Commodity Prices and the Business Cycle in resource-dependent Economies

Nikolas Kuschnig, Casper Engelen & Matthias Hagen
WU Vienna, 5741 - Money, Credit & Finance

06 August, 2018

Contents

Abstract

Abstract

Introduction

Intro

Principal component Analysis (PCA) can be used to reduce the dimensionality of a dataset with many interrelated variables, whilst also conserving as much of the variation present in the data set as possible (Abdi & Williams, 2010, Jolliffe, 2011, Stock & Watson, 2002a, Stock & Watson, 2002b). To achieve this, PCA focuses on variances, without neglecting covariances and correlations. Due to its versatile nature, PCA is one of the most widely used multivariate statistical techniques and has found application in a multitude of scientific disciplines (Abdi & Williams, 2010). The variables created in the PCA process can be useful for forecasting in macroeconomic analysis (Favero et al., 2005, Stock & Watson, 2002a, Stock & Watson, 2002b), where many different variables and time series are generally available. Since we want to model the effect of commodity prices on business cycles within a VAR framework, and there are many different commodities that are relevant to this analysis, it is useful to reduce the number of variables that are used in the estimation. VAR models benefit greatly from being set up parsimoniously and the reduction of dimensionality that can be achieved via PCA should lead to better results.

Methodology

Data

In pursuit of a scientifically sound foundation to build further analysis on, the initial concern of this paper was the definition and selection of *resource-dependent* countries. To avoid arbitrariness this was done in a data-based manner. For this, World Bank data on economic variables, resource rents and metal exports were utilised in the creation of four indicators of resource-dependence. These are the ratios of subsoil wealth to GDP, natural to total wealth, resource rents to GDP and metal exports to total exports. Together with constraints on size of the economy and ultimately data availability, these indicators were used to single out Australia, Chile, Norway and South Africa for our analysis. The major source for economic data was the OECD, with timeseries of quarterly GDP, trade, consumer prices, interest rates and many more. While a wide array of data is readily available, length and frequency of timeseries and comparability between countries posed some challenges. E.g. vital timeseries on monetary policy rates were mostly gathered from national central banks and - in the case of Australia and Norway - had to be extended with short-term interbank rates. Data on unemployment rates, exchange rates and other economic indicators were dropped due to issues with heterogeneity and availability. Market data was gathered via Bloomberg and Datastream and includes over forty commodity indices, individual resource prices and equity indices. However, the bulk of these timeseries only started in the 1990s, highlighting the dilemma of longevity versus abundance of data. For a more detailed overview of used data and its sources please consult Table 10.1 in the appendix.
To achieve stationarity the variables, apart from interest rates, were transformed. Three different approaches were up for consideration - namely first-differences, log-differences and a Hodrick-Prescott filter. Of these, log-differences seemed to perform the best, with the exception of GDP, which was filtered via an HP filter, and the principal components of the resource data, where first-differences were applied.

Vector Autoregressive Models

VAR

Variable Selection

If we consider the vector x with p random variables, where we are interested in the variances of the random variables and the structure of covariances and correlations between them. If the number of random variables is high, or the structure is complex, it is often helpful to look for few derived variables, that preserve the information of the variances, correlations and covariances, instead of examining them directly (Jolliffe, 2011). When using PCA these new, derived variables are called the principal components. They are uncorrelated and retain most of the variation present in the entire dataset (Jolliffe, 2011). Initially PCA finds a linear function $\alpha_1'x$ of the elements of x while maximizing variance, where α_1' is a vector of p constants, such that it can be described as $(\alpha_1 1, \alpha_1 2, ..., \alpha_1 p)'$. Hence:

$$\alpha_1'x = \alpha_1 1x_1 + \alpha_1 2x_2 + ... + \alpha_1 px_p = \sum_{j=1}^{p} \alpha_1 jx_j$$

Next, PCA looks for a linear function $\alpha_2'x$ that is uncorrelated with α_1' while also having maximum variance, and so on, so that the k-th stage yields the linear function $\alpha_k'x$, again uncorrelated with all linear function $\alpha_1'x, ..., \alpha_k - 1'x$ and with maximum variance. Up to p principal components can be distinguished, although the aim is to find an amount m, such that $m << p$, thus reducing the dimensionality of the dataset. The values of these resulting variables are called *factor scores*, and form their own timeseries (Abdi & Williams, 2010, Stock & Watson, 2002b).
For PCA to be successful, the considered variables should be correlated, but not perfectly so, since this would hinder a reduction in the dimensionality. To evaluate wether PCA can be used to reduce dimensionality one may use the *KMO-Criterion* and the p-value of *Bartlett's Test of Sphericity*. Afterwards, a balance has to be struck between providing a sufficient amount of observations to the VAR model, whilst also capturing as much information as possible from the commodity markets. There are several established criteria for handling this tradeoff, although the exact cut-off points are somewhat subjective (Abdi & Williams, 2010). The number of principal components is determined via the components' eigenvalues, which are equal to the sum of squared factor scores. The Kaiser Criterion (Kaiser, 1961) suggests a cut-off eigenvalue of one and is chosen due to its prevalence and sufficient reduction of dimensionality.

Furthermore, Bayesian Model Averaging is used as an additional way of assessing the model's structure. In short, *BMA* deals with the question of which independent variables to include in a model. Conceptionally this would be done by obtaining the results of every feasible model and averaging them (Koop 2003, 266). In practice it swiftly becomes an impossible task, as the number of possible combinations may very well be enormous. However, since the models considered here do not exceed nine variables, we are able to enumerate the entire model space using the R package *BMS* by Zeugner, Feldkircher, and others (2015).

Results

Variable Structure

For the optimal choice of dataset to be handled with PCA several starting points for the commodity timeseries were considered. Ultimately the second quarter of 1979 was selected, as it yielded the strongest test-statistics with regards to the possibility of using PCA, while also being covered by a sizeable amount of the available timeseries. A KMO-Criterion of 0.803 and a highly significant p-value of Bartlett's Test of Sphericity (see Appendix, Table 99.9) indicate the suitability of PCA. The Kaiser Criterion yields two components that add a satisfactory amount of information to the model (see Appendix, Figure 96.69). The components have eigenvalues of 7.79 and 2.68, which are both well above the cut-off. Additionally, the components capture 60 and 21 percent of the sample variance - which is considered sensible (Jolliffe, 2011). Based on the resulting factor loadings one can distinguish between the variables in each component. Whilst the first component comprises timeseries of precious metals, agriculture & livestock and a generic commodity index, the second one contanins industrial metals. Thus we will refer to them as *Commodities* and *Industrial Metals* (labelled *comm* and *industr*).

The iteratively computed posterior inclusion probabilities (PIPs) provide an interesting overview of the data (Appendix, Plots 14.1-14.5), but should be interpreted with care. In the case of Australia the commodities' principal components display very low PIPs, a picture that holds for all variables. The Chilean data however, appears to be rather interconnected. The commodity variables show decent inclusion in models explaining the monetary policy rate, the monetary aggregate, exports and even GDP. The Norwegian PIPs are generally rather low, although the commodity components turn out relatively high, with high inclusion when explaining GDP and decent values for equity prices. For South Africa the PIPs turn out slightly higher - the principal components show up when explaining the monetary policy rate, inflation and GDP. Based on these results we would expect higher impacts of the commodity components for Chile and South Africa, while one could presume the lower interplay of variables for Australia and Norway to lead to less convoluted models.

Impulse Responses

The impulse responses, as pictured in Plots 3.1 and 3.2 (as well as Appendix, Plots 17.1 - 17.6) show us the reaction of the model's variables in the rows to specific shocks in each column. The solid black line corresponds to the mean response; the dark- and light-gray areas cover 25-75 percent and 16-84 percent posterior confidence intervals among the 10,000 iterations.
In very loose terms the results do not hold any huge surprises. Technology shocks roughly turn out as expected, and do so rather significantly. In the cases of inflation and monetary aggregates the responses are pronounced, but dissipate very quickly. The picture is a similar one for monetary policy shocks, although the progression is harder to pinpoint due to broader confidence intervals. The impact of a shock to the first principal component (labelled *comm*) impacted GDP negatively in all countries, but Chile. The impacts on export were negative for all resource-dependent economies, but positive on the control-group of Germany and the US. The monetary policy rate reacted in an expansionary matter for Australia and the US, but insignificantly or contractionarily for the other countries. In German and South African models we can also observe a short, but significant jump in equity prices as a reaction to a shock on this first principal component. The second principal component (labelled *industr*) did not affect GDP as consistently; although, the initial effect was positive for all models, but Chile's. In the control group the positive GDP response even increased for a while and only faded after a considerable amount of time. In the resource-dependent countries the responses peaked rather early and subsequently tended to insignificance. Meanwhile, the responses of inflation, bond-yields, monetary aggregates and the monetary policy rates differed between countries, but turned out extremely similar in each country itself. Exports jumped significantly for about eight quarters for Australia and South Africa. Meanwhile they fell in the models of Germany and the US, only to rebound to a small but significant increase after about four to six quarters. The shock also had a significant, positive effect on South African equity prices.
Another set of models with country-specific commodities instead of principal components was estimated for comparison. These include industrial metals, agriculture and livestock, gold, copper, energy and precious metals (see appendix, Table 11.1). The resulting impulse responses were largely comparable, but the impacts and responses of commodity variables turned out rather insignificant. Ultimately this appraoch yielded less convincing results than one that simply used a single commodity price index. We interpret this as evidence for our approach of using principal components, as it combines the advantage of both approaches - the use of all available data, including rather country-specific timeseries, and the impactfulness of using few, but significant variables.

Discussion

Discuss

Literature

Koop, G. 2003. *Bayesian Econometrics*. Wiley. https://books.google.at/books?id=WRK3AAAAIAAJ.

Zeugner, Stefan, Martin Feldkircher, and others. 2015. "Bayesian Model Averaging Employing Fixed and Flexible Priors: The Bms Package for R." *Journal of Statistical Software* 68 (4). Foundation for Open Access Statistics: 1–37.