). The log CS standard deviation (Panel (b)) is larger around crisis periods, namely in 1991 (Gulf crisis), 1998 (LTCM crisis), 2000-2001 (tech bubble) and 2008-2009 (the subprime crisis). A value of factor $\log \sigma_t$ close to -2 corresponds to a standard deviation of the CS distribution of

[9] The asymptotic standard errors of the distributional macro-factors are computed with the results in Bai and Ng (2005).

returns of about 13.5%. The CS skewness is mostly positive, that is, the CS distributions are often right skewed. The log excess kurtosis varies between 1 and -6, which correspond to excess kurtosis values close to 3 and 0 respectively. The series of CS skewness and kurtosis can be used to compute the Jarque-Bera statistic for the CS distribution of returns for each month. Crisis periods are among the months characterized by the smallest values of the CS Jarque-Bera statistic, that are months in which the CS distribution is closer to a Gaussian one [see panel (b) of Figure 2.10 for August 1998 (LTCM Crisis), and Panel (d) of Figure 2.11 for October 2008 (Lehman Brothers crisis)]. This feature has already been noted by the econophysics literature for daily returns [see e.g. Borland (2012)].

2.5.2 The factor dynamics

Let us now specify the factor dynamics. We assume that the joint vector of distributional and positional macro-factors $F_t = (F'_{d,t}, F_{p,t})'$ follows a 5-dimensional Gaussian Vector Autoregressive process of order 1 [VAR(1)]:

$$F_t = a + AF_{t-1} + \eta_t, \qquad \eta_t \sim IIN(0, \Sigma), \tag{2.5.2}$$

where a is the vector of intercepts, A is the matrix of autoregressive coefficients, and Σ is the variance-covariance matrix of the innovations. We estimate parameters a, A and Σ in the joint VAR dynamics in equation (2.5.2) after replacing the unobservable values of the positional and distributional macro-factors with their estimates $\hat{F}_t = (\hat{F}'_{d,t}, \hat{F}_{p,t})'$, where $\hat{F}_{d,t}$ is defined in Section 2.5.1 and $\hat{F}_{p,t}$ is defined in equations (2.3.4)-(2.3.5).

In Table 2.2 we present the parameter estimates for the macro-factor VAR dynamics with their standard errors in parentheses. We also provide the estimated correlation matrix of the innovations vector.

[TABLE 2.2 : Estimates of the VAR (1) model for the macro-factor process.]

Five coefficients in the estimated autoregressive matrix are statistically significant (at the 1% level). As expected, the autoregressive coefficient of the log CS standard deviation is significant and large (0.84) pointing to a strong serial persistence in the dispersion of the CS distribution. The CS mean also features positive serial persistence, with estimated autoregressive coefficient 0.32. This multivariate regression coefficient has to be compared with the univariate autoregressive coefficient of the monthly return series of the CRSP Equally Weighted Index, that is equal to 0.28 in our sample period. We find a strong evidence for the analog of the Black leverage effect [Black (1976)], namely a negative regression coefficient of the current log CS standard deviation on the past CS mean return equal to -0.79. The estimated coefficient -0.61 of log CS excess kurtosis on lagged log CS standard deviation suggests that the tails in the CS distribution get thinner after a month characterized by a positive shock on the CS dispersion. This effect is likely related to the finding that the CS distribution is close to Gaussian in crisis periods. The multivariate autoregressive coefficient for the log CS excess kurtosis in Table 2.2 is not statistically significant. However, from the clustering in fat tails of the CS distributions observed in Figures 2.1 and 2.2, the CS (excess) kurtosis features serial persistence. Indeed, the univariate autoregressive coefficient of log CS excess kurtosis is equal to 0.37 and is statistically significant. The difference between the univariate and multivariate autoregressive coefficients is explained by the dynamic link between log excess CS kurtosis and log CS standard deviation, and by the contemporaneous correlation between the innovations on these two series. The autoregressive coefficient of the positional factor, and its regression coefficients on the lagged values of the cross-sectional factor, are not statistically significant. This finding is compatible with the marginal white noise property of the positional factor found empirically in Section 3.2 and used in Section 4. The eigenvalues of the estimated autoregressive matrix are 0.811, and two pairs of complex conjugate eigenvalues with modulus 0.187 and 0.071, respectively. Thus, the modulus of all eigenvalues is smaller than 1, which implies the stationarity of the estimated VAR process of the macro-factors driving both the dynamics of ranks and the dynamics of the cross-sectional distribution.

All the estimated contemporaneous covariances of the shocks are significantly different from 0. In particular, we observe a negative contemporaneous correlation equal to -0.32 between the shocks on CS mean and the positional persistence factor. Thus, a small cross-sectional mean of returns tends to be associated with a large positional persistence for those stocks having positive loadings on the positional factor. Such stocks are the majority in our sample (see Figure 2.6). The factor $F_{d,t}$ driving the univariate CS distributions and the factor $F_{p,t}$ driving the positional persistence are not independent. [10]

2.6 Efficient positional strategies

Let us now implement the efficient positional strategies defined in Section 2.2 using the complete dynamic model. Since the efficient positional or mean-variance management strategies demand the inversion of a n-by-n matrix, they can only be applied with a limited number n of assets, significantly smaller than 939. We apply these strategies to an investment universe corresponding to the $n = 57$ stocks in the industrial sector of utilities. The strategies are the following ones: *i)* The efficient positional strategy (EPS), with CARA positional utility function and positional risk aversion $\mathcal{A} = 3$. We implement the EPS strategy both in-sample (EPS IN) by using the entire available sample of returns from January 1990 to December 2010 to estimate the factor model, and out-of-sample (EPS OUT) by using an expanding window of past returns for estimation from January 1990 to the investment date; *ii)* The positional momentum strategy (PMS) based on the 20 stocks with the largest current ranks; *iii)* The positional reversal strategy (PRS) based on the 20 stocks with the smallest current ranks; *iv)* An expected positional momentum strategy (EPMS) based on the 20 stocks with the largest expected future ranks; *v)* The sectoral equally weighted (EW) portfolio; *vi)* The standard mean-variance (MV) strategy based on the unconditional moments. The financial literature reports poor out-of-sample properties for the MV strategy, which is often outperformed by the minimum-variance portfolio strategy [see e.g. Jagannathan and Ma (2003)]. For this reason, we include also the unconditional minimum-variance (MinV) strategy in our comparison. We implement strategies EMPS, MV and MinV out-of-sample [11].

We give in Appendix 6 a numerical algorithm to solve the constrained maximization problem (2.2.4) defining the EPS strategy. The algorithm involves the inversion of a $n \times n$ Hessian matrix, and computational costs grow quadratically in the number of stocks n in the investment universe. This explains the choice to restrict the investment universe compared to Section 2.4. An alternative optimization algorithm with linearly growing computational costs could be obtained by replacing the above Hessian matrix with a diagonal matrix

[10] The point estimate of the conditional variance of factor $F_{p,t}$ is slightly larger than 1. By taking into account the standard deviation of the estimate, this is compatible with the normalization of unconditional variance equal to 1.

[11] Given the relatively large cross-sectional dimension, we use a shrinkage estimator for the variance-covariance matrix of excess returns in both the mean-variance and minimum-variance strategies, as proposed by Ledoit and Wolf (2003). This estimator consists in an optimally weighted average of the sample covariance matrix of the excess returns and a single-index covariance matrix.

having the same diagonal elements.

We provide in Figure 2.13 the time series of cumulated portfolio excess returns for the above strategies. Summary statistics of the Gaussian ranks and excess returns series are presented in Table 2.3.

[TABLE 2.3 : Ex-post properties of the portfolio strategies, utilities sector, 2000-2009.]

[FIGURE 2.13 : Time series of cumulative returns of the portfolio strategies, utilities sector, 2000-2009.]

We consider different criteria to compare the performances of the strategies. They include: *i)* the mean and standard deviation of the Gaussian ranks, and the average positional utility; *ii)* the frequency of returns above a certain cross-sectional quantile of the investment universe; *iii)* summary statistics and Sharpe ratios of the excess returns. The in-sample EPS strategy outperforms the other allocation strategies according to most criteria. This finding is also confirmed by the series of the cumulated returns in Figure 2.13. The EPS strategy implemented out-of-sample outperforms the PMS, PRS, MV and MinV strategies according to both the average positional utility and Sharpe ratio. Along those dimensions, EPS OUT and the equally weighted portfolio perform similarly, but the former ensures larger probabilities to be well-ranked. For instance, the returns of the EPS OUT strategy is about 60% of the times above the CS median of the returns in the investment universe, and about 10% of the times above the CS 60% quantile. Instead, the equally weighted portfolio is above the CS 60% quantile only in 3% of the months. The standard (positional) momentum and reversal strategies PMS and PRS ensure large probabilities to be well ranked, but feature Gaussian ranks that are among the most volatile ones. A similar remark can be done for the MV strategy, which provides a large probability to be in the top 30%, but a very small value of the expected positional utility. In fact the MV strategy alternates very high ranks and very low ranks and appears as very risky. These remarks explain the low average positional utility of the PMS, PRS and MV strategies. Similarly as in Section 2.4, the EPMS strategy outperforms the PMS along all criteria, and features the largest Sharpe ratio. The EPMS overperforms also the PRS concerning the expected positional utility and the Sharpe ratio. In fact, the reversal strategy PRS, which is found to perform rather well for the larger investment universe in Section 2.4, features the smallest Sharpe ratio among the positional strategies for the investment universe of the utilities. The overperformance of the EPMS compared to traditional momentum and reversal strategies is likely due to the ability of the former strategy to exploit the time-varying and stock-specific positional persistence (see Figures 2.6 and 2.7). Indeed, the traditional momentum (resp. reversal) strategies implicitly assume that all stocks feature a positive (resp. negative) positional persistence, that is constant through time. Instead, the EPMS provides a combination of momentum and reversal strategies based on the stock-specific and time-specific information.

In order to better understand the similarities in the performances of the efficient positional strategy and the EW portfolio, in Figure 2.14 we display the relative discrepancy between the allocation vectors implied by these strategies overtime.

[FIGURE 2.14: Observed measure of relative discrepancy of optimal positional allocation from EW portfolio for the subsample of utilities]

The relative discrepancy at a given month t is measured as $n\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\alpha_{i,t}^{*}-1/n)^{2}}$, where the $\alpha_{i,t}^{*}$ are the efficient positional allocations in (2.2.4). This measure corresponds to the ratio between the standard deviation $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\alpha_{i,t}^{*}-1/n)^{2}}$ and the mean $1/n$ of the allocations across assets in the efficient positional portfolio. The relative discrepancy varies overtime between 0.4 and 1.2. In panel (b) of Figure 2.14 we see that the discrepancy increases when the predicted value $E(F_{p,t+1}|F_t)$ of the positional factor $F_{p,t+1}$, based on the full vector of macro-factors F_t, becomes larger in absolute value. We can understand qualitatively the pattern in panel (b) of Figure 2.14 by means of equation (2.3.9), which provides an approximation of the efficient positional allocation when the CS distribution of returns is close to Gaussian and the positional persistence is small. From (2.3.9), the quantity $\frac{1}{n}\sum_{i=1}^{n}(\alpha_{i,t}^{*}-1/n)^{2}$ is equal to the empirical cross-sectional variance of the conditional expected ranks, divided by $\mathcal{A}^2$. When the investment universe is large, i.e. n tends to infinity, the Law of Large Numbers (LLN) implies that this empirical cross-sectional variance converges to its theoretical counterpart, and we get [see Appendix 4 *vii)*]:

$$\frac{1}{n}\sum_{i=1}^{n}(\alpha_{i,t}^{*}-1/n)^{2} \simeq \frac{1}{\mathcal{A}^2}E[(\beta_i+\gamma_i E(F_{p,t+1}|F_t))^2\,|F_t], \tag{2.6.1}$$

where the expectation in the RHS is w.r.t. the distribution of the random effects (β_i, γ_i). We get a quadratic function of the conditional expectation $E(F_{p,t+1}|F_t)$ of the positional factor. This quadratic function is minimized when the macro-factor vector F_t is such that $E(F_{p,t+1}|F_t) = -E[\beta_i\gamma_i]/E[\gamma_i^2]$. From Figures 2.6 and 2.7, the individual effect β_i has a mean close to zero, and the covariance between the individual effects β_i and γ_i is positive, which explains why the minimum of the discrepancy measure in panel (b) of Figure 2.14 is attained for a negative value of the predicted positional factor.

Finally, Table 2.3 also provides information on the turnover of the strategies. The measure of the turnover is now weighted to account for the different weights introduced in an efficient portfolio, and is defined as follows: $Turnover_t = \sum_{i=1}^{57}|\alpha_{i,t}-\alpha_{i,t-1}|$, where $\alpha_{i,t}$ is the relative weight of stock i at date t for a certain strategy. Among the positional strategies, EPS OUT has the lowest and less volatile turnover.

2.7 Conclusions

In this paper we introduce different positional portfolio allocation strategies, that are the expected positional momentum strategies (EPMS) and the efficient positional strategies (EPS). We consider these allocation strategies in a big data framework, in which the investment universe consists of hundreds, or even thousands, of stocks. The implementation of expected positional momentum strategies simply requires a dynamic model for the ranks. This model is used to detect at each date the stocks with high expected future ranks, and the ones with low expected future ranks. This information is implicitly used in the EPMS strategies to mix in an efficient way momentum and reversal (or contrarian) strategies and leads to a closed-form characterization of the EPMS portfolios. The implementation of the efficient positional strategies is more demanding, since these strategies require a complete dynamic model for both the ranks and the cross-sectional distributions of returns. Moreover, the necessity of solving a constrained optimization problem implies that EPS can be applied only with more limited number of assets (57 in our example). As expected, these positional strategies have good properties in terms of the position of the portfolio returns. More surprising are their rather nice properties concerning the portfolio returns themselves. The main reason is that these strategies based on positions are robust to abnormal returns. It is well known that the standard mean-variance allocation strategy is very sensitive to outliers, especially when it is applied with a large number of assets. In particular, its performance can be much worse than the performance of the naive equally weighted portfolio, or $1/n$-strategy, giving the impression that sophisticated allocation strategies are not useful. Our analysis shows that, indeed, the equally weighted portfolio is difficult to outperform for portfolios invested in stocks. For instance, the $1/n$-strategy is clearly competitive with other strategies as the basic momentum, reversal and min-variance strategies. However, the positional strategies outperform the equally-weighted portfolio for performance criteria based on positions.

The positional strategies considered in this paper can be extended in various ways. For instance, we can consider a fund manager interested jointly in different rankings. Then, he or she will optimize a positional utility function depending on these different ranks associated with different universes. It is possible to manage jointly the ranking among the funds of the same management style and the ranking among all funds, and to weight differently the two associated universes. In the perspective of an analysis applied to funds managers' behaviors, it would be interesting to develop inference methods to test if the managers follow positional strategies, and to estimate their selected positioning universes and positional risk aversion parameters. Moreover, as noted in the mutual fund literature, if the rankings are published at the end of each year, the fund managers compete in annual tournaments that begin in January and end in December. They could follow a standard management at the beginning of the year and pass to a positional management with more risk in the second part of the year, if they performed poorly in the first part of the year [see Brown, Harlow, and Starks (1996), Chen and Pennacchi (2009) and Schwarz (2012)]. These questions are left for future research. Likely, the analysis will encounter an identification problem, especially if several fund managers follow such endogenous strategies [see the reflection problem highlighted by Manski (1993)].

2.8 Tables of Chapter 2

Table 2.1: Ex-post properties of the portfolio strategies, 2000-2009.

	PMS1	PMS2	PRS1	PRS2	EPMS1 IN	EPMS2 IN	EPMS1 OUT	EPMS2 OUT	EW
Gaussian ranks									
Mean	0.0276	0.0258	0.2043	0.1318	0.2171	0.1703	0.1105	0.1360	0.0646
St. dev.	0.4591	0.3574	0.4783	0.3748	0.4762	0.3181	0.4179	0.2587	0.1059
Excess returns									
Mean	0.0061	0.0059	0.0228	0.0158	0.0230	0.0184	0.0138	0.0155	0.0088
St. dev.	0.0690	0.0566	0.0818	0.0709	0.0681	0.0565	0.0680	0.0573	0.0490
Sharpe ratio (ann.)	0.3041	0.3587	0.9667	0.7736	1.1706	1.1295	0.7017	0.9357	0.6244
Skew.	0.1896	0.4073	-0.0891	-0.5103	0.2988	0.0581	0.1986	-0.1735	-0.6151
Exc. kurt.	1.1989	1.9350	1.1367	1.7235	1.1672	1.8834	1.5131	1.6940	2.2703
Quant. 5%	-0.1208	-0.0751	-0.1405	-0.1161	-0.0783	-0.0762	-0.1042	-0.0809	-0.0898
Quant. 25%	-0.0343	-0.0334	-0.0185	-0.0144	-0.0152	-0.0136	-0.0284	-0.0176	-0.0171
Quant. 50%	0.0106	0.0054	0.0214	0.0159	0.0203	0.0196	0.0141	0.0195	0.0093
Quant. 75%	0.0492	0.0488	0.0717	0.0573	0.0587	0.0508	0.0572	0.0498	0.0418
Quant. 95%	0.1010	0.0801	0.1445	0.1216	0.1372	0.1042	0.1119	0.1030	0.0737
Turnover									
Turn mean	0.8971	0.9415	0.8931	0.9356	0.8214	0.9080	0.8248	0.9094	0.0000
Turn std	0.0962	0.0933	0.1070	0.0927	0.1073	0.0931	0.0988	0.0940	0.0000

The table provides summary statistics for the monthly series of the Gaussian ranks, and of the excess returns, for the eight portfolio allocation strategies PMS1, PMS2, PRS1, PRS2, EPMS1 IN, EPMS2 IN, EPMS1 OUT and EPMS2 OUT, and for the equally weighted portfolio (EW), in the period 2000/1 - 2009/12. Strategy PMS1 (resp. PMS2) selects an equally weighted portfolio of all stocks whose current return is in the upper 5% quantile of the CS distribution (resp. between the upper 10% and 5% quantiles). Strategy PRS1 (resp. PRS2) selects an equally weighted portfolio of all stocks whose current return is in the lower 5% quantile of the CS distribution (resp. between the lower 10% and 5% quantiles). Strategy EMPS1 IN (resp. EPMS2 IN) selects an equally weighted portfolio of all stocks with the 5% largest expected future rank (resp., with the expected future rank between the upper 5% and 10% quantiles), with the parameters of the model estimated on the full sample (1990/1-2009/12). Strategy EMPS1 OUT (resp. EPMS2 OUT) selects an equally weighted portfolio of all stocks with the 5% largest expected future rank (resp., with the expected future rank between the upper 5% and 10% quantiles), with the parameters of the model estimated on the available sample up to the investment date. The investment universe consists of all the $n = 939$ NYSE, AMEX and NASDAQ stocks in our sample (see Section 2.2 for a description). The ranks are computed w.r.t. the CS distribution of the monthly returns of all the stocks in our sample. The Sharpe ratio is annualized. The table also provides the mean and the standard deviation of turnover. The turnover is measured by the proportion of selected stocks which are not kept in the portfolio between two consecutive dates.

Table 2.2: Estimates of the VAR (1) model for the macro-factor process.

The model for the dynamics of the factor $F_t = (\,\mu_t, \log\sigma_t, s_t, \log k_t^*, F_{p,t}\,)'$ is the Gaussian VAR(1) process:

$$F_t \;=\; a + AF_{t-1} + \eta_t, \qquad \eta_t \sim IIN(0,\Sigma)\,.$$

The estimates for the period 1990/01 - 2010/12 are given by:

$$\hat{a} = \begin{bmatrix} 0.0005 \\ (\mathit{0.0342}) \\ -0.3834^{***} \\ (\mathit{0.0948}) \\ 0.1084 \\ (\mathit{0.3787}) \\ -0.8249^{***} \\ (\mathit{0.2729}) \\ -0.1230 \\ (\mathit{0.8227}) \end{bmatrix}, \qquad \hat{A} = \begin{bmatrix} 0.3156^{***} & -0.0076 & -0.0096 & -0.0117 & -0.0026 \\ (\mathit{0.0914}) & (\mathit{0.0155}) & (\mathit{0.0080}) & (\mathit{0.0091}) & (\mathit{0.0029}) \\ -0.7889^{***} & 0.8406^{***} & -0.0128 & 0.0369 & 0.0066 \\ (\mathit{0.2532}) & (\mathit{0.0430}) & (\mathit{0.0221}) & (\mathit{0.0251}) & (\mathit{0.0080}) \\ 2.9916^{***} & -0.0966 & 0.0023 & -0.0911 & -0.0252 \\ (\mathit{1.0113}) & (\mathit{0.1718}) & (\mathit{0.0882}) & (\mathit{0.1004}) & (\mathit{0.0321}) \\ 0.2588 & -0.6060^{***} & -0.0427 & -0.0211 & -0.0294 \\ (\mathit{0.7290}) & (\mathit{0.1238}) & (\mathit{0.0636}) & (\mathit{0.0723}) & (\mathit{0.0232}) \\ -0.3748 & -0.0341 & -0.2034 & 0.1702 & -0.0112 \\ (\mathit{2.1973}) & (\mathit{0.3732}) & (\mathit{0.1917}) & (\mathit{0.2181}) & (\mathit{0.0698}) \end{bmatrix},$$

$$\hat{\Sigma} = \begin{bmatrix} 0.0017^{***} & & & & \\ (\mathit{0.0001}) & & & & \\ 0.0009^{***} & 0.0134^{***} & & & \\ (\mathit{0.0002}) & (\mathit{0.0009}) & & & \\ 0.0128^{***} & 0.0125^{***} & 0.2134^{***} & & \\ (\mathit{0.0011}) & (\mathit{0.0025}) & (\mathit{0.0138}) & & \\ -0.0015^{**} & -0.0153^{***} & -0.0359^{***} & 0.1109^{***} & \\ (\mathit{0.0006}) & (\mathit{0.0019}) & (\mathit{0.0072}) & (\mathit{0.0072}) & \\ -0.0135^{***} & 0.0177^{***} & -0.0574^{***} & -0.0416^{***} & 1.0073^{***} \\ (\mathit{0.0020}) & (\mathit{0.0054}) & (\mathit{0.0213}) & (\mathit{0.00154}) & (\mathit{0.0650}) \end{bmatrix},$$

with standard errors in parentheses. The asterisks ***, **, * denote that the coefficient is significant at the 1%, 5%, and 10% levels, respectively. The correlation matrix corresponding to the estimated variance-covariance matrix $\hat{\Sigma}$ is:

$$\begin{bmatrix} 1.0000 & & & & \\ 0.1892 & 1.0000 & & & \\ 0.6614 & 0.2348 & 1.0000 & & \\ -0.1075 & -0.3979 & -0.2336 & 1.0000 & \\ -0.3211 & 0.1528 & -0.1238 & -0.1246 & 1.0000 \end{bmatrix}.$$

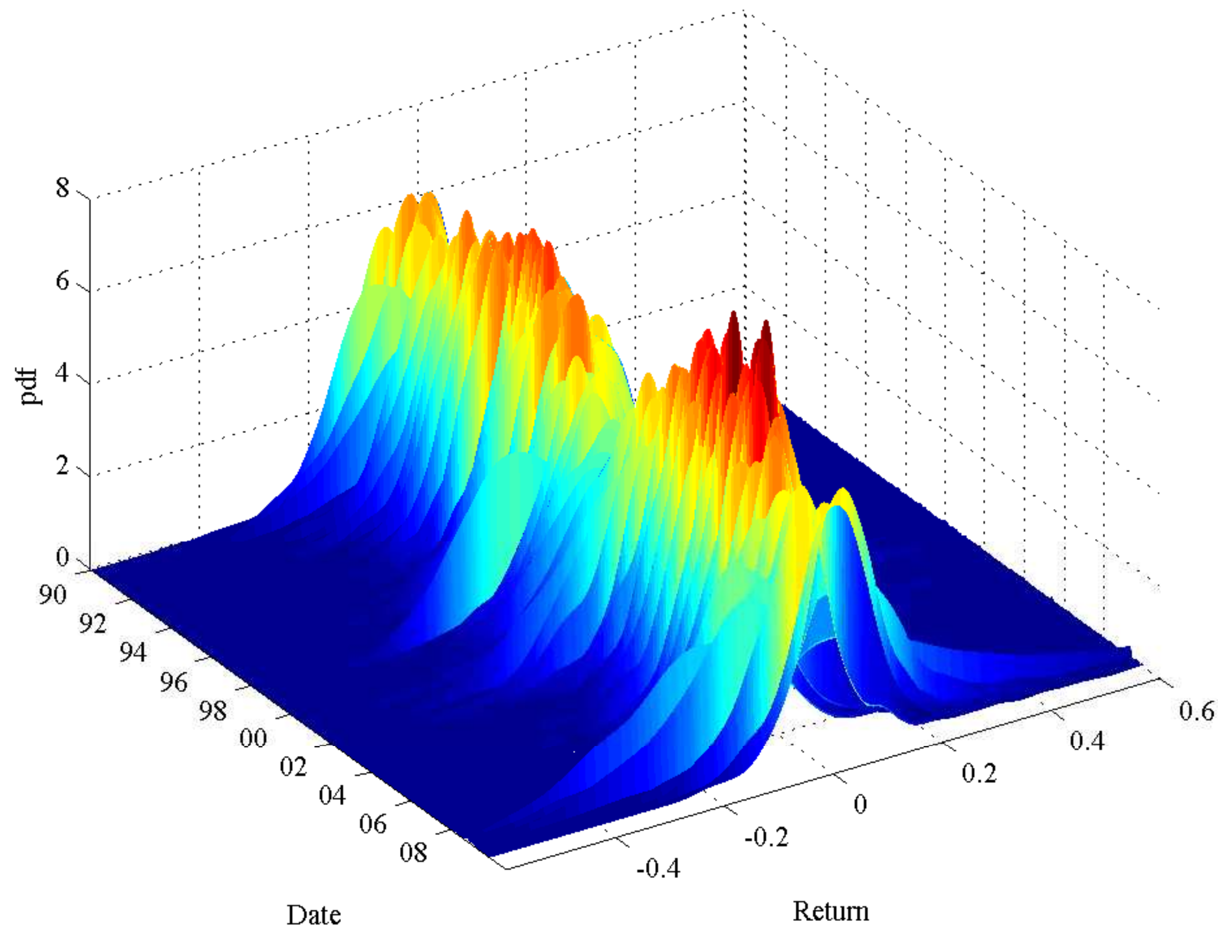
Table 2.3: Ex-post properties of the portfolio strategies, utilities sector, 2000-2009.

	EPS IN	EPS OUT	PMS	PRS	EPMS	EW	MV	MinV
Gaussian ranks								
Mean	0.0906	0.0534	0.0484	0.0589	0.0678	0.0566	-0.0556	0.0119
St. dev.	0.4090	0.4034	0.4406	0.4437	0.4224	0.4043	0.7532	0.4252
E(pos. utility)	-1.6012	-1.7596	-2.0378	-1.9668	-1.9388	-1.7471	-22.6449	-2.2134
Top positions								
$>$ 50% quant.	0.7000	0.6083	0.4833	0.6083	0.5833	0.6167	0.4000	0.4667
$>$ 60% quant.	0.1750	0.0917	0.1917	0.2833	0.2667	0.0333	0.3000	0.2750
$>$ 70% quant.	0.0250	0.0083	0.0667	0.0667	0.0500	0.0000	0.2500	0.2083
Excess returns								
Mean	0.0104	0.0076	0.0076	0.0076	0.0083	0.0079	0.0006	0.0052
St. dev.	0.0424	0.0410	0.0414	0.0476	0.0433	0.0414	0.0611	0.0339
Sharpe Ratio (ann.)	0.8529	0.6436	0.6371	0.5532	0.6614	0.6593	0.0353	0.5344
Skew.	-0.5238	-0.7922	-0.4227	-0.8576	-0.8488	-0.7210	0.1965	-0.2710
Exc. kurt.	1.9679	1.3576	1.3186	1.1699	1.6654	1.1882	2.2707	3.4109
Quant. 5%	-0.0759	-0.0720	-0.0735	-0.0818	-0.0754	-0.0703	-0.1074	-0.0503
Quant. 25%	-0.0090	-0.0111	-0.0135	-0.0099	-0.0100	-0.0138	-0.0320	-0.0137
Quant. 50%	0.0142	0.0124	0.0134	0.0142	0.0134	0.0141	-0.0013	0.0066
Quant. 75%	0.0381	0.0327	0.0321	0.0355	0.0358	0.0326	0.0336	0.0219
Quant. 95%	0.0701	0.0702	0.0706	0.0719	0.0683	0.0692	0.1028	0.0527
Turnover								
Mean	0.5745	0.4825	1.2379	1.1833	1.1818	0.0000	0.4398	0.0862
St. dev.	0.1559	0.1192	0.2378	0.2230	0.2953	0.0000	0.2636	0.1109

The table provides summary statistics for the monthly series of the Gaussian ranks, and of the excess returns, for the eight portfolio allocation strategies with investment universe being 57 stocks in the utilities sector in the period 2000/1 - 2009/12. The efficient positional strategy EPS IN uses the model estimated on the full sample (1990/1-2009/12). EPS OUT is the efficient positional strategy based on the model estimated on the available sample up to the investment date. The positional utility is a CARA function with positional risk aversion parameter $\mathcal{A} = 3$. The strategy PMS (resp. PRS) selects an equally weighted portfolio of the 20 stocks with largest current ranks (resp., smallest current ranks). The strategy EPMS selects an equally weighted portfolio of the 20 stocks with largest expected future ranks. The ranks are computed w.r.t. the CS distribution of the monthly returns of all the NYSE, AMEX and NASDAQ stocks in our sample. Strategies MV and MinV are mean-variance and minimum-variance strategies implemented using a shrinkage estimator for the variance-covariance matrix of excess returns. E(pos. utility) is the time series average of the positional utility of the portfolio returns. In the panel denoted “Top positions” we report the observed frequency of portfolio returns above a certain cross-sectional quantile of the stock returns in the investment universe. The Sharpe ratio is annualized. The table also provides the mean and the standard deviation of turnover, computed at each month t as follows: Turnover$_t = \sum_{i=1}^{n} |\alpha_{i,t} - \alpha_{i,t-1}|$, where $\alpha_{i,t}$ is the weight of stock i at date t.

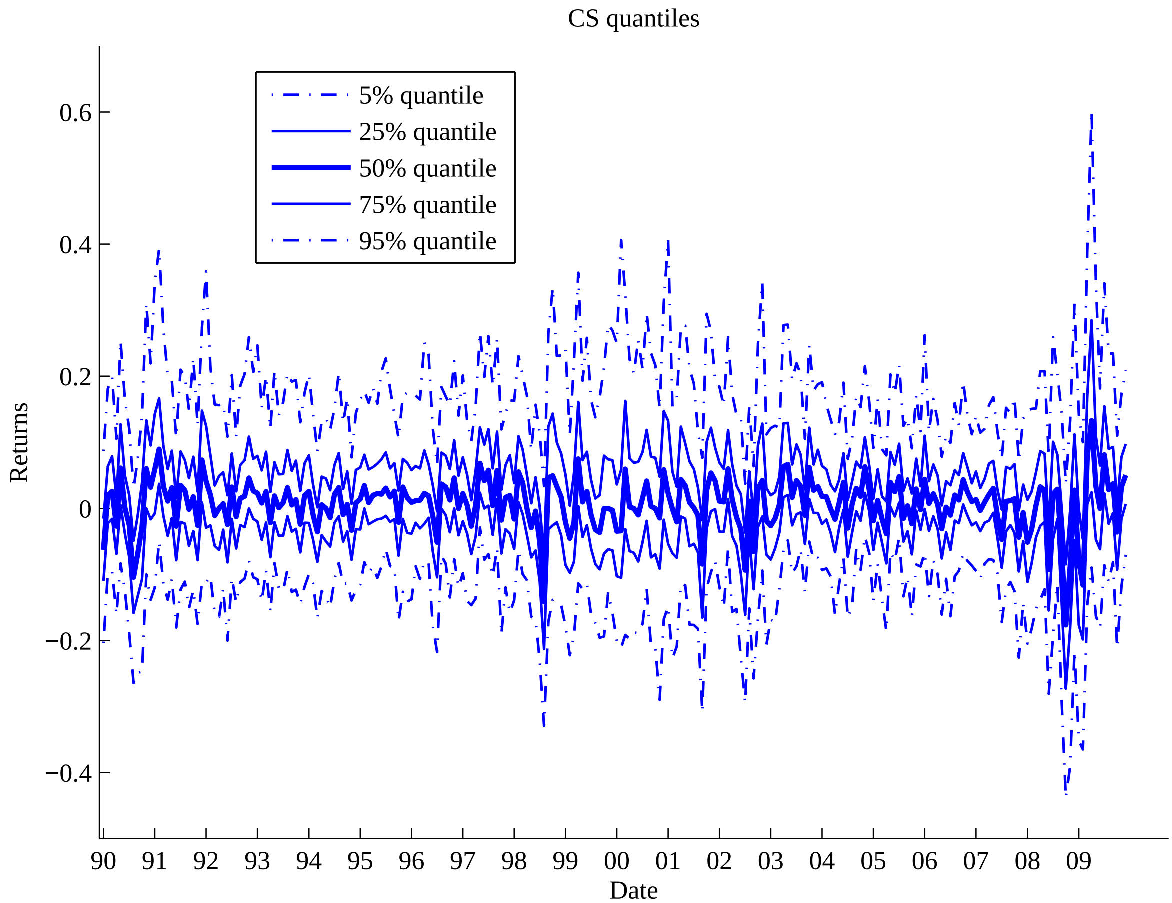
2.9 Figures of Chapter 2

Figure 2.1: Time series of cross-sectional distributions of monthly CRSP stock returns.



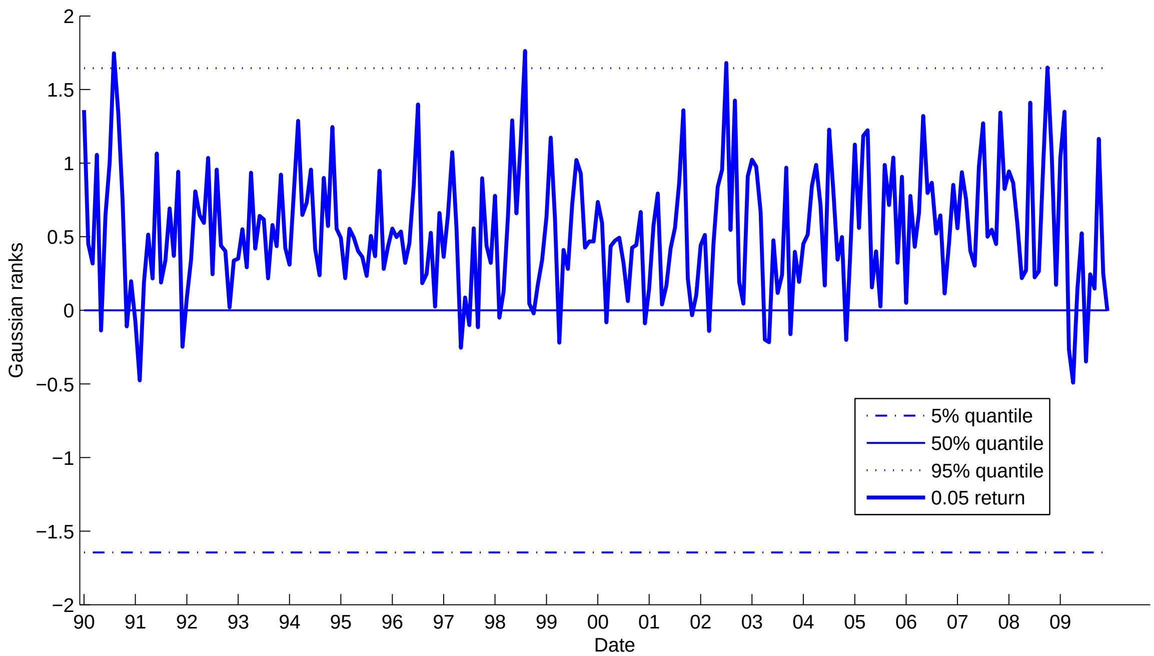
The figure displays the time series of cross-sectional distributions of monthly CRSP stock returns from January 1990 to December 2009. The monthly returns are computed as $y_{i,t} = p_{i,t}/p_{i,t-1} - 1$, where $p_{i,t}$ is the price of stock i at month t. The returns are not annualized and not in percentage. The CS probability density function (p.d.f.) are kernel estimates with Gaussian kernel and bandwidths selected by the rule of thumb in Silverman (1986).

Figure 2.2: Time series of quantiles of the CS distributions of monthly CRSP stock returns.



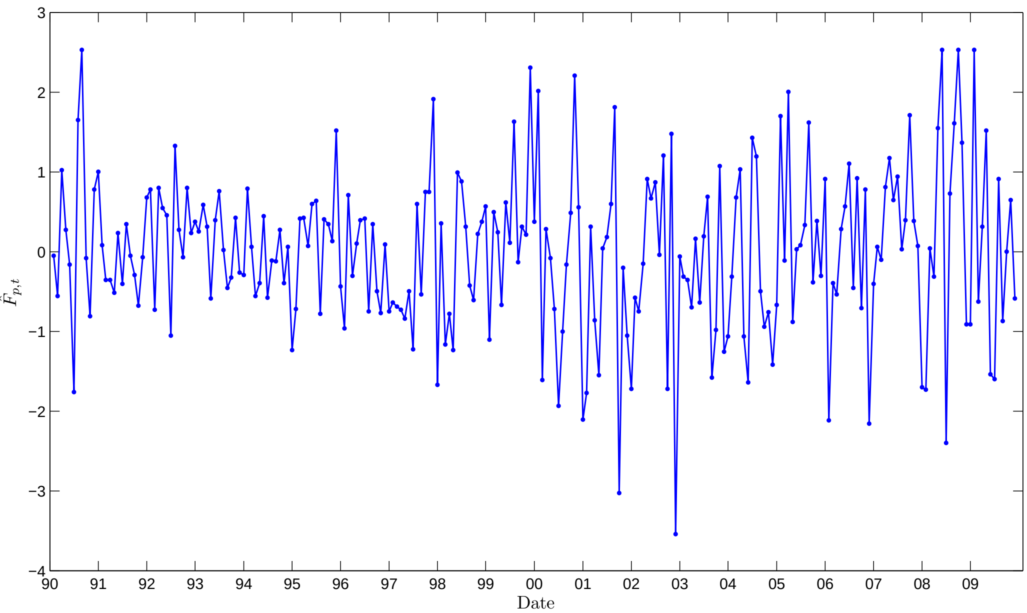
The figure displays the time series of the 5% CS quantile (lower dash-dotted line), the 25% CS quantile (lower solid line), the CS median (bold solid line), the 75% CS quantile (upper solid line), the 95% CS quantile (upper dash-dotted line).

Figure 2.3: Time series of ex-post Gaussian ranks associated with a constant monthly return of 0.05.



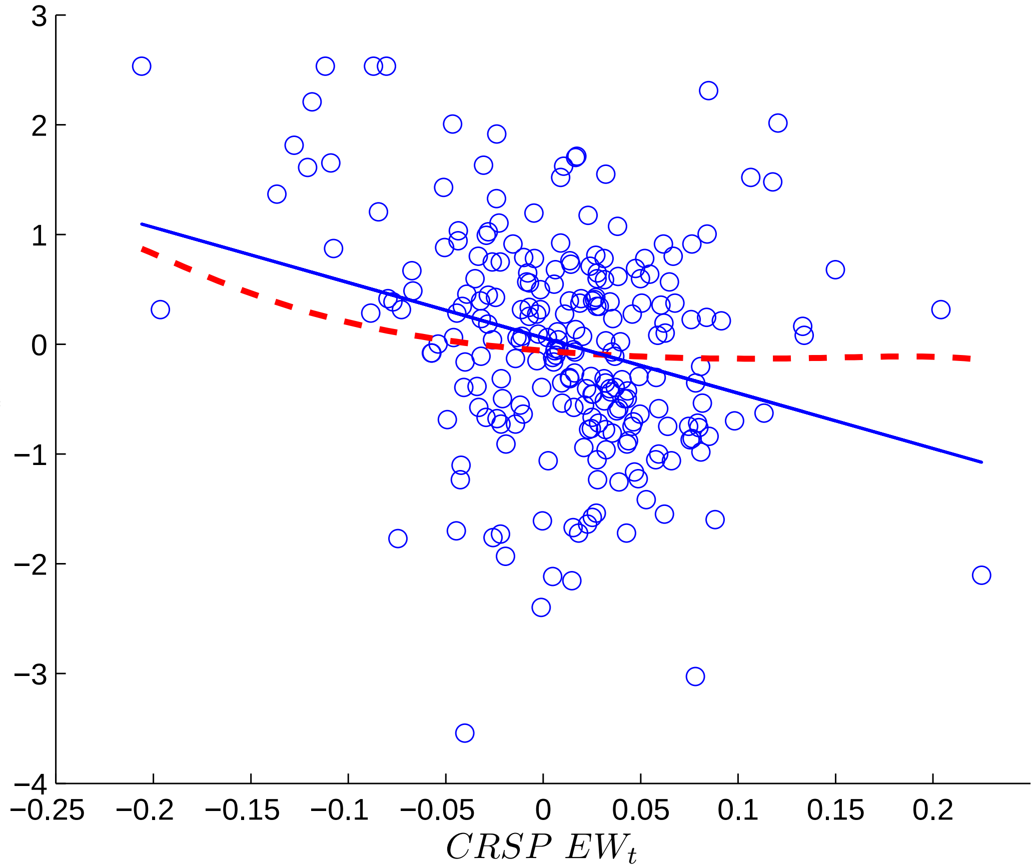
The solid bold line is the time series of ex-post Gaussian ranks of an asset with constant 0.05 monthly return. The dashed-dotted, thin solid and dotted horizontal lines represent the Gaussian ranks of a constant position at the 5%, 50% and 95% quantile of the cross-sectional distribution at each month, respectively.

Figure 2.4: Time series of positional factor estimates $\hat{F}_{p,t}$.



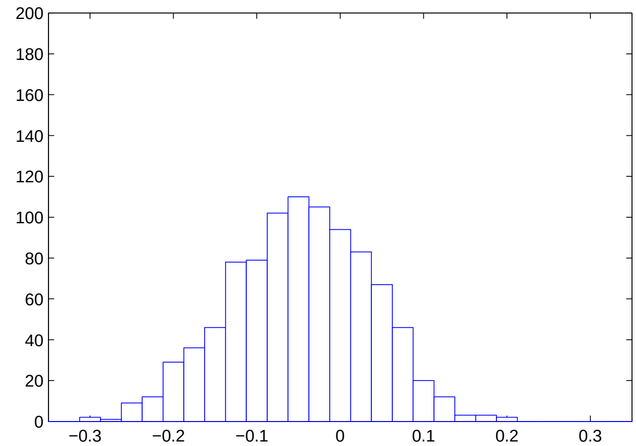
Monthly time series of estimates of the positional factor $\hat{F}_{p,t}$, obtained via the estimator in Equation (2.3.4), computed as described in Appendix A.4.

Figure 2.5: Positional factor vs. CRSP EW index returns.

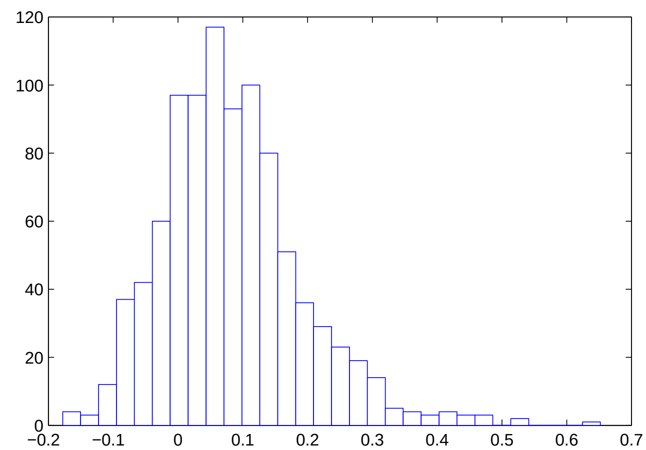


The figure displays a scatterplot of the estimates $\hat{F}_{p,t}$ of the positional factor versus the monthly returns of the equally weighted (EW) CRSP index. The solid and dashed lines correspond to a linear regression fit, and a nonparametric regression curve, respectively. The nonparametric regression curve is obtained as a kernel smoothing regression using a Gaussian kernel, with bandwidth equal to 0.0617, selected using the rule-of-thumb suggested by Bowman and Azzalini (1997).

Figure 2.6: Histograms of estimated individual effects.



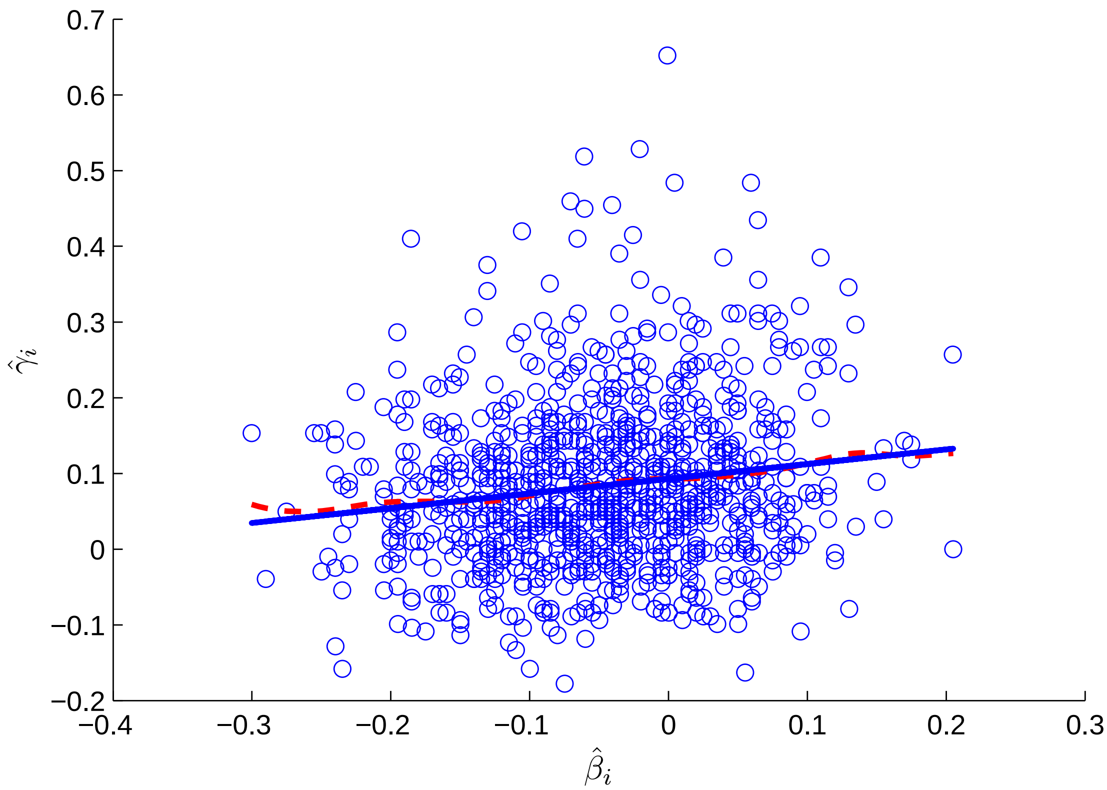
(a) $\hat{\beta}_i$



(b) $\hat{\gamma}_i$

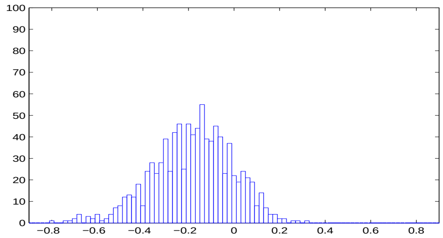
The figure displays the histograms of the estimated individual effects $\hat{\beta}_i$ and $\hat{\gamma}_i$ in panels (a) and (b), respectively. The estimates are obtained via the estimator in Equation (2.3.4), computed as described in Appendix A.4.

Figure 2.7: Scatterplot of $\hat{\gamma}_i$ vs. $\hat{\beta}_i$.

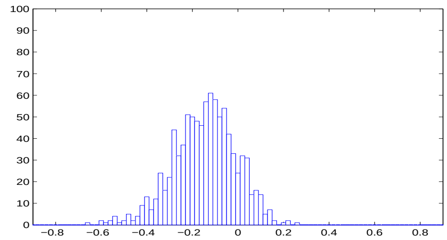


The figure displays the scatterplot of $\hat{\gamma}_i$ vs. $\hat{\beta}_i$, as well as the fitted linear regression line (solid) and the kernel smoothing regression line (dashed) corresponding to the regression of $\hat{\gamma}_i$ on $\hat{\beta}_i$. The smoothing regression is performed using a Gaussian kernel, with bandwidth equal to 0.0248, selected using the rule-of-thumb suggested by Bowman and Azzalini (1997).

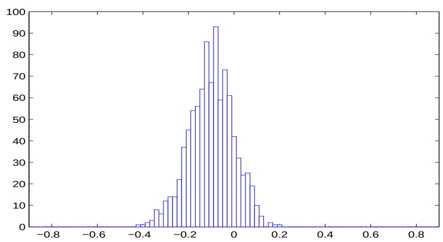
Figure 2.8: Histograms of positional persistence $\hat{\rho}_{i,t}$ as function of $F_{p,t}$.



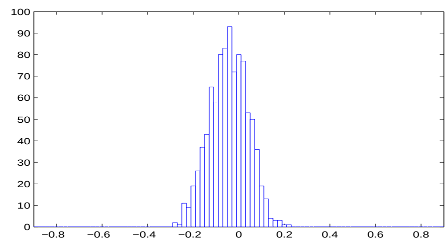
(a) $F_{p,t} = -1.7106$



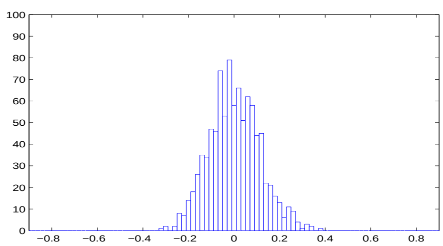
(b) $F_{p,t} = -1.2298$



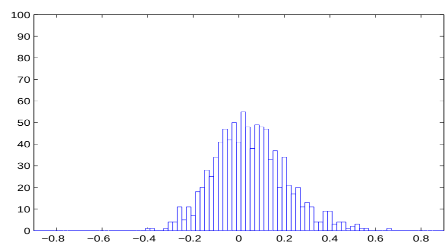
(c) $F_{p,t} = -0.6365$



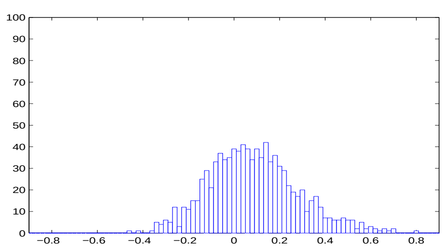
(d) $F_{p,t} = 0.0417$



(e) $F_{p,t} = 0.5985$



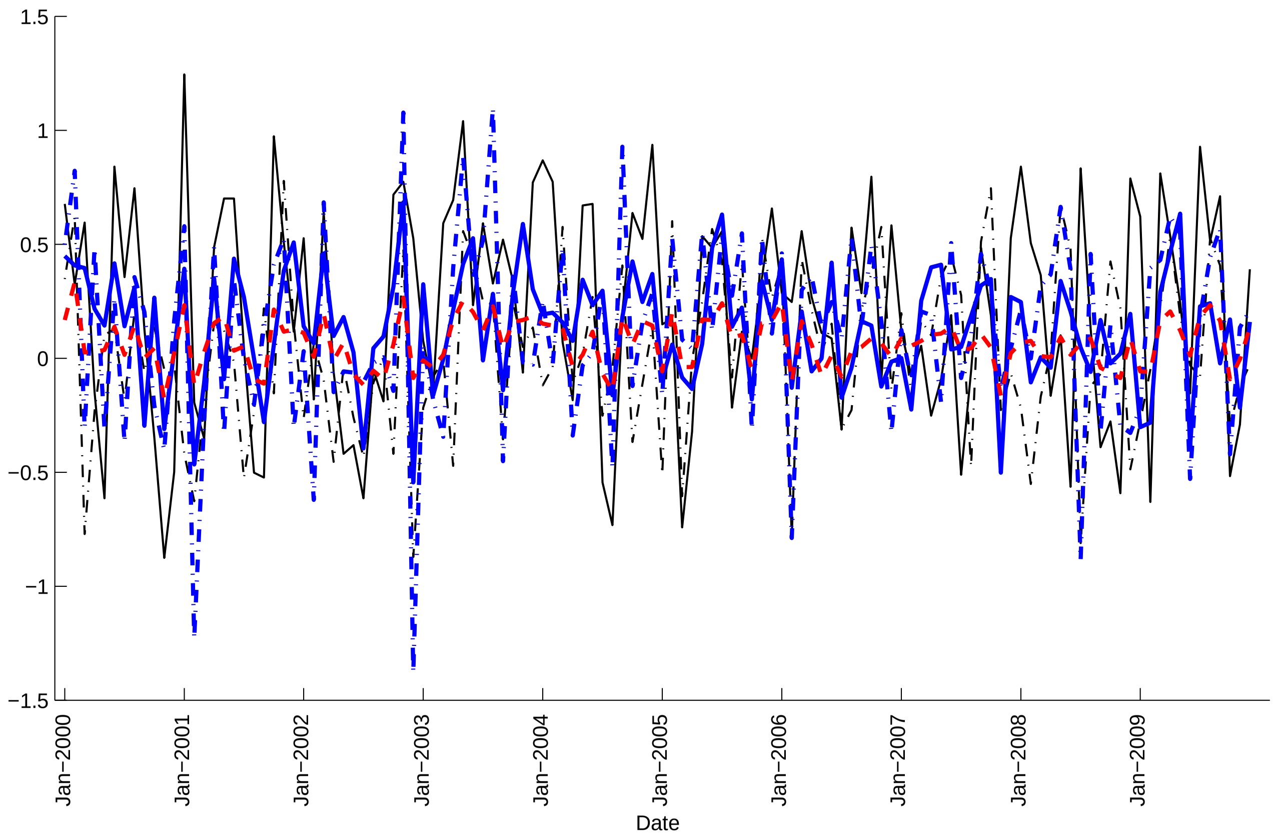
(f) $F_{p,t} = 1.2018$



(g) $F_{p,t} = 1.6792$

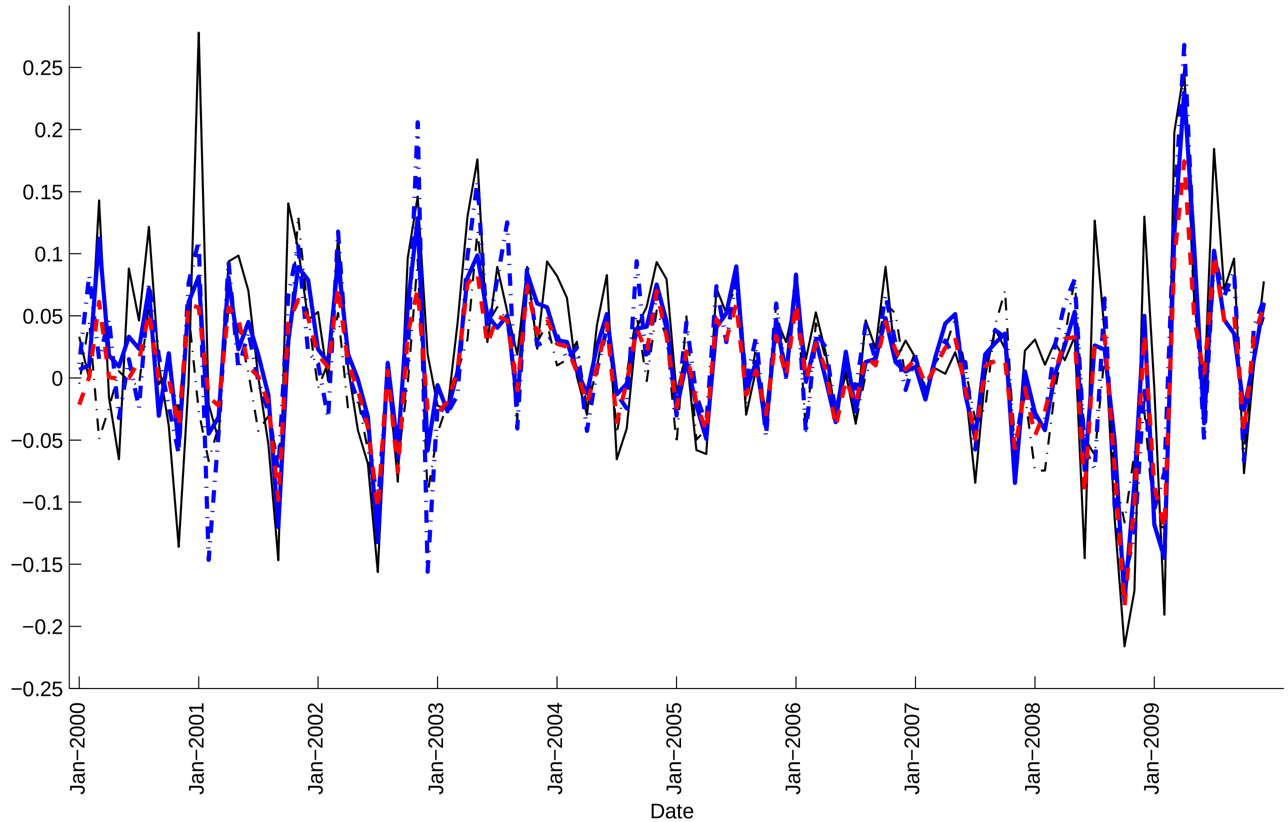
The figure displays the histograms of the positional persistence $\hat{\rho}_{i,t} = \Psi(\hat{\beta}_i + \hat{\gamma}_i F_{p,t})$ across stocks i, for different values of the positional factor $F_{p,t}$. These values of $F_{p,t}$ correspond to the 5%, 10%, 25%, 50%, 75%, 90% and 95% quantiles of the historical distribution of the positional factor in panels (a), (b), (c), (d), (e), (f) and (g), respectively.

Figure 2.9: Ex-post properties of the portfolio strategies, 2000-2009.



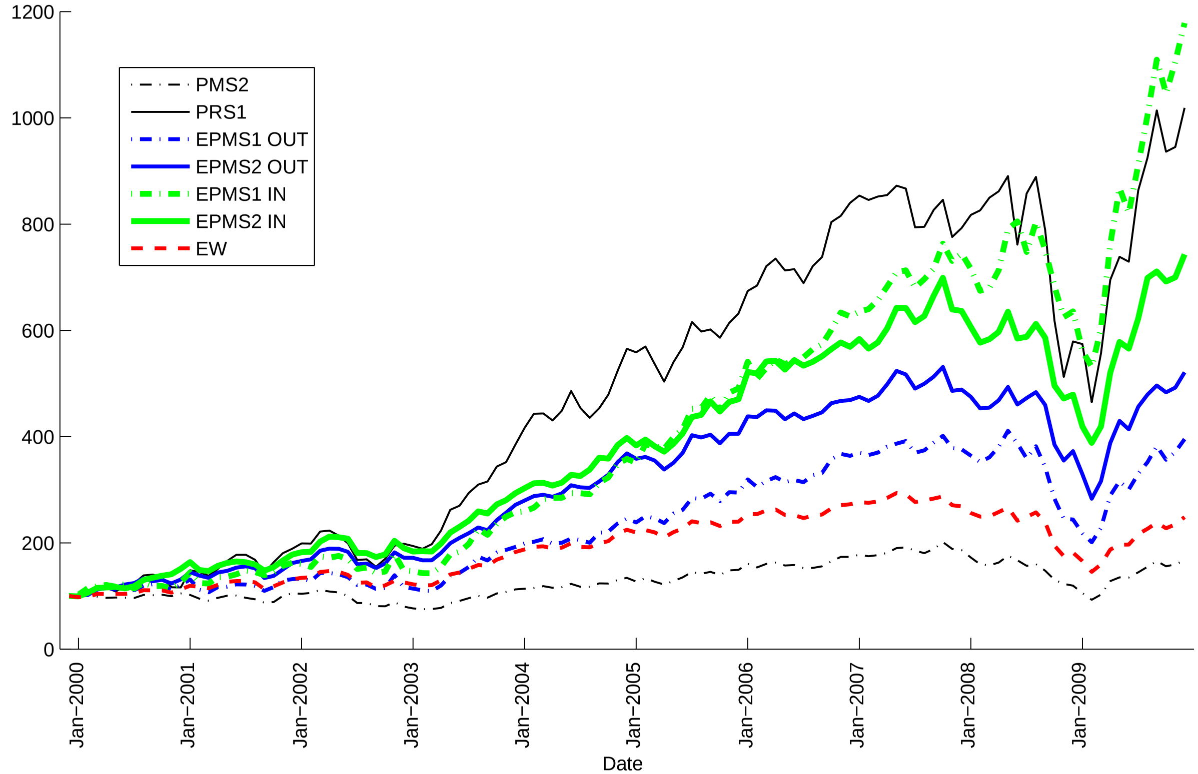
Panel (a): The figure displays the monthly time series of the Gaussian cross-sectional ranks of five portfolio strategies over the period from January 2000 to December 2009. The Gaussian ranks are computed w.r.t. the CS distribution of the monthly returns of all the NYSE, AMEX and NASDAQ stocks in our sample. The dashed-dotted thin black line corresponds to strategy PMS2, the solid thin black line corresponds to strategy PRS1, the dashed-dotted bold blue line corresponds to strategy EPMS1 OUT, the solid bold blue line corresponds to strategy EPMS2 OUT, and the dashed bold red line to the equally weighted portfolio. The strategies PMS1 and PRS2 perform worse than PMS2 and PRS1, respectively, in terms of the criteria in Table 2.1. For readability purpose, their series of Gaussian cross-sectional ranks are not displayed. The strategies are described in Section 4.2.

Figure 8: Ex-post properties of the portfolio strategies, 2000-2009.



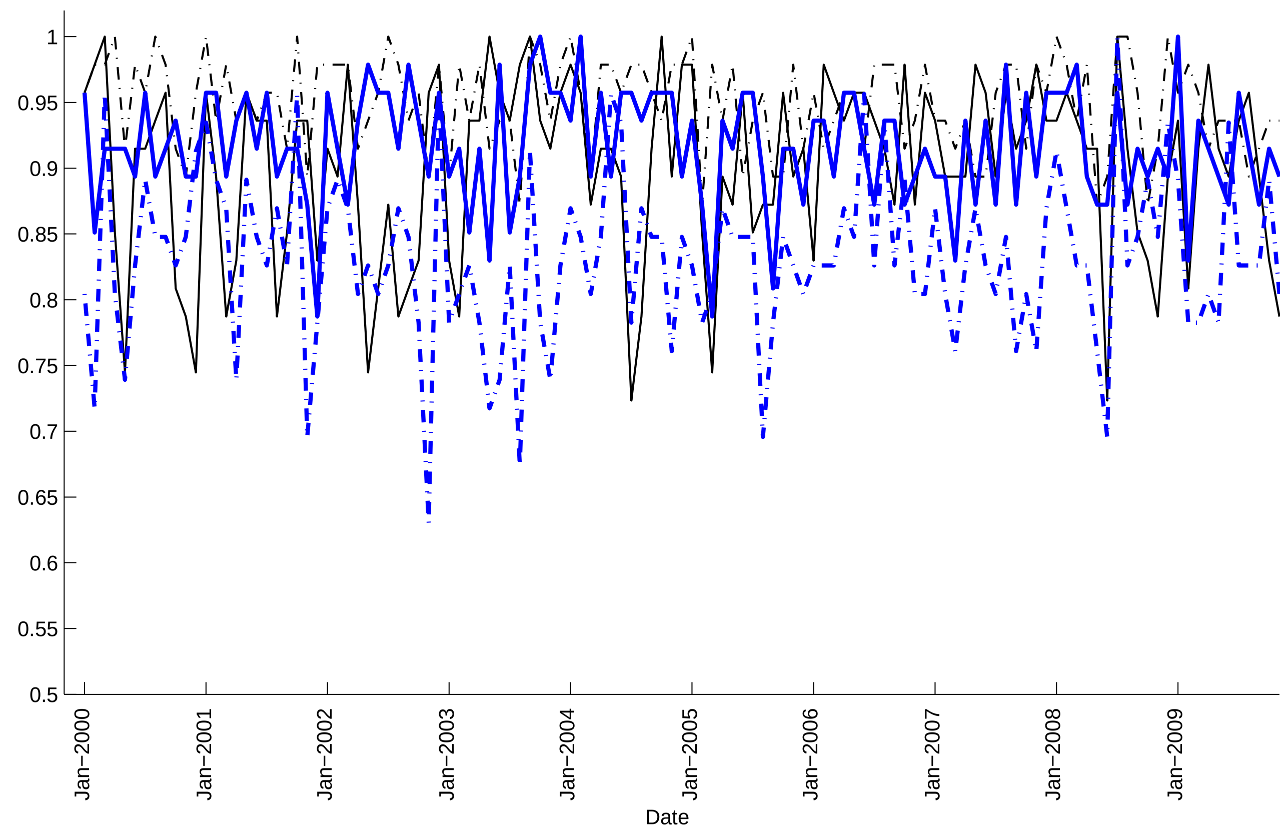
Panel (b): The figure displays the monthly time series of excess returns of five portfolio strategies over the period from January 2000 to December 2009. The dashed-dotted thin black line corresponds to strategy PMS2, the solid thin black line corresponds to strategy PRS1, the dashed-dotted bold blue line corresponds to strategy EPMS1 OUT, the solid bold blue line corresponds to strategy EPMS2 OUT, and the dashed bold red line to the equally weighted portfolio. The strategies PMS1 and PRS2 perform worse than PMS2 and PRS1, respectively, in terms of the criteria in Table 2.1. For readability purpose, their series of excess returns are not displayed. The strategies are described in Section 4.2.

Figure 8: Ex-post properties of the portfolio strategies, 2000-2009.



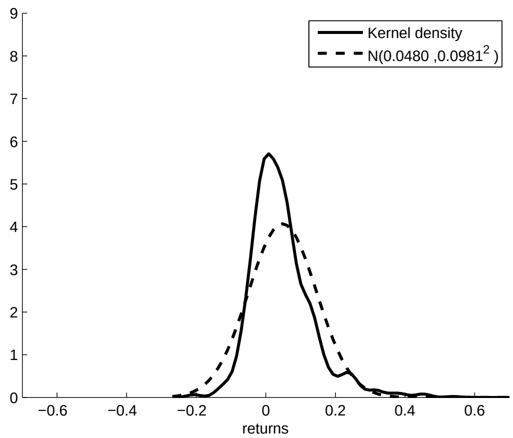
Panel (c): The figure displays the monthly time series of cumulated excess returns of five portfolio strategies over the period from January 2000 to December 2009. The dashed-dotted thin black line corresponds to strategy PMS2, the solid thin black line corresponds to strategy PRS1, the solid bold blue line corresponds to strategy EPMS2 OUT, the dashed-dotted bold blue line corresponds to strategy EPMS1 OUT, the dashed-dotted bold green line corresponds to strategy EPMS1 IN, the solid green line corresponds to strategy EPMS2 IN, and the dashed bold red line to the equally weighted portfolio. The strategies PMS1 and PRS2 perform worse than PMS2 and PRS1, respectively, in terms of the criteria in Table 2.1. For readability purpose, their series of cumulated returns are not displayed. The strategies are described in Section 4.2.

Figure 8: Ex-post properties of the portfolio strategies, 2000-2009.

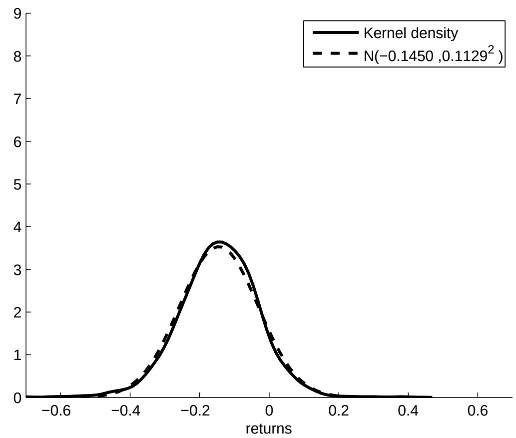


Panel (d): The figure displays the monthly time series of turnover of four portfolio strategies PMS2, PRS1, EPMS1, and EPMS2 over the period from January 2000 to December 2009. The dashed-dotted thin black line corresponds to strategy PMS2, the solid thin black line corresponds to strategy PRS1, the dashed-dotted bold blue line corresponds to strategy EPMS1 OUT, the solid bold blue line corresponds to strategy EPMS2 OUT. The turnover is computed as the proportion of selected stocks which are not kept in the portfolio between two consecutive dates. The strategies PMS1 and PRS2 perform worse than PMS2 and PRS1, respectively, in terms of the criteria in Table 2.1. For readability purpose, their turnover are not displayed. The strategies are described in Section 4.2.

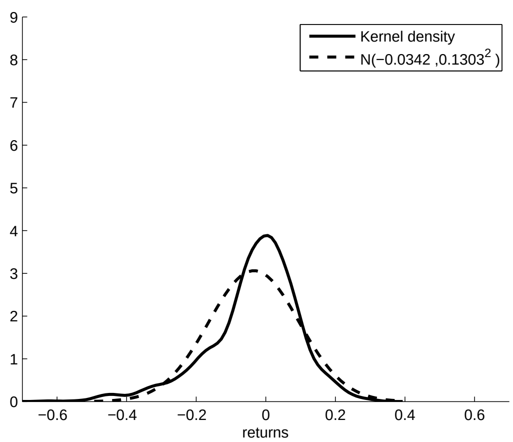
Figure 2.10: Cross-sectional distributions of monthly CRSP stock returns.



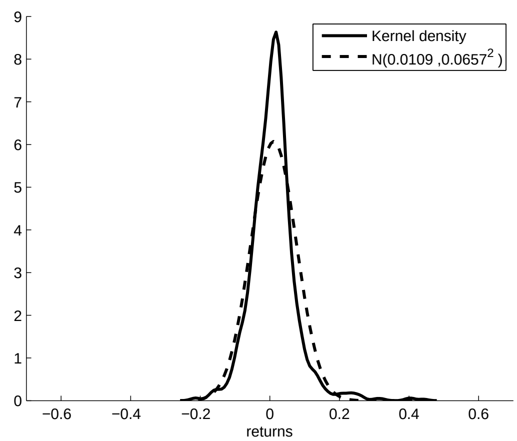
(a) Cross-sectional distribution on July 1995.



(b) Cross-sectional distribution on August 1998.



(c) Cross-sectional distribution on November 2000.

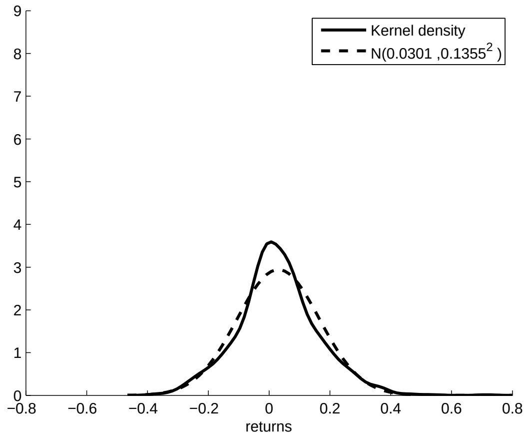


(d) Cross-sectional distribution on December 2006.

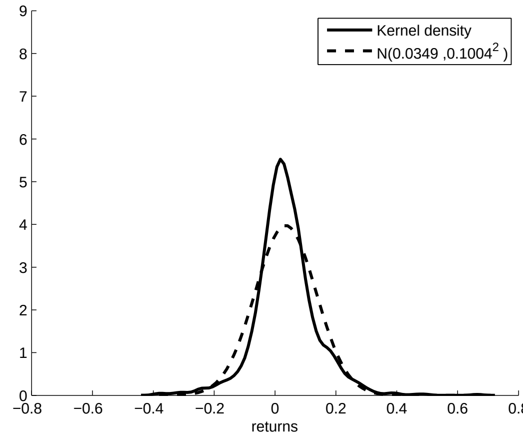
Each panel displays the kernel estimator of the CS density of the CRSP stock returns for a particular month (solid line), and compares it with a Gaussian distribution $N(\hat{\mu}_t, \hat{\sigma}_t^2)$ (dashed line), where $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ are the cross-sectional mean and variance:

$$\hat{\mu}_t = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}\,, \qquad \hat{\sigma}_t^2 = \frac{1}{n}\sum_{i=1}^{n}(y_{i,t} - \hat{\mu}_t)^2.$$

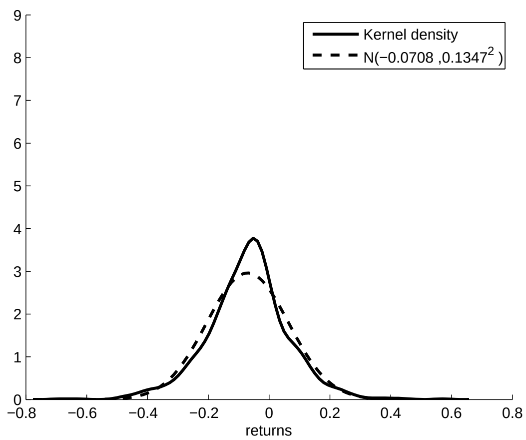
Figure 2.11: Cross-sectional distributions around the 2008 Lehman Brothers crisis.



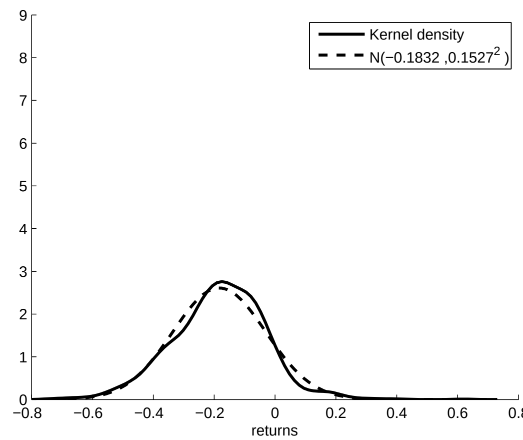
(a) Cross-sectional distribution on July 2008.



(b) Cross-sectional distribution on August 2008.



(c) Cross-sectional distribution on September 2008.

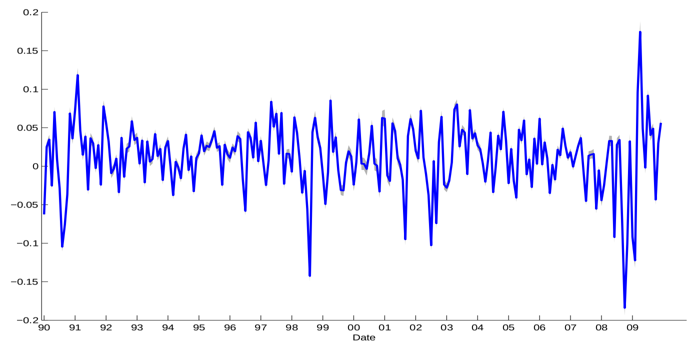


(d) Cross-sectional distribution on October 2008.

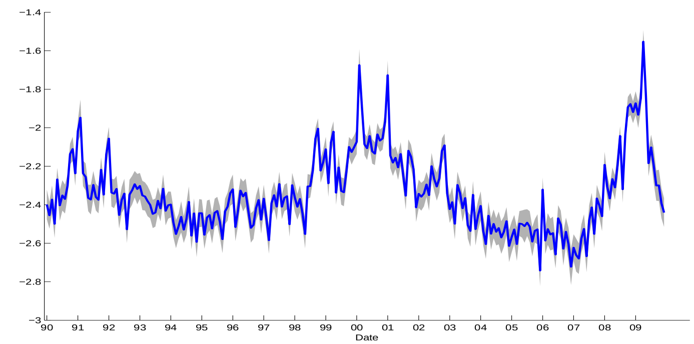
Each panel displays the kernel estimator of the CS density of the CRSP stock returns for a particular month (solid line), and compares it with a Gaussian distribution $N(\hat{\mu}_t, \hat{\sigma}_t^2)$ (dashed line), where $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ are the cross-sectional mean and variance:

$$\hat{\mu}_t = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}\,, \qquad \hat{\sigma}_t^2 = \frac{1}{n}\sum_{i=1}^{n}(y_{i,t} - \hat{\mu}_t)^2.$$

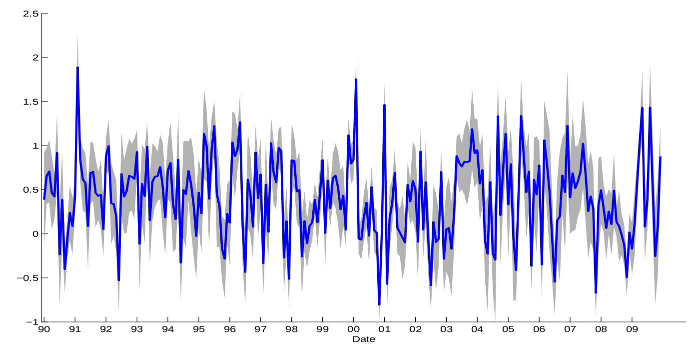
Figure 2.12: Time series of estimated distributional macro-factors.



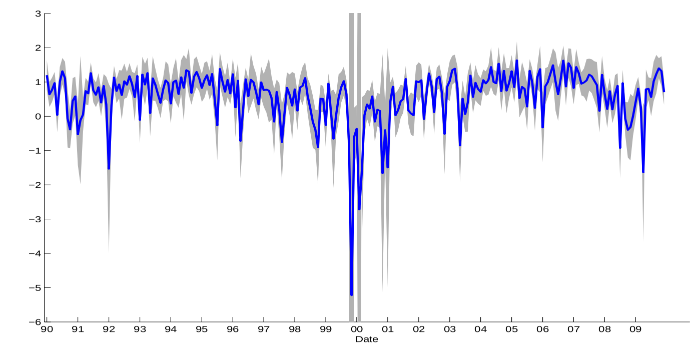
(a) CS mean ($\hat{\mu}_t$)



(b) log CS standard deviation ($\log \hat{\sigma}_t$)



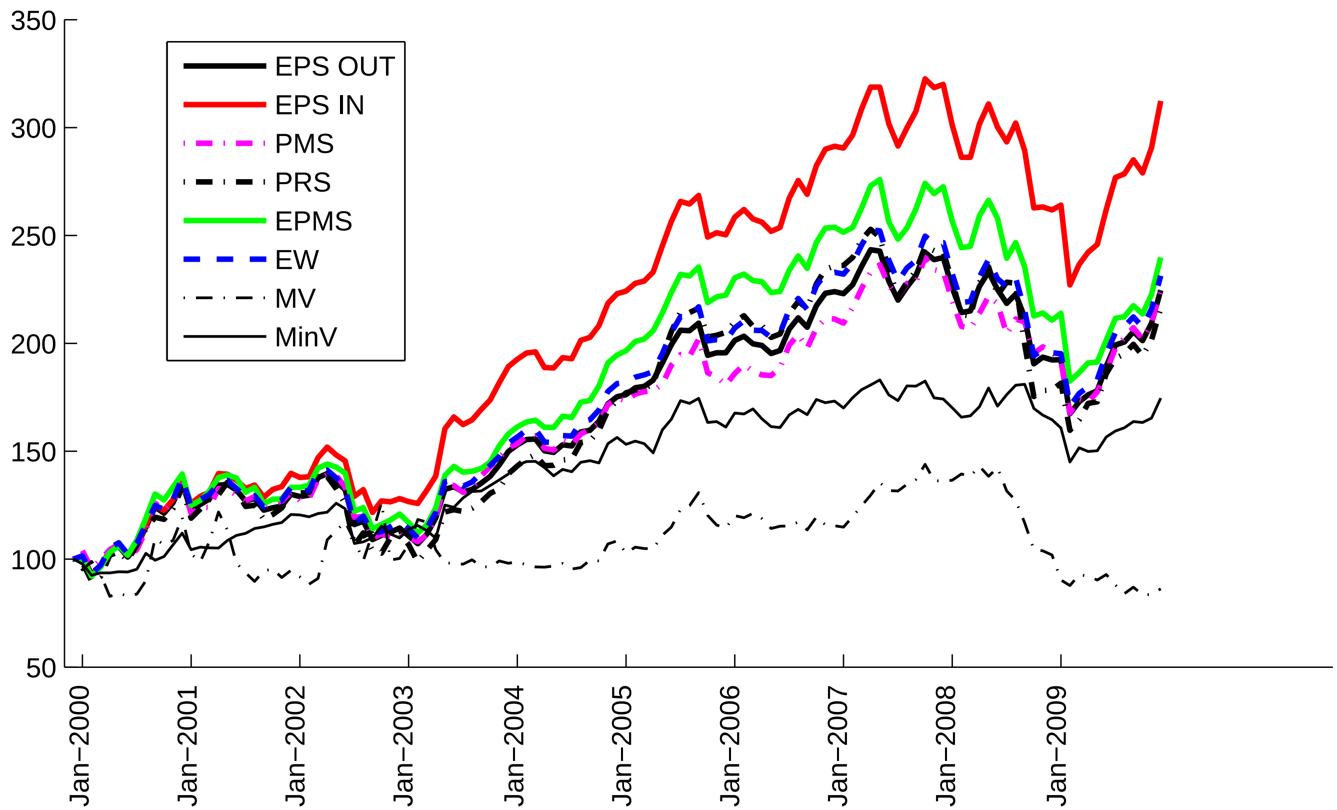
(c) CS skewness ($\hat{s}_t$)



(d) log CS excess kurtosis ($\log \hat{k}_t^*$)

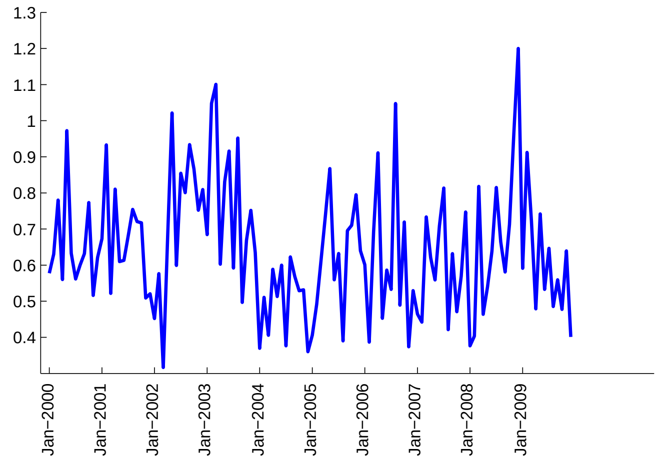
Panel (a) displays the time series of the CS mean $\hat{\mu}_t = \dfrac{1}{n}\sum_{i=1}^{n} y_{i,t}$ of the stock returns. Panel (b) displays the time series of the log CS standard deviation $\log \hat{\sigma}_t$, where $\hat{\sigma}_t^2 = \dfrac{1}{n}\sum_{i=1}^{n}(y_{i,t} - \hat{\mu}_t)^2$. Panel (c) displays the time series of the CS skewness $\hat{s}_t = \dfrac{1}{n}\sum_{i=1}^{n}(y_{i,t} - \hat{\mu}_t)^3/\hat{\sigma}_t^3$. Panel (d) displays the time series of the log CS excess kurtosis $\log \hat{k}_t^*$, where $\hat{k}_t^* = \hat{k}_t - 3(1 + \hat{s}_t^2/2)$, and $\hat{k}_t = \dfrac{1}{n}\sum_{i=1}^{n}(y_{i,t} - \hat{\mu}_t)^4/\hat{\sigma}_t^4$. The grey areas represent 95% pointwise confidence bands.

Figure 2.13: Time series of cumulative returns of the portfolio strategies in the utilities sector, 2000-2009.

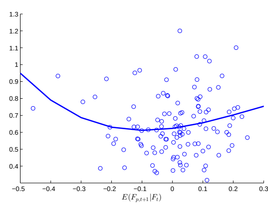


The Figure displays the cumulated excess returns of the eight portfolio strategies based on 57 stocks in the utilities sector over the period from January 2000 to December 2009. The solid bold black line corresponds to strategy EPS OUT, the solid bold red line corresponds to strategy EPS IN, the bold dash-dotted magenta line corresponds to strategy PMS, the bold dash-dotted black line corresponds to strategy PRS, the bold solid green line corresponds to strategy EPMS, the bold dashed blue line to the equally weighted portfolio, the thin solid black line corresponds to the minimum-variance strategy and the thin black dash-dotted line corresponds to the mean-variance portfolio.

Figure 2.14: Observed measure of relative discrepancy of optimal positional allocation from EW portfolio for the subsample of utilities.



(a) $n\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(a^*_{i,t}-\dfrac{1}{n})^2}$



(b) $n\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(a^*_{i,t}-\dfrac{1}{n})^2}$ vs. $E(F_{p,t+1}|F_t)$

Panel (a) displays the time series of the observed measure of relative discrepancy of the optimal positional allocation from the EW portfolio $n\sqrt{\frac{1}{n}\sum_{i=1}^{n}(a^*_{i,t}-\frac{1}{n})^2}$ for the subsample of utilities. Panel (b) displays the scatterplot of $n\sqrt{\frac{1}{n}\sum_{i=1}^{n}(a^*_{i,t}-\frac{1}{n})^2}$ vs. $E(F_{p,t+1}|F_t)$, as well as the 5-th order polynomial fit of the data (solid blue line). The conditional expectation $E(F_{p,t+1}|F_t)$ is computed by using the estimated VAR(1) model of the macro-factor process (see Section 2.5.2).

2.10 Appendix A

Appendix A.1: The factor model of stock returns and the ex-ante ranks

In this appendix we describe more formally the factor structure of stock returns introduced in Section 2, and give the definition of ex-ante ranks. We assume a factor structure for asset returns as in the assumption below.

Assumption A. 1. *i) The individual return histories $y_i = (y_{i,t})$, with $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) conditionally on the path of an unobservable factor (F_t).*

ii) The conditional distribution of the return $y_{i,t}$ given the past return history $\underline{y_{i,t-1}} = (y_{i,t-1}, y_{i,t-2}, ...)$ and the entire factor path (F_t) depends on the latter by means of the current and past factor values $\underline{F_t} = (F_t, F_{t-1}, ...)$ only.

The factor F_t can be multidimensional and corresponds to systematic, or common, risks. When the unobservable factor path (F_t) is integrated out, the individual asset returns histories become dependent. Under Assumption A.1 *i)*, the factor process (F_t) fully captures the dependence across assets returns. Assumption A.1 *ii)* implies that the conditional distribution of F_t given the past histories of the factor $\underline{F_{t-1}}$ and the returns $\underline{y_{i,t-1}}$, $i = 1, ..., n$, is independent of the latter, that is, the factor process is exogenous.

The unconditional distribution of assets returns is exchangeable, that is, invariant to asset permutations. This property corresponds to the ex-ante homogeneity of the population of assets. However, the assets are ex-post heterogeneous, as they have different distributions conditional on the past return histories. Indeed, under Assumption A.1, the model is compatible with assets having different individual unobservable characteristics (such as the factor sensitivities and idiosyncratic volatilities for stocks, or the manager's skill for fund portfolios) and the past return histories are informative for these individual unobservable characteristics.

Assumption A. 2. *The process (F_t) is strictly stationary and Markov.*

Under Assumption A.1, the returns at date t, that are $y_{1,t}, \ldots, y_{n,t}$, are conditionally i.i.d. variables admitting a cumulative distribution function (c.d.f.) H_t^* defined by $H_t^*(y) = \mathbb{P}(y_{i,t} \leq y | \underline{F_t})$. The distribution H_t^* is conditional on the current and past realizations $\underline{F_t} = (F_t, F_{t-1}, \ldots)$ of the systematic factor.

Assumption A. 3. *The cross-sectional returns c.d.f. H_t^* is continuous and strictly increasing.*

Under Assumption A.3, at any date t there is a one-to-one mapping between the stock returns and the ex-ante ranks, that are defined next.

Definition 1. *i) The uniform ex-ante ranks are defined as $u_{i,t}^* = H_t^*(y_{i,t})$.*

ii) The Gaussian ex-ante ranks are defined as $u_{i,t} = \Phi^{-1}(u_{i,t}^*) = H_t(y_{i,t})$, where Φ denotes the c.d.f. of the standard normal distribution, and $H_t = \Phi^{-1} \circ H_t^*$.

The ex-ante uniform ranks (resp. the ex-ante Gaussian ranks) at a given date are conditionally i.i.d. variables with cross-sectional uniform distribution on the interval $[0, 1]$ (resp. a standard Gaussian distribution).

The model introduced in Sections 2-5 can be cast in the framework of Assumptions A.1 - A.3 with multiple factor $F_t = (F_{d,t}', F_{p,t}')'$. The specification is such that the CS distribution $H_t^*(\cdot)$ depends on the current value of component $F_{d,t}$, and belongs to the Variance-Gamma family (Section 2.5.1 and Appendix

A.5). The component $F_{p,t}$ drives the positional persistence of the Gaussian ranks (Section 2.3.1). The unobservable characteristics $\delta_i = (\beta_i, \gamma_i)'$ introduce heterogeneity in the positional persistence of stocks (Section 3.1).

Appendix A.2: Positional management strategies

i) Derivation of the optimal positional allocation

In this section we derive the optimal positional allocation. For given budget w_r allocated in the risky assets, the future value of the risky part of the portfolio is $w_r + \gamma' y_{t+1} = w_r(\alpha' y_{t+1} + 1)$, where the dollar allocations vector γ (resp., the relative allocations vector α) is such that $\gamma' e = w_r$ (resp., $\alpha' e = 1$). The rank of this future portfolio value has to be computed with respect to the cross-sectional distribution of the values at month $t+1$ of portfolios with budget w_r invested at month t in any single risky asset i. These values are $w_r(y_{i,t+1} + 1)$ and their cross-sectional distribution is:

$$\tilde{H}_{t+1}(w) = \mathbb{P}[w_r(y_{i,t+1} + 1) \leq w | \underline{F_{t+1}}] = \mathbb{P}[y_{i,t+1} \leq w/w_r - 1 | \underline{F_{t+1}}] = H^*_{t+1}(w/w_r - 1),$$

where H^*_{t+1} is the cross-sectional distribution of stock returns at month $t+1$. Thus, the Gaussian rank of the future value of the risky part of the portfolio with dollar allocation γ is given by:

$$\begin{aligned} \Phi^{-1}\left[\tilde{H}_{t+1}(w_r + \gamma' y_{t+1})\right] &= \Phi^{-1}\left[H^*_{t+1}(\gamma' y_{t+1}/w_r)\right] \\ &= H_{t+1}(\gamma' y_{t+1}/w_r) = H_{t+1}(\alpha' y_{t+1}). \end{aligned}$$

The optimal positional dollar allocation γ^*_t is obtained by maximizing the expected positional utility of the Gaussian rank of the future portfolio value subject to the budget constraint:

$$\gamma^*_t = \underset{\gamma:\, \gamma' e = w_r}{\arg\max}\; E_t\left[\mathcal{U}\left(H_{t+1}(\gamma' y_{t+1}/w_r)\right)\right].$$

The solution is $\gamma^*_t = w_r \alpha^*_t$, where the optimal positional relative allocation α^*_t is given in equation (2.2.4).

ii) Aggregation of ranks

In this section we discuss the criterion function for positional allocation in terms of aggregation of ranks. There exist two ways to aggregate ranks. *a)* Let us consider a set of weights $\pi_1, \ldots, \pi_n$ with $\pi_i \geq 0$, for $i = 1, \ldots, n$, and $\sum_{i=1}^{n} \pi_i = 1$. It is usual to aggregate ranks by considering either the quantity $\sum_{i=1}^{n} \pi_i u^*_{i,t}$, or the quantity $\sum_{i=1}^{n} \pi_i u_{i,t}$. This ad-hoc approach is frequently used, for instance for selecting the stocks to include in a market index with a given number of assets in order to account jointly for the capitalization of the last month, the capitalization of the last three months and different liquidity measures. It has also been suggested in the latest draft released by the Basel Committee on Banking Supervision (BCBS, Basel Committee on Banking Supervision (2013)) to aggregate the scores for five categories of importance of risks:

size, cross-jurisdictional activity, interconnectedness, substitutability/financial institution infrastructure and complexity. *b)* An alternative consists in considering the rank of the associated weighted returns:

$$H_t^*\left(\sum_{i=1}^n \pi_i y_{i,t}\right) = H_t^*\left(\sum_{i=1}^n \pi_i H_t^{*-1}(u_{i,t}^*)\right), \tag{2.10.1}$$

or:

$$H_t\left(\sum_{i=1}^n \pi_i y_{i,t}\right) = H_t\left(\sum_{i=1}^n \pi_i H_t^{-1}(u_{i,t})\right). \tag{2.10.2}$$

We get H_t^*- and H_t-means of the individual ranks, respectively, instead of the time independent arithmetic means used in the first approach.

The second definition is more appealing in our framework. Indeed the set of weights π_i, $i = 1, \ldots, n$ can be considered as a portfolio allocation, for instance a portfolio of stocks, or a fund of funds. The average return $\sum_{i=1}^n \pi_i y_{i,t}$ is the portfolio (resp. fund of funds) return and is used to rank the new portfolio among the initial assets, that are the basic stocks (resp. funds). Moreover, definitions (2.10.1) and (2.10.2) are easily extended to negative π_i, or to π_i which are not summing up to 1. This is not the case with the first definition, since $\sum_{i=1}^n \pi_i u_{i,t}^*$ might be outside the unit interval $[0, 1]$ for negative weights for instance. It is seen that the second definition of rank aggregation corresponds to the criterion function in equation (2.2.5).

iii) Nash equilibrium with endogenous cross-sectional distribution

In this section we describe the Nash equilibrium for the positional asset allocation problem. The cross-sectional distribution that defines the position of the portfolio value of a given investor is no more considered as exogenous. Instead, this distribution becomes endogenous and is determined at the equilibrium by the portfolio allocations of the population of investors.

Let us consider a continuum of portfolio managers, indexed by j, with $j \in [0, 1]$. To simplify the exposition, let us assume that these managers have the same budget, normalized to 1, but different positional risk aversion parameters $\mathcal{A}_j$. These parameters admit a distribution $Q(d\mathcal{A})$. Each manager j allocates her budget in a portfolio of risky assets, with allocation vector α_j and portfolio return $\alpha_j' y_t$. We denote by $G(dy_t)$ the (exogenous) conditional distribution of the vector of asset returns, where for expository purpose we omit the time index $t-1$ of the conditional information. The distribution G is defined by the joint model of the assets returns (see Assumptions A.1-A.2 in Appendix A.1).

When solving the asset allocation problem, a given manager j considers the expected positioning of her portfolio with respect to the distribution of the portfolio values of the other managers. In the Nash equilibrium, the portfolio allocations of the other managers are considered as given by manager j. Let us denote by $\mathcal{H}_j$ the cumulative distribution function of the portfolio values of the other managers. The portfolio allocation problem of manager j becomes:

$$\begin{aligned}\alpha_j^* &= \arg\max_{\alpha:\alpha' e=1} E_G\left[\mathcal{U}(\mathcal{H}_j(\alpha' y_t); \mathcal{A}_j)\right] \\ &= \arg\max_{\alpha:\alpha' e=1} \int \mathcal{U}(\mathcal{H}_j(\alpha' y_t); \mathcal{A}_j) G(dy_t). \end{aligned} \tag{2.10.3}$$

Let us further assume that all managers have the same prior on the portfolio strategies of their competitors. Since any manager is negligible compared to the totality of the other managers, this assumption implies that the distribution $\mathcal{H}_j = \mathcal{H}$ is independent of j. Then, the solution of the constrained maximization problem (2.10.3) is such that:

$$\alpha_j^* = \alpha^*(\mathcal{A}_j, G, \mathcal{H}), \text{ say},$$

for some given function α^* of the positional risk aversion parameter, the asset returns distribution and the cross-sectional portfolio values distribution.

Let us now derive the Nash equilibrium condition.[12] The future portfolio value for manager j is $\alpha^*(\mathcal{A}_j, G, \mathcal{H})' y_t$. This portfolio value is stochastic due to its dependence on the positional risk aversion $\mathcal{A}_j$, with distribution Q, and on the asset returns vector y_t, with distribution G. Then, the distribution $\mathcal{V}$, say, of the portfolio value in the population of managers is:

$$\mathcal{V}(w) = \mathbb{P}_{Q,G}[\alpha^*(\mathcal{A}_j, G, \mathcal{H})' y_t \leq w] = \int 1\{\alpha^*(\mathcal{A}_j, G, \mathcal{H})' y_t \leq w\} Q(d\mathcal{A}_j) G(dy_t).$$

This distribution depends on Q, G and $\mathcal{H}$, and let us make this dependence explicit by writing $\mathcal{V}(\cdot; Q, G, \mathcal{H})$, say. Then, the Nash equilibrium condition requires that the prior distribution $\mathcal{H}$ corresponds to the distribution derived ex-post, that is:

$$\mathcal{H} = \mathcal{V}(\cdot; Q, G, \mathcal{H}). \tag{2.10.4}$$

The equilibrium distribution $\mathcal{H}^* = \mathcal{H}^*(Q, G)$ is obtained by solving the functional fixed-point equation (2.10.4). The discussion of the existence and uniqueness of the solution is beyond the scope of this paper.

Appendix A.3: ANOVA on Gaussian ranks

In order to motivate empirically the dynamic model for the Gaussian ranks with time variation and individual heterogeneity in positional persistence introduced in Section 3, let us perform a descriptive analysis of the empirical Gaussian rank processes. We consider the two-way panel regression:

$$\hat{u}_{i,t} \quad = \quad a + b_i + c_t + e_{i,t}, \tag{2.10.5}$$

where the empirical Gaussian ranks are explained in terms of a constant a, individual specific effects b_i, time specific effects c_t and disturbances $e_{i,t}$. For identification purpose, we set $\sum_{i=1}^{n} b_i = \sum_{t=1}^{T} c_t = 0$. The importance of individual and time effects to explain cross-sectional and time series variation of the Gaussian ranks can be assessed by testing the null hypotheses $H_0^1 : \{b_i = 0 \text{ for all } i\}$, $H_0^2 : \{c_t = 0 \text{ for all } t\}$, and the joint hypothesis $H_0^3 : \{b_i = 0 \text{ and } c_t = 0 \text{ for all } i \text{ and } t\}$. The values of the Fisher statistics $\mathcal{F}$ for the three hypotheses are provided below along with their corresponding critical values $\mathcal{F}^*$ at 95% level.

Eq. (2.10.5)	H_0^1	H_0^2	H_0^3
$\mathcal{F}$	0.725	0.005	0.578
$\mathcal{F}^*$	1.077	1.155	1.069

[12] Our approach can be compared with the equilibrium defined for two periods and two fund managers in Goriaev, Palomino, and Prat (2001).

The Fisher statistics fail to reject the three null hypotheses H_0^1, H_0^2 and H_0^3. This descriptive analysis suggests that the rank processes feature neither individual, nor time effects in their levels. The estimate of parameter a is 0.0014. The absence of time effects, and a small estimate of parameter a, were expected since the cross-sectional distribution of the Gaussian ranks $u_{i,t}$ is standard Gaussian at every date t.

In order to test for individual and time effects in positional persistence, we next consider the regression:

$$(\hat{u}_{i,t} - \bar{\hat{u}}_{i,\cdot})(\hat{u}_{i,t-1} - \bar{\hat{u}}_{i,\cdot,-1}) \quad = \quad a + b_i + c_t + e_{i,t}, \tag{2.10.6}$$

where $\bar{\hat{u}}_{i,\cdot} = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \hat{u}_{i,t}$ is the time average of the Gaussian ranks of stock i, and similarly for $\bar{\hat{u}}_{i,\cdot,-1}$. The explained variable in this regression is the cross-product of demeaned individual ranks at consecutive dates. We test the three hypotheses H_0^1, H_0^2 and H_0^3. The results for the test statistics are displayed next and show the presence of both individual and time effects in positional persistence.

Eq. (2.10.6)	H_0^1	H_0^2	H_0^3
$\mathcal{F}$	1.433	6.088	2.375
$\mathcal{F}^*$	1.077	1.155	1.069

Thus, in Section 3 we focus on the modelling of the positional persistence parameters.

Appendix A.4: The dynamics of ranks

i) Strict stationarity of the rank processes

Let us consider the rank dynamics in equations (2.3.1) and (2.3.2), and assume that the common factor $(F_{p,t})$ is a strictly stationary process (see Assumption A.2). Then, for any asset i, the rank process $(u_{i,t})$ is strictly stationary. Indeed, conditionally on any value $\delta_i = (\beta_i, \gamma_i)'$ of the random individual effect, the strict stationarity condition for a stochastic autoregressive process [see e.g. Bougerol and Picard (1992)], namely: $E\left[\log|\rho_{i,t}| \;|\; \delta_i\right] < 0$, is satisfied.

ii) Cross-sectional distribution of the ranks

Let us now verify that the cross-sectional distribution of the Gaussian ranks $u_{i,t}$, for i varying at date t, implied by equations (2.3.1) and (2.3.2) is standard Gaussian. By solving backward the autoregressive equation (2.3.1), we get an infinite-order Moving Average $MA(\infty)$ representation for process $u_{i,t}$, that is,

$$u_{i,t} = \sum_{\ell=0}^{\infty} \pi_{i,t}(\ell)\varepsilon_{i,t-\ell},$$

where the moving average coefficients $\pi_{i,t}(0) = \rho_{i,t}$ and $\pi_{i,t}(\ell) = \rho_{i,t}\rho_{i,t-1}...\rho_{i,t-\ell+1}\sqrt{1-\rho_{i,t-\ell}^2}$, for $\ell \geq 1$, are time-varying and stock-specific. Since the disturbances $(\varepsilon_{i,t})$ are independent Gaussian white noises and $\sum_{\ell=0}^{\infty} \pi_{i,t}(\ell)^2 = 1$, we get that variable $u_{i,t}$ admits a standard Gaussian $N(0,1)$ distribution conditional on the factor path (F_t) and individual heterogeneity δ_i. This implies that $u_{i,t}$ admits a standard Gaussian distribution conditional on the factor path only.

iii) A simple sequential updating procedure for numerical computation of the fixed effects estimators in equations (2.3.4)-(2.3.5)

Let us now provide an algorithm with a small degree of numerical complexity for computing of the fixed effects estimates of the factor values $F_{p,t}$, for $t = 1, ..., T$, and the individual effects β_i and γ_i for $i = 1, ..., n$ defined in equations (2.3.4)-(2.3.5). The Lagrangian function of the constrained maximization problem is:

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{i=1}^{n} \phi(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}) - \lambda \sum_{t=1}^{T} F_{p,t} - \mu \sum_{t=1}^{T} F_{p,t}^2,$$

where:

$$\phi(z, w; \rho) = -\frac{1}{2} \log(1 - \rho^2) - \frac{(z - \rho w)^2}{2(1 - \rho^2)},$$

$\rho_{i,t} = \Psi(\beta_i + \gamma_i F_{p,t}) = \Psi(\delta_i' x_t)$, with $\delta_i = (\beta_i, \gamma_i)'$ and $x_t = (1, F_{p,t})'$, as in equation (2.3.2), and λ and μ are the Lagrange multipliers for the constraints in (2.3.5). The first-order conditions for $F_{p,t}$, $t = 1, ..., T$ and δ_i, $i = 1, ..., n$ are given by:

$$\sum_{i=1}^{n} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}) \psi_{i,t} \gamma_i - \lambda - 2\mu F_{p,t} = 0, \quad t = 1, ..., T, \tag{2.10.7}$$

$$\sum_{t=1}^{T} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}) \psi_{i,t} x_t = 0, \quad i = 1, ..., n, \tag{2.10.8}$$

respectively, where $\psi_{i,t} = \Psi'(\beta_i + \gamma_i F_{p,t})$ and the partial derivative of the function ϕ w.r.t. ρ is given by:

$$\frac{\partial \phi}{\partial \rho}(z, w; \rho) = \frac{1}{1 - \rho^2} \left\{ (z - \rho w) w - \rho \left[\frac{(z - \rho w)^2}{1 - \rho^2} - 1 \right] \right\}.$$

By summing the equations in (2.10.7) over $t = 1, ..., T$, we get:

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}) \psi_{i,t} \gamma_i - T\lambda - 2\mu \sum_{t=1}^{T} F_{p,t} = 0.$$

The first term (resp. the third term) in the equation is equal to 0 from (2.10.8) [resp. from (2.3.5)]. It follows that $\lambda = 0$. Similarly, by multiplying both sides of equation (2.10.7) by $F_{p,t}$ and summing again over $t = 1, ..., T$, we get:

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}) \psi_{i,t} F_{p,t} \gamma_i - 2\mu \sum_{t=1}^{T} F_{p,t}^2 = 0.$$

The first term in the equation is equal to 0 from (2.10.8), while we have $\sum_{t=1}^{T} F_{p,t}^2 = T$ from (2.10.8). It follows that $\mu = 0$. The Lagrange multipliers are zero since the maximized function value is the same with or without the constraints (2.3.5). Thus, the estimators can be computed from the equations:

$$\sum_{i=1}^{n} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t})\psi_{i,t}\gamma_i = 0, \quad t = 1, ..., T, \tag{2.10.9}$$

$$\sum_{t=1}^{T} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t})\psi_{i,t}x_t = 0, \quad i = 1, ..., n, \tag{2.10.10}$$

imposing the identification constraints (2.3.5). We solve the system of equations (2.10.9) - (2.10.10) by a Newton-Raphson method, in which the updating is performed sequentially with respect to time and individual effects. In contrast to the joint updating, which would require the inversion of matrices of dimension $(2n+T, 2n+T)$ and has a large degree of numerical complexity, the sequential updating simplifies considerably the computation, since it allows to update the values of the effects $F_{p,t}$, and δ_i independently across dates and individuals without matrix inversions. Specifically, let $F_{p,t}^{(q)}$, $\delta_i^{(q)}$ denote the values of the parameters at step q satisfying the constraints (2.3.5), and let $x_t^{(q)}$, $\rho_{i,t}^{(q)}$ and $\psi_{i,t}^{(q)}$ be the corresponding values of x_t, $\rho_{i,t}$ and $\psi_{i,t}$. Let us expand equation (2.10.9) for date t w.r.t. $F_{p,t}$ around the solution at step q. We have:

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q)})\psi_{i,t}^{(q)}\gamma_i^{(q)} \\
&+ \left[\sum_{i=1}^{n} \left(\frac{\partial^2 \phi}{\partial \rho^2}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q)})[\psi_{i,t}^{(q)}]^2[\gamma_i^{(q)}]^2 + \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q)})\tau_{i,t}^{(q)}[\gamma_i^{(q)}]^2\right)\right](F_{p,t} - F_{p,t}^{(q)}) \simeq 0,
\end{aligned}$$

where $\tau_{i,t}^{(q)} = \Psi''(\beta_i^{(q)} + \gamma_i^{(q)}F_{p,t}^{(q)})$. By solving the above approximate equation, the new values of the time effects up to an additive constant and a multiplicative scale are given by:

$$\begin{aligned}
\tilde{F}_{p,t}^{(q+1)} &= F_{p,t}^{(q)} - \left[\sum_{i=1}^{n} \left(\frac{\partial^2 \phi}{\partial \rho^2}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q)})[\psi_{i,t}^{(q)}]^2 + \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q)})\tau_{i,t}^{(q)}\right)[\gamma_i^{(q)}]^2\right]^{-1} \\
&\cdot \left[\sum_{i=1}^{n} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q)})\psi_{i,t}^{(q)}\gamma_i^{(q)}\right]. \qquad (2.10.11)
\end{aligned}$$

Similarly, we update at step $q+1$ the individual effects by performing a Taylor expansion of the equations in (2.10.10) w.r.t. the β_i and γ_i individual by individual, by taking into account the update of the time effects at step $q+1$:

$$\begin{aligned}
\tilde{\delta}_i^{(q+1)} &= \delta_i^{(q)} - \left[\sum_{t=1}^{T} \left(\frac{\partial^2 \phi}{\partial \rho^2}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q+1/2)})[\psi_{i,t}^{(q+1/2)}]^2\right.\right. \\
&\left.\left. + \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q+1/2)})\tau_{i,t}^{(q+1/2)}\right) x_t^{(q+1)}[x_t^{(q+1)}]'\right]^{-1} \\
&\times \left[\sum_{t=1}^{T} \frac{\partial \phi}{\partial \rho}(\hat{u}_{i,t}, \hat{u}_{i,t-1}; \rho_{i,t}^{(q+1/2)})\psi_{i,t}^{(q+1/2)}x_t^{(q+1)}\right], \qquad (2.10.12)
\end{aligned}$$

where $x_t^{q+1} = (1, \tilde{F}_{p,t}^{(q+1)})'$, $\rho_{i,t}^{(q+1/2)} = \Psi(\beta_i^{(q)} + \gamma_i^{(q)} \tilde{F}_{p,t}^{(q+1)})$ and similarly for $\psi_{i,t}^{(q+1/2)}$ and $\tau_{i,t}^{(q+1/2)}$. Finally, we get the estimates at step $q+1$ by recentering and rescaling the values in (2.10.11) - (2.10.12) to account for the constraints:

$$
\begin{aligned}
\hat{F}_{p,t}^{(q+1)} &= \frac{\tilde{F}_{p,t}^{(q+1)} - \frac{1}{T}\sum_{t=1}^{T} \tilde{F}_{p,t}^{(q+1)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\left[\tilde{F}_{p,t}^{(q+1)}\right]^2 - \frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{p,t}^{(q+1)}\right)^2}}, \quad t = 1, ..., T, \\
\hat{\gamma}_i^{(q+1)} &= \tilde{\gamma}_i^{(q+1)} \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\left[\tilde{F}_{p,t}^{(q+1)}\right]^2 - \frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{p,t}^{(q+1)}\right)^2}, \quad i = 1, ..., n, \\
\hat{\beta}_i^{(q+1)} &= \tilde{\beta}_i^{(q+1)} + \tilde{\gamma}_i^{(q+1)} \frac{1}{T}\sum_{t=1}^{T} \tilde{F}_{p,t}^{(q+1)}, \quad i = 1, ..., n.
\end{aligned}
$$

iv) Proof of equation (2.3.6)

From the assumption of Gaussian CS distribution, the future position of the risky portfolio return is [see equation (2.2.8)]:

$$
\begin{aligned}
H_{t+1}(\alpha' y_{t+1}) &= \alpha' u_{t+1} \\
&= \sum_{i=1}^{n} \alpha_i \rho_{i,t+1} u_{i,t} + \sum_{i=1}^{n} \alpha_i \sqrt{1 - \rho_{i,t+1}^2}\, \varepsilon_{i,t+1}. \qquad (2.10.13)
\end{aligned}
$$

Then, by using that the $(\varepsilon_{i,t})$ are independent Gaussian white noise processes, the expected positional utility is:

$$
\begin{aligned}
& -E\left[\exp\left(-\mathcal{A} H_{t+1}(\alpha' y_{t+1})\right) | \underline{F_t}, \underline{y_t}\right] \\
= & -E\left\{E\left[\exp\left(-\mathcal{A} H_{t+1}(\alpha' y_{t+1})\right) | \underline{F_{t+1}}, \underline{y_t}\right] | \underline{F_t}, \underline{y_t}\right\} \\
= & -E\left[\exp\left(-\mathcal{A}\sum_{i=1}^{n} \alpha_i \rho_{i,t+1} u_{i,t} + \frac{1}{2}\mathcal{A}^2 \sum_{i=1}^{n} \alpha_i^2 (1 - \rho_{i,t+1}^2)\right) | \underline{F_t}, \underline{y_t}\right].
\end{aligned}
$$

The Lagrangian function for the maximization of the expected positional utility w.r.t. the portfolio allocation α subject to the constraint $\alpha' e = 1$ is:

$$
\mathcal{L} = -E\left[\exp\left(-\mathcal{A}\sum_{i=1}^{n} \alpha_i \rho_{i,t+1} u_{i,t} + \frac{1}{2}\mathcal{A}^2 \sum_{i=1}^{n} \alpha_i^2 (1 - \rho_{i,t+1}^2)\right) | \underline{F_t}, \underline{y_t}\right] + \lambda(\alpha' e - 1),
$$

where λ is the Lagrange multiplier. The first-order condition for α_i is:

$$
\begin{aligned}
& \mathcal{A} E\left[\left(\rho_{i,t+1} u_{i,t} - \mathcal{A}(1 - \rho_{i,t+1}^2)\alpha_i\right) \exp\left(-\mathcal{A}\sum_{i=1}^{n} \alpha_i \rho_{i,t+1} u_{i,t} + \frac{1}{2}\mathcal{A}^2 \sum_{i=1}^{n} \alpha_i^2 (1 - \rho_{i,t+1}^2)\right) | \underline{F_t}, \underline{y_t}\right] \\
& + \lambda = 0.
\end{aligned}
$$

In the conditional expectation, the exponential function can be replaced by $\exp\left(-\mathcal{A}H_{t+1}(\alpha' y_{t+1})\right)$. We deduce that the solution $\alpha = \alpha_t^*$ satisfies the implicit equation:

$$\alpha_{i,t}^* = \frac{1}{\mathcal{A}^2}\frac{\lambda}{E_t^\alpha(1-\rho_{i,t+1}^2)} + \frac{1}{\mathcal{A}}\xi_{i,t}, \tag{2.10.14}$$

where $\xi_{i,t}$ is defined in equation (2.3.8). From the constraint $\sum_{i=1}^n \alpha_{i,t}^* = 1$, we get that the Lagrange multiplier is such that:

$$\frac{\lambda}{\mathcal{A}^2} = \frac{1}{\sum_{i=1}^n [E_t^\alpha(1-\rho_{i,t+1}^2)]^{-1}}\left(1 - \frac{1}{\mathcal{A}}\sum_{i=1}^n \xi_{i,t}\right).$$

By replacing this expression into equation (2.10.14), we get equation (2.3.6).

v) Portfolio with least risky future rank

In this subsection, we assume that positional persistence is time invariant, i.e. $\rho_{i,t} = \rho_i$. Then, from (2.10.13) the conditional variance of the future portfolio rank given the individual assets ranks is:

$$V\left[H_{t+1}(\alpha' y_{t+1}) \mid u_t\right] = \sum_{i=1}^n \alpha_i^2(1-\rho_i^2).$$

By maximizing this conditional variance w.r.t. α subject to the constraint $\alpha' e = 1$, we get the portfolio allocation with conditionally least risky future rank:

$$\alpha_i = \frac{(1-\rho_i^2)^{-1}}{\sum_{i=1}^n (1-\rho_i^2)^{-1}}.$$

vi) Proof of equation (2.3.9)

At first-order in the persistence parameter we have:

$$E_t^\alpha(1-\rho_{i,t+1}^2) \simeq 1,$$

and:

$$\begin{aligned} E_t^\alpha(\rho_{i,t+1}) &= \frac{E\left[\rho_{i,t+1}\exp\left(-\mathcal{A}\sum_{i=1}^n \alpha_i\rho_{i,t+1}u_{i,t} + \frac{1}{2}\mathcal{A}^2\sum_{i=1}^n \alpha_i^2(1-\rho_{i,t+1}^2)\right)|\underline{F_t}, \underline{y_t}\right]}{E\left[\exp\left(-\mathcal{A}\sum_{i=1}^n \alpha_i\rho_{i,t+1}u_{i,t} + \frac{1}{2}\mathcal{A}^2\sum_{i=1}^n \alpha_i^2(1-\rho_{i,t+1}^2)\right)|\underline{F_t}, \underline{y_t}\right]} \\ &\simeq E\left[\rho_{i,t+1}|\underline{F_t}, \underline{y_t}\right] \simeq E\left[\rho_{i,t+1}|F_t, \delta_i\right]. \end{aligned}$$

Then, from (2.3.7) and (2.3.8) we get $w_{i,t} \simeq 1/n$ and $\xi_{i,t} \simeq E_t(\rho_{i,t+1})u_{i,t}$. By plugging these approximations in (2.3.6), we get equation (2.3.9).

vii) Proof of approximation (2.6.1)

From equation (2.3.9) we have:

$$\frac{1}{n}\sum_{i=1}^{n}(\alpha_{i,t}^{*}-1/n)^2=\frac{1}{\mathcal{A}^2}\frac{1}{n}\sum_{i=1}^{n}\left(E_t(\rho_{i,t+1})u_{i,t}-\frac{1}{n}\sum_{i=1}^{n}E_t(\rho_{i,t+1})u_{i,t}\right)^2. \qquad (2.10.15)$$

When $n\to\infty$, the strong Law of Large Numbers implies:

$$\frac{1}{n}\sum_{i=1}^{n}\left(E_t(\rho_{i,t+1})u_{i,t}-\frac{1}{n}\sum_{i=1}^{n}E_t(\rho_{i,t+1})u_{i,t}\right)^2 \quad\to\quad V(E_t(\rho_{i,t+1})u_{i,t}|\underline{F_t}), \qquad (2.10.16)$$

where the convergence is almost surely. Moreover, we have:

$$V(E_t(\rho_{i,t+1})u_{i,t}|\underline{F_t})=E(E_t(\rho_{i,t+1})^2u_{i,t}^2|\underline{F_t})=E(E_t(\rho_{i,t+1})^2|F_t), \qquad (2.10.17)$$

where we use that the Gaussian ranks are such that $u_{i,t}\sim N(0,1)$ conditional on $\underline{F_t},\delta_i$, and the process of the macro-factors is Markov. Let us now approximate the quantity $E(E_t(\rho_{i,t+1})^2|F_t)$ for small positional persistence. By a first-order Taylor expansion of function $\Psi(s)=(e^{2s}-1)/(e^{2s}+1)$ around $s=0$, the conditional expected positional persistence is such that:

$$E_t(\rho_{i,t+1})\simeq\Psi(0)+\Psi'(0)[\beta_i+\gamma_iE(F_{p,t+1}|F_t)]=\beta_i+\gamma_iE(F_{p,t+1}|F_t). \qquad (2.10.18)$$

By combining (2.10.15)-(2.10.18), approximation (2.6.1) follows.

Appendix A.5: Parametrization of the Variance-Gamma distribution

The Variance-Gamma (VG) is a parametric family of distributions yielding a flexible yet tractable specification of third and fourth order moments. The VG distribution was first used in Finance by Madan and Seneta (1990) to describe the historical distribution of security returns. In our paper, we use the VG family to model the theoretical CS distribution of CRSP stock returns. The theoretical CS p.d.f. $h_t^*(y)=\partial H_t^*(y)/\partial y$ at month t is given by [see Seneta (2004), p. 180]:

$$h_t^*(y)=\frac{2\exp\{\gamma_t(y-c_t)/\omega_t\}}{\sqrt{2\pi\omega_t}(1/\lambda_t)^{\lambda_t}\Gamma(\lambda_t)}\left(\frac{|y-c_t|}{\sqrt{2\omega_t\lambda_t+\gamma_t^2}}\right)^{\lambda_t-1/2}K_{\lambda_t-1/2}\left(\frac{|y-c_t|\sqrt{2\omega_t\lambda_t+\gamma_t^2}}{\omega_t}\right), \qquad (2.10.19)$$

where $K_\lambda(\cdot)$ denotes the Bessel function of the third kind [13] with index λ, $\Gamma(\cdot)$ is the Gamma function[14] and $c_t\in\mathbb{R}$, $\omega_t>0$, $\gamma_t\in\mathbb{R}$ and $\lambda_t>0$ are the parameters for month t. The four VG parameters c_t, ω_t, γ_t, λ_t are time-varying and stochastic. They correspond to transformations of the elements of a four-dimensional common stochastic factor $F_{d,t}$ that drives the pattern of the theoretical CS distribution of stock returns. The

[13] The Bessel function of the third kind with index λ is defined as $K_\lambda(x)=\frac{1}{2}\int_0^{+\infty}t^{\lambda-1}e^{-\frac{1}{2}x(t+t^{-1})}dt$, for $x>0$.

[14] The Gamma function is defined as $\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt$, for $x>0$.

VG distribution in equation (2.10.19) is the distribution of returns $y_{i,t}$ at month t, for i varying, conditional on the observed factor $F_{d,t}$. Since the VG family of distributions can be parameterized in several alternative ways, vector $F_{d,t}$ is defined up to a one-to-one transformation. We select this transformation such that the parameters, i.e. the components of vector $F_{d,t}$, admit simple interpretations and vary without constraints in the domain $\mathbb{R}^4$. The latter condition eases the specification of a dynamic model for process (F_t) in Section 2.5.2. To define the parameter transformation, let us consider the first four standardized cross-sectional power moments at month t. They are given by [see Seneta (2004)]:

$$\begin{aligned}
\mu_t &= E[y_{i,t}|F_{d,t}] = c_t + \gamma_t, && (2.10.20)\\
\sigma_t^2 &= V[y_{i,t}|F_{d,t}] = \gamma_t^2/\lambda_t + \omega_t, && (2.10.21)\\
s_t &= \frac{E[(y_{i,t}-\mu_t)^3|F_{d,t}]}{\sigma_t^3} = \frac{\gamma_t}{\lambda_t}\frac{2\gamma_t^2/\lambda_t + 3\omega_t}{(\gamma_t^2/\lambda_t+\omega_t)^{3/2}}, && (2.10.22)\\
k_t &= \frac{E[(y_{i,t}-\mu_t)^4|F_{d,t}]}{\sigma_t^4} = 3 + \frac{3}{\lambda_t}\frac{\omega_t^2 + 4\omega_t\gamma_t^2/\lambda_t + 2\gamma_t^4/\lambda_t^2}{(\gamma_t^2/\lambda_t+\omega_t)^2}. && (2.10.23)
\end{aligned}$$

We have the following result, which is proved at the end of this Appendix.

Lemma 1. *i) In the VG family, the kurtosis k_t is lower bounded, with the lower bound depending on the skewness s_t:*

$$k_t > 3(1 + s_t^2/2). \tag{2.10.24}$$

ii) Define:

$$k_t^* = k_t - 3(1 + s_t^2/2). \tag{2.10.25}$$

Then, the parameters $\mu_t \in \mathbb{R}$, $\sigma_t > 0$, $s_t \in \mathbb{R}$ and $k_t^* > 0$ vary independently on their domains, and are jointly in a one-to-one relationship with the original parameters $c_t \in \mathbb{R}$, $\omega_t > 0$, $\gamma_t \in \mathbb{R}$ and $\lambda_t > 0$.

The inequality (2.10.24) on the third and fourth order moments of the VG distribution is more restrictive than the condition valid for any distribution, namely $k_t > 1 + s_t^2$ [see Pearson (1916)]. Moreover, in the VG model the kurtosis is larger than 3, that is, the kurtosis of a Gaussian distribution. A Gaussian distribution is the limit of the VG distribution when $s_t = 0$ and $k_t^* \to 0$ [see Seneta (2004)]. Lemma 1 *i)* suggests to consider k_t^* defined in (2.10.25) as a measure of excess kurtosis. Then, we define the factor $F_{d,t}$ as in equation (2.5.1), namely $F_{d,t} = (\mu_t, \log\sigma_t, s_t, \log k_t^*)'$. Its components are in one-to-one relationship with the parameters of the VG family from Lemma 1 *ii)*, and they are free to vary in the unbounded domain $\mathbb{R}^4$.

Proof of Lemma 1: We omit the time index of the parameters as it is not relevant here. Define the parameter transformations:

$$\xi = \frac{\gamma/\sqrt{\lambda}}{\sqrt{\gamma^2/\lambda+\omega}}, \qquad \eta = \frac{1}{\sqrt{\lambda}}. \tag{2.10.26}$$

The parameters $\mu \in \mathbb{R}$, $\sigma > 0$, $\xi \in (-1,1)$ and $\eta > 0$ vary independently on their domains, and are in a one-to-one relationship with the original parameters $c \in \mathbb{R}$, $\omega > 0$, $\gamma \in \mathbb{R}$ and $\lambda > 0$. Indeed, the original parameters can be written as $c = \mu - \xi\sigma/\eta$, $\omega = \sigma^2(1-\xi^2)$, $\gamma = \xi\sigma/\eta$ and $\lambda = 1/\eta^2$. Moreover, the skewness and kurtosis can be written as:

$$\begin{aligned}
s &= \eta\xi(3-\xi^2), && (2.10.27)\\
k &= 3 + 3\eta^2(1+2\xi^2-\xi^4), && (2.10.28)
\end{aligned}$$

and are functions of parameters η and ξ only. From equation (2.10.27), when $\xi \neq 0$ we have $\eta = s/[\xi(3 - \xi^2)]$. By replacing this expression of η into equation (2.10.28) we get:

$$k = 3 + 3s^2 g(\xi^2), \tag{2.10.29}$$

where function g is defined by $g(z) = \dfrac{1 + 2z - z^2}{z(3-z)^2}$, for $z > 0$. The function g is monotonic decreasing on the interval $(0,1)$, with $g(z) \to \infty$ as $z \to 0$ and $g(1) = 1/2$. We deduce inequality (2.10.24).

Defining k^* as $k^* = k - 3(1 + s^2/2)$, the parameters $\mu \in \mathbb{R}$, $\sigma > 0$, $s \in \mathbb{R}$ and $k^* > 0$ are in a one-to-one relationship with the original parameters. Indeed, given the values of $s \in \mathbb{R}$ and $k^* > 0$, we can determine uniquely the values of $\eta > 0$ and $\xi \in (-1,1)$:
i) If $s = 0$, from equations (2.10.27), (2.10.28) and the definition of k^* it follows $\xi = 0$ and $\eta = \sqrt{k^*/3}$.
ii) If $s \neq 0$, we can use equations (2.10.29), the definition of k^*, and the monotonicity of function g to get $\xi^2 = g^{-1}[k^*/(3s^2) + 1/2] \in (0,1)$. From equation (2.10.27), the sign of ξ is the same of that of s. Then, $\eta = s/[\xi(3 - \xi^2)]$. QED.

Appendix A.6: Numerical implementation of efficient positional strategies

In this appendix we provide a feasible numerical algorithm for the computation of the optimal positional portfolio allocation defined in equation (2.2.4). The algorithm consists in the application of the Newton-Raphson method for the solution of a maximization problem with equality constraints [see, e.g. Boyd and Vandenberghe (2004)]. Then, the conditional expectations in the gradient and the Hessian of the criterion are computed using the estimated joint model for Gaussian ranks, cross-sectional distribution and macro-factor dynamics in Sections 3 and 5.

i) Newton-Raphson algorithm with equality constraints

The maximization problem associated with equation (2.2.4) can be written as:

$$\begin{aligned} &\max_{\alpha} V_t(\alpha) \\ &s.t.\ \ \alpha' e = 1 \end{aligned}$$

where $\alpha = (\alpha_1, ..., \alpha_n)'$, e is an n-vector of ones, and

$$V_t(\alpha) = E_t\left[\mathcal{U}\left(H_{t+1}(\alpha' y_{t+1})\right)\right] = E_t\left[\mathcal{U}\left(H_{t+1}\left(\sum_{i=1}^{n} \alpha_i H_{t+1}^{-1}(u_{i,t+1})\right)\right)\right].$$

The associated Lagrangian function for the constrained maximization problem is:

$$\mathcal{L}_t(\alpha, \lambda) = V_t(\alpha) + \lambda(\alpha' e - 1),$$

and the first-order conditions are:

$$\begin{cases} \dfrac{\partial V_t(\alpha)}{\partial \alpha} + \lambda e = 0 \\ \alpha' e - 1 = 0 \end{cases}.$$

By applying an extended Newton's procedure based on the Taylor's expansion of the first-order conditions with respect to $(\alpha', \lambda)'$, the solution of the problem is obtained by an iterative algorithm with $(q+1)$-th step given by [see, e.g. Boyd and Vandenberghe (2004)]:

$$\begin{bmatrix} \alpha^{(q+1)} \\ \lambda^{(q+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(q)} \\ 0 \end{bmatrix} - \begin{bmatrix} \dfrac{\partial^2 V_t(\alpha^{(q)})}{\partial\alpha\partial\alpha'} & e \\ e' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\partial V_t(\alpha^{(q)})}{\partial\alpha} \\ \alpha^{(q)\prime}e - 1 \end{bmatrix}.$$

The initial step of the algorithm uses $\alpha^{(0)} = \dfrac{1}{n}e$, that is, the equally weighted portfolio.

ii) Formulas for the gradient vector and Hessian matrix of the expected CARA positional utility function

In the case of a CARA positional utility function with $\mathcal{U}(v;\mathcal{A}) = -\exp(-\mathcal{A}v)$, where $\mathcal{A} > 0$ is the positional risk aversion parameter, written on the Gaussian rank of the portfolio return, the expected positional utility $V_t(\alpha)$ is:

$$V_t(\alpha) = -E_t\left[\exp\left\{-\mathcal{A}H_{t+1}\left(\alpha' y_{t+1}\right)\right\}\right].$$

The gradient vector of the expected CARA positional utility function is:

$$\frac{\partial V_t(\alpha)}{\partial\alpha} = \mathcal{A}E_t\left[\exp\left\{-\mathcal{A}H_{t+1}\left(\alpha' y_{t+1}\right)\right\} H'_{t+1}(\alpha' y_{t+1})y_{t+1}\right], \tag{2.10.30}$$

where $H'_{t+1}(y) = \dfrac{dH_{t+1}(y)}{dy}$. The Hessian matrix of the expected CARA positional utility function is:

$$\frac{\partial^2 V_t(\alpha)}{\partial\alpha\partial\alpha'} = \mathcal{A}E_t\left[\exp\left\{-\mathcal{A}H_{t+1}\left(\alpha' y_{t+1}\right)\right\}\left(-\mathcal{A}H'_{t+1}(\alpha' y_{t+1})^2 + H''_{t+1}(\alpha' y_{t+1})\right) y_{t+1}\, y'_{t+1}\right] \tag{2.10.31}$$

where $H''_{t+1}(y) = \dfrac{d^2 H_{t+1}(y)}{dy^2}$.

Let us now compute the two functions $H'_{t+1}(y)$ and $H''_{t+1}(y)$. Recall from Definition 1 *ii)* in Appendix 1 that:

$$H_t(y) = \Phi^{-1}(H_t^*(y)), \tag{2.10.32}$$

where H_t^* is the cross-sectional c.d.f. of the VG family at date t. Therefore, we have:

$$H'_t(y) = \frac{1}{\phi[\Phi^{-1}(H_t^*(y))]} h_t^*(y),$$

where $h_t^*(y) \equiv h^*(y|F_{d,t}) = dH_t^*(y)/dy$ is the VG p.d.f. in equation (2.10.19) and $\phi(\cdot) = \Phi'(\cdot)$ is the p.d.f. of the standard Gaussian distribution. This allows to compute:

$$H''_t(y) = \frac{\Phi^{-1}(H_t^*(y))}{(\phi[\Phi^{-1}(H_t^*(y))])^2}(h_t^*(y))^2 + \frac{1}{\phi[\Phi^{-1}(H_t^*(y))]}\frac{dh_t^*(y)}{dy}.$$

Finally, we can re-write equation (2.10.31) as:

$$\frac{\partial^2 V_t(\alpha)}{\partial\alpha\partial\alpha'} = \mathcal{A}E_t\left[\exp\left\{-\mathcal{A}H_{t+1}\left(\alpha' y_{t+1}\right)\right\}\xi_{t+1}\left(\alpha' y_{t+1}\right) y_{t+1}\, y'_{t+1}\right], \tag{2.10.33}$$

where:

$$\begin{aligned}\xi_t\,(y) &= -\mathcal{A}H'_t(y)^2 + H''_t(y) \\ &= \frac{\Phi^{-1}(H^*_t(y)) - \mathcal{A}}{(\phi[\Phi^{-1}(H^*_t(y))])^2}(h^*_t(y))^2 + \frac{1}{\phi[\Phi^{-1}(H^*_t(y))]}\frac{dh^*_t(y)}{dy}.\end{aligned} \tag{2.10.34}$$

iii) Estimation of the gradient vector and Hessian matrix of the expected CARA positional utility function

Let us finally discuss the estimation of the conditional expectations in the gradient vector and Hessian matrix of the criterion given in equations (2.10.30) and (2.10.33).

The conditioning information at date t includes the past history of assets returns $\underline{y_t}$ and systematic factors $\underline{F_t}$, and the individual effects $\delta_i = (\beta_i, \gamma_i)'$ for all assets. We use that functions $H_{t+1}(\cdot)$ and $\xi_{t+1}(\cdot)$ in equations (2.10.32) and (2.10.34) involve the future factor value F_{t+1}, the asset returns are $y_{i,t+1} = H^{-1}_{t+1}(u_{i,t+1})$, and the joint process $(F_t, u_{1,t}, ..., u_{n,t})$ is Markov conditional on the individual effects. Then, the conditional expectations in equations (2.10.30) and (2.10.31) are taken with respect to future factor value F_{t+1} and Gaussian ranks $u_{i,t+1}$, given F_t, $u_{i,t}$ and δ_i for all assets in the investment universe. These conditional expectations are computed by Monte Carlo integration by simulating future ranks and factor values according to their estimated models in Sections 3 and 5. The current values of the factor F_t and ranks $u_{i,t}$, and the individual effects δ_i in the conditioning set are replaced by their estimated values.

2.11 Bibliograpghy of Chapter 2

ARAGON, G. O., AND V. NANDA (2012): "Tournament Behavior in Hedge Funds: High-Water Marks, Fund Liquidation, and Managerial Stake," *The Review of Financial Studies*, 25, 937–974.

AVRAMOV, D., T. CHORDIA, AND A. GOYAL (2006): "Liquidity and Autocorrelations in Individual Stock Returns," *The Journal of Finance*, 61, 2365 – 2394.

BAI, J., AND S. NG (2005): "Tests for Skewness, Kurtosis, and Normality for Time Series Data," *Journal of Business and Economic Statistics*, 23, 49–60.

BASEL COMMITTEE ON BANKING SUPERVISION (2013): "Global Systemically Important Banks: Updated Assessment Methodology and the Higher Loss Absorbency Requirement," July.

BELEZNAY, A., M. MARKOV, AND A. PANCHEKHA (2012): "Hidden Benefits of Equal Weighting : The Case of Hedge Fund Indices," *Working Paper*.

BIAIS, B., T. FOUCAULT, AND S. MOINAS (2013): "Equilibrium Fast Trading," *Toulouse University Working Paper*.

BLACK, F. (1976): "Studies of Stock Price Volatility Changes," *Proceedings of the Business and Economics Sections of the American Statistical Association*, pp. 177–181.

BLUME, L. E., W. A. BROCK, S. N. DURLAUF, AND R. JAYARAMAN (2013): "Linear Social Interactions Models," *NBER Working Paper 19212*.

BORLAND, L. (2012): "Statistical Signatures in Times of Panic: Markets as a Self-organizing System," *Quantitative Finance*, 12(9), 1367–1379.

BOUGEROL, P., AND N. PICARD (1992): "Strict Stationarity of Generalized Autoregressive Processes," *The Annals of Probability*, 20, 1714–1730.

BOWMAN, A. W., AND A. AZZALINI (1997): *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford University Press.

BOYD, S., AND L. VANDENBERGHE (2004): *Convex Optimization*. Cambridge University Press.

BROWN, K. C., W. V. HARLOW, AND L. STARKS (1996): "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry," *Journal of Finance*, 51, 85–110.

CALVO-ARMENGOL, A., E. PATACCHINI, AND Y. ZENOU (2009): "Peer Effects and Social Networks in Education," *Review of Economic Studies*, 76, 1239–1267.

CHAN, L. K. C., N. JEGADEESH, AND J. LAKONISHOK (1996): "Momentum Strategies," *The Journal of Finance*, 51, 1681–1713.

CHEN, H., AND G. PENNACCHI (2009): “Does Prior Performance Affect a Mutual Fund’s Choice of Risk? Theory and Further Empirical Evidence,” *Journal of Financial and Quantitative Analysis*, 44, 745–775.

DAROLLES, S., AND C. GOURIEROUX (2014): “The Effects of Management and Provision Accounts on Hedge Fund Returns. Part 1: High-Water Mark Allocation Scheme; Part 2: The Loss Carry Forward Scheme,” *in Advances in Intelligent Systems and Computing, 251*, V. Huyun et al. eds, Springer, 23–46 and 47–72.

DAVEZIES, L., X. D’HAULTFOEUILLE, AND D. FOUGÈRE (2009): “Identification of Peer Effects Using Group Size Variation,” *The Econometrics Journal*, 12, 397–413.

DE BONDT, W., AND R. THALER (1985): “Does the Stock Market Overreact,” *The Journal of Finance*, 40, 793–805.

DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009): “Optimal Versus Naive Diversification: How Inefficient is the $1/N$ Portfolio Strategy?,” *Review of Financial Studies*, 22, 1915–1953.

EASTERLIN, R. A. (1995): “Will Raising the Incomes of All Increase the Happiness of All?,” *Journal of Economic Behavior and Organization*, 27, 35–47.

FRANK, R. (1991): “Positional Externalities,” *in "Strategy and Choice. Essays in Honor of T. Schelling" (R. Zeckhauser ed.), Cambridge MIT Press*.

FRANK, R. H. (1999): *Luxury Fever: Why Money Fails to Satisfy in an Era of Excess*. New York, Free Press.

FREY, B. S., AND A. STUTZER (2002): *Happiness and Economics: How the Economy and Institutions Affect Human Well-Being*. Princeton N.J., Princeton University Press.

GABAIX, X., AND A. LANDIER (2008): “Why Has CEO Pay Increased so Much?,” *Quarterly Journal of Economics*, 123, 49–100.

GORIAEV, A., T. NIJMAN, AND B. J. M. WERKER (2005): “Yet Another Look at Mutual Fund Tournaments,” *Journal of Empirical Finance*, 12, 127–133.

GORIAEV, A., F. PALOMINO, AND A. PRAT (2001): “Mutual Fund Tournament: Risk Taking Incentives Induced by Ranking Objectives,” *CEPR Discussion Papers*.

HIRSCH, F. (1976): *Social Limits to Growth*. Cambridge MA, Harvard University Press.

INGERSOLL, J. (1987): *Theory of Financial Decision Making*. Rowman & Littlefield.

JAGANNATHAN, R., AND T. MA (2003): “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps,” *The Journal of Finance*, 58(4), 1651–1683.

JEGADEESH, N. (1990): “Evidence of Predictable Behavior of Security Returns,” *The Journal of Finance*, 45, 881–898.

JEGADEESH, N., AND S. TITMAN (1993): “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency,” *The Journal of Finance*, 48, 65–91.

KAHNEMAN, D., AND A. TVERSKY (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

KHANDANI, A., AND A. LO (2011): "What Happened to the Quants in August 2007? Evidence from Factors and Transactions Data," *Journal of Financial Markets*, 14, 1–46.

LEDOIT, O., AND M. WOLF (2003): "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10, 603–621.

LEHMANN, B. (1990): "Fads, Martingales, and Market Efficiency," *Quarterly Journal of Economics*, 105, 1–28.

MADAN, D., AND E. SENETA (1990): "The Variance Gamma (V.G.) Model for Share Market Returns," *Journal of Business*, 63, 511–524.

MANSKI, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542.

MOSKOWITZ, T., Y. OOI, AND L. PEDERSEN (2012): "Time Series Momentum," *Journal of Financial Economics*, 104, 228–250.

NEUMARK, D., AND A. POSTLEWAITE (1998): "Relative Income Concerns and the Rise in Married Women's Employment," *Journal of Public Economics*, 70, 157–183.

PASQUARIELLO, P. (2014): "Financial Market Dislocations," *Review of Financial Studies*, 27, 1868–1914.

PEARSON, K. (1916): "Mathematical Contributions to the Theory of Evolution, XIX. Second supplement to a memoir on skew variation," *Philosophical Transactions of the Royal Society*, 216, 432.

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

SCHWARZ, C. (2012): "Mutual Fund Tournaments: The Sorting Bias and New Evidence," *Review of Financial Studies*, 25, 913–936.

SENETA, E. (2004): "Fitting the Variance-Gamma Model to Financial Data," *Journal of Applied Probability*, 41, 177–187.

SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. London, Chapman and Hall.

THANASSOULIS, J. (2012): "The Case of Intervening in Banking Pay," *The Journal of Finance*, 67, 849–895.

VON NEUMANN, J., AND O. MORGENSTERN (1944): *Theory of Games and Economic Behavior*. Princeton University Press.

Chapter 3

Is Industrial Production Still the Dominant Factor for the US Economy?

Abstract

We propose a new class of approximate factor models which enable us to study the full spectrum of quarterly IP sector data combined with annual non-IP sectors of the economy. We derive the large sample properties of the estimators for the new class of factor models involving mixed frequency data. Despite the growth of service sectors, we find that a single common factor explaining 90% of the variability in IP output growth index also explains 60% of total GDP output growth fluctuations. A single low frequency factor unrelated to manufacturing explains 14% of GDP growth. The picture with a structural factor model featuring technological innovations is quite different. IP sectors technology shocks do not play a dominant role.

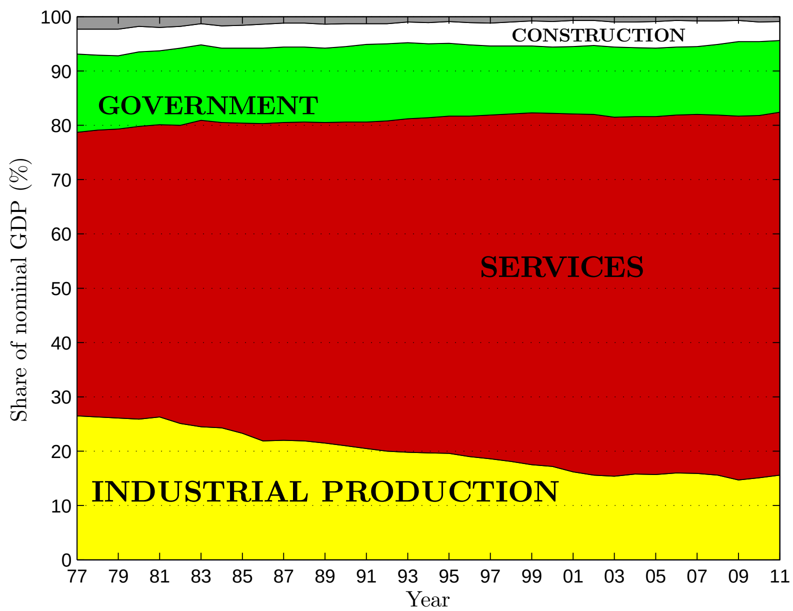
JEL Codes: E23, E32, C32, C38, C55.

Keywords: Approximate Factor Model, Principal Component Analysis, Canonical Correlations, Mixed-frequency Data, Sectoral Output Growth, Industrial Production, Gross Domestic Product.

3.1 Introduction

In the public arena it is often claimed that manufacturing has been in decline in the US and most jobs have migrated overseas to lower wage countries. First, we would like to nuance this observation somewhat. It is true, as the figure below clearly shows, that the share of the industrial production sector has been in decline since the late 70's, which is the beginning of our sample period. However, does size matter? The fact that the size shrank does not necessarily exclude the possibility that the industrial production sector still is a key factor, or even the dominant factor, of total US output. We study the validity of this question using novel econometric methods designed to deal with some of the challenging data issues one encounters when trying to address the problem.

Figure 3.1: Sectoral decomposition of US nominal GDP.



The figure displays the evolution from 1977 to 2011 of the sectoral decomposition of US nominal GDP. We aggregate the shares of different sectors available from the website of the US Bureau of Economic Analysis, according to their North American Industry Classification System (NAICS) codes, in 5 different *macro* sectors: Industrial Production (yellow), Services (red), Government (green), Construction (white), Others (grey).

When studying the role of the industrial production sector we face a conundrum. On the one hand, we have fairly extensive data on industrial production (IP) which consists of 117 sectors that make up aggregate IP, each sector roughly corresponding to a four-digit industry classification using NAICS. These data are published monthly, and therefore cover a rich time series and cross-section. In our analysis we use the data sampled at quarterly frequency, for reasons explained later in the paper, and consists of over 16,000 data points counting all quarters from 1977 until 2011 (end of our data set) across all sectors. On the other hand, contrary to IP, we do not have monthly or quarterly data about the cross-section of US output across non-IP sectors, but we do so on an annual basis. Indeed, the US Bureau of Economic Analysis provides Gross Domestic Product (GDP) and Gross Output by industry - not only IP sectors - annually. In our empirical analysis we use data on 42 non-IP sectors. If we were to study all sectors annually, we would be left with roughly 4000 data points for IP - a substantial loss of information.

Economists have proposed different models to explain how various sectors in the economy interact. Some rely on aggregate shocks which affect all sectors at once. Foerster, Sarte, and Watson (2011), who use an approximate factor model estimated with quarterly data, find that nearly all of IP variability is associated with a small number of common factors - even a single common factor suffices according to their findings. To what extend does the single common factor which drives the cross-sectional variation of IP sectors also affect the rest of the economy, in particular in light of the fact that the services sector grew in relative size? To put it differently, can we maintain a common factor view if we expand beyond IP sectors? Or should we think about sector-specific shocks affecting aggregate US output? If so, are these IP sector shocks, or rather services sector ones?

We propose a new class of factor models able to address these key questions of interest using *all* the data - despite the mixed sampling frequency setting. Empirical research generally avoids the direct use of mixed frequency data by either first aggregating higher frequency series and then performing estimation and testing at the low frequency common across the series, or neglecting the low frequency data and working only on the high frequency series. The literature on large scale factor models is no exception to this practice, see e.g. Forni and Reichlin (1998), Stock and Watson (2002a,b) and Stock and Watson (2010). Using the terminology of the approximate factor model literature, we have a panel consisting of N_H cross-sectional IP sector growth series sampled across MT time periods, where M = 4 for quarterly data and M = 12 for monthly data, with T the number of years. Moreover, we also have a panel of N_L non-IP sectors - such as services and construction for example - which is only observed over T periods. Hence, generically speaking we have a high frequency panel data set of size $N_H \times MT$ and a corresponding low frequency panel data set of size $N_L \times T$. The issue we are interested in can be thought of as follows. There are three types of factors: (1) those which explain variations in both panels - say g^C, and therefore are economy-wide factors, (2) those exclusively pertaining to IP sector movements - say g^H, and finally (3) those exclusively affecting non-IP, denoted by g^L. Hence, we have (1) common, (2) high frequency and (3) low frequency factors. We use superscripts C, H and L because the theory we develop is generic and pertains to common (C), high frequency (H) and low frequency (L) factors. The question how to extract common factors from a mixed frequency panel data set is of general interest and has many applications in economics and other fields. In fact our analysis covers an even broader class of group factor models, as will be explained shortly, which is of general interest beyond the mixed frequency setting considered in the empirical application.

The purpose of this paper is to propose large scale approximate factor models in the spirit of Bai and Ng (2002), Stock and Watson (2002a), Bai (2003), Bai and Ng (2006), and extend their analysis to mixed frequency data settings. A number of mixed frequency factor models have been proposed in the literature, although they almost exclusively rely on small cross-sections.[1]

We approach the problem from a different angle. We start with a setup which identifies factors common

[1]See for example, Mariano and Murasawa (2003), Nunes (2005), Aruoba, Diebold, and Scotti (2009) Frale and Monteforte (2010), Marcellino and Schumacher (2010) and Banbura and Rünstler (2011), among others. Stock and Watson (2002b) in their Appendix A, propose a modification of the EM algorithm of Dempster, Laird, and Rubin (1977) to estimate high frequency factors from potentially large unbalanced panels, with mixed-frequency being a special case. Moreover, Jungbacker, Koopman, and der Wel (2011) introduce a computationally efficient EM algorithm for the maximum likelihood estimation of a high-dimensional linear factor model with missing data.

to both high and low frequency data panels, the aforementioned g^C, and factors specific to the high and low frequency data. Our approach amounts to writing the model as a grouped factor model. The idea to apply grouped factor analysis to mixed frequency data is novel and has many advantages in terms of identification and estimation. In the proposed identification strategy, the groups correspond to panels observed at different sampling frequencies. While there is a literature on how to estimate factors in a grouped model setting, there does not exist a general unifying asymptotic theory for large panel data.[2] We propose estimators for the common and group specific factors, and an inference procedure for the number of common and group specific factors based on canonical correlation analysis of the principal components (PCs) estimators on each subgroup. The procedure is therefore general in scope and also of interest in many applications other than the one considered in the current paper. We study the large sample properties of our estimators and inference procedure as $T, N_H, N_L \to \infty$.

Our empirical application revisits the analysis of Foerster, Sarte, and Watson (2011) who use factor analytic methods to decompose industrial production (IP) into components arising from aggregate shocks and idiosyncratic sector-specific shocks. They focus exclusively on the industrial production sectors of the US economy. We find that a single common factor explains around 90% of the variability in the aggregate IP output growth index, and a factor specific to IP has very little additional explanatory power. This implies that the single common factor can be interpreted as an Industrial Production factor. Moreover, more than 60% of the variability of GDP output growth in service sectors, such as Transportation and Warehousing services, is also explained by the common factor. A single low frequency factor unrelated to manufacturing, explaining around 14 % of GDP growth fluctuations, drives the comovement of non-IP sectors such as Construction and Government. Note the great advantage of the mixed frequency setting - compared to the single frequency one - in the context of our IP and GDP sector application. The mixed frequency panel setting allows us to identify and estimate the *high frequency* observations of factors common to IP and non-IP sectors. With IP (i.e. high frequency) data only we cannot assess what is common with non-IP. With low frequency data only, we cannot estimate the high frequency common factors.

We re-examine whether the common factor reflects sectoral shocks that have propagated by way of input-output linkages between service sectors and manufacturing. A structural factor analysis indicates that both low and high frequency aggregate shocks continue to be the dominant source of variation in the US economy. The propagation mechanisms are very different, however, from those identified by Foerster, Sarte, and Watson (2011). Looking at technology shocks instead of output growth, it does not appear that a common factor explaining IP fluctuations is a dominant one for the entire economy. A factor specific to technological innovations in IP sectors is more important for the IP sector shocks and a low frequency factor which appears to explain variation in information industry as well as professional and business services innovations plays relatively speaking a more important role. Hence, when it comes to innovation shocks, IP is no longer the dominant factor.

The rest of the paper is organized as follows. In section 3.2 we introduce the formal model and discuss identification. In section 3.3 we study estimation and inference on the number of common factors. The large sample theory appears in section 3.4. Section 3.5 covers the empirical application. Section 3.6 concludes the paper. Readers who are only interested in the empirical applications can go directly to section 3.5 which starts with a summary of the novel econometric procedure. Appendix C contains additional material including an alternative identification strategy for the common factor space, a discussion of some properties of an iterative

[2] In section 3.2.3 we provide a short review of the literature on group factor models, including recent contributions related to the topic of the current paper.

estimator, dataset description, a procedure for the extraction of technology shocks, and additional empirical results.

3.2 Model specification and identification

We consider a setting where both low and high frequency data are available. Let $t = 1, 2, \ldots, T$ be the low frequency (LF) time units. Each period $(t-1, t]$ is divided into M subperiods with high frequency (HF) dates $t - 1 + m/M$, with $m = 1, \ldots, M$. Moreover, we assume a panel data structure with a cross-section of size N_H of high frequency data and N_L of low frequency data. It will be convenient to use a double time index to differentiate low and high frequency data. Specifically, we let $x^{Hi}_{m,t}$, for $i = 1, \ldots, N_H$, be the high frequency data observation i during subperiod m of low frequency period t. Likewise, we let x^{Li}_t, with $i = 1, \ldots, N_L$, be the observation of the i^{th} low-frequency series at t. These observations are gathered into the N_H-dimensional vectors $x^H_{m,t}$, $\forall\, m$, and the N_L-dimensional vector x^L_t, respectively.

We have a latent factor structure in mind to explain the panel data variation for both the low and high frequency data. To that end, we assume that there are three types of factors, which we denote by respectively $g^C_{m,t}$, $g^H_{m,t}$ and $g^L_{m,t}$. The former represents factors which affect both high and low frequency data (throughout we use superscript C for common), whereas the other two types of factors affect exclusively high (superscript H) and low (marked by L) frequency data. We denote by k^C, k^H and k^L, the dimensions of these factors. The latent factor model with high frequency data sampling is:

$$\begin{array}{rcl} x^H_{m,t} & = & \Lambda_{HC} g^C_{m,t} + \Lambda_H g^H_{m,t} + e^H_{m,t}, \\ x^{L*}_{m,t} & = & \Lambda_{LC} g^C_{m,t} + \Lambda_L g^L_{m,t} + e^L_{m,t}, \end{array} \tag{3.2.1}$$

where $m = 1, ..., M$ and $t = 1, ..., T$, and Λ_{HC}, Λ_H, Λ_{LC} and Λ_L are matrices of factor loadings. The vector $x^{L*}_{m,t}$ is not observable for each high frequency subperiod and the measurements, denoted by x^L_t, depend on the observation scheme, which can be either flow sampling or stock sampling (or some general linear scheme). In the remainder of this section we study identification of the model for the case of flow sampling, corresponding to the empirical application covered later in the paper.[3]

In the case of flow sampling, the low frequency observations are the sum (or average) of all $x^{L*}_{m,t}$ in each high frequency subperiod m, that is: $x^L_t = \sum_{m=1}^M x^{L*}_{m,t}$. Then, model (3.2.1) implies:

$$\begin{array}{rcl} x^H_{m,t} & = & \Lambda_{HC} g^C_{m,t} + \Lambda_H g^H_{m,t} + e^H_{m,t}, \quad m = 1, ..., M, \\ x^L_t & = & \Lambda_{LC} \sum_{m=1}^M g^C_{m,t} + \Lambda_L \sum_{m=1}^M g^L_{m,t} + \sum_{m=1}^M e^L_{m,t}. \end{array} \tag{3.2.2}$$

Let us define the aggregated variables and innovations $x^H_t := \sum_{m=1}^M x^H_{m,t}$, $\bar{e}^U_t := \sum_{m=1}^M e^U_{m,t}$, $U = H, L$, and the aggregated factors:

$$\bar{g}^U_t \;\; := \;\; \sum_{m=1}^M g^U_{m,t}, \qquad U = C, H, L.$$

[3]The identification with stock sampling is discussed in Appendix 3.7.1. It is worth noting though that any sampling scheme leading to a representation of the model analogous to the group-factor model in equation (3.2.3) or (3.2.4) - discussed shortly - is compatible with the identification and estimation strategies of this paper.

Then we can stack the observations x_t^H and x_t^L and write:

$$\left[\begin{array}{c} x_t^H \\ x_t^L \end{array}\right] = \left[\begin{array}{ccc} \Lambda_{HC} & \Lambda_H & 0 \\ \Lambda_{LC} & 0 & \Lambda_L \end{array}\right] \left[\begin{array}{c} \bar{g}_t^C \\ \bar{g}_t^H \\ \bar{g}_t^L \end{array}\right] + \left[\begin{array}{c} \bar{e}_t^H \\ \bar{e}_t^L \end{array}\right]. \tag{3.2.3}$$

The last equation corresponds to a group factor model, with common factor $\bar{g}_t^C$ and "group-specific" factors $\bar{g}_t^H$, $\bar{g}_t^L$.

To further generalize the setup, and draw directly upon the group-factor structure, we will consider the generic specification. To separate the specific from the generic case, we will change notation slightly. Namely, we keep the notation introduced so far with high and low frequency data, temporal aggregation, etc. for the mixed frequency setting further used in the empirical application and use the following notation for the generic grouped factor model setting:

$$\left[\begin{array}{c} y_{1,t} \\ y_{2,t} \end{array}\right] = \left[\begin{array}{ccc} \Lambda_1^c & \Lambda_1^s & 0 \\ \Lambda_2^c & 0 & \Lambda_2^s \end{array}\right] \left[\begin{array}{c} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{array}\right] + \left[\begin{array}{c} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{array}\right], \tag{3.2.4}$$

where $y_{j,t} = [y_{j,1t}, ..., y_{j,N_jt}]'$, $\Lambda_j^c = [\lambda_{j,1}^c, ..., \lambda_{j,N_j}^c]'$, $\Lambda_j^s = [\lambda_{j,1}^s, ..., \lambda_{j,N_j}^s]'$ and $\varepsilon_{j,t} = [\varepsilon_{j,1t}, ..., \varepsilon_{j,N_jt}]'$, with $j = 1, 2$. The dimensions of the common factor f_t^c and the group-specific factors $f_{1,t}^s$, $f_{2,t}^s$ are k^c, k_1^s and k_2^s, respectively. In the case of no common factors, we set $k^c = 0$, while in the case of no group-specific factors we set $k_j^s = 0$, $j = 1, 2$.[4] The group-specific factors $f_{1,t}^s$ and $f_{2,t}^s$ are orthogonal to the common factor f_t^c. Since the unobservable factors can be standardized, we assume:

$$E\left[\begin{array}{c} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right], \tag{3.2.5}$$

and

$$V\left[\begin{array}{c} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{array}\right] = \left[\begin{array}{ccc} I_{k^c} & 0 & 0 \\ 0 & I_{k_1^s} & \Phi \\ 0 & \Phi' & I_{k_2^s} \end{array}\right], \tag{3.2.6}$$

where Φ is the covariance between the group-specific factors.[5]

3.2.1 Separation of common and group-specific factors

In standard linear latent factor models, the normalization induced by an identity factor variance-covariance matrix identifies the factor process up to a rotation (and change of signs). Let us now show that, under suitable identification conditions, the rotational invariance of model (3.2.4) - (3.2.6) allows only for separate rotations among the components of $f_{1,t}^s$, among those of $f_{2,t}^s$, and among those of f_t^c. The rotational invariance of

[4] The case of more than two groups is a relatively straightforward generalization. Note that would also handle situations with more than two sampling frequencies. In the interest of conciseness, we do not consider this type of generalization in the current paper.

[5] In the main body of the text we only highlight some of the key assumptions. In Appendix A section 3.7.2 we provide a detailed list of the assumptions.

model (3.2.4) - (3.2.6) therefore maintains the interpretation of common factor and specific factors. More formally, let us consider the following transformation of the stacked factor process:

$$\begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} \tilde{f}_t^c \\ \tilde{f}_{1,t}^s \\ \tilde{f}_{2,t}^s \end{bmatrix} \tag{3.2.7}$$

where $(\tilde{f}_t^{c\prime}, \tilde{f}_{1,t}^{s\prime}, \tilde{f}_{2,t}^{s\prime})'$ is the transformed stacked factor vector, and the block matrix $A = (A_{ij})$ is nonsingular.

Definition 1. *The model is identifiable if: the data $y_{1,t}$ and $y_{2,t}$ satisfy a factor model of the same type as (3.2.4) - (3.2.6) with $(f_t^{c\prime}, f_{1,t}^{s\prime}, f_{2,t}^{s\prime})'$ replaced by $(\tilde{f}_t^{c\prime}, \tilde{f}_{1,t}^{s\prime}, \tilde{f}_{2,t}^{s\prime})'$ only when matrix A is a block-diagonal orthogonal matrix.*

The following proposition gives a sufficient condition for the identification of the model with common and group-specific factors.[6]

Proposition 1. *If matrices $\Lambda_1 = \begin{bmatrix} \Lambda_1^c \vdots \Lambda_1^s \end{bmatrix}$ and $\Lambda_2 = \begin{bmatrix} \Lambda_2^c \vdots \Lambda_2^s \end{bmatrix}$ are full column-rank (for N_j large enough), then the model is identifiable in the sense of Definition 1.*

Proof: See Appendix 3.8.1.

Therefore the common factor f_t^c and the group-specific factors $f_{1,t}^s$, $f_{2,t}^s$ and the factor loadings Λ_j^c, Λ_j^s, are identifiable up to a linear transformation, since the variables $y_{j,t}$ are observable. By the same token in the mixed frequency setting of equation (3.2.3), the aggregated factors $\bar{g}_t^C$, $\bar{g}_t^H$, $\bar{g}_t^L$, and the factor loadings Λ_{HC}, Λ_{LC}, Λ_H, Λ_L, are identified. Once the factor loadings are identified from (3.2.3), the values of the common and high frequency factors for subperiods $m = 1, ..., M$ are identifiable by cross-sectional regression of the high frequency data on loadings Λ_{HC} and Λ_H in (3.2.1). More precisely, $g_{m,t}^C$ and $g_{m,t}^H$ are identified by regressing $x_{m,t}^{Hi}$ on $\lambda_{HC,i}$ and $\lambda_{H,i}$ across $i = 1, 2, ...,$ for any $m = 1, ..., M$ and any t. Hence, with flow sampling, we can identify the common factor $g_{m,t}^C$ and the high frequency factor $g_{m,t}^H$ at all high frequency subperiods. On the other hand, only $\bar{g}_t^L = \sum_{m=1}^{M} g_{m,t}^L$, i.e. the within-period sum of the low frequency factor, is identifiable by the paired panel data set consisting of x_t^H combined with x_t^L. This is not surprising, since we have no HF observation available for the LF process.

3.2.2 Identification of the (common) factor space from canonical correlations and directions

In the interest of generality, let us again consider the generic setting of equation (3.2.4) and let $k_j = k^c + k_j^s$, for $j = 1, 2$, be the dimensions of the factor spaces for the two groups, and define $\underline{k} = \min(k_1, k_2)$. We collect the factors of each group in the k_j-dimensional vectors $h_{j,t}$:

$$h_{j,t} := \begin{bmatrix} f_t^c \\ f_{j,t}^s \end{bmatrix}, \qquad j = 1, 2, \qquad t = 1, ..., T, \tag{3.2.8}$$

[6] See also results in e.g. Schott (1999), Wang (2012), Chen (2010, 2012). Proposition 1 is implied by Proposition 1 in Wang (2012).

and the loadings in the k_j-dimensional vectors $\lambda_{j,i}$:

$$\lambda_{j,i} := \begin{bmatrix} \lambda_{j,i}^c \\ \lambda_{j,i}^s \end{bmatrix}, \qquad j = 1,2, \qquad i = 1,...,N_j.$$

Using these definitions, model (3.2.4) can equivalently be written as:

$$y_{j,it} = \lambda_{j,i}' h_{j,t} + \varepsilon_{j,it}, \qquad j = 1,2, \qquad i = 1,...,N_j, \qquad t = 1,...,T,$$

We also stack the factors $h_{j,t}$, $j = 1,2$, into the K-dimensional vector $h_t = (h_{1,t}', h_{2,t}')'$, with $K = k_1 + k_2$. Moreover, let us express the (K, K)-dimensional matrix $V(h_t)$ as:

$$V(h_t) = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \tag{3.2.9}$$

where:

$$V_{j\ell} := E(h_{j,t} h_{\ell,t}'), \qquad j, \ell = 1,2. \tag{3.2.10}$$

Let us first recall a few basic results from canonical analysis (see e.g. Anderson (2003) and Magnus and Neudecker (2007)). Let ρ_ℓ, $\ell = 1,...,\underline{k}$ denote the canonical correlations between $h_{1,t}$ and $h_{2,t}$. The largest $\underline{k}$ eigenvalues of matrices

$$R = V_{11}^{-1} V_{12} V_{22}^{-1} V_{21}, \qquad \text{and} \quad R^* = V_{22}^{-1} V_{21} V_{11}^{-1} V_{12},$$

are the same, and are equal to the squared canonical correlations ρ_ℓ^2, $\ell = 1,...,\underline{k}$ between $h_{1,t}$ and $h_{2,t}$. The associated eigenvectors $w_{1,\ell}$ (resp. $w_{2,\ell}$), with $\ell = 1,...,\underline{k}$, of matrix R (resp. R^*) standardized such that $w_{1,\ell}' V_{11} w_{1,\ell} = 1$ (resp. $w_{2,\ell}' V_{22} w_{2,\ell} = 1$) are the canonical directions which allow to construct the canonical variables from vector $h_{1,t}$ (resp. $h_{2,t}$). The matrices $w_j = (w_{j,1},...,w_{j,\underline{k}})$, $j = 1,2$, are such that $w_j' V_{jj} w_j = I_{\underline{k}}$, $j = 1,2$. Moreover, if $\rho_\ell \neq 0$, then

$$\begin{aligned} w_{1,\ell} &= \frac{1}{\rho_\ell} V_{11}^{-1} V_{12} w_{2,\ell}, \\ w_{2,\ell} &= \frac{1}{\rho_\ell} V_{22}^{-1} V_{21} w_{1,\ell}. \end{aligned} \tag{3.2.11}$$

Proposition 2. *The following hold:*
i) If $k^c > 0$, the largest k^c canonical correlations between $h_{1,t}$ and $h_{2,t}$ are equal to 1, and the remaining $\underline{k} - k^c$ canonical correlations are strictly smaller than 1.
ii) Let W_j be the (k_j, k^c) matrix whose columns are the canonical directions for $h_{j,t}$ associated with the k^c canonical correlations equal to 1, with $j = 1,2$. Then, we have $f_t^c = W_j' h_{j,t}$ (up to a rotation matrix), for $j = 1,2$.
iii) If $k^c = 0$, all canonical correlations between $h_{1,t}$ and $h_{2,t}$ are strictly smaller than 1.
iv) Let W_1^s (resp. W_2^s) be the (k_1, k_1^s) (resp. (k_2, k_2^s)) matrix whose columns are the eigenvectors of matrix R (resp. R^*) associated with the smallest k_1^s (resp. k_2^s) eigenvalues. Then $f_{j,t}^s = W_j^{s\prime} h_{j,t}$ (up to a rotation matrix) for $j = 1,2$.

Proof: See Appendix 3.8.2.

Proposition 2 shows that the number of common factors k^c, the common factor space spanned by f_t^c, and the spaces spanned by group specific factors, can be identified from the canonical correlations and canonical variables of $h_{1,t}$ and $h_{2,t}$. Therefore, the dimension k^c, and factors f_t^c and $f_{j,t}^s$, $j = 1, 2$, (up to a rotation) are identifiable from information that can be inferred by disjoint principal component analysis (PCA) on the two subgroups. Note that disjoint PCA on the two subgroups allows us to identify $h_{1,t}$ and $h_{2,t}$ up to linear transformations. This fact does not prevent identifiability of the common and group-specific factors from Proposition 2. More precisely, from the subpanel j we can identify the vector $h_{j,t}$ up to a non-singular matrix U_j, say, $j = 1, 2$. Under the transformation $h_{j,t} \to U_j h_{j,t}$, the matrices R and R^* are transformed such that $R \to (U_1')^{-1} R U_1'$ and $R^* \to (U_2')^{-1} R^* U_2'$. Therefore, the matrices of canonical directions W_1 and W_2 are transformed such as $W_j \to (U_j')^{-1} W_j$, $j = 1, 2$. Therefore, the quantities $W_j' h_{j,t}$, $j = 1, 2$, are invariant under such transformations.

One may wonder why we do not apply canonical correlation analysis directly to the (aggregated) high and low frequency data - avoiding the first step of computing PCs since the extra step considerably complicates the asymptotics and actually entails a novel contribution of the paper. What makes the first step of computing PCs necessary is the fact that canonical correlations applied to the raw data may not necessarily uncover pervasive factors.[7] One may also wonder why we cannot stack all groups into one panel and apply standard PCA to estimate common factors as in Bekaert, Hodrick, and Zhang (2009) and Korajczyk and Sadka (2008), for instance. Unfortunately, this is also not a solution either, as discussed in Boivin and Ng (2006), Goyal, Pérignon, and Villa (2008), Wang (2012) and Breitung and Eickmeier (2014). In fact, in the case of a model with k^c common factors, a finite number of groups, and a positive number of group-specific factors, the estimate of the common factor obtained from the first k^c principal components of the pooled data is inconsistent due to the correlation in the residuals terms arising from the group specific factors.

3.2.3 What is new and different?

There already exist a number of papers on group factor models, sometimes also named "multilevel factor models" or "hierarchical factor models". Many, rooted in the statistics literature, deal with large T and finite cross-sections, i.e. $N = \max_j (N_j) < \infty$. See for example, Dauxois and Pousse (1975), Krzanowski (1979), Flury (1984), Flury (1988), Schott (1988), Schott (1991), Schott (1999), Chen and Robinson (1989), Dauxois, Romain, and Viguier (1993), Gregory and Head (1999), Kose, Otrok, and Whiteman (2008), and Viguier-Pla (2004). Recently, Bayesian methods for state space models have been applied by Moench and Ng (2011) and Moench, Ng, and Potter (2013) for relatively large scale hierarchical factor models. Moreover, Hallin and Liska (2011) extend the estimator based on dynamic PCs to their dynamic factor model with block structure, which is similar to the grouped factor models studied by the above literature. The main contribution of Goyal, Pérignon, and Villa (2008) is the extension of the results of the classical statistical literature on group factor models to the case of approximate group factor models, even if they do not derive analytical asymptotic results.

Note that one novel contribution of our analysis is to characterize mixed frequency data panel data models as group factor models. This connection actually eliminates one topic of concern in group factor models - the so called classification issue - we do not deal with. Namely, in a mixed data sampling setting we know a

[7] A simple example would be to add an anomalous series to one panel and repeat the series to the other one. The canonical correlation analysis applied to the raw data will uncover the presence of the anomalous series in both panels.

priori to which group observations in the panel belong. Some analysis in group factor models assumes that the researcher has first to figure out the allocation of observations across the different groups.

Our work is most closely related to Chen (2010, 2012), Wang (2012), and Breitung and Eickmeier (2014), who handle the large dimensional T and N cases. There are gaps in the existing analysis, however, and to the best of our knowledge there is no comprehensive asymptotic treatment yet of grouped factor models in a large dimensional setting. Hence, our contribution is to provide a general comprehensive analysis.

It is worth explaining the relationship with the aforementioned existing literature. Wang (2012) and Breitung and Eickmeier (2014) propose an iterative solution derived from a Least Square (LS) estimation problem.[8] In particular, Wang (2012) derives the first order conditions for a principal component estimator and describes an iterative PC estimator (see Wang (2012)). The large sample properties of the iterated PC estimator are, however, not provided. In addition, we show in defin section 3.9.2 that the iterative procedure is not operational as the resulting equations do not have a unique solution.

Chen (2010, 2012) computes for different numbers of common and specific factors a modification of Bai and Ng (2002) information criteria.[9] He proves that choosing the number of common and specific factors by minimizing this new criterion is a consistent selection procedure for the number of common and specific factors. Let us consider the case with two groups: this procedure tells that the number of common factors in the two groups corresponds to the number of eigenvalues equal to 2 of the variance covariance matrix of the stacked pervasive factors (i.e. PCs) extracted in the two groups. Consequently, the common factors can be estimated using the associated eigenvectors. We provide in Appendix C the identification of the common factor space from variance-covariance matrix of stacked factors (see Section 3.9.1). Chen (2012) uses this theory to show how to estimate the common factors, and derives the asymptotic distribution of the estimators - assuming that their number is known. However, the asymptotic theory for the distribution of the eigenvalues equal to two is not developed. Our paper fills this gap and provides (1) a comprehensive asymptotic theory for testing the number of common factors and specific factors, and (2) a large sample theory for the factor estimates.

More specifically, we provide a theory for the number of canonical correlations equal to one. Note that if the PCs in the two groups were observed, then the problem of testing for unit canonical correlations would have a degenerate feature, because it involves testing for deterministic relationships between random vectors. The estimation errors of the PCs drive the asymptotic distribution of the statistic, with an unusual convergence rate of $N\sqrt{T}$. Moreover, we also have to deal with a bias adjustment term, due to a problem akin to errors in the variables. The positive bias adjustment term re-centers the distribution to yield an asymptotic Gaussian density. The next section provides the details of these derivations.[10]

[8] Similarly, Ando and Bai (2013) suggest and iterative procedure based on the minimization of a least square criterion with a shrinkage penalty for the estimation of grouped factor models featuring both observable and unobservable factors.

[9] Similarly, Hallin and Liska (2011) extend the information criteria of Hallin and Liska (2007) for their dynamic factor model with block structure.

[10] Using our theory developed in the next sections, it is straightforward to derive the asymptotic distribution for the number of eigenvalues equal to two of the variance covariance matrix of the stacked PCs. We also note that, the results of our paper could be used to derive the asymptotic properties of the criteria for the selection of the number of common and group-specific factors in Goyal, Pérignon, and Villa (2008).

3.3 Estimation and inference on the number of common factors

In Section 3.3.1 we provide estimators of the common and group-specific factors, based on canonical correlations and canonical directions, when the true number of group-specific and common factors are known. In Section 3.3.2 we propose a sequential testing procedure for determining the number of common factors when only the dimensions k_1 and k_2 are known. The test statistic is based on the canonical correlations between the estimated factors in each subgroup of observables. In Section 3.3.3 we explain why the asymptotic results concerning the test statistic and the factors estimators obtained under the assumption that the number of pervasive factors k_1 and k_2 in each group is known, remain unchanged when the number of pervasive factors is consistently estimated. Finally, in Section 3.3.4 we use these results to define estimators and test statistics for the mixed frequency factor model.

3.3.1 Estimation of common and group-specific factors when the number of common and group-specific factors is known

Let us assume that the true number of factors $k_j > 0$ in each subgroup, $j = 1, 2$ is known, and also that the true number of common factors $k^c > 0$, is known. Proposition 2 suggests the following estimation procedure for the common factor. Let $h_{1,t}$ and $h_{2,t}$ be estimated (up to a rotation) by extracting the first k_j Principal Components (PCs) from each subpanel j, and denote by $\hat{h}_{j,t}$ these PC estimates of the factors, $j = 1, 2$. Let $\hat{H}_j = [\hat{h}_{j,1}, ..., \hat{h}_{j,T}]'$ be the (T, k_j) matrix of estimated PCs extracted from panel $Y_j = [y_{j,1}, ..., y_{j,T}]'$ associated with the largest k_j eigenvalues of matrix $\frac{1}{N_j T} Y_j Y_j'$, $j = 1, 2$. Let $\hat{V}_{j\ell}$ denote the empirical covariance matrix of the estimated vectors $\hat{h}_{j,t}$ and $\hat{h}_{\ell,t}$, with $j, \ell = 1, 2$:

$$\hat{V}_{j\ell} = \frac{\hat{H}_j' \hat{H}_\ell}{T} = \frac{1}{T}\sum_{t=1}^{T} \hat{h}_{j,t}\hat{h}_{\ell,t}', \qquad j, \ell = 1, 2, \tag{3.3.1}$$

and let matrices $\hat{R}$ and $\hat{R}^*$ be defined as:

$$\hat{R} := \hat{V}_{11}^{-1}\hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21}, \text{ and } \hat{R}^* := \hat{V}_{22}^{-1}\hat{V}_{21}\hat{V}_{11}^{-1}\hat{V}_{12}. \tag{3.3.2}$$

Matrices $\hat{R}$ and $\hat{R}^*$ have the same non-zero eigenvalues. From Anderson (2003) and Magnus and Neudecker (2007), we know that the largest k^c eigenvalues of $\hat{R}$ (resp. $\hat{R}^*$), denoted by $\hat{\rho}_\ell^2$, $\ell = 1, ..., k^c$, are the first k^c squared sample canonical correlation between $\hat{h}_{1,t}$ and $\hat{h}_{2,t}$. We also know that the associated k^c canonical directions, collected in the (k_1, k^c) (resp. (k_2, k^c)) matrix $\hat{W}_1$ (resp. $\hat{W}_2$), are the eigenvectors associated with the largest k^c eigenvalues of matrix $\hat{R}$ (resp. $\hat{R}^*$), normalized to have length 1 w.r.t. matrix $\hat{V}_{11}$ (resp. $\hat{V}_{22}$). It also holds:

$$\hat{W}_1'\hat{V}_{11}\hat{W}_1 = I_{k^c}, \text{ and } \hat{W}_2'\hat{V}_{22}\hat{W}_2 = I_{k^c}.$$

Definition 2. *Two estimators of the common factors vector are* $\hat{f}_t^c = \hat{W}_1'\hat{h}_{1,t}$ *and* $\hat{f}_t^{c*} = \hat{W}_2'\hat{h}_{2,t}$.

Let matrix $\hat{W}_1^s$ (resp. $\hat{W}_2^s$) be the (k_1, k_1^s) (resp. (k_2, k_2^s)) matrix collecting k_1^s (resp. k_2^s) eigenvectors associated with the k_1^s (resp. k_2^s) smallest eigenvalues of matrix $\hat{R}$ (resp. $\hat{R}^*$), normalized to have length 1 w.r.t. matrix $\hat{V}_{11}$ (resp. $\hat{V}_{22}$). It also holds:

$$\hat{W}_1^{s\prime}\hat{V}_{11}\hat{W}_1^s = I_{k_1^s}, \text{ and } \hat{W}_2^{s\prime}\hat{V}_{22}\hat{W}_2^s = I_{k_2^s}.$$

The estimators of the group-specific factors can be defined analogously to the definition of the common factors.

Definition 3. *Two estimators of the specific factors vector are* $\breve{f}^s_{1,t} = \hat{W}^s_1{}'\hat{h}_{1,t}$ *and* $\breve{f}^s_{2,t} = \hat{W}^s_2{}'\hat{h}_{2,t}$.

Let $\hat{F}^c = [\hat{f}^c_1{}', ..., \hat{f}^c_T{}']'$ and $\hat{F}^{c*} = [\hat{f}^{c\,*}_1{}', ..., \hat{f}^{c\,*}_T{}']'$ be the (T, k^c) matrices of estimated common factors, and $\breve{F}^s_j = [\breve{f}^s_{j,1}{}', ..., \breve{f}^s_{j,T}{}']'$ be the (T, k^s_j), for $j = 1, 2$, be the matrices of estimated group-specific factors. Then, $\hat{F}^c$ (resp. $\hat{F}^{c*}$) and $\breve{F}^s_1$ (resp. $\breve{F}^s_2$) are orthogonal in sample.

An alternative estimator for the group-specific factors $f^s_{1,t}$ (resp. $f^s_{2,t}$) is obtained by computing the first k^s_1 (resp. k^s_2) principal components of the variance-covariance matrix of the residuals of the regression of $y_{1,t}$ (resp. $y_{2,t}$) on the estimated common factors.[11] More specifically, let $\hat{\Lambda}^c_j = [\hat{\lambda}^c_{j,1}, ..., \hat{\lambda}^c_{j,N_j}]'$ be the (N_j, k^c) matrix collecting the loadings estimators:

$$\hat{\Lambda}^c_j \;=\; Y_j'\hat{F}^c(\hat{F}^c{}'\hat{F}^c)^{-1}, \qquad j = 1, 2. \tag{3.3.3}$$

Let $\xi_{j,it} = y_{j,it} - \hat{\lambda}^c_{j,i}{}'\hat{f}^c_t$ be the residuals of the regression of $y_{j,t}$ on the estimated common factor $\hat{f}^c_t$, and let $\xi_{j,t} = [\xi_{j,1t}, ..., \xi_{j,N_j t}]'$, for $j = 1, 2$. Let $\Xi_j = [\xi_{j,1}, ..., \xi_{j,T}]'$ be the (T, N_j) matrix of the regression residuals, for $j = 1, 2$.

Definition 4. *An alternative estimator of the specific factor vector is* $\hat{f}^s_{1,t}$ *(resp.* $\hat{f}^s_{2,t}$*), defined as the first* k^s_1 *(resp.* k^s_2*) Principal Components of subpanel* Ξ_1 *(resp.* Ξ_2*).*

We denote by $\hat{F}^s_j = [\hat{f}^s_{j,1}{}', ..., \hat{f}^s_{j,T}{}']'$ the (T, k^s_j) matrix of estimated group-specific factors, corresponding to the PCs extracted from panel Ξ_j associated with the largest k^s_j eigenvalues of matrix $\dfrac{1}{N_jT}\Xi_j\Xi_j'$, for $j = 1, 2$. Then, $\hat{F}^c$ is orthogonal in sample both to $\hat{F}^s_1$ and to $\hat{F}^s_2$. Moreover, we define $\hat{\Lambda}^s_j = [\hat{\lambda}^s_{j,1}, ..., \hat{\lambda}^s_{j,N_j}]'$ as the (N_j, k^s_j) matrix collecting the loadings estimators:

$$\hat{\Lambda}^s_j \;=\; Y_j'\hat{F}^s_j(\hat{F}^s_j{}'\hat{F}^s_j)^{-1} \;=\; \Xi_j'\hat{F}^s_j(\hat{F}^s_j{}'\hat{F}^s_j)^{-1}, \qquad j = 1, 2, \tag{3.3.4}$$

where the second equality follows from the in-sample orthogonality of $\hat{F}^c$ and $\hat{F}^s_j$, for $j = 1, 2$.

3.3.2 Inference on the number of common factors based on canonical correlations

Suppose that the number of factors k_1 and k_2 in each subpanel is known, and hence $\underline{k} = \min(k_1, k_2)$ as well, and let us consider the problem of inferring the dimension k^c of the common factor space. From Proposition 2, this dimension is the number of unit canonical correlations between $h_{1,t}$ and $h_{2,t}$. We consider the following set of hypotheses:

$$\begin{aligned}
H(0) &= \left\{1 > \rho_1 \geq ... \geq \rho_{\underline{k}}\right\}, \\
H(1) &= \left\{\rho_1 = 1 > \rho_2 \geq ... \geq \rho_{\underline{k}}\right\} \\
&\;\;\; ... \\
H(k^c) &= \left\{\rho_1 = ... = \rho_{k^c} = 1 > \rho_{k^c+1} \geq ... \geq \rho_{\underline{k}}\right\}, \\
&\;\;\; ... \\
H(\underline{k}) &= \left\{\rho_1 = ... = \rho_{\underline{k}} = 1\right\},
\end{aligned}$$

[11] This alternative estimation method for the group-specific factors corresponds to the method proposed by Chen (2012).

where $\rho_1, ..., \rho_{\underline{k}}$ are the canonical correlations of $h_{1,t}$ and $h_{2,t}$. Hypothesis $H(0)$ corresponds to the case of no common factor in the two groups of observables Y_1 and Y_2. Generically, $H(k^c)$ corresponds to the case of k^c common factor and $k_1 - k^c$ and $k_2 - k^c$ group-specific factors in each group. The largest possible number of common factors is the minimum between k_1 and k_2, i.e. $\underline{k}$, and corresponds to hypothesis $H(\underline{k})$. In order to select the number of common factors, let us consider the following sequence of tests:

$$H_0 = H(k^c) \qquad \text{against} \qquad H_1 = \bigcup_{0 \leq r < k^c} H(r),$$

for each $k^c = \underline{k}, \underline{k} - 1, ..., 1$. We propose the following statistic to test H_0 against H_1, for any given $k^c = \underline{k}, \ \underline{k} - 1, \ ..., \ 1$:

$$\hat{\xi}(k^c) \ = \ \sum_{\ell=1}^{k^c} \hat{\rho}_\ell. \tag{3.3.5}$$

The statistic $\hat{\xi}(k^c)$ corresponds to the sum of the k^c largest sample canonical correlations. We reject the null $H_0 = H(k^c)$ when $\hat{\xi}(k^c) - k^c$ is negative and large. The critical value is deduced by the large sample distribution provided in Section 3.4.

3.3.3 Inference on the number of common factors when k_1 and k_2 are unknown

The tests defined in Section 3.3.2 require the knowledge of the true number of pervasive factors $k_j > 0$ in each subgroup, $j = 1, 2$. When the true number of pervasive factors is not known, but consistent estimators $\hat{k}_1$ and $\hat{k}_2$, say, are available, the asymptotic distributions and rates of convergence for the test statistic $\hat{\xi}(k^c)$ based on $\hat{k}_1$ and $\hat{k}_2$ are the same as those of the test based on the true number of factors. Intuitively, this holds because the consistency of estimators $\hat{k}_j$, implies that $P(\hat{k}_j = k_j) \to 1$ for $j = 1, 2$, which means that the error due to the estimation of the number of pervasive factors is (asymptotically) negligible.[12]

The estimators based on the penalized information criteria of Bai and Ng (2002) applied on the two subgroups, are examples of consistent estimators for the numbers of pervasive factors. Therefore, in the next Section 3.4, the asymptotic distributions and rates of convergence of the test statistic and factors estimators are derived assuming that the true numbers of factors $k_j > 0$ in each subgroup, $j = 1, 2$, are known.

3.3.4 Estimation and inference in the mixed frequency factor model

The estimators and test statistics defined in Sections 3.3.1 - 3.3.3 for the group factor model (3.2.4) allow to define estimators for the loadings matrices Λ_{HC}, Λ_H, Λ_{LC}, Λ_L, the aggregated factor values $\overline{g}_t^U$, $U = C, H, L$ and the test statistic for the common factor space dimension k^C in equation (3.2.3). We denote these estimators $\hat{\Lambda}_{HC}$, $\hat{\Lambda}_H$, $\hat{\Lambda}_{LC}$, $\hat{\Lambda}_L$, $\hat{\overline{g}}_t^U$, and the test statistic $\hat{\xi}(k^C)$. The estimators of the common and high frequency factor values are:

$$\begin{pmatrix} \hat{g}_{m,t}^C \\ \hat{g}_{m,t}^H \end{pmatrix} \ = \ \left(\hat{\Lambda}_1'\hat{\Lambda}_1\right)^{-1} \hat{\Lambda}_1' x_{m,t}^H, \qquad m = 1, ..., M, \qquad t = 1, ..., T, \tag{3.3.6}$$

where $\hat{\Lambda}_1 = [\hat{\Lambda}_{HC} \vdots \hat{\Lambda}_H]$.

[12]This argument is formalized using similar arguments as, for instance, in footnote 5 of Bai (2003).

3.4 Large sample theory

In this section we derive the large sample distributions of the estimators of factor spaces and factor loadings, and of the test statistic for the dimension of the common factor space. We consider the joint asymptotics $N_1, N_2, T \to \infty$ under Assumptions 1-8 provided in Appendices 3.7.2 and 3.7.3. From the asymptotic theory of principal component analysis (PCA) estimators in large panels (see e.g. Bai and Ng (2002), Stock and Watson (2002a), Bai (2003), Bai and Ng (2006)) we know that:

$$\hat{h}_{j,t} \simeq \hat{\mathcal{H}}_j \left(h_{j,t} + \frac{1}{\sqrt{N_j}} u_{j,t} + \frac{1}{T} b_{j,t} \right), \quad j = 1, 2, \tag{3.4.1}$$

where $b_{j,t}$ is a deterministic bias term, the matrix $\hat{\mathcal{H}}_j$ converges to a non-singular matrix as $N_j, T \to \infty$, and:

$$\begin{aligned} u_{j,t} &:= \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i} \varepsilon_{j,it} \\ &= \left(\frac{\Lambda'_j \Lambda_j}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \Lambda'_j \varepsilon_{j,t}. \end{aligned} \tag{3.4.2}$$

Note that the terms $u_{j,t}$ depend also from the cross-sectional dimension N_j, but for notational convenience, we omit the index N_j in $u_{j,t}$. From Assumptions 2 and 5 *d)* the error terms $u_{j,t}$ are asymptotically Gaussian as $N_j \to \infty$:

$$u_{j,t} \xrightarrow{d} N(0, \Sigma_{u,j}), \tag{3.4.3}$$

where the asymptotic variance is:

$$\Sigma_{u,j} = \Sigma_{\Lambda,j}^{-1} \Omega_j \Sigma_{\Lambda,j}^{-1}, \tag{3.4.4}$$

and

$$\Sigma_{\Lambda,j} = \lim_{N_j \to \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda'_{j,i}, \tag{3.4.5}$$

$$\Omega_j = \lim_{N_j \to \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \sum_{\ell=1}^{N_j} \lambda_{j,i} \lambda'_{j,\ell} Cov(\varepsilon_{j,i,t}, \varepsilon_{j,\ell,t}), \quad j = 1, 2. \tag{3.4.6}$$

Without loss of generality, let $N_2 \leq N_1$. We assume $\sqrt{N_1}/T = o(1)$ (Assumption 6), which allows to neglect the bias terms $b_{j,t}/T$ in the asymptotic expansion (3.4.1). We also assume $T/N_2 = o(1)$, which further simplifies the asymptotic distributions derived in the next section.

3.4.1 Main asymptotic results for the group factor model

In this section we collect the main results concerning the asymptotic distributions of estimators and test statistics for the group factor model. Define the matrices:

$$
\begin{aligned}
\Omega_{j,k}(h) &= \lim_{N_j,\, N_k \to \infty} \frac{1}{\sqrt{N_j N_k}} \sum_{i=1}^{N_j} \sum_{\ell=1}^{N_k} \lambda_{j,i} \lambda'_{k,\ell} Cov(\varepsilon_{j,i,t}, \varepsilon_{k,\ell,t-h}), \qquad (3.4.7)\\
\Sigma_{u,jk}(h) &= \Sigma_{\Lambda,j}^{-1} \Omega_{jk}(h) \Sigma_{\Lambda,k}^{-1}, \qquad (3.4.8)
\end{aligned}
$$

for $j,k = 1,2$, and $h = ..., -1, 0, 1, ...$ Matrix $\Sigma_{u,jk}(h)$ is the asymptotic covariance between $u_{j,t}$ and $u_{k,t-h}$. Moreover, we have $\Omega_j \equiv \Omega_{j,j}(0)$ and $\Sigma_{u,j} \equiv \Sigma_{u,jj}(0)$, and similarly we define $\Sigma_{u,12} \equiv \Sigma_{u,12}(0) = \Sigma'_{u,21}$. Let us denote $N = \min\{N_1, N_2\} = N_2$ the minimal cross-sectional dimension among the two groups, and $\mu_N^2 = N_2/N_1 \leq 1$. Let $\mu_N \to \mu$, with $\mu \in [0,1]$. The boundary value $\mu = 0$ accounts for the possibility that N_1 grows faster than N_2.

Theorem 1. *Under Assumptions 1 - 6, and the null hypothesis $H_0 = H(k^c)$ of k^c common factors, we have:*

$$
N\sqrt{T}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N} tr\left\{\tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N}\right\}\right] \xrightarrow{d} N\left(0, \frac{1}{4}\Omega_U\right), \qquad (3.4.9)
$$

where

$$
\begin{aligned}
\tilde{\Sigma}_{cc} &= \frac{1}{T}\sum_{t=1}^{T} f_t^c f_t^{c\prime}, \qquad (3.4.10)\\
\Omega_U &= 2\sum_{h=-\infty}^{\infty} tr\left\{\Sigma_U(h)\Sigma_U(h)'\right\}, \qquad (3.4.11)\\
\Sigma_U(h) &= \mu^2 \Sigma_{u,11}^{(cc)}(h) + \Sigma_{u,22}^{(cc)}(h) - \mu\Sigma_{u,12}^{(cc)}(h) - \mu\Sigma_{u,21}^{(cc)}(h), \qquad (3.4.12)\\
\Sigma_{U,N} &= \mu_N^2 \Sigma_{u,1}^{(cc)} + \Sigma_{u,2}^{(cc)} - \mu_N \Sigma_{u,12}^{(cc)} - \mu_N \Sigma_{u,21}^{(cc)}, \qquad (3.4.13)
\end{aligned}
$$

and the upper index (c,c) denotes the upper-left (k^c, k^c) block of a matrix.

Proof: See Appendix 3.8.3.

The asymptotic distribution of $\hat{\xi}(k^c) - k^c$ after appropriate recentering and rescaling is Gaussian. The convergence rate is $N\sqrt{T}$. The asymptotic expansion of $\hat{\xi}(k^c) - k^c$ involves a time series average of squared estimation errors on group factors. Since these estimation errors are of order $1/\sqrt{N}$, the expected value of their square will be of order $1/N$, originating a recentering term of the second order analogous to an error-in-variable bias adjustment. Moreover, the averaging over time of the recentered squared estimation errors allows to apply a root-T central limit theorem for weakly dependent processes, originating a total estimation uncertainty for the test statistic of order $1/(N\sqrt{T})$.

Theorem 2. *Under Assumptions 1 - 6 we have:*

$$
\begin{aligned}
\sqrt{N_1}(\hat{\mathcal{H}}_c \hat{f}_t^c - f_t^c) &\xrightarrow{d} N\left(0, \Sigma_{u,1}^{(cc)}\right), \qquad (3.4.14)\\
\sqrt{N_2}(\hat{\mathcal{H}}_c^* \hat{f}_t^{c\,*} - f_t^c) &\xrightarrow{d} N\left(0, \Sigma_{u,2}^{(cc)}\right), \qquad (3.4.15)\\
\sqrt{N_j}\left[\hat{\mathcal{H}}_{s,j} \hat{f}_{j,t}^s - \left(f_{j,t}^s - (F_j^{s\prime} F^c)(F^{c\prime} F^c)^{-1} f_t^c\right)\right] &\xrightarrow{d} N\left(0, (\Sigma_{\Lambda,j}^{(ss)})^{-1} \Omega_j^{(ss)} (\Sigma_{\Lambda,j}^{(ss)})^{-1}\right), \qquad (3.4.16)
\end{aligned}
$$

for any j, t, where $\hat{\mathcal{H}}_c$, $\hat{\mathcal{H}}_c^*$ and $\hat{\mathcal{H}}_{s,j}$ are non-singular matrices, $F^c = [f_1^c, ..., f_T^c]'$, $F_j^s = [f_{j,1}^s, ..., f_{j,T}^s]'$ and the upper index (ss) denotes the lower-right (k_j^s, k_j^s) block of a matrix.

Proof: See Appendix 3.8.4.

From Theorem 2 a linear transformation of vector $\hat{f}_t^c$ (resp. $\hat{f}_t^{c*}$) estimates the common factor f_t^c at a rate $1/\sqrt{N_1}$ (resp. $1/\sqrt{N_2}$). The variance of the asymptotic Gaussian distribution is the upper-left (c, c) block of matrix $\Sigma_{u,1}$ (resp. $\Sigma_{u,2}$), i.e. the asymptotic variance of the estimation error $u_{1,t}$ (resp. $u_{2,t}$) for the PC vector in group 1 (resp. group 2). The estimation error for recovering the common factors from the group PC's is of order $1/\sqrt{NT}$, and therefore asymptotically negligible. The estimator $\hat{f}_{j,t}^s$ approximates the residual of the sample projection of the group-j specific factor on the common factor, up to a linear transformation, at rate $1/\sqrt{N_j}$.

Let us now derive the asymptotic distribution of the factor loadings estimators.[13] Define the matrices:

$$\Phi_{j,i} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{r=1}^{T} E[f_{j,t} f_{j,r}'] cov(\varepsilon_{j,i,t},\ \varepsilon_{j,i,r}), \tag{3.4.17}$$

$$\Psi_j = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{r=1}^{T} E\left[f_{j,t}^s f_{j,r}^{s\prime} \otimes f_t^c f_r^{c\prime}\right]. \tag{3.4.18}$$

Theorem 3. *Under Assumptions 1 - 6 we have:*

$$\sqrt{T}\left[\left(\hat{\mathcal{H}}_c'\right)^{-1} \hat{\lambda}_{j,i}^c - \lambda_{j,i}^c\right] \xrightarrow{d} N\left(0, \Phi_{j,i}^{(cc)} + (\lambda_{j,i}^{s\prime} \otimes I_{k^c})\Psi_j(\lambda_{j,i}^s \otimes I_{k^c})\right), \tag{3.4.19}$$

$$\sqrt{T}\left[\left(\hat{\mathcal{H}}_{s,j}'\right)^{-1} \hat{\lambda}_{j,i}^s - \lambda_{j,i}^s\right] \xrightarrow{d} N\left(0, \Phi_{j,i}^{(ss)}\right), \tag{3.4.20}$$

for any j, i, where $\hat{\mathcal{H}}_c$ and $\hat{\mathcal{H}}_{s,j}$, $j = 1, 2$, are the same non-singular matrices of Theorem 2.

Proof: See Appendix 3.8.4.

The factor loadings are estimated at rate $\sqrt{T}$. To get a feasible distributional result for the statistic $\hat{\xi}(k^c)$, we need consistent estimators for the unknown matrices $\tilde{\Sigma}_{cc}$, $\Sigma_{U,N}$ and Ω_U in Theorem 1. To simplify the analysis, we assume at this stage that the errors $\varepsilon_{j,it}$ are uncorrelated across subpanels j, individuals i and dates t (Assumption 7).[14] Then, we have:

$$\Sigma_{U,N} = \mu_N^2 \Sigma_{u,1}^{(cc)} + \Sigma_{u,2}^{(cc)}, \quad \Sigma_U(0) = \mu^2 \Sigma_{u,1}^{(cc)} + \Sigma_{u,2}^{(cc)}, \quad \Omega_U = 2tr\left\{\Sigma_U(0)^2\right\}. \tag{3.4.21}$$

In Theorem 4 below, we replace $\tilde{\Sigma}_{cc}$, $\Sigma_{U,N}$ and $\Sigma_U(0)$ by consistent estimators, such that the estimation error for $tr(\tilde{\Sigma}_{cc}^{-1}\Sigma_{U,N})$ in the bias adjustment is $o_p(1/\sqrt{T})$. Therefore, the asymptotic distribution of the statistic is unchanged.

---

[13] We assume that $\hat{f}_t^c$ is used for the estimation of the factor loadings. The distribution of the loadings estimators is analogous when using $\hat{f}_t^{c\,*}$ as common factor estimator.

[14] If the errors are weakly correlated across series and/or time, consistent estimation of $\Sigma_{U,N}$ and Ω_U requires thresholding of estimated cross-sectional covariances and/or HAC-type estimators.

Theorem 4. *Let* $\hat{\Sigma}_U = (N_2/N_1)\hat{\Sigma}_{u,1}^{(cc)} + \hat{\Sigma}_{u,2}^{(cc)}$, *with*

$$\hat{\Sigma}_{u,j} = \left(\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j}\right)^{-1}\left(\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j\hat{\Lambda}_j\right)\left(\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j}\right)^{-1}, \qquad j = 1, 2, \tag{3.4.22}$$

where $\hat{\Gamma}_j = diag(\hat{\gamma}_{j,ii},\ i = 1, ..., N_j)$, *and* $\hat{\Lambda}_j = [\hat{\Lambda}_j^c \vdots \hat{\Lambda}_j^s]$, *where* $\hat{\Lambda}_j^c$ *and* $\hat{\Lambda}_j^s$, *with* $j = 1, 2$, *are the loadings estimators defined in equations (3.3.3) and (3.3.4), and*

$$\hat{\gamma}_{j,ii} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{j,it}^2, \tag{3.4.23}$$

where $\hat{\varepsilon}_{j,it} = y_{j,it} - \hat{\lambda}_{j,i}^{c\,\prime}\hat{f}_t^c - \hat{\lambda}_{j,i}^{s\,\prime}\hat{f}_{j,t}^s$. *Moreover, let* $\hat{\Sigma}_{cc} = \frac{1}{T}\sum_{t=1}^{T}\hat{f}_t^c\hat{f}_t^{c\,\prime}$ *be the estimator of* $\tilde{\Sigma}_{cc}$. *Then, under Assumptions 1 - 7, and the null hypothesis* $H_0 = H(k^c)$ *of* k^c *common factors, we have:*

$$\tilde{\xi}(k^c) := N\sqrt{T}\left(\frac{1}{2}tr\{\hat{\Sigma}_U^2\}\right)^{-1/2}\left[\hat{\xi}(k^c) - k^c + \frac{1}{2N}tr\left\{\hat{\Sigma}_{cc}^{-1}\hat{\Sigma}_U\right\}\right] \xrightarrow{d} N(0,1). \tag{3.4.24}$$

Proof: See Appendix 3.8.5.

The feasible asymptotic distribution in Theorem 4 is the basis for a one-sided test of the null hypothesis of k^c common factors. If $\tilde{\xi}(k^c) < -1.64$, this null hypothesis is rejected at 5% level against the alternative hypothesis of less than k^c common factors.

3.4.2 Main asymptotic results for the mixed frequency factor model

In this section we give the asymptotic distribution for estimators of factor values in the mixed frequency factor model. The asymptotics is for $N_H, N_L, T \to \infty$, such that $N_L \leq N_H$, $\sqrt{N_H}/T = o(1)$, $N_L/T = o(1)$. Define the matrices:

$$\Omega_{\Lambda,m}^* = \lim_{N_H\to\infty}\frac{1}{N_H}\sum_{i=1}^{N_H}\sum_{\ell=1}^{N_H}\lambda_{1,i}\lambda_{1,\ell}' Cov(e_{m,t}^{i,H}, e_{m,t}^{\ell,H}), \qquad m = 1, ..., M, \tag{3.4.25}$$

where $\lambda_{1,i}'$ is the i-th row of the $(N_H, k^C + k^H)$ matrix $\Lambda_1 = [\ \Lambda_{HC} \vdots \Lambda_H\ ]$.

Theorem 5. *Under Assumptions 1 - 8 we have:*

$$\sqrt{N_H}(\hat{\mathcal{H}}_c\hat{g}_{m,t}^C - g_{m,t}^C) \xrightarrow{d} N\left(0, [\Sigma_{\Lambda,1}^{-1}\Omega_{\Lambda,m}^*\Sigma_{\Lambda,1}^{-1}]^{(CC)}\right), \tag{3.4.26}$$

$$\sqrt{N_H}\left[\hat{\mathcal{H}}_{1,s}\hat{g}_{m,t}^H - (g_{m,t}^H - (\bar{g}^{H\prime}\bar{g}^C)(\bar{g}^{C\prime}\bar{g}^C)^{-1})g_{m,t}^C\right] \xrightarrow{d} N\left(0, [\Sigma_{\Lambda,1}^{-1}\Omega_{\Lambda,m}^*\Sigma_{\Lambda,1}^{-1}]^{(HH)}\right) \tag{3.4.27}$$

for any m, t, *where* $\hat{\mathcal{H}}_c$ *and* $\hat{\mathcal{H}}_{1,s}$ *are the same non-singular matrices of Theorem 2,* $\bar{g}^C = [\bar{g}_1^C, ..., \bar{g}_T^C]'$, $\bar{g}^H = [\bar{g}_1^H, ..., \bar{g}_T^H]'$, $\Sigma_{\Lambda,1} = \lim_{N_H\to\infty}\frac{1}{N_H}\sum_{i=1}^{N_H}\lambda_{1,i}\lambda_{1,i}'$, *and indices* (CC) *and* (HH) *denote the upper-left* (k^C, k^C) *block and lower-right* (k^H, k^H) *block of a matrix, respectively.*

Proof: See Appendix 3.8.6.

From Theorem 5, a linear transformation of vector $\hat{g}^C_{m,t}$, resp. $\hat{g}^H_{m,t}$, estimates the common factor $g^C_{m,t}$, resp. the residual of the sample projection of the high-frequency factor on the common factor. The estimation rate is $\sqrt{N_H}$. There is no asymptotic effect from the error-in-variable problem induced by using estimated factor loadings in the cross-sectional regression when $T/N_H = o(1)$. The asymptotic distribution of the estimator $\hat{\bar{g}}^L_t$ of the aggregated low-frequency factor is deduced from Theorem 2.

3.5 Empirical application

It is worth summarizing the procedure underpinning the empirical analysis, for the benefit of the readers who skipped the previous sections. This is done in a first subsection.

3.5.1 Practical implementation of the procedure

We first assume that k^C, k^H, k^L, the number of respectively common, high and low frequency factors in equation (3.2.1), are known and all strictly larger than zero. The identification strategy presented in Section 3.2 directly implies a simple estimation procedure for the factor values and the factor loadings, which consists of the three following steps:

1. *PCA performed on the HF and LF panels separately*
   Define the (T, N_H) matrix of temporally aggregated (in our application flow-sampled) HF observables as $X^H = [x^H_1, ..., x^H_T]'$, and the (T, N_L) matrix of LF observables as $X^L = [x^L_1, ..., x^L_T]'$. The estimated pervasive factors of the HF data, which are collected in $(T, k^C + k^H)$ matrix $\hat{h}_H = [\hat{h}_{H,1}, ..., \hat{h}_{H,T}]'$, are obtained performing PCA on the HF data:
   $$\left(\frac{1}{TN_H} X^H X^{H\prime}\right) \hat{h}_H = \hat{h}_H \hat{V}_H, \tag{3.5.1}$$
   where $\hat{V}_H$ is the diagonal matrix of the eigenvalues of $(TN_H)^{-1} X^H X^{H\prime}$. Analogously, the estimated pervasive factors of the LF data, which are collected in the $(T, k^C + k^L)$ matrix $\hat{h}_L = [\hat{h}_{L,1}, ..., \hat{h}_{L,T}]'$, are obtained performing PCA on the LF data:
   $$\left(\frac{1}{TN_L} X^L X^{L\prime}\right) \hat{h}_L = \hat{h}_L \hat{V}_L, \tag{3.5.2}$$
   where $\hat{V}_L$ is the diagonal matrix of the eigenvalues of $(TN_L)^{-1} X^L X^{L\prime}$.

2. *Canonical correlation analysis performed on estimated principal components*
   Let $\hat{W}^C_U$ be the $(k^C + k^U, k^C)$ matrix whose columns are the canonical directions for $\hat{h}_{U,t}$ associated with the k^C largest canonical correlations between $\hat{h}_H$ and $\hat{h}_L$, for $U = H, L$. Then, the estimator of the (in our application flow sampled) common factor is $\hat{\bar{g}}^C_t = \hat{W}^{C\,\prime}_U \hat{h}_{U,t}$, for $U = H, L$ and $t = 1, ..., T$, and the estimated loadings matrices $\hat{\Lambda}_{HC}$ and $\hat{\Lambda}_C$ are obtained from the least squares

regressions of x_t^H and x_t^L on estimated factor $\hat{\hat{g}}_t^C$. Collect the residuals of these regressions:

$$\begin{aligned}\hat{\hat{\xi}}_t^H &= x_t^H - \hat{\Lambda}_{HC}\hat{\hat{g}}_t^C,\\ \hat{\hat{\xi}}_t^L &= x_t^L - \hat{\Lambda}_{LC}\hat{\hat{g}}_t^C,\end{aligned}$$

in the following (T, N_U), with $U = H, L$, matrices:

$$\hat{\hat{\Xi}}^U = \left[\hat{\hat{\xi}}_1^{U\,\prime}, ..., \hat{\hat{\xi}}_T^{U\,\prime}\right]', \qquad U = H, L.$$

Then the estimators of the HF-specific and LF-specific factors, collected in the (T, k^U), $U = H, L$, matrices:

$$\hat{G}^U = \left[\hat{\hat{g}}_1^{U\,\prime}, ..., \hat{\hat{g}}_T^{U\,\prime}\right]', \qquad U = H, L,$$

are obtained extracting the first k^H and k^L PCs from the matrices:

$$\left(\frac{1}{TN_U}\hat{\hat{\Xi}}^U\hat{\hat{\Xi}}^{U\,\prime}\right)\hat{G}^U = \hat{G}^U\hat{V}_S^U, \qquad U = H, L,$$

where $\hat{V}_S^U$, with $U = H, L$ are the diagonal matrices of the associated eigenvalues. Next, the estimated loadings matrices $\hat{\Lambda}_H$ and $\hat{\Lambda}_C$ are obtained from the least squares regression of $\hat{\xi}_t^H$ and $\hat{\xi}_t^L$ on respectively the estimated factors $\hat{\hat{g}}_t^H$ and $\hat{\hat{g}}_t^L$.

3. *Reconstruction of the common and high frequency-specific factors*
   The estimates of the common and HF-specific factors for each HF subperiod, denoted by $\hat{g}_{m,t}^C$ and $\hat{g}_{m,t}^H$, for any $m = 1, ..., M$ and $t = 1, ..., T$, are obtained by cross-sectional regression of $x_{m,t}$ on the estimated loadings $[\hat{\Lambda}_{HC} \vdots \hat{\Lambda}_H]$ obtained from the second step.

Inference on the number of common, low and high-frequency specific factors proceeds as follows:

- Suppose that $k_X := k^C + k^H$ and $k_Y := k^C + k^L$, i.e. the numbers of pervasive factors in panels X and Y, are known (consistent estimators: IC_{p1} and IC_{p2} criteria of Bai and Ng(2002)).

- Let $k^* := \min(k_X, k_Y)$, we develop a test for:

  $$H_0 : k^C = r \qquad \text{against} \qquad H_1 : k^C < r,$$

  for any given $r = k^*,\ k^* - 1,\ ...,\ 1$.

- We use the statistic defined in equation (3.3.5), namely: $\hat{\xi}(r) = \sum_{\ell=1}^{r} \hat{\rho}_\ell$, where $\hat{\rho}_\ell$, $\ell = 1, ..., r$, are the r largest canonical correlations between $\hat{h}_{H,t}$ and $\hat{h}_{L,t}$ (i.e. the empirical analogs of $h_{H,t}$ and $h_{L,t}$).

3.5.2 Data description

The data consists of a combination of IP and non-IP sectors. For industrial production we use the same data on 117 IP sectoral indices considered by Foerster, Sarte, and Watson (2011), sampled at quarterly frequency from 1977.Q1 to 2011.Q4.[15] These indices correspond to the finest level of disaggregation for the sectoral components of the IP aggregate index which can be matched with the available sectors in the *Input-Output* and *Capital Use* tables used in the structural analysis in Section 3.5.4. The data for all the remaining non-IP sectors consist of the annual growth rates of real GDP for the following 42 sectors: 35 services, Construction, Farms, Forestry-Fishing and related activities, General government (federal), Government enterprises (federal), General government (state and local) and Government enterprises (state and local). These LF data are available from 1977 until 2011 and are published by the Bureau of Economic Analysis (BEA).[16] Moreover, as IP is a Gross Output measure, in the structural analysis it is convenient to consider the yearly growth rates of real Gross Output (GO) for the non-IP sectors. These data are available from 1988 until 2011 and are also published by the BEA. Following the sectoral productivity literature, in the structural analysis we focus exclusively on the private sectors, and therefore exclude four Government Gross Output indices, reducing the sample size to 38 non-IP sectors indices. All growth rates refer to seasonally adjusted real output indices, and are expressed in percentage points.[17]

Figure 3.2: Growth rates of the Industrial Production and Gross Domestic Product indices

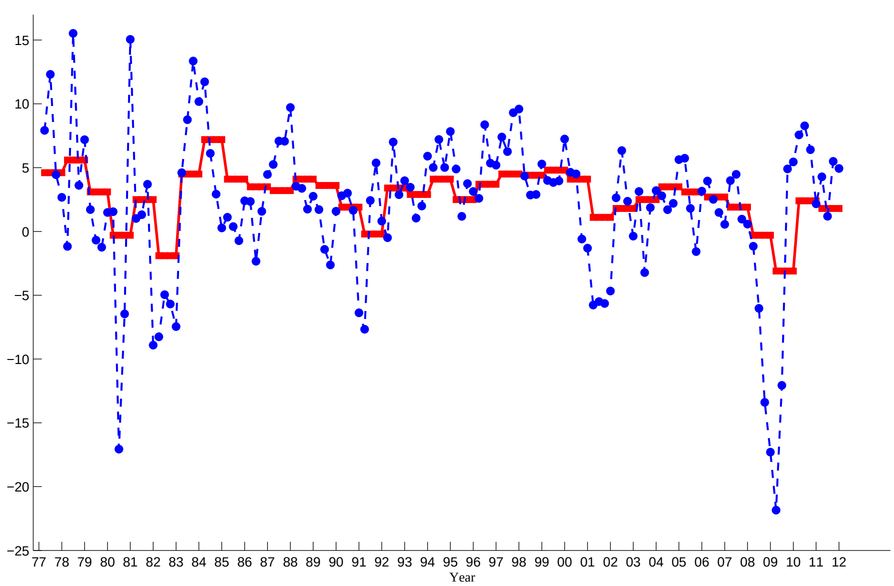


Figure 3.2 displays the growth rates of the aggregate Industrial Production (dotted (blue) quarterly data) and

[15]The IP data are available also at monthly frequency. Following Foerster, Sarte, and Watson (2011), we focus only on quarterly IP data, as they share the main feature of the monthly ones, but are less noisy.

[16]GDP data are available at quarterly frequency for the aggregate index, but not for sectoral ones. As in the remaining part of the paper we study comovements among different sectors, we consider the panel of yearly GDP sectoral data.

[17]A detailed description of the dataset is provided in Appendix C 3.9.3.

Gross Domestic Product (solid line (red) annual data) indices over the sample period from 1977 until 2011. The objective of this empirical application is to use our mixed frequency factor model to capture the major sources of comovement among the sectoral constituents of these two indices, which are the most reliable measures of US economic activity.

3.5.3 Factors common to all US sectors

We assume that our dataset follows the factor structure for flow sampling as in equation (3.2.2), with $x_{m,t}^H$ and x_t^L corresponding to respectively quarterly IP and annual non-IP data. Let $X^H = [x_1^H, ..., x_T^H]'$, with $x_t^H := \sum_{m=1}^4 x_{m,t}^H$, be the (T, N_H) panel of the yearly observations of the IP indices growth rates (computed as the sum of the quarterly growth rates $x_{m,t}^H$, $m = 1, ..., 4$ for year t), and let $X^L = [x_1^L, ..., x_T^L]'$ be the (T, N_L) panel of the yearly growth rates of the non-IP indices. Let also $X_{HF} = [x_{1,1}^H, x_{2,1}^H, ..., x_{m,t}^H, ..., x_{4,T}^H]'$ be the $(4T, N_H)$ panel of IP indices quarterly growth rates.

We start by selecting the number of factors in each subpanel, which are of dimensions $k_X = k^C + k^H$ and $k_Y = k^C + k^L$, respectively. We use the IC_{p2} information criteria of Bai and Ng (2002), and report the results in Table 3.1. Results for other criteria are in Appendix C 3.9.5.

Table 3.1: Estimated number of factors

	X_{HF}	X^H	X^L	$[\ X^H\ \ X^L\ ]$
IP data: 1977.Q1-2011.Q. Non-IP data: Gross Domestic Product, 1977-2011				
IC_{p2}	1	2	1	1
IP data: 1988.Q1-2011.Q4. Non-IP data: Gross Output, 1988-2011				
IC_{p2}	1	1	2	2

The number of latent pervasive factors selected by the IC_{p2} information criteria is reported for different subpanels. Subpanels X_{HF} and X^H correspond to IP data sampled at quarterly and yearly frequency, respectively. Panels X^L and $[\ X^H\ \ X^L\ ]$ correspond to non-IP data, and the stacked panels of IP and non-IP data, respectively. We use $k_{max} = 15$ as maximum number of factors when computing IC_{p2}.

Table 3.1 corroborates the evidence in Foerster, Sarte, and Watson (2011) suggesting that there is either one or perhaps two pervasive factors in the IP data ($k_X = 1$ or $k_X = 2$). Likewise, for the non-IP data, we also find evidence in favor of either one or two pervasive factors ($k_Y = 1$ or $k_Y = 2$).

Table 3.2: Canonical Correlations and Tests for Common Factors

$\hat{\rho}_1$	$\hat{\rho}_2$	$\tilde{\xi}(2)$	$\tilde{\xi}(1)$
0.84	0.06	-3.56	-1.56

IP data: 1977.Q1-2011.Q4, Non-IP data: GDP, 1977-2011. We report the two largest canonical correlations among the PCs computed from each subpanel of IP and non-IP data, and the values of $\tilde{\xi}(r)$, the feasible standardized value of the test statistic $\hat{\xi}(r)$, for the null hypothesis of $r = 2$ or $r = 1$ common factors, respectively.

Table 3.3: Adjusted R^2 of regressions on common factors from indices growth rates

Panel A

Factors	$\bar{R}^2$: Quantile 10%	25%	50%	75%	90%
Observables: Gross Domestic Product, 1977-2011					
common	-2.2	-0.5	11.5	28.9	42.9
common, LF-spec.	0.1	9.2	25.4	34.5	60.3
LF-spec.	-2.8	-2.3	5.7	15.7	22.4
Observables: IP, 1977.Q1-2011.Q4					
common	0.3	4.8	20.3	36.0	60.0
common, HF-spec.	1.1	6.8	28.7	45.3	63.4
HF-spec.	-0.7	-0.1	3.0	11.2	23.5

Panel B

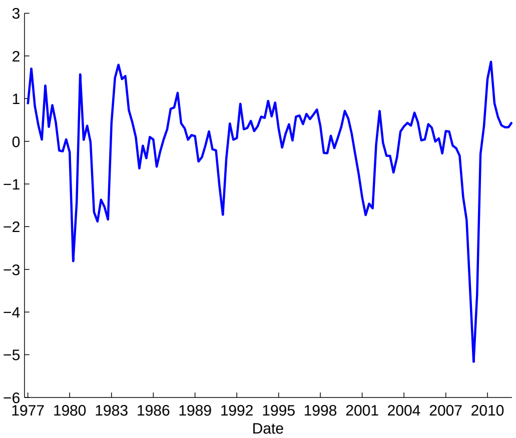
Factors	$\bar{R}^2$: Quantile 10%	25%	50%	75%	90%
Observables: Gross Output, 1988-2011					
common	-2.0	6.6	28.2	45.6	64.5
common, LF-spec.	2.8	15.2	45.0	63.7	70.8
LF-spec.	-4.5	-3.8	3.2	13.4	40.7
Observables: IP, 1988.Q1-2011.Q4					
common	0.1	3.5	10.5	29.8	48.2
common, HF-spec.	0.8	7.9	28.2	43.2	65.4
HF-spec.	-0.8	2.0	10.0	21.9	33.9

Panel A. The regressions in the first three lines involve the growth rates of the 42 non-IP sectors as dependent variables, while those in the last tree lines involve the growth rates of the 117 IP indices as dependent variables. The explanatory variables are factors estimated from the same indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1977-2011. *Panel B*. The regressions in the first three lines involve the Gross Output growth rates growth of the 38 non-IP as dependent variables, while those in the last tree lines involve the growth of the 117 IP indices as dependent variables. The explanatory variables are factors estimated from the same indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1988-2011.

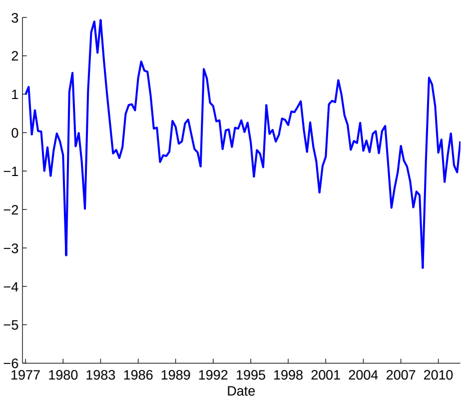
Let us consider the dataset where the HF data are quarterly IP indices, and the LF data are annual GDP non-IP indices. In order to select the number of common and frequency-specific factors, we follow the procedure detailed in Section 3.5.1. In Table 3.2 we report the estimated canonical correlations of the first two PC's estimated in each subpanel X^H and X^L, which are used to compute the value of the test statistic $\hat{\xi}(r)$, for the null hypothesis of r = 2 or r = 1 common factors.[18] We note that the first canonical correlation is close to one, which is consistent with the presence of one common factor in each of the two mixed frequency datasets considered. The test rejects the null hypothesis r = 2, i.e. the presence of two common factors, for any significance level, while we cannot reject the null of one common factor with a 5% significance level. In light of the results in Tables 3.1 and 3.2 we select a model with $k^C = k^H = k^L = 1$. The factors are then obtained using the estimation procedure of Section 3.5.1. In Figure 3.3 we plot the estimated factors from the panels of 42 GDP sectors and 117 IP indices on the entire sample going from 1977 to 2011. All factors are standardized to have zero mean and unit variance, and their sign is chosen so that the majority of the associated loadings are positive. A visual inspection of the plots in Figure 3.3 reveals that the common factor in Panel (a) resembles the IP index of Figure 3.2, with a large decline corresponding to the Great Recession following the financial crisis of 2007-2008 and the positive spike associated to the recent economic recovery. On the other hand, the LF-specific factor features a less dramatic fall during the Great Recession, and actually features a positive spike in 2008, followed by large negative values in the following years. This constitutes preliminary evidence suggesting that some non-IP sectors could feature different responses to the financial crisis of 2007-2008. The interpretation of factors is easier when they are used as explanatory variables in standard regression analysis. We start with a disaggregated analysis, and

[18]We extract the first two PC's in each subgroup, compute the matrix $\hat{R}$ as defined in equation 3.3.2 and compute the canonical correlations as the square root of its two largest eigenvalues.

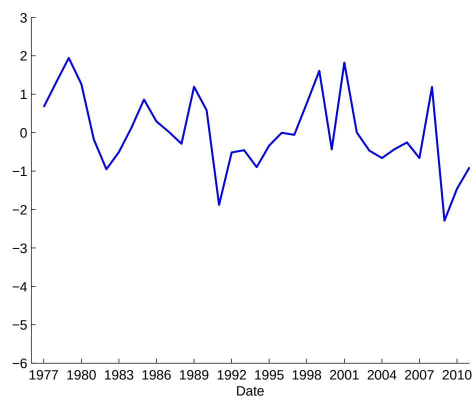
Figure 3.3: Sample paths of the estimated common and specific factors



(a) Common factor



(b) HF specific factor



(c) LF specific factor

Panel (a) displays the time series plot of the estimated common factor. Panel (b) displays that of the HF-specific factor and finally Panel (c) that of the LF-specific factor. The factors are estimated from the panels of 42 non-IP GDP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period is 1977.Q1-2011.Q4.

look at the relative importance of the common and frequency specific factors in explaining the variability across all sectoral growth rates. For each sector in the panel, we regress the index growth rates on (i) the common factor only, (ii) on the specific factor only, and (iii) on both common and specific factors. In Table 3.3 we report the quantiles of the empirical distribution of the adjusted R^2 (denoted $\bar{R}^2$) of these regressions. In the first and fourth rows of Panels A and B we report the quantiles of $\bar{R}^2$ of the regressions involving as explanatory variable the common factor only, in the second and fifth rows we report the quantiles of $\bar{R}^2$ when the explanatory variables are the common and frequency-specific factors. Finally, the quantiles of $\bar{R}^2$ in the third and sixth rows refer to regressions where the explanatory variable is the frequency-specific factor only.[19]

From the first three lines of Panel A we observe that adding the LF specific factor to the common factor regressions for the non-IP indices yields an increment of the median $\bar{R}^2$ around 14%, going from 11.5% to 25.4%, and for more than 10% of the sectors the $\bar{R}^2$ increases at least by 17%. On the other hand, the HF-specific factor, when added to the common factor, contributes less to the increments in $\bar{R}^2$ for the IP sectors. In Panel B we note that for at least half of both the IP and non-IP Gross Output sectoral indices, the frequency-specific factors contribute to an increase in $\bar{R}^2$ of at least 15% when added to the common factor. Overall, Table 3 confirms that the common and frequency-specific factors explain a significant part of the variability of output growth for the majority of the sectors of the US economy. Moreover, the common factor is pervasive for most of the IP and non-IP sectors alike.

In order to give economic interpretation to the estimated factors, we list in Table 3.4 the top and bottom ten GDP non-IP sectors in terms of $\bar{R}^2$ when regressed on the common factor only, and both the common and LF-specific factors. We also report the top and bottom ten GDP non-IP sectors with the highest and lowest absolute increments in $\bar{R}^2$ when the LF-specific factor is added to the common one.[20] From Panel A we first note that the common factor explains most of the variability of service sectors with direct economic links to industrial production sectors like Transportation and Warehousing: for instance, Truck Transportation, Other Transportation & Support Activities, and Warehousing & Storage have an $\bar{R}^2$ of 63%, 43% and 41%, respectively, when regressed on the common factor only. This is a clear indication that the common factor could be interpreted as IP factor. On the other hand, the common factor is completely unrelated to Agriculture, forestry, fishing & hunting, most of the Financial and Information services sectors.

Turning to Panel C, we note that the LF-specific factor explains more than 20% of the variability of output for very heterogeneous services sectors like Miscellaneous professional, scientific, & technical services, Administrative & support services, Legal services, Real Estate, some important financial services like Credit intermediation, & Related activities, Rental & Leasing Services but also Government (state & local). Interpreting these results, we can conclude that the LF-specific factor is completely unrelated to service sectors which depend almost exclusively on IP output, and is a common factor driving the comovement of non-IP sectors such as some Services, Construction and Government.

In Table 3.4 we highlight further differences in the dynamics of output growth between the two subsectors of the financial services industry which are particularly revealing: "Securities" and "Credit intermediation", extensively studied by Greenwood and Scharfstein (2013). We find that the subsectors "Funds, trusts, & other financial vehicles" and "Securities, commodity contracts, & investments" are unrelated to both the common and LF-specific factors, indicating that their output growth is uncorrelated with the common component of real output growth across the other sectors of the US economy. In contrast, the "credit intermediation" industry comoves with the other IP and non-IP sectors.[21]

Up to this point, we looked at the explanatory power of the factors for sectoral output indices. For both the non-IP GDP and Gross Output, these indices correspond to the finest level of disaggregation of output growth by sector. In Table 3.5 we report the results of regressions with aggregated indices instead. In particular, we regress the output of each aggregate index either on the estimated common factor or the common and frequency specific factors, and focus on the adjusted R^2s of these regressions. It is also important to note

[19]The regressions in the second and third rows are restricted MIDAS regressions. The regressions in fourth, fifth and sixth rows impose the estimated coefficients of the common and HF-specific factors to be the same for each quarter, as they are estimated as HF regressions.

[20]The entire lists of ordered non-IP sectors for the three panels in Table 3.4 is available in Tables 3.17-3.19 in Appendix C.

[21]See also Tables 3.17 and 3.18 in Appendix C.

Table 3.4: Regression of yearly sectoral GDP growth on the common and LF-specific factors: adjusted R^2

Panel A. Regressor: common factor

Sector	$\bar{R}^2$
Ten sectors with largest $\bar{R}^2$	
Truck transportation	63.10
Accommodation	62.43
Construction	44.05
Other transp. & support activ.	43.31
Administrative & support services	42.69
Other services, except government	42.53
Warehousing & storage	40.95
Air transportation	31.58
Retail trade	30.70
Amusem., gambling, & recr. ind.	29.17
Ten sectors with smallest $\bar{R}^2$	
Funds, trusts, & other finan. vehicles	-1.23
Motion picture & sound record. ind.	-1.68
Pipeline transportation	-1.74
Information & data processing services	-1.84
Transit & ground passenger transp.	-2.05
General government (state & local)	-2.12
Forestry, fishing, & related activities	-2.33
Water transportation	-2.94
Securities, commodity contracts, & investm.	-2.99
Insurance carriers & related activities	-3.03

Panel B. Regressors: common and LF spec. factors

Sector	$\bar{R}^2$
Ten sectors with largest $\bar{R}^2$	
Misc. prof., scient., & tech. serv.	66.67
Administrative & support services	62.63
Truck transportation	62.51
Accommodation	61.48
Construction	59.75
Warehousing & storage	52.53
Government enterprises (state & local)	45.78
Other services, except government	41.75
Other transportation & support activities	41.71
Government enterprises (federal)	37.78
Ten sectors with smallest $\bar{R}^2$	
Ambulatory health care services	7.76
Management of companies & enterprises	7.52
Funds, trusts, & other fin. vehicles	6.15
Information & data processing services	1.96
Educational services	1.35
Insurance carriers & related activities	0.36
Water transportation	-0.64
Farms	-1.87
Forestry, fishing, & related activities	-5.31
Securities, commodity contracts, & investm.	-5.99

Panel C. Increment in adjusted R^2

Sector	$\Delta\bar{R}^2$
Ten sectors with largest change in $\bar{R}^2$	
Misc. prof., scient., & tech. serv.	49.69
Government enterprises (state & local)	34.69
Rental & leasing serv. & lessors of int. assets	29.52
General government (state & local)	24.90
Legal services	24.32
Motion picture & sound recording ind.	22.77
Fed. Reserve banks, credit interm., & rel. activ.	20.31
Administrative & support services	19.95
Social assistance	19.91
Real estate	18.14
Ten sectors with smallest change in $\bar{R}^2$	
Accommodation	-0.96
Rail transportation	-1.16
Other transportation & support activities	-1.59
Air transportation	-1.77
Retail trade	-2.15
Amusements, gambling, & recreation ind.	-2.15
Educational services	-2.62
Farms	-2.80
Forestry, fishing, & related activities	-2.98
Securities, commodity contracts, & investm.	-3.00

In the table we report the adjusted R^2, denoted $\bar{R}^2$, for restricted MIDAS regressions of the growth rates of 42 GDP non-IP sectoral indices on the estimated factors. The factors are estimated from the panel of 42 GDP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1977-2011. Regressions in *Panel A* involve a LF explained variable and the estimated common factor. Regressions in *Panel B* involve a LF explained variable and both the common and LF-specific estimated factors. The regressions in both tables are restricted MIDAS regressions. In *Panel C* we report the difference in $\bar{R}^2$ (denoted as $\Delta\bar{R}^2$) between the regressions in *Panel B* and regressions in *Panel A*.

Table 3.5: Adj. R^2 of aggregate IP and selected GDP indices growth rates on estimated factors

Panel A *Quarterly observations, 1977.Q1-2011.Q4*

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(H)$	(3) $\bar{R}^2(C+H)$	(3) - (1)
Industrial Production	89.06	5.02	90.26	1.20

Panel B *Yearly observations, 1977-2011*

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(C+L)$	(3) - (1)
GDP	60.54	8.59	74.21	13.67
GDP - Manufacturing	81.88	-3.03	81.53	-0.35
GDP - Agriculture, forestry, fishing, and hunting	1.43	-2.52	-1.26	-2.69
GDP - Construction	44.05	11.22	59.75	15.70
GDP - Wholesale trade	20.35	7.90	30.83	10.48
GDP - Retail trade	30.70	-2.86	28.56	-2.15
GDP - Transportation and warehousing	62.14	-2.95	60.97	-1.17
GDP - Information	12.14	22.28	37.57	25.43
GDP - Finance, insurance, real estate, rental, and leasing	-1.42	21.22	21.11	22.53
GDP - Professional and business serv.	30.02	30.21	65.61	35.59
GDP - Educational serv., health care, and social assist.	-1.38	18.38	18.18	19.56
GDP - Arts, entert., recreat., accomm., and food serv.	53.51	-2.23	53.70	0.18
GDP - Government	-2.12	22.37	20.47	22.59

In the table we report the adjusted R^2, denoted $\bar{R}^2$, of the regression of growth rates of the aggregate IP index and selected aggregated sectoral GDP non-IP output indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(H)$ and $\bar{R}^2(L)$) only, and the common and frequency-specific factors together (column (3)). The last column displays the difference between the values in the third and first columns. The factors are estimated from the panel of 42 GDP non-IP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1977-2011.

that we also include the GDP Manufacturing aggregate index which is *not* used in the estimation of the factors. This will help us with the interpretation of the factors - common and frequency-specific - which we obtained.

Panel A of Table 3.5 shows that the common factor explains around 90% of the variability in the aggregate IP index. This implies that the common factor can be interpreted as an Industrial Production factor. This is further corroborated in Panel B where we find an $\bar{R}^2$ around 82% for the regression of the GDP Manufacturing Index on the common factor only. As most of the sectors included in the Industrial Production index are Manufacturing sectors, this result is not surprising, but is still worth noting because, as noted earlier, the GDP data on Manufacturing have not been used in the factor estimation, in order not to double-count these sectors in our mixed frequency sectoral panel.[22] As expected from the results in Table 3.4, more than 60% of the variability of GDP of Transportation and Warehousing services index is explained by the common

[22] A detailed discussion of the difference in the sectoral components of the IP index and the GDP Manufacturing index is provided in Appendix 3.9.3.

factor only, and the LF-specific factor has no explanatory power. On the other hand, the HF-specific factor seems not to be important in explaining the aggregate IP index, as the $\bar{R}^2$ increases only by 1% when it is added as a regressor to the common factor.[23] This suggests that the HF-specific factor is pervasive only for a subgroup of IP sectors which have relatively low weights in the index, meaning that their aggregate output is a negligible part of the output of the entire IP sector and, consequently, also the entire US economy.[24]

Looking at the aggregate GDP index, we first note that even if the weight of Industrial Production sectors in the aggregate nominal GDP index has always been below 30%, as evident from Figure 3.1, still 60% of its total variability can be explained exclusively by the common factor which - as shown in Panel A - is primarily an IP factor. This implies that there must be substantial comovement between IP and some important service sectors. Moreover, it appears from the first entry in Panel B that a relevant part of the variability of the aggregate GDP index not due to the common factor is explained by the LF-specific factor (the $\bar{R}^2$ increases by about 14% to 74%).[25] This indicates that significant comovements are present among the most important sectors of the US economy which are not related to manufacturing. Indeed, Panel B in Table 3.5 indicates that some services sectors such as Professional & Business Services and Information and Construction load significantly both on the common and the LF-specific factor, while some other sectors like Finance and Government load exclusively on the LF-specific factor.[26]

3.5.4 Structural model and productivity shocks

The macroeconomics literature, with the works of Long and Plosser (1983), Horvath (1998) and Carvalho (2010), among many others, has recognized that input-output linkages in both intermediate materials and capital goods lead to propagation of sector-specific shocks in a way that generates comovements across sectors. An important contribution of the work of Foerster, Sarte, and Watson (2011) is to describe the conditions under which an approximate linear factor structure for sectoral output growth arises from standard neoclassical multisector models including those linkages. In particular, they develop a generalized version of the multisector growth model of Horvath (1998), which allows them to filter out the effects of these linkages, and reconstruct the time series of productivity shocks using sector data on output growth when input-output tables for intermediate materials and capital goods are available. We can characterize this as statistical versus structural factor analysis.

The main objective of this section is to verify the presence of a common factor in the *innovations* of productivity for *all* the sectors (not just IP) of the US economy by means of our mixed frequency factor model. If a common factor is present also in the productivity shocks, then the factor structure uncovered by the reduced form analysis of output growth in Section 3.5.3 is not only due to interlinkages in materials and capital use among different sectors.

We rely on the same multi-industry real business cycle model described in Section IV of Foerster, Sarte, and Watson (2011) to extract productivity shocks from the time series of the growth rates of the same 117 IP indices considered in the previous section, and the growth rates of 38 non-IP Gross Output of private sectors,

[23]See also Table 3.20 in Appendix 3.9.5 for the $\bar{R}^2$ of the regression of all GDP indices on the HF factor only, and all the 3 factors together.

[24]These results corroborate the findings of Foerster, Sarte, and Watson (2011), who claim that the main results of their paper are qualitatively the same when considering either one or two common factors extracted from the same 117 IP indices of our study.

[25]See the results in Table 3.20 in Appendix C.

[26]The results change when we look at Finance sector disaggregated in Credit Intermediation, “Securities”, Insurance and Real estate, as evident in Table 3.4.

therefore excluding the 4 Government indices considered previously.[27] One challenge due to the mixed frequency nature of our output growth dataset consists in the extraction of mixed frequency technological shocks. In Appendix C 3.9.4 we explain how to adapt the algorithm proposed by Foerster, Sarte, and Watson (2011) to estimate technological shocks for our mixed frequency output series. Specifically, the multi-sector business cycle model that we use to filter out the technological shocks correspond to the "Benchmark" model considered by Foerster, Sarte, and Watson (2011) in their Section IV, while the data on input-output and capital use matrices necessary to estimate the model are built from the BEA's 1997 "use table" and "capital flow table", respectively.[28] Using the extracted productivity shocks for the IP and non-IP sectors, denoted $\hat{\varepsilon}_{m,t}^{H}$ and $\hat{\varepsilon}_{t}^{L}$, respectively, we estimate a mixed frequency factor model with these productivity shock series. The sample period for the estimation of both the factor model and the regressions is 1989-2011, because the productivity shocks can not be computed for the first year of the sample (see Foerster, Sarte, and Watson (2011), especially equation (B38) on page 10 of their Appendix B). For a direct comparison between the statistical factor model covered in the previous subsection and the structural factor analysis, we need to first re-estimate our model with one common, one HF-specific and one LF-specific factors on the panels of growth rates of annual Gross Output non-IP indices (as opposed to the GDP growth indices in Table 3.5) and the same 117 quarterly sectoral IP indices. The results are reported in Table 3.6.

For the moment we focus exclusively on the shaded areas of Table 3.6, as the non-shaded areas pertain to the productivity shocks which will be covered later. We expect some difference with the previous results for at least two reasons. First, the dataset in which the non-IP data are Gross Output indices, refers to shorter time period going from 1988, instead of 1977, to the end of 2011, as Gross Output indices are not available before 1988. Second, as the panel in Table 3.6 does not include the four governmental sectors, we expect that the common and frequency-specific factors may have different dynamics when compared to those extracted from the panel with GDP non-IP sectors.

We obtain qualitatively similar results, as shown in Table 3.6. There appear to be only two notable differences with the results reported in Table 3.5. We see an increased importance of the HF-specific factor in explaining the variability of the IP aggregate index (see Panel A in Table 3.6), at the expense of a lower explanatory power for the common factor. Moreover, there is also an increased importance of both the common and LF-specific factors in explaining the total variability of total aggregate output (measured as total Gross Output, in the first line of Panel B in Table 3.6). Still the common factor explains roughly 65 % of the variation in the panel of IP data.

What do we learn from the structural analysis with common and frequency-specific factors of productivity shocks? First, it is remarkable to find that again there is one common factor in productivity shocks. Indeed, the selection of the number of common factors is performed as in the previous section, and our testing methodology suggests the presence of one common factor. Therefore we estimate a model for the productivity innovations with $k^C = k^H = k^L = 1$.[29] As in the previous section, we start with a disaggregated analysis and look at the relative importance of the new common and frequency specific factors in explaining the variability of the constituents of the panel of productivity innovations, and the panels of all output growth rates used for the extraction of the productivity innovation themselves. For each sector, we regress both the productivity innovations and the index growth rates on the common factor only, on the specific factor only, and on both common and specific factors. In Table 3.7 we report the quantiles of the empirical distribution

[27]The exclusion of the public sector from the analysis is a standard choice in the sectoral productivity literature.

[28]The last year for which sectoral capital use tables have been constructed by the BEA is 1997.

[29]The values of the penalized selection criteria of Bai and Ng (2002) performed on different subpanels and the test for the number of common factors are available in Tables 3.22 and 3.23 in Appendix C 3.9.5.

Table 3.6: Adj. R^2 of aggregate IP and selected Gross Output indices growth rates on estimated factors (shaded) and estimated factors from productivity innovations

Panel A *Quarterly observations, 1988.Q1-2011.Q4*

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(H)$	(3) $\bar{R}^2(C+H)$	(3) - (1)
Industrial Production	63.71	38.32	89.48	25.78
	31.21	50.15	77.25	46.05

Panel B *Yearly observations, 1988-2011*

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(C+L)$	(3) - (1)
GO (all sectors)	68.54	12.20	89.66	21.12
	42.17	13.97	57.60	15.43
GO - Manufacturing	86.08	-3.05	88.94	2.86
	62.29	-0.20	64.42	2.13
GO - Agriculture, forestry, fishing, and hunting	-3.21	3.35	-0.25	2.96
	0.96	-4.23	-3.35	-4.31
GO - Wholesale trade	80.82	-3.85	79.97	-0.85
	74.73	-3.08	74.74	0.01
GO - Construction	25.30	34.16	67.15	41.84
	6.64	20.55	27.78	21.14
GO - Retail trade	64.72	-4.50	63.15	-1.57
	47.02	-4.35	45.04	-1.98
GO - Transportation and warehousing	83.82	-4.51	83.22	-0.60
	70.42	-2.69	70.58	0.15
GO - Information	33.70	38.59	81.54	47.84
	17.78	42.45	61.76	43.98
GO - Finance, insurance, real estate, rental, and leasing	3.37	50.30	59.29	55.92
	-4.09	17.55	13.96	18.05
GO - Professional and business services	45.13	21.97	75.48	30.36
	25.17	44.89	71.81	46.64
GO - Educational serv., health care, and social assist.	-4.19	-1.58	-6.17	-1.98
	-4.73	-4.48	-9.66	-4.93
GO - Arts, entert., recreat., accomm., and food serv.	71.06	-3.74	71.90	0.84
	55.64	-2.29	55.49	-0.16

In the table we display the adjusted R^2, denoted $\bar{R}^2$, of the regressions of growth rates of the aggregate IP index and selected aggregated sectoral Gross Output non-IP output indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(H)$ and $\bar{R}^2(L)$)only, and the common and frequency-specific factor together (column (3)). The last column displays the difference between the values in the third and first columns. The factors are estimated from the panel of 38 Gross Output non-IP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1988-2011.

Table 3.7: Adjusted R^2 of regressions on common factors from productivity innovations

Panel A

Factors	Adjusted R^2: Quantile 10%	25%	50%	75%	90%
Observables: Gross Output productivity innovations, 1989-2011					
common	-3.3	-0.3	11.0	33.6	46.1
common, LF-spec.	-2.6	4.8	26.3	45.0	60.7
LF-spec.	-4.2	-3.6	-0.1	17.7	33.1
Observables: IP productivity innovations, 1989.Q1-2011.Q4					
common	-1.0	-0.4	1.5	12.1	22.4
common, HF-spec.	-0.6	3.1	13.1	28.4	40.1
HF-spec.	-0.7	0.6	6.2	18.7	28.2

Panel B

Factors	Adjusted R^2: Quantile 10%	25%	50%	75%	90%
Observables: Gross Output, 1988-2011					
common	-2.4	3.7	21.2	31.5	55.8
common, LF-spec.	-0.9	7.8	28.2	56.9	68.0
LF-spec.	-4.6	-3.3	1.3	20.6	43.8
Observables: IP,1988.Q1-2011.Q4					
common	-0.8	0.2	4.5	17.7	34.7
common, HF-spec.	1.2	5.9	25.7	40.8	63.8
HF-spec.	-0.3	2.2	14.7	29.2	37.8

Panel A: The regressions in the first three lines involve the productivity innovations of the 38 non-IP sectors as dependent variables, while the regressions in the last tree lines involve the productivity innovations of the 117 IP indices as dependent variables. Productivity innovations are computed using the panel of Gross Output growth rates for the LF observables. The explanatory variables are factors estimated from a mixed frequency factor model with $k^C = k^H = k^L = 1$, on the panels of productivity innovations filtered adapting the procedure of Foerster, Sarte, and Watson (2011). The sample period for the estimation of both the factor model and the regressions is 1989.Q1-2011.Q4. Panel B: The regressions in the first three lines involve the Gross Output growth rates of the 38 non-IP sectors as dependent variables, while the regressions in the last tree lines involve the growth of the 117 IP indices as dependent variables. The explanatory variables are the same factors used in the regressions of Panel A. The sample period for the estimation of both the factor model and the regressions is 1989.Q1-2011.Q4. Productivity innovations are computed using the panel of Gross Output growth rates for the LF observables.

of $\bar{R}^2$ of these regressions.[30]

Panel A of Table 3.7 confirms that both the common and the frequency-specific factors are pervasive for the panels of productivity innovations. From the first two rows we note that the common factor alone explains at least 11% of the variability of half of the non-IP series considered, and this fraction increases to more than 26 % when the LF-specific factor is added as regressor to the common one. On the other hand, from the last three rows of we note that for the panels of IP the high frequency specific factor seems to explain the majority of the variability of the productivity indices, while the explanatory power of the common factor only seem to be significant only for 50% of the IP sectors. Panel B reports the $\bar{R}^2$ of the regressions of the GO indices growth rates on the factors estimated on the panels of productivity shocks themselves. Therefore, they give an indication of the fraction of variability of the indices explained by the common components of the output growth which is not due to input-output linkages between sectors, as captured by the structural "Benchmark" of Foerster, Sarte, and Watson (2011). Panel B of Table 3.7 can be compared with Panel B of Table 3.3. As expected, as part of the comovement among different sectors is due to input-output and capital use linkages, all the $\bar{R}^2$ in Panel B of Table 3.7 are strictly lower than those in Table 3.3, if we exclude the negative ones and those very close to zero. For instance the median $\bar{R}^2$ of regressions including the common only factor for the non-IP sectors decrease from 28% to 21%, and median $\bar{R}^2$ of regressions including the common and LF-specific factors decreases from 45% to 28%. A similar pattern is observed for the higher quantiles, and for the IP indices. Overall, Panel B gives a first indication of the presence of commonality in

[30]The regressions in the second and third rows are restricted MIDAS regressions. The regressions in fourth, fifth and sixth rows impose the estimated coefficients of the common and HF-specific factors to be the same at each quarter, as they are estimated as HF regressions.

the comovement on the majority of the sectors of the US economy even when the output growth rates are purged of the input-output linkages in both intermediate materials and capital goods.

We conclude the analysis repeating the same exercise of Table 3.6 shaded areas, and regress the Industrial Production and aggregate (mostly non-IP) Gross Output indices growth on the factors extracted from productivity innovations and look at the adjusted R^2s in the non-shaded rows of Table 3.6. From Panel A we observe that the common extracted from productivity innovations explains around 31% of the variability of the aggregate IP index, i.e. around half of the variability explained by the common factor extracted directly from the output series. Moreover, when the high frequency-specific productivity factor is added as explanatory variable, the $\bar{R}^2$ increases to 77% which is also significantly smaller than the 89% $\bar{R}^2$ obtained using as regressors the factors extracted from the output series.[31] Hence, the case of a common pervasive factor in innovation shocks across the entire economy mainly related to IP sector technology shocks is less compelling. From Panel B we observe that 42% of the variability of the aggregate Gross Output of the US economy can be explained by the common factor of productivity shocks, and when the factor specific to non-IP sector is added, the $\bar{R}^2$ grows to 57%.

From this analysis we learn something interesting which Foerster, Sarte, and Watson (2011) were not able to address since they exclusively examined IP sectors. Overall there is a difference in the explanatory power of factors in structural versus non-structural factor models - as they found. However, it seems that looking at technology shocks instead of output, it does not appear that a common factor explaining IP fluctuations is a dominant factor for the entire economy. A factor specific to technological innovations in IP sectors is more important for the IP sector shocks and a low frequency factor which appears to explain variation in information industry as well as professional and business services innovations plays, relatively speaking, a more important role.

[31] See in particular Panel A of Table 3.6. This result is in line with the findings of Foerster, Sarte, and Watson (2011) in their Section IV C.

3.5.5 Subsample analysis

Our sample covers what is known as the Great Moderation, which refers to a reduction in the volatility of business cycle fluctuations starting in the mid-1980s. We turn therefore to analyzing subsamples. We start by selecting the number of pervasive factors in each subpanel, using the IC_{p2} information criteria, and report the results in Table 3.8. In Table 3.9 we report the canonical correlation analysis common factor tests. We consider two subsample configurations: 1984.Q1-2007.Q4 and 1984.Q1-2011.Q4. The former is the Great Moderation sample considered by Foerster, Sarte, and Watson (2011) whereas the second is an augmented subsample including the Great Depression. In light of the results in Tables 3.8 and 3.9 we select a model with $k^C = k^H = k^L = 1$, for both subsamples. The factors for both datasets are obtained using the estimation procedure described in Section 3.5.1.[32]

Table 3.8: Estimated number of factors for different subsamples

	X_{HF}	X^H	X^L	$[\ X^H\ \ X^L\ ]$
IP data: 1984.Q1-2007.Q4. Non-IP data: Gross Domestic Product, 1984-2007				
IC_{p2}	1	2	1	1
IP data: 1984.Q1-2011.Q4. Non-IP data: Gross Domestic Product, 1984-2011				
IC_{p2}	1	2	1	1

The number of latent pervasive factors selected by the IC_{p2} information criteria is reported for different subpanels and different sample periods. Subpanels X_{HF} and X^H correspond to IP data sampled at quarterly and yearly frequency, respectively. Panels X^L and $[\ X^H\ \ X^L\ ]$ correspond to non-IP data, and the stacked panels of IP and non-IP data, respectively. We use $k_{max} = 15$ as maximum number of factors when computing IC_{p2}.

Table 3.9: Canonical Correlations and Tests for Common Factors

$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\tilde{\xi}(3)$	$\tilde{\xi}(2)$	$\tilde{\xi}(1)$
IP data: 1984.Q1-2007.Q4. Non-IP data: Gross Domestic Product, 1984-2007					
0.81	0.13	-	-	-6.61	-2.98
0.87	0.57	0.45	-3.15	-2.74	-1.03
IP data: 1984.Q1-2011.Q4. Non-IP data: Gross Domestic Product, 1984-2011					
0.70	0.33	-	-	-1.67	-1.28

We report the canonical correlations of the first two PCs computed in each subpanel of IP and non-IP data, and the values of $\tilde{\xi}(r)$, the estimated value of the test statistic $\hat{\xi}(r)$, for the null hypothesis of $r = 3, 2, 1$ common factors, respectively.

[32]For the shorter sample 1984.Q1-2007.Q4, selecting a model with $k_1 = k_2 = 3$ pervasive factors in each subpanel, we reject the null hypotheses of 3 and 2 common factor, while we cannot reject the null of 1 common factor. Regression results for $k^C = 1$ and $k^H = k^L = 2$ are very similar than those presented in Table 3.10, i.e. for a model with $k^C = k^H = k^L = 1$ factors, and therefore are omitted.

Table 3.10: Adj. R^2 of aggregate IP and selected GDP indices growth rates on estimated factors

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(H)$	(3) $\bar{R}^2(C+H)$	(3) - (1)
Panel A *Quarterly observations IP*				
IP 1984.Q1-2007.Q4	72.48	10.58	80.02	7.54
IP 1984.Q1-2011.Q4	80.11	16.83	88.87	8.76

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(C+L)$	(3) - (1)
Panel B *Yearly observations, 1984-2007*				
GDP	29.22	39.24	76.71	47.49
GDP - Manufacturing	70.69	-3.85	71.18	0.50
GDP - Agriculture, forestry, fishing, and hunting	0.81	-0.87	0.51	-0.30
GDP - Construction	13.02	50.30	70.39	57.37
GDP - Wholesale trade	-4.40	21.36	18.09	22.49
GDP - Retail trade	-0.44	58.14	62.65	63.09
GDP - Transportation and warehousing	41.43	11.16	52.02	10.59
GDP - Information	-4.37	-4.10	-8.83	-4.46
GDP - Finance, insurance, real estate, rental, and leasing	-3.78	-0.60	-4.78	-1.00
GDP - Professional and business services	4.89	56.09	67.06	62.18
GDP - Educational serv., health care, and social assist.	-3.81	3.31	-0.20	3.61
GDP - Arts, entert., recreat., accomm., and food serv.	13.66	37.32	57.01	43.35
GDP - Government	0.74	14.51	14.83	14.09
Panel C *Yearly observations, 1984-2011*				
GDP	56.33	14.88	77.87	21.55
GDP - Manufacturing	83.78	-3.85	83.37	-0.41
GDP - Agriculture, forestry, fishing, and hunting	-3.64	-2.65	-6.59	-2.95
GDP - Construction	40.54	21.76	68.61	28.07
GDP - Wholesale trade	23.62	10.48	37.71	14.09
GDP - Retail trade	20.70	6.76	30.39	9.69
GDP - Transportation and warehousing	65.17	1.10	67.14	1.97
GDP - Information	6.20	9.23	17.35	11.14
GDP - Finance, insurance, real estate, rental, and leasing	-1.95	5.04	3.68	5.64
GDP - Professional and business services	27.59	30.75	64.39	36.80
GDP - Educational serv., health care, and social assist.	-0.73	-0.90	-2.00	-1.27
GDP - Arts, entert., recreat., accomm., and food serv.	56.94	1.56	62.97	6.03
GDP - Government	0.50	18.75	19.03	18.53

In the table we report the adjusted R^2, denoted $\bar{R}^2$, of the regression of growth rates of the aggregate IP index and selected aggregated sectoral GDP non-IP output indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(H)$ and $\bar{R}^2(L)$) only, and the common and frequency-specific factor together(column (3)). The last column displays the difference between the values in the third and first columns. The factors are estimated from the panel of 42 GDP non-IP sectors and 117 IP indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample periods for the estimation of both factor model and regressions are 1984-2007 (Great Moderation), and 1984-2011.

In Table 3.10 we report the results of regressions of aggregated version of the indices used for the estimation on the same factors considered in the full samples. This allows us to understand if, and to what extent, the most important sectors of the US economy comoved over the different subsamples. Again, we regress the output of each aggregate index on the estimated common factor only, the common and frequency specific factors, and concentrate our attention on the adjusted R^2s of these regressions. The results in Table 3.10 indicate that in general there is a deterioration of the overall fit of approximate factor models during the Great Moderation, i.e. during the sample starting in 1984 and ending 2007 – a finding also reported by Foerster, Sarte, and Watson (2011) – and that the common factor plays a lesser role during the Great Moderation. According to the results in Panel A, the common factor only explains roughly 72 % of the variation across IP sectors, but interestingly when the financial crisis is added to the Great Moderation subsample, we see again a pattern closer to the full sample results reported in the previous subsection. This also transpires from Panels B and C, when examining the total GDP variations projected on the common factor. During the Great Moderation the common factor only explained around 30 %, which goes to 56 % when we add the Great Depression. The other patterns, i.e. the exposure of the various subindices, appear to be similar to those in the full sample.

3.6 Conclusions

Panels with data sampled at different frequencies are the rule rather than the exception in economic applications. We develop a novel approximate factor modeling approach which allows us to estimate factors which are common across all data regardless of their sample frequency, versus factors which are specific to subpanels stratified by sampling frequency. To develop the generic theoretical framework, we cast our analysis into a group factor structure and develop a unified asymptotic theory for the identification of common and group- or frequency-specific factors, for the determination of the number of common and specific factors, for the estimation of loadings and the factors via principal component analysis in a setting with large dimensional data sets, using asymptotic expansions both in the cross-sections and the time series.

There are a plethora of applications to which our theoretical analysis applies. We selected a specific example based on the work of Foerster, Sarte, and Watson (2011) who analyzed the dynamics of comovements across 117 industrial production sectors using both statistical and structural factor models. We revisit their analysis and incorporate the rest, and most dominant part of the US economy, namely the non-IP sectors which we only observe annually.

Despite the generality of our analysis, we can think of many possible extensions, such as models with loadings which change across subperiods (i.e. periodic loadings) or loading which vary stochastically or feature structural breaks. All these extensions are left for future research.

3.7 Appendix A: Identification with stock sampling and assumptions

3.7.1 Identification: stock sampling

In the case of stock sampling, the low frequency observations of $x_{m,t}^{L*}$ in the factor model (3.2.1) are the values of $x_{M,t}^{L*}$, i.e. $x_t^L = y_{M,t}^{L*}$. Then, the model for the observable variables becomes:

$$\begin{aligned} x_{m,t}^H &= \Lambda_{HC} g_{m,t}^C + \Lambda_H g_{m,t}^H + e_{m,t}^H, \quad m = 1, ..., M, \\ x_t^L &= \Lambda_{LC} g_{M,t}^C + \Lambda_L g_{M,t}^L + e_{M,t}^L. \end{aligned}$$

We stack the observations $x_{m,t}$ and y_t of the last high frequency subperiod and write:

$$\begin{bmatrix} x_t^H \\ x_t^L \end{bmatrix} = \begin{bmatrix} \Lambda_{HC} & \Lambda_H & 0 \\ \Lambda_{LC} & 0 & \Lambda_L \end{bmatrix} \begin{bmatrix} g_{M,t}^C \\ g_{M,t}^H \\ g_{M,t}^L \end{bmatrix} + \begin{bmatrix} e_{M,t}^H \\ e_{M,t}^L \end{bmatrix}. \tag{3.7.1}$$

This equation corresponds to a group factor model, with common factor $g_{M,t}^C$ and "group-specific" factors $g_{M,t}^H$, $g_{M,t}^L$. Therefore, the factor values $g_{M,t}^C$, $f_{M,t}^H$, $f_{M,t}^L$, and the factor loadings Λ_{HC}, Λ_{LC}, Λ_H, Λ_L, are identifiable up to a sign as proved in Section 3.2.1 (see also results in e.g. Schott (1999), Wang (2012), Chen (2010, 2012)).

Once the factor loadings are identified from (3.7.1), the values of the common and high frequency factors for subperiods $m = 1, ..., M-1$ are identifiable by cross-sectional regression of the high frequency data on loadings Λ_{HC} and Λ_H in (3.2.1). More precisely, $g_{m,t}^C$ and $g_{m,t}^H$ are identified by regressing $x_{m,t}^{Hi}$ on $\lambda_{HC,i}$ and $\lambda_{H,i}$ across $i = 1, 2, ..., N_H$, for any $m = 1, ..., M-1$ and any t. To summarize, with stock sampling, we can identify the common factor $g_{m,t}^C$ and the high frequency factor $g_{m,t}^H$ at all high frequency subperiods. We cannot estimate $g_{m,t}^L$, for $m < M$, as only $g_{M,t}^L$ is identified by the last paired panel data set consisting of $x_{M,t}^H$ combined with x_t^L. This is not surprising, since we have no HF observation available for the LF process.

3.7.2 Assumptions: group factor model

Let $\|A\| = \sqrt{tr(A'A)}$ denote the Frobenius norm of matrix A. Let $k^F = k^c + k_1^s + k_2^s$, and define the k_F-dimensional vector of factors: $F_t = [\ f_t^{c\,\prime},\ \ f_{1,t}^{s\,\prime},\ \ f_{2,t}^{s\,\prime}\ ]'$, and the (T, k_F) matrix $F = [\ F_1',\ ...,\ F_T'\ ]'$. We make the following assumptions:

Assumption A. 1. *The unobservable factor process is such that $F'F/T = \Sigma_F + O_p(1/\sqrt{T})$ as $T \to \infty$, where Σ_F is a positive definite $(k^F \times k^F)$ matrix defined as:*

$$\Sigma_F = \begin{bmatrix} I_{k^c} & 0 & 0 \\ 0 & I_{k_1^s} & \Phi \\ 0 & \Phi' & I_{k_2^s} \end{bmatrix}. \tag{3.7.2}$$

Assumption A. 2. *The loadings matrices $\Lambda_1 = \begin{bmatrix} \Lambda_1^c \vdots \Lambda_1^s \end{bmatrix}$ and $\Lambda_2 = \begin{bmatrix} \Lambda_2^c \vdots \Lambda_2^s \end{bmatrix}$ are full column-rank, for N_1, N_2 large enough. The loadings $\lambda_{j,i}$ are such that:*

$$\frac{\Lambda_j'\Lambda_j}{N_j} = \Sigma_{\Lambda,j} + O\left(\frac{1}{\sqrt{N_j}}\right), \qquad j = 1, 2, \tag{3.7.3}$$

where $\Sigma_{\Lambda,j} \; := \lim_{N_j \to \infty} \left(\frac{\Lambda_j' \Lambda_j}{N_j} \right)$ *is a p.d.* (k_j, k_j) *matrix, for* $j = 1, 2$.

Assumption A. 3. *The error terms* $(\varepsilon_{1,it} \;\; \varepsilon_{2,it})'$ *are weakly dependent across* i *and* t*, and such that* $E[\varepsilon_{j,it}] = 0$.

Assumption A. 4. *There exists a constant* C_ε *such that* $E[\varepsilon_{j,it}^4] \leq C_\varepsilon$ *for all* j*,* i *and* t.

Assumption A. 5. *a) The variables* F_t *and* $\varepsilon_{j,is}$ *are independent, for all* i, j, t *and* s.
b) The processes $\{\varepsilon_{j,it}\}$ *are stationary, for all* j, i.
c) The process $\{F_t\}$ *is stationary and weakly dependent over time.*
d) For each j *and* t*, as* $N_j \to \infty$*, it holds:*

$$\frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i} \varepsilon_{j,it} \;\xrightarrow{d}\; N(0, \Omega_j), \tag{3.7.4}$$

where $\Omega_j = \lim_{N_j \to \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \sum_{\ell=1}^{N_j} \lambda_{j,i} \lambda_{j,\ell}' E[\varepsilon_{j,it} \varepsilon_{j,\ell t}]$.

Assumption A. 6. *The asymptotic analysis is for* $N_1, N_2, T \to \infty$ *such that* $N_2 \leq N_1$*,* $T/N_2 = o(1)$*,* $\sqrt{N_1}/T = o(1)$.

The following Assumption 7 simplifies the derivation of the feasible asymptotic distribution of the statistic used to test the dimension of the common factor space k^c.

Assumption A. 7. *The error terms* $\varepsilon_{j,it}$ *are uncorrelated across* j*,* i *and* t*, and* $\varepsilon_{j,it} \sim (0, \gamma_{j,ii})$.

Assumption 7 is a stronger condition than Assumptions 3 and 5 *b)*. Moreover, under Assumption 7, the matrix Ω_j in Assumption 5 *d)* simplifies to $\Omega_j = \lim_{N_j \to \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \lambda_{j,i} \lambda_{j,i}' \gamma_{j,ii}$.

3.7.3 Assumptions: mixed frequency factor model

Let $\lambda_{1,i}'$ be the i-th row of the $(N_H, k^C + k^H)$ matrix $\Lambda_1 = [\, \Lambda_{HC} \vdots \Lambda_H \,]$. We make the following assumption:

Assumption A. 8. *The variables* $\lambda_{1,i}$ *and* $e_{m,t}^{i,H}$ *are such that:*

$$\frac{1}{\sqrt{N_H}} \sum_{i=1}^{N_H} \lambda_{1,i} e_{m,t}^{i,H} \;\xrightarrow{d}\; N(0, \Omega_{\Lambda,m}^*), \tag{3.7.5}$$

where

$$\Omega_{\Lambda,m}^* \quad = \quad \lim_{N_H \to \infty} \frac{1}{N_H} \sum_{i=1}^{N_H} \sum_{\ell=1}^{N_H} \lambda_{1,i} \lambda_{1,\ell}' Cov(e_{m,t}^{i,H}, e_{m,t}^{\ell,H}), \qquad m = 1, ..., M. \tag{3.7.6}$$

3.8 Appendix B: Proofs of Propositions, Theorems and Lemmas

3.8.1 Proof of Proposition 1

By replacing equation (3.2.7) into model (3.2.4), we get

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \Lambda_1^c A_{11} + \Lambda_1^s A_{21} & \Lambda_1^c A_{12} + \Lambda_1^s A_{22} & \Lambda_1^c A_{13} + \Lambda_1^s A_{23} \\ \Lambda_2^c A_{11} + \Lambda_2^s A_{31} & \Lambda_2^c A_{12} + \Lambda_2^s A_{32} & \Lambda_2^c A_{13} + \Lambda_2^s A_{33} \end{bmatrix} \begin{bmatrix} \tilde{f}_t^c \\ \tilde{f}_{1,t}^s \\ \tilde{f}_{2,t}^s \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}. \tag{3.8.1}$$

This factor model satisfies the restrictions in the loading matrix appearing in equation (3.2.4) if, and only if,

$$\Lambda_1^c A_{13} + \Lambda_1^s A_{23} = 0, \tag{3.8.2}$$

$$\Lambda_2^c A_{12} + \Lambda_2^s A_{32} = 0. \tag{3.8.3}$$

Equations (3.8.2) and (3.8.3) can be written as linear homogeneous systems of equations for the elements of matrices $[A_{13}'\ A_{23}']'$ and $[A_{12}'\ A_{32}']'$:

$$\left[\Lambda_1^c \vdots \Lambda_1^s\right] \begin{bmatrix} A_{13} \\ A_{23} \end{bmatrix} = 0, \quad \text{and} \quad \left[\Lambda_1^c \vdots \Lambda_2^s\right] \begin{bmatrix} A_{12} \\ A_{32} \end{bmatrix} = 0.$$

Since $\left[\Lambda_1^c \vdots \Lambda_1^s\right]$ and $\left[\Lambda_2^c \vdots \Lambda_2^s\right]$ are full column rank, it follows that

$$A_{13} = 0,\ A_{23} = 0, \tag{3.8.4}$$

$$A_{12} = 0,\ A_{32} = 0. \tag{3.8.5}$$

Therefore, the transformation of the factors that is compatible with the restrictions on the loading matrix in equation (3.2.4) is:

$$\begin{bmatrix} f_t^c \\ f_{1,t}^s \\ f_{2,t}^s \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ A_{31} & 0 & A_{33} \end{bmatrix} \begin{bmatrix} \tilde{f}_t^c \\ \tilde{f}_{1,t}^s \\ \tilde{f}_{2,t}^s \end{bmatrix}.$$

We can invert this transformation and write:

$$\begin{aligned} \tilde{f}_t^c &= A_{11}^{-1} f_t^c, \\ \tilde{f}_{1,t}^s &= A_{22}^{-1} f_{1,t}^s - A_{22}^{-1} A_{21} A_{11}^{-1} f_t^c, \\ \tilde{f}_{2,t}^s &= A_{33}^{-1} f_{2,t}^s - A_{33}^{-1} A_{31} A_{11}^{-1} f_t^c. \end{aligned}$$

The transformed factors satisfy the normalization restrictions in (3.2.6) if, and only if,

$$Cov(\tilde{f}_{1,t}^s,\ \tilde{f}_t^c) = -A_{22}^{-1} A_{21} A_{11}^{-1} (A_{11}^{-1})' = 0, \tag{3.8.6}$$

$$Cov(\tilde{f}_{2,t}^s,\ \tilde{f}_t^c) = -A_{33}^{-1} A_{31} A_{11}^{-1} (A_{11}^{-1})' = 0, \tag{3.8.7}$$

$$V(\tilde{f}_t^c) = A_{11}^{-1} (A_{11}^{-1})' = I_{k^c}, \tag{3.8.8}$$

$$V(\tilde{f}_{1,t}^s) = A_{22}^{-1} (A_{22}^{-1})' + A_{22}^{-1} A_{21} A_{11}^{-1} (A_{11}^{-1})' A_{21}' (A_{22}^{-1})' = I_{k_1^s}, \tag{3.8.9}$$

$$V(\tilde{f}_{2,t}^s) = A_{33}^{-1} (A_{33}^{-1})' + A_{33}^{-1} A_{31} A_{11}^{-1} (A_{11}^{-1})' A_{31}' (A_{33}^{-1})' = I_{k_2^s}, \tag{3.8.10}$$

Since the matrices A_{11}, A_{22} and A_{33} are nonsingular, equations (3.8.6) and (3.8.7) imply

$$A_{21} = 0, \text{ and } A_{31} = 0. \tag{3.8.11}$$

Then, from equations (3.8.8) - (3.8.10), we get that matrices A_{11}, A_{22} and A_{33} are orthogonal.

Q.E.D.

3.8.2 Proof of Proposition 2

From equation (3.2.6) we have

$$R = \begin{pmatrix} I_{k^c} & 0 \\ 0 & \Phi\Phi' \end{pmatrix} \text{ and } R^* = \begin{pmatrix} I_{k^c} & 0 \\ 0 & \Phi'\Phi \end{pmatrix}.$$

Matrix R is block diagonal, and the upper-left block I_{k^c} has eigenvalue 1 with multiplicity k^c. The associated eigenspace is $\{(\xi', 0')', \xi \in \mathbb{R}^{k^c}\}$. The lower-right block $\Phi\Phi'$ is a positive semi-definite matrix, and its largest eigenvalue is $\tilde{\rho}^2$, where $\tilde{\rho}^2 = \sup\left\{\xi_1'\Phi\Phi'\xi_1 : \xi_1 \in \mathbb{R}^{k_1^s}, \|\xi_1\| = 1\right\} < 1$ is the first squared canonical correlation of vectors $f_{1,t}^s$ and $f_{2,t}^s$. Therefore, we deduce that the largest eigenvalue of matrix R is equal to 1, with multiplicity k^c, and the associated eigenspace, denoted by $\mathcal{E}_c$, is spanned by vectors $(\xi', 0')'$, with $\xi \in \mathbb{R}^{k^c}$. Let S_1 be an orthogonal (k^c, k^c) matrix, then the columns of the (k_1, k^c) matrix

$$W_1 = \begin{pmatrix} S_1 \\ 0_{k_1^s \times k^c} \end{pmatrix}$$

are an orthonormal basis of the eigenspace $\mathcal{E}_c$. We have:

$$W_1' h_{1,t} = S_1' f_t^c. \tag{3.8.12}$$

Analogous arguments allow to show that the largest eigenvalue of matrix R^* is equal to 1, with multiplicity k^c and that the associated eigenspace , denoted by $\mathcal{E}_c^*$, is spanned by vectors $(\xi^{*\prime}, 0')'$, with $\xi^* \in \mathbb{R}^{k^c}$. Let S_2 be an orthogonal (k^c, k^c) matrix. Then, the columns of the (k_2, k^c) matrix

$$W_2 = \begin{pmatrix} S_2 \\ 0_{k_2^s \times k^c} \end{pmatrix}$$

are an orthonormal basis of the eigenspace $\mathcal{E}_c^*$. We have:

$$W_2' h_{2,t} = S_2' f_t^c, \tag{3.8.13}$$

which yields parts *i)* and *ii)*.

When there is no common factor, the matrix R becomes $R = \Phi\Phi'$, and matrix R^* becomes $R^* = \Phi'\Phi$. By the above arguments, the largest eigenvalue of matrix R, which is equal to the largest eigenvalue of matrix R^*, is not larger than $\tilde{\rho}^2$, where $\tilde{\rho}^2 < 1$ is the first squared canonical correlation between the two group-specific factors. This yields part *iii)*.

Finally, we prove part *iv)*. We showed that the lower-right block $\Phi\Phi'$ of matrix R is a positive semi-definite matrix and all its $k_1^s = k_1 - k^c$ eigenvalues are strictly smaller than one. These are also eigenvalues of matrix R. Let us denote the space spanned by the associated k_1^s eigenvectors of matrix R by $\mathcal{E}_{s,1}$. This space is spanned by vectors $(0', \tilde{\xi}')'$ with $\tilde{\xi} \in \mathbb{R}^{k_1^s}$. We note that, by construction, the vectors $(0', \tilde{\xi}')'$ are linearly independent of the vectors $(\xi', 0')'$ spanning the eigenspace $\mathcal{E}_c$. Let Q_1 be an orthogonal (k_1^s, k_1^s) matrix, then the columns of matrix

$$W_1^s = \begin{pmatrix} 0_{k^c \times k_1^s} \\ Q_1 \end{pmatrix}$$

are an orthonormal basis of the eigenspace $\mathcal{E}_{s,1}$. We have:

$$W_1^{s\prime} h_{1,t} = Q_1' f_{1,t}^s. \tag{3.8.14}$$

Analogously, we have that the lower-right block $\Phi'\Phi$ of matrix R^* is a positive semi-definite matrix and all its $k_2^s = k_2 - k^c$ eigenvalues are strictly smaller than one. These are also eigenvalues of matrix R^*. Let us denote the space spanned by the associated k_2^s eigenvectors of matrix R^* by $\mathcal{E}_{s,2}$. This space is spanned by vectors $(0', \tilde{\xi}^{*\prime})'$ with $\tilde{\xi}^* \in \mathbb{R}^{k_2^s}$. We note that, by construction, the vectors $(0', \tilde{\xi}^{*\prime})'$ are linearly independent of the vectors $(\xi^{*\prime}, 0')'$ spanning the eigenspace $\mathcal{E}_c^*$. Let Q_2 be an orthogonal (k_2^s, k_2^s) matrix, then the columns of matrix

$$W_2^s = \begin{pmatrix} 0_{k^c \times k_2^s} \\ Q_2 \end{pmatrix}$$

are an orthonormal basis of the eigenspace $\mathcal{E}_{s,2}$. We have:

$$W_2^{s\prime} h_{2,t} = Q_2' f_{2,t}^s. \tag{3.8.15}$$

Q.E.D.

3.8.3 Proof of Theorem 1

Asymptotic expansion of $\hat{R}$

In order to derive the asymptotic distribution of the test statistic $\hat{\xi}(k^c)$ defined in equation (3.3.5), and common factor estimator introduced in Definition 2, we consider a perturbation of matrix $\hat{R}$ and its eigenvalues and eigenvectors. More precisely, the perturbation of the eigenvalues will allow us to derive the asymptotic distribution of the test statistic $\hat{\xi}(k^c)$, while the perturbation of the eigenvectors will allow us to derive the asymptotic distribution of the common factor estimator.

The canonical correlations and the canonical directions are invariant to one-to-one transformations of the vectors $\hat{h}_{1,t}$ and $\hat{h}_{2,t}$ (see, among others, Anderson (2003)). Therefore, without loss of generality, for the asymptotic analysis of the estimator of the dimension of the common factor space statistic $\hat{\xi}(k^c)$, we can set $\hat{\mathcal{H}}_j = I_{k_j}$, $j = 1, 2$, in approximation (3.4.1). Moreover, under Assumption 6 the bias term is negligible, and we get:

$$\hat{h}_{j,t} \simeq h_{j,t} + \frac{1}{\sqrt{N_j}} u_{j,t}, \quad j = 1, 2. \tag{3.8.16}$$

By using approximation (3.8.16), and $N_2 = N$, $N_1 = N/\mu_N{}^2$, we have:

$$\begin{aligned}
\hat{V}_{12} &= \frac{1}{T}\sum_{t=1}^{T} \hat{h}_{1,t}\hat{h}'_{2,t} \\
&\simeq \frac{1}{T}\sum_{t=1}^{T} \left(h_{1,t} + \frac{1}{\sqrt{N}}\mu_N u_{1,t}\right)\left(h_{2,t} + \frac{1}{\sqrt{N}}u_{2,t}\right)' \\
&= \tilde{V}_{12} + \hat{X}_{12},
\end{aligned}$$

where:

$$\begin{aligned}
\tilde{V}_{12} &= \frac{1}{T}\sum_{t=1}^{T} h_{1,t}h'_{2,t}, \\
\hat{X}_{12} &= \frac{1}{T\sqrt{N}}\sum_{t=1}^{T}(h_{1,t}u'_{2,t} + \mu_N u_{1,t}h'_{2,t}) + \frac{\mu_N}{TN}\sum_{t=1}^{T} u_{1,t}u'_{2,t}. \qquad (3.8.17)
\end{aligned}$$

Similarly:

$$\begin{aligned}
\hat{V}_{jj} &= \frac{1}{T}\sum_{t=1}^{T} \hat{h}_{j,t}\hat{h}'_{j,t} \\
&\simeq \frac{1}{T}\sum_{t=1}^{T} \left(h_{j,t} + \frac{1}{\sqrt{N_j}}u_{j,t}\right)\left(h_{j,t} + \frac{1}{\sqrt{N_j}}u_{j,t}\right)' \\
&= \tilde{V}_{jj} + \hat{X}_{jj} \qquad (3.8.18) \\
&= \tilde{V}_{jj}\left(Id + \tilde{V}_{jj}^{-1}\hat{X}_{jj}\right), \quad j = 1, 2, \qquad (3.8.19)
\end{aligned}$$

where:

$$\begin{aligned}
\tilde{V}_{jj} &= \frac{1}{T}\sum_{t=1}^{T} h_{j,t}h'_{j,t}, \quad j = 1, 2, \qquad (3.8.20) \\
\hat{X}_{11} &= \frac{\mu_N}{T\sqrt{N}}\sum_{t=1}^{T}(h_{1,t}u'_{1,t} + u_{1,t}h'_{1,t}) + \frac{\mu_N^2}{TN}\sum_{t=1}^{T} u_{1,t}u'_{1,t}, \qquad (3.8.21) \\
\hat{X}_{22} &= \frac{1}{T\sqrt{N}}\sum_{t=1}^{T}(h_{2,t}u'_{2,t} + u_{2,t}h'_{2,t}) + \frac{1}{TN}\sum_{t=1}^{T} u_{2,t}u'_{2,t}. \qquad (3.8.22)
\end{aligned}$$

Therefore, we get:

$$\hat{R} \simeq \left(Id + \tilde{V}_{11}^{-1}\hat{X}_{11}\right)^{-1}\tilde{V}_{11}^{-1}\left(\tilde{V}_{12} + \hat{X}_{12}\right)\left(Id + \tilde{V}_{22}^{-1}\hat{X}_{22}\right)^{-1}\tilde{V}_{22}^{-1}\left(\tilde{V}_{21} + \hat{X}_{21}\right).$$

Let us expand $\hat{R}$ at first order in the $\hat{X}_{j,k} = O_p\left(\frac{1}{\sqrt{NT}}\right)$. By using $(Id + X)^{-1} \simeq Id - X$ for $X \simeq 0$, we have:

$$\begin{aligned}
\hat{R} &\simeq \left(Id - \tilde{V}_{11}^{-1}\hat{X}_{11}\right)\tilde{V}_{11}^{-1}\left(\tilde{V}_{12} + \hat{X}_{12}\right)\left(Id - \tilde{V}_{22}^{-1}\hat{X}_{22}\right)\tilde{V}_{22}^{-1}\left(\tilde{V}_{21} + \hat{X}_{21}\right) \\
&\simeq \tilde{V}_{11}^{-1}\tilde{V}_{12}\tilde{V}_{22}^{-1}\tilde{V}_{21} \\
&\quad -\tilde{V}_{11}^{-1}\hat{X}_{11}\tilde{V}_{11}^{-1}\tilde{V}_{12}\tilde{V}_{22}^{-1}\tilde{V}_{21} + \tilde{V}_{11}^{-1}\hat{X}_{12}\tilde{V}_{22}^{-1}\tilde{V}_{21} - \tilde{V}_{11}^{-1}\tilde{V}_{12}\tilde{V}_{22}^{-1}\hat{X}_{22}\tilde{V}_{22}^{-1}\tilde{V}_{21} + \tilde{V}_{11}^{-1}\tilde{V}_{12}\tilde{V}_{22}^{-1}\hat{X}_{21}.
\end{aligned}$$

Defining the following quantities:

$$\tilde{A} = \tilde{V}_{11}^{-1}\tilde{V}_{12}, \tag{3.8.23}$$

$$\tilde{B} = \tilde{V}_{22}^{-1}\tilde{V}_{21}, \tag{3.8.24}$$

$$\tilde{R} = \tilde{V}_{11}^{-1}\tilde{V}_{12}\tilde{V}_{22}^{-1}\tilde{V}_{21} = \tilde{A}\tilde{B}, \tag{3.8.25}$$

$$\hat{\Psi}^* = -\hat{X}_{11}\tilde{R} + \hat{X}_{12}\tilde{B} - \tilde{B}'\hat{X}_{22}\tilde{B} + \tilde{B}'\hat{X}_{21}, \tag{3.8.26}$$

$$\hat{\Psi} = \tilde{V}_{11}^{-1}\hat{\Psi}^*, \tag{3.8.27}$$

we get the asymptotic expansion of matrix $\hat{R}$:

$$\hat{R} = \tilde{R} + \hat{\Psi} + O_p\left(\frac{1}{NT}\right). \tag{3.8.28}$$

Matrix $\tilde{R}$ and its eigenvalues and eigenvectors

Let us now compute matrix $\tilde{R}$ and its eigenvalues, that are $\tilde{\rho}_1^2, ..., \tilde{\rho}_{k_1}^2$, i.e. the squared sample canonical correlations of vectors $h_{1,t}$ and $h_{2,t}$, under the null hypothesis of $k^c > 0$ common factors among the 2 groups of observables. Since the vectors $h_{1,t}$ and $h_{2,t}$ have a common component of dimension k^c, we know that $\tilde{\rho}_1 = ... = \tilde{\rho}_{k^c} = 1$ a.s.. Using the notation:

$$\begin{aligned}
\tilde{\Sigma}_{cc} &= \frac{1}{T}\sum_{t=1}^{T} f_t^c f_t^{c\prime}, \\
\tilde{\Sigma}_{cj} &= \frac{1}{T}\sum_{t=1}^{T} f_t^c f_{j,t}^{s\prime}, \quad \tilde{\Sigma}_{jc} = \tilde{\Sigma}_{cj}', \quad j = 1,2, \\
\tilde{\Sigma}_{jj} &= \frac{1}{T}\sum_{t=1}^{T} f_{j,t}^s f_{j,t}^{s\prime}, \quad j = 1,2, \\
\tilde{\Sigma}_{12} &= \frac{1}{T}\sum_{t=1}^{T} f_{1,t}^s f_{2,t}^{s\prime},
\end{aligned}$$

we can write matrices $\tilde{V}_{jj}$, with $j = 1, 2$, and $\tilde{V}_{12}$ as:

$$\tilde{V}_{jj} = \begin{pmatrix} \tilde{\Sigma}_{cc} & \tilde{\Sigma}_{c,j} \\ \tilde{\Sigma}_{j,c} & \tilde{\Sigma}_{jj} \end{pmatrix}, \quad j = 1,2, \tag{3.8.29}$$

$$\tilde{V}_{12} = \begin{pmatrix} \tilde{\Sigma}_{cc} & \tilde{\Sigma}_{c,2} \\ \tilde{\Sigma}_{1,c} & \tilde{\Sigma}_{12} \end{pmatrix} = \tilde{V}_{21}'. \tag{3.8.30}$$

By matrix algebra we get:

$$\tilde{V}_{11}^{-1} = \left[\begin{array}{cc} \Sigma_*^{-1} & -\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1} \\ -\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_*^{-1} & \tilde{\Sigma}_{11}^{-1} + \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1} \end{array} \right], \tag{3.8.31}$$

where

$$\tilde{\Sigma}_* = \tilde{\Sigma}_{cc} - \tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}. \tag{3.8.32}$$

From assumption 1, we have:

$$\tilde{\Sigma}_{c1} = O_p(1/\sqrt{T}), \tag{3.8.33}$$
$$\tilde{\Sigma}_{cc} = I_{k^c} + O_p(1/\sqrt{T}), \tag{3.8.34}$$
$$\tilde{\Sigma}_{11} = I_{k_1^s} + O_p(1/\sqrt{T}), \tag{3.8.35}$$
$$\tilde{\Sigma}_{22} = I_{k_2^s} + O_p(1/\sqrt{T}), \tag{3.8.36}$$
$$\tilde{\Sigma}_{12} = \Phi + O_p(1/\sqrt{T}), \tag{3.8.37}$$

which imply:

$$\tilde{\Sigma}_* = \tilde{\Sigma}_{cc} + O_p(1/T), \tag{3.8.38}$$

$$\tilde{\Sigma}_*^{-1} = \tilde{\Sigma}_{cc}^{-1} + O_p(1/T), \tag{3.8.39}$$

$$\begin{aligned} -\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1} &= -\tilde{\Sigma}_{cc}^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1} + O_p(1/T), \\ &= -\tilde{\Sigma}_{c1} + O_p(1/T), \end{aligned} \tag{3.8.40}$$

$$\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1} = O_p(1/T). \tag{3.8.41}$$

Substituting results (3.8.38) - (3.8.41) into equation (3.8.31) we get:

$$\tilde{V}_{11}^{-1} = \left[\begin{array}{cc} \tilde{\Sigma}_{cc}^{-1} & -\tilde{\Sigma}_{c1} \\ -\tilde{\Sigma}_{1c} & \tilde{\Sigma}_{11}^{-1} \end{array} \right] + O_p(1/T). \tag{3.8.42}$$

Equation (3.8.31) allows to compute $\tilde{A}$:

$$\begin{aligned} \tilde{A} &= \tilde{V}_{11}^{-1}\tilde{V}_{12} \\ &= \left[\begin{array}{cc} \tilde{\Sigma}_*^{-1} & -\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1} \\ -\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_*^{-1} & \tilde{\Sigma}_{11}^{-1} + \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1} \end{array} \right] \left[\begin{array}{cc} \tilde{\Sigma}_{cc} & \tilde{\Sigma}_{c2} \\ \tilde{\Sigma}_{1c} & \tilde{\Sigma}_{12} \end{array} \right] \\ &= \left[\begin{array}{cc} I_{k^c} & \tilde{A}_{cs} \\ 0 & \tilde{A}_{ss} \end{array} \right], \end{aligned} \tag{3.8.43}$$

where:

$$\tilde{A}_{cs} = \tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c2} - \tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} = O_p\left(\frac{1}{\sqrt{T}}\right), \tag{3.8.44}$$

$$\begin{aligned}\tilde{A}_{ss} &= -\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c2} + \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} + \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{1c}\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} \\ &= \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} + O_p\left(\frac{1}{T}\right) \\ &= \Phi + O_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned} \tag{3.8.45}$$

Remark 1. *Matrices $\tilde{V}_{12}$ and $\tilde{V}_{11}$ have the same first k^c columns, therefore also matrices $\tilde{V}_{11}^{-1}\tilde{V}_{12}$ and $\tilde{V}_{11}^{-1}\tilde{V}_{11} = I_{k_1}$ have the first k^c columns, which implies:*

$$\tilde{V}_{11}^{-1}\tilde{V}_{12} = \begin{bmatrix} I_{k^c} & * \\ 0 & * \end{bmatrix}.$$

Let us compute:

$$\tilde{V}_{22}^{-1} = \begin{bmatrix} \tilde{\Sigma}_{*2}^{-1} & -\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1} \\ -\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1} & \tilde{\Sigma}_{22}^{-1} + \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1} \end{bmatrix}, \tag{3.8.46}$$

where

$$\tilde{\Sigma}_{*2} = \tilde{\Sigma}_{cc} - \tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}.$$

Equation (3.8.46) allows to compute $\tilde{B}$:

$$\begin{aligned}\tilde{B} &= \tilde{V}_{22}^{-1}\tilde{V}_{21} \\ &= \begin{bmatrix} \tilde{\Sigma}_{*2}^{-1} & -\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1} \\ -\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1} & \tilde{\Sigma}_{22}^{-1} + \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\Sigma}_{cc} & \tilde{\Sigma}_{c1} \\ \tilde{\Sigma}_{2c} & \tilde{\Sigma}_{21} \end{bmatrix} \\ &= \begin{bmatrix} I_{k^c} & \tilde{B}_{cs} \\ 0 & \tilde{B}_{ss} \end{bmatrix}, \end{aligned} \tag{3.8.47}$$

where:

$$\tilde{B}_{cs} = \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c1} - \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} = O_p\left(\frac{1}{\sqrt{T}}\right), \tag{3.8.48}$$

$$\begin{aligned}\tilde{B}_{ss} &= -\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c1} + \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{2c}\tilde{\Sigma}_*^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} \\ &= \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right) \\ &= \Phi' + O_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned} \tag{3.8.49}$$

Finally, using results (3.8.43) and (3.8.47) we can compute:

$$\tilde{R} = \tilde{A}\tilde{B} \tag{3.8.50}$$

$$\begin{aligned} &= \begin{pmatrix} I_{k^c} & \tilde{A}_{cs} \\ 0 & \tilde{A}_{ss} \end{pmatrix} \begin{pmatrix} I_{k^c} & \tilde{B}_{cs} \\ 0 & \tilde{B}_{ss} \end{pmatrix} \\ &= \begin{pmatrix} I_{k^c} & \tilde{R}_{cs} \\ 0 & \tilde{R}_{ss} \end{pmatrix}, \end{aligned} \tag{3.8.51}$$

where

$$\begin{aligned}
\tilde{R}_{cs} &= \tilde{B}_{cs} + \tilde{A}_{cs}\tilde{B}_{ss} = O_p(1/\sqrt{T}), && (3.8.52)\\
\tilde{R}_{ss} &= \tilde{A}_{ss}\tilde{B}_{ss} \\
&= \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(1/T\right) \\
&= \Phi\Phi' + O_p(1/\sqrt{T}). && (3.8.53)
\end{aligned}$$

The eigenvalues of matrix $\tilde{R}$ are $\tilde{\rho}_1^2 = ... = \tilde{\rho}_{k^c}^2 = 1 > \tilde{\rho}_{k^c+1}^2 \geq ... \geq \tilde{\rho}_{k_1}^2$. The eigenvectors associated with the first k^c eigenvalues are spanned by the columns of matrix:

$$\underset{(k_1 \times k^c)}{E_c} = \left[\begin{array}{c} I_{k^c} \\ 0 \end{array} \right]. \tag{3.8.54}$$

Define:

$$\underset{(k_1 \times (k_1 - k^c))}{E_s} = \left[\begin{array}{c} 0 \\ I_{k_1 - k^c} \end{array} \right]. \tag{3.8.55}$$

We note:

$$I_{k_1} = \left[\begin{array}{ccc} E_c & \vdots & E_s \end{array} \right],$$

so that the columns of matrices E_c and E_s span the space $\mathbb{R}^{k_1}$. The estimators of the first k^c canonical correlations are such that $\hat{\rho}_\ell^2$, with $\ell = 1, ..., k^c$ are the k^c largest eigenvalues of matrix $\hat{R}$. We derive their asymptotic expansion using perturbations arguments.

Perturbation of the eigenvalues and eigenvectors of matrix $\hat{R}$

Under the null hypothesis $H(k^c)$, let $\hat{W}_1^*$ be a (k_1, k^c) matrix whose columns are eigenvectors of matrix $\hat{R}$ associated with the eigenvalues $\hat{\rho}_\ell^2$, with $\ell = 1, ..., k^c$. We have:

$$\hat{R}\hat{W}_1^* = \hat{W}_1^*\hat{\Lambda}, \tag{3.8.56}$$

where:

$$\hat{\Lambda} = diag(\hat{\rho}_\ell^2, \ell = 1, ..., k^c), \tag{3.8.57}$$

is the (k^c, k^c) diagonal matrix containing the k^c largest eigenvalues of $\hat{R}$. We know from the previous subsection that the eigenspace associated with the largest eigenvalue of $\tilde{R}$ (equal to 1) has dimension k^c and is spanned by the columns of matrix E_c. Since the columns of E_c and E_s span $\mathbb{R}^{k_1}$, we can write the following expansions:

$$\begin{aligned}
\hat{W}_1^* &= E_c\,\hat{\mathcal{U}} + E_s\alpha, && (3.8.58)\\
\hat{\Lambda} &= I_{k^c} + \hat{M}, && (3.8.59)
\end{aligned}$$

where E_c and E_s are defined in equations (3.8.54) and (3.8.55), $\hat{\mathcal{U}}$ is a (k^c, k^c) nonsingular matrix, $\hat{M} = diag(\hat{\mu}_1, ..., \hat{\mu}_{k^c})$, and α is a $(k_1 - k^c, k^c)$ matrix, with α, $\hat{\mu}_1$, ..., $\hat{\mu}_{k^c}$ converging to zero as $N_1, N_2, T \to \infty$. Substituting the expansions in equations (3.8.28) and (3.8.56) we get:

$$(\tilde{R} + \hat{\Psi})(E_c\hat{\mathcal{U}} + E_s\alpha) \quad \simeq \quad (E_c\hat{\mathcal{U}} + E_s\alpha)(I_{k^c} + \hat{M}),$$

which implies:

$$\tilde{R}E_c\hat{\mathcal{U}} + \tilde{R}E_s\alpha + \hat{\Psi}E_c\hat{\mathcal{U}} + \hat{\Psi}E_s\alpha \quad \simeq \quad E_c\hat{\mathcal{U}} + E_s\alpha + E_c\hat{\mathcal{U}}\hat{M} + E_s\alpha\hat{M}.$$

By using $\tilde{R}E_c = E_c$, and keeping only the terms at first order, we get:

$$\tilde{R}E_s\alpha + \hat{\Psi}E_c\hat{\mathcal{U}} \quad \simeq \quad E_s\alpha + E_c\hat{\mathcal{U}}\hat{M}. \tag{3.8.60}$$

Pre-multiplying equation (3.8.60) by E_c', we get:

$$\begin{aligned} &E_c'\tilde{R}E_s\alpha + E_c'\hat{\Psi}E_c\hat{\mathcal{U}} \simeq \hat{\mathcal{U}}\hat{M} \\ &\Leftrightarrow \hat{M} \simeq \hat{\mathcal{U}}^{-1}\left(\tilde{R}_{cs}\alpha + \hat{\Psi}_{cc}\hat{\mathcal{U}}\right), \end{aligned} \tag{3.8.61}$$

where we use the fact that $\hat{\mathcal{U}}$ is non-singular and

$$\hat{\Psi}_{cc} = E_c'\hat{\Psi}E_c.$$

Pre-multiplying equation (3.8.60) by E_s', we get:

$$\begin{aligned} &E_s'\tilde{R}E_s\alpha + E_s'\hat{\Psi}E_c\hat{\mathcal{U}} \simeq \alpha \\ &\Leftrightarrow \alpha \simeq \tilde{R}_{ss}\alpha + \hat{\Psi}_{sc}\hat{\mathcal{U}}, \end{aligned} \tag{3.8.62}$$

where

$$\hat{\Psi}_{sc} = E_s'\hat{\Psi}E_c.$$

This implies:

$$\alpha \quad \simeq \quad (I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\hat{\mathcal{U}}. \tag{3.8.63}$$

Substituting the first order approximation of α from equation (3.8.63) into equation (3.8.58) we get:

$$\hat{W}_1^* \quad \simeq \quad \left(E_c + E_s(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\right)\hat{\mathcal{U}}. \tag{3.8.64}$$

The normalized eigenvectors corresponding to the canonical directions are:

$$\hat{W}_1 \quad = \quad \hat{W}_1^* \cdot diag(\hat{W}_1^{*\prime}\hat{V}_{11}\hat{W}_1^*)^{-1/2}. \tag{3.8.65}$$

Substituting the first order approximation of α from equation (3.8.63) into (3.8.61), we get the first order approximation of matrix $\hat{M}$:

$$\hat{M} \quad \simeq \quad \hat{\mathcal{U}}^{-1}\left(\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\right)\hat{\mathcal{U}}. \tag{3.8.66}$$

Substituting the first order approximation of $\hat{M}$ from equation (3.8.66) into (3.8.59), matrix $\hat{\Lambda}$ can be approximated as:

$$\hat{\Lambda} \simeq I_{k^c} + \hat{\mathcal{U}}^{-1}\left(\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\right)\hat{\mathcal{U}}.$$

Note that this first order approximation holds for the terms in the main diagonal, as matrix $\hat{\Lambda}$ has been defined to be diagonal, and the out-of-diagonal terms are of higher order. Up to higher order terms we have:

$$\hat{\Lambda}^{1/2} \simeq I_{k^c} + \frac{1}{2}\hat{\mathcal{U}}^{-1}\left[\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\right]\hat{\mathcal{U}},$$

which implies:

$$\begin{aligned}
\sum_{\ell=1}^{k^c}\hat{\rho}_\ell &= tr(\hat{\Lambda}^{1/2}) \\
&= k^c + \frac{1}{2}tr\left[\hat{\mathcal{U}}^{-1}\left(\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\right)\hat{\mathcal{U}}\right] + O_p\left(\frac{1}{NT}\right), \\
&= k^c + \frac{1}{2}tr\left[\hat{\Psi}_{cc} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\hat{\Psi}_{sc}\right] + O_p\left(\frac{1}{NT}\right), \qquad (3.8.67)
\end{aligned}$$

by the commutative property of the trace.

Asymptotic distribution of $\sum_{\ell=1}^{k^c}\hat{\rho}_\ell$.

Equation (3.8.67) can be written as:

$$\begin{aligned}
\sum_{\ell=1}^{k^c}\hat{\rho}_\ell &= k^c + \frac{1}{2}tr\left\{\left[I_{k^c} \;\vdots\; \tilde{R}_{cs}(I_{(k_1-k^c)} - \tilde{R}_{ss})^{-1}\right]\hat{\Psi}E_c\right\} + O_p\left(\frac{1}{NT}\right) \\
&= k^c + \frac{1}{2}tr\left\{\left[I_{k^c} \;\vdots\; \tilde{R}_{cs}(I_{(k_1-k^c)} - \tilde{R}_{ss})^{-1}\right]\tilde{V}_{11}^{-1}\hat{\Psi}^* E_c\right\} + O_p\left(\frac{1}{NT}\right). \qquad (3.8.68)
\end{aligned}$$

Substituting equation (3.8.27), we get:

$$\sum_{\ell=1}^{k^c}\hat{\rho}_\ell = k^c + \frac{1}{2}tr\left\{\left[I_{k^c} \;\vdots\; \tilde{R}_{cs}(I_{(k_1-k^c)} - \tilde{R}_{ss})^{-1}\right]\tilde{V}_{11}^{-1}\left[\begin{array}{c}\hat{\Psi}^*_{cc} \\ \hat{\Psi}^*_{sc}\end{array}\right]\right\} + O_p\left(\frac{1}{NT}\right) \qquad (3.8.69)$$

where:

$$\hat{\Psi}^*_{cc} = \left[-\hat{X}_{11}\tilde{R} + \hat{X}_{12}\tilde{B} - \tilde{B}'\hat{X}_{22}\tilde{B} + \tilde{B}'\hat{X}_{21}\right]_{(11)}, \qquad (3.8.70)$$

$$\hat{\Psi}^*_{sc} = \left[-\hat{X}_{11}\tilde{R} + \hat{X}_{12}\tilde{B} - \tilde{B}'\hat{X}_{22}\tilde{B} + \tilde{B}'\hat{X}_{21}\right]_{(21)}, \qquad (3.8.71)$$

with $M_{(ij)}$ denoting the block in position (i, j) of matrix M. As matrices $\tilde{R}$ and $\tilde{B}$ have the same structure $[\ E_c\ \vdots\ *\ ]$, we have:

$$\hat{\Psi}^*_{cc} = \left[-\hat{X}_{11} + \hat{X}_{12} - \tilde{B}'(\hat{X}_{22} - \hat{X}_{21})\right]_{(11)}, \tag{3.8.72}$$

$$\hat{\Psi}^*_{sc} = \left[-\hat{X}_{11} + \hat{X}_{12} - \tilde{B}'(\hat{X}_{22} - \hat{X}_{21})\right]_{(21)}. \tag{3.8.73}$$

Moreover as $\tilde{B}' = \begin{bmatrix} I_{k^c} & 0 \\ \tilde{B}'_{cs} & \tilde{B}'_{ss} \end{bmatrix}$, equation (3.8.72) further simplifies to:

$$\hat{\Psi}^*_{cc} = \left[-\hat{X}_{11} + \hat{X}_{12} - \hat{X}_{22} + \hat{X}_{21}\right]_{(11)}. \tag{3.8.74}$$

Equations (3.8.73) and (3.8.74) allow to perform the asymptotic expansion of terms $\hat{\Psi}^*_{sc}$ and $\hat{\Psi}^*_{cc}$, respectively. Let us compute the asymptotic expansions of the terms $\hat{X}_{11}$, $\hat{X}_{12}$, $\hat{X}_{22}$ and $\hat{X}_{21}$. Vectors $u_{j,t}$, with $j = 1, 2$, can be partitioned into the k^c-dimensional vector $u^{(c)}_{jt}$ and the k^s_j-dimensional vector $u^{(s)}_{jt}$:

$$u_{jt} = \begin{bmatrix} u^{(c)}_{jt} \\ u^{(s)}_{jt} \end{bmatrix}, \quad j = 1, 2\,, \tag{3.8.75}$$

and from Assumption 5 we can express $\Sigma_{u,j}$, $j = 1, 2$, as: [33]

$$\Sigma_{u,j} = E[u_{jt}u'_{jt}] = E\begin{bmatrix} u^{(c)}_{jt}u^{(c)\prime}_{jt} & u^{(c)}_{jt}u^{(s)\prime}_{jt} \\ u^{(s)}_{jt}u^{(c)\prime}_{jt} & u^{(s)}_{jt}u^{(s)\prime}_{jt} \end{bmatrix} = \begin{bmatrix} \Sigma^{(cc)}_{u,j} & \Sigma^{(cs)}_{u,j} \\ \Sigma^{(sc)}_{u,j} & \Sigma^{(ss)}_{u,j} \end{bmatrix}, \quad j = 1, 2\,. \tag{3.8.76}$$

We also define:

$$\Sigma_{u,12} := E[u_{1t}u'_{2t}] := E\begin{bmatrix} u^{(c)}_{1t}u^{(c)\prime}_{2t} & u^{(c)}_{1t}u^{(s)\prime}_{2t} \\ u^{(s)}_{1t}u^{(c)\prime}_{2t} & u^{(s)}_{1t}u^{(s)\prime}_{2t} \end{bmatrix} = \begin{bmatrix} \Sigma^{(cc)}_{u,12} & \Sigma^{(cs)}_{u,12} \\ \Sigma^{(sc)}_{u,12} & \Sigma^{(ss)}_{u,12} \end{bmatrix}, \tag{3.8.77}$$

and

$$\Sigma_{u,21} = \Sigma'_{u,12}. \tag{3.8.78}$$

From equation (3.8.21) we have:

$$\begin{aligned}
\hat{X}_{11} &= \frac{\mu_N}{T\sqrt{N}}\sum_{t=1}^{T}(h_{1,t}u'_{1,t} + u_{1,t}h'_{1,t}) + \frac{\mu_N^2}{TN}\sum_{t=1}^{T}u_{1,t}u'_{1,t} \\
&= \frac{\mu_N}{T\sqrt{N}}\sum_{t=1}^{T}\left(\begin{bmatrix} f^c_t \\ f^s_{1t} \end{bmatrix}\begin{bmatrix} u^{(c)\prime}_{1t} & u^{(s)\prime}_{1t} \end{bmatrix} + \begin{bmatrix} u^{(c)}_{1t} \\ u^{(s)}_{1t} \end{bmatrix}\begin{bmatrix} f^{c\prime}_t & f^{s\prime}_{1t} \end{bmatrix}\right) \\
&\quad + \frac{\mu_N^2}{TN}\sum_{t=1}^{T}\begin{bmatrix} u^{(c)}_{1t} \\ u^{(s)}_{1t} \end{bmatrix}\begin{bmatrix} u^{(c)\prime}_{1t} & u^{(s)\prime}_{1t} \end{bmatrix} \\
&= \frac{\mu_N}{\sqrt{TN}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix} f^c_t u^{(c)\prime}_{1t} + u^{(c)}_{1t}f^{c\prime}_t & f^c_t u^{(s)\prime}_{1t} + u^{(c)}_{1t}f^{s\prime}_{1t} \\ f^s_{1t}u^{(c)\prime}_{1t} + u^{(s)}_{1t}f^{c\prime}_t & f^s_{1t}u^{(s)\prime}_{1t} + u^{(s)}_{1t}f^{s\prime}_{1t} \end{bmatrix}\right) + \frac{\mu_N^2}{TN}\sum_{t=1}^{T}\begin{bmatrix} u^{(c)}_{1t}u^{(c)\prime}_{1t} & u^{(c)}_{1t}u^{(s)\prime}_{1t} \\ u^{(s)}_{1t}u^{(c)\prime}_{1t} & u^{(s)}_{1t}u^{(s)\prime}_{1t} \end{bmatrix},
\end{aligned}$$

[33] Matrix $\Sigma_{u,j}$ is the asymptotic variance of $u_{j,t}$ as $N_j \to \infty$. We omit the limit for expository purpose.

and from assumption 5 *b)* we have:

$$\begin{aligned}
\hat{X}_{11} &= \frac{\mu_N}{\sqrt{TN}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} f_t^c u_{1t}^{(c)\prime}+u_{1t}^{(c)}f_t^{c\prime} & f_t^c u_{1t}^{(s)\prime}+u_{1t}^{(c)}f_{1t}^{s\prime} \\ f_{1t}^s u_{1t}^{(c)\prime}+u_{1t}^{(s)}f_t^{c\prime} & f_{1t}^s u_{1t}^{(s)\prime}+u_{1t}^{(s)}f_{1t}^{s\prime}\end{array}\right]\right)+\frac{\mu_N^2}{N}E\left[\begin{array}{cc} u_{1t}^{(c)}u_{1t}^{(c)\prime} & u_{1t}^{(c)}u_{1t}^{(s)\prime} \\ u_{1t}^{(s)}u_{1t}^{(c)\prime} & u_{1t}^{(s)}u_{1t}^{(s)\prime}\end{array}\right] \\
&\quad +\frac{\mu_N^2}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} u_{1t}^{(c)}u_{1t}^{(c)\prime}-E[u_{1t}^{(c)}u_{1t}^{(c)\prime}] & u_{1t}^{(c)}u_{1t}^{(s)\prime}-E[u_{1t}^{(c)}u_{1t}^{(s)\prime}] \\ u_{1t}^{(s)}u_{1t}^{(c)\prime}-E[u_{1t}^{(s)}u_{1t}^{(c)\prime}] & u_{1t}^{(s)}u_{1t}^{(s)\prime}-E[u_{1t}^{(s)}u_{1t}^{(s)\prime}]\end{array}\right]\right) \\
&= \frac{\mu_N}{\sqrt{TN}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} f_t^c u_{1t}^{(c)\prime}+u_{1t}^{(c)}f_t^{c\prime} & f_t^c u_{1t}^{(s)\prime}+u_{1t}^{(c)}f_{1t}^{s\prime} \\ f_{1t}^s u_{1t}^{(c)\prime}+u_{1t}^{(s)}f_t^{c\prime} & f_{1t}^s u_{1t}^{(s)\prime}+u_{1t}^{(s)}f_{1t}^{s\prime}\end{array}\right]\right) \\
&\quad +\frac{\mu_N^2}{N}\left[\begin{array}{cc} \Sigma_{u,1}^{(cc)} & \Sigma_{u,1}^{(cs)} \\ \Sigma_{u,1}^{(sc)} & \Sigma_{u,1}^{(ss)}\end{array}\right]+\frac{\mu_N^2}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} u_{1t}^{(c)}u_{1t}^{(c)\prime}-\Sigma_{u,1}^{(cc)} & u_{1t}^{(c)}u_{1t}^{(s)\prime}-\Sigma_{u,1}^{(cs)} \\ u_{1t}^{(s)}u_{1t}^{(c)\prime}-\Sigma_{u,1}^{(sc)} & u_{1t}^{(s)}u_{1t}^{(s)\prime}-\Sigma_{u,1}^{(ss)}\end{array}\right]\right). \qquad (3.8.79)
\end{aligned}$$

Analogously, from (3.8.22) we have:

$$\begin{aligned}
\hat{X}_{22} &= \frac{1}{\sqrt{TN}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} f_t^c u_{2t}^{(c)\prime}+u_{2t}^{(c)}f_t^{c\prime} & f_t^c u_{2t}^{(s)\prime}+u_{2t}^{(c)}f_{2t}^{s\prime} \\ f_{2t}^s u_{2t}^{(c)\prime}+u_{2t}^{(s)}f_t^{c\prime} & f_{2t}^s u_{2t}^{(s)\prime}+u_{2t}^{(s)}f_{2t}^{s\prime}\end{array}\right]\right) \\
&\quad +\frac{1}{N}\left[\begin{array}{cc} \Sigma_{u,2}^{(cc)} & \Sigma_{u,2}^{(cs)} \\ \Sigma_{u,2}^{(sc)} & \Sigma_{u,2}^{(22)}\end{array}\right]+\frac{1}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} u_{2t}^{(c)}u_{2t}^{(c)\prime}-\Sigma_{u,2}^{(cc)} & u_{2t}^{(c)}u_{2t}^{(s)\prime}-\Sigma_{u,2}^{(cs)} \\ u_{2t}^{(s)}u_{2t}^{(c)\prime}-\Sigma_{u,2}^{(sc)} & u_{2t}^{(s)}u_{2t}^{(s)\prime}-\Sigma_{u,2}^{(ss)}\end{array}\right]\right). \qquad (3.8.80)
\end{aligned}$$

From equation (3.8.17), the term $\hat{X}_{12}$ results to be:

$$\begin{aligned}
\hat{X}_{12} &= \frac{1}{T\sqrt{N}}\sum_{t=1}^{T}(h_{1,t}u_{2,t}'+\mu_N u_{1,t}h_{2,t}')+\frac{\mu_N}{TN}\sum_{t=1}^{T}u_{1,t}u_{2,t}' \\
&= \frac{1}{T\sqrt{N}}\sum_{t=1}^{T}\left(\left[\begin{array}{c} f_t^c \\ f_{1t}^s\end{array}\right]\left[\begin{array}{cc} u_{2t}^{(c)\prime} & u_{2t}^{(s)\prime}\end{array}\right]+\mu_N\left[\begin{array}{c} u_{1t}^{(c)} \\ u_{1t}^{(s)}\end{array}\right]\left[\begin{array}{cc} f_t^{c\prime} & f_{2t}^{s\prime}\end{array}\right]\right) \\
&\quad +\frac{\mu_N}{TN}\sum_{t=1}^{T}\left[\begin{array}{c} u_{1t}^{(c)} \\ u_{1t}^{(s)}\end{array}\right]\left[\begin{array}{cc} u_{2t}^{(c)\prime} & u_{2t}^{(s)\prime}\end{array}\right] \\
&= \frac{1}{\sqrt{TN}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} f_t^c u_{2t}^{(c)\prime}+\mu_N u_{1t}^{(c)}f_t^{c\prime} & f_t^c u_{2t}^{(s)\prime}+\mu_N u_{1t}^{(c)}f_{2t}^{s\prime} \\ f_{1t}^s u_{2t}^{(c)\prime}+\mu_N u_{1t}^{(s)}f_t^{c\prime} & f_{1t}^s u_{2t}^{(s)\prime}+\mu_N u_{1t}^{(s)}f_{2t}^{s\prime}\end{array}\right]\right)+\frac{\mu_N}{N}\left[\begin{array}{cc} \Sigma_{u,12}^{(cc)} & \Sigma_{u,12}^{(cs)} \\ \Sigma_{u,12}^{(sc)} & \Sigma_{u,12}^{(ss)}\end{array}\right] \\
&\quad +\frac{\mu_N}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} u_{1t}^{(c)}u_{2t}^{(c)\prime}-\Sigma_{u,12}^{(cc)} & u_{1t}^{(c)}u_{2t}^{(s)\prime}-\Sigma_{u,12}^{(cs)} \\ u_{1t}^{(s)}u_{2t}^{(c)\prime}-\Sigma_{u,12}^{(sc)} & u_{1t}^{(s)}u_{2t}^{(s)\prime}-\Sigma_{u,12}^{(ss)}\end{array}\right]\right). \qquad (3.8.81)
\end{aligned}$$

Finally we have:

$$\begin{aligned}
\hat{X}_{21} &= \hat{X}_{12}' \\
&= \frac{1}{\sqrt{TN}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} u_{2t}^{(c)}f_t^{c\prime}+\mu_N f_t^c u_{1t}^{(c)\prime} & u_{2t}^{(c)}f_{1t}^{s\prime}+\mu_N f_t^c u_{1t}^{(s)\prime} \\ u_{2t}^{(s)}f_t^{c\prime}+\mu_N f_{2t}^s u_{1t}^{(c)\prime} & u_{2t}^{(s)}f_{1t}^{s\prime}+\mu_N f_{2t}^s u_{1t}^{(s)\prime}\end{array}\right]\right)+\frac{\mu_N}{N}\left[\begin{array}{cc} \Sigma_{u,21}^{(cc)} & \Sigma_{u,21}^{(cs)} \\ \Sigma_{u,21}^{(sc)} & \Sigma_{u,21}^{(22)}\end{array}\right] \\
&\quad +\frac{\mu_N}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\begin{array}{cc} u_{2t}^{(c)}u_{1t}^{(c)\prime}-\Sigma_{u,21}^{(cc)} & u_{2t}^{(c)}u_{1t}^{(s)\prime}-\Sigma_{u,21}^{(cs)} \\ u_{2t}^{(s)}u_{1t}^{(c)\prime}-\Sigma_{u,21}^{(sc)} & u_{2t}^{(s)}u_{1t}^{(s)\prime}-\Sigma_{u,21}^{(ss)}\end{array}\right]\right). \qquad (3.8.82)
\end{aligned}$$

We can now compute directly term $\hat{\Psi}^*_{cc}$. From equation (3.8.74), we get:

$$\begin{aligned}
& \hat{\Psi}^*_{cc} \qquad (3.8.83)\\
= & \left[-\hat{X}_{11} + \hat{X}_{12} - \hat{X}_{22} + \hat{X}_{21}\right]_{(11)},\\
= & \frac{1}{\sqrt{TN}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[-\mu_N f^c_t u^{(c)\prime}_{1t} - \mu_N u^{(c)}_{1t} f^{c\prime}_t + f^c_t u^{(c)\prime}_{2t} + \mu_N u^{(c)}_{1t} f^{c\prime}_t - f^c_t u^{(c)\prime}_{2t} - u^{(c)}_{2t} f^{c\prime}_t + u^{(c)}_{2t} f^{c\prime}_t + \mu_N f^c_t u^{(c)\prime}_{1t}\right]\right)\\
& + \frac{1}{N}[-\mu_N^2 \Sigma^{(cc)}_{u,1} - \Sigma^{(cc)}_{u,2} + \mu_N \Sigma^{(cc)}_{u,12} + \mu_N \Sigma^{(cc)}_{u,21}]\\
& + \frac{1}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[-\mu_N^2[u^{(c)}_{1t}u^{(c)\prime}_{1t} - \Sigma^{(cc)}_{u,1}] + \mu_N[u^{(c)}_{1t}u^{(c)\prime}_{2t} - \Sigma^{(cc)}_{u,12}] - [u^{(c)}_{2t}u^{(c)\prime}_{2t} - \Sigma^{(cc)}_{u,2}] + \mu_N[u^{(c)}_{2t}u^{(c)\prime}_{1t} - \Sigma^{(cc)}_{u,12}]\right]\right)\\
= & -\frac{1}{N}E[(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})']\\
& -\frac{1}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})' - E[(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})']\right]\right). \qquad (3.8.84)
\end{aligned}$$

Using the limit $\mu_N \to \mu$, we get:

$$\begin{aligned}
\hat{\Psi}^*_{cc} = & -\frac{1}{N}E[(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})']\\
& -\frac{1}{N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})(\mu_N u^{(c)}_{1t} - u^{(c)}_{2t})' - E[(\mu u^{(c)}_{1t} - u^{(c)}_{2t})(\mu u^{(c)}_{1t} - u^{(c)}_{2t})']\right]\right)\\
& + o_p\left(\frac{1}{N\sqrt{T}}\right). \qquad (3.8.85)
\end{aligned}$$

Before computing $\hat{\Psi}^*_{sc}$ and substituting it into equation (3.8.69), we note that some of the terms of this equation can be further simplified. Let us consider the asymptotic expansion of the following term of equation (3.8.69):

$$\left[I_{k^c} \;\vdots\; \tilde{R}_{cs}(I_{(k_1-k^c)} - \tilde{R}_{ss})^{-1}\right]\tilde{V}_{11}^{-1}.$$

Using equation (3.8.42), we get:

$$\begin{aligned}
& \left[I_{k^c} \;\vdots\; \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\right]\tilde{V}_{11}^{-1}\\
& = \left[I_{k^c} \;\vdots\; \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\right]\begin{bmatrix} \tilde{\Sigma}_{cc}^{-1} & -\tilde{\Sigma}_{c1} \\ -\tilde{\Sigma}_{1c} & \tilde{\Sigma}_{11}^{-1}\end{bmatrix} + O_p\left(\frac{1}{T}\right)\\
& = \left[\tilde{\Sigma}_{cc}^{-1} - \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\tilde{\Sigma}_{1c} \;\vdots\; -\tilde{\Sigma}_{c1} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\tilde{\Sigma}_{11}^{-1}\right] + O_p\left(\frac{1}{T}\right)\\
& = \left[\tilde{\Sigma}_{cc}^{-1} \;\vdots\; -\tilde{\Sigma}_{c1} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\right] + O_p\left(\frac{1}{T}\right), \qquad (3.8.86)
\end{aligned}$$

where the last equality follows form the fact that $\tilde{R}_{cs} = O_p(1/\sqrt{T})$, $\tilde{\Sigma}_{1c} = O_p(1/\sqrt{T})$ and $\tilde{\Sigma}_{11} = I_{k_1} + O_p(1/\sqrt{T})$. Note that equation (3.8.86) can be further simplified, considering the asymptotic expansion of term $\tilde{R}_{cs}$. Let us consider the different terms in the equations of $\tilde{R}_{cs}$ and $\tilde{R}_{ss}$:

$$\tilde{R}_{cs} = \tilde{B}_{cs} + \tilde{A}_{cs}\tilde{B}_{ss}, \tag{3.8.87}$$

$$\tilde{R}_{ss} = \tilde{A}_{ss}\tilde{B}_{ss}, \tag{3.8.88}$$

where:

$$\tilde{A}_{cs} = \tilde{\Sigma}_{*}^{-1}\tilde{\Sigma}_{c2} - \tilde{\Sigma}_{*}^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}, \tag{3.8.89}$$

$$\tilde{A}_{ss} = \Sigma_{11}^{-1}\tilde{\Sigma}_{12} + O_p\left(\frac{1}{T}\right), \tag{3.8.90}$$

$$\tilde{B}_{cs} = \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c1} - \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}, \tag{3.8.91}$$

$$\tilde{B}_{ss} = \tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right). \tag{3.8.92}$$

Substituting equations (3.8.89) - (3.8.92) into equations (3.8.87) and (3.8.88) we get:

$$\begin{aligned} \tilde{R}_{cs} &= \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c1} - \tilde{\Sigma}_{*2}^{-1}\tilde{\Sigma}_{c2}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + \left[\tilde{\Sigma}_{*}^{-1}\tilde{\Sigma}_{c2} - \tilde{\Sigma}_{*}^{-1}\tilde{\Sigma}_{c1}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\right]\left[\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right)\right] \\ &= \tilde{\Sigma}_{c1}\left[I_{k_1^s} - \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}\right] + O_p\left(\frac{1}{T}\right), \end{aligned} \tag{3.8.93}$$

and

$$\begin{aligned} \tilde{R}_{ss} &= \left[\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12} + O_p\left(\frac{1}{T}\right)\right]\left[\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right)\right] \\ &= \tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + O_p\left(\frac{1}{T}\right). \end{aligned} \tag{3.8.94}$$

Therefore we have:

$$\tilde{R}_{cs} = \tilde{\Sigma}_{c1}(I_{k_1-k^c} - \tilde{R}_{ss}) + O_p\left(\frac{1}{T}\right), \tag{3.8.95}$$

which implies :

$$-\tilde{\Sigma}_{c1} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1} = O_p\left(\frac{1}{T}\right). \tag{3.8.96}$$

Equation (3.8.96) and $\hat{\Psi}_{sc}^{*} = O_p\left(\frac{1}{\sqrt{NT}}\right)$, together with the assumption $\sqrt{N}/T = o(1)$, imply:

$$\left[-\tilde{\Sigma}_{c1} + \tilde{R}_{cs}(I_{k_1-k^c} - \tilde{R}_{ss})^{-1}\right]\hat{\Psi}_{sc}^{*} = o_p\left(\frac{1}{N\sqrt{T}}\right). \tag{3.8.97}$$

Therefore, substituting results (3.8.84), (3.8.86), and (3.8.97) into equation (3.8.69), and rearranging terms, we get:

$$\begin{aligned} \sum_{\ell=1}^{k^c}\hat{\rho}_\ell &= k^c - \frac{1}{N}\frac{1}{2}tr\left\{\tilde{\Sigma}_{cc}^{-1}E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})']\right\} \\ &\quad - \frac{1}{N\sqrt{T}}\frac{1}{2}tr\left\{\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu u_{1t}^{(c)} - u_{2t}^{(c)})(\mu u_{1t}^{(c)} - u_{2t}^{(c)})' - E[(\mu u_{1t}^{(c)} - u_{2t}^{(c)})(\mu u_{1t}^{(c)} - u_{2t}^{(c)})']\right]\right)\right\} \\ &\quad + o_p\left(\frac{1}{N\sqrt{T}}\right). \end{aligned} \tag{3.8.98}$$

From the definition of matrix $\Sigma_{U,N}$ we have;

$$E[(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})(\mu_N u_{1t}^{(c)} - u_{2t}^{(c)})'] = \Sigma_{U,N}. \tag{3.8.99}$$

Moreover, let us define:

$$U_t \quad := \quad \mu u_{1t}^{(c)} - u_{2t}^{(c)}. \tag{3.8.100}$$

Definition (3.8.100) together with the commutativity and linearity properties of the trace operator allow to write the fourth term in the r.h.s. of equation (3.8.98) as:

$$\begin{aligned} & tr\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\mu u_{1t}^{(c)} - u_{2t}^{(c)})(\mu u_{1t}^{(c)} - u_{2t}^{(c)})' - E[(\mu u_{1t}^{(c)} - u_{2t}^{(c)})(\mu u_{1t}^{(c)} - u_{2t}^{(c)})']\right]\right\} \\ & = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{U_t'U_t - E(U_t'U_t)\right\}. \end{aligned} \tag{3.8.101}$$

Equations (3.8.99) and (3.8.101) allow to write equation (3.8.98) as:

$$\sum_{\ell=1}^{k^c}\hat{\rho}_\ell \quad = \quad k^c - \frac{1}{2N}tr\left\{\tilde{\Sigma}_{cc}^{-1}\Sigma_{U,N}\right\} - \frac{1}{2N\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[U_t'U_t - E(U_t'U_t)\right]\right) + o_p\left(\frac{1}{N\sqrt{T}}\right). \tag{3.8.102}$$

By a CLT for weakly dependent data we have:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[U_t'U_t - E(U_t'U_t)\right] \xrightarrow{d} N\left(0, \Omega_U\right), \tag{3.8.103}$$

where:

$$\Omega_U \quad = \quad \lim_{T\to\infty} V\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_t'U_t\right) = \sum_{h=-\infty}^{\infty}Cov(U_t'U_t, U_{t-h}'U_{t-h}). \tag{3.8.104}$$

From equation (3.8.103) we get that the asymptotic distribution of $\sum_{\ell=1}^{k^c}\hat{\rho}_\ell$, under the hypothesis of k^c common factors in each group is:

$$N\sqrt{T}\left[\sum_{\ell=1}^{k^c}\hat{\rho}_\ell - k^c + \frac{1}{2N}tr\left\{\tilde{\Sigma}_{cc}^{-1}\Sigma_U\right\}\right] \xrightarrow{d} N\left(0, \frac{1}{4}\Omega_U\right). \tag{3.8.105}$$

To conclude the proof, let us derive the expression of matrix Ω_U in equation (3.4.11). For this purpose, note that vector $(U_t', U_{t-h}')'$ is asymptotically Gaussian for any h:

$$\begin{pmatrix} U_t \\ U_{t-h} \end{pmatrix} \xrightarrow{d} N\begin{pmatrix} \Sigma_U(0) & \Sigma_U(h) \\ \Sigma_U(h)' & \Sigma_U(0) \end{pmatrix}. \tag{3.8.106}$$

We use the following lemma.

Lemma 1. *Let the* $(n,1)$ *random vector* x *and the* $(m,1)$ *random vector* y *be such that*

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N \begin{pmatrix} \Omega_{xx} & \Omega_{xy} \\ \Omega_{xy}' & \Omega_{yy} \end{pmatrix}, \tag{3.8.107}$$

and let A *and* B *be symmetric* (n,n) *and* (m,m) *matrices, respectively. Then:*

$$\begin{array}{ll} i) & V[x'Ax] = 2tr\left\{(A\Omega_{xx})^2\right\}, \\ ii) & Cov(x'Ax, y'By) = 2tr\left\{A\Omega_{xy}B\Omega_{xy}'\right\}. \end{array}$$

Proof of Lemma 1: For point i), see Theorem 12 p. 284 in Magnus and Neudecker (2007). Point ii) is a consequence of point i) applied to vectors x, y and $(x', y')'$, see also Theorem 10.21 in Schott (2005).

From Lemma 1 we get (asymptotically):

$$Cov(U_t'U_t, U_{t-h}'U_{t-h}) = 2tr\left\{\Sigma_U(h)\Sigma_U(h)'\right\}, \tag{3.8.108}$$

and the conclusion follows.

Q.E.D.

3.8.4 Proof of Theorems 2 and 3

Asymptotic distribution of $\hat{f}_t^c$ and $\hat{f}_t^{c\,*}$

Equation (3.8.64) and $\hat{\Psi}_{sc} = O_p\left(\frac{1}{\sqrt{NT}}\right)$ imply:

$$\hat{W}_1^* = E_c\,\hat{\mathcal{U}} + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{3.8.109}$$

Recall from equation (3.8.65) that the normalized eigenvectors corresponding to the canonical directions are:

$$\hat{W}_1 = \hat{W}_1^*\hat{D},$$

where $\hat{D} = diag(\hat{W}_1^{*\,\prime}\hat{V}_{11}\hat{W}_1^*)^{-1/2}$. Then, we get:

$$\begin{aligned} \hat{f}_t^c &= \hat{W}_1'\hat{h}_{1,t} \\ &= \hat{D}\hat{\mathcal{U}}'E_c'\left(h_{1,t} + \frac{1}{\sqrt{N_1}}u_{1,t}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \\ &= \hat{D}\hat{\mathcal{U}}'\left(f_t^c + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \end{aligned} \tag{3.8.110}$$

Therefore the estimated factor can be written as:

$$\hat{f}_t^c = \hat{\mathcal{H}}_c^{-1}\left(f_t^c + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{3.8.111}$$

where $\hat{\mathcal{H}}_c^{-1} = \hat{D}\hat{\mathcal{U}}'$. Equation (3.8.111) implies:

$$\sqrt{N_1}\left(\hat{\mathcal{H}}_c\hat{f}_t^c - f_t^c\right) = u_{1,t}^{(c)} + o_p(1) \xrightarrow{d} N\left(0, \Sigma_{u,1}^{(cc)}\right).$$

The derivation of the asymptotic distribution of $\sqrt{N_2}\left(\hat{\mathcal{H}}_c^*\hat{f}_t^{c\,*} - f_t^c\right)$ obtained from the canonical direction $\hat{W}_2$ is analogous, and therefore is omitted.

Asymptotic distribution of $\hat{\lambda}^c_{j,i}$

Let us derive the asymptotic expansion of the loading estimator $\hat{\lambda}^c_{j,i} = (\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\prime}y_{j,i}$, where $y_{j,i}$ is the i-th column of matrix Y_j. From equation (3.8.111) we can express $\hat{F}^c = [\hat{f}^c_1, ..., \hat{f}^c_T]'$ as:

$$\begin{aligned} \hat{F}^c &= \left(F^c + \frac{1}{\sqrt{N_1}}U_1^{(c)}\right)\left(\hat{\mathcal{H}}_c^{-1}\right)' + O_p\left(\frac{1}{\sqrt{NT}}\right) \\ &= F^c\left(\hat{\mathcal{H}}_c^{-1}\right)' + \frac{1}{\sqrt{N_1}}U_1^{(c)}\left(\hat{\mathcal{H}}_c^{-1}\right)' + O_p\left(\frac{1}{\sqrt{NT}}\right), \end{aligned} \tag{3.8.112}$$

where $U_1^{(c)} = [u^{(c)}_{1,1}, ..., u^{(c)}_{1,T}]'$. Equation (3.8.112) implies:

$$\hat{F}^c\hat{\mathcal{H}}_c' - F^c = \frac{1}{\sqrt{N_1}}U_1^{(c)} + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{3.8.113}$$

Then, denoting with $\xi_{j,i}$ the i-th column of matrix Ξ_j, we get:

$$\begin{aligned} \hat{\lambda}^c_{j,i} &= (\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\,\prime}y_{j,i} \\ &= (\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\,\prime}\left(F^c\lambda^c_{j,i} + F^s_j\lambda^s_{j,i} + \varepsilon_{j,i}\right) \\ &= (\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\,\prime}\left[\left(F^c - \hat{F}^c\hat{\mathcal{H}}_c' + \hat{F}^c\hat{\mathcal{H}}_c'\right)\lambda^c_{j,i} + F^s_j\lambda^s_{j,i} + \varepsilon_{j,i}\right] \\ &= \hat{\mathcal{H}}_c'\lambda^c_{j,i} + (\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\,\prime}\varepsilon_{j,i} \\ &\quad + (\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\,\prime}\left(F^c - \hat{F}^c\hat{\mathcal{H}}_c'\right)\lambda^c_{j,i} + (\hat{F}^{c\prime}\hat{F}^c)^{-1}\hat{F}^{c\,\prime}F^s_j\lambda^s_{j,i}, \qquad j = 1, 2. \end{aligned} \tag{3.8.114}$$

We first note that

$$\begin{aligned} \frac{\hat{F}^{c\prime}\hat{F}^c}{T} &= \frac{1}{T}\hat{\mathcal{H}}_c^{-1}\left(F^c + \frac{1}{\sqrt{N_1}}U_1^{(c)}\right)'\left(F^c + \frac{1}{\sqrt{N_1}}U_1^{(c)}\right)(\hat{\mathcal{H}}_c^{-1})' + O_p\left(\frac{1}{\sqrt{NT}}\right) \\ &= \hat{\mathcal{H}}_c^{-1}\frac{F^{c\prime}F^c}{T}(\hat{\mathcal{H}}_c^{-1})' + \frac{1}{\sqrt{N_1}}\hat{\mathcal{H}}_c^{-1}\frac{U_1^{(c)\prime}F^c_t}{T}(\hat{\mathcal{H}}_c^{-1})' \\ &\quad + \frac{1}{\sqrt{N_1}}\hat{\mathcal{H}}_c^{-1}\frac{F^{c\prime}U_1^{(c)}}{T}(\hat{\mathcal{H}}_c^{-1})' + \frac{1}{N_1}\hat{\mathcal{H}}_c^{-1}\frac{U_1^{(c)\prime}U_1^{(c)}}{T}(\hat{\mathcal{H}}_c^{-1})' + O_p\left(\frac{1}{\sqrt{NT}}\right) \\ &= \hat{\mathcal{H}}_c^{-1}\frac{F^{c\prime}F^c}{T}(\hat{\mathcal{H}}_c^{-1})' + O_p\left(\frac{1}{\sqrt{NT}}\right), \end{aligned}$$

where we use $\frac{1}{\sqrt{T}}F^{c\prime}U_1^{(c)} = O_p(1)$, $\frac{1}{T}U_1^{(c)\prime}U_1^{(c)} = O_p(1)$ and $T/N_1 = o(1)$. We also have:

$$\left(\frac{\hat{F}^{c\prime}\hat{F}^c}{T}\right)^{-1} = \hat{\mathcal{H}}_c'\left(\frac{F^{c\prime}F^c}{T}\right)^{-1}\hat{\mathcal{H}}_c + O_p\left(\frac{1}{\sqrt{TN}}\right). \tag{3.8.115}$$

Equations (3.8.112) and (3.8.113) allow to compute:

$$\begin{aligned} \frac{1}{T}\hat{F}^{c\,\prime}\left(F^c - \hat{F}^c\hat{\mathcal{H}}_c'\right) &\simeq -\frac{1}{T\sqrt{N_1}}\hat{\mathcal{H}}_c^{-1}F^{c\,\prime}U_1^{(c)} - \frac{1}{N_1T}\hat{\mathcal{H}}_c^{-1}U_1^{(c)\,\prime}U_1^{(c)} \\ &= O_p\left(\frac{1}{\sqrt{NT}}\right), \end{aligned} \tag{3.8.116}$$

and:

$$\frac{1}{T}\hat{F}^{c\,\prime}\varepsilon_{j,i} = \hat{\mathcal{H}}_c^{-1}\left(\frac{1}{T}F^{c\prime}\varepsilon_{j,i} + \frac{1}{T\sqrt{N_1}}U_1^{(c)\prime}\varepsilon_{j,i}\right)$$
$$= \hat{\mathcal{H}}_c^{-1}\frac{1}{T}F^{c\prime}\varepsilon_{j,i} + O_p\left(\frac{1}{\sqrt{NT}}\right). \quad (3.8.117)$$

We also have:

$$\frac{1}{T}\hat{F}^{c\,\prime}F_j^s = \hat{\mathcal{H}}_c^{-1}\left(\frac{1}{T}F^{c\prime}F_j^s + \frac{1}{T\sqrt{N_1}}U_1^{(c)\prime}F_j^s\right)$$
$$= \hat{\mathcal{H}}_c^{-1}\frac{1}{T}F^{c\prime}F_j^s + O_p\left(\frac{1}{\sqrt{NT}}\right). \quad (3.8.118)$$

Substituting approximations (3.8.115) - (3.8.118) into equation (3.8.114) we get:

$$\hat{\lambda}_{j,i}^c \simeq \hat{\mathcal{H}}_c'\lambda_{j,i}^c + \hat{\mathcal{H}}_c'\left(\frac{F^{c\prime}F^c}{T}\right)^{-1}\frac{1}{T}F^{c\,\prime}\varepsilon_{j,i}$$
$$+\hat{\mathcal{H}}_c'\left(\frac{F^{c\prime}F^c}{T}\right)^{-1}\frac{1}{T}F^{c\,\prime}F_j^s\lambda_{j,i}^s + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

The last equation implies:

$$\sqrt{T}\left[\left(\hat{\mathcal{H}}_c'\right)^{-1}\hat{\lambda}_{j,i}^c - \lambda_{j,i}^c\right] = \varphi_{j,i} + K_j\lambda_{j,i}^s + o_p(1), \quad (3.8.119)$$

where:

$$\varphi_{j,i} = \left(\frac{F^{c\prime}F^c}{T}\right)^{-1}\frac{1}{\sqrt{T}}F^{c\,\prime}\varepsilon_{j,i}, \quad (3.8.120)$$
$$K_j = \left(\frac{F^{c\prime}F^c}{T}\right)^{-1}\frac{1}{\sqrt{T}}F^{c\,\prime}F_j^s. \quad (3.8.121)$$

Since $(F^{c\,\prime}F^c/T)^{-1} = I_{k^c} + o_p(1)$, the r.h.s. of equation (3.8.119) can be rewritten to get:

$$\sqrt{T}\left[\left(\hat{\mathcal{H}}_c'\right)^{-1}\hat{\lambda}_{j,i}^c - \lambda_{j,i}^c\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t^c(\varepsilon_{j,it} + f_{j,t}^{s\,\prime}\lambda_{j,i}^s) + o_p(1) \equiv w_{j,i}^c + o_p(1). \quad (3.8.122)$$

Then, since the errors and the factors are independent (Assumption 5 *a)*) , a CLT for weakly dependent data yields equation (3.4.19).

Asymptotic distribution of $\hat{f}_{j,t}^s$ and $\hat{\lambda}_{j,i}^s$

Let us now derive the asymptotic expansion of term $\hat{f}_{j,t}^s$. We start by computing the asymptotic expansion of the regression residuals $y_{j,it} - \hat{f}_t^{c\,\prime}\hat{\lambda}_{j,i}^c$:

$$y_{j,it} - \hat{f}_t^{c\,\prime}\hat{\lambda}_{j,i}^c = f_{j,t}^{s\,\prime}\lambda_{j,i}^s + \varepsilon_{j,it} - \left(\hat{f}_t^{c\,\prime}\hat{\lambda}_{j,i}^c - f_t^{c\,\prime}\lambda_{j,i}^c\right)$$
$$= f_{j,t}^{s\,\prime}\lambda_{j,i}^s + \varepsilon_{j,it} - \left[\left(f_t^c + \frac{1}{\sqrt{N_1}}u_{1,t}^{(c)}\right)'\left(\lambda_{j,i}^c + \frac{1}{\sqrt{T}}\varphi_{j,i} + \frac{1}{\sqrt{T}}K_j\lambda_{j,i}^s\right) - f_t^{c\,\prime}\lambda_{j,i}^c\right]$$
$$\simeq g_{j,t}'\lambda_{j,i}^s + e_{j,it}, \quad (3.8.123)$$

where:

$$g_{j,t} \quad := \quad f_{j,t}^{s} - \frac{1}{\sqrt{T}} K_j' f_t^c = f_{j,t}^s - (F_j^{s\prime} F^c)(F^{c\prime} F^c)^{-1} f_t^c, \tag{3.8.124}$$

$$e_{j,it} \quad := \quad \varepsilon_{j,it} - \frac{1}{\sqrt{T}} f_t^{c\prime} \varphi_{j,i}. \tag{3.8.125}$$

Then, the residuals $y_{j,it} - \hat{f}_t^{c\prime} \hat{\lambda}_{j,i}^c$ satisfy an approximate factor structure with factors $g_{j,t}$ and errors $e_{j,it}$. From asymptotic theory of the PC estimators in large panels, we know that:

$$\sqrt{N} \left[\hat{\mathcal{H}}_{s,j} \hat{f}_{j,t}^s - g_{j,t} \right] \quad = \quad v_{j,t}^{*s} + o_p(1), \qquad j = 1, 2, \tag{3.8.126}$$

where $\hat{\mathcal{H}}_{s,j}$, $j = 1, 2$, is a non-singular matrix and:

$$\begin{aligned} v_{j,t}^{*s} &= \left(\frac{\Lambda_j^{s\prime} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \Lambda_j^{s\prime} e_{j,t} \\ &= \left(\frac{\Lambda_j^{s\prime} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,it} - \left(\frac{\Lambda_j^{s\prime} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_j} \lambda_{j,i}^s f_t^{c\prime} \left(\frac{1}{\sqrt{T}} \sum_{r=1}^{T} f_r^c \varepsilon_{j,ir} \right) \\ &= \left(\frac{\Lambda_j^{s\prime} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,it} + o_p(1). \end{aligned}$$

Therefore we have

$$\sqrt{N} \left[\hat{\mathcal{H}}_{s,j} \hat{f}_{j,t}^s - (f_{j,t}^s - (F_j^{s\prime} F^c)(F^{c\prime} F^c)^{-1} f_t^c) \right] \quad = \quad v_{j,t}^s + o_p(1), \qquad j = 1, 2, \tag{3.8.127}$$

where $v_{j,t}^s = \left(\frac{\Lambda_j^{s\prime} \Lambda_j^s}{N_j} \right)^{-1} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \lambda_{j,i}^s \varepsilon_{j,it}$, which proves equation (3.4.16).

From asymptotic theory of the PC estimators in large panels, we also know that the following result must hold for the loadings estimator of factor model (3.8.123):

$$\sqrt{T} \left[\left(\hat{\mathcal{H}}_{s,j}' \right)^{-1} \hat{\lambda}_{j,i}^s - \lambda_{j,i}^s \right] \quad = \quad w_{j,i}^{*s} + o_p(1), \qquad j = 1, 2 \tag{3.8.128}$$

where $\hat{\mathcal{H}}_{s,j}$, $j = 1, 2$ are the same non-singular matrices in equation (3.8.126), and

$$\begin{aligned}
w_{j,i}^{*s} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(f_{j,t}^{s} + \frac{1}{\sqrt{T}} K_j' f_t^c \right) e_{j,it}, \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(f_{j,t}^{s} + \frac{1}{\sqrt{T}} K_j' f_t^c \right) \left(\varepsilon_{j,it} - \frac{1}{\sqrt{T}} f_t^{c\prime} \varphi_{j,i} \right), \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_{j,t}^{s} \varepsilon_{j,it} - \frac{1}{T} \sum_{t=1}^{T} f_{j,t}^{s} f_t^{c\prime} \varphi_{j,i} \\
&\quad + K_j' \frac{1}{T} \sum_{t=1}^{T} f_t^{c\prime} \varepsilon_{j,it} - K_j' \frac{1}{T\sqrt{T}} \sum_{t=1}^{T} f_t^c f_t^{c\prime} \varphi_{j,i} \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_{j,t}^{s} \varepsilon_{j,it} + o_p(1), \qquad (3.8.129)
\end{aligned}$$

since $\frac{1}{T} \sum_{t=1}^{T} f_{j,t}^{s} f_t^{c\prime} = o_p(1)$. Therefore, we get:

$$\sqrt{T} \left[\left(\hat{\mathcal{H}}_{s,j}' \right)^{-1} \hat{\lambda}_{j,i}^{s} - \lambda_{j,i}^{s} \right] = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_{j,t}^{s} \varepsilon_{j,it} + o_p(1) \equiv w_{j,i}^{s} + o_p(1), \qquad (3.8.130)$$

which yields equation (3.4.20).

Q.E.D.

3.8.5 Proof of Theorem 4

Theorem 4 follows from Theorem 1 since we have:

$$\begin{aligned}
tr \left\{ \hat{\Sigma}_{cc}^{-1} \hat{\Sigma}_U \right\} &= tr \left\{ \tilde{\Sigma}_{cc}^{-1} \Sigma_{U,N} \right\} + o_p(1/\sqrt{T}), \qquad (3.8.131) \\
tr \left\{ \hat{\Sigma}_U^2 \right\} &= tr \left\{ \Sigma_U(0)^2 \right\} + o_p(1). \qquad (3.8.132)
\end{aligned}$$

These expansions are proved next.

Asymptotic expansion of $\hat{\Sigma}_{cc}^{-1}$

Substituting the expression of $\hat{f}_t^c$ from equation (3.8.111) into $\hat{\Sigma}_{cc} = \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t^c \hat{f}_t^{c\prime}$ we get:

$$\begin{aligned}
\hat{\Sigma}_{cc} &= \frac{1}{T} \sum_{t=1}^{T} \hat{\mathcal{H}}_c^{-1} \left(f_t^c + \frac{1}{\sqrt{N_1}} u_{j,t}^{(c)} \right) \left(f_t^c + \frac{1}{\sqrt{N_1}} u_{j,t}^{(c)} \right)' \left(\hat{\mathcal{H}}_c^{-1} \right)' + O_p\left(\frac{1}{\sqrt{NT}} \right) \\
&= \hat{\mathcal{H}}_c^{-1} \tilde{\Sigma}_{cc} \left(\hat{\mathcal{H}}_c^{-1} \right)' + O_p\left(\frac{1}{\sqrt{NT}} \right).
\end{aligned}$$

This implies:

$$\hat{\Sigma}_{cc}^{-1} \;=\; \hat{\mathcal{H}}_c' \tilde{\Sigma}_{cc}^{-1} \hat{\mathcal{H}}_c + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{3.8.133}$$

Asymptotic expansion of $\hat{\Sigma}_U$

i) Asymptotic expansion of $\frac{\hat{\Lambda}_j' \hat{\Lambda}_j}{N_j}$

To derive the asymptotic expansion of matrix $\hat{\Lambda}_j' \hat{\Lambda}_j / N_j$, it is useful to write the matrix versions of the quantities defined in equations (3.8.122) and (3.8.130). Stacking the loadings $\hat{\lambda}_{j,i}^c$ in matrix $\hat{\Lambda}_j^c = [\hat{\lambda}_{j,1}^c, ..., \hat{\lambda}_{j,N}^c]'$ we get:

$$\hat{\Lambda}_j^c \;=\; \left[\Lambda_j^c + \frac{1}{\sqrt{T}} G_j^c\right] \hat{\mathcal{H}}_c + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where

$$G_j^c \;=\; \frac{1}{\sqrt{T}} \varepsilon_j' F^c + \Lambda_j^s \left(\frac{1}{\sqrt{T}} F_j^{s\prime} F^c\right) \tag{3.8.134}$$

$$\;=\; \frac{1}{\sqrt{T}} \varepsilon_j' F^c + \Lambda_j^s \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_{j,t}^s f_t^{c\prime}\right). \tag{3.8.135}$$

Similarly, stacking the loadings $\hat{\lambda}_{j,i}^s$ in matrix $\hat{\Lambda}_j^s = [\hat{\lambda}_{j,1}^s, ..., \hat{\lambda}_{j,N}^s]'$ we get:

$$\hat{\Lambda}_j^s \;=\; \left[\Lambda_j^c + \frac{1}{\sqrt{T}} G_j^s\right] \hat{\mathcal{H}}_{j,s} + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where

$$G_j^s \;=\; \frac{1}{\sqrt{T}} \varepsilon_j' F_j^s. \tag{3.8.136}$$

By gathering these expansions, we get:

$$\hat{\Lambda}_j \;\simeq\; \left(\Lambda_j + \frac{1}{\sqrt{T}} G_j\right) \hat{\mathcal{U}}_j, \qquad j = 1, 2, \tag{3.8.137}$$

where

$$G_j \;=\; \left[\begin{array}{ccc} G_j^c & \vdots & G_j^s \end{array}\right], \tag{3.8.138}$$

$$\hat{\mathcal{U}}_j \;=\; \left[\begin{array}{cc} \hat{\mathcal{H}}_c & 0 \\ 0 & \hat{\mathcal{H}}_{s,j} \end{array}\right]. \tag{3.8.139}$$

We start by computing the asymptotic expansion of $\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j}$. From Assumptions 1, 2 and 5 we get:

$$\frac{1}{N_j}\left[\Lambda_j + \frac{1}{\sqrt{T}}G_j\right]'\left[\Lambda_j + \frac{1}{\sqrt{T}}G_j\right] \simeq \frac{1}{N_j}\Lambda_j'\Lambda_j + \frac{1}{N\sqrt{T}}\left(\Lambda_j'G_j + G_j'\Lambda_j\right) + \frac{1}{NT}G_j'G_j. \tag{3.8.140}$$

Let us compute the asymptotic expansion of $\frac{1}{N\sqrt{T}}\Lambda_j'G_j$:

$$\frac{1}{N_j\sqrt{T}}\Lambda_j'G_j = \frac{1}{N_j\sqrt{T}}\begin{bmatrix} \Lambda_j^{c\prime}G_j^c & \Lambda_j^{c\prime}G_j^s \\ \Lambda_j^{s\prime}G_j^c & \Lambda_j^{s\prime}G_j^s \end{bmatrix}. \tag{3.8.141}$$

Using equation (3.8.134) we get:

$$\begin{aligned}\frac{1}{N_j\sqrt{T}}\Lambda_j^{c\prime}G_j^c &= \frac{1}{N_j\sqrt{T}}\Lambda_j^{c\prime}\left[\frac{1}{\sqrt{T}}\varepsilon_j'F^c + \Lambda_j^s\left(\frac{1}{\sqrt{T}}F_j^{s\prime}F^c\right)\right] \\ &= \frac{1}{N_jT}\Lambda_j^{c\prime}\varepsilon_j'F^c + \frac{1}{N_jT}\Lambda_j^{c\prime}\Lambda_j^s\left(F_j^{s\prime}F^c\right) \\ &= \left(\frac{\Lambda_j^{c\prime}\Lambda_j^s}{N_j}\right)\frac{1}{T}\sum_{t=1}^{T} f_{j,t}^s f_t^{c\prime} + O_p\left(\frac{1}{\sqrt{N_jT}}\right),\end{aligned} \tag{3.8.142}$$

Using analogous arguments and equation (3.8.136), we get:

$$\frac{1}{N_j\sqrt{T}}\Lambda_j^{s\prime}G_j^c = \left(\frac{\Lambda_j^{s\prime}\Lambda_j^s}{N_j}\right)\frac{1}{T}\sum_{t=1}^{T} f_{j,t}^s f_t^{c\prime} + O_p\left(\frac{1}{\sqrt{N_jT}}\right), \tag{3.8.143}$$

$$\frac{1}{N_j\sqrt{T}}\Lambda_j^{c\prime}G_j^s = \frac{1}{N_j\sqrt{T}}\Lambda_j^{c\prime}\varepsilon_j'F^s = O_p\left(\frac{1}{\sqrt{N_jT}}\right), \tag{3.8.144}$$

$$\frac{1}{N_j\sqrt{T}}\Lambda_j^{s\prime}G_j^s = \frac{1}{N_j\sqrt{T}}\Lambda_j^{s\prime}\varepsilon_j'F^s = O_p\left(\frac{1}{\sqrt{N_jT}}\right). \tag{3.8.145}$$

The last four equations imply:

$$\begin{aligned}\frac{1}{N_j\sqrt{T}}\Lambda_j'G_j &= \begin{bmatrix} \left(\frac{\Lambda_j^{c\prime}\Lambda_j^s}{N_j}\right)\frac{1}{T}\sum_{t=1}^{T} f_{j,t}^s f_t^{c\prime} & 0 \\ \left(\frac{\Lambda_j^{s\prime}\Lambda_j^s}{N_j}\right)\frac{1}{T}\sum_{t=1}^{T} f_{j,t}^s f_t^{c\prime} & 0 \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N_jT}}\right) \\ &= \begin{bmatrix} \left(\frac{\Lambda_j'\Lambda_j^s}{N_j}\right)\frac{1}{T}\sum_{t=1}^{T} f_{j,t}^s f_t^{c\prime} & \vdots & 0_{(k_j\times k_j^s)} \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N_jT}}\right).\end{aligned} \tag{3.8.146}$$

Using analogous arguments, we have:

$$\begin{aligned}\frac{1}{N_jT}G_j^{c\prime}G_j^c &= \frac{1}{N_jT}\left[\frac{1}{\sqrt{T}}\varepsilon_j'F^c+\Lambda_j^s\left(\frac{1}{\sqrt{T}}F_j^{s\prime}F^c\right)\right]'\left[\frac{1}{\sqrt{T}}\varepsilon_j'F^c+\Lambda_j^s\left(\frac{1}{\sqrt{T}}F_j^{s\prime}F^c\right)\right]\\ &= o_p\left(\frac{1}{\sqrt{T}}\right)\end{aligned} \tag{3.8.147}$$

and

$$\frac{1}{N_jT}G_j'G_j = o_p\left(\frac{1}{\sqrt{T}}\right). \tag{3.8.148}$$

Substituting (3.8.146) and (3.8.148) into equation (3.8.140) we get:

$$\frac{1}{N_j}\left[\Lambda_j+\frac{1}{\sqrt{T}}G_j\right]'\left[\Lambda_j+\frac{1}{\sqrt{T}}G_j\right] \simeq \Sigma_{\Lambda,j}+\frac{1}{\sqrt{T}}\left(L_{1,j}+L_{1,j}'\right)+O_p\left(\frac{1}{\sqrt{N}}\right) \tag{3.8.149}$$

where

$$L_{1,j} = \left[\ \left(\frac{\Lambda_j'\Lambda_j^s}{N_j}\right)\left(\frac{1}{\sqrt{T}}F_j^{s\prime}F^c\right)\ \vdots\ 0_{(k_j\times k_j^s)}\ \right]. \tag{3.8.150}$$

Therefore we have:

$$\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j} = \hat{\mathcal{U}}_j'\left[\Sigma_{\Lambda,j}+\frac{1}{\sqrt{T}}\left(L_{1,j}+L_{1,j}'\right)\right]\hat{\mathcal{U}}_j+o_p\left(\frac{1}{\sqrt{T}}\right). \tag{3.8.151}$$

ii) Asymptotic expansion of $\hat{\Gamma}_j$

The approximations in Propositions 2 and 3 allow to compute the asymptotic expansion of $\hat{\varepsilon}_{j,it}$:

$$\begin{aligned}\hat{\varepsilon}_{j,it} &= y_{j,it}-\hat{\lambda}_{j,i}^{c\,\prime}\hat{f}_t^c-\hat{\lambda}_{j,i}^{s\,\prime}\hat{f}_{j,t}^s\\ &= \varepsilon_{j,it}-\left[\hat{\lambda}_{j,i}^{c\,\prime}\hat{f}_t^c-\lambda_{j,i}^{c\,\prime}f_t^c\right]-\left[\hat{\lambda}_{j,i}^{s\,\prime}\hat{f}_{j,t}^s-\lambda_{j,i}^{s\,\prime}f_{j,t}^s\right]\\ &\simeq \varepsilon_{j,it}-\left[\left(\lambda_{j,i}^c+\frac{1}{\sqrt{T}}w_{j,i}^c\right)'\left(f_t^c+\frac{1}{\sqrt{N_1}}u_{1,t}^{(c)}\right)-\lambda_{j,i}^{c\,\prime}f_t^c\right]\\ &\quad -\left[\left(\lambda_{j,i}^s+\frac{1}{\sqrt{T}}w_{j,i}^s\right)'\left(f_{j,t}^s-\frac{1}{\sqrt{T}}K_j'f_t^c+\frac{1}{\sqrt{N_j}}v_{j,t}^s\right)-\lambda_{j,i}^{s\,\prime}f_{j,t}^s\right]\\ &\simeq \varepsilon_{j,it}-\left(\frac{1}{\sqrt{N_1}}\lambda_{j,i}^{c\,\prime}u_{1,t}^{(c)}+\frac{1}{\sqrt{T}}w_{j,i}^{c\prime}f_t^c\right)-\left(\frac{1}{\sqrt{N_j}}\lambda_{j,i}^{s\,\prime}v_{j,t}^s+\frac{1}{\sqrt{T}}w_{j,i}^{s\prime}f_{j,t}^s\right)\\ &\quad +\lambda_{j,i}^{s\,\prime}\frac{1}{\sqrt{T}}K_j'f_t^c.\end{aligned} \tag{3.8.152}$$

Since $T/N_j=o(1)$, we keep only the terms of order $1/\sqrt{T}$ in equation (3.8.152), and we get:

$$\hat{\varepsilon}_{j,it} = \varepsilon_{j,it}-\frac{1}{\sqrt{T}}\left(w_{j,i}^{c\prime}f_t^c+w_{j,i}^{s\prime}f_{j,t}^s\right)+\lambda_{j,i}^{s\,\prime}\frac{1}{\sqrt{T}}K_j'f_t^c+o_p\left(\frac{1}{\sqrt{T}}\right). \tag{3.8.153}$$

From the definition of $w^c_{j,i}$ in Proposition 3 we get:

$$w^{c\prime}_{j,i} f^c_t = \frac{1}{\sqrt{T}} \left(\sum_{r=1}^{T} \varepsilon_{j,ir} f^{c\prime}_r \right) f^c_t + \lambda^{s\prime}_{j,i} K'_j f^c_t, \tag{3.8.154}$$

which implies:

$$\hat{\varepsilon}_{j,it} = \varepsilon_{j,it} - \frac{1}{\sqrt{T}} \left(\tilde{w}^{c\prime}_{j,i} f^c_t + w^{s\prime}_{j,i} f^s_{j,t} \right) + o_p \left(\frac{1}{\sqrt{T}} \right), \tag{3.8.155}$$

where:

$$\tilde{w}^c_{j,i} = \frac{1}{\sqrt{T}} \sum_{r=1}^{T} f^c_r \varepsilon_{j,ir}. \tag{3.8.156}$$

Equation (3.8.153) allows us to compute:

$$\begin{aligned}
\hat{\gamma}_{j,ii} &= \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}^2_{j,it} \\
&\simeq \frac{1}{T} \sum_{t=1}^{T} \left[\varepsilon_{j,it} - \frac{1}{\sqrt{T}} \left(\tilde{w}^{c\prime}_{j,i} f^c_t + w^{s\prime}_{j,i} f^s_{j,t} \right) \right]^2 \\
&= \frac{1}{T} \sum_{t=1}^{T} \varepsilon^2_{j,it} - \frac{2}{T\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{j,it} \left(\tilde{w}^{c}_{j,i}{}' f^c_t + w^{s}_{j,i}{}' f^s_{j,t} \right) + \frac{1}{T^2} \sum_{t=1}^{T} \left(\tilde{w}^{c}_{j,i}{}' f^c_t + w^{s}_{j,i}{}' f^s_{j,t} \right)^2 .
\end{aligned} \tag{3.8.157}$$

Using $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{j,it} f^c_t = O_p(1)$ and $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{j,it} f^s_{j,t} = O_p(1)$ we get:

$$\hat{\gamma}_{j,ii} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon^2_{j,it} + O_p \left(\frac{1}{T} \right), \tag{3.8.158}$$

which implies:

$$\begin{aligned}
\hat{\gamma}_{j,ii} &= \frac{1}{T} \sum_{t=1}^{T} \varepsilon^2_{j,it} + o_p \left(\frac{1}{\sqrt{T}} \right) \\
&= \gamma_{j,ii} + \frac{1}{\sqrt{T}} w_{j,i} + o_p \left(\frac{1}{\sqrt{T}} \right),
\end{aligned} \tag{3.8.159}$$

where

$$w_{j,i} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon^2_{j,it} - \gamma_{j,ii}) = O_p(1), \tag{3.8.160}$$

from Assumptions 4 and 7. Therefore, we have:

$$\hat{\Gamma}_j \simeq \Gamma_j + \frac{1}{\sqrt{T}} W_j. \tag{3.8.161}$$

where $\Gamma_j = diag(\gamma_{j,ii},\ i = 1, ..., N)$ and $W_j = diag(w_{j,i},\ i = 1, ..., N)$, for $j = 1, 2$.

iii) Asymptotic expansion of $\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j\hat{\Lambda}_j$

Let us define

$$
\begin{aligned}
\hat{\Omega}_j^* &:= \frac{1}{N_j}\left(\Lambda_j + \frac{1}{\sqrt{T}}G_j\right)'\hat{\Gamma}\left(\Lambda_j + \frac{1}{\sqrt{T}}G_j\right) \\
&= \frac{1}{N_j}\left(\Lambda_j + \frac{1}{\sqrt{T}}G_j\right)'\left(\Gamma_j + \frac{1}{\sqrt{T}}W_j\right)\left(\Lambda_j + \frac{1}{\sqrt{T}}G_j\right) \\
&= \frac{1}{N_j}\Lambda_j'\Gamma_j\Lambda_j + \hat{\Omega}_{j,I}^* + \hat{\Omega}_{j,II}^* + \hat{\Omega}_{j,III}^* + \hat{\Omega}_{j,II}^{*\,\prime} + \hat{\Omega}_{j,III}^{*\,\prime} + \hat{\Omega}_{j,IV}^* + \hat{\Omega}_{j,V}^*, \qquad (3.8.162)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\Omega}_{j,I}^* &= \frac{1}{N_j\sqrt{T}}\Lambda_j'W_j\Lambda_j = O_p\left(\frac{1}{\sqrt{NT}}\right), && (3.8.163) \\
\hat{\Omega}_{j,III}^* &= \frac{1}{N_jT}\Lambda_j'W_jG_j = O_p\left(\frac{1}{T}\right), && (3.8.164) \\
\hat{\Omega}_{j,IV}^* &= \frac{1}{N_jT}G_j'\Gamma_jG_j = O_p\left(\frac{1}{T}\right), && (3.8.165) \\
\hat{\Omega}_{j,V}^* &= \frac{1}{N_jT\sqrt{T}}G_j'W_jG_j = O_p\left(\frac{1}{T\sqrt{T}}\right). && (3.8.166)
\end{aligned}
$$

Moreover, similarly as for (3.8.146) we have:

$$
\begin{aligned}
\hat{\Omega}_{j,II}^* &= \frac{1}{N_j\sqrt{T}}\Lambda_j'\Gamma_jG_j \\
&= \left[\ \frac{1}{N_j}\Lambda_j'\Gamma_j\Lambda_j^s\left(\frac{1}{T}\sum_{t=1}^{T}f_{j,t}^s f_t^{c\prime}\right)\ \vdots\ 0_{(k_j\times k_j^s)}\ \right] + o_p\left(\frac{1}{\sqrt{T}}\right), && (3.8.167) \\
&= \frac{1}{\sqrt{T}}\left[\ \left(\frac{1}{N_j}\Lambda_j'\Gamma_j\Lambda_j^s\right)\left(\frac{1}{\sqrt{T}}F_j^{s\,\prime}F_c\right)\ \vdots\ 0_{(k_j\times k_j^s)}\ \right] + o_p\left(\frac{1}{\sqrt{T}}\right), && (3.8.168) \\
&= \frac{1}{\sqrt{T}}L_{2,j} + o_p\left(\frac{1}{\sqrt{T}}\right), && (3.8.169) \\
& && (3.8.170)
\end{aligned}
$$

where

$$
L_{2,j} = \left[\ \left(\frac{1}{N}\Lambda_j'\Gamma_j\Lambda_j^s\right)\left(\frac{1}{\sqrt{T}}F_j^{s\,\prime}F_c\right)\ \vdots\ 0_{(k_j\times k_j^s)}\ \right].
$$

Collecting the previous results, using $T/N = o_p(1)$, and defining $\Omega_j^* = \lim_{N\to\infty}\frac{1}{N}\Lambda_j'\Gamma_j\Lambda_j$ we get:

$$
\begin{aligned}
\hat{\Omega}_j^* &= \frac{1}{N}\Lambda_j'\Gamma_j\Lambda_j + \frac{1}{\sqrt{T}}\left(L_{2,j} + L_{2,j}'\right) + o_p\left(\frac{1}{\sqrt{T}}\right) \\
&= \Omega_j^* + \frac{1}{\sqrt{T}}\left(L_{2,j} + L_{2,j}'\right) + o_p\left(\frac{1}{\sqrt{T}}\right). && (3.8.171)
\end{aligned}
$$

Substituting equation (3.8.137) into $\frac{1}{N_j}\hat{\Lambda}_j'\hat{\Gamma}_j\hat{\Lambda}_j$, and using equation (3.8.171) we get:

$$\begin{aligned}\hat{\Omega}_j &= \hat{\mathcal{U}}_j'\hat{\Omega}_j^*\hat{\mathcal{U}}_j \\ &= \hat{\mathcal{U}}_j'\left[\Omega_j^* + \frac{1}{\sqrt{T}}\left(L_{2,j}+L_{2,j}'\right)\right]\hat{\mathcal{U}}_j + o_p\left(\frac{1}{\sqrt{T}}\right), \qquad j=1,2. \end{aligned} \tag{3.8.172}$$

iv) Asymptotic expansion of $\hat{\Sigma}_U$

The estimator of $\Sigma_{u,j}$ is given in equation (3.4.22). Equation (3.8.151) allows to compute the asymptotic approximation of $\left(\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j}\right)^{-1}$:

$$\left(\frac{\hat{\Lambda}_j'\hat{\Lambda}_j}{N_j}\right)^{-1} \simeq \hat{\mathcal{U}}_j^{-1}\left[\Sigma_{\Lambda,j}^{-1} - \frac{1}{\sqrt{T}}\Sigma_{\Lambda,j}^{-1}\left(L_{1,j}+L_{1,j}'\right)\Sigma_{\Lambda,j}^{-1}\right]\left(\hat{\mathcal{U}}_j'\right)^{-1}. \tag{3.8.173}$$

Substituting equations (3.8.173) and (3.8.172) into equation (3.4.22), we get:

$$\begin{aligned}\hat{\Sigma}_{u,j} &\simeq \hat{\mathcal{U}}_j^{-1}\left[\Sigma_{\Lambda,j}^{-1} - \frac{1}{\sqrt{T}}\Sigma_{\Lambda,j}^{-1}\left(L_{1,j}+L_{1,j}'\right)\Sigma_{\Lambda,j}^{-1}\right]\left[\Omega_j^* + \frac{1}{\sqrt{T}}\left(L_{2,j}+L_{2,j}'\right)\right] \\ &\quad \times\left[\Sigma_{\Lambda,j}^{-1} - \frac{1}{\sqrt{T}}\Sigma_{\Lambda,j}^{-1}\left(L_{1,j}+L_{1,j}'\right)\Sigma_{\Lambda,j}^{-1}\right]\left(\hat{\mathcal{U}}_j'\right)^{-1} \\ &\simeq \hat{\mathcal{U}}_j^{-1}\Sigma_{\Lambda,j}^{-1}\left[I - \frac{1}{\sqrt{T}}\left(L_{1,j}+L_{1,j}'\right)\Sigma_{\Lambda,j}^{-1}\right]\left[\Omega_j^* + \frac{1}{\sqrt{T}}\left(L_{2,j}+L_{2,j}'\right)\right] \\ &\quad \times\left[I - \frac{1}{\sqrt{T}}\Sigma_{\Lambda,j}^{-1}\left(L_{1,j}+L_{1,j}'\right)\right]\Sigma_{\Lambda,j}^{-1}\left(\hat{\mathcal{U}}_j'\right)^{-1} \\ &\simeq \hat{\mathcal{U}}_j^{-1}\Sigma_{\Lambda,j}^{-1}\left[\Omega_j^* + \frac{1}{\sqrt{T}}\left(L_{2,j}+L_{2,j}'\right) - \frac{1}{\sqrt{T}}\Omega_j^*\Sigma_{\Lambda,j}^{-1}\left(L_{1,j}+L_{1,j}'\right) - \frac{1}{\sqrt{T}}\left(L_{1,j}+L_{1,j}'\right)\Sigma_{\Lambda,j}^{-1}\Omega_j^*\right] \\ &\quad \times\Sigma_{\Lambda,j}^{-1}\left(\hat{\mathcal{U}}_j'\right)^{-1},\end{aligned}$$

which implies:

$$\hat{\Sigma}_{u,j} = \hat{\mathcal{U}}_j^{-1}\Sigma_{u,j}\left(\hat{\mathcal{U}}_j'\right)^{-1} + \frac{1}{\sqrt{T}}\hat{\mathcal{U}}_j^{-1}L_{3,j}\left(\hat{\mathcal{U}}_j'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where

$$L_{3,j} = \Sigma_{\Lambda,j}^{-1}\left[\left(L_{2,j}+L_{2,j}'\right) - \Omega_j^*\Sigma_{\Lambda,j}^{-1}\left(L_{1,j}+L_{1,j}'\right) - \left(L_{1,j}+L_{1,j}'\right)\Sigma_{\Lambda,j}^{-1}\Omega_j^*\right]\Sigma_{\Lambda,j}^{-1}. \tag{3.8.174}$$

From equation (3.8.139) we have:

$$\begin{aligned}\hat{\Sigma}_U &= \mu_N^2\hat{\Sigma}_{u,1}^{(cc)} + \hat{\Sigma}_{u,2}^{(cc)} \\ &= \hat{\mathcal{H}}_c^{-1}\left[\mu_N^2\Sigma_{u,1}+\Sigma_{u,2}\right]^{(cc)}\left(\hat{\mathcal{H}}_c'\right)^{-1} + \frac{1}{\sqrt{T}}\hat{\mathcal{H}}_c^{-1}\left(\mu_N^2 L_{3,1}+L_{3,2}\right)^{(cc)}\left(\hat{\mathcal{H}}_c'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right) \\ &= \hat{\mathcal{H}}_c^{-1}\Sigma_{U,N}\left(\hat{\mathcal{H}}_c'\right)^{-1} + \frac{1}{\sqrt{T}}\hat{\mathcal{H}}_c^{-1}\left(\mu_N^2 L_{3,1}+L_{3,2}\right)^{(cc)}\left(\hat{\mathcal{H}}_c'\right)^{-1} + o_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned} \tag{3.8.175}$$

This expansion, the convergence $\Sigma_{U,N} \to \Sigma_U(0)$ and the commutative property of the trace, imply equation (3.8.132).

Asymptotic expansion of $tr\left\{\tilde{\Sigma}_{cc}^{-1}\hat{\Sigma}_U\right\}$

Results (3.8.133) and (3.8.175), and the commutative property of the trace, imply:

$$tr\left\{\hat{\Sigma}_{cc}^{-1}\hat{\Sigma}_U\right\} \; = \; tr\left\{\tilde{\Sigma}_{cc}^{-1}\Sigma_{U,N}\right\} + \frac{1}{\sqrt{T}} tr\left\{\tilde{\Sigma}_{cc}^{-1}\left(\mu_N^2 L_{3,1} + L_{3,2}\right)^{(cc)}\right\} + o_p\left(\frac{1}{\sqrt{T}}\right).$$

Noting that $L_{3,j} = O_p(1)$, for $j = 1, 2$, and recalling that $\tilde{\Sigma}_{cc} = I_{k^c} + O_p(1/\sqrt{T})$ and $\mu_N = \mu + o(1)$, the last equation can be further simplified to

$$tr\left\{\hat{\Sigma}_{cc}^{-1}\hat{\Sigma}_U\right\} \; = \; tr\left\{\tilde{\Sigma}_{cc}^{-1}\Sigma_{U,N}\right\} + \frac{1}{\sqrt{T}} tr\left\{\left(\mu^2 L_{3,1} + L_{3,2}\right)^{(cc)}\right\} + o_p\left(\frac{1}{\sqrt{T}}\right). \tag{3.8.176}$$

Let us compute $L_{3,j}$ explicitly. From equation (3.8.150) we get:

$$\begin{aligned} L_{1,j} \;&=\; \begin{bmatrix} \left(\frac{\Lambda_j^{c\prime}\Lambda_j^s}{N}\right)\left(\frac{1}{\sqrt{T}}F_j^{s\,\prime}F^c\right) & 0_{(k^c\times k_j^s)} \\ \left(\frac{\Lambda_j^{s\prime}\Lambda_j^s}{N}\right)\left(\frac{1}{\sqrt{T}}F_j^{s\,\prime}F^c\right) & 0_{(k_j^s\times k_j^s)} \end{bmatrix} \\ &=\; \begin{bmatrix} \Sigma_{\Lambda,j,cs}\left(\frac{1}{\sqrt{T}}F_j^{s\,\prime}F^c\right) & 0 \\ \Sigma_{\Lambda,j,ss}\left(\frac{1}{\sqrt{T}}F_j^{s\,\prime}F^c\right) & 0 \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &=\; \Sigma_{\Lambda,j}\begin{bmatrix} 0_{(k^c\times k^c)} & 0_{(k^c\times k_j^s)} \\ K_j' & 0_{(k_j^s\times k_j^s)} \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \tag{3.8.177}$$

Equation (3.8.177) implies:

$$\Omega_j^*\Sigma_{\Lambda,j}^{-1}L_{1,j} \;=\; L_{2,j} + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{3.8.178}$$

Substituting results (3.8.177) and (3.8.178) into equation (3.8.174) we get:

$$\begin{aligned} L_{3,j} \;&=\; -\Sigma_{\Lambda,j}^{-1}\left[\Omega_j^*\Sigma_{\Lambda,j}^{-1}L_{1,j}' + L_{1,j}\Sigma_{\Lambda,j}^{-1}\Omega_j^*\right]\Sigma_{\Lambda,j}^{-1} \\ &=\; -\Sigma_{u,j}L_{1,j}'\Sigma_{\Lambda,j}^{-1} - \Sigma_{\Lambda,j}^{-1}L_{1,j}\Sigma_{u,j} + O_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \tag{3.8.179}$$

Moreover, noting that:

$$\Sigma_{\Lambda,j}^{-1}L_{1,j} \;=\; \begin{bmatrix} 0_{(k^c\times k^c)} & 0_{(k^c\times k_j^s)} \\ \left(\frac{1}{\sqrt{T}}F_j^{s\,\prime}F^c\right) & 0_{(k_j^s\times k_j^s)} \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{3.8.180}$$

we get:

$$\Sigma_{\Lambda,j}^{-1}L_{1,j}\Sigma_{u,j} \;=\; \begin{bmatrix} 0_{(k^c\times k^c)} & 0_{(k^c\times k_j^s)} \\ * & * \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{3.8.181}$$

Equation (3.8.181) implies:

$$(L_{3,j})^{(cc)} = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{3.8.182}$$

Finally, substituting result (3.8.182) into equation (3.8.176), equation (3.8.131) follows.

Q.E.D.

3.8.6 Proof of Theorem 5

Let us re-write the model for the high frequency observables $x^H_{m,t}$, where $m = 1, ..., M$, and $t = 1, ..., T$ in equation (3.2.1) as:

$$\begin{aligned} x^H_{m,t} &= \Lambda_{HC} g^C_{m,t} + \Lambda_H g^H_{m,t} + e^H_{m,t}, \\ &= \Lambda_1 g_{m,t} + e^H_{m,t}, \\ &= \hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} g_{m,t} - \left(\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1\right) g_{m,t} + e^H_{m,t}, \end{aligned} \tag{3.8.183}$$

where $g_{m,t} = [\, g^{C\,\prime}_{m,t} \vdots g^{H\,\prime}_{m,t}\,]'$, $\Lambda_1 = [\Lambda_{HC} \vdots \Lambda_H] = [\Lambda_1^c \vdots \Lambda_1^s]$, $\hat{\Lambda}_1 = [\hat{\Lambda}_{HC} \vdots \hat{\Lambda}_H] = [\hat{\Lambda}_1^c \vdots \hat{\Lambda}_1^s]$, and $\hat{\mathcal{U}}_1$ has been defined in equation (3.8.139). Let us also define the estimator $\hat{g}_{m,t} = [\, \hat{g}^{C\,\prime}_{m,t} \vdots \hat{g}^{H\,\prime}_{m,t}\,]'$ as in equation (3.3.6):

$$\hat{g}_{m,t} = \begin{bmatrix} \hat{g}^C_{m,t} \\ \hat{g}^H_{m,t} \end{bmatrix} = \left(\hat{\Lambda}_1' \hat{\Lambda}_1\right)^{-1} \hat{\Lambda}_1' x^H_{m,t}, \qquad m = 1, ..., M, \qquad t = 1, ..., T. \tag{3.8.184}$$

Substituting equation (3.8.183) into equation (3.8.184), and rearranging terms, we get:

$$\hat{g}_{m,t} = \hat{\mathcal{U}}_1^{-1} g_{m,t} - \left(\frac{\hat{\Lambda}_1'\hat{\Lambda}_1}{N_H}\right)^{-1} \frac{1}{N_H} \hat{\Lambda}_1' \left(\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1\right) g_{m,t} + \left(\frac{\hat{\Lambda}_1'\hat{\Lambda}_1}{N_H}\right)^{-1} \frac{1}{N_H} \hat{\Lambda}_1' e^H_{m,t}. \tag{3.8.185}$$

From equations (3.8.150) and (3.8.151) we have:

$$\frac{\hat{\Lambda}_1'\hat{\Lambda}_1}{N_H} = \hat{\mathcal{U}}_1' \Sigma_{\Lambda,1} \hat{\mathcal{U}}_1 + O_p\left(\frac{1}{\sqrt{T}}\right),$$

which implies:

$$\left(\frac{\hat{\Lambda}_1'\hat{\Lambda}_1}{N_H}\right)^{-1} = \hat{\mathcal{U}}_1^{-1} \Sigma_{\Lambda,1}^{-1} \left(\hat{\mathcal{U}}_1'\right)^{-1} + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{3.8.186}$$

From equations (3.8.134) - (3.8.138) we get:

$$\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1 \simeq \frac{1}{\sqrt{T}} G_1, \tag{3.8.187}$$

where

$$G_1 = \left[G_1^c \vdots G_1^s \right], \tag{3.8.188}$$

with

$$G_1^c = \frac{1}{\sqrt{T}} \bar{e}^{H\prime} \bar{g}^C + \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H\prime} \bar{g}^C \right), \tag{3.8.189}$$

$$G_1^s = \frac{1}{\sqrt{T}} \bar{e}^{H\prime} \bar{g}^H, \tag{3.8.190}$$

$\bar{e}^H = \left[\bar{e}_1^H, ..., \bar{e}_T^H\right]'$, $\bar{g}^C = \left[\bar{g}_1^C, ..., \bar{g}_T^C\right]'$ and $\bar{g}^H = \left[\bar{g}_1^H, ..., \bar{g}_T^H\right]'$. Moreover, we have:

$$\hat{\Lambda}_1 \simeq \Lambda_1 \hat{\mathcal{U}}_1 + \frac{1}{\sqrt{T}} G_1 \hat{\mathcal{U}}_1. \tag{3.8.191}$$

From equations (3.8.187) and (3.8.191) it follows:

$$\begin{aligned}
\frac{1}{N_H} \hat{\Lambda}_1' \left(\hat{\Lambda}_1 \hat{\mathcal{U}}_1^{-1} - \Lambda_1 \right) &\simeq \frac{1}{N_H} \left(\Lambda_1 \hat{\mathcal{U}}_1 + \frac{1}{\sqrt{T}} G_1 \hat{\mathcal{U}}_1 \right)' \frac{1}{\sqrt{T}} G_1 \\
&= \frac{1}{N_H \sqrt{T}} \hat{\mathcal{U}}_1' \Lambda_1' G_1 + \frac{1}{N_H T} \hat{\mathcal{U}}_1' G_1' G_1. \qquad (3.8.192)
\end{aligned}$$

Equations (3.8.186) and (3.8.192) allow to express the second term in the r.h.s. of equation (3.8.185) as:

$$\left(\frac{\hat{\Lambda}_1' \hat{\Lambda}_1}{N_H} \right)^{-1} \frac{1}{N_H} \hat{\Lambda}_1' \left(\hat{\Lambda}_1' \hat{\mathcal{U}}_1^{-1} - \Lambda_1 \right) g_{m,t} \simeq \hat{\mathcal{U}}_1^{-1} \Sigma_{\Lambda,1}^{-1} \frac{1}{N_H \sqrt{T}} \Lambda_1' G_1 g_{m,t} + \hat{\mathcal{U}}_1^{-1} \Sigma_{\Lambda,1}^{-1} \frac{1}{N_H T} G_1' G_1 g_{m,t}. \tag{3.8.193}$$

From equation (3.8.146) we have:

$$\frac{1}{N_H \sqrt{T}} \Lambda_1' G_1 = \left[\left(\frac{\Lambda_1' \Lambda_H}{N_H} \right) \frac{1}{T} \sum_{t=1}^{T} \bar{g}_t^H \bar{g}_t^{C\prime} \ \vdots \ 0_{(k_1 \times k^H)} \right] + O_p \left(\frac{1}{\sqrt{N_H T}} \right), \tag{3.8.194}$$

where $k_1 = k^C + k^H$. From equation (3.8.188) we have:

$$\frac{1}{N_H T} G_1' G_1 = \frac{1}{N_H T} \begin{bmatrix} G_1^{c\prime} G_1^c & G_1^{c\prime} G_1^s \\ G_1^{s\prime} G_1^c & G_1^{s\prime} G_1^s \end{bmatrix}. \tag{3.8.195}$$

Equation (3.8.189) implies:

$$\begin{aligned}
\frac{1}{N_H T} G_1^{c\prime} G_1^c &= \frac{1}{N_H T} \left[\frac{1}{\sqrt{T}} \bar{e}^{H\prime} \bar{g}^C + \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H\prime} \bar{g}^C \right) \right]' \left[\frac{1}{\sqrt{T}} \bar{e}^{H\prime} \bar{g}^C + \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H\prime} \bar{g}^C \right) \right] \\
&= \frac{1}{N_H T^2} \bar{g}^{C\prime} \bar{e}^H \bar{e}^{H\prime} \bar{g}^C + \frac{1}{N_H T \sqrt{T}} \bar{g}^{C\prime} \bar{e}^H \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H\prime} \bar{g}^C \right) \\
&\quad + \frac{1}{N_H T \sqrt{T}} \left(\frac{1}{\sqrt{T}} \bar{g}^{H\prime} \bar{g}^C \right)' \Lambda_H' \bar{e}^{H\prime} \bar{g}^C + \frac{1}{N_H T} \left(\frac{1}{\sqrt{T}} \bar{g}^{H\prime} \bar{g}^C \right)' \Lambda_H' \Lambda_H \left(\frac{1}{\sqrt{T}} \bar{g}^{H\prime} \bar{g}^C \right) \\
&= O_p \left(\frac{1}{T} \right), \qquad (3.8.196)
\end{aligned}$$

where the last equality follows from the assumption $T/N_H = o(1)$. Equation (3.8.196) and the assumption $\sqrt{N_H}/T = o(1)$ imply:

$$\frac{1}{N_H T} G_1^{c\prime} G_1^c \;\; = \;\; o_p\left(\frac{1}{\sqrt{N_H}}\right). \tag{3.8.197}$$

Similar arguments applied to the other blocks of the matrix in the r.h.s. of (3.8.195) yield:

$$\frac{1}{N_H T} G_1' G_1 \;\; = \;\; o_p\left(\frac{1}{\sqrt{N_H}}\right). \tag{3.8.198}$$

Substituting equations (3.8.194) and (3.8.198) into equation (3.8.193) we get:

$$\left(\frac{\hat{\Lambda}_1'\hat{\Lambda}_1}{N_H}\right)^{-1} \frac{1}{N_H}\hat{\Lambda}_1'\left(\hat{\Lambda}_1\hat{\mathcal{U}}_1^{-1} - \Lambda_1\right) g_{m,t} \;\; \simeq \;\; \hat{\mathcal{U}}_1^{-1}\Sigma_{\Lambda,1}^{-1}\left(\frac{\Lambda_1'\Lambda_H}{N_H}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\bar{g}_t^H \bar{g}_t^{C\prime}\right) g_{m,t}^C + o_p\left(\frac{1}{\sqrt{N_H}}\right). \tag{3.8.199}$$

Let us now focus on the third term in the r.h.s. of equation (3.8.185). From equation (3.8.191) we have:

$$\begin{aligned} \frac{1}{N_H}\hat{\Lambda}_1' e_{m,t}^H \;\; &\simeq \;\; \frac{1}{N_H}\left(\Lambda_1\hat{\mathcal{U}}_1 + \frac{1}{\sqrt{T}}G_1\hat{\mathcal{U}}_1\right)' e_{m,t}^H \\ &= \;\; \hat{\mathcal{U}}_1' \frac{1}{N_H}\Lambda_1' e_{m,t}^H + \hat{\mathcal{U}}_1' \frac{1}{N_H\sqrt{T}} G_1' e_{m,t}^H. \end{aligned} \tag{3.8.200}$$

The second term in the r.h.s. of equation (3.8.200) can be written as:

$$\frac{1}{N_H\sqrt{T}} G_1' e_{m,t}^H \;\; = \;\; \frac{1}{N_H\sqrt{T}}\begin{bmatrix} G_1^{c\prime} e_{m,t}^H \\ G_1^{s\prime} e_{m,t}^H \end{bmatrix}. \tag{3.8.201}$$

Using equation (3.8.189) we get:

$$\begin{aligned} \frac{1}{N_H\sqrt{T}} G_1' e_{m,t}^H \;\; &= \;\; \frac{1}{N_H T}\bar{g}^{C\prime}\bar{e}^H e_{m,t}^H + \frac{1}{N_H\sqrt{T}}\left(\frac{1}{\sqrt{T}}\bar{g}^{C\prime}\bar{g}^H\right)\Lambda_H' e_{m,t}^H \\ &= \;\; O_p\left(\frac{1}{\sqrt{N_H T}}\right). \end{aligned} \tag{3.8.202}$$

Equation (3.8.190) implies:

$$\frac{1}{N_H\sqrt{T}} G_1^{s\prime} e_{m,t}^H \;\; = \;\; \frac{1}{N_H T}\bar{g}^{H\prime}\bar{e}^H e_{m,t}^H \; = \; O_p\left(\frac{1}{\sqrt{N_H T}}\right). \tag{3.8.203}$$

Substituting results (3.8.202) and (3.8.203) into equations (3.8.201) and (3.8.200) we get:

$$\frac{1}{N_H}\hat{\Lambda}_1' e_{m,t}^H \;\; = \;\; \hat{\mathcal{U}}_1' \frac{1}{N_H}\Lambda_1' e_{m,t}^H + O_p\left(\frac{1}{\sqrt{N_H T}}\right). \tag{3.8.204}$$

Substituting results (3.8.186), (3.8.199), and (3.8.204) into equation (3.8.185), and rearranging terms we get:

$$\hat{\mathcal{U}}_1\hat{g}_{m,t} - g_{m,t} \;\; = \;\; -\Sigma_{\Lambda,1}^{-1}\left(\frac{\Lambda_1'\Lambda_H}{N_H}\right)\left(\frac{1}{T}\bar{g}^{H\prime}\bar{g}^C\right) g_{m,t}^C + \Sigma_{\Lambda,1}^{-1}\frac{1}{N_H}\Lambda_1' e_{m,t}^H + o_p\left(\frac{1}{\sqrt{N_H}}\right) \tag{3.8.205}$$

Let us denote the last k^H columns of matrix $\Sigma_{\Lambda,1}$ as $\Sigma_{\Lambda,1}^{(\cdot\ s)}$. The term $\frac{\Lambda_1'\Lambda_H}{N_H}$ in equation (3.8.205) can be written as:

$$\begin{aligned}\frac{\Lambda_1'\Lambda_H}{N_H} &= \Sigma_{\Lambda,1}^{(\cdot\ s)} + \frac{1}{N_H}\sum_{i=1}^{N_H}\lambda_{1,i}\lambda_{H,i}' - \Sigma_{\Lambda,1}^{(\cdot\ s)} \\ &= \Sigma_{\Lambda,1}^{(\cdot\ s)} + O_p\left(\frac{1}{\sqrt{N_H}}\right),\end{aligned} \tag{3.8.206}$$

where the last equality follows from Assumption 2. Equation (3.8.206) implies:

$$\Sigma_{\Lambda,1}^{-1}\left(\frac{\Lambda_1'\Lambda_H}{N_H}\right) = \begin{bmatrix} 0_{(k^C\times k^H)} \\ I_{k^H}\end{bmatrix} + O_p\left(\frac{1}{\sqrt{N_H}}\right). \tag{3.8.207}$$

Substituting equation (3.8.207) into equation (3.8.205) we have:

$$\hat{\mathcal{U}}_1\hat{g}_{m,t} - g_{m,t} = -\begin{bmatrix} 0_{(k^C\times k^H)} \\ I_{k^H}\end{bmatrix}\left(\frac{1}{T}\bar{g}^{H\prime}\bar{g}^C\right)g_{m,t}^C + \Sigma_{\Lambda,1}^{-1}\frac{1}{N_H}\Lambda_1' e_{m,t}^H + o_p\left(\frac{1}{\sqrt{N_H}}\right) \tag{3.8.208}$$

Recalling the expression of $\hat{\mathcal{U}}_1$ from equation (3.8.139):

$$\hat{\mathcal{U}}_1 = \begin{bmatrix} \hat{\mathcal{H}}_c & 0 \\ 0 & \hat{\mathcal{H}}_{s,1}\end{bmatrix}, \tag{3.8.209}$$

from equation (3.8.208) we get the asymptotic expansions:

$$\hat{\mathcal{H}}_c\hat{g}_{m,t}^C - g_{m,t}^C \simeq \left[\Sigma_{\Lambda,1}^{-1}\frac{1}{N_H}\Lambda_1' e_{m,t}^H\right]^{(C)}, \tag{3.8.210}$$

$$\hat{\mathcal{H}}_{1,s}\hat{g}_{m,t}^H - g_{m,t}^H \simeq -\left(\frac{1}{T}\bar{g}^{H\prime}\bar{g}^C\right)g_{m,t}^C + \left[\Sigma_{\Lambda,1}^{-1}\frac{1}{N_H}\Lambda_1' e_{m,t}^H\right]^{(H)}, \tag{3.8.211}$$

where $\left[\Sigma_{\Lambda,1}^{-1}\frac{1}{N_H}\Lambda_1' e_{m,t}^H\right]^{(C)}$ and $\left[\Sigma_{\Lambda,1}^{-1}\frac{1}{N_H}\Lambda_1' e_{m,t}^H\right]^{(H)}$ denote the upper k^C rows, resp. the lower k^H rows, of vector $\Sigma_{\Lambda,1}^{-1}\frac{1}{N_H}\Lambda_1' e_{m,t}^H$. Since $\bar{g}^{C\prime}\bar{g}^C/T = I_{k^C} + o_p(1)$, we can rewrite equation (3.8.211) as:

$$\hat{\mathcal{H}}_{1,s}\hat{g}_{m,t}^H - (g_{m,t}^H - (\bar{g}^{H\prime}\bar{g}^C)(\bar{g}^{C\prime}\bar{g}^C)^{-1}g_{m,t}^C) \simeq [\Sigma_{\Lambda,1}^{-1}\frac{1}{N_H}\Lambda_1' e_{m,t}^H]^{(H)}. \tag{3.8.212}$$

From Assumption 8 we have:

$$\frac{1}{\sqrt{N_H}}\Lambda_1' e_{m,t}^H \xrightarrow{d} N(0, \Omega_{\Lambda,m}^*), \tag{3.8.213}$$

where

$$\Omega_{\Lambda,m}^* = \lim_{N_H\to\infty}\frac{1}{N_H}\sum_{i=1}^{N_H}\sum_{\ell=1}^{N_H}\lambda_{1,i}\lambda_{1,\ell}' Cov(e_{m,t}^{i,H}, e_{m,t}^{\ell,H}). \tag{3.8.214}$$

Equations (3.8.210) and (3.8.213) imply:

$$\sqrt{N_H}\left(\hat{\mathcal{H}}_c \hat{g}^C_{m,t} - g^C_{m,t}\right) \quad \xrightarrow{d} \quad N\left(0, \left[\Sigma^{-1}_{\Lambda,1}\Omega^*_{\Lambda,m}\Sigma^{-1}_{\Lambda,1}\right]^{(CC)}\right).$$

Similarly, equation (3.8.212) and (3.8.213) imply:

$$\sqrt{N_H}\left[\hat{\mathcal{H}}_{1,s}\hat{g}^H_{m,t} - (g^H_{m,t} - (\bar{g}^{H\prime}\bar{g}^C)(\bar{g}^{C\prime}\bar{g}^C)^{-1})g^C_{m,t}\right] \quad \xrightarrow{d} \quad N\left(0, \left[\Sigma^{-1}_{\Lambda,1}\Omega^*_{\Lambda,m}\Sigma^{-1}_{\Lambda,1}\right]^{(HH)}\right). \tag{3.8.215}$$

Q.E.D.

3.9 Appendix C: Additional results

3.9.1 Identification of the common factor space from variance-covariance matrix of stacked factors

Let us normalize the principal components such that $V(h_{j,t}) = I_{k_j}$, $j = 1, 2$.

Lemma 2. *Let $h_t = [h'_{1,t}\ \ h'_{2,t}]'$, be a random vector, such that $V_{11} = V(h_{1,t}) = I_{k_1}$, $V_{22} = V(h_{2,t}) = I_{k_2}$, $V_{12} = Cov(h_{1,t}, h_{2,t})$ and let $V(h_t)$ be the variance-covariance matrix of h_t:*

$$V(h_t) \quad = \quad \begin{bmatrix} I_{k_1} & V_{12} \\ V_{21} & I_{k_2} \end{bmatrix}.$$

Let $r = rank(V_{21})$, with $r \leq k_1$. Then, matrix $V(h_t)$ has $2r$ eigenvalues $1 \pm \rho_\ell$, $\ell = 1, ..., r$, with multiplicity 1, corresponding to the non-zero canonical correlations, $\rho_\ell \neq 0$, and the eigenvalue 1 with multiplicity $k_1 + k_2 - 2r$. The eigenvectors of $V(h_t)$ associated with the eigenvalues $1 \pm \rho_\ell$, $\ell = 1, ..., r$ are

$$v^{\pm}_\ell = \begin{bmatrix} w_{1,\ell} \\ \pm w_{2,\ell} \end{bmatrix}, \ \ell = 1, ..., r,$$

where $w_{1,\ell}$ (resp. $w_{2,\ell}$), are the normalized eigenvectors of $R = V_{12}V_{21}$ (resp. $R^* = V_{21}V_{12}$), associated with eigenvalues ρ^2_ℓ.

From Proposition 2 and Lemma 2 we get the next corollary.

Corollary 1. *i) The number k^c of common factors is equal to the multiplicity of the eigenvalue 2 of matrix $V(h_t)$. ii) Let W be the $(k_1 + k_2, k^c)$ matrix whose columns are the orthonormal eigenvectors associated with the k^c eigenvalues of $V(h_t)$ equal to 2. Then, $f^c_t = \dfrac{1}{\sqrt{2}} W' h_t$ (up to a one-to-one transformation).*

We note that Corollary 1 is analogous to Proposition 3.1 in Chen (2012).

Proof of Lemma 2: Let ρ_i, $i = 1, ..., k_1$, be the canonical correlations between $h_{1,t}$ and $h_{2,t}$. From Anderson (2003) and Magnus and Neudecker (2007), ρ^2_i corresponds to the i-th ordered eigenvalue of matrix $R =$

$V_{12}V_{21}$. Let $1+\mu$, say, be an eigenvalue of matrix $V(h_t)$, and $Z = [Z_1' \; Z_2']' \in \mathbb{R}^K$ be the associated (normalized) eigenvector. We have:

$$V(h_t)Z \;=\; (1+\mu)Z\,.$$

Rewriting matrix $V(h_t)$ as:

$$V(h_t) \;=\; I_K + \begin{bmatrix} 0 & V_{12} \\ V_{21} & 0 \end{bmatrix}, \tag{3.9.1}$$

we get:

$$\begin{bmatrix} 0 & V_{12} \\ V_{21} & 0 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \;=\; \mu \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}. \tag{3.9.2}$$

The last equation implies:

$$V_{12}Z_2 \;=\; \mu Z_1, \tag{3.9.3}$$
$$V_{21}Z_1 \;=\; \mu Z_2, \tag{3.9.4}$$

and:

$$V_{12}V_{21}Z_1 \;=\; \mu^2 Z_1, \tag{3.9.5}$$
$$V_{21}V_{12}Z_2 \;=\; \mu^2 Z_2. \tag{3.9.6}$$

If $Z_1 \neq 0$, then μ^2 is an eigenvalue of $V_{12}V_{21}$, i.e. a squared canonical correlation, and if $Z_2 \neq 0$, then μ^2 is an eigenvalue of $V_{21}V_{12}$. From assumption $rank(V_{21}) = r$, $r \leq k_1$, there are r canonical correlations different from zero: $\rho_1 \geq ... \geq \rho_r > 0$. Let $w_{1,\ell}$, $\ell = 1, ..., r$, be the associated eigenvectors of $R = V_{12}V_{21}$, and $w_{2,\ell}$, $\ell = 1, ..., r$ the corresponding eigenvectors of $R^* = V_{21}V_{12}$. Then, the scalars

$$\mu_{\ell,\pm} \;=\; \pm\rho_\ell, \qquad \ell = 1, ..., r,$$

and the vectors

$$v_\ell^\pm \;=\; \begin{bmatrix} w_{1,\ell} \\ \pm\dfrac{1}{\rho_\ell}V_{21}w_{1,\ell} \end{bmatrix} = \begin{bmatrix} w_{1,\ell} \\ \pm w_{2,\ell} \end{bmatrix} \tag{3.9.7}$$

solve equation (3.9.2). Here, we use $\dfrac{1}{\rho_\ell}V_{21}w_{1,\ell} = w_{2,\ell}$, from property (3.2.11). Thus, $1 \pm \rho_\ell$ are eigenvalues of $V(h_t)$ associated with eigenvectors $v_\ell^\pm$, with $\ell = 1, ..., r$.

Let us now consider the solutions of equation (3.9.2) with $\mu = 0$. We have:

$$V_{12}Z_2 \;=\; 0, \tag{3.9.8}$$
$$V_{21}Z_1 \;=\; 0. \tag{3.9.9}$$

From $rank(V_{12}) = r$, with $r \leq k_1 \leq k_2$, the null space of matrix V_{12} is $(k_2 - r)$-dimensional. Let the columns of the $(k_2, k_2 - r)$ full column rank matrix $\tilde{Z}_2$ span the $(k_2 - r)$-dimensional space of solutions of equation (3.9.8). Similarly, let the columns of the $(k_1, k_1 - r)$ full column rank matrix $\tilde{Z}_1$ span the $(k_1 - r)$-dimensional space of solutions of equation (3.9.9). Define the $(k_1 + k_2, 2(k_2 - r))$ matrix:

$$\tilde{Z}_0 \quad = \quad \left[\begin{array}{cc|cc} \tilde{Z}_1 & O_{k_1 \times (k_2 - k_1)} & -\tilde{Z}_1 & O_{k_1 \times (k_2 - k_1)} \\ & \tilde{Z}_2 & & \tilde{Z}_2 \end{array} \right].$$

Any column of this matrix is a solution of (3.9.2) with $\mu = 0$. Since matrices $\tilde{Z}_1$ and $\tilde{Z}_2$ are full column rank, the column rank of matrix Z_0 is $2(k_1 - r) + (k_2 - k_1) = k_1 + k_2 - 2r$. Therefore, there are $k_1 + k_2 - 2r$ linearly independent eigenvectors of $\begin{bmatrix} 0 & V_{12} \\ V_{21} & 0 \end{bmatrix}$ associated with the eigenvalue 0. These vectors are eigenvectors of $V(h_t)$ associated with the eigenvalue 1.

Q.E.D.

Proof of Corollary 1: From Lemma 2, $V(h_t)$ has eigenvalue 2 if, and only if, there is a canonical correlation equal to 1. Part *i)* follows from Proposition 2 *i)*. Moreover, from Proposition 2 and Lemma 2 the columns of matrix $W = \frac{1}{\sqrt{2}} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$ are orthonormal eigenvectors of $V(h_t)$ associated with eigenvalue 2, since $W'W = \frac{1}{2}(W_1'W_1 + W_2'W_2) = I_{k^c}$. Finally, $\frac{1}{\sqrt{2}} W'h_t = \frac{1}{2}(W_1'h_{1,t} + W_2'h_{2,t}) = \frac{1}{2}(S_1' + S_2')f_t^c$, from (3.8.12) and (3.8.13), which implies part *ii)*.

Q.E.D.

3.9.2 Estimator based on fixed point iteration

In this Appendix we consider the estimator for group factor models based on the Least Squares (LS) method suggested by Wang (2012). The estimator uses fixed point iteration to solve the first-order conditions (FOC). We discuss here some issues concerning the uniqueness of the fixed point.

The group factor model is:

$$Y_1 \quad = \quad F^c \Lambda_1^{c\,\prime} + F_1^s \Lambda_1^{s\,\prime} + \varepsilon_1, \tag{3.9.10}$$

$$Y_2 \quad = \quad F^c \Lambda_2^{c\,\prime} + F_2^s \Lambda_2^{s\,\prime} + \varepsilon_2. \tag{3.9.11}$$

The estimators of factor values and factor loadings are defined by minimizing the LS criterion

$$Q = \sum_{j=1}^{2} Tr[(Y_j - F^c \Lambda_j^{c\,\prime} - F_j^s \Lambda_j^{s\,\prime})'(Y_j - F^c \Lambda_j^{c\,\prime} + F_j^s \Lambda_j^{s\,\prime})], \tag{3.9.12}$$

w.r.t. arguments F^c, F_j^s, Λ_j^c, Λ_j^s, $j = 1, 2$, subject to the constraints:

$$F^{c\,\prime} F^c / T = I_{k^c}, \quad F_j^{s\,\prime} F_j^s / T = I_{k_j^s}, \quad F^{c\,\prime} F_j^s = 0, \quad j = 1, 2. \tag{3.9.13}$$

The first-order conditions (FOC) for this constrained minimization problem yield the following eigenvalue-eigenvector problems (see the proof at the end of the appendix):

- F_j^s is the $T \times k_j^s$ matrix of standardized eigenvectors of matrix

  $$M_{F^c}(Y_j Y_j'/N) M_{F^c} \tag{3.9.14}$$

  associated with the k_j^s largest eigenvalues, for $j = 1, 2$,

- F^c is the $T \times k^c$ matrix of standardized eigenvectors of matrix

$$M_{F^s}(Y_1Y_1'/N + Y_2Y_2'/N)M_{F^s} \tag{3.9.15}$$

associated with the k^c largest eigenvalues,

where $M_{F^c} = I_T - F^c(F^{c\,\prime}F^c)^{-1}F^{c\,\prime}$ and $M_{F^s} = I_T - F^s(F^{s\prime}F^s)^{-1}F^{s\,\prime}$, with $F^s = [F_1^s \;\; F_2^s]$. The eigenvectors are normalized such that $F^{c\,\prime}F^c/T = I_{k^c}$, $F_j^{s\,\prime}F_j^s/T = I_{k_j^s}$, for $j = 1, 2$, and satisfy automatically the identification restrictions $F^{c\,\prime}F_j^s = 0$, for $j = 1, 2$.

Wang (2012) suggests to solve the FOC by an iterative procedure. Given an estimate $\tilde{F}^c$, the estimate $\hat{F}_j^s$ is computed by the spectral decomposition of the matrix in (3.9.14) with $F^c = \tilde{F}^c$, for $j = 1, 2$. The estimate $\hat{F}^s = [\hat{F}_1^s \; \hat{F}_2^s]$ is used to compute the matrix in (3.9.15), whose spectral decomposition yields a new estimate $\hat{F}^c$. This procedure defines the (stochastic) mapping $\tilde{F}^c \to \Psi(\hat{F}^c)$.

Let us now investigate the properties of the mapping Ψ. For this purpose we consider the setting with scalar factors, i.e. $k^c = k_1^s = k_2^s = 1$, and the next assumption.

Assumption A. 9. *a) The errors are $\varepsilon_1 = \varepsilon_2 = 0$, and the true factor values are such that $F^{c\,\prime}F^c/T = F_j^{s\,\prime}F_j^s/T = 1$, $F^{c\,\prime}F_j^s = 0$, for $j = 1, 2$. b) $F_j^{s\,\prime}F_j^s = 0$, $j = 1, 2$.*

Assumption 9 defines a specific realization of the errors and the factors. In part a), we shut down the errors to mimic the large N, T, setting where the impact of the idiosyncratic shocks vanishes. The factor values match in sample the theoretical normalization restrictions. For expository purpose, we assume that the group-specific factors are orthogonal, and part b) matches this condition in sample.

Proposition 3. *Under Assumption 9, any vector $\tilde{F}^c$, that is a linear combination of F^c, F_1^s, F_2^s (true factor values), is a fixed point of the mapping Ψ (up to a sign change).*

Proof of Proposition 3: Define the $T \times 3$ matrix $H = [F^c \;\; F_1^s \;\; F_2^s]$. Under Assumption 9 we have $H'H/T = I_3$, and the data can be written as $Y_1 = H[\Lambda_1^c \;\; \Lambda_1^s \;\; 0]'$ and $Y_2 = H[\Lambda_2^c \;\; 0 \;\; \Lambda_2^s]'$. Then, we get:

$$Y_1Y_1'/N = H \begin{bmatrix} \Lambda_1^{c\,\prime}\Lambda_1^c/N & \Lambda_1^{c\,\prime}\Lambda_1^s/N & 0 \\ \Lambda_1^{s\,\prime}\Lambda_1^c/N & \Lambda_1^{s\,\prime}\Lambda_1^s/N & 0 \\ 0 & 0 & 0 \end{bmatrix} H' \equiv H\Pi_1H'.$$

Similarly, we have $Y_2Y_2'/N = H\Pi_2H'$ for a suitable 3×3 matrix Π_2, and $Y_1Y_1'/N + Y_2Y_2'/N = H\Pi H'$ with $\Pi = \Pi_1 + \Pi_2$.

Now, let

$$\tilde{F}^c = F^c\beta_1 + F_1^s\beta_2 + F_2^s\beta_3 = H\beta,$$

where $\beta = (\beta_1, \beta_2, \beta_3)'$ is such that $\beta'\beta = 1$. Then:

$$\begin{aligned} M_{\tilde{F}^c} &= I_T - \frac{1}{T}\tilde{F}^c\tilde{F}^{c\,\prime} \\ &= I_T - \frac{1}{T}H\beta\beta'H' = M_H + \frac{1}{T}HM_\beta H', \end{aligned}$$

where $M_\beta = I_3 - \beta(\beta'\beta)^{-1}\beta' = I_3 - \beta\beta'$. Thus, the matrix in (3.9.14) corresponding to $\tilde{F}^c$ can be written as:

$$
\begin{aligned}
M_{\tilde{F}^c}(Y_j Y_j'/N) M_{\tilde{F}^c} &= (M_H + \frac{1}{T} H M_\beta H') H \Pi_j H' (M_H + \frac{1}{T} H M_\beta H') \\
&= H M_\beta \Pi_j M_\beta H'.
\end{aligned}
$$

The eigenvector associated with the largest eigenvalue of matrix $H M_\beta \Pi_j M_\beta H'$ is in the column space of H:

$$\hat{F}_j^s = H\alpha_j,$$

where the 3×1 vector α_j is the normalized eigenvector of matrix $M_\beta \Pi_j M_\beta$ associated with the largest eigenvalue, $j = 1, 2$. In particular, α_j is orthogonal to β, $j = 1, 2$. The vectors α_1 and α_2 are not collinear (verify this).

Let $\hat{F}^s = [\hat{F}_1^s \; \hat{F}_2^s] = H\alpha$, where $\alpha = [\alpha_1 \; \alpha_2]$. The matrix in (3.9.15) corresponding to $\hat{F}^s$ is

$$M_{\hat{F}^s}(Y_1 Y_1'/N + Y_2 Y_2'/N) M_{\hat{F}^s} = H M_\alpha \Pi M_\alpha H'.$$

The eigenvector $\hat{F}^c$ of this matrix associated with the largest eigenvalue is $\hat{F}^c = H\gamma$, where γ is the eigenvector of matrix $M_\alpha \Pi M_\alpha$ associated with the largest eigenvalue. This implies that γ is orthogonal to α_1 and α_2, and thus is collinear to β. By normalization we have either $\gamma = \beta$, or $\gamma = -\beta$. Thus, either $\hat{F}^c = \tilde{F}^c$, or $\hat{F}^c = -\tilde{F}^c$. Q.E.D.

Proof of the FOC for the constrained minimization (3.9.12)-(3.9.13): The Lagrange multipliers for the identification restrictions (3.9.13) are zero. The FOC for the factor loadings under the constraints yield:

$$
\begin{aligned}
\Lambda_j^c &= Y_j' F^c (F^{c\,\prime} F^c)^{-1}, \\
\Lambda_j^s &= Y_j' F_j^s (F_j^{s\,\prime} F_j^s)^{-1}, \quad j = 1, 2.
\end{aligned}
$$

From these equations, the residuals are

$$Y_j - F^c \Lambda_j^{c\,\prime} - F_j^s \Lambda_j^{s\,\prime} = (I_T - P_{F^c} - P_{F_j^s}) Y_j, \quad j = 1, 2,$$

where $P_{F^c} = F^c (F^{c\,\prime} F^c)^{-1} F^{c\,\prime} = I_T - M_{F^c}$ and $P_{F_j^s} = F_j^s (F_j^{s\,\prime} F_j^s)^{-1} F_j^{s\,\prime} = I_T - M_{F_j^s}$. From the orthogonality $F^{c\,\prime} F_j^s = 0$ in (3.9.13), matrices M_{F^c} and $M_{F_j^s}$ commute, and matrices

$$I_T - P_{F^c} - P_{F_j^s} = M_{F_j^s} M_{F^c} = M_{F^c} M_{F_j^s}, \quad j = 1, 2,$$

are idempotent. Thus, the concentrated LS criterion becomes:

$$Q = \sum_{j=1}^{2} Tr[Y_j' M_{F^c} M_{F_j^s} Y_j]. \tag{3.9.16}$$

From the constraints (3.9.13) and the commutative property of the trace, the concentrated LS criterion can be rewritten as:

$$
\begin{aligned}
Q &= \sum_{j=1}^{2} Tr[M_{F_j^s} M_{F^c} Y_j Y_j' M_{F^c}] \\
&= \sum_{j=1}^{2} Tr[M_{F^c} Y_j Y_j' M_{F^c}] - \sum_{j=1}^{2} \frac{1}{T} Tr[F_j^{s\,\prime} M_{F^c} Y_j Y_j' M_{F^c} F_j^s].
\end{aligned}
$$

For $j = 1, 2$, the minimization of this concentrated criterion w.r.t. F_j^s is equivalent to the maximization of $Tr[F_j^s{}' M_{F^c} Y_j Y_j' M_{F^c} F_j^s]$. Under the constraint $F^{s\prime} F_j^s / T = I_{k_j^s}$, this problem is solved by the matrix of normalized eigenvectors of matrix $M_{F^c} Y_j Y_j' M_{F^c}$ associated with the k_j^s largest eigenvalues.

Similarly, from the constraints (3.9.13) and the commutative property of the trace, the concentrated LS criterion (3.9.16) can be rewritten as:

$$
\begin{aligned}
Q &= \sum_{j=1}^{2} Tr[M_{F^c} M_{F_j^s} Y_j Y_j' M_{F_j^s}] \\
&= \sum_{j=1}^{2} Tr[M_{F_j^s} Y_j Y_j' M_{F_j^s}] - \sum_{j=1}^{2} \frac{1}{T} Tr[F^c{}' M_{F_j^s} Y_j Y_j' M_{F_j^s} F^c] \\
&= \sum_{j=1}^{2} Tr[M_{F_j^s} Y_j Y_j' M_{F_j^s}] - \frac{1}{T} Tr[F^c{}' M_{F^s} (\sum_{j=1}^{2} Y_j Y_j') M_{F^s} F^c].
\end{aligned}
$$

The minimization of this concentrated criterion w.r.t. F^c is equivalent to the maximization of $Tr[F^c{}' M_{F^s} (\sum_{j=1}^{2} Y_j Y_j') M_{F^s} F^c]$. Under the constraint $F^c{}' F^c / T = I_{k^c}$, this problem is solved by the matrix of normalized eigenvectors of matrix $M_{F^s} (\sum_{j=1}^{2} Y_j Y_j') M_{F^s}$ associated with the k^c largest eigenvalues. Q.E.D.

3.9.3 Dataset description

High Frequency dataset: Industrial Production sectors

Our high frequency dataset includes the same 117 sectors constituting the aggregate Industrial Production index considered by Foerster, Sarte, and Watson (2011) for the years 1977-2011. This sample period coincides with the maximum number of years for which the data for the 42 non-Industrial Production sectors of our low frequency panel were available and therefore - differently from Foerster, Sarte, and Watson (2011) - we do not consider the entire time series available for IP data starting in 1972. We download the monthly level of the 117 IP indices from Board of Governors of the Federal Reserve System (FED)[34]. From these raw data, which are indices of real output, we compute the corresponding quarterly growth rates.

The 117 sectors roughly corresponds to a four-digit industry in the North American Industry Classification System (NAICS) for year 2002, which is the finest level of disaggregation allowing to match the sectors to the input-output and capital use tables used to calibrate the structural models. The IP Sectors are classified by the FED according in the following subsectors: *Manufacturing*, *Mining* and *Utilities*. *Manufacturing* comprises those industries included in NAICS definition of manufacturing plus the logging and newspaper, periodical, book, and directory publishing industries that have traditionally been considered manufacturing, and is divided in *Durable*, *Nondurable* and *Other manufacturing*. *Durable manufacturing* includes three-digit NAICS codes 321, 327, 331-337, and 339. *Nondurable manufacturing* includes three-digit NAICS codes 311-316 and 322-326. *Other manufacturing* includes NAICS codes 1133 and 5111. *Mining* includes three-digit NAICS codes 211-213. *Utilities* include electric utilities and natural gas distribution, corresponding to NAICS codes 2211 and 2212.[35] We refer to Foerster, Sarte, and Watson (2011), and especially to their Appendix A, for a list of the names of the 117 sectors, and the methodology used to approximate the missing data for some sectors.

Low Frequency dataset: non-Industrial Production sectors

The US Bureau of Economic Analysis (BEA) publishes at yearly frequency the growth rates for the real Gross Domestic Product and real Gross Output for all the sectors of the US economy, not only for the sectors included in the IP index. We use the *Release Date November 13, 2012* dataset as downloaded for the BEA website[36]. The period 1977-2011 coincides with the maximum number of years for which the data for the 42 non-Industrial Production sectors in our low frequency panel were available at the date of download of the dataset. [37] Our original BEA dataset includes the time series for the output growth rates of 65 mutually exclusive sectors constituting the entire US economy, for the sample period 1977-2011. These sectors are aggregates of either 2 or 3 digits 2002 NAICS codes. Out of these 65 sectors, 19 are *Manufacturing* sectors (NAICS 2002 codes: 31-33), 3 are *Mining* sectors (NAICS 2002 codes: 211-213) and one is *Utilities* (NAICS 2002 code: 22). These 23 sectors are all included in the IP dataset, and therefore are excluded from our LF panel to avoid duplication of sectors in the two panels. The IP sectors *Logging*, *Newspaper Publishers* and *Periodical, Book, and Other Publishers* (NAICS 1133, 5111, 5112) are subsectors of the 2 BEA sectors

[34] See `http://www.federalreserve.gov/releases/G17/default.htm`.

[35] For a detailed description of the IP constituents see `http://www.federalreserve.gov/releases/g17/About.htm`.

[36] See `http://www.bea.gov/industry/gdpbyind_data.htm`.

[37] Time series for 22 aggregates of our 42 sectors are also available form the BEA website since 1947, and time series for a more disaggregated version of our 42 indices, but only for Gross Output, is available only from 1997.

Publishing industries (includes software) and *Forestry, fishing, and related activities*. We keep these 2 sectors in the the low frequency panel. Therefore our non-IP low frequency panel includes the 42 sectors listed in Table 3.11 together with the corresponding NAICS 2002 codes.

Table 3.11: List of non-Industrial Production sectors. (Source: BEA)

Sector	**NAICS 2002 codes**
Farms	111, 112
Forestry, fishing, and related activities	113, 114, 115
Construction	23
Wholesale trade	42
Retail trade	44, 45
Air transportation	481
Rail transportation	482
Water transportation	483
Truck transportation	484
Transit and ground passenger transportation	485
Pipeline transportation	486
Other transportation and support activities	487, 488, 492
Warehousing and storage	493
Publishing industries (includes software)	511, 516
Motion picture and sound recording industries	512
Broadcasting and telecommunications	515, 517
Information and data processing services	518, 519
Federal Reserve banks, credit intermediation, and related activities	521, 522
Securities, commodity contracts, and investments	523
Insurance carriers and related activities	524
Funds, trusts, and other financial vehicles	525
Real estate	531
Rental and leasing services and lessors of intangible assets	532, 533
Legal services	5411
Computer systems design and related services	5415
Miscellaneous professional, scientific, and technical services	5412-5414, 5416-5419
Management of companies and enterprises	55
Administrative and support services	561
Waste management and remediation services	562
Educational services	61
Ambulatory health care services	621
Hospitals and nursing and residential care facilities	622, 623
Social assistance	624
Performing arts, spectator sports, museums, and related activities	711, 712
Amusements, gambling, and recreation industries	713
Accommodation	721
Food services and drinking places	722
Other services, except government	81
Federal Government - General government	-
Federal Government - Government enterprises	- (includes 491)
State and Local Government - General government	-
State and Local Government - Government enterprises	-

In Table 3.12 we report names of the sectors corresponding to the aggregated version of the yearly indices used in Tables 3.5, 3.6, 3.7, and in the analogous Tables in the subsample analysis, together with their corresponding first or first two NAICS 2002 codes. The yearly growth rates of these real aggregated indices are downloaded directly by the BEA website.

Table 3.12: List of aggregates of non-Industrial Production sectors. (Source: BEA)

Sector	**NAICS 2002 codes**
GDP / GO (all sectors)	all sectors included
Manufacturing	31, 32, 33
Agriculture, forestry, fishing, and hunting	11
Construction	23
Wholesale trade	42
Retail trade	44, 45
Transportation and warehousing	48, 49 (except 491)
Information	51
Finance, insurance, real estate, rental, and leasing	52, 53
Professional and business services	54
Educational services, health care, and social assistance	6
Arts, entertainment, recreation, accommodation, and food services	7
Government	- (includes 491)

3.9.4 Filtration of structural productivity shocks from mixed frequency output growth data.

In this section we propose an adaptation of the methodology proposed by Foerster, Sarte, and Watson (2011) to estimate mixed-frequency technological shocks from the time series of quarterly Industrial Production growth rates, and yearly non-IP Gross Output growth rates for all the sectors described in the previous section.

Foerster, Sarte, and Watson (2011) consider a multi-industry real business cycle model, with an economy characterized by N_H distinct sectors. As they consider only IP sectors, they work with quarterly observations of measures of output growth. The generic sector j, with $j = 1, ..., N_H$ produces at quarter t, a certain quantity Y_{jt} of good j, using labor, capital and intermediate goods. The novelty in the model consists in the fact that both capital goods and intermediate goods used by a certain sector are other sectors' output, generating interconnections among the outputs of all sectors. One of the assumptions in the model is that the N-dimensional vector A_t collecting the productivity indices of the N sectors follows a random walk process:

$$\ln(A_t) = \ln(A_{t-1}) + \varepsilon_t,$$

where ε_t is the N_H-dimensional vector of productivity (or technological) innovations. The authors specify the utility function of a representative agent defined on the consumption of the N_H goods and labor allocated to each sector, and solve the utility maximization problem of the representative agent, after specifying the goods and investments production technologies, the low of motion of capital stock and the resource constraints. In their Appendix B they show that "the deterministic steady state of the model continues to be analytically tractable" even with a large number of sectors. The log-linearization of the equilibrium equations, obtained using the results of King and Watson (2002), produce an approximation of the model's first order conditions and resource constraints around the steady state, yielding a vector ARMA(1,1) model for quarterly sectoral output growth N_H-dimensional vector $X_t = [\Delta \ln Y_{1t}, ..., \Delta \ln Y_{Nt}]'$:

$$(I - \Phi L)X_t \quad = \quad (\Pi_0 + \Pi_1 L)\varepsilon_t, \tag{3.9.17}$$

where L is the lag operator, and Φ, Π_0, Π_1 are $N \times N$ matrices that depend only on the model parameters, the input-output matrix Γ and the sectoral investment matrix Θ, also named "capital use" matrix. The input-output and capital use matrices are constructed using the BEA's 1997 "use table" and "capital flow table", which are available for the $N_H = 117$ IP sectors studied by Foerster, Sarte, and Watson (2011). As the time series of X_t is observable, the VARMA(1,1) model (3.9.17) can easily be used to compute $\hat{\varepsilon}_t$ as a filtered version of X_t:

$$\hat{\varepsilon}_t \quad = \quad (\Pi_0 + \Pi_1 L)^{-1}(I - \Phi L)X_t. \tag{3.9.18}$$

Let us now consider our mixed-frequency data settings, where $X^q_{H,t}$ is the N_H-dimensional vector of growth rates of Industrial Production, for the $N_H = 117$ IP sectors of our empirical analysis, for the generic quarter $q = 1, 2, 3, 4$ in year t. Let also $X^Y_{H,t}$ be the vector of the yearly growth rates for the same 117 IP sectors, for the generic year t. Moreover let $X_{L,t}$ be the N_L-dimensional vector of growth rates of the $N_L = 38$ non-IP sectors also considered in our empirical analysis. The high frequency technological shocks $\hat{\varepsilon}^{X,q}_t$ are obtained as a filtered version of the time series of stacked vectors $\left[X^q_{H,t}, X_{L,t}\right]'$:

$$\left[\begin{array}{c} \hat{\varepsilon}^{X,q}_t \\ \tilde{\varepsilon}^{Y}_t \end{array}\right] = (\tilde{\Pi}_0 + \tilde{\Pi}_1 L)^{-1}(I - \tilde{\Phi} L) \left[\begin{array}{c} 4 \cdot X^q_{H,t} \\ X_{L,t} \end{array}\right], \tag{3.9.19}$$

for each $q = 1, 2, 3, 4$. The idea is to consider the LF time series of $X^q_{H,t}$ as an approximation for the yearly growth rate of Industrial Production. Note that as there are 4 quarters in a year, we have 4 possible proxies for the yearly growth rates of IP. This allows us to have an approximation of $\varepsilon^{X,q}_t$ the high frequency productivity innovation for each quarter quarter q. Finally, the low frequency technological shocks $\hat{\varepsilon}^Y_t$ are obtained as a filtered version of the time series of stacked vectors $\left[X^Y_{H,t}, X_{L,t}\right]'$:

$$\left[\begin{array}{c} \tilde{\varepsilon}^{X,q}_t \\ \hat{\varepsilon}^{Y}_t \end{array}\right] = (\tilde{\Pi}_0 + \tilde{\Pi}_1 L)^{-1}(I - \tilde{\Phi} L) \left[\begin{array}{c} X^Y_{H,t} \\ X_{L,t} \end{array}\right], \tag{3.9.20}$$

i.e. we use the actual yearly growth rates of IP sectors as an input of the filtering algorithm. When we filter productivity innovations as in equation (3.9.19) and (3.9.20), we are implicitly assuming the same structural model for our economy as the one described by Foerster, Sarte, and Watson (2011). Moreover, filtering from equations (3.9.19) and (3.9.20) is feasible because matrices $\tilde{\Pi}_0$, $\tilde{\Pi}_1$, and $\tilde{\Phi}$ are functions of $\tilde{\Gamma}$ and $\tilde{\Theta}$, which are the input-output and capital use matrices for both IP and non-IP sectors, also available for the year 1997 from the BEA. In particular, we note that the 117×117 matrices Γ and Θ are the upper-left blocks of the 155×155 matrices $\tilde{\Gamma}$ and $\tilde{\Theta}$, respectively.

3.9.5 Additional Empirical results

In this section we report additional empirical results.

Table 3.13: Estimated number of factors: results for IC_{p1} information criterion

	X_{HF}	X^H	X^L	$[\ X^H\ \ X^L\ ]$
IP data: 1977.Q1-2011.Q. Non-IP data: Gross Domestic Product, 1977-2011				
IC_{p1}	2	2	1	2
IP data: 1988.Q1-2011.Q4. Non-IP data: Gross Output, 1988-2011				
IC_{p1}	1	2	15	2

The number of latent pervasive factors selected by the IC_{p1} information criteria is reported for different subpanels and different sample periods. Subpanels X_{HF} and X^H correspond to IP data sampled at quarterly and yearly frequency, respectively. Panels X^L and $[\ X^H\ \ X^L\ ]$ correspond to non-IP data, and the stacked panels of IP and non-IP data, respectively. We use $k_{max} = 15$ as maximum number of factors when computing IC_{p1}.

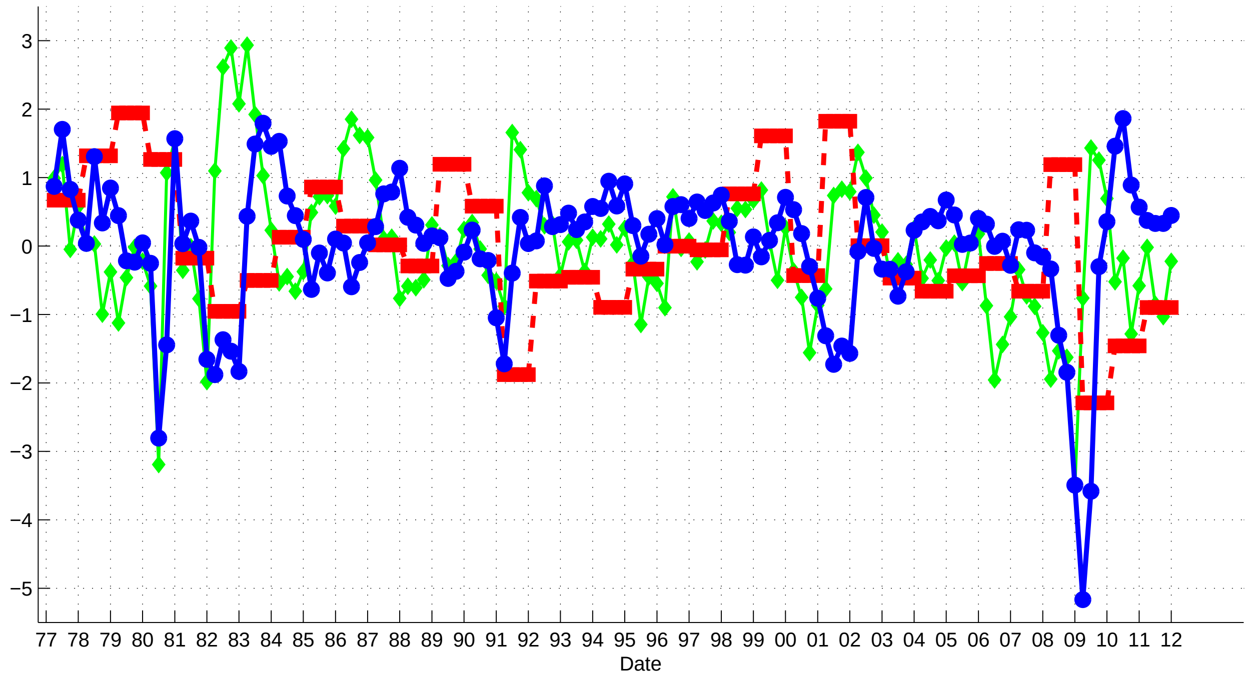
Table 3.14: Eigenvalues of the variance-covariance matrix of the stacked PC's

1^{st} eig.	2^{nd} eig.	3^{rd} eig.	4^{th} eig.
IP data: 1977.Q1-2011.Q4. Non-IP data: Gross Domestic Product, 1977-2011			
1.84	1.06	0.93	0.15
IP data: 1988.Q1-2011.Q4. Non-IP data: Gross Output, 1988-2011			
1.81	1.11	0.80	0.19

In this table we report the eigenvalues of the sample variance-covariance matrix of the stacked PC's estimated in each subpanel of IP (X^H) and non-IP data (X^L). We extract the first 2 PC's in each subgroup, and compute the variance-covariance matrix of these 4 stacked PC's.

We find an eigenvalue close to two for both datasets, which is consistent with the presence of one common factor in each of the two mixed frequency datasets considered.

Figure 3.4: Trajectories of the estimated common and LF-specific factors.



The Figure displays the time series of estimated values of the common factor (blue circles), the LF-specific factor (red squares) and the HF-specific factor (green rotated squares). For each year we represent the LF factor as 4 squares corresponding to the 4 quarters, assuming the same value. The factors are estimated from the panel of real output growth rates of 42 GDP sectors and 117 industrial production indices, using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of the factor model is 1977.Q1-2011.Q4.

Table 3.15: Correlation matrix of the estimated factors, computed at low frequency.

	$\hat{\hat{g}}_t^C$	$\hat{\hat{g}}_t^H$	$\hat{\hat{g}}_t^L$
$\hat{\hat{g}}_t^C$	1.0000		
$\hat{\hat{g}}_t^H$	-0.0425	1.0000	
$\hat{\hat{g}}_t^L$	-0.0486	-0.1473	1.0000

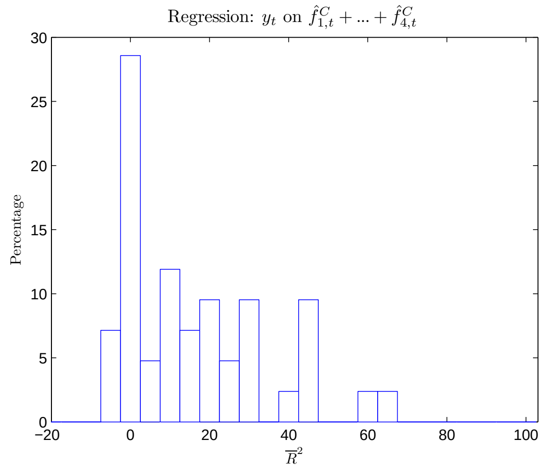
In the table we display the correlation matrix of the stacked vector of estimated factors $(\hat{\hat{g}}_t^C, \hat{\hat{g}}_t^H, \hat{\hat{g}}_t^L)$. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1977.Q1-2011.Q4.

Table 3.16: Correlation matrix of the estimated common, HF-specific and LF-specific factors: quarterly observations.

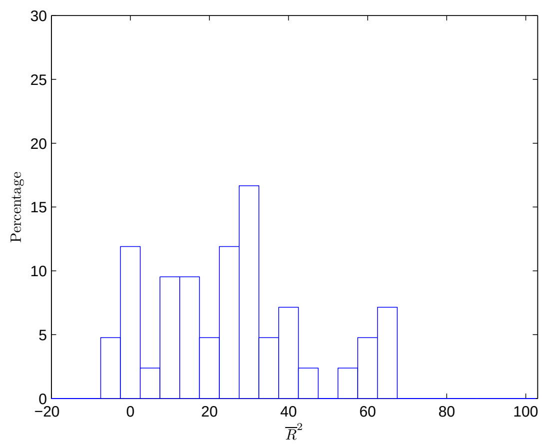
	$\hat{g}^C_{1,t}$	$\hat{g}^C_{2,t}$	$\hat{g}^C_{3,t}$	$\hat{g}^C_{4,t}$	$\hat{g}^H_{1,t}$	$\hat{g}^H_{2,t}$	$\hat{g}^H_{3,t}$	$\hat{g}^H_{4,t}$	$\hat{\hat{g}}^L_t$
$\hat{g}^C_{1,t}$	1.00	0.75	0.41	0.20	0.12	-0.38	-0.53	-0.29	0.28
$\hat{g}^C_{2,t}$	0.75	1.00	0.77	0.30	0.35	0.11	-0.41	-0.26	-0.05
$\hat{g}^C_{3,t}$	0.41	0.77	1.00	0.66	0.40	0.23	-0.19	-0.03	-0.35
$\hat{g}^C_{4,t}$	0.20	0.30	0.66	1.00	0.34	-0.03	0.11	0.43	-0.16
$\hat{g}^H_{1,t}$	0.12	0.35	0.40	0.34	1.00	0.56	0.44	0.41	-0.01
$\hat{g}^H_{2,t}$	-0.38	0.11	0.23	-0.03	0.56	1.00	0.65	0.47	-0.27
$\hat{g}^H_{3,t}$	-0.53	-0.41	-0.19	0.11	0.44	0.65	1.00	0.79	-0.12
$\hat{g}^H_{4,t}$	-0.29	-0.26	-0.03	0.43	0.41	0.47	0.79	1.00	-0.06
$\hat{\hat{g}}^L_t$	0.28	-0.05	-0.35	-0.16	-0.01	-0.27	-0.12	-0.06	1.00

In the table we display the correlation matrix of the stacked vector of estimated factors $(\hat{g}^C_{1,t}, \hat{g}^C_{2,t}, \hat{g}^C_{3,t}, \hat{g}^C_{4,t}, \hat{g}^H_{1,t}, \hat{g}^H_{2,t}, \hat{g}^H_{3,t}, \hat{g}^H_{4,t}, \hat{\hat{g}}^L_t)$. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1977-2011.

Figure 3.5: Adj. R^2 of the regression of yearly sectoral GDP growth rates on estimated factors.



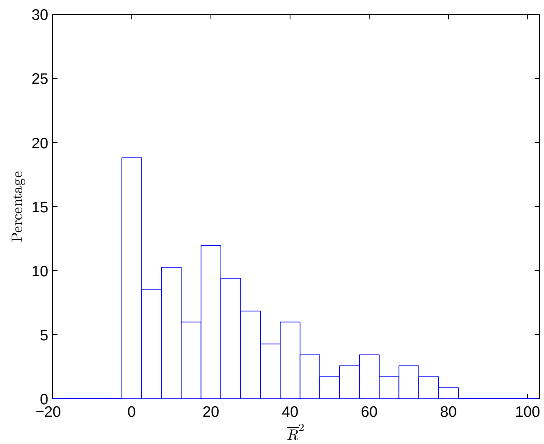
(a) Adjusted R^2 of the regression of yearly sectoral GDP growth on the common factor.



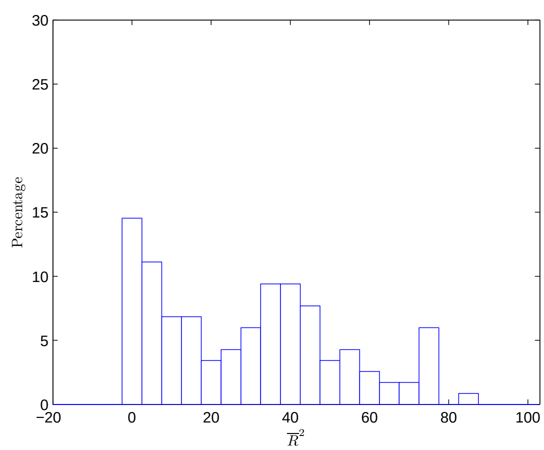
(b) Adjusted R^2 of the regression of yearly sectoral GDP growth on the common and LF-specific factors.

In Panel (a) we show the histogram of the adjusted R^2, denoted $\bar{R}^2$, of the regressions of the yearly growth rates of sectoral GDP indices on the estimated common factor. In Panel (b) we show the histogram of the adjusted R^2 of the regressions of the same growth rates on the estimated common and LF-specific factors. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1977-2011.

Figure 3.6: Adj. R^2 of the regression of sectoral IP growth rates on estimated factors.



(a) Adjusted R^2 of the regression of quarterly industrial production growth on the common factor.



(b) Adjusted R^2 of the regression of quarterly industrial production growth on the common and HF-specific factors.

In Panel (a) we show the histogram of the adjusted R^2, denoted $\bar{R}^2$, of the regressions of the quarterly growth rates of the industrial production indices on the estimated common factor. In Panel (b) we show the histogram of the adjusted R^2 of the regressions of the same growth rates on the estimated common and HF-specific factors. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1977-2011.

Table 3.17: Adjusted R^2 of the regression of yearly sectoral GDP growth on the common factor.

Sector	$\bar{R}^2$
Truck transportation	63.10
Accommodation	62.43
Construction	44.05
Other transportation and support activities	43.31
Administrative and support services	42.69
Other services, except government	42.53
Warehousing and storage	40.95
Air transportation	31.58
Retail trade	30.70
Amusements, gambling, and recreation industries	29.17
Government enterprises (federal)	28.91
Rail transportation	24.84
Performing arts, spectator sports, museums, and related activities	22.63
Publishing industries (includes software)	22.02
Computer systems design and related services	21.24
Food services and drinking places	20.59
Wholesale trade	20.35
Miscellaneous professional, scientific, and technical services	16.98
Waste management and remediation services	14.79
Social assistance	12.91
General government (federal)	11.97
Government enterprises (state & local)	11.10
Real estate	10.39
Legal services	10.19
federal Reserve banks, credit intermediation, and related activities	9.74
Educational services	3.97
Rental and leasing services and lessors of intangible assets	2.81
Broadcasting and telecommunications	1.24
Ambulatory health care services	1.01
Farms	0.93
Hospitals and nursing and residential care facilities	0.64
Management of companies and enterprises	-0.45
Funds, trusts, and other financial vehicles	-1.23
Motion picture and sound recording industries	-1.68
Pipeline transportation	-1.74
Information and data processing services	-1.84
Transit and ground passenger transportation	-2.05
General government (state & local)	-2.12
Forestry, fishing, and related activities	-2.33
Water transportation	-2.94
Securities, commodity contracts, and investments	-2.99
Insurance carriers and related activities	-3.03

In the table we display the adjusted R^2, denoted $\bar{R}^2$, for the time series regressions of each of the of 42 GDP sectors on the estimated common factor. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1977.Q1-2011.Q4. The regressions in this table are restricted MIDAS regressions.

Table 3.18: Adjusted R^2 of the regression of yearly sectoral GDP growth on the common and LF-specific factors.

Sector	$\bar{R}^2$
Miscellaneous professional, scientific, and technical services	66.67
Administrative and support services	62.63
Truck transportation	62.51
Accommodation	61.48
Construction	59.75
Warehousing and storage	52.53
Government enterprises (STATES AND LOCAL)	45.78
Other services, except government	41.75
Other transportation and support activities	41.71
Government enterprises (federal)	37.78
Legal services	34.51
Social assistance	32.82
Rental and leasing services and lessors of intangible assets	32.32
Wholesale trade	30.83
Performing arts, spectator sports, museums, and related activities	30.49
federal Reserve banks, credit intermediation, and related activities	30.05
Air transportation	29.81
Retail trade	28.56
Real estate	28.53
Computer systems design and related services	27.07
Amusements, gambling, and recreation industries	27.02
Publishing industries (includes software)	23.85
Rail transportation	23.68
General government (STATES AND LOCAL)	22.78
Food services and drinking places	21.67
Motion picture and sound recording industries	21.10
Hospitals and nursing and residential care facilities	17.47
Broadcasting and telecommunications	14.46
Waste management and remediation services	14.24
Pipeline transportation	14.13
General government (federal)	11.11
Transit and ground passenger transportation	9.18
Ambulatory health care services	7.76
Management of companies and enterprises	7.52
Funds, trusts, and other financial vehicles	6.15
Information and data processing services	1.96
Educational services	1.35
Insurance carriers and related activities	0.36
Water transportation	-0.64
Farms	-1.87
Forestry, fishing, and related activities	-5.31
Securities, commodity contracts, and investments	-5.99

In the table we display the adjusted R^2, denoted $\bar{R}^2$, for the time series regressions of each of the of 42 GDP sectors on the estimated common and LF-specific factors. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1977.Q1-2011.Q4. The regressions in this table are restricted MIDAS regressions.

Table 3.19: Change in adjusted R^2 of the regression of yearly sectoral GDP growth on the common factor and the LF-specific factors vs. the regression on the common factor only.

Sector	change in $\bar{R}^2$
Miscellaneous professional, scientific, and technical services	49.69
Government enterprises (STATES AND LOCAL)	34.69
Rental and leasing services and lessors of intangible assets	29.52
General government (STATES AND LOCAL)	24.90
Legal services	24.32
Motion picture and sound recording industries	22.77
federal Reserve banks, credit intermediation, and related activities	20.31
Administrative and support services	19.95
Social assistance	19.91
Real estate	18.14
Hospitals and nursing and residential care facilities	16.84
Pipeline transportation	15.87
Construction	15.70
Broadcasting and telecommunications	13.23
Warehousing and storage	11.58
Transit and ground passenger transportation	11.23
Wholesale trade	10.48
Government enterprises (federal)	8.87
Management of companies and enterprises	7.98
Performing arts, spectator sports, museums, and related activities	7.87
Funds, trusts, and other financial vehicles	7.39
Ambulatory health care services	6.76
Computer systems design and related services	5.83
Information and data processing services	3.80
Insurance carriers and related activities	3.39
Water transportation	2.30
Publishing industries (includes software)	1.83
Food services and drinking places	1.07
Waste management and remediation services	-0.54
Truck transportation	-0.60
Other services, except government	-0.78
General government (federal)	-0.86
Accommodation	-0.96
Rail transportation	-1.16
Other transportation and support activities	-1.59
Air transportation	-1.77
Retail trade	-2.15
Amusements, gambling, and recreation industries	-2.15
Educational services	-2.62
Farms	-2.80
Forestry, fishing, and related activities	-2.98
Securities, commodity contracts, and investments	-3.00

In the table we display the difference in the adjusted R^2 ($\bar{R}^2$) from the regressions of each industrial production index growth on the common and LF-specific estimated factors and on the HF factor only. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both factor model and regressions is 1977.Q1-2011.Q4. The regressions in this table are restricted MIDAS regressions.

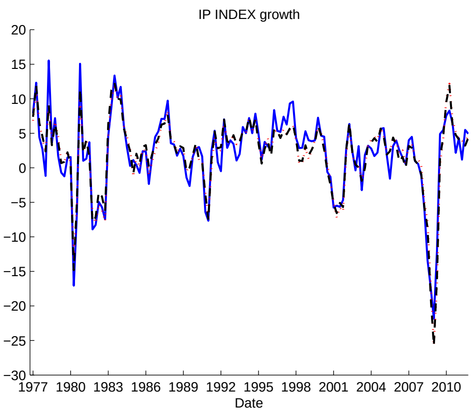
Table 3.20: Adj. R^2 of selected GDP and IP indices growth rates on the estimated factors

Panel a. *Yearly observations, 1977-2011*

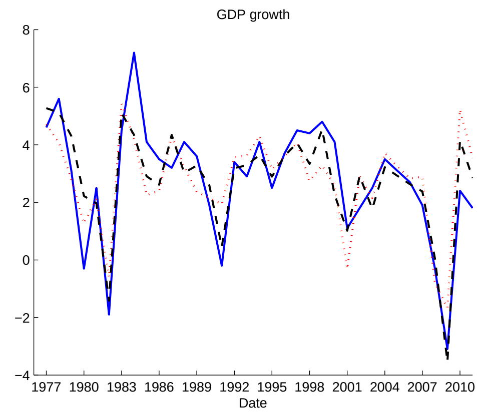
Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(H)$	(4) $\bar{R}^2(C+L+H)$	(4) - (1)
PANEL a) LF observations					
GDP	60.54	8.59	-2.04	73.39	12.85
GDP - Manufacturing	81.88	-3.03	0.22	82.90	1.02
GDP - Agriculture, forestry, fishing, and hunting	1.43	-2.52	-0.85	-2.16	-3.59
GDP - Construction	44.05	11.22	-2.84	58.70	14.64
GDP - Wholesale trade	20.35	7.90	-2.94	29.77	9.41
GDP - Retail trade	30.70	-2.86	2.67	33.71	3.00
GDP - Transportation and warehousing	62.14	-2.95	0.19	61.95	-0.19
GDP - Information	12.14	22.28	-3.03	36.53	24.39
GDP - Finance, insurance, real estate, rental, and leasing	-1.42	21.22	-2.19	18.58	20.00
GDP - Professional and business services	30.02	30.21	-1.98	64.52	34.50
GDP - Educational services, health care, and social assistance	-1.38	18.38	-0.60	16.25	17.63
GDP - Arts, entert., recreat., accommodation, and food serv.	53.51	-2.23	-0.50	57.00	3.49
GDP - Government	-2.12	22.37	-2.95	18.96	21.08

In the table we display the adjusted R^2, denoted $\bar{R}^2$, of the regression of growth rates of selected HF and LF indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(L)$ and $\bar{R}^2(H)$) and on these three factors together (column (4)). The last column displays the difference between the values in the fourth and the first columns, i.e. the increment in the adjusted R^2 when both specific factors are added as a regressors to the common factor.

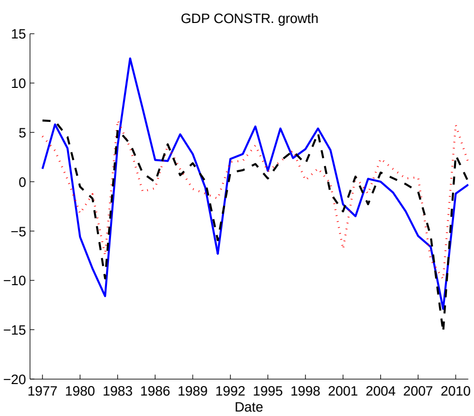
Figure 3.7: Regression of LF and HF indices on estimated factors.



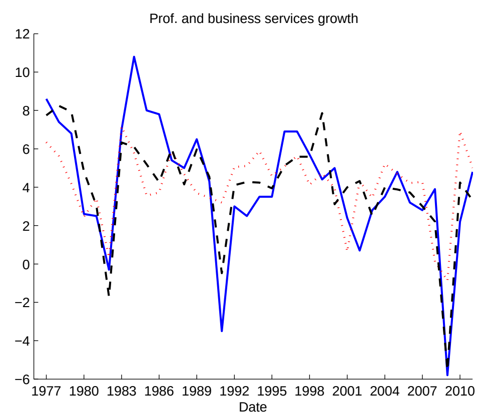
(a) HF Index: Industrial Production Index growth.



(b) LF Index: Aggregate GDP Index growth.



(c) LF Index: GDP-Construction Index growth.



(d) LF Index: Prof. and Busn. Serv. Index growth.

Each panel displays the time series of the growth rate an observed index (solid line) and its fitted value obtained from a regression on the common factor (dotted line). Fitted values from a regression on multiple factors (dashed line) are also displayed. In the first panel we regress the IP index on both the common and HF-specific factors, in the second panel we regress the aggregate GDP Index (LF) on $\hat{\hat{g}}_t^C$, $\hat{\hat{g}}_t^H$ and $\hat{\hat{g}}_t^L$. In the third panel and fourth we regress the growth rates of the LF Construction Index and of Professional and Business Services Index, respectively, on both the common and LF-specific factors. In the fourth we regress the growth rates of Professional and Business Services Index on the common and LF-specific factors. The first, second and fourth indices reported in the panels are aggregates of the indices used to estimate the factors. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1977.Q1-2011.Q4.

Table 3.21: Regression of yearly sectoral IPgrowth on the common and HF-specific factors: adjusted R^2

Panel a. Regressor: common factor		**Panel b. Regressors: common and HF spec. factors**		**Panel c. Absolute increment in adj. R^2**	
Sector	$\bar{R}^2$	Sector	$\bar{R}^2$	Sector	$\Delta\bar{R}^2$
Ten sectors with largest $\bar{R}^2$		*Ten sectors with largest $\bar{R}^2$*		*Ten sectors with largest change in $\bar{R}^2$*	
Forging & stamping	79.86	Forging & stamping	82.87	Mining & oil & gas field machinery	44.03
Other fabricated metal product	74.67	Com. & serv. ind. machin. & other gen. purp. machin.	76.77	Veneer, plywood, & engineered wood product	37.63
Coating, engraving, heat treating, & allied activities	74.28	Plastics product	76.72	Millwork	30.46
Com. & serv. ind. machin. & other gen. purp. machin.	72.01	Metalworking machinery	75.77	All other wood product	26.31
Machine shops, turned prod., & screw, nut, & bolt	71.53	Other fabricated metal product	75.69	Sawmills & wood preservation	24.87
Foundries	69.78	Coating, engraving, heat treating, & allied activities	74.68	Major appliance	24.78
Other electrical equipment	66.79	Household & institutional furniture & kitchen cabinet	73.36	Resin & synthetic rubber	23.86
Metalworking machinery	66.00	Machine shops, turned product, & screw, nut, & bolt	73.14	Support activities for mining	23.85
Plastics product	61.36	Millwork	71.19	Paperboard container	23.46
Household & institutional furniture & kitchen cabinet	60.74	Foundries	69.61	Fiber, yarn, & thread mills Tobacco	22.94
Ten sectors with smallest $\bar{R}^2$		*Ten sectors with smallest $\bar{R}^2$*		*Ten sectors with smallest change in $\bar{R}^2$*	
Other Food Except Coffee & Tea	-0.28	Other Food Except Coffee & Tea	0.79	Audio & video equipment	-0.63
Pharmaceutical & medicine	-0.46	Sugar & confectionery product	0.75	Soap, cleaning compound, & toilet preparation	-0.63
Grain & oilseed milling	-0.46	Ice Cream & Frozen Desserts	0.19	Other transportation equipment	-0.65
Aerospace product & parts	-0.50	Wineries & Distilleries	-0.49	Aerospace product & parts	-0.66
Ice Cream & Frozen Desserts	-0.56	Grain & oilseed milling	-0.69	Ship & boat building	-0.69
Oil & gas extraction	-0.56	Animal slaughtering & processing	-0.81	Petroleum refineries	-0.70
Wineries & Distilleries	-0.56	Aerospace product & parts	-1.16	Animal food	-0.71
Nonferrous metal (except aluminum) smelt. & refin.	-0.58	Nonferrous metal (except aluminum) smelt. & ref.	-1.31	Fruit & vegetable preserving & specialty food	-0.73
Dairy product (except frozen)	-0.66	Dairy product (except frozen)	-1.39	Nonferrous metal (except aluminum) smelt. & ref.	-0.73
Fruit & vegetable preserving & specialty food	-0.71	Fruit & vegetable preserving & specialty food	-1.44	Dairy product (except frozen)	-0.73

In the table we display the adjusted R^2, denoted $\bar{R}^2$, for the time series regressions of the growth rates of the of 117 industrial production indices on the estimated factors. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The regressions in *Panel a* involve a HF explained variable and the estimated common factor. The regressions in *Panel b* involve a HF explained variable and both the common and HF-specific estimated factors. In *Panel c* we display the absolute increment $\bar{R}^2$, when the HF-specific factor is added to the common factor.

Table 3.22: Estimated number of factors

	$\varepsilon_{X,HF}$	ε_X^H	ε_X^L	$[\ \varepsilon_X^H\ \ \varepsilon_X^L\ ]$
IP data: 1989.Q1-2011.Q4. Non-IP data: Gross Output, 1989-2011				
IC_{p1}	1	1	15	1
IC_{p2}	1	1	1	1

In this table we report the number of latent pervasive factors selected by the IC_{p1} and IC_{p2} information criteria on different subpanels. Subpanels $\varepsilon_{X,HF}$ and ε_X^H correspond to productivity innovations of IP data sampled at quarterly and yearly frequency, respectively. Panels ε_X^L and $[\ \varepsilon_X^H\ \ \varepsilon_X^L\ ]$ correspond to to productivity innovations of non-IP data, and the stacked panels of to productivity innovations of IP and non-IP data, respectively. We use $k_{max} = 15$ as maximum number of factors when computing IC_{p1} and IC_{p2}.

In Table 3.23 we display the estimated eigenvalues of the variance-covariance matrix of the stacked PC's estimated in each subpanel ε_X^H and ε_X^L.

Table 3.23: eigenvalues of the variance-covariance matrix of the stacked PC's

1^{st} eig.	2^{nd} eig.	3^{rd} eig.	4^{th} eig.
IP data: 1988.Q1-2011.Q4. Non-IP data: Gross Output, 1988-2011			
1.93	1.28	0.72	0.07

We report the eigenvalues of the the variance-covariance matrix of the first two stacked PCs computed in each subpanel of IP and non-IP productivity innovations.

Table 3.24: Adj. R^2 of selected Gross Output indices on factors estimated from indices growth rates

Panel a. *Yearly observations, 1988-2011*

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(H)$	(4) $\bar{R}^2(C+L+H)$	(4) - (1)
PANEL a) LF observations					
GO (all sectors)	68.54	12.20	-3.66	90.95	22.41
GO - Manufacturing	86.08	-3.05	-0.45	94.81	8.73
GO - Agriculture, forestry, fishing, and hunting	-3.21	3.35	5.91	5.85	9.06
GO - Construction	25.30	34.16	-4.30	66.04	40.74
GO - Wholesale trade	80.82	-3.85	-4.41	78.97	-1.85
GO - Retail trade	64.72	-4.50	9.50	79.78	15.06
GO - Transportation and warehousing	83.82	-4.51	-4.49	82.80	-1.02
GO - Information	33.70	38.59	-4.50	80.62	46.91
GO - Finance, insurance, real estate, rental, and leasing	3.37	50.30	-3.12	59.08	55.71
GO - Professional and business services	45.13	21.97	-2.33	75.88	30.76
GO - Educational services, health care, and social assistance	-4.19	-1.58	0.56	-6.00	-1.81
GO - Arts, entert., recreat., accommodation, and food serv.	71.06	-3.74	-4.47	70.50	-0.56

In the table we display the adjusted R^2, denoted $\bar{R}^2$, of the regression of growth rates of selected HF and LF indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(L)$ and $\bar{R}^2(H)$) and different combinations of these three factors (column (4)). The last column displays the difference between the values in the fourth and first column, i.e. the increment in the adjusted R^2 when the specific factors are added as a regressors to the common factor. Factors are estimated from the panel of 38 non-IP sectors and 117 industrial production indices considered by Foerster, Sarte, and Watson (2011), using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1988-2011.

Table 3.25: Adj. R^2 of selected Gross Output indices on factors estimated from productivity innovations (contemporaneous values only)

Panel a. *Yearly observations, 1989-2011*

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(H)$	(4) $\bar{R}^2(C+L+H)$	(4) - (1)
PANEL a) LF observations					
GO (all sectors)	42.17	13.97	5.40	79.84	37.67
GO - Manufacturing	62.29	-0.20	13.40	91.84	29.55
GO - Agriculture, forestry, fishing, and hunting	0.96	-4.23	4.61	4.22	3.25
GO - Construction	6.64	20.55	0.82	42.14	35.50
GO - Wholesale trade	74.73	-3.08	-1.46	79.39	4.66
GO - Retail trade	47.02	-4.35	25.68	76.76	29.74
GO - Transportation and warehousing	70.42	-2.69	3.01	81.72	11.29
GO - Information	17.78	42.45	-2.82	74.21	56.42
GO - Finance, insurance, real estate, rental, and leasing	-4.09	17.55	-0.36	24.36	28.45
GO - Professional and business services	25.17	44.89	-4.34	79.57	54.40
GO - Educational services, health care, and social assistance	-4.73	-4.48	-1.87	-11.24	-6.51
GO - Arts, entert., recreat., accommodation, and food serv.	55.64	-2.29	-0.74	60.66	5.02

Panel b. *Quarterly observations, 1989.Q1-2011.Q4*

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(H)$	(4) $\bar{R}^2(C+H)$	(4) - (1)
PANEL c) HF observations					
Industrial Production	31.21		50.15	77.25	46.05

In the table we display the adjusted R^2, denoted $\bar{R}^2$, of the regression of growth rates of selected HF and LF indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(L)$ and $\bar{R}^2(H)$) and different combinations of these three factors (column (4)). Only contemporaneous values of the factors are used as regressors. The last column displays the difference between the values in the fourth and first column, i.e. the increment in the adjusted R^2 when the specific factors are added as a regressors to the common factor. Factors are estimated from the panels of productivity innovations of 38 non-IP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1988-2011 because the productivity shocks can not be computed for the first year of the sample (see Foerster, Sarte, and Watson (2011), especially their equation (B38) on page 10 of their Appendix B).

Table 3.26: Adj. R^2 of selected Gross Output indices on factors estimated from productivity innovations (contemporaneous values and first lag)

Panel a. *Yearly observations, 1990-2011*

Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(H)$	(4) $\bar{R}^2(C+L+H)$	(4) - (1)
PANEL a) LF observations					
GO (all sectors)	42.09	12.39	57.42	84.99	42.90
GO - Manufacturing	61.02	2.42	63.69	92.34	31.33
GO - Agriculture, forestry, fishing, and hunting	-0.23	-5.19	-0.22	-7.52	-7.29
GO - Construction	11.35	21.18	35.99	54.99	43.64
GO - Wholesale trade	74.13	-1.50	59.86	84.78	10.64
GO - Retail trade	55.17	-8.95	59.99	82.78	27.61
GO - Transportation and warehousing	71.66	10.46	71.68	88.28	16.63
GO - Information	21.87	41.54	23.39	83.94	62.07
GO - Finance, insurance, real estate, rental, and leasing	-2.16	14.83	3.08	21.62	23.78
GO - Professional and business services	37.79	46.62	31.84	82.58	44.78
GO - Educational services, health care, and social assistance	-10.37	-10.06	-2.20	-17.33	-6.95
GO - Arts, entert., recreat., accommodation, and food serv.	59.11	15.31	62.66	70.32	11.21

Panel b. *Yearly observations, 1990-2011*

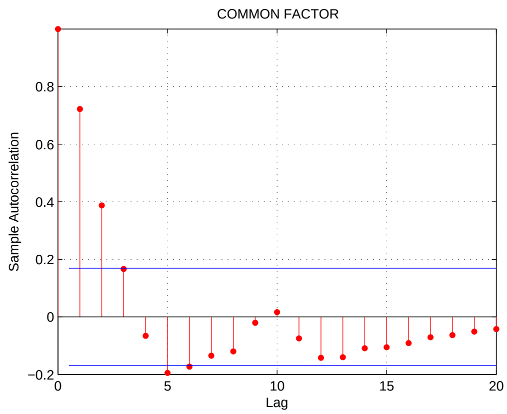
Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(H)$	(4) $\bar{R}^2(C+L)$	(4) - (1)
PANEL b) LF observations					
GO (all sectors)	42.09	12.39		54.12	12.02
GO - Manufacturing	61.02	2.42		60.54	-0.48
GO - Agriculture, forestry, fishing, and hunting	-0.23	-5.19		-5.92	-5.69
GO - Construction	11.35	21.18		20.03	8.69
GO - Wholesale trade	74.13	-1.50		75.58	1.44
GO - Retail trade	55.17	-8.95		51.76	-3.41
GO - Transportation and warehousing	71.66	10.46		69.56	-2.09
GO - Information	21.87	41.54		66.72	44.85
GO - Finance, insurance, real estate, rental, and leasing	-2.16	14.83		7.41	9.57
GO - Professional and business services	37.79	46.62		73.22	35.42
GO - Educational services, health care, and social assistance	-10.37	-10.06		-22.67	-12.30
GO - Arts, entert., recreat., accommodation, and food serv.	59.11	15.31		54.99	-4.13

Panel c. *Quarterly observations, 1990-2011*

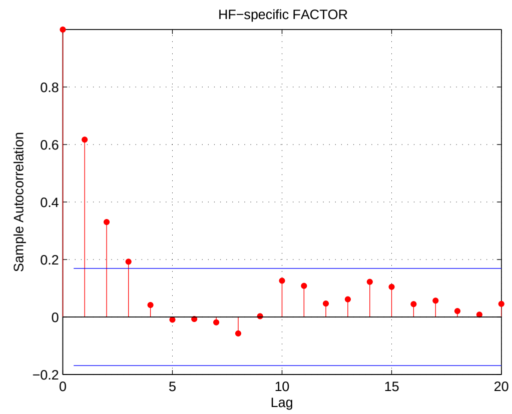
Sector	(1) $\bar{R}^2(C)$	(2) $\bar{R}^2(L)$	(3) $\bar{R}^2(H)$	(4) $\bar{R}^2(C+H)$	(4) - (1)
PANEL c) HF observations					
Industrial Production (Q)	35.89		63.22	77.69	41.81

In the table we display the adjusted R^2, denoted $\bar{R}^2$, of the regression of growth rates of selected HF and LF indices on the common factor (column $\bar{R}^2(C)$), the specific HF and LF factors (columns $\bar{R}^2(L)$ and $\bar{R}^2(H)$) and different combinations of these three factors (column (4)). Both the contemporaneous and the lagged values (only lag 1 is included) of the factors are used as regressors. The last column displays the difference between the values in the fourth and first column, i.e. the increment in the adjusted R^2 when the specific factors are added as a regressors to the common factor. Factors are estimated from the panels of productivity innovations of 38 non-IP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of both the factor model and the regressions is 1990-2011 because the productivity shocks can not be computed for the first year of the sample (see Foerster, Sarte, and Watson (2011), especially their equation (B38) on page 10 of their Appendix B).

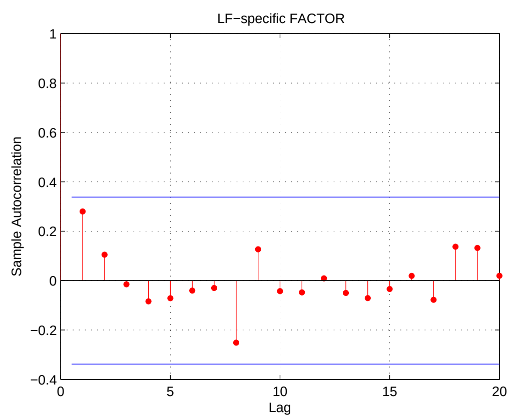
Figure 3.8: Autocorrelation functions of the estimated common and specific factors.



(a) Common factor: autocorrelation function.



(b) HF factor: autocorrelation function.



(c) LF factor: autocorrelation function.

Panel (a) displays the autocorrelation function of the estimated values of the common factor. Panel (b) displays the autocorrelation function of the estimated values of the HF factor. Panel (c) displays the autocorrelation function of the estimated values of the LF factor. The horizontal lines are asymptotic 95% confidence bands. The factors are estimated from the panel of 42 GDP sectors and 117 industrial production indices using a mixed frequency factor model with $k^C = k^H = k^L = 1$. The sample period for the estimation of the factor model is 1977.Q1-2011.Q4.

3.9.6 Granularity analysis

In this section we perform a Granularity analysis in line with Foerster, Sarte, and Watson (2011) and Gabaix (2011).

Quarterly IP Index

Table 3.27: Share weight decomposition of Industrial Production quarterly index.

Series	1977-2011	1977-1983	1984-2007	2008-2011
$g_t = \sum w_{it} x_{it}$	5.7	7.9	3.6	9.7
$(1/N) \sum x_{it}$	7.3	9.9	4.2	13.7
$\sum(\bar{w}_i - (1/N))x_{it}$	2.2	3.3	1.5	2.9
$\sum(w_{it} - \bar{w}_i)x_{it}$	1.0	1.0	0.5	2.1

Entries are the sample standard deviations of the quarterly growth rates of the quarterly Industrial Production index growth (g_t) and its components. Percentage points are at annual rates. The table corresponds to Table 1 in Foerster, Sarte, and Watson (2011).

Table 3.28: Average pairwise correlations of sectoral Industrial Production indices.

1977-2011	1977-1983	1984-2007	2008-2011
0.21	0.25	0.12	0.34

Entries are the average pairwise sample correlations of the quarterly growth rates of the 117 industrial production indices considered in the paper. The table corresponds to Table 2 in Foerster, Sarte, and Watson (2011).

Table 3.29: Standard deviation of Industrial Production indices growth with and without sectoral covariance

	1977-2011	1977-1983	1984-2007	2008-2011
	A. Using Actual w_{it} Share Weights			
With Sectoral covariation	5.6	7.8	3.5	9.5
Without Sectoral covariation	1.9	2.5	1.6	2.5
	B. Using Equal $(1/N)$ Share Weights			
With Sectoral covariation	7.3	9.9	4.2	13.7
Without Sectoral covariation	1.9	2.7	1.4	2.4

The entries for rows labeled "with sectoral covariation" are sample standard deviations of $\sum w_{it} x_{it}$ (Panel A) and $N^{-1} \sum x_{it}$ (Panel B). The entries labeled "without sectoral covariation" are computed as: $\sqrt{T^{-1} \sum_t \sum_i h_{it}^2 (x_{it} - \bar{x}_i)^2}$, where $h_{it} = w_{it}$ in panel A and $h_{it} = N^{-1}$ in panel B. Percentage points are at annual rates. The table corresponds to Table 3 in Foerster, Sarte, and Watson (2011).

Annual GDP sectoral indices (all)

Table 3.30: Share weight decomposition of annual GDP index.

Series(GDP)	1977-2011	1977-1983	1984-2007	2008-2011
$g_t = \sum w_{it} x_{it}$	2.1	2.6	1.4	2.3
$(1/N) \sum x_{it}$	2.7	3.4	1.7	5.2
$\sum(\bar{w}_i - (1/N))x_{it}$	0.9	1.0	0.8	1.8
$\sum(w_{it} - \bar{w}_i)x_{it}$	0.4	0.2	0.2	1.2

Entries are the sample standard deviations of the annual growth rates of annual GDP index growth (g_t) and its components. The index is constructed using weights of nominal GDP. The table corresponds to Table 1 in Foerster, Sarte, and Watson (2011).

Table 3.31: Average pairwise correlations of all GDP sectoral indices.

1977-2011	1977-1983	1984-2007	2008-2011
0.18	0.29	0.11	0.19

Entries are the average pairwise sample correlations of the annual growth rates of the all 61 GDP sectoral indices. The table corresponds to Table 2 in Foerster, Sarte, and Watson (2011).

Table 3.32: Standard deviation of GDP sectoral indices growth with and without sectoral covariance

	1977-2011	1977-1983	1984-2007	2008-2011
	A. Using Actual w_{it} Share Weights			
With Sectoral covariation	2.0	2.6	1.4	2.3
Without Sectoral covariation	0.8	0.9	0.7	1.0
	B. Using Equal $(1/N)$ Share Weights			
With Sectoral covariation	2.7	3.4	1.7	5.2
Without Sectoral covariation	1.2	1.1	1.0	1.9

The entries for rows labeled "with sectoral covariation" are sample standard deviations of $\sum w_{it}x_{it}$ (Panel A) and $N^{-1}\sum x_{it}$ (Panel B). The entries labeled "without sectoral covariation" are computed as: $\sqrt{T^{-1}\sum_t \sum_i h_{it}^2 (x_{it} - \bar{x}_i)^2}$, where $h_{it} = w_{it}$ in panel A and $h_{it} = N^{-1}$ in panel B. The table corresponds to Table 3 in Foerster, Sarte, and Watson (2011).

Annual IP sectors in GDP sectoral indices

Table 3.33: Share weight decomposition of IP sectors in GDP annual indices.

Series(GDP)	1977-2011	1977-1983	1984-2007	2008-2011
$g_t = \sum w_{it} x_{it}$	4.5	5.8	3.5	7.6
$(1/N) \sum x_{it}$	5.3	6.9	3.4	10.9
$\sum (\bar{w}_i - (1/N)) x_{it}$	1.2	0.9	1.1	2.2
$\sum (w_{it} - \bar{w}_i) x_{it}$	1.6	0.6	0.4	4.8

Entries are the sample standard deviations of the annual growth rate and the components of the annual index (g_t) created from the IP sectors in GDP indices. The index is constructed using weights of nominal GDP. The table corresponds to Table 1 in Foerster, Sarte, and Watson (2011).

Table 3.34: Average pairwise correlations of all IP sectors in GDP indices.

1977-2011	1977-1983	1984-2007	2008-2011
0.35	0.48	0.27	0.29

Entries are the average pairwise sample correlations of the annual growth rates of the 19 IP sectors in GDP sectoral indices. The table corresponds to Table 2 in Foerster, Sarte, and Watson (2011).

Table 3.35: Standard deviation of IP sectors in GDP sectoral indices growth with and without sectoral covariance

	1977-2011	1977-1983	1984-2007	2008-2011
	A. Using Actual w_{it} Share Weights			
With Sectoral covariation	4.5	5.8	3.5	7.6
Without Sectoral covariation	2.7	2.6	2.3	3.9
	B. Using Equal $(1/N)$ Share Weights			
With Sectoral covariation	5.3	6.9	3.4	10.9
Without Sectoral covariation	2.8	2.9	1.9	5.3

The entries for rows labeled "with sectoral covariation" are sample standard deviations of $\sum w_{it} x_{it}$ (Panel A) and $N^{-1} \sum x_{it}$ (Panel B). The entries labeled "without sectoral covariation" are computed as: $\sqrt{T^{-1} \sum_t \sum_i h_{it}^2 (x_{it} - \bar{x}_i)^2}$, where $h_{it} = w_{it}$ in panel A and $h_{it} = N^{-1}$ in panel B. The table corresponds to Table 3 in Foerster, Sarte, and Watson (2011).

Annual non-IP sectors in GDP sectoral indices

Table 3.36: Share weight decomposition of non-IP sectors in GDP annual indices.

Series(GDP)	1977-2011	1977-1983	1984-2007	2008-2011
$g_t = \sum w_{it}x_{it}$	1.7	2.1	1.3	1.9
$(1/N)\sum x_{it}$	2.1	2.6	1.3	3.4
$\sum(\bar{w}_i - (1/N))x_{it}$	0.9	0.5	1.0	1.5
$\sum(w_{it} - \bar{w}_i)x_{it}$	0.2	0.3	0.2	0.2

Entries are the sample standard deviations of the annual growth rate and the components of the annual index (g_t) created from the non-IP sectors in GDP indices. The index is constructed using weights of nominal GDP. The table corresponds to Table 1 in Foerster, Sarte, and Watson (2011).

Table 3.37: Average pairwise correlations of all non-IP sectors in GDP indices.

1977-2011	1977-1983	1984-2007	2008-2011
0.18	0.32	0.10	0.21

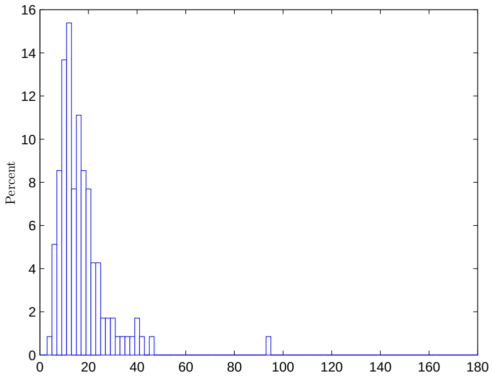
Entries are the average pairwise sample correlations of the annual growth rates of the 42 non-IP sectors in GDP sectoral indices considered in the paper. The table corresponds to Table 2 in Foerster, Sarte, and Watson (2011).

Table 3.38: Standard deviation of non-IP sectors in GDP sectoral indices growth with and without sectoral covariance

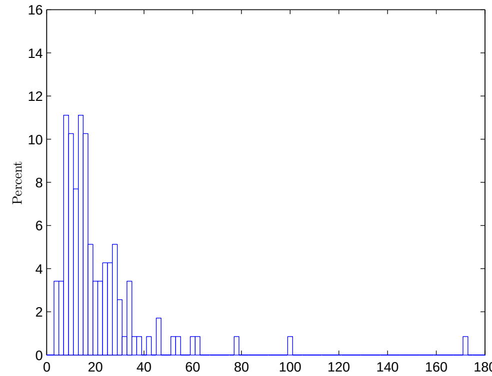
	1977-2011	1977-1983	1984-2007	2008-2011
	A. Using Actual w_{it} Share Weights			
With Sectoral covariation	1.7	2.1	1.3	1.9
Without Sectoral covariation	0.9	0.9	0.8	0.9
	B. Using Equal $(1/N)$ Share Weights			
With Sectoral covariation	2.1	2.6	1.3	3.4
Without Sectoral covariation	1.2	0.9	1.1	1.4

The entries for rows labeled "with sectoral covariation" are sample standard deviations of $\sum w_{it}x_{it}$ (Panel A) and $N^{-1}\sum x_{it}$ (Panel B). The entries labeled "without sectoral covariation" are computed as: $\sqrt{T^{-1}\sum_t \sum_i h_{it}^2(x_{it} - \bar{x}_i)^2}$, where $h_{it} = w_{it}$ in panel A and $h_{it} = N^{-1}$ in panel B. The table corresponds to Table 3 in Foerster, Sarte, and Watson (2011).

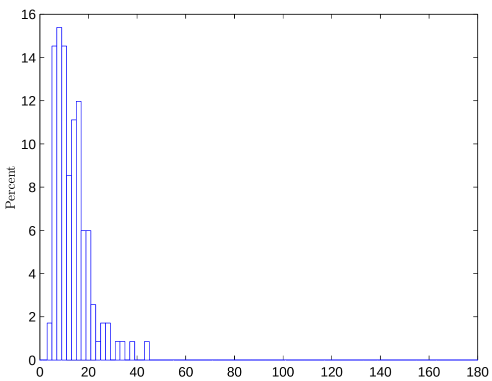
Figure 3.9: Standard deviations of quarterly growth rates of sectoral Industrial Production indices.



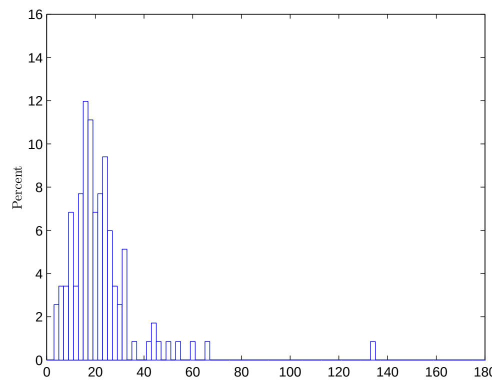
(a) 1977.Q1-2011.Q4



(b) 1977.Q1-1983.Q4



(c) 1984.Q1-2007.Q4

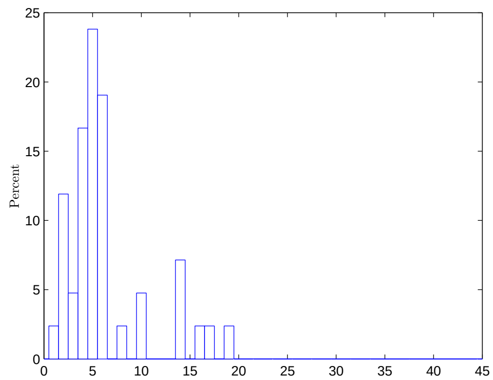


(d) 2008.Q1-2011.Q4

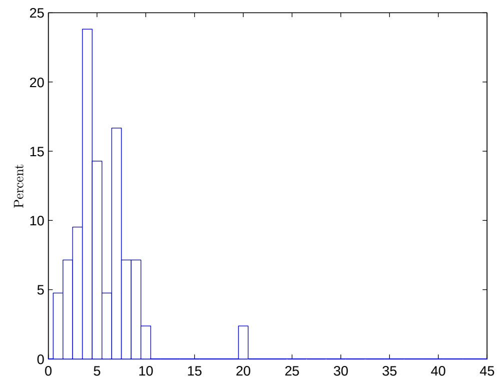
Each panel displays the histogram of the standard deviations of quarterly growth rates of sectoral IP indices. Percentage points are at annual rates. The graphs correspond to Figure 2 in Foerster, Sarte, and Watson (2011).

Panel	25th Percentile	Median	75th Percentile
(a) 1977.Q1-2011.Q4	10.77	14.19	19.71
(b) 1977.Q1-1983.Q4	10.60	15.92	25.29
(c) 1984.Q1-2007.Q4	8.32	11.48	16.71
(d) 2008.Q1-2011.Q4	14.40	18.91	25.65

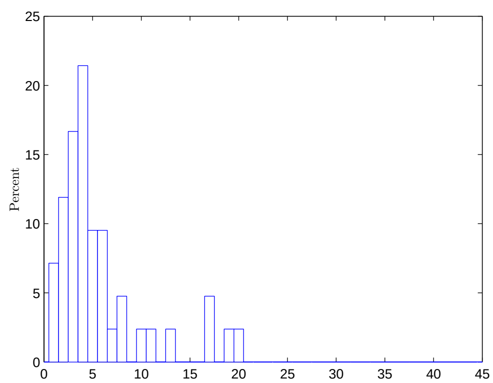
Figure 3.10: Standard deviations of annual growth rates of NON-IP sectoral GDP indices.



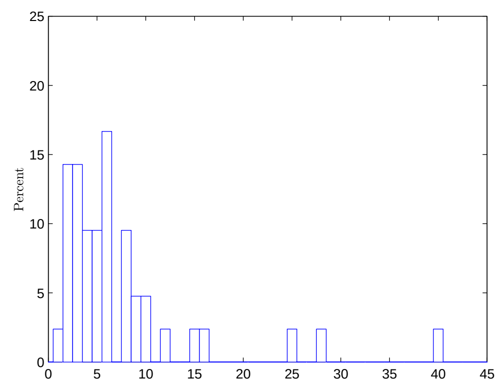
(a) 1977-2011



(b) 1977-1983



(c) 1984-2007

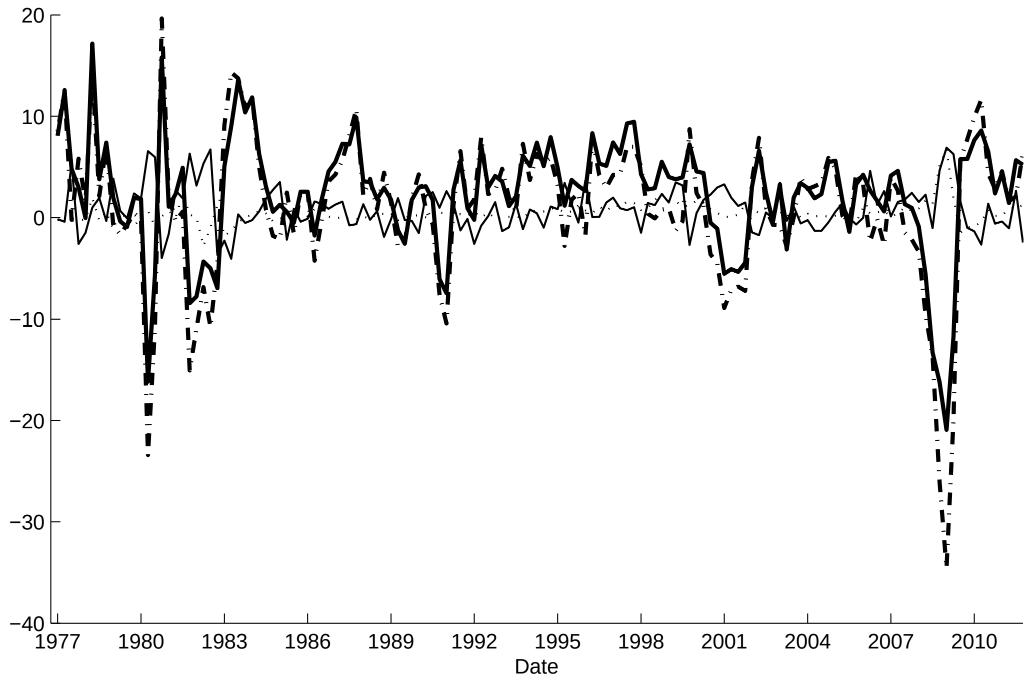


(d) 2008-2011

Each panel displays the histogram of the standard deviations of the annual growth rates of the 42 non-IP sectoral GDP indices. The graphs correspond to Figure 2 in Foerster, Sarte, and Watson (2011).

Panel	25th Percentile	Median	75th Percentile
(a) 1977-2011	3.92	4.91	6.39
(b) 1977-1983	3.65	4.89	7.05
(c) 1984-2007	3.22	4.31	6.19
(d) 2008-2011	3.22	5.47	8.39

Figure 3.11: Share weight decomposition of quarterly IP index.

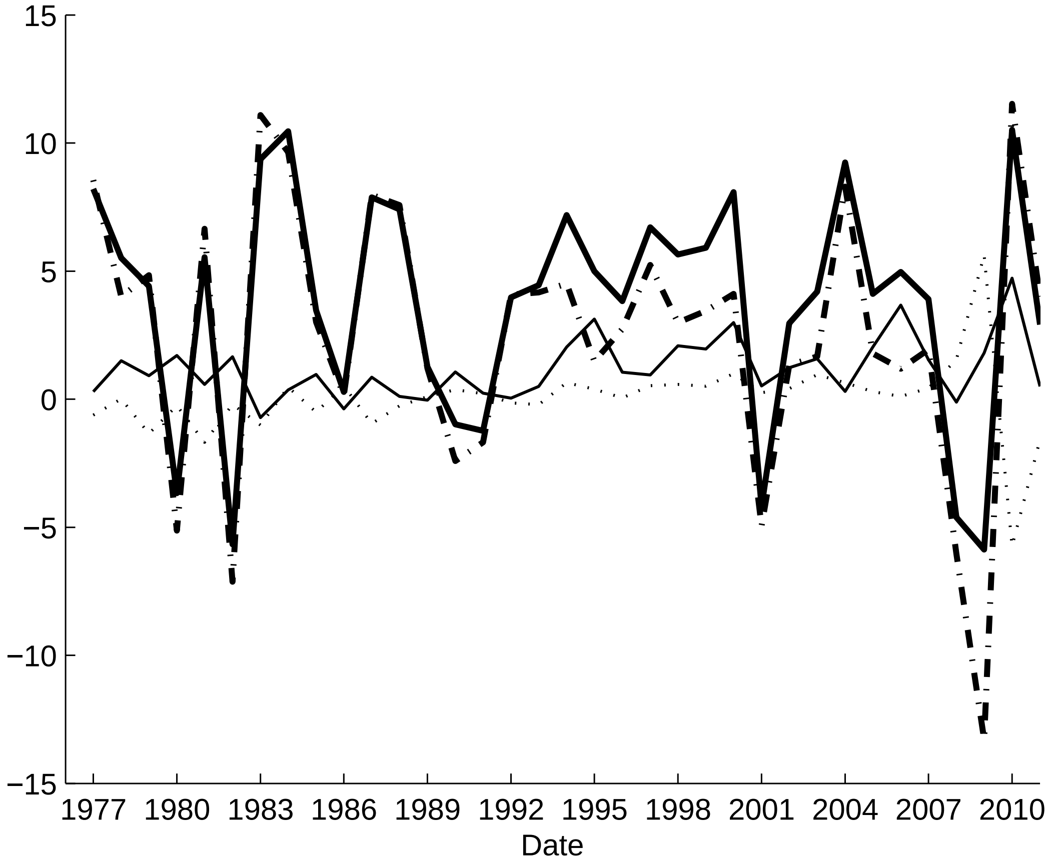


The figure displays the Share weight decomposition of quarterly IP index growth rates. Percentage points are at annual rates. This figure corresponds to Figure 3 in Foerster, Sarte, and Watson (2011). The bold solid line corresponds to $\sum w_{it}x_{it}$ (i.e. the aggregate IP index). The bold dash-dotted line corresponds to $\sum(1/N)x_{it}$. The thin solid line corresponds to $\sum(w_i - (1/N))x_{it}$. The thin dotted line corresponds to $\sum(w_{it} - w_i)x_{it}$.

Figure 3.12: Share weight decomposition of annual GDP index.

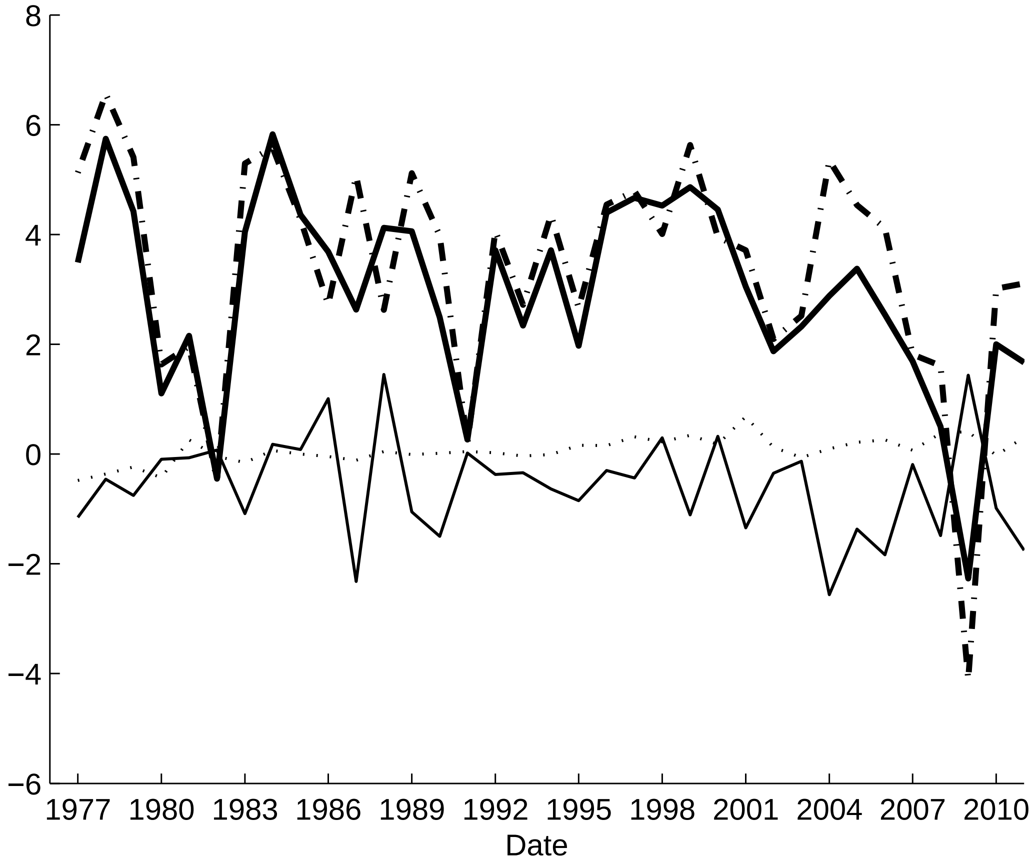
The figure displays the Share weight decomposition of annual GDP index. The index is constructed using weights (w_{it}) of nominal GDP. This figure corresponds to Figure 3 in Foerster, Sarte, and Watson (2011). The bold solid line corresponds to $\sum w_{it}x_{it}$ (i.e. the aggregate real GDP index). The bold dash-dotted line corresponds to $\sum(1/N)x_{it}$. The thin solid line corresponds to $\sum(w_i - (1/N))x_{it}$. The thin dotted line corresponds to $\sum(w_{it} - w_i)x_{it}$.

Figure 3.13: Share weight decomposition of annual IP sectors in GDP indices.



The figure displays the Share weight decomposition of annual GDP index of IP sectors. The index is constructed using weights (w_{it}) of nominal GDP. This figure corresponds to Figure 3 in Foerster, Sarte, and Watson (2011). The bold solid line corresponds to $\sum w_{it}x_{it}$ (i.e. the aggregate real GDP index for IP sectors). The bold dash-dotted line corresponds to $\sum(1/N)x_{it}$. The thin solid line corresponds to $\sum(w_i - (1/N))x_{it}$. The thin dotted line corresponds to $\sum(w_{it} - w_i)x_{it}$.

Figure 3.14: Share weight decomposition of annual non-IP sectors in GDP indices.



The figure displays the Share weight decomposition of annual GDP index of of non-IP sectors . The index is constructed using weights (w_{it}) of nominal GDP. growth rates. This figure corresponds to Figure 3 in Foerster, Sarte, and Watson (2011). The bold solid line corresponds to $\sum w_{it}x_{it}$ (i.e. the aggregate real GDP index for non-IP sectors). The bold dash-dotted line corresponds to $\sum(1/N)x_{it}$. The thin solid line corresponds to $\sum(w_i - (1/N))x_{it}$. The thin dotted line corresponds to $\sum(w_{it} - w_i)x_{it}$.

3.10 Bibliograpghy of Chapter 3

ANDERSON, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience.

ANDO, T., AND J. BAI (2013): "Multifactor Asset Pricing with a Large Number of Observable Risk Factors and Unobservable Common and Group-specific Factors," *Working Paper*.

ARUOBA, S. B., F. X. DIEBOLD, AND C. SCOTTI (2009): "Real-time Measurement of Business Conditions," *Journal of Business and Economic Statistics*, 27(4), 417–427.

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74(4), 1133–1150.

BANBURA, M., AND G. RÜNSTLER (2011): "A Look into the Factor Model Black Box: Publication Lags and the Role of Hard and Soft Data in Forecasting GDP," *International Journal of Forecasting*, 27, 333–346.

BEKAERT, G., R. J. HODRICK, AND X. ZHANG (2009): "International Stock Return Comovements," *Journal of Finance*, 64, 2591–2626.

BOIVIN, J., AND S. NG (2006): "Are More Data Always Better for Factor Analysis?," *Journal of Econometrics*, 132(1), 169–194.

BREITUNG, J., AND S. EICKMEIER (2014): "Analyzing Business and Financial Cycles Using Multi-level Factor Models," *Deutsche Bundesbank, Research Centre, Discussion Papers 11/2014*.

CARVALHO, V. M. (2010): "Aggregate Fluctuations and the Network Structure of Intersectoral Trade," *Unpublished Manuscript*.

CHEN, K. H., AND J. ROBINSON (1989): "Comparison of Factor Spaces of two Related Populations," *Journal of Multivariate Analysis*, 28, 190–203.

CHEN, P. (2010): "A Grouped Factor Model," *Working Paper*.

——— (2012): "Common Factors and Specific Factors," *Working Paper*.

DAUXOIS, J., AND A. POUSSE (1975): "Une Extension de l'Analyse Canonique. Quelques Applications.," *Ann. Inst. Henri Poincaré*, XI, 355–379.

DAUXOIS, J., Y. ROMAIN, AND S. VIGUIER (1993): "Comparison of Two Factor Subspaces," *Journal of Multivariate Analysis*, 44(1), 160–178.

DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977): "Maximum Likelihood from Incomplete Data via di EM Algorithm," *Journal of the Royal Statistical Society Series B*, 39(1), 1–38.

FLURY, B. N. (1984): "Common Principal Components in k Groups," *Journal of the American Statistical Association*, 79, 892–898.

FLURY, B. N. (1988): *Common Principal Components and Related Multivariate Models*. Wiley, New York.

FOERSTER, A. T., P.-D. G. SARTE, AND M. W. WATSON (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119(1), 1–38.

FORNI, M., AND L. REICHLIN (1998): "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics," *The Review of Economic Studies*, 65, 453–473.

FRALE, C., AND L. MONTEFORTE (2010): "FaMIDAS: A Mixed Frequency Factor Model with MIDAS Structure," *Government of the Italian Republic (Italy), Ministry of Economy and Finance, Department of the Treasury Working Paper*, 3.

GABAIX, X. (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

GOYAL, A., C. PÉRIGNON, AND C. VILLA (2008): "How Common are Common Return Factors across the NYSE and Nasdaq?," *Journal of Financial Economics*, 90, 252–271.

GREENWOOD, R., AND D. SCHARFSTEIN (2013): "The Growth of Finance," *Journal of Economic Perspectives*, 27(2), 3–28.

GREGORY, A. W., AND A. C. HEAD (1999): "Common and Country-specific Fluctuations in Productivity, Investment, and the Current Account," *Journal of Monetary Economics*, 44(3), 423 – 451.

HALLIN, M., AND R. LISKA (2007): "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102(478), 603–617.

——— (2011): "Dynamic Factors in the Presence of Blocks," *Journal of Econometrics*, 163, 29–41.

HORVATH, M. (1998): "Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks," *Review of Economic Dynamics*, 1, 781–808.

JUNGBACKER, B., S. KOOPMAN, AND M. V. DER WEL (2011): "Maximum Likelihood Estimation for Dynamic Factor Models with Missing Data," *Journal of Economic Dynamics and Control*, 35(8), 1358 – 1368.

KING, R. G., AND M. WATSON (2002): "System Reduction and Model Solution Algorithms for Singular Linear Rational Expectations Models," *Computational Economics*, 20, 57–86.

KORAJCZYK, R. A., AND R. SADKA (2008): "Pricing the Commonality across Alternative Measures of Liquidity," *Journal of Financial Economics*, 87(1), 45 – 72.

KOSE, A. M., C. OTROK, AND C. H. WHITEMAN (2008): "Understanding the Evolution of World Business Cycles," *Journal of International Economics*, 75, 110–130.

KRZANOWSKI, W. (1979): "Between-groups Comparison of Principal Components," *Journal of the American Statistical Association*, 74, 703–707.

LONG, J. B. J., AND C. I. PLOSSER (1983): "Real Business Cycles," *Journal of Political Economy*, 91, 39–69.

MAGNUS, J. R., AND H. NEUDECKER (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons: Chichester/New York.

MARCELLINO, M., AND C. SCHUMACHER (2010): "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP," *Oxford Bulletin of Economics and Statistics*, 72(4), 518–550.

MARIANO, R. S., AND Y. MURASAWA (2003): "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics*, 18(4), 427–443.

MOENCH, E., AND S. NG (2011): "A Hierarchical Factor Analysis of US Housing Market Dynamics," *The Econometrics Journal*, 14, C1–C24.

MOENCH, E., S. NG, AND S. POTTER (2013): "Dynamic Hierarchical Factor Models," *Review of Economics and Statistics*, 95, 1811–1817.

NUNES, L. C. (2005): "Nowcasting Quarterly GDP Growth in a Monthly Coincident Indicator Model," *Journal of Forecasting*, 24(8), 575–592.

SCHOTT, J. R. (1988): "Common Principal Components Subspaces in Two Groups," *Biometrika*, 75, 229–236.

——— (1991): "Some Tests for Common Principal Component Subspaces in Several Groups," *Biometrika*, 78, 771–777.

SCHOTT, J. R. (1999): "Partial Common Principal Component Subspaces," *Biometrika*, 86, 899–908.

——— (2005): *Matrix Analysis for Statistic*. Wiley, New York, 2 edn.

STOCK, J. H., AND M. W. WATSON (2002a): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167 – 1179.

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–162.

——— (2010): "Dynamic Factor Models," in *Oxford Handbook of Economic Forecasting*, ed. by E. Graham, C. Granger, and A. Timmerman, pp. 87–115. Michael P. Clements and David F. Hendry (eds), Oxford University Press, Amsterdam.

VIGUIER-PLA, S. (2004): "Factor-based Comparison of k Populations," *Statistics*, 38(1), 1–15.

WANG, P. (2012): "Large Dimensional Factor Models with a Multi-Level Factor Structure: Identifcation, Estimation, and Inference," *Working Paper*.

Chapter 4

Indirect Inference Estimation of Mixed Frequency Stochastic Volatility State Space Models using MIDAS Regressions and ARCH Models

Abstract

We examine the relationship between MIDAS regressions and the estimation of state space models applied to mixed frequency data. While in some cases the binding function is known, in general it is not, and therefore indirect inference is called for. The approach is appealing when we consider state space models which feature stochastic volatility, or other non-Gaussian and nonlinear settings where maximum likelihood methods require computationally demanding approximate filters. The stochastic volatility feature is particularly relevant when considering high frequency financial series. In addition, we propose a filtering scheme which relies on a combination of reprojection methods and nowcasting MIDAS regressions with ARCH models. We assess the efficiency of our indirect inference estimator for the stochastic volatility model by comparing it with the Maximum Likelihood (ML) estimator in Monte Carlo simulation experiments. The ML estimate is computed with a simulation-based Expectation-Maximization (EM) algorithm, in which the smoothing distribution required in the E step is obtained via a particle forward-filtering/backward-smoothing algorithm. Our Monte Carlo simulations show that the Indirect Inference procedure is very appealing, as its statistical accuracy is close to that of MLE but the former procedure has clear advantages in terms of computational efficiency. An application to forecasting quarterly GDP growth in the Euro area with monthly macroeconomic indicators illustrates the usefulness of our procedure in empirical analysis.

JEL Codes: C15, C31, C53, E37.

Keywords: Indirect Inference, Reprojection, Mixed-frequency Data, State Space Model, Stochastic Volatility, GDP Forecasting.

4.1 Introduction

Econometric models that take into account the unbalanced nature of datasets have attracted substantial attention recently. Policy makers and practitioners alike need to assess in real-time the current state of the economy, with at best mixed frequency data at their disposal. For example, one of the key indicators of macroeconomic activity, the Gross Domestic Product (GDP), is released quarterly, while a range of leading and coincident indicators is timely available at a monthly or even higher frequency. Hence, we may want to construct a forecast of the current quarter GDP growth (a so called nowcast) based on the available higher frequency information.

Econometric models with mixed frequency data can be classified into two broad classes: (1) likelihood-based involving latent processes and (2) purely regression-based. The former category consists primarily of state space models, studied by Harvey and Pierse (1984), Harvey (1989), Zadrozny (1990), Bernanke, Gertler, and Watson (1997), Mariano and Murasawa (2003), Mittnik and Zadrozny (2005), Aruoba, Diebold, and Scotti (2009), Ghysels and Wright (2009), Kuzin, Marcellino, and Schumacher (2011), among others. The regression-based methods involve Mixed Data Sampling (MIDAS) regressions; see e.g. Ghysels, Santa-Clara, and Valkanov (2006), Andreou, Ghysels, and Kourtellos (2010). As one considers high frequency data, the issue of time-varying volatility becomes increasingly relevant. Dealing with stochastic volatility (SV) in state space models is doable but poses challenges both statistical and computational in nature. One possibility is to consider Bayesian approaches in this context, as done by Carriero, Clark, and Marcellino (2013) and Marcellino, Porqueddu, and Venditti (2015). However, when it comes to classical inference one typically relies on the Expectation-Maximization (EM) algorithm to compute numerically the ML estimate in a model with unobservable variables (Dempster, Laird, and Rubin (1977)). The likelihood function of the model involves a large-dimensional integral with respect to the latent factor paths as the latent factors appear in the conditional mean and volatility of the high frequency data series. This integral representation of the likelihood is impractical for the computation of the ML estimate.

If the objective is to estimate state space models with mixed frequency data - of which there are many examples - featuring stochastic volatility, using classical inference methods, is there perhaps a simpler way to do so? This is the contribution of our paper. We introduce indirect inference estimation procedures proposed by Gouriéroux, Monfort, and Renault (1993), Smith (1993) and Gallant and Tauchen (1996), to estimate the models of interest using MIDAS regressions augmented with ARCH-type models as well as mixed frequency Vector Autoregressive (VAR) models (see e.g. Ghysels (2014)) as auxiliary models. Same frequency data settings are a special case of mixed frequency ones. The analysis in this paper is therefore also applicable to standard state space models. Moreover, the idea of estimating SV-type models using ARCH-type auxiliary models has a long history starting with Engle and Lee (1999) and Pastorello, Renault, and Touzi (2000). Our paper combines insights from the literature on SV models with those from the mixed frequency data literature.

It is worth noting that in some specific cases we know the binding function between the state space model and the implied MIDAS regression, as discussed in Bai, Ghysels, and Wright (2013). However, these cases are rather too simple to be practical, so that the use of indirect inference is a natural way to tackle the unknown binding function. The methods we propose are fairly easy to implement and involve auxiliary model-based estimators involving MIDAS regressions combined with ARCH specifications for the errors. In addition, we filter latent variables, given observables, using reprojection methods proposed by Gallant and Tauchen (1998).

We compare the two estimation methods, namely (1) Maximum Likelihood (ML) and (2) indirect inference, via Monte Carlo simulations. To implement the former method in the mixed frequency SV model, we consider a simulation-based estimator relying on the EM algorithm. The smoothing distribution required in the Expectation step is computed via a particle forward-filtering/backward-smoothing algorithm. We compare the two estimation methods on the basis of (a) statistical criteria - mean/bias/quantiles of sampling distributions, (b) filtering accuracy - both conditional mean and volatility and (c) computational time. Our results show that there are clear advantages in terms of computational time to the new indirect inference procedure put forward in this paper, while the losses in statistical efficiency compared to MLE are very limited. Even in the linear Gaussian case, we find our indirect inference methods remarkably accurate, when compared to the standard MLE based on the Kalman filter.

The paper is organized as follows. Section 4.2 introduces state space models with mixed frequency data and stochastic volatility. Section 4.3 defines our indirect inference estimator. This section covers the linear Gaussian state space model with mixed frequency data as a special case of the general specification, and discusses its relation with MIDAS regressions. This link yields useful insights to define the auxiliary model for indirect inference in the general SV case. Section 4.3 also describes the estimation of the SV model with ML via a simulation-based EM algorithm. Section 4.4 discusses filtering via reprojection, followed by Section 4.5 which reports the results of an extensive Monte Carlo study. Section 4.6 presents an empirical application of our model to the problem of forecasting at short horizons Euro-area quarterly GDP growth using monthly macroeconomic indicators. The dataset is the same as the one considered in the empirical study of Marcellino, Porqueddu, and Venditti (2015). Section 4.7 concludes the paper.

4.2 State Space Models with mixed frequency data and stochastic volatility

There is a burgeoning literature on nowcasting using either MIDAS regressions or state space models, see e.g. Mariano and Murasawa (2003), Nunes (2005), Giannone, Reichlin, and Small (2008), Aruoba, Diebold, and Scotti (2009), Marcellino and Schumacher (2010), Andreou, Ghysels, and Kourtellos (2013) and Banbura and Modugno (2014), among others. Recent surveys include Andreou, Ghysels, and Kourtellos (2011), Foroni and Marcellino (2013) and Banbura, Giannone, Modugno, and Reichlin (2013), where the latter paper has a stronger focus on more complex Kalman filter-based factor modeling techniques.

State space models have been widely used in econometrics as well as other scientific disciplines, in particular engineering where the Gaussian state space model and its Kalman filtering algorithm originated.[1] A key starting point is that observations are driven by some latent process. Moreover, it is also assumed that data are contaminated by measurement errors. To accommodate the mixed frequency sampling scheme, we adopt a time scale expressed in a form that easily represents such mixtures. We will focus on small values of m, the number of high frequency subperiods, such as for example $m = 3$ for monthly data sampled every quarter. We consider a dynamic model for the latent factors as follows:

Assumption 1. *Let (F) be a $n_f \times 1$ dimensional vector process satisfying*

$$F_{t+j/m} = \sum_{l=1}^{p} \Phi_l F_{t+(j-l)/m} + \eta_{t+j/m} \qquad \forall t = 1, \ldots, T, \qquad j = 0, \ldots, m-1, \tag{4.2.1}$$

where Φ_l are $n_f \times n_f$ matrices, the eigenvalues of the autoregressive matrix in the stacked AR(1) representation lie inside the unit circle, and (η) is an i.i.d. zero mean Gaussian error process with diagonal covariance

[1] The econometric literature on the topic is vast, see e.g. Harvey (1989), Hamilton (1994), among others.

matrix Σ_η = *diag*($\sigma^2_{i,\eta}, i = 1, \ldots, n_f$)*. Finally, the number of factors* n_f*, is assumed to be known.*

We have two types of data: (1) time series sampled at a low frequency (LF) - every integer date t, and (2) time series sampled at high frequency (HF) - every $t + j/m$, with $j = 0, \ldots, m-1$. Bai, Ghysels, and Wright (2013) make two convenient simplifications which depart from generality. First, they assume that there is only one low-frequency process and call it y_t, and second, consider the combination of only two sampling frequencies. We will proceed with the same simplifications and also assume - for the sake of simplicity - that there is only one high-frequency series, denoted $x_{t+j/m}$. It is fairly easy to extend the methods proposed in this paper to cases involving multiple low and high frequency series - which we will not cover explicitly.

If the low-frequency process were observed at high frequency, it would relate to the factors as follows:

$$y^*_{t+j/m} = \gamma_1' F_{t+j/m} + u_{1,t+j/m} \qquad \forall t, \qquad j = 0, \ldots, m-1, \tag{4.2.2}$$

where y^* denotes the process which is not directly observed and γ_1 is a $n_f \times 1$ vector of factor loadings. The error process $u_{1,t+j/m}$ has an $AR(k)$ representation:

$$d_1(L^{1/m}) u_{1,t+j/m} = \varepsilon_{1,t+j/m}, \qquad d_1(L^{1/m}) \equiv 1 - d_{11} L^{1/m} - \ldots - d_{k1} L^{k/m}, \tag{4.2.3}$$

where the lag operator $L^{1/m}$ applies to high-frequency data, i.e. $L^{1/m} u_t \equiv u_{t-1/m}$. The observed low-frequency process y relates to the process y^* via a linear aggregation scheme:

$$y^c_{t+j/m} = \Psi_j y^c_{t+(j-1)/m} + \lambda_j y^*_{t+j/m} \tag{4.2.4}$$

where y_t is equal to the cumulator variable y^c_t for integer t, and is not observed otherwise. The above scheme, also used by Harvey (1989) and Nunes (2005), covers both stock and flow aggregation. We get the case of a stock variable by setting Ψ_j = $1(j \neq 0, m, 2m...)$ and λ_j = $1(j = 0, m, 2m, ...)$, where $1(.)$ denotes the indicator function. If we pick instead Ψ_j = $1(j \neq 1, m+1, 2m+1, ...)$ and λ_j = $1/m$ for all j, then we get a flow variable.

The high frequency process $x_{t+j/m}$ relates to the factors as follows:

$$x_{t+j/m} = \gamma_2' F_{t+j/m} + u_{2,t+j/m} \qquad \forall t, \quad j = 0, \ldots, m-1, \tag{4.2.5}$$

where γ_2 is a $n_f \times 1$ vector and:

$$d_2(L^{1/m}) u_{2,t+j/m} = \varepsilon_{2,t+j/m}, \qquad d_2(L^{1/m}) \equiv 1 - d_{12} L^{1/m} - \ldots - d_{k2} L^{k/m}. \tag{4.2.6}$$

As usual in latent factor models, factor loadings γ_1, γ_2 and the parameters of the factor dynamics are subject to identification restrictions.

The standard approach is to assume that the innovation processes (ε_k) are i.i.d. Gaussian with mean zero and variance $\sigma^2_{\varepsilon_k}$, for $k = 1, 2$. Indeed, the literature typically ignores the presence of time-varying volatility, yet the high frequency data often involve financial and other series which feature conditional heteroskedasticity. This means that the state space models are no longer Gaussian. There is a substantial literature on non-Gaussian state space models tailored for the analysis of financial returns data (see e.g. Ghysels, Harvey, and Renault (1996), Shephard (2005) and references therein). The type of models of interest to us are rather state space models with stochastic volatility in measurement equations. Hence, our analysis relates more directly to recent work by Clark (2011), Carriero, Clark, and Marcellino (2012), Carriero, Clark, and Marcellino (2013), or Marcellino, Porqueddu, and Venditti (2015).

We augment equations (4.2.5)-(4.2.6) for high frequency data with time-varying volatility:

$$\varepsilon_{2,t+j/m} \sim \mathcal{N}(0, h_{t+j/m}), \tag{4.2.7}$$

where the log volatility follows a Gaussian autoregressive process:

$$\ln h_{t+j/m} = c + \rho_{SV} \ln h_{t+(j-1)/m} + \xi_{t+j/m}, \qquad \xi_{t+j/m} \sim i.i.\mathcal{N}(0, \nu_2^2), \tag{4.2.8}$$

and parameter ρ_{SV} is smaller than 1 in absolute value. We obtain a SV-type volatility specification without common factor structure.

While our analysis relates to recent work by Marcellino, Porqueddu, and Venditti (2015), among others, as noted before, there are also subtle but important differences. In their model the factor process features stochastic volatility. Instead, in equation (4.2.7) we assume that the measurement error features stochastic volatility. When dealing with low frequency macroeconomic series exposed to factors, we think it is more appropriate to assume that those factors do not feature volatility clustering, while the high frequency series are conditionally heteroskedastic. Ideally one could consider models where SV is featured in both the observation and state equations. We leave this as a topic for future research.

Assumptions 1 and 2 (below) define the parametric models of interest in this paper. We denote by θ the vector of unknown parameters in these models.

Assumption 2. *The observable processes* (y) *and* (x) *are such that:*

$$\begin{aligned}
y^*_{t+j/m} &= \gamma_1' F_{t+j/m} + u_{1,t+j/m}, \\
d_1(L^{1/m}) u_{1,t+j/m} &= \varepsilon_{1,t+j/m}, \qquad d_1(L^{1/m}) \equiv 1 - d_{11} L^{1/m} - \ldots - d_{k1} L^{k/m}, \\
y^c_{t+j/m} &= \Psi_j y^c_{t+(j-1)/m} + \lambda_j y^*_{t+j/m}, \\
y_t &= y^c_t, \\
x_{t+j/m} &= \gamma_2' F_{t+j/m} + u_{2,t+j/m}, \\
d_2(L^{1/m}) u_{2,t+j/m} &= h^{1/2}_{t+j/m} \varepsilon_{2,t+j/m}, \qquad d_2(L^{1/m}) \equiv 1 - d_{12} L^{1/m} - \ldots - d_{k2} L^{k/m}, \\
\ln h_{t+j/m} &= c + \rho_{SV} \ln h_{t+(j-1)/m} + \xi_{t+j/m}, \qquad \forall t, \qquad j = 0, \ldots, m-1
\end{aligned}$$

where $|\rho_{SV}| < 1$*, and* (ε_1)*,* (ε_2)*,* (ξ) *are mutually independent i.i.d. Gaussian processes, with distributions* $\mathcal{N}(0, \sigma^2_{\varepsilon_1})$*,* $\mathcal{N}(0,1)$*,* $\mathcal{N}(0, \nu_2^2)$ *respectively, and independent of process* (η)*.*

4.3 Indirect Inference estimation

Estimating via Maximum Likelihood (ML) the mixed frequency models with SV presented in the previous section is rather involved. Indeed, the likelihood function involves a large-dimensional integral with respect to the latent factors path. This integral representation of the likelihood is impractical for computation of the ML estimate, and numerical filtering techniques are necessary.

In this section we introduce indirect inference estimation methods - proposed by Gouriéroux, Monfort, and Renault (1993), Smith (1993) and Gallant and Tauchen (1996) - to estimate the mixed frequency SV models. Indirect inference can be used to estimate virtually any model from which it is possible to simulate data. This obviously includes state space models. Indirect inference estimation in fact involves two types of models - a model of interest already specified in the previous section - and an auxiliary model which is easy to estimate. Both models are linked - in terms of parameter spaces - by a binding function.

4.3.1 Linear setting with known binding function

To explain our estimation approach it is worth starting with a setting where the binding function is known. This setting is provided by a linear state space model with Gaussian errors. This model is a special case of the general specification in Assumptions 1 and 2 when there is no SV.[2] In this linear state space model, the Kalman filter can be applied for prediction and filtering. Bai, Ghysels, and Wright (2013) show that for a model with a single latent factor ($n_f = 1$) having a AR(1) dynamics and persistence parameter ρ, and $m = 3$ as for instance for a monthly/quarterly mixture of data, one obtains (see Appendix 4.10.2 for details):

$$E\left[y_{t+h}|I_t^M\right] = \rho^{3h}\kappa_{3,1}\sum_{j=0}^{\infty}\vartheta^j y_{t-j} + \rho^{3h}\sum_{j=0}^{\infty}\vartheta^j x(\theta_x)_{t-j} \tag{4.3.1}$$

where I_t^M denotes the information in the available low and high frequency data up to time t, ϑ = $[(\rho - \rho\kappa_1)(\rho - \rho\kappa_2)(\rho - \rho\kappa_3)]$, and κ_i, $\kappa_{3,i}$ are steady state Kalman gain parameters. Moreover, one has:

$$x(\theta_x)_t \equiv [\kappa_{3,2} + (\rho - \rho\kappa_3)\kappa_2 L^{1/3} + (\rho - \rho\kappa_3)(\rho - \rho\kappa_2)\kappa_1 L^{2/3}]x_t \tag{4.3.2}$$

which is a parameter-driven low-frequency process composed of high-frequency data aggregated at the quarterly level.

The above equation relates to the multiplicative MIDAS regression models considered by Chen and Ghysels (2010) and Andreou, Ghysels, and Kourtellos (2013). In particular consider the following ADL-MIDAS regression:

$$y_{t+h} = \beta_y\sum_{j=0}^{K_y} w_j(\theta_y)y_{t-j} + \beta_x\sum_{j=0}^{K_x} w_j(\theta_x^1)^j x(\theta_x^2)_{t-j} + \varepsilon_{t+h} \tag{4.3.3}$$

where $w_j(\theta_y)$, $w_j(\theta_x^1)$ follow an exponential Almon scheme and

$$x(\theta_x^2)_{t-j} \equiv \sum_{k=0}^{m-1} w_k(\theta_x^2)L^{k/m}x_{t-k/m}$$

also follows an exponential Almon scheme.[3] Provided that $\rho > 0$, equation (4.3.1) is a special case of this model with $K_y = K_x = \infty$, $w_j(\theta_y) \propto \exp(\log(\vartheta)j)$, $w_j(\theta_x^1) \propto \exp(\log(\vartheta)j)$ and $w_k(\theta_x^2) \propto \exp(\theta_{x,1}^2 k + \theta_{x,2}^2 k^2)$ where $\theta_{x,1}^2$ and $\theta_{x,2}^2$ are parameters that solve the equations:

$$\begin{aligned}\log\{(\rho - \rho\kappa_3)\kappa_2/\kappa_{3,2}\} &= \theta_{x,1}^2 + \theta_{x,2}^2,\\ \log\{(\rho - \rho\kappa_3)(\rho - \rho\kappa_2)\kappa_1/\kappa_{3,2}\} &= 2\theta_{x,1}^2 + 4\theta_{x,2}^2.\end{aligned} \tag{4.3.4}$$

Equations (4.3.3) and (4.3.4) implicitly define a binding function between the parameters of the state space model and those of the MIDAS regression. Note, however, that the mapping under-identifies the parameters

[2] It corresponds to the parameter constraints $\nu_2 = 0$ and $\rho_{SV} = 1$.

[3] The constructed low-frequency regressor is estimated jointly with the other (MIDAS) regression parameters. Hence, one can view $x(\theta_x^2)_{t-j}$ as the best *aggregator* that yields the best *prediction*. This ADL-MIDAS regression involves more parameters than the usual specification involving only one polynomial.

of the state space model if we rely on a standard multiplicative MIDAS regression scheme. Moreover, the mapping is only valid for a single factor state space model with i.i.d. measurement errors. What do we do for multi-factor models or single factor models with autoregressive errors? Bai, Ghysels, and Wright (2013) show that MIDAS regressions still provide very accurate approximations, although there is no exact (underidentified) mapping.

4.3.2 Auxiliary models: U-MIDAS and ARCH

A departure from the setup in Bai, Ghysels, and Wright (2013) is that we replace equation (4.3.3) with a U-MIDAS - meaning unrestricted MIDAS - specification suggested by Foroni, Marcellino, and Schumacher (2013), namely:

$$y_{t+1} = \bar{\beta}_0 + \sum_{k=0}^{\tilde{K}_y} \beta_k y_{t-k} + \sum_{j=0}^{m\tilde{K}_x} \gamma_j x_{t-j/m} + \varepsilon_{t+1}. \tag{4.3.5}$$

Note that we estimate $\tilde{K}_y + m\tilde{K}_x + 2$ parameters (not including intercept and residual variance). When m is small, as shown by Foroni, Marcellino, and Schumacher (2013), we are able to estimate these parameters with reasonable precision using sample sizes typically encountered in economic applications. One attractive feature of U-MIDAS mispecification is the fact that estimation is numerically straightforward, as it can be performed by OLS.

Suppose we collect all the parameters of the U-MIDAS regression into the vector $\phi \in \Phi$. Assuming $dim(\theta) \leq dim(\phi) \equiv \tilde{K}_y + m\tilde{K}_x + 4$ we may be able to identify and estimate the parameters via indirect inference.[4]

Since the models of interest feature SV, we can consider as auxiliary models the following U-MIDAS regressions augmented with ARCH errors:

$$\begin{aligned} y_{t+1} &= \bar{\beta}_0 + \sum_{k=0}^{\tilde{K}_y} \beta_k y_{t-k} + \sum_{j=0}^{m\tilde{K}_x} \gamma_j x_{t-j/m} + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma_{t+1}^2) \\ \sigma_t^2 &= \omega + \sum_{k=1}^{p} \alpha_k \varepsilon_{t-k}^2 \end{aligned} \tag{4.3.6}$$

which has the advantage of being simple to implement as it only involves a linear regression specification with ARCH(p) errors. The idea for this auxiliary model is that heteroskedasticity in the high frequency data affects the residuals of the reduced form MIDAS regressions. Obviously, the ARCH model in the above equation is only estimated at low frequency, and therefore the ARCH effects may not be particularly strong.

4.3.3 Auxiliary models: mixed frequency VAR and ARCH

The auxiliary U-MIDAS regressions considered in the previous subsection do not fully exploit all features of the data since the link between latent factors and high frequency data is not being taken into account. In this subsection we remedy to this shortcoming by considering mixed frequency VAR models. It is worth noting from the start that there might be some confusion about the characterization of mixed frequency VAR

[4]Note that we added a constant and residual variance in the MIDAS regressions parameter count.

models. The analysis below serves two purposes: (1) it generalizes the U-MIDAS setup discussed so far and (2) it enables us to consider a suitable approach for state space models with stochastic volatility.

A number of authors, including Zadrozny (1988), Zadrozny (1990) and more recently Kuzin, Marcellino, and Schumacher (2011), Schorfheide and Song (2013), among others, start from a *latent* high frequency VAR process, namely:

$$\begin{pmatrix} y^*_{t+(j+1)/m} \\ x_{t+(j+1)/m} \end{pmatrix} = C_0 + \sum_{k=1}^{k^{max}} C_k \begin{pmatrix} y^*_{t+(j+1-k)/m} \\ x_{t+(j+1-k)/m} \end{pmatrix} + \begin{pmatrix} \varepsilon^{y*}_{t+(j+1)/m} \\ \varepsilon^{x}_{t+(j+1)/m} \end{pmatrix} \tag{4.3.7}$$

where $y^*_{t+j/m}$ is defined in equation (4.2.2). The above latent VAR model is related to observables via a measurement equation and therefore cast in state space framework with missing observations.

State space models are, using the terminology of Cox (1981), parameter-driven models whereas VAR models are, using again the same terminology, observation-driven models as they are formulated exclusively in terms of observable data. Ghysels (2014) introduces a class of observation-driven mixed frequency VAR models which provides an alternative to commonly used state space models involving latent processes. In addition, the mixed frequency VAR model is a multivariate extension of MIDAS regressions.

The mixed frequency VAR considered by Ghysels (2014), tailored towards the current application, can be written as follows:

$$\begin{pmatrix} x_{t+1} \\ \vdots \\ x_{t+1+(m-1)/m} \\ y_{t+1} \end{pmatrix} = \tilde{C}_0 + \sum_{k=1}^{\tilde{K}_{max}} \tilde{C}_k \begin{pmatrix} x_{t+1-k} \\ \vdots \\ x_{t+1-k+(m-1)/m} \\ y_{t+1-k} \end{pmatrix} + \begin{pmatrix} \varepsilon^1_{t+1} \\ \vdots \\ \varepsilon^m_{t+1} \\ \varepsilon^y_{t+1} \end{pmatrix}. \tag{4.3.8}$$

Hence, it involves a VAR of dimension $m + 1$ (with single high and low frequency series) where the high and low frequency data for low frequency period (quarter, say) t are stacked into a vector whose dynamics is described by a linear multivariate autoregressive structure. Note, that elements of the matrices $\tilde{C}_k$ now describe within-period (intra-quarterly) time series dependencies.[5] The stacking implies that, if we read across a particular row of the mixed frequency VAR, we have high frequency processes predicted by past high and low frequency series and vice versa.

The unrestricted VAR model in equation (4.3.8) includes $(m+1) + \tilde{K}_{max}(m+1)^2 + m(m+1)/2$ parameters which can be estimated by OLS. Ghysels (2014) proposes a parsimonious parametrization which can be estimated by Maximum Likelihood, at the expense of a higher computational cost, as the likelihood has to be maximized numerically either using classical or Bayesian techniques. Therefore, this restricted VAR is not suitable as auxiliary model, because of the heavy computational cost.

A parsimonious auxiliary model, which ensures computational speed for indirect inference estimation, can be obtained by considering an AR model for the high frequency data, which can be easily estimated by OLS. The following model will be used as the auxiliary model in our Monte Carlo simulation exercise for

[5] Most notably Granger causal patterns as discussed in Ghysels, Hill, and Motegi (2014), Ghysels, Hill, and Motegi (2016), Götz and Hecq (2014a) and Götz and Hecq (2014b).

DGPs without SV:

$$\begin{cases} y_{t+1} & = & \bar{\beta}_0 + \sum_{k=0}^{\tilde{K}_y} \beta_k y_{t-k} + \sum_{j=0}^{m(\tilde{K}_x+1)-1} \gamma_j x_{t-j/m} + \zeta_{t+1}^y \\ x_{t+(j+1)/m} & = & c_0 + \sum_{k=1}^{m(\tilde{K}_x+1)} c_k x_{t+(j+1-k)/m} + \zeta_{t+(j+1)/m}^x \end{cases} . \tag{4.3.9}$$

The first equation of this auxiliary model corresponds to the U-MIDAS specification in equation (4.3.5), while the second equation is an AR of order $m(\tilde{K}_x + 1)$ specified on the high frequency (HF) data only. The second equation can be obtained from one of the equations for the HF observables of a structural mixed frequency VAR model, where the high frequency variables do not depend explicitly on the lagged low frequency ones (see Ghysels (2014)). Model (4.3.9) can be estimated by OLS, and the correlation between the innovations ζ_{t+1}^y and $\zeta_{t+(j+1)/m}^x$, which can only be computed at low frequency, could be included as an auxiliary parameter to estimate, or can be set to zero.

In order to handle the DGP with SV, we can add ARCH-type augmentations to the auxiliary models. In particular, the complete auxiliary model used in the Monte Carlo simulation for DGPs with SV is:

$$\begin{cases} y_{t+1} & = & \bar{\beta}_0 + \sum_{k=0}^{\tilde{K}_y} \beta_k y_{t-k} + \sum_{j=0}^{m(\tilde{K}_x+1)-1} \gamma_j x_{t-j/m} + \zeta_{t+1}^y \\ x_{t+(j+1)/m} & = & c_0 + \sum_{k=1}^{m(\tilde{K}_x+1)} c_k x_{t+(j+1-k)/m} + \zeta_{t+(j+1)/m}^x, \qquad \zeta_{t+(j+1)/m}^x \sim \mathcal{N}(0, \sigma_{t+(j+1)/m}^x) \\ \sigma_{t+(j+1)/m}^x & = & \omega + \sum_{k=1}^{p} \alpha_k (\zeta_{t+(j+1-k)/m}^x)^2 \end{cases} \tag{4.3.10}$$

where the errors ζ^x and ζ^y can be correlated.

4.3.4 Estimation via Indirect Inference

The parameter vectors for the auxiliary model will be denoted respectively ϕ^{Mi} for the U-MIDAS specification appearing in equation (4.3.6), and ϕ^V for the mixed frequency VAR model in equation (4.3.8) in general - and more specifically in equation (4.3.10). Given a sample of size Tm we obtain OLS estimates $\hat{\phi}_{Tm}^{Mi}$ and $\hat{\phi}_{Tm}^V$.

We simulate mixed frequency data with the state space model in Assumptions 1 and 2, given a particular structural parameter value θ, by drawing S independent samples of size $T^S m$ from the model:

$$\begin{aligned}
F_{s,t+j/m}(\theta) &= \sum_{l=1}^{p} \Phi_l(\theta) F_{s,t+(j-l)/m}(\theta) + \eta_{s,t+j/m}(\theta) \\
\ln h_{s,t+j/m}(\theta) &= c(\theta) + \rho_{SV}(\theta) \ln h_{s,t+(j-1)/m}(\theta) + \xi_{s,t+j/m}(\theta) \\
y^*_{s,t+j/m}(\theta) &= \gamma_1(\theta)' F_{t+j/m}(\theta) + u_{s,1,t+j/m}(\theta) \\
d_1(L^{1/m}, \theta) u_{s,1,t+j/m}(\theta) &= \varepsilon_{s,1,t+j/m}(\theta) \\
y^c_{s,t+j/m}(\theta) &= \Psi_j y^c_{s,t+(j-1)/m}(\theta) + \lambda_j y^*_{s,t+j/m}(\theta) \\
y_{s,t}(\theta) &= y^c_{s,t}(\theta) \\
x_{s,t+j/m}(\theta) &= \gamma_2(\theta)' F_{s,t+j/m}(\theta) + u_{s,2,t+j/m}(\theta) \\
d_2(L^{1/m}, \theta) u_{s,2,t+j/m}(\theta) &= h_{s,t+j/m}(\theta)^{1/2} \varepsilon_{s,2,t+j/m}(\theta) \\
& \forall t = 1, \ldots, T^S, \qquad j = 0, \ldots, m-1, \qquad s = 1, \ldots, S,
\end{aligned} \tag{4.3.11}$$

where innovation processes $\eta_s(\theta)$, $\varepsilon_{s,1}(\theta)$, $\varepsilon_{s,2}(\theta)$ and $\xi_s(\theta)$ are independent i.i.d. processes with Gaussian distributions $\mathcal{N}(0, \Sigma_\eta(\theta)), \mathcal{N}(0, \sigma^2_{\varepsilon_1}(\theta)), \mathcal{N}(0, 1), \mathcal{N}(0, \nu_2^2(\theta))$. Given the S simulated samples, we compute the following estimators:

- The Indirect Inference (II) estimator of Gouriéroux, Monfort, and Renault (1993) and Smith (1993), using the U-MIDAS auxiliary model, denoted by $\hat{\theta}^{IIMi}_{TmS}$;
- The II estimator of Gouriéroux, Monfort, and Renault (1993) and Smith (1993), using the mixed frequency VAR auxiliary model, denoted by $\hat{\theta}^{IIV}_{TmS}$.

The II estimators for auxiliary models Mi and V are obtained via:

$$\hat{\theta}^{IIMi}_{TmS} = \arg\min_\theta \left(\hat{\phi}^{Mi}_{Tm} - \frac{1}{S} \sum_s \hat{\phi}^{Mi}_{Tm,s}(\theta) \right)' \Omega^{Mi} \left(\hat{\phi}^{Mi}_{Tm} - \frac{1}{S} \sum_s \hat{\phi}^{Mi}_{Tm,s}(\theta) \right) \tag{4.3.12}$$

and:

$$\hat{\theta}^{IIV}_{TmS} = \arg\min_\theta \left(\hat{\phi}^{V}_{Tm} - \frac{1}{S} \sum_s \hat{\phi}^{V}_{Tm,s}(\theta) \right)' \Omega^{V} \left(\hat{\phi}^{V}_{Tm} - \frac{1}{S} \sum_s \hat{\phi}^{V}_{Tm,s}(\theta) \right) \tag{4.3.13}$$

respectively, with $\hat{\phi}^{Mi}_{Tm,s}(\theta)$ and $\hat{\phi}^{V}_{Tm,s}(\theta)$ being the U-MIDAS and VAR auxiliary model parameter estimates for generated sample s and structural parameter value θ, and Ω^{Mi} and Ω^V being (optimal) weighting matrices.

Assumptions 1-2 and standard regularity conditions (see e.g. Gouriéroux and Monfort (1997)) imply that the indirect inference estimators are consistent and asymptotically normal as T and $S \to \infty$:

$$\sqrt{Tm}(\hat{\theta}^{EST}_{TmS} - \theta_0) \to_d \mathcal{N}(0, V^{EST}) \tag{4.3.14}$$

for $EST \equiv IIMi$ and IIV respectively, where θ_0 denotes the true value of the structural parameter.

4.3.5 EM algorithm for mixed frequency SV model

We assess the efficiency of our indirect inference estimators for the SV model with mixed frequency data by comparing their performances with that of the Maximum Likelihood (ML) estimator in a Monte Carlo experiment (see Section 4.5). Due to the latent factor processes (F) and (h) in the dynamics of the data, the likelihood function of the model involves a large-dimensional integral with respect to the latent factors path. This integral representation of the likelihood is impractical for computation of the ML estimate. We consider instead a simulation-based estimator relying on the Expectation Maximization (EM) algorithm. The smoothing distribution required in the Expectation step is computed via a particle forward-filtering/backward-smoothing algorithm.[6] In this section we describe the main steps of the procedure, and refer to Appendix B for the detailed definition of the estimation algorithm.

The Expectation-Maximization (EM) algorithm is an iterative procedure to compute numerically the ML estimate in a model with unobservable variables (Dempster, Laird, and Rubin (1977)). Let $Y_t = (y_t, x_{t-j/m}, j = 0, 1, ..., m-1)'$ be the vector of stacked observable variables (measurements) and $f_t = (F_{t-j/m}, h_{t-j/m}, j = 0, 1, ..., m-1)'$ the Markov vector of stacked latent factors, for $t = 1, ..., T$. The EM algorithm relies on the complete-observation log-likelihood function, that is the log of the joint density of observable and unobservable variables in the structural model:

$$\begin{aligned}
\mathcal{L}^*(\theta) &= \log \ell(\underline{Y_T},\ \underline{f_T}; \theta) \\
&= \sum_{t=1}^{T} \log h(Y_t | Y_{t-1}, \underline{f_t}; \theta) + \sum_{t=1}^{T} \log g(f_t | f_{t-1}; \theta),
\end{aligned}$$

where $\underline{Y_T}$ denotes the history of Y_t up to T, and similarly for $\underline{f_T}$ and $\underline{f_t}$. Here, h is the measurement density and g is the transition density in the state space representation (see Appendix 4.11.2 for the expression of $\mathcal{L}^*(\theta)$ in the mixed-frequency SV model). Let $\hat{\theta}_{Tm}^{EM,(i)}$ be the estimate of parameter θ at iteration i of the EM algorithm. The update $i \to i+1$ consists of two steps:

1. *Expectation (E) step*. Compute function $Q(\theta|\tilde{\theta})$, with $\tilde{\theta} = \hat{\theta}_{Tm}^{EM,(i)}$, where:

$$Q\left(\theta|\tilde{\theta}\right) = E_{\tilde{\theta}}\left[\mathcal{L}^*(\theta)|\underline{Y_T}\right]$$

   and $E_{\tilde{\theta}}\left[\cdot|\underline{Y_T}\right]$ denotes the expectation w.r.t. the conditional distribution of $\underline{f_T}$ given $\underline{Y_T}$ for parameter value $\tilde{\theta}$.

2. *Maximization (M) step*. Compute the estimate for iteration $i+1$ as:

$$\hat{\theta}_{Tm}^{EM,(i+1)} := \arg\max_{\theta} Q\left(\theta|\hat{\theta}_{Tm}^{EM,(i)}\right).$$

The iteration is performed until a criterion for numerical convergence of the estimate is met, and $\hat{\theta}_{Tm}^{ML} = \hat{\theta}_{Tm}^{EM,(\infty)}$. The details for the E-step and the M-step in the mixed frequency SV model are provided in Appendix 4.11.3.

[6] Other approaches have been proposed in the literature to implement the MLE in nonlinear state space models with SV and could be adapted to our mixed frequency framework, for instance the Monte Carlo ML approach in Sandmann and Koopman (1998).

The E-step in the EM algorithm requires the smoothing distribution of the unobservable factor path for given parameter value $\tilde{\theta}$ to compute the conditional expectation $E_{\tilde{\theta}}\left[\cdot|\underline{Y_T}\right]$. This smoothing distribution cannot be characterized analytically for a nonlinear state space specification as the mixed frequency SV model. We approximate the smoothing distribution via a large sample of draws from it, called particles. The smoothing algorithm we adopt uses a sample of particles from the filtering distribution as an input. Specifically, for the E-step of the i-th iteration in the EM algorithm, we generate samples $f_{t+j/m}^{s,(i)} = (F_{t+j/m}^{s,(i)}, h_{t+j/m}^{s,(i)})'$, $s = 1, ..., S$, from the filtering distribution of the latent factors at each date $t + j/m$, for parameter value $\hat{\theta}_{Tm}^{EM,(i)}$. For this task we use a sequential algorithm based on the auxiliary particle filter method running from the first sample date to the last sample date. We refer to Pitt and Shephard (1999) for the auxiliary particle filter; see also e.g. Douc, Moulines, and Olsson (2009), Carvalho, Johannes, Lopes, and Polson (2010), Doucet (2010), Lopes and Tsay (2011), Creal (2012), Kantas, Doucet, Singh, Maciejowski, and Chopin (2015) for recent developments and applications. The algorithm is described in detail in Section B.4.2. Then, we use a backward algorithm to generate sample paths $(\tilde{f}_{t+j/m}^{s,(i)}, \forall t, j)$, $s = 1, ..., S$, from the smoothing distribution; see e.g. Kim and Stoffer (2008) and Godsill, Doucet, and West (2004). Appendix 4.11.4 provides the detailed simulation procedure. The sample paths $(\tilde{f}_{t+j/m}^{s,(i)}, \forall t, j)$, $s = 1, ..., S$, are approximate draws from the distribution of $(f_{t+j/m}, \forall t, j)$ given $\underline{Y_T}$ for parameter value $\hat{\theta}_{Tm}^{EM,(i)}$ when the number of particles S is large. We use averages across these sample paths to approximate the conditional expectation $E_{\tilde{\theta}}\left[\cdot|\underline{Y_T}\right]$ for $\tilde{\theta} = \hat{\theta}_{Tm}^{EM,(i)}$.

4.4 Filtering via reprojection and nowcasting

State space models do not only involve parameter estimation but also filtering of the latent states, for which the Kalman filter is the standard scheme in the linear Gaussian case. In this section we present alternative methods which easily extend to, say, the non-Gaussian case involving stochastic volatility. Our approach relies on the reprojection method of Gallant and Tauchen (1998) to produce filtering estimates of the latent factors.

The procedure is fairly simple to implement. Let $\hat{\theta}_{TmS}^{EST}$ be the parameter estimate obtained by one of the Indirect Inference estimators introduced in Section 3.4. We start again with simulating a long sample of size $T^{reproj}m$, say, from the model of interest as in equation (4.3.11), using parameter value $\theta = \hat{\theta}_{TmS}^{EST}$. Then, the simulated sample is used to estimate a specification for the conditional expectation of the latent factors given the observable data. Finally, the estimated specification for this conditional expectation is applied to the original sample of observations y, x, and used as a filter.

To develop further insight in the methodology, we start with the Gaussian case (the model in Assumptions 1 and 2 without SV, i.e. with $\rho_{SV} = 1$ and $\nu_2 = 0$). Next, we discuss the filtering algorithm for the non-Gaussian case with stochastic volatility.

In a Gaussian linear state space model, the conditional expectation of the latent factor given the measurements is linear (see equation (4.10.15) in Appendix A). In our mixed frequency setting, in analogy to equation (4.3.5), this remark suggests estimating a U-MIDAS regression on the simulated sample:

$$\begin{aligned} F_{t+j/m}(\hat{\theta}_{TmS}^{EST}) &= b_0(\hat{\theta}_{TmS}^{EST}) + \sum_{k=0}^{\tilde{K}_y} b_k(\hat{\theta}_{TmS}^{EST}) y_{t-k}(\hat{\theta}_{TmS}^{EST}) \\ & + \sum_{k=0}^{m\tilde{K}_x} c_k(\hat{\theta}_{TmS}^{EST}) x_{t+(j-k)/m}(\hat{\theta}_{TmS}^{EST}) + \epsilon_{t+j/m}(\hat{\theta}_{TmS}^{EST}) \\ & t = 1, 2, ..., T^{reproj}, \quad j = 0, 1, ..., m-1, \end{aligned} \tag{4.4.1}$$

which amounts to regressing latent factors onto observables. Note that the observables have a nowcasting feature, i.e. contemporaneous period $t + j/m$ high frequency data is used. Once we have the parameters of the above regression, we can apply the scheme to observed data y and x and therefore use it as a filter. We denote by $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{TmS}^{EST})$ the reprojection factor values.

Likewise, the mixed frequency VAR framework of Ghysels (2014) could be modified to perform the task as filter, namely we run the system of regressions:

$$\begin{aligned} \bar{C}(\hat{\theta}_{TmS}^{EST}) \begin{pmatrix} F_{t+1}(\hat{\theta}_{TmS}^{EST}) \\ \vdots \\ F_{t+1+(m-1)/m}(\hat{\theta}_{TmS}^{EST}) \\ y_{t+1}(\hat{\theta}_{TmS}^{EST}) \end{pmatrix} &= \tilde{C}_0(\hat{\theta}_{TmS}^{EST}) + \sum_{k=1}^{\tilde{K}_{max}} \tilde{C}_k(\hat{\theta}_{TmS}^{EST}) \begin{pmatrix} x_{t+1-k}(\hat{\theta}_{TmS}^{EST}) \\ \vdots \\ x_{t+1-k+(m-1)/m}(\hat{\theta}_{TmS}^{EST}) \\ y_{t+1-k}(\hat{\theta}_{TmS}^{EST}) \end{pmatrix} \\ & + \begin{pmatrix} \epsilon_{t+1}^{1}(\hat{\theta}_{TmS}^{EST}) \\ \vdots \\ \epsilon_{t+1}^{m}(\hat{\theta}_{TmS}^{EST}) \\ \epsilon_{t+1}^{y}(\hat{\theta}_{TmS}^{EST}) \end{pmatrix} \end{aligned} \tag{4.4.2}$$

where $\bar{C}(\hat{\theta}_{TmS}^{EST})$ is a lower triangular matrix to accommodate nowcasting - see Ghysels (2014) for further details. Here again, once we estimate the system of equations over a long simulated sample, we can treat the resulting estimates as weights for a filtering scheme.

In the nonlinear state space model with stochastic volatility, the conditional expectation of the latent factors given the current and past values of the observable variables is no more linear in the conditioning variables. Therefore, in such framework the regressions in (4.4.1) and (4.4.2) do not provide exact filters (up to a truncation of the number of lags). However, we can interpret these regressions as numerically feasible linear approximations of the unknown exact filter for the latent factor F in the conditional mean. A second-order approximation is obtained by including quadratic terms in low and high frequency observations. Similar approximate filters can be developed for the stochastic volatility factor h. In this case, the filter can be based on squared measurement errors. For instance, in a model without AR effects in the measurement errors at high frequency (to simplify), we can run the regression:

$$h_{t+j/m}(\hat{\theta}_{TmS}^{EST}) = \bar{d}_0 + \sum_{k=0}^{\tilde{K}_u} \bar{d}_k (u_{2,t+(j-k)/m}(\hat{\theta}_{TmS}^{EST}))^2 + \epsilon_{t+j/m}^{u}(\hat{\theta}_{TmS}^{EST}), \tag{4.4.3}$$

possibly including also higher-order terms.

In the linear Gaussian case, we can make direct comparisons of the filters based on reprojection with the Kalman filter in order to gauge the reliability of the proposed method. In the non-Gaussian case with stochastic volatility, a benchmark for comparison is obtained by first estimating the model by Monte Carlo EM as described in Section 3.5, and then compute the filtered factor value $\hat{f}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML}) = [\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML}), \hat{h}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML})]'$, say, by averaging the particles $f_{t+j/m}^{s}$, with $s = 1, ..., S$, from the filtering distribution for parameter value $\hat{\theta}_{Tm}^{ML}$. We perform these comparisons in the Monte Carlo simulations presented in Section 5. There, we keep the reprojections quite simple in fact, namely we implement the filter for F in equation (4.4.1) in both the Gaussian and stochastic volatility settings, and we use a filter for volatility factor h based on squared residuals such as (4.4.3) in the latter setting. These filters could be improved upon by considering the setup in equation (4.4.2), and adding higher-order terms in the SV case.

4.5 Monte Carlo Simulations

We conduct a Monte Carlo simulation to appraise the small and large sample properties of the indirect inference procedures proposed in earlier sections. A first subsection covers the design of the simulations. A second subsection covers the Gaussian state space model where the Kalman filter and maximum likelihood are the natural benchmarks. In a final subsection we consider non-Gaussian cases with stochastic volatility, where we compare our indirect inference procedure with a simulation-based EM algorithm.

4.5.1 Design

We consider three designs for the MC experiments. In all of them we have $m = 3$, corresponding to - for instance - a mixture of monthly and quarterly data, and stock sampling of the low frequency variable. In the first MC design, we consider a linear Gaussian state space model. The DGP has a single Gaussian AR(1) latent factor process ($n_f = 1$), and Gaussian AR(1) measurement errors for both the high and low frequency data, with the same persistence parameter.

DGP 1: Single factor linear Gaussian state space model

The data (y) and (x), and the single latent factor (F), are such that:

$$\begin{aligned}
F_{t+j/3} &= \rho F_{t+(j-1)/3} + \eta_{t+j/3}, \\
y^*_{t+j/3} &= \gamma_1 F_{t+j/3} + u_{y,t+j/3}, \\
u_{y,t+j/3} &= d \cdot u_{y,t+(j-1)/3} + \sigma_y \varepsilon_{y,t+j/3}, \\
x_{t+j/3} &= \gamma_2 F_{t+j/3} + u_{x,t+j/3}, \\
u_{x,t+j/3} &= d \cdot u_{x,t+(j-1)/3} + \sigma_x \varepsilon_{x,t+j/3}, \qquad t = 1, ..., T, \ j = 0, 1, 2,
\end{aligned}$$

where the low frequency variable y is stock-sampled, and (η), (ε_y) and (ε_x) are mutually independent i.i.d. standard Gaussian processes. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $d = 0$, $\sigma_y = \sigma_x = 1$. We consider two values for the persistence of the latent factor, that are $\rho = 0.5$ and $\rho = 0.9$.

The number of structural parameters in DGP1 is 6. In each Monte Carlo simulation, we draw from this DGP samples of sizes $T = 100$ (corresponding to 25 years of quarterly data), $T = 200$ and $T = 500$. We perform 1000 Monte Carlo repetitions. On each simulated sample we compute the Indirect Inference (II) estimator $\hat{\theta}_{TmS}^{IIV}$ of Gouriéroux, Monfort, and Renault (1993) and Smith (1993) as described in Section 3.4, and the associated reprojections $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{TmS}^{IIV})$ as described in Section 4. The auxiliary model is a U-MIDAS regression for the low frequency data with $\tilde{K}_x = \tilde{K}_y = 3$ and an AR(9) process for the high frequency data (see equation (4.3.9)). This auxiliary model has 30 parameters and yields an overidentifed II setting. Instead of running S simulations from the DGP of length T, we simulate a unique long path, i.e. we set $S = 1$ and $T^S = 50000$. Moreover, we use the identity weighting matrix. The reprojection of the latent factor is computed by regression on a simulated sample of size $T^{reproj} = 100000$.

In this linear Gaussian state space model, the MLE estimator of the model parameters $\hat{\theta}_{Tm}^{ML}$ and the Kalman filter of the latent factor values - which we denote $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML})$ - serve as the natural benchmark. We compute the Kalman filter and the ML estimates using the algorithm presented in Appendix A.

In the second Monte-Carlo design, the DGP is a two-factor linear state space model ($n_f = 2$). The two latent factors follow independent AR(1) processes, with same autoregressive parameter.

DGP 2: Two-factor linear Gaussian state space model

The data (y) and (x), and the bivariate latent factor (F), are such that:

$$
\begin{aligned}
F_{t+j/m} &= \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} F_{t+(j-1)/3} + \eta_{t+j/3}, \\
y^*_{t+j/3} &= \gamma_1' F_{t+j/3} + u_{y,t+j/3}, \\
u_{y,t+j/3} &= d \cdot u_{y,t+(j-1)/3} + \sigma_y \varepsilon_{y,t+j/3}, \\
x_{t+j/3} &= \gamma_2' F_{t+j/3} + u_{x,t+j/3}, \\
u_{x,t+j/3} &= d \cdot u_{x,t+(j-1)/3} + \sigma_x \varepsilon_{x,t+j/3}, \qquad t = 1, ..., T, \ j = 0, 1, 2,
\end{aligned}
$$

where the low frequency variable y is stock-sampled, and (η), (ε_y) and (ε_x) are mutually independent i.i.d. Gaussian processes, with distribution $\mathcal{N}(0, I_2)$ for (η), and distribution $\mathcal{N}(0, 1)$ for (ε_y) and (ε_x). The true values of the parameters are $\rho = 0.9$, $\gamma_1 = (1, 0.2)'$, $\gamma_2 = (0.2, 1)'$, $d = 0$, $\sigma_y = \sigma_x = 1$.

The number of structural parameters in DGP2 is 8. The sample sizes are $T = 100$, $T = 200$ and $T = 500$. We compute the II estimator $\hat{\theta}_{TmS}^{IIV}$ of Gouriéroux, Monfort, and Renault (1993) and Smith (1993) and the associated reprojections $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{TmS}^{IIV})$ with the same auxiliary model and the same simulation length as for DGP1. We also compute the MLE $\hat{\theta}_{Tm}^{ML}$ and the Kalman filter estimates $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML})$ with the algorithm in Appendix A.

The third DGP is a mixed frequency state space model with stochastic volatility. This DGP features a single Gaussian AR(1) factor in the mean of high frequency and low frequency observables ($n_f = 1$). The measurement error of the low frequency variable is a Gaussian AR(1) process. The measurement error of the high frequency variable is a conditionally heteroskedastic process.

The number of structural parameters in DGP3 is 8. The SV specification for the high frequency innovations in equations (4.5.1) and (4.5.2) is a reparametrization of the one proposed in equations (4.2.7) and (4.2.8). This specification is analogous to the one used by Monfardini (1998), Marcellino, Porqueddu, and Venditti (2015) and Clark (2011), among others. In particular, here h is the log volatility process, and is normalized to have mean zero. In this parameterization, both latent factors have a linear autoregressive dynamics.

DGP 3: Stochastic volatility model

The data (y) and (x), and the scalar latent factors (F) and (h), are such that:

$$
\begin{aligned}
F_{t+j/3} &= \rho F_{t+(j-1)/3} + \eta_{t+j/3}, \\
y^*_{t+j/3} &= \gamma_1 F_{t+j/3} + u_{y,t+j/3}, \\
u_{y,t+j/3} &= d \cdot u_{y,t+(j-1)/3} + \sigma_y \varepsilon_{y,t+j/3}, \\
x_{t+j/3} &= \gamma_2 F_{t+j/3} + \sigma_x \exp\left\{\frac{1}{2} h_{t+j/3}\right\} \varepsilon_{x,t+j/3}, \qquad (4.5.1) \\
h_{t+j/3} &= \rho_{SV} h_{t+(j-1)/3} + \nu \cdot \xi_{t+j/3}, \qquad t = 1, ..., T,\ j = 0, 1, 2, \qquad (4.5.2)
\end{aligned}
$$

where the low frequency variable y is stock-sampled, and (η), (ε_y), (ε_x) and (ξ) are mutually independent i.i.d. standard Gaussian processes. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $d = 0$, $\sigma_y = \sigma_x = 1$, $\rho_{SV} = 0.95$, $\nu = 0.3$. We consider two values for the persistence of the latent factor in the conditional mean, that are $\rho = 0.5$ and $\rho = 0.9$.

Again, the sizes of the simulated samples are $T = 100$, $T = 200$ and $T = 500$. We have tried different auxiliary models for the indirect inference procedure, including GARCH(1,1) for the squared high frequency residuals, an AR(10) model on the logarithm of the squared high frequency residuals and an AR(10) model on the logarithm of the squared high frequency observables (similarly as in Monfardini (1998)). Barigozzi, Halbleib-Chiriac, and Veredas (2014) show that the GARCH(1,1) model is the best auxiliary model for estimating a stochastic volatility model with Indirect Inference, in the sense that it provides the best trade-off between efficiency and estimation noise. The GARCH(1,1) auxiliary model reduces, however, the computational speed of the indirect inference estimator, as it requires estimation via maximum likelihood. We therefore prefer an AR-ARCH specification in the auxiliary model, since this allows for estimation via a simple two-step approach based on OLS regressions. Specifically, we compute the indirect inference estimator $\hat{\theta}^{IIV}_{TmS}$ of Gouriéroux, Monfort, and Renault (1993) and Smith (1993) as described in Section 3.4, using the auxiliary model in equation (4.3.10), with $\tilde{K}_x = \tilde{K}_y = 4$ in the U-MIDAS regression for the low frequency data, and an AR(9)-ARCH(10) specification for the high frequency data.[7] We compare the distribution of our indirect inference estimator with the distribution of the MLE in the Monte Carlo simulations. In the nonlinear state space model of DGP 3, we implement the MLE via a simulation-based EM algorithm as described in Section 4.3.5 (see Appendix B for the detailed algorithm).

4.5.2 Monte Carlo results in the linear Gaussian state space model

Tables 4.1 through 4.3 report the results for the linear Gaussian state space models in DGP1 and DGP2. For each combination of DGP parameters and sample size, we provide the results of the Indirect Inference (II) procedure for parameter estimation and filtering of the latent factor path. As a benchmark, we also provide the estimation and filtering results using the Maximum Likelihood (ML) procedure based on the Kalman

[7] In this paper we do not consider the moment matching procedure of Gallant and Tauchen (1996). However, adopting their procedure, which is computationally even more attractive, could make the use of GARCH-type auxiliary models more attractive. As the Gallant-Tauchen procedure is based on the score, it would not require iterated ML estimates (see for instance, Sentana, Calzolari, and Fiorentini (2008)).

filter.[8]

In Table 4.1, we consider DGP1 where the single latent factor process is mildly persistent with autocorrelation ρ = 0.5. The finite sample performance of the II estimator is remarkably good. First, it has only a small bias for most configurations. Second - as expected - the ML estimator based on the Kalman filter is more efficient, but the efficiency loss of the II estimator is rather limited. The bias of both the II and MLE is more pronounced for parameter σ_y, that is the volatility of the low frequency measurement error. For this parameter, the efficiency loss of the II estimator compared to MLE is a bit larger. As expected, the dispersions of the estimators decrease with the sample size T. Moreover, the reprojection procedure provides rather accurate estimates of the latent factor values. Indeed, the average correlation between true and filtered factor values is about 0.80 for all sample sizes, which is close to the performance of the Kalman filter.

In unreported MC results we compared the performance of the above II estimator - which uses the U-MIDAS/AR auxiliary model for high/low frequency data - with the performance of the II estimator using only the low-frequency U-MIDAS specification as auxiliary model. The II estimator of the standard deviation parameter for the high-frequency data σ_x based on the low-frequency U-MIDAS auxiliary model has a large bias. As shown in Table 4.1, this problem does not arise when we include high-frequency data in the auxiliary model via the mixed frequency VAR specification. These findings confirm the intuition that using data at both frequencies provides a more informative auxiliary model.

In Table 4.2, the autocorrelation of the latent factor in DGP1 is set equal to ρ = 0.9. Both the ML and II estimators have smaller dispersions in this MC design compared to Table 4.1. This effect is due to the more favorable signal-to-noise setting when ρ is changed from 0.5 to 0.9 in our parameterization of the DGP. Indeed, with ρ = 0.9 the factor has a larger unconditional variance relative to the noise variance, which is fixed across the two cases. Hence, the signal-to-noise ratio is larger for the DGP in Table 4.2 compared to Table 4.1.

In Table 4.3 we report the simulation results for DGP2, which features two latent factors, with loadings equal to $\gamma_1 = (1.0,\ 0.2)'$ and $\gamma_2 = (0.2,\ 1.0)'$. Compared to the one-factor case in Tables 4.1 and 4.2, the loadings are estimated rather precisely, with the dispersion of the loadings equal to 0.2 being larger than that of the loadings equal to 1.0. Also in this case we find that the II estimator has a very good performance, with the exception of the estimator of the low frequency volatility σ_y, which has a bias of around 20% for small sample sizes ($T = 200, 100$), and a large dispersion. Nevertheless, the reprojection procedure produces accurate estimates of both factors. As expected, the factor which loads mainly on the high frequency observables (that is F_2) is estimated more precisely (average correlation with the true factor equal to 0.88 for $T = 100$) than the factor which loads mainly on the low frequency observables (average correlation with the true factor equal to 0.74 for $T = 100$).

Overall, the results in Tables 4.1 through 4.3 are remarkably impressive, since they show that the performance of the II procedure is rather close to the efficient benchmark in the linear Gaussian state space model.

4.5.3 Monte Carlo results for the state space stochastic volatility model

We now consider the more challenging state space model with stochastic volatility in DGP3. Tables 4.4 and 4.5 report the results of Monte Carlo simulations comparing the II estimator with the MLE (implemented

[8] All Monte Carlo simulations in Section 5 have been performed using Matlab 7.10.0 (R2010a) on a laptop with a 1.60 GHz processor and 4 GB RAM. Optimization problems involved in parameter estimation have been solved using the Matlab procedure 'fminunc'.

via a simulation-based EM algorithm) for sample sizes $T = 500$ and $T = 200$ respectively. These tables compare the two estimation methods on the basis of (a) statistical criteria - mean/bias/quantiles of sampling distributions, (b) filtering accuracy - both for conditional mean and volatility factors, and (c) computational time. Compared to the linear Gaussian state space model in DGP1, the structural model now has two additional parameters, which are the autoregressive coefficient ρ_{SV} and the volatility parameter ν of the log stochastic volatility process. A first encouraging finding is that the estimation results for parameters γ_1, γ_2, ρ, d, σ_y, σ_x are comparable to those of the Gaussian state space model displayed in Tables 4.1 and 4.2, with slightly larger dispersions in Tables 4.4 and 4.5, as expected. The latter effect is more pronounced for parameter d, the autoregressive coefficient of the low frequency measurement error, for both the II estimator and the MLE. The stochastic volatility parameters ρ_{SV} and ν are estimated with rather small biases. Note that sample size $T = 200$ corresponds to 600 high frequency observations, and for such sample sizes the estimation of ARCH and SV specifications can be inaccurate, even in absence of latent factors in the mean. Yet, comparing the distributions of II and ML estimates, we observe that also in the stochastic volatility case the efficiency loss of the former estimator is limited.

It is worth noting that the reprojection method provides rather accurate estimates of the latent factor values also in the stochastic volatility model. Results are less good for the log volatility factor (average correlation between estimated and true factor values equal to 0.55 for sample size $T = 500$ in the design with $\rho = 0.5$). This result is not surprizing, because there is no obvious choice for the transformations of the observable variables, whose linear combination provides the best approximation of the conditional expectation of the volatility factor in this nonlinear state space model. In Tables 4.4 and 4.5 we use current and past values of log squared high frequency residuals, but other choices could yield better results.

The II procedure provides a substantial reduction in computational time compared to the simulation-based EM procedure used to obtain the ML estimates. For instance, the computation of the II estimates for one Monte Carlo repetition in the stochastic volatility design with $\rho = 0.5$ and sample size $T = 200$ takes on average about 18 minutes, against the 24 minutes required on average for the ML estimates. The difference is larger with sample size $T = 500$, for which the average computational times are 16 minutes for II and 61 minutes for ML. Here, the computational time for the II procedure is less than 21 minutes in 75% of the MC replications, while the computational time for ML is more than one hour in more than 25% of the MC replications. Sample sizes such as $T = 500$ or even larger are often encontered in financial datasets, if the lower frequency is weekly or monthly.

To summarize the findings of the MC simulations with the stochastic volatility design, the II procedure offers a substantial gain in computational time compared to the ML procedure implemented via Monte Carlo EM, while the cost in terms of efficiency loss is limited.

4.6 Empirical study

We present an empirical application of our model to the problem of forecasting at short horizons the Euro-area quarterly GDP growth using monthly macroeconomic indicators.

4.6.1 Data and model specification

The dataset is the same as the one considered in the empirical study of Marcellino, Porqueddu, and Venditti (2015).[9] The data consists of the quarterly GDP growth rates for the Euro-area (GDP) observed from 1991-Q1 to 2011-Q1, and the monthly observations for the same period, i.e. from 1991-M1 to 2011-M3, for the following 8 macroeconomic indicators: (1) the aggregate European Industrial Production index for all sectors of the European economy: IP, (2) the European Industrial Production index for "Pulp and Paper sector": IP-Pulp/Paper, (3) the Germany IFO Business Climate Index: IFO, (4) the Euro-area Economic Sentiment Index: ESI, (5) the Euro-area Composite Purchasing Manager Index: PMI, (6) the bilateral dollar-euro exchange rate, measured as year-on-year percentage growth: EXC, (7) the difference between 3-month and 10-year US Treasury bond yield: SPR, and (8) the University of Michigan consumer sentiment index for the US: MICH. In line with the empirical study of Bai, Ghysels, and Wright (2013), we consider the first difference of the series (3) to (8) to induce stationarity, and we normalize all series by their full sample mean and standard deviation.[10]

We estimate the mixed-frequency stochastic volatility model defined as DGP 3 in Section 4.5 and the linear Gaussian factor model defined as DGP 1 on eight different pairs of mixed frequency observables.[11] In each model we include GDP as the low frequency observable, and one of the eight monthly indicators listed above as the high frequency variable. We assume the presence of one high frequency latent factor ($n_f = 1$), and that the observed quarterly GDP is the sum of three unobservable monthly growth rates: $y_t = y_t^* + y_{t-1/3}^* + y_{t-2/3}^*$. Thus, we have $m = 3$ and the low frequency variable is flow sampled. We estimate the SV model by the Indirect Inference (II) procedure, using the same auxiliary model as in the MC simulations of Section 4.5, and deploy the II estimates in the reprojection procedure to filter the latent factors. We estimate the Gaussian state space model without SV by adapting the Kalman filter for periodic state space models proposed in Bai, Ghysels, and Wright (2013) to accommodate flow sampling (see Appendix C).

4.6.2 Estimation and in-sample explanatory power

Before performing the forecasting exercise, we discuss the estimation results of the models for the entire data sample ending in 2011-Q1. In Table 4.6 we report the values of the R^2 of the regression of both GDP and the five monthly indicators (1)-(5) on the filtered values of the latent factor F in each model.[12] For all five considered models, the factor explains a substantial fraction of the variability of both the GDP and

[9] We thank M. Marcellino, M. Poqueddu and F. Venditti for sharing their dataset with us.

[10] Augmented Dickey-Fuller tests failed to reject the null hypothesis of a unit root for series (3) to (8).

[11] Differently from the specification of DGP 1 in Section 4.5, we allow the autoregressive parameters d_x and d_y of the idiosyncratic error terms u_x and u_y to be different.

[12] The regression of GDP on the factor is a special type of MIDAS regression, in which we regress the value of GDP growth at the end of quarter t on the sum of the filtered values of the factor in the months of the same quarter:

$$\begin{aligned} y_t &= y_t^* + y_{t-1/3}^* + y_{t-2/3}^* \\ &= \gamma_1(F_t + F_{t-1/3} + F_{t-2/3}) + u_{y,t} + u_{y,t-1/3} + u_{y,t-2/3}. \end{aligned}$$

the respective HF monthly indicator. For the factor model with stochastic volatility estimated with the IP indicator, the common factor explains 74% of the variability of GDP and 48% of Industrial Production. When the factor model without SV is estimated on the same data, the explanatory power of the common factor for GDP is slightly higher, as the R^2 increases to 82%. On the other hand, the explanatory power for GDP (resp. the HF indicator) of the factor extracted using the IFO and ESI survey indices, are higher (resp. lower) for the SV model than for the linear Gaussian one.[13] The factor extracted using the IP-Pulp/Paper index explains only 11% of the variability of this HF index for both models, but this is not surprising as the Pulp and Paper sector represents a small fraction of the total Industrial Production. Moving to the estimation of our model using the EXC, SPR and MICH indicators, both in the full sample and in the shorter subsamples considered in the forecasting exercise below, the regressions produce loadings of the HF observables on the factor close to zero, and a filtered factor uncorrelated with the corresponding HF variable, with no forecasting power for GDP. For this reason we report only results for the five monthly indicators (1)-(5). It should be noted that our mixed frequency model admits only the contemporaneus impact of the common latent factor on the HF variable, and the impact of the factor values within a quarter on the flow-sampled LF variable. It could be that more general specifications, such as a factor model in which the observables load on more lags of the latent factor - on the last 12 months, for instance - might be more appropriate to assess the forecasting power for the European GDP of the 2 US macroeconomic indicators SPR and MICH, and the Euro-dollar exchange rate EXC.[14]

Figures 4.1 to 4.5 display the time series of the observable variables used to estimate the factor models, and the filtered mean and stochastic volatility factor paths obtained via reprojection, corresponding to HF indicators (1) to (5). Visual inspection of the estimated factor paths $\hat{F}$ in Panels (c) of the five figures reveal commonalities across models, like the major drop and the successive rebound following the financial crisis of 2008. Nevertheless, the relative size of this drop appears to be more pronounced for the two IP series than for IFO, ESI and PMI.[15] The trajectories of the filtered stochastic volatility factor are represented in Panels (d) of Figures 4.1 to 4.5. For all but one of the monthly indicators, the estimated idiosyncratic volatility factor oscillates around zero before 2008 and then increases to values larger than 1.5, indicating that the idiosyncratic volatility of the monthly macroeconomic series more than doubled during the recent financial crisis.[16] Only the IP-Pulp/Paper idiosyncratic volatility shows a different behavior, being much larger in the first half of the sample, than in the second one. We stress that we do not impose any dependence structure between the mean factor F and the stochastic volatility factor h specific to each HF series, and this fact might be relevant for the situations like the one of the IP-Pulp/Paper monthly indicator in which the large drop of the mean factor in 2009, corresponding to the drop in DGP, is not associated with a spike in the volatility of the high frequency index.

On the other hand, each high frequency indicator is regressed only on the contemporaneous value of the factor. In Table 4.6 we report R^2 instead of the values of the loadings of the factor on observables, as they are more easily interpretable.

[13]These results are robust to the choice of the starting point of the estimation algorithms, which did not show convergence problems for the series reported in the tables.

[14]See Marcellino, Porqueddu, and Venditti (2015), in particular Section B of their online Appendix, for an example of a richer dependence structure between the observables and the factor. Nevertheless, this result is not surprising as the loadings of EXC, SPR and MICH on their common latent factor summarizing the current state of the business cycle, are much smaller - in absolute value - than the loadings of the other five macroeconomic indicators.

[15]As the estimated loadings of the latent factor F on the observables have positive signs, a drop in the factor is associated with a drop in both GDP and the monthly indicator in the same quarter.

[16]Indeed, the value of volatility $\exp(0.5 \cdot h_t)$ increases from 1 to more than 2.1, when factor h_t goes from 0 to 1.5.

4.6.3 Forecasting

As the in-sample estimates of our five factor models are different, we expect the models to have different forecasting power for the GDP. Similarly to Marcellino, Porqueddu, and Venditti (2015), we perform an out-of-sample forecasting exercise where at the end of each quarter we estimate the models with and without stochastic volatility, and use them to forecast GDP up to an horizon of $H = 4$ quarters ahead of the estimation sample final date. The first estimation window is from 1991-Q1 to 2005-Q4, and is recursively expanded up to 2010-Q4. In Table 4.7 we report the Root Mean Squared Forecasting Errors (RMSFE) as ratios to the RMSFE of a forecasting model assuming constant growth of the GDP. An entry below one in Table 4.7 indicates that the factor model outperforms the naive constant growth benchmark. This choice allows us to have comparable results across different models, forecasting horizons, but also with the results of (Marcellino, Porqueddu, and Venditti, 2015, Figures 8 and 9).

We immediately note that the forecasting ability of all models, relative to the naive benchmark, is limited to short horizons up to 2 quarters ahead. Indeed, all the RMSFE ratios reported in Table 4.7 are very close to, or even larger than, 1 for forecasting horizons $H = 3, 4$ quarters. Note that Marcellino, Porqueddu, and Venditti (2015) report the RMSFE ratios for a maximum of 7 months ahead, and that the RMSFE for 6 months (i.e. $H = 2$ quarters) is always very close to 1 for all their models. For the factor models estimated using the aggregate Industrial Production index, in Table 4.7 the linear Gaussian model seems to outperform the model incorporating SV when used to forecast GDP at 1 quarter horizon, as the RMSFE ratio for the latter model is 0.7, which is smaller than the value slightly below 0.8 reported by both our SV model, and by Marcellino, Porqueddu, and Venditti (2015) for all their specifications. On the other hand, the results are completely different when considering the 1 quarter ahead forecasting accuracy of our SV models estimated on the IFO and ESI indexes (RMSFE around 0.7 for $H = 1$, and 0.9 for $H = 2$, in both models), which clearly outperform our models without SV (only the IFO model has a RMSFE lower than 1, equal to 0.9 for $H = 1$) and the model of Marcellino, Porqueddu, and Venditti (2015) (RMSFE around 0.8 for $H = 1$, and around 1.0 for $H = 1$) at both 1 and 2 quarters ahead horizons. Finally, the models estimated on IP-Pulp/paper and PMI show some forecasting power at 1 quarter horizon, yet with larger RMSFE compared to all models discussed above.

Overall, the results of this empirical exercise demostrate the importance of considering stochastic volatility when estimating mixed frequency factor models both for the in-sample explanatory power of the extracted factors, which might be important when constructing coincident indexes of the economy as in Marcellino, Porqueddu, and Venditti (2015), and for the out-of sample predictive ability of the estimated model. Moreover, the estimation of our SV models on GDP (LF series) and only one monthly macroeconomic indicator (HF series) showed, that the forecasting accuracy of the different macroeconomic variables can be different across different variables, horizons and model specifications.

There is scope for even further improvements - despite the fact that some of our models already outperform the approach suggested by Marcellino, Porqueddu, and Venditti (2015) - using the same data and sample configurations. In our approach, we followed Bai, Ghysels, and Wright (2013) who focused exclusively on bivariate specifications, whereas Marcellino, Porqueddu, and Venditti (2015) build one joint model for the eight series considered. We have in principle 8 forecasts obtained from the paired bivariate models - with some outperforming and some mostly at par with the single large model they consider. In light with Andreou, Ghysels, and Kourtellos (2013) we could further improve the forecasting output by constructing forecast combinations of our 8 predictions - ultimately producing a single combination forecast. Since the scope of our paper is not to produce the best forecasting model, but rather show the possibilities of estimating and implementing state space models with SV using a new indirect inference approach, we refrain from adding these further improvements.

Finally, the procedures we implemented lend themselves easily to nowcasting simply by adopting a MIDAS with leads regression approach - see Andreou, Ghysels, and Kourtellos (2013) for further details. As noted in Section 4.4, this is only done at the reprojection stage. Hence, the model parameter estimates suffice to run another simulation to obtain the nowcasting models.

4.7 Conclusions

We proposed a fairly simple and remarkably accurate indirect inference estimation procedure for state space models with either Gaussian errors or stochastic volatility. We consider a mixed frequency data setting as it is a typical situation where stochastic volatility is relevant due to the use of high frequency data. We confined our attention to settings involving only a single high and low frequency data series. Yet, the methods can easily be extended to more series of either type as the mixed frequency VAR auxiliary model can straightforwardly accommodate such settings. A more challenging extension involves larger values of m - the differences in low and high frequencies. The use of U-MIDAS regressions makes our approach extremely computationally attractive due to the use of OLS. With larger values of m we know that U-MIDAS becomes over-parameterized. While regular MIDAS regressions are a feasible alternative - they require non-linear estimation and are therefore less appealing. It should also be noted that we only covered indirect estimation procedures. It would also be fairly straightforward to apply the moment matching procedure of Gallant and Tauchen (1996) instead. As is well known, this would make our procedures potentially computationally even more attractive, while maintaining the same asymptotic properties. This would also broaden the potential set of auxiliary models, including GARCH and EGARCH, as the Gallant and Tauchen (1996) procedure is based on the empirical score and does not require repeated ML estimates. An interesting extension in this regard would be to use the criteria introduced by Barigozzi, Halbleib-Chiriac, and Veredas (2014) for choosing the best auxiliary model.

Last but not least, it should be noted that we assumed that the number of factors is known. In practice, one should of course also consider testing for the number of factors. There is a considerable literature on testing for the number of factors. In terms of testing, it is worth noting that the indirect inference procedures should not pose any additional issues in terms of testing the number of factors. See in particular Guay and Scaillet (2003) who study a hypothesis testing problem quite similar to determining the number of factors - namely involving unidentified parameters under the null - in the context of indirect inference.

4.8 Tables of Chapter 4

Table 4.1: MC simulations for the single-factor linear Gaussian state space model (persistence parameter of the latent factor $\rho = 0.5$)

T = 500	MLE (Kalman filter)					Indirect Inference (Auxiliary model: U-MIDAS / AR)				
Coeff.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
γ_2	0.99	-0.01	0.94	0.98	1.06	0.98	-0.02	0.86	0.97	1.10
γ_1	1.01	0.01	0.94	1.00	1.06	1.00	0.00	0.90	0.97	1.05
ρ	0.50	0.00	0.48	0.50	0.52	0.50	0.00	0.47	0.50	0.54
d	-0.01	-0.01	-0.07	0.01	0.05	-0.02	-0.02	-0.10	0.01	0.07
σ_x	1.00	-0.00	0.92	1.01	1.07	0.99	-0.01	0.87	1.04	1.15
σ_y	0.95	-0.05	0.90	0.99	1.06	0.91	-0.09	0.87	1.02	1.11
$corr(\hat{F}, F)$	0.80	-	0.79	0.80	0.81	0.78	-	0.78	0.79	0.80
T = 200										
γ_2	0.99	-0.01	0.86	1.01	1.12	0.94	-0.06	0.75	0.93	1.09
γ_1	1.03	0.03	0.91	1.02	1.13	1.01	0.01	0.85	0.94	1.16
ρ	0.50	-0.00	0.45	0.50	0.54	0.51	0.01	0.45	0.51	0.58
d	-0.03	-0.03	-0.13	-0.00	0.09	0.00	0.00	-0.13	0.05	0.14
σ_x	0.97	-0.03	0.84	1.00	1.12	1.00	0.00	0.85	1.07	1.23
σ_y	0.86	-0.14	0.79	0.97	1.05	0.79	-0.21	0.50	1.00	1.12
$corr(\hat{F}, F)$	0.79	-	0.78	0.80	0.81	0.77	-	0.75	0.78	0.79
T = 100										
γ_2	0.97	-0.03	0.82	0.96	1.12	0.89	-0.11	0.69	0.85	1.11
γ_1	1.01	0.01	0.88	0.99	1.11	0.92	-0.08	0.69	0.89	1.11
ρ	0.50	-0.00	0.43	0.50	0.59	0.51	0.01	0.40	0.54	0.67
d	-0.04	-0.04	-0.20	0.01	0.11	-0.01	-0.01	-0.21	0.07	0.18
σ_x	0.95	-0.05	0.77	1.00	1.16	0.98	-0.02	0.76	1.11	1.26
σ_y	0.81	-0.19	0.67	0.94	1.08	0.84	-0.16	0.54	1.05	1.17
$corr(\hat{F}, F)$	0.78	-	0.75	0.78	0.81	0.75	-	0.73	0.76	0.78

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the ML (left) and Indirect Inference (II, right) estimators in 1000 MC replications. The data generating process is DGP1 in Section 5.1, corresponding to a mixed frequency linear state space model with a single AR(1) latent factor, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $\rho = 0.5$, $d = 0$, $\sigma_y = \sigma_x = 1$. The simulated samples have size $T = 500$ (top), $T = 200$ (middle), $T = 100$ (bottom). The auxiliary model for the indirect inference estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 3$ and an $AR(9)$ model for the high frequency data (see equation (4.3.9)), with the correlation between the errors of the two equations as a free auxiliary parameter. The Indirect Inference estimator uses a single long simulated sample of the structural model ($S = 1$ and T^S = 50000) and an identity weighting matrix. We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true factor paths. The estimated factor paths are obtained by Kalman filter with the ML estimate (left) and the reprojection method with the II estimate (right), using $T^{reproj} = 100000$.

Table 4.2: MC simulations for the single-factor linear Gaussian state space model (persistence parameter of the latent factor $\rho = 0.9$)

T = 500	MLE (Kalman filter)					Indirect Inference (Auxiliary model: U-MIDAS / AR)				
Coeff.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
γ_2	1.00	-0.00	0.98	1.00	1.02	0.98	-0.02	0.94	0.97	1.02
γ_1	1.00	-0.00	0.96	1.00	1.03	0.98	-0.02	0.91	0.98	1.04
ρ	0.90	0.00	0.89	0.90	0.91	0.90	0.00	0.87	0.91	0.93
d	-0.00	-0.00	-0.03	-0.00	0.03	0.02	0.02	-0.02	0.02	0.07
σ_x	1.00	-0.00	0.96	1.00	1.03	1.02	0.02	0.98	1.04	1.08
σ_y	0.99	-0.01	0.96	1.00	1.03	1.03	0.03	0.97	1.04	1.10
$corr(\hat{F}, F)$	0.95	-	0.94	0.95	0.95	0.94	-	0.94	0.94	0.95
T = 200										
γ_2	1.01	0.01	0.96	1.00	1.05	0.95	-0.05	0.87	0.95	1.02
γ_1	1.00	-0.00	0.96	1.00	1.05	0.95	-0.05	0.85	0.96	1.02
ρ	0.90	-0.00	0.88	0.90	0.91	0.90	0.00	0.85	0.91	0.95
d	-0.01	-0.01	-0.08	-0.00	0.06	0.03	0.03	-0.04	0.04	0.12
σ_x	0.99	-0.01	0.94	0.98	1.04	1.03	0.03	0.97	1.05	1.11
σ_y	0.99	-0.01	0.93	0.99	1.05	1.05	0.05	0.95	1.09	1.18
$corr(\hat{F}, F)$	0.95	-	0.94	0.95	0.95	0.94	-	0.94	0.94	0.95
T = 100										
γ_2	0.99	-0.01	0.94	0.99	1.06	0.92	-0.08	0.79	0.90	1.02
γ_1	0.99	-0.01	0.94	0.98	1.05	0.89	-0.11	0.77	0.86	1.00
ρ	0.89	-0.01	0.87	0.90	0.91	0.89	-0.01	0.81	0.92	0.99
d	-0.02	-0.02	-0.11	-0.01	0.09	0.05	0.05	-0.16	0.09	0.23
σ_x	0.98	-0.02	0.91	0.99	1.05	1.03	0.03	0.89	1.09	1.20
σ_y	0.98	-0.02	0.90	0.97	1.08	1.13	0.13	0.99	1.19	1.29
$corr(\hat{F}, F)$	0.94	-	0.93	0.94	0.95	0.93	-	0.92	0.93	0.94

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the ML (left) and Indirect Inference (II, right) estimators in 1000 MC replications. The data generating process is DGP1 in Section 5.1, corresponding to a mixed frequency linear state space model with a single AR(1) latent factor, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $\rho = 0.9$, $d = 0$, $\sigma_y = \sigma_x = 1$. The simulated samples have size $T = 500$ (top), $T = 200$ (middle), $T = 100$ (bottom). The auxiliary model for the indirect inference estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 3$ and an $AR(9)$ model for the high frequency data (see equation (4.3.9)), with the correlation between the errors of the two equations as a free auxiliary parameter. The Indirect Inference estimator uses a single long simulated sample of the structural model ($S = 1$ and T^S = 50000) and an identity weighting matrix. We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true factor paths. The estimated factor paths are obtained by Kalman filter with the ML estimate (left) and the reprojection method with the II estimate (right), using $T^{reproj} = 100000$.

Table 4.3: MC simulations for the two-factor linear Gaussian state space model

Indirect Inference	T = 500					T = 200					T = 100				
Coeff.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
$\gamma_{2,1}$	0.17	-0.03	0.00	0.00	0.37	0.16	-0.04	0.00	0.00	0.33	0.18	-0.02	0.00	0.12	0.36
$\gamma_{2,2}$	0.98	-0.02	0.92	0.99	1.05	0.98	-0.02	0.87	0.98	1.10	1.00	-0.00	0.85	1.00	1.17
$\gamma_{1,1}$	0.99	-0.01	0.91	0.98	1.05	1.05	0.05	0.94	1.03	1.17	1.01	0.01	0.81	0.99	1.22
$\gamma_{1,2}$	0.21	0.01	0.00	0.30	0.37	0.22	0.02	0.00	0.26	0.38	0.15	-0.05	0.00	0.12	0.27
ρ	0.90	-0.00	0.87	0.90	0.92	0.88	-0.02	0.84	0.89	0.92	0.88	-0.02	0.85	0.89	0.93
d	0.01	0.01	-0.07	0.00	0.07	-0.01	-0.01	-0.10	0.01	0.10	-0.05	-0.05	-0.18	-0.02	0.10
σ_x	1.00	-0.00	0.93	1.01	1.05	1.00	0.00	0.92	1.01	1.09	0.94	-0.06	0.85	1.00	1.09
σ_y	0.93	-0.07	0.85	0.98	1.09	0.76	-0.24	0.52	0.83	1.07	0.81	-0.19	0.36	0.96	1.21
$corr(\hat{F}_1, F_1)$	0.78	-	0.76	0.78	0.80	0.76	-	0.74	0.76	0.79	0.74	-	0.72	0.76	0.78
$corr(\hat{F}_2, F_2)$	0.92	-	0.91	0.92	0.92	0.91	-	0.90	0.92	0.93	0.88	-	0.89	0.92	0.93

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the Indirect Inference (II) estimator in 1000 MC replications. The data generating process is DGP2 in Section 5.1, corresponding to a mixed frequency linear state space model with two independent AR(1) latent factors, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = (1.0, 0.2)'$, $\gamma_2 = (0.2, 1.0)'$, $\rho = 0.9$, $d = 0$, $\sigma_y = \sigma_x = 1$. The simulated samples have size $T = 500$ (left), $T = 200$ (middle), $T = 100$ (right). The auxiliary model for the indirect inference estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 3$ and an $AR(9)$ model for the high frequency data (see equation (4.3.9)), with the correlation between the errors of the two equations as a free auxiliary parameter. The II estimator uses a single long simulated sample of the structural model ($S = 1$ and T^S = 50000) and an identity weighting matrix. We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true paths for each factor. The estimated factor paths are obtained by the reprojection method with the II estimate, using $T^{reproj} = 100000$.

Table 4.4: MC simulations for the stochastic volatility model (sample size $T = 500$)

$\rho = 0.5$	Indirect Inference (Auxiliary model: U-MIDAS/AR-ARCH)					MLE (Monte Carlo EM)				
Coeff.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
γ_2	0.97	-0.03	0.88	0.97	1.05	0.89	-0.11	0.86	0.89	0.92
γ_1	0.95	-0.05	0.87	0.94	1.01	1.07	0.07	1.02	1.06	1.12
ρ	0.51	0.01	0.47	0.51	0.55	0.52	0.02	0.51	0.53	0.54
d	-0.01	-0.01	-0.27	-0.02	0.29	-0.04	-0.04	-0.28	0.00	0.21
σ_y	0.95	-0.05	0.89	1.00	1.09	0.83	-0.17	0.75	0.84	0.91
σ_x	1.05	0.05	0.94	1.07	1.18	0.93	-0.07	0.91	0.94	0.96
ρ_{SV}	0.95	0.00	0.93	0.96	0.97	0.94	-0.01	0.93	0.94	0.95
ν	0.25	-0.05	0.19	0.25	0.31	0.29	-0.01	0.28	0.29	0.29
$corr(\hat{F}, F)$	0.74	-	0.72	0.74	0.75	0.79	-	0.78	0.79	0.80
$corr(\hat{h}, h)$	0.55	-	0.51	0.55	0.59	0.74	-	0.71	0.74	0.77
Comp. time (min)	15.89	-	9.01	14.69	20.94	61.05	-	38.49	54.11	78.05
$\rho = 0.9$										
γ_2	0.96	-0.04	0.91	0.96	1.00	0.97	-0.03	0.96	0.97	0.98
γ_1	0.95	-0.05	0.90	0.96	1.01	1.00	-0.00	0.98	1.00	1.02
ρ	0.91	0.01	0.89	0.91	0.93	0.90	0.00	0.89	0.90	0.91
d	0.00	0.00	-0.26	-0.05	0.38	-0.05	-0.05	-0.34	0.00	0.22
σ_y	0.97	-0.03	0.90	0.99	1.07	0.88	-0.12	0.81	0.88	0.95
σ_x	1.08	0.08	1.02	1.08	1.16	0.86	-0.14	0.82	0.87	0.91
ρ_{SV}	0.94	-0.01	0.93	0.95	0.97	0.95	-0.00	0.94	0.95	0.96
ν	0.26	-0.04	0.19	0.25	0.32	0.29	-0.01	0.28	0.29	0.29
$corr(\hat{F}, F)$	0.93	-	0.92	0.93	0.93	0.95	-	0.95	0.95	0.96
$corr(\hat{h}, h)$	0.51	-	0.47	0.51	0.56	0.75	-	0.72	0.76	0.79
Comp. time (min)	13.87	-	9.05	12.32	16.86	60.27	-	35.67	54.86	81.67

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the Indirect Inference (II, left) and ML (right) estimators in 200 MC replications. The data generating process is DGP3 in Section 5.1, corresponding to a mixed frequency stochastic volatility model with a single AR(1) latent factor in the mean, an AR(1) log SV process, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $d = 0$, $\sigma_y = \sigma_x = 1$, $\rho_{SV} = 0.95$, $\nu = 0.3$. The autoregressive coefficient of the factor in the mean is $\rho = 0.5$ in the upper panel and $\rho = 0.9$ in the lower panel. The simulated samples have size $T = 500$. The auxiliary model for the II estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 4$ and an $AR(9) - ARCH(10)$ model for the high frequency data (see equation (4.3.10)), with the correlation between the errors of the two equations as a free auxiliary parameter. The II estimator uses a single long simulated sample of the structural model ($S = 1$ and T^S = 50000) and an identity weighting matrix. The MLE is computed by Monte Carlo EM, using a particle forward-filtering backward-smoothing algoritm in the E step (see Appendix B for the detailed algorithm). We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true paths of the mean and volatility factors. The estimated factor paths are obtained by the reprojection method with the Indirect Inference estimate (left), using $T^{reproj} = 100000$, and by the average across the particles of the filtering algorithm with the ML estimate (right). Finally, we report the mean and 25%, 50%, 75% quantiles of the computational time (in minutes) for obtaining the parameter estimates and the filtered factor paths in a single simulation.

Table 4.5: MC simulations for the stochastic volatility model (sample size $T = 200$)

$\rho = 0.5$	Indirect Inference (Auxiliary model: U-MIDAS/AR-ARCH)					MLE (Monte Carlo EM)				
Coeff.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
γ_2	0.99	-0.01	0.81	0.96	1.14	0.90	-0.10	0.86	0.91	0.94
γ_1	0.88	-0.12	0.75	0.86	0.96	1.07	0.07	1.01	1.07	1.13
ρ	0.50	0.00	0.44	0.51	0.60	0.52	0.02	0.49	0.52	0.56
d	-0.05	-0.05	-0.35	-0.06	0.19	-0.03	-0.03	-0.31	0.00	0.27
σ_y	0.96	-0.04	0.90	1.06	1.17	0.81	-0.19	0.71	0.82	0.92
σ_x	1.03	0.03	0.90	1.09	1.24	0.94	-0.06	0.92	0.95	0.98
ρ_{SV}	0.94	-0.01	0.93	0.96	0.99	0.94	-0.01	0.92	0.94	0.96
ν	0.23	-0.07	0.08	0.22	0.34	0.29	-0.01	0.28	0.29	0.29
$corr(\hat{F}, F)$	0.72	-	0.69	0.73	0.76	0.79	-	0.77	0.79	0.81
$corr(\hat{h}, h)$	0.54	-	0.48	0.55	0.61	0.73	-	0.67	0.74	0.79
Comp. time (min)	18.23	-	12.17	15.41	22.88	24.76	-	13.11	21.84	33.52
$\rho = 0.9$										
γ_2	0.92	-0.08	0.84	0.93	1.00	0.97	-0.03	0.96	0.98	0.99
γ_1	0.89	-0.11	0.79	0.89	0.98	1.00	0.00	0.98	1.00	1.03
ρ	0.91	0.01	0.87	0.92	0.95	0.90	-0.00	0.89	0.90	0.91
d	0.04	0.04	-0.25	-0.04	0.48	-0.01	-0.01	-0.28	0.00	0.28
σ_y	1.03	0.03	0.88	1.06	1.18	0.87	-0.13	0.78	0.88	0.98
σ_x	1.13	0.13	0.99	1.14	1.25	0.89	-0.11	0.85	0.91	0.95
ρ_{SV}	0.93	-0.02	0.92	0.96	0.99	0.94	-0.01	0.92	0.94	0.96
ν	0.22	-0.08	0.10	0.21	0.30	0.29	-0.01	0.28	0.29	0.29
$corr(\hat{F}, F)$	0.92	-	0.91	0.92	0.93	0.95	-	0.95	0.95	0.96
$corr(\hat{h}, h)$	0.51	-	0.46	0.52	0.58	0.74	-	0.70	0.74	0.80
Comp. time (min)	16.45	-	12.11	14.19	19.94	22.50	-	11.50	18.33	32.10

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the Indirect Inference (II, left) and ML (right) estimators in 200 MC replications. The data generating process is DGP3 in Section 5.1, corresponding to a mixed frequency stochastic volatility model with a single AR(1) latent factor in the mean, an AR(1) log SV process, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $d = 0$, $\sigma_y = \sigma_x = 1$, $\rho_{SV} = 0.95$, $\nu = 0.3$. The autoregressive coefficient of the factor in the mean is $\rho = 0.5$ in the upper panel and $\rho = 0.9$ in the lower panel. The simulated samples have size $T = 200$. The auxiliary model for the II estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 4$ and an $AR(9) - ARCH(10)$ model for the high frequency data (see equation (4.3.10)), with the correlation between the errors of the two equations as a free auxiliary parameter. The II estimator uses a single long simulated sample of the structural model ($S = 1$ and T^S = 50000) and an identity weighting matrix. The MLE is computed by Monte Carlo EM, using a particle forward-filtering backward-smoothing algoritm in the E step (see Appendix B for the detailed algorithm). We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true paths of the mean and volatility factors. The estimated factor paths are obtained by the reprojection method with the II estimate (left), using $T^{reproj} = 100000$, and by the average across the particles of the filtering algorithm with the ML estimate (right). Finally, we report the mean and 25%, 50%, 75% quantiles of the computational time (in minutes) for obtaining the parameter estimates and the filtered factor paths in a single simulation.

Table 4.6: In-sample R^2 of GDP and HF indicator on latent factor

	Stochastic volatility (Indirect Inference)		Gaussian state space (Kalman filter)	
HF Indicator	$R^2(GDP)$	R^2(HF indicator)	$R^2(GDP)$	R^2(HF indicator)
(1) Industrial Production	0.74	0.48	0.82	0.50
(2) Industrial Production - Pulp/paper	0.69	0.11	0.80	0.11
(3) Business Climate - IFO	0.56	0.43	0.42	0.65
(4) Economic Sentiment Index	0.48	0.71	0.36	0.86
(5) PMI Composite	0.49	0.46	0.80	0.04

This table reports the R^2 for the regressions of both GDP and the monthly indicators on the filtered values of the latent factor F for different mixed frequency models. We estimate the mixed-frequency stochastic volatility model defined as DGP 3 in Section 4.5 and the linear Gaussian state space model defined as DGP 1 with different pairs of mixed frequency observables. In each model, GDP is the low frequency (quarterly) observable, and is treated as a flow sampled variable. The table reports results for 10 different models, which differ for the high frequency (monthly) observable and the presence/absence of stochastic volatility. Columns 2 and 3 (resp. 4 and 5) display the R^2 for the regression of the GDP and the HF observable on the filtered values of F obtained from the models with (resp. without) stochastic volatility. We estimate the SV models via Indirect Inference using the same auxiliary models as in the MC simulations of Section 4.5. The mean and volatility factors are filtered by reprojection. We estimate the Gaussian state space model by adapting the Kalman filter for periodic state space models proposed in Bai, Ghysels, and Wright (2013), see Section C. All models are estimated on the full data sample from 1991-Q1 to 2011-Q1.

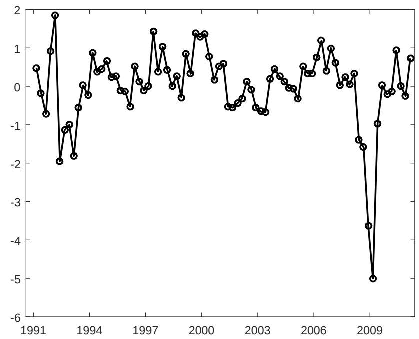
Table 4.7: Root Mean Squared Forecasting Error (RMSFE) for GDP

	Stochastic volatility (Indirect Inference)				Gaussian state space (Kalman filter)			
	Forecast horizon H (Quarters ahead)				Forecast horizon H (Quarters ahead)			
HF Indicator	1	2	3	4	1	2	3	4
(1) Industrial Production	0.79	1.00	1.12	1.16	0.70	1.02	1.06	1.06
(2) Industrial Production - Pulp/paper	0.83	1.30	1.00	1.05	0.80	1.02	1.03	1.02
(3) Business Climate - IFO	0.68	0.89	0.98	1.02	0.91	1.12	1.14	1.08
(4) Economic Sentiment Index	0.70	0.92	1.01	1.04	1.00	0.98	0.97	0.97
(5) PMI Composite	0.92	0.99	0.99	0.99	0.78	1.00	1.01	1.02

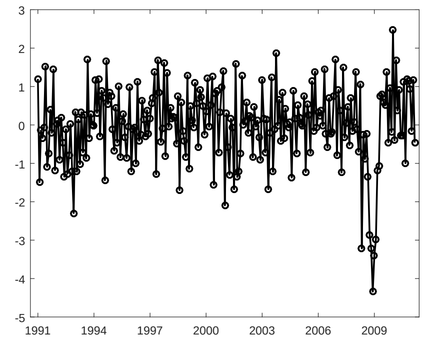
This table reports the Root Mean Squared Forecasting Error (RMSFE) for GDP in different mixed frequency models. The RMSFE is reported as the ratio to the RMSFE of the naive forecasting model assuming constant GDP growth rate. We consider the mixed-frequency stochastic volatility model defined as DGP 3 in Section 4.5 and the linear Gaussian state space model defined as DGP 1 with different pairs of mixed frequency observables. In each model, GDP is the low frequency (quarterly) observable, and is treated as a flow sampled variable. The table reports the forecasting results for 10 different models, which differ for the high frequency (monthly) observable and the presence/absence of stochastic volatility. To produce the forecasts, the models are estimated on the estimation window, and then used for prediction up to 4 quarters ahead of the estimation final date. The first estimation window is from 1991-Q1 to 2005-Q4, and is recursively expanded up to 2010-Q4. Columns 2 to 5 (resp. 6 and 9) display the RMSFE ratios at horizons $H = 1, 2, 3, 4$ quarters ahead for the models with (resp. without) stochastic volatility. We estimate the SV models via Indirect Inference using the same auxiliary models as in the MC simulations of Section 4.5. We estimate the Gaussian state space model by adapting the Kalman filter for periodic state space models proposed in Bai, Ghysels, and Wright (2013), see Section C.

4.9 Figures of Chapter 4

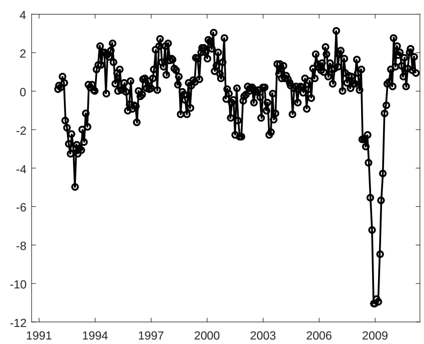
Figure 4.1: Time series of observables and estimated factors: stochastic volatility model estimated on GDP and IP.



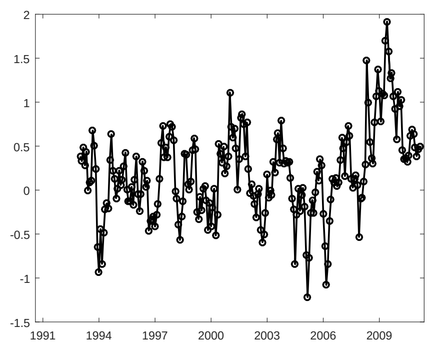
(a) Quartely GDP.



(b) Monthly IP index.



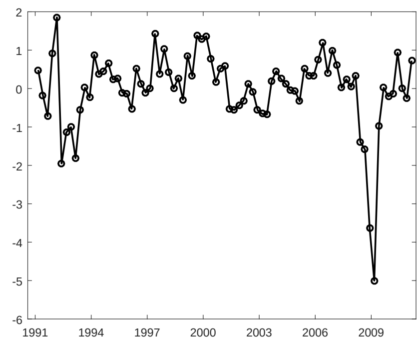
(c) Estimated common factor $\hat{F}$.



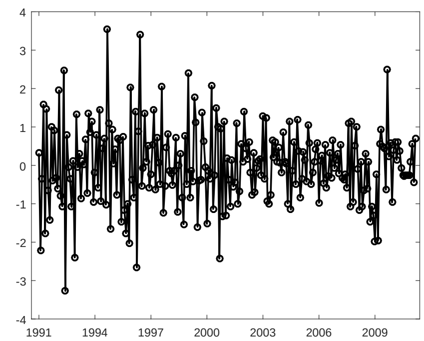
(d) Estimated SV factor $\hat{h}$.

Panels (a) and (b) display the time series of the (standardized) quarterly growth rate of European GDP, and the (standardized) monthly growth rate of aggregate European Industrial Production index. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable. The sample is from 1991-Q1 to 2011-Q1. Panels (c) and (d) display the estimated mean and idiosyncratic volatility factors $\hat{F}$ and $\hat{h}$. The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

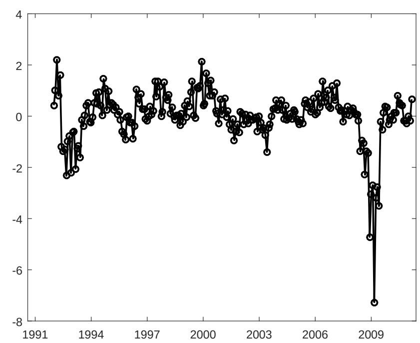
Figure 4.2: Time series of observables and estimated factors: stochastic volatility model estimated on GDP and IP - Pulp/Paper.



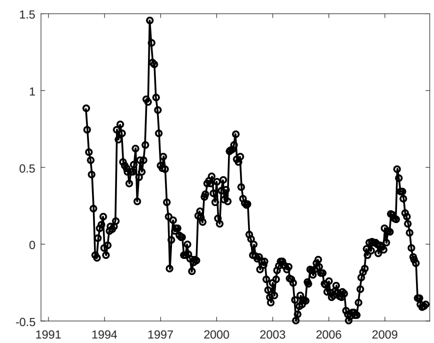
(a) Quartely GDP.



(b) Monthly IP - Pulp/Paper index.



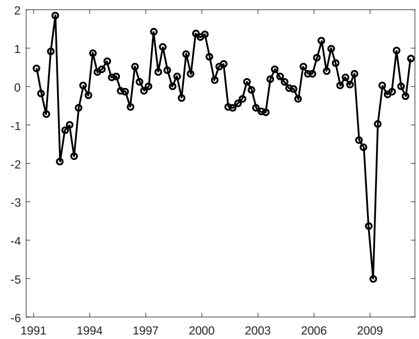
(c) Estimated common factor $\hat{F}_{t+j/3}$.



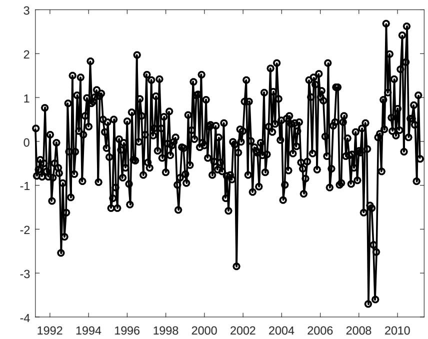
(d) Estimated SV factor $\hat{h}_{t+j/3}$.

Panels (a) and (b) display the time series of the (standardized) quarterly growth rate of European GDP, and the (standardized) monthly growth rate European Industrial Production index for "Pulp and Paper sector". These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable. The sample is from 1991-Q1 to 2011-Q1. Panels (c) and (d) display the estimated mean and idiosyncratic volatility factors $\hat{F}$ and $\hat{h}$. The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

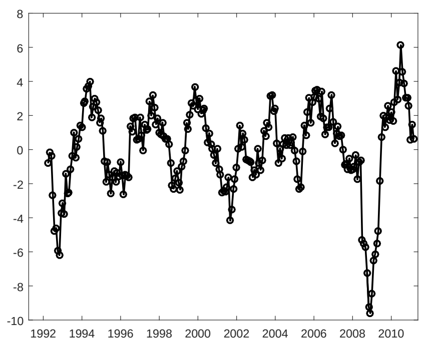
Figure 4.3: Time series of observables and estimated factors: stochastic volatility model estimated on GDP and IFO.



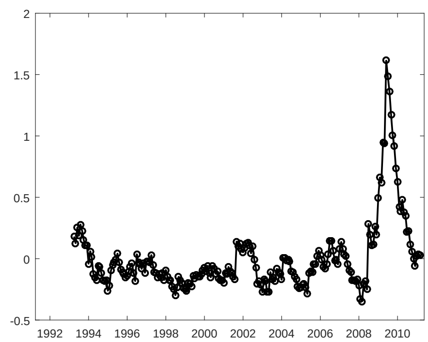
(a) Quartely GDP.



(b) Monthly IFO index.



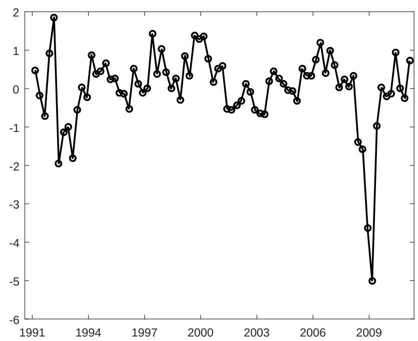
(c) Estimated common factor $\hat{F}_{t+j/3}$.



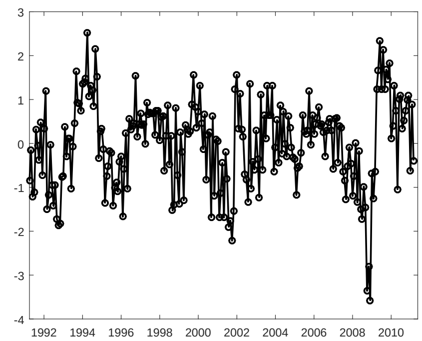
(d) Estimated SV factor $\hat{h}_{t+j/3}$.

Panels (a) and (b) display the time series of the (standardized) quarterly growth rate of European GDP, and the monthly (standardized first difference of) Germany IFO Business Climate index. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable. The sample is from 1991-Q1 to 2011-Q1. Panels (c) and (d) display the estimated mean and idiosyncratic volatility factors $\hat{F}$ and $\hat{h}$. The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

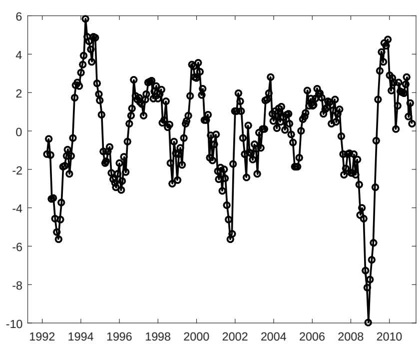
Figure 4.4: Time series of observables and estimated factors: stochastic volatility model estimated on GDP and ESI.



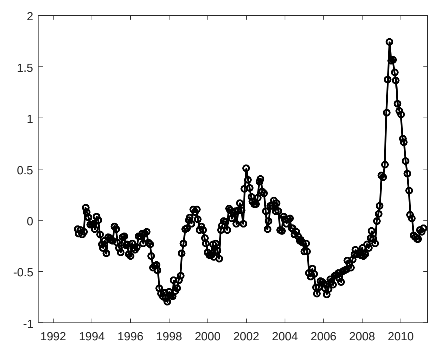
(a) Quartely GDP.



(b) Monthly ESI index.



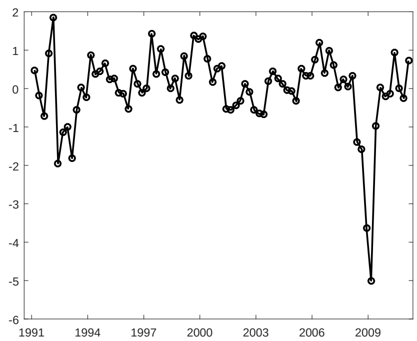
(c) Estimated common factor $\hat{F}_{t+j/3}$.



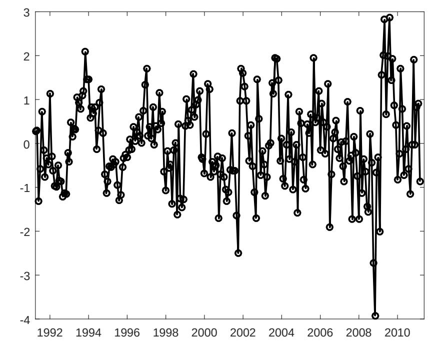
(d) Estimated SV factor $\hat{h}_{t+j/3}$.

Panels (a) and (b) display the time series of the (standardized) quarterly growth rate of European GDP, and the (standardized first difference of) monthly Euro-area Economic Sentiment Index. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable. The sample is from 1991-Q1 to 2011-Q1. Panels (c) and (d) display the estimated mean and idiosyncratic volatility factors $\hat{F}$ and $\hat{h}$. The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

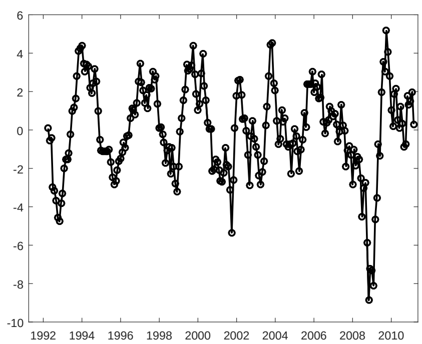
Figure 4.5: Time series of observables and estimated factors: stochastic volatility model estimated on GDP and PMI.



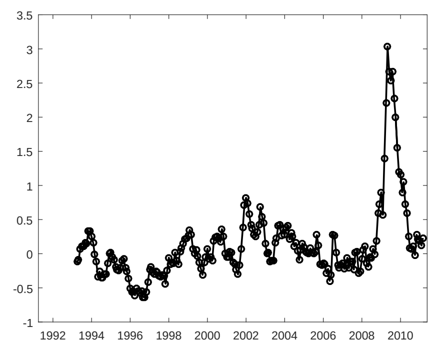
(a) Quartely GDP.



(b) Monthly PMI index.



(c) Estimated common factor $\hat{F}_{t+j/3}$.



(d) Estimated SV factor $\hat{h}_{t+j/3}$.

Panels (a) and (b) display the time series of the (standardized) quarterly growth rate of European GDP, and the (standardized first difference of) monthly Euro-area Composite Purchasing Manager Index. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable. The sample is from 1991-Q1 to 2011-Q1. Panels (c) and (d) display the estimated mean and idiosyncratic volatility factors $\hat{F}$ and $\hat{h}$. The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

4.10 Appendix A: Linear state space models with mixed frequency data

In this Appendix we summarize some results from Bai, Ghysels, and Wright (2013) concerning linear state space models with mixed frequency data. These results are useful to obtain the binding function linking our structural state space model and the auxiliary MIDAS regressions when the structural model does not feature SV (see Section 4.3.1). They also provide the Kalman filter algorithm for ML estimation of the structural model without SV used in the MC simulations (see Section 4.5).

4.10.1 Model setup

The linear state space model presented in Section 3.1 can be summarized as follows. The latent factor F follows a $VAR(p)$ process:

$$F_{t+j/m} = \sum_{l=1}^{p} \Phi_l F_{t+(j-l)/m} + \eta_{t+j/m} \qquad \forall t = 1, \ldots, T, \qquad j = 0, \ldots, m-1. \tag{4.10.1}$$

The low frequency data is related to factors as follows:

$$y^*_{t+j/m} = \gamma_1' F_{t+j/m} + u_{1,t+j/m} \qquad \forall t, \qquad j = 0, \ldots, m-1, \tag{4.10.2}$$

with $u_{1,t+j/m}$ having an $AR(k)$ representation:

$$d_1(L^{1/m}) u_{1,t+j/m} = \varepsilon_{1,t+j/m}, \qquad d_1(L^{1/m}) \equiv 1 - d_{11} L^{1/m} - \ldots - d_{k1} L^{k/m}, \tag{4.10.3}$$

and the lag operator $L^{1/m}$ applying to high-frequency data, i.e. $L^{1/m} u_t \equiv u_{t-1/m}$. The observed low-frequency process y relates to the latent process y^* via a linear aggregation scheme:

$$y^c_{t+j/m} = \Psi_j y^c_{t+(j-1)/m} + \lambda_j y^*_{t+j/m} \tag{4.10.4}$$

where y_t is equal to y^c_t for integer t, and is not observed otherwise. The high frequency process $x_{t+j/m}$ relates to the factors as follows:

$$x_{t+j/m} = \gamma_2' F_{t+j/m} + u_{2,t+j/m} \qquad \forall t, \quad j = 0, \ldots, m-1, \tag{4.10.5}$$

where:

$$d_2(L^{1/m}) u_{2,t+j/m} = \varepsilon_{2,t+j/m}, \qquad d_2(L^{1/m}) \equiv 1 - d_{12} L^{1/m} - \ldots - d_{k2} L^{k/m}. \tag{4.10.6}$$

This model corresponds to a restricted version of the specification in Assumptions 1 and 2 with $\rho_{SV} = 1$ and $\nu_2 = 0$.

4.10.2 State space representation and Kalman filter

The above equations yield a periodic state space model with measurement equation:

$$Y^j_t = Z_j \alpha_{t+j/m} \qquad \begin{cases} Y^j_t = (y_t, x_t)' & j = 0 \\ Y^j_t = x_{t+j/m} & 0 < j \leq m-1 \end{cases}, \tag{4.10.7}$$

where

$$Z_0 = \begin{bmatrix} \gamma_1' & O_{2\times n_f(p-1)} & I_2 & O_{2\times 2(k-1)} \\ \gamma_2' & & & \end{bmatrix}$$

$$Z_j = \begin{bmatrix} \gamma_2' & O_{1\times n_f(p-1)} & 1 & O_{1\times 2(k-1)} \end{bmatrix}$$

for $0 < j \leq m-1$ and state vector

$$\alpha_{t+j/m} = \left(F'_{t+j/m}, \ldots, F'_{t+(j-p+1)/m}, u'_{t+j/m}, \ldots, u'_{t+(j-k+1)/m}\right)'$$

where $u_{t+j/m} = (u_{1,t+j/m}, u_{2,t+j/m})'$.

The transition equation is:

$$\alpha_{t+j/m} = R\alpha_{t+(j-1)/m} + Q\zeta_{t+j/m} \tag{4.10.8}$$

where

$$R = \begin{bmatrix} \Phi_1 \ldots \Phi_{p-1} & \Phi_p & O_{n_f\times 2(k-1)} & O_{n_f\times 2} \\ I_{(p-1)n_f} & O_{(p-1)n_f\times n_f} & O_{(p-1)n_f\times 2(k-1)} & O_{(p-1)n_f\times n} \\ O_{2\times(p-1)n_f} & O_{2\times n_f} & D_1 \ldots D_{k-1} & D_k \\ O_{2(k-1)\times(p-1)n_f} & O_{2(k-1)\times n_f} & I_{2(k-1)} & O_{2(k-1)\times 2} \end{bmatrix}$$

$$Q = \begin{bmatrix} I_{n_f} & O_{n_f\times 2} \\ O_{(p-1)n_f\times n_f} & O_{(p-1)n_f\times 2} \\ O_{2\times n_f} & I_2 \\ O_{2(k-1)\times n_f} & O_{2(k-1)\times 2} \end{bmatrix}$$

D_i = diag$(d_{li}, l = 1, 2)$ and $\zeta_{t+j/m} = (\eta'_{t+j/m}, \varepsilon_{1,t+j/m}, \varepsilon_{2,t+j/m})'$. Let Σ_ζ denote the variance-covariance matrix of $\zeta_{t+j/m}$.

The above state space model is periodic as it cycles to the data release pattern that repeats itself every m periods. Such systems have a (periodic) steady state (see e.g. Assimakis and Adam (2009)). If we let $P_{j|j-1}$ denote the steady state covariance matrix of $\alpha_{t+j/m|t+(j-1)/m}$, then the equations:

$$\begin{aligned} P_{j+1|j} &= Q\Sigma_\zeta Q' + RP_{j|j-1}R' - RP_{j|j-1}Z_j'[Z_jP_{j|j-1}Z_j']^{-1}Z_jP_{j|j-1}R' \qquad j = 0, \ldots, m-2 \\ P_{0|-1} &= Q\Sigma_\zeta Q' + RP_{2|1}R' - RP_{2|1}Z_j'[Z_jP_{2|1}Z_j']^{-1}Z_jP_{2|1}R' \qquad j = m-1 \end{aligned} \tag{4.10.9}$$

must be satisfied and $P_{j|j-1} = P_{j+m|j+m-1}, \forall\, j$. The periodic steady state Kalman gain is therefore:

$$K_{j|j-1} = P_{j|j-1}Z_j'[Z_jP_{j|j-1}Z_j']^{-1} \tag{4.10.10}$$

with $K_{j|j-1} \equiv K_{j+m|j-1+m}, \forall\, j$. Let us define the information set $I^M_{t+j/m}$ as the linear space generated by $\{Y^j_{t+(j-k)/m} | k \geq 0\}$. When we define the extraction of the state vector as:

$$\hat{\alpha}_{(t+j/m)|(t+j/m)} = E[\alpha_{t+j/m}|I^M_{t+j/m}] \tag{4.10.11}$$

the filtered states are:

$$\hat{\alpha}_{(t+j/m)|(t+j/m)} = A_{j|j-1}\hat{\alpha}_{t+(j-1)/m|t+(j-1)/m} + K_{j|j-1}Y_t^j \tag{4.10.12}$$

where $A_{j|j-1} = R - K_{j|j-1} Z_j R$ and $Y_t^m = Y_{t+1}^0$.

Suppose we are interested in predicting at low-frequency intervals only, namely $\hat{\alpha}_{(t+k)|t}$, for k integer valued, using all available low and high-frequency data. First we note that:

$$\hat{\alpha}_{(t+k)|(t+k)} = [\tilde{A}_1^m]^k \hat{\alpha}_{t|t} + \sum_{i=1}^{m} \sum_{j=1}^{k} [\tilde{A}_1^m]^{k-j} \tilde{A}_{i+1}^m K_{i|i-1} Y_{t+j-1}^i \tag{4.10.13}$$

where

$$\tilde{A}_j^i = \begin{cases} A_{i|i-1} A_{i-1|i-2} \dots A_{j|j-1} & i \geq j \\ I & i < j \end{cases}$$

Expression (4.10.13) can be obtained via straightforward algebra - see Assimakis and Adam (2009). Given Assumption 1, all the eigenvalues of $A_{j|j-1}$, $j = 1, \dots, m-1$, are inside the unit circle, as are the eigenvalues of the product matrices $\{\tilde{A}_j^i\}$ (see again Assimakis and Adam (2009)). This implies that we can rewrite (4.10.13) as:

$$\begin{aligned} \hat{\alpha}_{t|t} &= \sum_{j=0}^{\infty} \sum_{i=1}^{m} [\tilde{A}_1^m]^j \tilde{A}_{i+1}^m K_{i|i-1} Y_{t-j}^i \\ &= \sum_{j=0}^{\infty} [\tilde{A}_1^m]^j K_{m|m-1} \begin{pmatrix} y_{t-j} \\ x_{t-j} \end{pmatrix} + \sum_{j=0}^{\infty} \sum_{i=1}^{m-1} [\tilde{A}_1^m]^j \tilde{A}_{i+1}^m K_{i|i-1} x_{t-1-j+i/m} \end{aligned} \tag{4.10.14}$$

from which forecasts can easily be constructed as $E_t[y_{t+h}] = Z_{0,1} R^{mh} \hat{\alpha}_{t|t}$, where $Z_{0,1}$ denotes the first row of the matrix Z_0. When factor F is scalar with autoregressive coefficient ρ, and $m = 3$, the latter equation yields equation (4.3.1) in Section 3.1.

4.10.3 ML estimation

To proceed to maximum likelihood estimation, let $\theta \in \Theta$ be the parameter vector governing the parameters of the state space model, i.e. $\theta = ((\gamma_i)_{i=1}^2, (\Psi_i)_{i=1}^p, (D_i)_{i=1}^k, \Sigma_\zeta)$ (accounting for identification constraints). Consider the vector Y_t^j defined for $j = 0, \dots, m-1$, in equation (4.10.7) and the information set $I_{t+j/m}^M$ in equation (4.10.11). Then:

$$Y_{t+(j+1)/m}^j | I_{t+j/m}^M; \theta \sim \mathcal{N}(\mu_{t+(j+1)/m}(\theta), \Sigma_{t+(j+1)/m}(\theta)) \tag{4.10.15}$$

with $\mu_{t+(j+1)/m}(\theta) \equiv Z_{j+1}(\theta) \hat{\alpha}_{t+(j+1)/m|t+j/m}(\theta)$ and

$$\Sigma_{t+(j+1)/m}(\theta) \equiv Z_{j+1}(\theta)' P_{t+(j+1)/m|t+j/m}(\theta) Z_{j+1}(\theta) + Q(\theta).$$

The value of the log likelihood for a sample of size Tm is then:

$$\begin{aligned} \sum_{t=1}^{T} \sum_{j=0}^{m-1} \log \ell(Y_{t+(j+1)/m}^j | I_{t+j/m}^M; \theta) = -\frac{Tm}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log |\Sigma_{t+(j+1)/m}(\theta)| \\ -\frac{1}{2} \sum_{t=1}^{T} \sum_{j=0}^{m-1} (Y_{t+(j+1)/m}^j - \mu_{t+(j+1)/m})' (\Sigma_{t+(j+1)/m}(\theta))^{-1} (Y_{t+(j+1)/m}^j - \mu_{t+(j+1)/m}(\theta)) \end{aligned} \tag{4.10.16}$$

We denote the estimator that maximizes this log likelihood function by $\hat{\theta}_{Tm}^{ML}$. Standard regularity conditions imply that as $T \to \infty$:

$$\sqrt{Tm}(\hat{\theta}_{Tm}^{ML} - \theta_0) \to_d \mathcal{N}(0, V^{ML}), \tag{4.10.17}$$

where θ_0 denotes the true parameter value.

4.11 Appendix B: Estimation of the mixed-frequency SV model by Monte Carlo EM algorithm

In this Appendix we describe a Monte Carlo Expectation Maximization (EM) algorithm for estimation of the state space model with mixed frequency data and stochastic volatility (see Section 3.5). In this algorithm, the smoothing distribution of the latent factors necessary in the Expectation step is obtained using a Forward Filtering-Backward Smoothing simulation-based procedure.

4.11.1 Model Setup

In this appendix we only consider models with unidimensional observables y_t and $x_{t+j/m}$, and unidimensional latent factors $F_{t+j/m}$ and $h_{t+j/m}$. The generalization to multivariate observables and latent factors is relatively straightforward, at the expense of a more involved notation. We consider a model with autocorrelated innovations $u_{y,t+j/m}$ and stock sampled LF variables $y^*_{t+j/m}$:

$$\begin{aligned}
y^*_{t+j/m} &= \gamma_1 F_{t+j/m} + u_{y,t+j/m}, && (4.11.1)\\
x_{t+j/m} &= \gamma_2 F_{t+j/m} + u_{x,t+j/m}, && (4.11.2)\\
F_{t+j/m} &= \rho F_{t+(j-1)/m} + \eta_{t+j/m}, && (4.11.3)\\
h_{t+j/m} &= \rho_{SV} h_{t+(j-1)/m} + \nu \xi_{t+j/m}, && (4.11.4)\\
u_{y,t+j/m} &= d\, u_{y,t+(j-1)/m} + \sigma_y \varepsilon_{y,t+j/m}, && (4.11.5)\\
u_{x,t+j/m} &= \sigma_x \exp\{\frac{1}{2} h_{t+j/m}\} \varepsilon_{x,t+j/m}, && (4.11.6)\\
(\eta_{t+j/m}, \xi_{t+j/m}, \varepsilon_{y,t+j/m}, \varepsilon_{x,t+j/m})' &\sim i.i.\mathcal{N}(0, I_4), && (4.11.7)\\
y^*_{t+j/m} & \quad \text{is stock-sampled at } j = 0. && (4.11.8)
\end{aligned}$$

We focus on the setting with $m = 3$ as in the Monte-Carlo analysis of Section 5.

In Section B.2 we derive the state space representation of the SV model in low frequency. In Section B.3 we describe the E-step and the M-step of the EM algorithm. In Section B.4 we provide the simulation-based procedure to obtain the smoothing distribution of the latent factor process required in the E-step of the EM algorithm. Throughout this Appendix, $\ell(\cdot)$ denotes the (conditional) density of the indicated random variables.

4.11.2 State space representation

We derive a state space representation of model (4.11.1)-(4.11.8) in low frequency. For this purpose, we define the vector of stacked measurements Y_t and the vector of stacked latent factors f_t as follows:

$$\begin{aligned}
Y_t &:= (y_t, x_t, x_{t-1/3}, x_{t-2/3})',\\
f_t &:= \begin{bmatrix} \tilde{F}_t \\ \tilde{h}_t \end{bmatrix}, \qquad \tilde{F}_t := \begin{bmatrix} F_t \\ F_{t-1/3} \\ F_{t-2/3} \end{bmatrix}, \qquad \tilde{h}_t := \begin{bmatrix} h_t \\ h_{t-1/3} \\ h_{t-2/3} \end{bmatrix}.
\end{aligned}$$

Measurement density

Let us first derive the distribution of Y_t given $\underline{Y_{t-1}}$ and $\underline{f_t}$. From equations (4.11.1)-(4.11.8) we get:

$$Y_t = \Gamma \tilde{F}_t + u_t, \tag{4.11.9}$$

where

$$u_t \quad := \quad (u_{y,t}, u_{x,t}, u_{x,t-1/3}, u_{x,t-2/3})', \qquad \Gamma := \begin{bmatrix} \gamma_1 & 0 & 0 \\ \gamma_2 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_2 \end{bmatrix}.$$

To derive the dynamics of innovation u_t, we use that equation (4.11.5) and backward iteration imply $u_{y,t} = d^3 u_{y,t-1} + \sigma_y(\varepsilon_{y,t} + d\varepsilon_{y,t-1/3} + d^2\varepsilon_{y,t-2/3})$. This equation can be written as:

$$u_{y,t} \quad = \quad d^3 u_{y,t-1} + \sigma_y \sqrt{1 + d^2 + d^4}\, \varepsilon^*_{y,t}, \qquad \varepsilon^*_{y,t} \sim i.i.\mathcal{N}(0,1), \;\; t = 1, 2, ..., T,$$

where $(\varepsilon^*_{y,t})$ is independent from $(\varepsilon_{x,t-j/3})$, $(\eta_{t-j/3})$ and $(\xi_{t-j/3})$. Thus, innovation process (u_t) is such that:

$$u_t \quad = \quad A u_{t-1} + B_t \tilde{\varepsilon}^*_t, \tag{4.11.10}$$

where

$$\tilde{\varepsilon}^*_t \quad = \quad \begin{bmatrix} \varepsilon^*_{y,t} \\ \varepsilon_{x,t} \\ \varepsilon_{x,t-1/3} \\ \varepsilon_{x,t-2/3} \end{bmatrix}, \qquad A \quad = \quad \begin{bmatrix} d^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$B_t \quad = \quad \begin{bmatrix} \sigma_y\sqrt{1+d^2+d^4} & 0 & 0 & 0 \\ 0 & \sigma_x \exp\{\frac{1}{2}h_t\} & 0 & 0 \\ 0 & 0 & \sigma_x \exp\{\frac{1}{2}h_{t-1/3}\} & 0 \\ 0 & 0 & 0 & \sigma_x \exp\{\frac{1}{2}h_{t-2/3}\} \end{bmatrix}.$$

Equations (4.11.9) and (4.11.10) imply:

$$Y_t - AY_{t-1} \quad = \quad \Gamma\tilde{F}_t - A\Gamma\tilde{F}_{t-1} + B_t\tilde{\varepsilon}^*_t,$$

and thus:

$$Y_t \;=\; \begin{bmatrix} y_t \\ x_t \\ x_{t-1/3} \\ x_{t-2/3} \end{bmatrix} \quad = \quad \begin{bmatrix} d^3 y_{t-1} + \gamma_1(F_t - d^3 F_{t-1}) \\ \gamma_2 F_t \\ \gamma_2 F_{t-1/3} \\ \gamma_2 F_{t-2/3} \end{bmatrix} + B_t\tilde{\varepsilon}^*_t.$$

From the last equation we get the measurement distribution:

$$Y_t \mid \underline{Y_{t-1}},\; \underline{f_t} \quad \sim \quad \mathcal{N}\left(\begin{bmatrix} d^3 y_{t-1} + \gamma_1(F_t - d^3 F_{t-1}) \\ \gamma_2 F_t \\ \gamma_2 F_{t-1/3} \\ \gamma_2 F_{t-2/3} \end{bmatrix},\; B_t^2 \right), \tag{4.11.11}$$

and the measurement density:

$$\begin{aligned}
\ell(Y_t|\underline{Y_{t-1}}, \underline{f_t}; \theta) &= \frac{1}{\sqrt{(2\pi)^4\, \sigma_y^2\, (1+d^2+d^4)\, (\sigma_x^2)^3 \exp\{h_t + h_{t-1/3} + h_{t-2/3}\}}} \\
&\quad \times \exp\left\{ -\frac{[y_t - d^3 y_{t-1} - \gamma_1(F_t - d^3 F_{t-1})]^2}{2(1+d^2+d^4)\sigma_y^2} \right. \\
&\quad \left. - \frac{(x_t - \gamma_2 F_t)^2}{2\sigma_x^2 \exp\{h_t\}} - \frac{(x_{t-1/3} - \gamma_2 F_{t-1/3})^2}{2\sigma_x^2 \exp\{h_{t-1/3}\}} - \frac{(x_{t-2/3} - \gamma_2 F_{t-2/3})^2}{2\sigma_x^2 \exp\{h_{t-2/3}\}} \right\} \\
&=: h(Y_t|Y_{t-1}, \underline{f_t}; \theta). \qquad (4.11.12)
\end{aligned}$$

The measurement density depends on the past measurement Y_{t-1}, and on the current and past factor values f_t, f_{t-1}.

Transition density

Let us now derive the distribution of f_t given $\underline{Y_{t-1}}$ and $\underline{f_{t-1}}$. From equations (4.11.3)-(4.11.4) and being $(\eta_{t+j/3})$, $(\xi_{t+j/3})$ independent Gaussian White Noise processes, we have:

$$\ell(f_t|\underline{Y_{t-1}}, \underline{f_{t-1}}; \theta) = \ell(f_t|f_{t-1}; \theta) = \ell(\tilde{F}_t|\tilde{F}_{t-1}; \theta)\ell(\tilde{h}_t|\tilde{h}_{t-1}; \theta).$$

Thus, process (f_t) is exogenous and first-order Markov, with transition density:

$$\begin{aligned}
g(f_t|f_{t-1}; \theta) &= g(\tilde{F}_t|\tilde{F}_{t-1}; \theta)\, g(\tilde{h}_t|\tilde{h}_{t-1}; \theta) \\
&= g(F_t|F_{t-1/3}; \theta)\, g(F_{t-1/3}|F_{t-2/3}; \theta)\, g(F_{t-2/3}|F_{t-1}; \theta) \\
&\quad \times g(h_t|h_{t-1/3}; \theta)\, g(h_{t-1/3}|h_{t-2/3}; \theta)\, g(h_{t-2/3}|h_{t-1}; \theta),
\end{aligned}$$

where:

$$g(F_{t-j/3}|F_{t-(j+1)/3}; \theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(F_{t-j/3} - \rho F_{t-(j+1)/3})^2}{2} \right\}, \qquad (4.11.13)$$

$$g(h_{t-j/3}|h_{t-(j+1)/3}; \theta) = \frac{1}{\sqrt{2\pi\nu^2}} \exp\left\{ -\frac{(h_{t-j/3} - \rho_{SV} h_{t-(j+1)/3})^2}{2\nu^2} \right\}, \qquad (4.11.14)$$

for $j = 0, 1, 2$.

The likelihood function

The density of $(\underline{Y_T}, \underline{f_T})$, conditioning on Y_0 and f_0, is:

$$\ell(\underline{Y_T}, \underline{f_T}; \theta) = \prod_{t=1}^{T} \ell(Y_t|\underline{Y_{t-1}}, \underline{f_t}; \theta)\ell(f_t|\underline{Y_{t-1}}, \underline{f_{t-1}}; \theta) = \prod_{t=1}^{T} h(Y_t|Y_{t-1}, \underline{f_t}; \theta)\, g(f_t|f_{t-1}; \theta).$$

The likelihood function $\ell(\underline{Y_T};\theta)$, conditioning on y_0 and f_0, is obtained by integrating out the path of the unobservable factor:

$$\begin{aligned}\ell(\underline{Y_T};\theta) &= \int \ell(\underline{Y_T},\underline{f_T};\theta)d\underline{f_T} \\ &= \int ... \int \prod_{t=1}^{T}\{h(Y_t|Y_{t-1},\underline{f_t};\theta)\ g(f_t|f_{t-1};\theta)\}\prod_{t=1}^{T}df_t.\end{aligned}$$

The large-dimensional integral with respect to the factor path makes this expression of the likelihood function intractable for the computation of the Maximum Likelihood (ML) estimate. The EM algorithm defined in the next section relies instead on the so-called *complete-observation log-likelihood function*, i.e., the log-density function of both the observable and unobservable variables:

$$\begin{aligned}\mathcal{L}^*(\theta) &:= \log \ell(\underline{Y_T},\ \underline{f_T};\theta) \\ &= \sum_{t=1}^{T}\log h(Y_t|Y_{t-1},\underline{f_t};\theta) + \sum_{t=1}^{T}\log g(f_t|f_{t-1};\theta) \\ &= \sum_{t=1}^{T}\log h(Y_t|Y_{t-1},\underline{f_t};\theta) \\ &\quad + \sum_{t=1}^{T}\sum_{j=0}^{2}\left[\log g(F_{t-j/3}|F_{t-(j+1)/3};\theta) + \log g(h_{t-j/3}|h_{t-(j+1)/3};\theta)\right]. \qquad (4.11.15)\end{aligned}$$

Substituting equations (4.11.12), (4.11.13) and (4.11.14) into equation (4.11.15) we get:

$$\begin{aligned}\mathcal{L}^*(\theta) &= -\frac{1}{2}\Bigg(T\log(1+d^2+d^4) + T\log\sigma_y^2 + 3T\log\sigma_x^2 + 3T\log\nu^2 \\ &\quad + \sum_{t=1}^{T}\frac{[y_t - d^3y_{t-1} - \gamma_1(F_t - d^3F_{t-1})]^2}{(1+d^2+d^4)\sigma_y^2} \\ &\quad + \sum_{t=1}^{T}\sum_{j=0}^{2}\left\{h_{t-j/3} + \frac{(x_{t-j/3} - \gamma_2 F_{t-j/3})^2}{\sigma_x^2\exp\left\{h_{t-j/3}\right\}} + (F_{t-j/3} - \rho F_{t-(j+1)/3})^2\right. \\ &\quad \left. + \frac{1}{\nu^2}(h_{t-j} - \rho_{SV}h_{t-(j+1)/3})^2\right\}\Bigg), \qquad (4.11.16)\end{aligned}$$

up to an additive constant.

4.11.3 The EM algorithm

The Expectation-Maximization (EM) algorithm is an iterative procedure to compute numerically the ML estimate in a model with unobservable variables (Dempster, Laird, and Rubin (1977)). Let $\hat{\theta}^{(i)} \equiv \hat{\theta}_{Tm}^{EM,(i)}$ be the estimate of parameter θ at iteration i of the EM algorithm. The update $i \to i+1$ consists of two steps:

1. *Expectation (E) step*. Compute function $Q(\theta|\tilde{\theta})$, with $\tilde{\theta} = \hat{\theta}^{(i)}$, where:

$$\begin{aligned} Q\left(\theta|\tilde{\theta}\right) &:= E_{\tilde{\theta}}\left[\mathcal{L}^*(\theta)|\underline{Y_T}\right] \\ &= \sum_{t=1}^{T} E_{\tilde{\theta}}\left[h(Y_t|Y_{t-1}, \underline{f_t};\theta)|\underline{Y_T}\right] \\ &\quad + \sum_{t=1}^{T}\sum_{j=0}^{2} E_{\tilde{\theta}}\left[\log g(F_{t-j/3}|F_{t-(j+1)/3};\theta) + \log g(h_{t-j/3}|h_{t-(j+1)/3};\theta)|\underline{Y_T}\right], \end{aligned}$$

   and $E_{\tilde{\theta}}\left[\cdot|\underline{Y_T}\right]$ denotes the expectation w.r.t. the conditional distribution of $\underline{f_T}$ given $\underline{Y_T}$ for parameter value $\tilde{\theta}$.

2. *Maximization (M) step*. Compute the estimate for iteration $i+1$ as:

$$\hat{\theta}^{(i+1)} := \arg\max_{\theta\in\Theta} Q\left(\theta|\hat{\theta}^{(i)}\right).$$

The iteration is performed until numerical convergence of the estimate is achieved.

We detail below the E-step and the M-step of the EM algorithm for the SV model with mixed frequency.

The E-step

Let us compute explicitly $Q\left(\theta|\tilde{\theta}\right)$, with $\tilde{\theta} = \hat{\theta}^{(i)}$, for model (4.11.1)-(4.11.8). From (4.11.16), we have:

$$\begin{aligned} Q\left(\theta|\tilde{\theta}\right) &:= E_{\tilde{\theta}}\left[\mathcal{L}^*(\theta)|\underline{Y_T}\right] \\ &= -\frac{1}{2}\Bigg(T\log(1+d^2+d^4) + T\log\sigma_y^2 + 3T\log\sigma_x^2 + 3T\log\nu^2 + \sum_{t=1}^{T}\sum_{j=0}^{2} E_{\tilde{\theta}}\left[h_{t-j/3}|\underline{Y_T}\right] \\ &\quad + \frac{1}{(1+d^2+d^4)\sigma_y^2}\sum_{t=1}^{T} E_{\tilde{\theta}}\left[(y_t - d^3 y_{t-1} - \gamma_1(F_t - d^3 F_{t-1}))^2\middle|\,\underline{Y_T}\right] \\ &\quad + \frac{1}{\sigma_x^2}\sum_{t=1}^{T}\sum_{j=0}^{2}\Big\{E_{\tilde{\theta}}\left[(x_{t-j/3} - \gamma_2 F_{t-j/3})^2 e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right] \\ &\quad + E_{\tilde{\theta}}\left[(F_{t-j/3} - \rho F_{t-(j+1)/3})^2\middle|\,\underline{Y_T}\right] \\ &\quad + \frac{1}{\nu^2}E_{\tilde{\theta}}\left[(h_{t-j/3} - \rho_{SV} h_{t-(j+1)/3})^2\middle|\,\underline{Y_T}\right]\Big\}\Bigg), \end{aligned}$$

up to an additive constant. The last equation can be expressed as:

$$\begin{aligned}
Q\left(\theta|\tilde{\theta}\right) \quad = \quad & -\frac{1}{2}\Bigg(T\log(1+d^2+d^4)+T\log\sigma_y^2+3T\log\sigma_x^2+3T\log\nu^2 \\
& +\sum_{t=1}^{T}\sum_{j=0}^{2}E_{\tilde{\theta}}\left[h_{t-j/3}|\underline{Y_T}\right] \\
& +\frac{1}{(1+d^2+d^4)\sigma_y^2}\sum_{t=1}^{T}\Bigg\{(y_t-d^3y_{t-1})^2 \\
& -2\gamma_1\Bigg(E_{\tilde{\theta}}\left[F_t|\,\underline{Y_T}\right]y_t-d^3\left(E_{\tilde{\theta}}\left[F_{t-1}|\,\underline{Y_T}\right]y_t+E_{\tilde{\theta}}\left[F_t|\,\underline{Y_T}\right]y_{t-1}\right)+d^6E_{\tilde{\theta}}\left[F_{t-1}|\,\underline{Y_T}\right]y_{t-1}\Bigg) \\
& +\gamma_1^2\Bigg(E_{\tilde{\theta}}\left[F_t^2|\,\underline{Y_T}\right]-2d^3E_{\tilde{\theta}}\left[F_tF_{t-1}|\,\underline{Y_T}\right]+d^6E_{\tilde{\theta}}\left[F_{t-1}^2|\,\underline{Y_T}\right]\Bigg)\Bigg\} \\
& +\frac{1}{\sigma_x^2}\sum_{t=1}^{T}\sum_{j=0}^{2}\Bigg\{x_{t-j/3}^2E_{\tilde{\theta}}\left[e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right]-2\gamma_2x_{t-j/3}E_{\tilde{\theta}}\left[F_{t-j/3}e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right] \\
& +\gamma_2^2E_{\tilde{\theta}}\left[F_{t-j/3}^2e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right] \\
& +E_{\tilde{\theta}}\left[F_{t-j/3}^2\middle|\,\underline{Y_T}\right]-2\rho E_{\tilde{\theta}}\left[F_{t-j/3}F_{t-(j+1)/3}\middle|\,\underline{Y_T}\right]+\rho^2E_{\tilde{\theta}}\left[F_{t-(j+1)/3}^2\middle|\,\underline{Y_T}\right] \\
& +\frac{1}{\nu^2}E_{\tilde{\theta}}\left[h_{t-j/3}^2\middle|\,\underline{Y_T}\right]-2\rho_{SV}E_{\tilde{\theta}}\left[h_{t-j/3}h_{t-(j+1)/3}\middle|\,\underline{Y_T}\right]+\rho_{SV}^2E_{\tilde{\theta}}\left[h_{t-(j+1)/3}^2\middle|\,\underline{Y_T}\right]\Bigg\}\Bigg).
\end{aligned} \tag{4.11.17}$$

From equation (4.11.17) we note that the estimation step requires the smoothing distribution of the factor path, in order to compute the conditional expectations $E_{\tilde{\theta}}\left[\cdot|\,\underline{Y_T}\right]$. As an exact smoother is not available, in Section 4.11.4 we propose a recursive particle smoother to compute $E_{\tilde{\theta}}\left[\cdot|\,\underline{Y_T}\right]$.

The M-step

By maximizing function $\theta \to Q\left(\theta|\tilde{\theta}\right)$, for $\tilde{\theta} = \hat{\theta}^{(i)}$ in equation (4.11.17), we get the following estimates of the model parameters collected in vector $\hat{\theta}^{(i+1)}$:

$$
\begin{aligned}
\hat{\gamma}_2 &= \frac{\sum_{t=1}^{T}\sum_{j=0}^{2} E_{\tilde{\theta}}\left[F_{t-j/3}e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right]x_{t-j/3}}{\sum_{t=1}^{T}\sum_{j=0}^{2} E_{\tilde{\theta}}\left[F^2_{t-j/3}e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right]},\\
\hat{\sigma}_x^2 &= \frac{1}{3T}\sum_{t=1}^{T}\sum_{j=0}^{2}\Bigg(E_{\tilde{\theta}}\left[e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right]x^2_{t-j/3} - 2\hat{\gamma}_2 E_{\tilde{\theta}}\left[F_{t-j/3}e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right]x_{t-j/3}\\
&\quad + \hat{\gamma}_2^2 E_{\tilde{\theta}}\left[F^2_{t-j/3}e^{-h_{t-j/3}}\middle|\,\underline{Y_T}\right]\Bigg),\\
\hat{\rho} &= \frac{\sum_{t=1}^{T}\sum_{j=0}^{2} E_{\tilde{\theta}}\left[F_{t-j/3}F_{t-(j+1)/3}\middle|\,\underline{Y_T}\right]}{\sum_{t=1}^{T}\sum_{j=0}^{2} E_{\tilde{\theta}}\left[F^2_{t-(j+1)/3}\middle|\,\underline{Y_T}\right]},\\
\hat{\rho}_{SV} &= \frac{\sum_{t=1}^{T}\sum_{j=0}^{2} E_{\tilde{\theta}}\left[h_{t-j/3}h_{t-(j+1)/3}\middle|\,\underline{Y_T}\right]}{\sum_{t=1}^{T}\sum_{j=0}^{2} E_{\tilde{\theta}}\left[h^2_{t-(j+1)/3}\middle|\,\underline{Y_T}\right]},\\
\hat{\nu}^2 &= \frac{1}{3T}\sum_{t=1}^{T}\sum_{j=0}^{2}\Bigg(E_{\tilde{\theta}}\left[h^2_{t-j/3}\middle|\,\underline{Y_T}\right] - 2\hat{\rho}_{SV}E_{\tilde{\theta}}\left[h_{t-j/3}h_{t-(j+1)/3}\middle|\,\underline{Y_T}\right]\\
&\quad + \hat{\rho}_{SV}^2 E_{\tilde{\theta}}\left[h^2_{t-(j+1)/3}\middle|\,\underline{Y_T}\right]\Bigg),
\end{aligned}
$$

and:

$$
\begin{aligned}
(\hat{\gamma}_1, \hat{d}, \hat{\sigma}_y) &= \underset{\gamma_1, d, \sigma_y}{\arg\min}\Bigg[T\log(1+d^2+d^4) + T\log\sigma_y^2 + \frac{1}{(1+d^2+d^4)\sigma_y^2}\sum_{t=1}^{T}\Bigg\{(y_t - d^3 y_{t-1})^2\\
&\quad -2\gamma_1\Bigg(E_{\tilde{\theta}}\left[F_t|\,\underline{Y_T}\right]y_t - d^3\left(E_{\tilde{\theta}}\left[F_{t-1}|\,\underline{Y_T}\right]y_t + E_{\tilde{\theta}}\left[F_t|\,\underline{Y_T}\right]y_{t-1}\right) + d^6 E_{\tilde{\theta}}\left[F_{t-1}|\,\underline{Y_T}\right]y_{t-1}\Bigg)\\
&\quad +\gamma_1^2\Bigg(E_{\tilde{\theta}}\left[F_t^2|\,\underline{Y_T}\right] - 2d^3 E_{\tilde{\theta}}\left[F_tF_{t-1}|\,\underline{Y_T}\right] + d^6 E_{\tilde{\theta}}\left[F^2_{t-1}|\,\underline{Y_T}\right]\Bigg)\Bigg\}\Bigg].
\end{aligned} \tag{4.11.18}
$$

The estimates $\hat{\gamma}_2$, $\hat{\sigma}_x^2$, $\hat{\rho}$, $\hat{\rho}_{SV}$, $\hat{\nu}^2$ of the parameters in the M-step are available in closed form, therefore they do not contribute any substantial computational cost. The parameters $\hat{\gamma}_1$, $\hat{d}$ and $\hat{\sigma}_y$, are estimated solving numerically the minimization problem in equation (4.11.18), with a negligible computational cost compared that of the filtering and smoothing algorithms proposed in the next Section 4.11.4.

4.11.4 Sequential particle filtering and smoothing algorithms

The E-step in the EM algorithm involves the smoothing distribution of the latent factors paths to compute the conditional expectation $E_{\tilde{\theta}}\left[\cdot|\,\underline{Y_T}\right]$. As an exact smoother is not available for our nonlinear SV model, we propose a sequential backward smoothing algorithm to approximate these conditional expectations. The smoothing algorithm requires, at each date $t - j/3$, for $t = 1, ..., T$ and $j = 0, 1, 2$, samples from the filtering distribution of the latent factors. For this reason we start with the description of the sequential filtering algorithm based on simulation, before describing the smoothing algorithm. The filtering algorithm proposed in the next section is based on Appendix A.1 in Kim and Stoffer (2008), in particular see their pages 816, 817, 828 and 829, and the references therein, mainly Kitagawa (1996) and Kitagawa and Sato (2001). The idea is to approximate the filtering distribution by a sample of S draws ("particles") from it, with S large. This requires an algorithm to draw from the specific distributions of our model. At the E-step of the i-th iteration of the EM algorithm, the estimate of the model parameter $\hat{\theta}^{(i)}$ is available from the previous iteration $(i-1)$-th.

In this Section, it is convenient to write the model in state space at high frequency. Let $\tau = t - j/3$, for $t = 1, ..., T$ and $j = 0, 1, 2$. The measurement is $Y_\tau = (y_\tau, x_\tau)'$ if $\tau = t$, and $Y_\tau = x_\tau$ if $\tau = t - j/3$, $j = 1, 2$. The latent factor is $f_\tau = (F_\tau, h_\tau)'$. The transition equation can be written as:

$$f_\tau = \begin{bmatrix} F_\tau \\ h_\tau \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ 0 & \rho_{SV} \end{bmatrix} \begin{bmatrix} F_{\tau-1/3} \\ h_{\tau-1/3} \end{bmatrix} + \begin{bmatrix} \eta_\tau \\ \xi_\tau \end{bmatrix}, \qquad \begin{bmatrix} \eta_\tau \\ \xi_\tau \end{bmatrix} \sim i.i.\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & \nu^2 \end{bmatrix} \right). \tag{4.11.19}$$

Sequential filtering based on importance sampling

We propose an algorithm to obtain the samples $f_\tau^{s,(i)} = \left[F_\tau^{s,(i)}, h_\tau^{s,(i)}\right]'$, with $s = 1, ..., S$, from the filtering distribution of the latent factors for parameter value $\hat{\theta}^{(i)}$, denoted as $\ell\left(f_\tau | \underline{Y_\tau}; \hat{\theta}^{(i)} \right)$, for any τ. The following steps constitute the filtering algorithm based on importance sampling with resampling:

1. Start at the first date $\tau = t - j/3 = 0$ by drawing a sample $f_0^{s,(i)}$, for $s = 1, ..., S$, from the stationary distribution of f_τ for parameter value $\hat{\theta}^{(i)}$, denoted $\ell\left(f_\tau; \hat{\theta}^{(i)} \right)$:

$$\ell\left(f_\tau; \hat{\theta}^{(i)} \right) \sim \mathcal{N}\left(0, \begin{bmatrix} \frac{1}{1-\hat{\rho}^{(i),2}} & 0 \\ 0 & \frac{\hat{\nu}^{(i),2}}{1-\hat{\rho}_{SV}^{(i),2}} \end{bmatrix} \right). \tag{4.11.20}$$

2. At date $\tau = t - j/3 \geq 1/3$, let the input be an approximation of the filtering distribution $\ell\left(f_{\tau-1/3} | \underline{Y_{\tau-1/3}}; \hat{\theta}^{(i)} \right)$ via particles $f_{\tau-1/3}^{s,(i)}$, for $s = 1, ..., S$.

   (a) Generate a sample $f_\tau^{0,s,(i)}$, $s = 1, ..., S$, from $\ell(f_\tau | \underline{Y_{\tau-1/3}}; \hat{\theta}^{(i)})$. We use $f_\tau | f_{\tau-1/3}, \underline{Y_{\tau-1/3}} \sim g\left(\cdot | f_{\tau-1/3} \right)$ where g is the transition density (see Section B.2.2).
   Hence, we draw $f_\tau^{0,s,(i)}$ from $g\left(\cdot | f_{\tau-1/3}^{s,(i)} \right)$. This is achieved by the following steps:
   (a.1)
   Generate independent random numbers:

$$\eta_\tau^{s,(i)} \sim \mathcal{N}(0, 1), \qquad \xi_\tau^{s,(i)} \sim \mathcal{N}(0, \hat{\nu}^{(i),2}),$$

for $s = 1, ..., S$.

(a.2)

Compute

$$f_\tau^{0,s,(i)} = \begin{bmatrix} F_\tau^{0,s,(i)} \\ h_\tau^{0,s,(i)} \end{bmatrix} = \begin{bmatrix} \hat{\rho}^{(i)} & 0 \\ 0 & \hat{\rho}_{SV}^{(i)} \end{bmatrix} \begin{bmatrix} F_{\tau-1/3}^{s,(i)} \\ h_{\tau-1/3}^{s,(i)} \end{bmatrix} + \begin{bmatrix} \eta_\tau^{s,(i)} \\ \xi_\tau^{s,(i)} \end{bmatrix},$$

for $s = 1, ..., S$.

(b) Generate a sample from the filtering distribution $\ell(f_\tau | \underline{Y_\tau}; \hat{\theta}^{(i)})$. We use $\ell(f_\tau | \underline{Y_\tau}; \hat{\theta}^{(i)}) \propto \ell\left(Y_\tau | \underline{Y_{\tau-1/3}}, \underline{f_\tau}; \hat{\theta}^{(i)}\right) \ell(f_\tau | \underline{Y_{\tau-1/3}}; \hat{\theta}^{(i)})$ and the importance sampling principle. Compute the weights:

$$\begin{aligned} w_\tau^{s,(i)} &\propto l\left(Y_\tau | \underline{Y_{\tau-1/3}}, \underline{f_\tau^{0,s,(i)}}; \hat{\theta}^{(i)}\right) \\ &= \begin{cases} \frac{1}{\sqrt{(2\pi)^2 \hat{\sigma}_y^{(i),2} \hat{\sigma}_x^{(i),2} \exp\{h_t^{0,s,(i)}\}}} \\ \times \exp\left\{ -\frac{[y_t - d^3 y_{t-1} - \hat{\gamma}_1^{(i)} (F_t^{0,s,(i)} - d^3 F_{t-1}^{0,s,(i)})]^2}{2\hat{\sigma}_y^{(i),2}} - \frac{(x_t - \hat{\gamma}_2^{(i)} F_t^{0,s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\left\{ h_t^{0,s,(i)} \right\}} \right\} & \tau = t, \\ \frac{1}{\sqrt{2\pi \hat{\sigma}_x^{(i),2} \exp\{h_\tau^{0,s,(i)}\}}} \exp\left\{ -\frac{(x_{t-j/3} - \hat{\gamma}_2^{(i)} F_\tau^{0,s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\left\{ h_\tau^{0,s,(i)} \right\}} \right\} & \tau = t - j/3,\ j = 1, 2, \end{cases} \end{aligned}$$

for $s = 1, ..., S$.

Then, generate $f_\tau^{s,(i)} = [F_\tau^{s,(i)}, h_\tau^{s,(i)}]'$ by resampling from $f_\tau^{0,s,(i)} = [F_\tau^{0,s,(i)}, h_\tau^{0,s,(i)}]'$ with weights $w_\tau^{s,(i)}$, for $s = 1, ..., S$.

This filtering algorithm is straightforward to implement for our model because it only requires *(i)* to simulate from the state transition density and *(ii)* evaluate the measurement density. In unreported Monte Carlo experiments, we find that the direct application of this filtering algorithm produces, in a non negligible fraction of the MC replications, degenerate filtered distribution of the latent factors. This degeneracy problem has been solved by modifying the algorithm presented in this section as an auxiliary particle filter algorithm, similarly as Pitt and Shephard (1999). See, among others, Douc, Moulines, and Olsson (2009), Carvalho, Johannes, Lopes, and Polson (2010), Doucet (2010), Lopes and Tsay (2011), Creal (2012), Kantas, Doucet, Singh, Maciejowski, and Chopin (2015), and the reference therein, for a more extensive description of auxiliary particle filter. In Section 4.11.4 we describe the auxiliary particle filter used to produce the MC results in the main body of this paper.

Sequential filtering based on auxiliary particle filter

The following steps constitute our auxiliary particle filter:

1. Start at the first date $\tau = t - j/3 = 0$ by drawing a sample $f_0^{s,(i)}$, for $s = 1, ..., S$, from the stationary distribution of f_τ for parameter value $\hat{\theta}^{(i)}$, denoted $\ell\left(f_\tau; \hat{\theta}^{(i)}\right)$ and given in (4.11.20).

2. At date $\tau = t - j/3 \geq 1/3$, let the input be an approximation of the filtering distribution $\ell\left(f_{\tau-1/3} | \underline{Y_{\tau-1/3}}; \hat{\theta}^{(i)}\right)$ via particles $f_{\tau-1/3}^{s,(i)}$, for $s = 1, ..., S$.

(a) Generate auxiliary particles $\bar{f}_\tau^{s,(i)} = \left[\bar{F}_\tau^{s,(i)}, \bar{h}_\tau^{s,(i)}\right]'$, where $\bar{F}_\tau^{s,(i)} = \hat{\rho}^{(i)} F_{\tau-1/3}^{s,(i)}$ and $\bar{h}_\tau^{(i)} = \hat{\rho}_{SV}^{(i)} h_{\tau-1/3}^{s,(i)}$, i.e. $\bar{F}_\tau^{s,(i)} = E\left[F_\tau | \, F_{\tau-1/3} = F_{\tau-1/3}^{s,(i)}; \hat{\theta}^{(i)}\right]$ and $\bar{h}_\tau^{s,(i)} = E\left[h_\tau | \, h_{\tau-1/3} = h_{\tau-1/3}^{s,(i)}; \hat{\theta}^{(i)}\right]$.

(b) The auxiliary particles are used to define weights and resample from the old particles $f_{\tau-1/3}^{s,(i)}$. Specifically, compute the weights:

$$\begin{aligned}
\breve{w}_\tau^{s,(i)} &\propto \ell\left(Y_\tau | \, \underline{Y_{\tau-1/3}}, \underline{\bar{f}_\tau^{s,(i)}}; \hat{\theta}^{(i)}\right) \\
&= \begin{cases}
\frac{1}{\sqrt{(2\pi)^2 \hat{\sigma}_y^{(i),2} \hat{\sigma}_x^{(i),2} \exp\{\bar{h}_t^{s,(i)}\}}} & \\
\times \exp\left\{-\frac{[y_t - d^3 y_{t-1} - \hat{\gamma}_1^{(i)}(\bar{F}_t^{s,(i)} - d^3 \bar{F}_{t-1}^{(i)})]^2}{2\hat{\sigma}_y^{(i),2}} - \frac{(x_t - \hat{\gamma}_2^{(i)} \bar{F}_t^{s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\left\{\bar{h}_t^{s,(i)}\right\}}\right\} & \tau = t, \\
\frac{1}{\sqrt{2\pi \hat{\sigma}_x^{(i),2} \exp\{\bar{h}_\tau^{s,(i)}\}}} \exp\left\{-\frac{(x_\tau - \hat{\gamma}_2^{(i)} \bar{F}_\tau^{s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\left\{\bar{h}_\tau^{s,(i)}\right\}}\right\} & \tau = t - j/3, \ j = 1, 2
\end{cases},
\end{aligned}$$

for $s = 1, ..., S$. Generate particles $\breve{f}_{\tau-1/3}^{s,(i)} = [\breve{F}_{\tau-1/3}^{s,(i)}, \breve{h}_{\tau-1/3}^{s,(i)}]'$ by resampling $f_{\tau-1/3}^{s,(i)} = [F_{\tau-1/3}^{s,(i)}, h_{\tau-1/3}^{s,(i)}]'$ with weights $\breve{w}_\tau^{s,(i)}$, $s = 1, ..., S$.

(c) Generate a sample from $\ell(f_\tau | \underline{Y_{\tau-1/3}}; \hat{\theta}^{(i)})$. We use $f_\tau | \, f_{\tau-1/3}, \underline{Y_{\tau-1/3}} \sim g\left(\cdot | \, f_{\tau-1/3}\right)$. We draw $f_\tau^{0,s,(i)}$ from $g\left(\cdot | \, \breve{f}_{\tau-1/3}^{s,(i)}\right)$. This is achieved by:

(c.1)

Generate independent random numbers:

$$\eta_\tau^{s,(i)} \sim \mathcal{N}(0, 1), \qquad \xi_\tau^{s,(i)} \sim \mathcal{N}(0, \hat{\nu}^{(i),2}),$$

for $s = 1, ..., S$.

(c.2)

Compute

$$f_\tau^{0,s,(i)} = \begin{bmatrix} F_\tau^{0,s,(i)} \\ h_\tau^{0,s,(i)} \end{bmatrix} = \begin{bmatrix} \hat{\rho}^{(i)} & 0 \\ 0 & \hat{\rho}_{SV}^{(i)} \end{bmatrix} \begin{bmatrix} \breve{F}_{\tau-1/3}^{s,(i)} \\ \breve{h}_{\tau-1/3}^{s,(i)} \end{bmatrix} + \begin{bmatrix} \eta_\tau^{s,(i)} \\ \xi_\tau^{s,(i)} \end{bmatrix},$$

for $s = 1, ..., S$.

(d) Compute the weights:

$$\begin{aligned}
w_\tau^{s,(i)} &\propto \frac{\ell\left(Y_\tau | \, \underline{Y_{\tau-1/3}} \underline{f_\tau^{0,s,(i)}}; \hat{\theta}^{(i)}\right)}{\ell\left(Y_\tau | \, \underline{Y_{\tau-1/3}}, \underline{\bar{f}_\tau^{s,(i)}}; \hat{\theta}^{(i)}\right)} \\
&= \begin{cases}
\frac{\sqrt{\exp\{\bar{h}_\tau^{s,(i)}\}}}{\sqrt{\exp\{h_\tau^{0,s,(i)}\}}} \exp\left\{-\frac{[y_t - d^3 y_{t-1} - \hat{\gamma}_1^{(i)}(F_t^{0,s,(i)} - d^3 F_{t-1}^{0,s,(i)})]^2}{2\hat{\sigma}_y^{(i),2}} - \frac{(x_t - \hat{\gamma}_2^{(i)} F_t^{0,s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\left\{h_t^{0,s,(i)}\right\}}\right\} & \\
\times \exp\left\{+\frac{[y_t - d^3 y_{t-1} - \hat{\gamma}_1^{(i)}(\bar{F}_t^{s,(i)} - d^3 \bar{F}_{t-1}^{s,(i)})]^2}{2\hat{\sigma}_y^{(i),2}} + \frac{(x_t - \hat{\gamma}_2^{(i)} \bar{F}_t^{s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\left\{\bar{h}_t^{s,(i)}\right\}}\right\} & \tau = t, \\
\frac{\sqrt{\exp\{\bar{h}_{\tau-1/3}^{s,(i)}\}}}{\sqrt{\exp\{h_\tau^{0,s,(i)}\}}} & \\
\times \exp\left\{-\frac{(x_\tau - \hat{\gamma}_2^{(i)} F_\tau^{0,s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\left\{h_\tau^{0,s,(i)}\right\}} + \frac{(x_\tau - \hat{\gamma}_2^{(i)} \bar{F}_\tau^{s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\left\{\bar{h}_\tau^{s,(i)}\right\}}\right\} & \tau = t - j/3, \ j = 1, 2,
\end{cases}
\end{aligned}$$

Generate $f_{\tau}^{s,(i)} = [F_{\tau}^{s,(i)}, h_{\tau}^{s,(i)}]'$ by resampling from $f_{\tau}^{0,s,(i)} = [F_{\tau}^{0,s,(i)}, h_{\tau}^{0,s,(i)}]'$ with weights $w_{\tau}^{s,(i)}$, for $s = 1, ..., S$.

Sequential smoothing with importance sampling

Any EM algorithm requires the computation of the smoothing distribution of the latent factor path only once at each iteration i. In our specific case, we need to compute only some moments of the smoothing distribution of the latent factor path. Specifically we need to sample from the smoothing distribution of the factors at each high frequency date $\tau = t - j/3$, with $t = 1, ..., T$, and $j = 0, 1, 2$. Let $\tilde{f}_{\tau}^{s,(i)} = \left(\tilde{F}_{\tau}^{s,(i)}, \tilde{h}_{\tau}^{s,(i)}\right)'$, with $s = 1, ..., S$, be a sample from the smoothing distribution of the latent factors at time τ, obtained using the estimated model parameters $\hat{\theta}^{(i)}$, i.e. during the i-th iteration of the EM algorithm. Then, the conditional expectations $E_{\tilde{\theta}}\left[\cdot | \underline{Y_T}\right]$ of the E-step for $\tilde{\theta} = \hat{\theta}^{(i)}$ can be approximated by sample averages of the S particles.

The smoothing algorithm is based on Appendix A.2 in Kim and Stoffer (2008) and the reference therein, mainly Godsill, Doucet, and West (2004), and is free from degeneracy. It uses the particles of the filtering distribution as input. Specifically, let $f_{\tau}^{s,(i)} = \left(F_{\tau}^{s,(i)}, h_{\tau}^{s,(i)}\right)'$, with $s = 1, ..., S$, be draws from the filtering distribution $\ell\left(f_{\tau} | \underline{Y_{\tau}}; \hat{\theta}^{(i)}\right)$ for date $\tau = t - j/3$, $j = 0, 1, 2$ (see Section B.4.2). The following steps constitute the backward sequential smoothing algorithm. For any $s = 1, ..., S$:

1. Start at the last date $\tau = T$, and draw $\tilde{f}_{T}^{s,(i)}$ from set $\{f_{T}^{r,(i)},\ r = 1, ..., S\}$ with equal weights $1/S$. In other words, we obtain one draw from the filtering distribution $\ell\left(f_T | \underline{Y_T}; \hat{\theta}^{(i)}\right)$ at the last sample date.

2. For any date τ, from $\tau = T - 1/3$ to $\tau = 0$:

   (a) Compute the weights:

   $$w_{\tau,\tau+1/3}^{r,(i)} \quad \propto \quad g(\tilde{f}_{\tau+1/3}^{s,(i)} | f_{\tau}^{r,(i)}; \hat{\theta}^{(i)}),$$

   for $r = 1, ..., S$.

   (b) Draw $\tilde{f}_{\tau}^{s,(i)}$ from $\{f_{\tau}^{r,(i)},\ r = 1, ..., S\}$ with probability weights $\{w_{\tau,\tau+1/3}^{r,(i)},\ r = 1, ..., S\}$.

At the end of the smoothing algorithm we have S simulated paths $\{(\tilde{f}_{0}^{s,(i)}, \tilde{f}_{1/3}^{s,(i)}, ..., \tilde{f}_{T}^{s,(i)}), s = 1, ..., S\}$ from the smoothing distribution $\ell\left(f_0, f_{1/3}, ..., f_T \middle| \underline{Y_T}; \hat{\theta}^{(i)}\right)$. Note that the second step of our backward sequential smoothing algorithm requires only *(i)* a sample from the filtering distribution which is already available from the filtering algorithm, and *(ii)* to be able to evaluate the transition density.

Stopping rule

For the Monte Carlo EM algorithm to converge to the MLE estimate, the number of particles S needs to increase with the number of EM iterations, see for instance Olsson, Cappé, Douc, and Moulines (2008), Neath (2013) and the references therein. Moreover, a rule needs to be set in order to stop the algorithm and assess its convergence. We follow the same procedure of Kim and Stoffer (2008). On the basis of the work of Chan and Ledolter (1995), Kim and Stoffer (2008) start the EM algorithm with a small value of S to save computing time, at the end of each EM iteration compute ϵ - the estimated change in log-likelihood with respect to the previous EM iteration - and increase S when ϵ is below a certain small lower bound. The EM algorithm in our paper is implemented starting with a number of particles $S = 500$, then the value of S is increased to 1000 as soon as $\epsilon < 0.10$ for a certain iteration, then S is increased to 1500 when $\epsilon < 0.05$, and finally the algorithm is stopped at the first iteration in which $\epsilon < 0.01$. The values of ϵ and S, together with the stopping rule, were calibrated in preliminary unreported MC experiments. See Kim and Stoffer (2008) and Chan and Ledolter (1995) for an in-depth analysis of this procedure and the concept of "Relative Likelihood".

4.12 Appendix C: Linear state space model with one flow sampled low frequency variable

In this Appendix we show one way to adapt the measurement and transition equations of the linear state space model with mixed frequency data in Bai, Ghysels, and Wright (2013), for the case of one low frequency variable which is flow sampled. This adaptation is necessary to implement the Kalman filter algorithm for ML estimation of the structural model without SV used in the empirical application of Section 4.6.

4.12.1 Model setup

The linear state space model without SV considered in the empirical application in Section 4.6 has one flow sampled low frequency variable y_t, with $t = 1, ..., T$, $m = 3$ high frequency subperiods and a single latent factor (i.e. $n_f = 1$). The latent factor follows an $AR(1)$ process:

$$F_{t+j/3} = \Phi_1 F_{t+(j-1)/3} + \eta_{t+j/3} \qquad t = 1, \ldots, T, \qquad j = 0, 1, 2, \tag{4.12.1}$$

where Φ_1 is a scalar parameter to be estimated. Latent process y^* is related to the factor as follows:

$$y^*_{t+j/3} = \gamma_1 F_{t+j/3} + u_{1,t+j/3} \qquad t = 1, \ldots, T, \qquad j = 0, 1, 2, \tag{4.12.2}$$

with $u_{1,t+j/3}$ having an $AR(1)$ representation:

$$u_{1,t+j/3} = d_1 u_{1,t+(j-1)/3} + \varepsilon_{1,t+j/3}. \tag{4.12.3}$$

The observed low-frequency process y is flow sampled, i.e. it relates to the latent process y^* in the following way:

$$y_t = y^*_t + y^*_{t-1/3} + y^*_{t-2/3}, \qquad t = 1, .., T, \qquad j = 0, 1, 2. \tag{4.12.4}$$

The high frequency process $x_{t+j/3}$ relates to the factor as follows:

$$x_{t+j/3} = \gamma_2 F_{t+j/3} + u_{2,t+j/3} \qquad t = 1, \ldots, T, \qquad j = 0, 1, 2, \tag{4.12.5}$$

where:

$$u_{2,t+j/3} = d_2 u_{2,t+(j-1)/3} + \varepsilon_{2,t+j/3}. \tag{4.12.6}$$

The innovations (η), (ε_1), (ε_2) are mutually independent i.i.d. Gaussian processes, with distributions $\mathcal{N}(0, 1)$, $\mathcal{N}(0, \sigma^2_{\varepsilon_1})$ and $\mathcal{N}(0, \sigma^2_{\varepsilon_2})$.

4.12.2 State space representation and Kalman filter

The above equations yield a periodic state space model with measurement equation:

$$Y^j_t = Z_j \alpha_{t+j/m} \qquad \begin{cases} Y^j_t = (y_t, x_t)' & j = 0 \\ Y^j_t = x_{t+j/m} & j = 1, 2 \end{cases}, \tag{4.12.7}$$

for $t = 1, ..., T$, where

$$Z_0 = \begin{bmatrix} \gamma_1 & \gamma_1 & \gamma_1 & 1 & 1 & 1 & 0 \\ \gamma_2 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$Z_j = \left[\begin{array}{ccccccc} \gamma_2 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right], \qquad j = 1, 2,$$

and state vector:

$$\alpha_{t+j/3} = \left(F_{t+j/3}, F_{t+(j-1)/3}, F_{t+(j-2)/3}, u_{1,t+j/3}, u_{1,t+(j-1)/3}, u_{1,t+(j-2)/3}, u_{2,t+j/3}\right)'.$$

The transition equation is:

$$\alpha_{t+j/m} = R\alpha_{t+(j-1)/3} + Q\zeta_{t+j/3}, \qquad t = 1, .., T, \qquad j = 0, 1, 2, \tag{4.12.8}$$

where

$$R = \left[\begin{array}{ccccccc} \Phi_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & d_2 \end{array} \right], \qquad Q = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{array} \right],$$

$\zeta_{t+j/m}$ = $(\eta_{t+j/m}, \varepsilon_{1,t+j/m}, \varepsilon_{2,t+j/m})'$, and $\Sigma_\zeta = diag(1, \sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_2})$ denotes the variance-covariance matrix of $\zeta_{t+j/m}$. Then, the Kalman filter algorithm presented in Appendix A can be performed after replacing matrices Z_0, Z_j, R, Q and Σ_ζ by the new definitions.

4.13 Bibliograpghy of Chapter 4

ANDREOU, E., E. GHYSELS, AND A. KOURTELLOS (2010): “Regression Models with Mixed Sampling Frequencies,” *Journal of Econometrics*, 158, 246–261.

——— (2011): “Forecasting with Mixed-frequency Data,” in *Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements, and D. F. Hendry, pp. 225–245.

——— (2013): “Should Macroeconomic Forecasters use Daily Financial Data and how?,” *Journal of Business and Economic Statistics*, 31, 240–251.

ARUOBA, S. B., F. X. DIEBOLD, AND C. SCOTTI (2009): “Real-time Measurement of Business Conditions,” *Journal of Business and Economic Statistics*, 27(4), 417–427.

ASSIMAKIS, N., AND M. ADAM (2009): “Steady state Kalman Filter for Periodic Models: a New Approach,” *International Journal of Contemporary Mathematical Sciences*, 4, 201–218.

BAI, J., E. GHYSELS, AND J. H. WRIGHT (2013): “State Space Models and MIDAS Regressions,” *Econometric Reviews*, 32, 779–813.

BANBURA, M., D. GIANNONE, M. MODUGNO, AND L. REICHLIN (2013): “Now-casting and the Real-time Data Flow,” ECB working paper.

BANBURA, M., AND M. MODUGNO (2014): “Maximum Likelihood Estimation of Factor Models on Data Sets with Arbitrary Pattern of Missing Data,” *Journal of Applied Econometrics*, 29(1), 133–160.

BARIGOZZI, M., R. HALBLEIB-CHIRIAC, AND D. VEREDAS (2014): “Which Model to Match?,” Available at SSRN 1986419.

BERNANKE, B., M. GERTLER, AND M. WATSON (1997): “Systematic Monetary Policy and the Effects of Oil Price Shocks,” *Brookings Papers on Economic Activity*, 1, 91–157.

CARRIERO, A., T. CLARK, AND M. MARCELLINO (2012): “Common Drifting Volatility in Large Bayesian VARs,” CEPR Discussion Paper No. 8894.

——— (2013): “Real-Time Nowcasting with Large Bridging Equations with Time-Varying Volatility,” CEPR Discussion Paper No. 9312.

CARVALHO, C. M., M. S. JOHANNES, H. F. LOPES, AND N. G. POLSON (2010): “Particle Learning and Smoothing,” *Statistical Science*, 25(1), 88–106.

CHAN, K. S., AND J. LEDOLTER (1995): “Monte Carlo EM Estimation for Time Series Models Involving Counts,” *Journal of the American Statistical Association*, 90(429), 242–252.

CHEN, X., AND E. GHYSELS (2010): “News - Good or Bad - and its Impact on Volatility Predictions over Multiple Horizons,” *Review of Financial Studies*, pp. 46–81.

CLARK, T. E. (2011): "Real-time Density Forecasts from Bayesian Vector Autoregressions with Stochastic Volatility," *Journal of Business and Economic Statistics*, 29, 327–341.

COX, D. (1981): "Statistical Analysis of Time Series: Some Recent Developments [with Discussion and Reply]," *Scandinavian Journal of Statistics*, 8, 93–115.

CREAL, D. (2012): "A survey of Sequential Monte Carlo Methods for Economics and Finance," *Econometric Reviews*, 31(3), 245–296.

DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977): "Maximum Likelihood from Incomplete Data via di EM Algorithm," *Journal of the Royal Statistical Society Series B*, 39(1), 1–38.

DOUC, R., E. MOULINES, AND J. OLSSON (2009): "Optimality of the Auxiliary Particle Filter," *Probability and Mathematical Statistics*, 29, 1–28.

DOUCET, A., J. A. (2010): "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," in *Handbook of Nonlinear Filtering*, ed. by D. Crisan, and B. Rozovsky. Oxford University Press: Oxford.

ENGLE, R. F., AND G. G. LEE (1999): "A Permanent and Transitory Component Model of Stock Return Volatility," in *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, ed. by R. Engle, and H. White, pp. 475–497. Oxford University Press, Oxford.

FORONI, C., AND M. MARCELLINO (2013): "A Survey of Econometric Methods for Mixed-frequency Data," Available at SSRN 2268912.

FORONI, C., M. MARCELLINO, AND C. SCHUMACHER (2013): "Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials," Journal of the Royal Statistical Society: Series A.

GALLANT, A. R., AND G. TAUCHEN (1996): "Which Moments to Match?," *Econometric Theory*, 12, 657–681.

——— (1998): "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions," *Journal of the American Statistical Association*, 93, 10–24.

GHYSELS, E. (2014): "Macroeconomics and the Reality of Mixed Frequency Data," *Journal of Econometrics* (forthcoming).

GHYSELS, E., A. HARVEY, AND E. RENAULT (1996): "Stochastic Volatility," in *Handbook of Statistics, Volume 14 - Statistical Methods in Finance*, ed. by G. Maddala, and C. Rao. North Holland.

GHYSELS, E., J. B. HILL, AND K. MOTEGI (2014): "Regression-based Mixed Frequency Granger Causality Tests," Discussion paper, Technical report, Department of Economics, University of North Carolina, Chapel Hill.

——— (2016): "Granger Causality in Mixed Frequency Vector Autoregressive Models," *Journal of Econometrics* - forthcoming.

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2006): "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies," *Journal of Econometrics*, 131, 59–95.

GHYSELS, E., AND J. WRIGHT (2009): "Forecasting Professional Forecasters," *Journal of Business and Economic Statistics*, 27, 504–516.

GIANNONE, D., L. REICHLIN, AND D. SMALL (2008): "Nowcasting: the Real-time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, 55, 665–676.

GODSILL, S., A. DOUCET, AND M. WEST (2004): "Monte Carlo Smoothing for Non-linear Time Series," *Journal of the American Statistical Association*, 199, 156–168.

GÖTZ, T. B., AND A. HECQ (2014a): "Nowcasting Causality in Mixed Frequency Vector Autoregressive Models," *Economics Letters*, 122, 74–78.

——— (2014b): "Testing for Granger Causality in Large Mixed-Frequency VARs," Discussion paper, Technical report, Department of Economics, University of Maastricht.

GOURIÉROUX, C., AND A. MONFORT (1997): *Simulation-based Econometric Methods*. Oxford University Press.

GOURIÉROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.

GUAY, A., AND O. SCAILLET (2003): "Indirect Inference, Nuisance Parameter, and Threshold Moving Average Models," *Journal of Business and Economic Statistics*, 21, 122–132.

HAMILTON, J. D. (1994): "State-space Models," in *Handbook of Econometrics - Volume 4*, ed. by R. Engle, and D. McFadden, pp. 3039–3080. Elsevier.

HARVEY, A. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

HARVEY, A. C., AND R. G. PIERSE (1984): "Estimating Missing Observations in Economic Time Series," *Journal of the American Statistical Association*, 79, 125–131.

KANTAS, N., A. DOUCET, S. S. SINGH, J. M. MACIEJOWSKI, AND N. CHOPIN (2015): "On Particle Methods for Parameter Estimation in General State-Space Models," *Statistical Science*, 30(3), 328–351.

KIM, J., AND D. STOFFER (2008): "Fitting Stochastic Volatility Models in the Presence of Irregular Sampling Via Particle Methods and the EM Algorithm," *Journal of Time Series Analysis*, 29, 811–833.

KITAGAWA, G. (1996): "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State-space Models," *Journal of Computational and Graphical Statistics*, 5, 1–25.

KITAGAWA, G., AND S. SATO (2001): "Sequential Monte Carlo Methods in Practice," in *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

KUZIN, V., M. MARCELLINO, AND C. SCHUMACHER (2011): "MIDAS versus Mixed-frequency VAR: Nowcasting GDP in the Euro area," *International Journal of Forecasting*, 27, 529–542.

LOPES, H. F., AND R. TSAY (2011): "Particle filters and Bayesian Inference in Financial Econometrics," *Journal of Forecasting*, 30, 168–209.

MARCELLINO, M., M. PORQUEDDU, AND F. VENDITTI (2015): "Short-term GDP Forecasting with a Mixed Frequency Dynamic Factor Model with Stochastic Volatility," *Journal of Business and Economic Statistics* - forthcoming.

MARCELLINO, M., AND C. SCHUMACHER (2010): "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP," *Oxford Bulletin of Economics and Statistics*, 72(4), 518–550.

MARIANO, R. S., AND Y. MURASAWA (2003): "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics*, 18(4), 427–443.

MITTNIK, S., AND P. ZADROZNY (2005): "Forecasting Quarterly German GDP at Monthly Intervals Using Monthly Ifo Business Conditions Data," in *Ifo Survey Data in Business Cycle and Monetary Policy Analysis*, ed. by J.-E. Sturm, and W. T. Springer.

MONFARDINI, C. (1998): "Estimating Stochastic Volatility Models through Indirect Inference," *The Econometrics Journal*, 1, 113–128.

NEATH, R. C. (2013): "On Convergence Properties of the Monte Carlo EM Algorithm," *Institute of Mathematical Statistics Collections*, 10, 43–62.

NUNES, L. (2005): "Nowcasting Quarterly GDP Growth in a Monthly Coincident Indicator Model," *Journal of Forecasting*, 24, 575–592.

OLSSON, J., O. CAPPÉ, R. DOUC, AND E. MOULINES (2008): "Sequential Monte Carlo Smoothing with Application to Parameter Estimation in Nonlinear State Space Models," *Bernoulli*, 14(1), 155–179.

PASTORELLO, S., E. RENAULT, AND N. TOUZI (2000): "Statistical Inference for Random-variance Option Pricing," *Journal of Business and Economic Statistics*, 18, 358–367.

PITT, M. K., AND N. SHEPHARD (1999): "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94(446), 590–599.

SANDMANN, G., AND S. J. KOOPMAN (1998): "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood," *Journal of Econometrics*, 87, 271–301.

SCHORFHEIDE, F., AND D. SONG (2013): "Real-time Forecasting with a Mixed-frequency VAR," Discussion paper, National Bureau of Economic Research.

SENTANA, E., G. CALZOLARI, AND G. FIORENTINI (2008): "Indirect Estimation of Large Conditionally Heteroskedastic Factor Models, with an Application to the Dow 30 Stocks," *Journal of Econometrics*, 146, 10–25.

SHEPHARD, N. (2005): *Stochastic Volatility: Selected Readings*. Oxford University Press.

SMITH, A. A. (1993): "Estimating Nonlinear Time-series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, 63–84.

ZADROZNY, P. (1988): "Gaussian Likelihood of Continuous-time ARMAX Models when Data are Stocks and Flows at Different Frequencies," *Econometric Theory*, 4, 108–124.

ZADROZNY, P. (1990): "Forecasting US GNP at Monthly Intervals with an Estimated Bivariate Time Series Model," *Federal Reserve Bank of Atlanta Economic Review*, 75, 2–15.