Is the US Economy Still Mostly Driven by Manufacturing?*

E. Andreou[†] P. Gagliardini[‡] E. Ghysels[§] M. Rubin[¶]

First version: October, 2014
This version: April 13, 2015

April 13, 2015

Abstract

We study the decomposition of US output and examine whether it is driven by aggregate or sector-specific shocks. While the use of factor models has been found convenient, the challenge one faces is that sectoral data beyond industrial production (IP) is only available annually. This imbalance of sampling frequencies poses challenges which existing methods have not been able to resolve. We propose a new class of mixed frequency data factor models which enable us to study to full spectrum of monthly or quarterly IP sector data combined with the annual non-IP sectors of the economy. We derive the large sample properties of estimator for the new class of factor models involving mixed frequency data. Using our new approximate factor model, we find that nearly all of US output variability is associated with common factors, as opposed to sector-specific factors. We also use a multi-sector growth model to incorporate input-output linkages in the factor analysis.

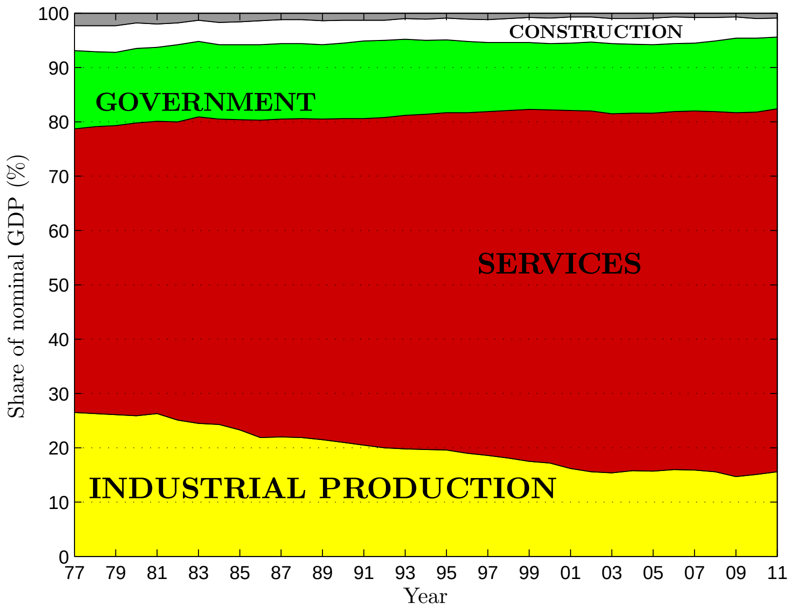
*We thank M. Deistler, D. Giannone, T. Götz, M. Del Negro, K. A. Aastveit, C. Hurlin, C. Pérignon and participants at the 2013 $(EC)^2$ Conference on "The Econometric Analysis of Mixed Frequency Data" in Nicosia, the 2014 "Eighth ECB Workshop on Forecasting Techniques" in Frankfurt am Main, the 2014 "68^{th} European Meeting of the Econometric Society" in Toulouse, and the "8^{th} Financial Risk International Forum on Scenarios, Stress and Forecast in Finance" in Paris for useful comments.

[†]University of Cyprus and CEPR (elena.andreou@ucy.ac.cy).
[‡]Università della Svizzera Italiana and Swiss Finance Institute (patrick.gagliardini@usi.ch).
[§]University of North Carolina - Chapel Hill and CEPR (eghysels@unc.edu).
[¶]Università della Svizzera Italiana and Swiss Finance Institute (mirco.rubin@usi.ch).

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 question somewhat. It is true, as the figure below clearly shows, that the share of the manufacturing sector has been in decline since the late 70's, which is the beginning of our sample period.[1] The fact that the size shrank does not necessarily exclude that the manufacturing sector still is a key driver 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 1: Sectoral decomposition of US nominal GDP.



To address the role of the manufacturing sector in the US economy 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 using NAICS. This data is available monthly and therefore covers a rich time series and cross-section of over 47,000 data points counting all months from 1977 until 2011 (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 GDP 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.

[1]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 NAICS codes, in 5 different *macro* sectors: Industrial Production (yellow), Services (red), Government (green), Construction (white), Others (grey).

Economists have proposed different models about 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. Do the factors which drive 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 say 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 manufacturing shocks, or rather services sector ones? Various models featuring sector-specific shocks exist. For example, Gabaix (2011) proposes a model with a handful of very large sectors which are the source of fluctuations in the economy. Another class of models postulates complementarities across sectors, such as input-output linkages, that propagate sector specific shocks throughout the economy.

Does the paradigm of common shocks affecting IP as identified by Foerster, Sarte, and Watson (2011) extend to the economy at large? Or, are sector-specific shocks, either originated in IP or else non-IP sectors the driving force of the US economy? Our analysis is much inspired by Foerster, Sarte, and Watson (2011), who use a multi-sector growth model to adjust for the effects of input-output linkages in the factor analysis. However, to conduct our investigation we need to be analyze all sectors of the economy in a unified framework. If we only use annual data, as it would appear to be our only choice, we forgo a vast wealth of IP sector data available at monthly or quarterly frequency.

We propose a new class of factor models able to address the key question of interest using *all* the data - despite the mixed sampling frequency. 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 f^C, and therefore are economy-wide drivers, (2) those exclusively pertaining to IP sector movements - say f^H, and finally (3) those exclusively affecting non-IP, denoted by f^L. Hence, we have (1) common, (2) high frequency and (3) low frequency factors.

The purpose of this paper is to propose large scale 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.[2] Under suitable regularity conditions, the factor values and loadings can be

[2] A number of mixed frequency factor models have been proposed in the literature, although they almost exclusively rely on small cross-sections. 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, Banbura and Modugno (2014) build on the work of Doz, Giannone, and Reichlin (2012) and Giannone, Reichlin, and Small (2008) and extend the EM algorithm of Rubin and Thayer (1982) for the estimation by (quasi) maximum likelihood of a factor model from the same type of unbalanced and mixed frequency panels.

identified and estimated exploiting results from a literature on grouped factor models, see for example Krzanowski (1979), Flury (1984), Kose, Otrok, and Whiteman (2008), Goyal, Pérignon, and Villa (2008), Bekaert, Hodrick, and Zhang (2009), Wang (2012), Moench and Ng (2011), Moench, Ng, and Potter (2013), Ando and Bai (2013) and Breitung and Eickmeier (2014) among others. In the proposed identification strategy, the groups correspond to panels observed at different sampling frequencies. Our theoretical contributions are twofold. First, while our work is most closely related to Wang (2012) and Chen (2010, 2012) we provide a more comprehensive asymptotic treatment of grouped factor models.[3]. We do not rely on the iterative solution from a Least Square (LS) problem, since the resulting equations do not have a unique solution. 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. Second, the idea to apply grouped factor analysis to mixed frequency data is novel and has many advantages in terms of identification and estimation which are discussed in the paper.

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. Foerster, Sarte, and Watson (2011) 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. The US economy has become more of 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. The US Bureau of Economic Analysis provides GDP by industry - not only IP sectors - annually. We identify three factors in a 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. We re-examine whether the 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 as it omitted a major ingredient of US economic activity. 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 mechanism are very different, however, from those identified by Foerster, Sarte, and Watson (2011).

[3]We investigate in depth the link between canonical analysis and the Principal Components (PC) estimated on each group on the one hand, and spectral analysis of the variance-covariance matrix of stacked PC's on the other hand.

References

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 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.

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.

Breitung, J., and S. Eickmeier (2014): "Analyzing business and financial cycles using multi-level factor models," *Deutsche Bundesbank, Research Centre, Discussion Papers 11/2014*.

Chen, P. (2010): "A Grouped Factor Model," *Working Paper*.

——— (2012): "Common Factors and Specific Factors," *Working Paper*.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum likelihood estimation from incomplete data," *Journal of the Royal Statistical Society*, 14, 1–38.

Doz, C., D. Giannone, and L. Reichlin (2012): "A quasi maximum likelihood approach for large approximate dynamic factor models," *Review of Economics and Statistics*, 94(4), 1014–1024.

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., 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.

GIANNONE, D., L. REICHLIN, AND D. SMALL (2008): "Nowcasting: the Real-time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, 55(4), 665–676.

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.

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.

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.

RUBIN, D., AND D. THAYER (1982): "EM Algorithms for ML Factor Analysis," *Psychometrica*, 47, 69–76.

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.

WANG, P. (2012): "Large Dimensional Factor Models with a Multi-Level Factor Structure: Identifcation, Estimation, and Inference," *Working Paper*.