# High-Dimensional Macroeconomic Forecasting:

## A Partial-Correlation Based Panel Vector Autoregressive Model Estimation Method


**Rongxuan Zhang**

**442941rz**


Bachelor Thesis
Econometrics and Operations Research


Erasmus School of Economics
Erasmus University Rotterdam

supervised by: **Dr. Maria Grith**
co-read by: **Dr. Andrea Naghi**


July 6, 2019

**ABSTRACT**

The Vector Autoregressive Model(VAR) has been successful in forecasting macroeconomic time series. With the growing inter-dependencies between economies, econometricians have developed the Panel-VAR model to account for the international linkages between macroeconomic variables. However, the number of explanatory variables of a Panel-VAR model often greatly exceeds the number of observations, which inhibits the use of ordinary estimation methods. Utilized the persistent-pattern robust inference construction method introduced by [Müller and Watson(2018)], this paper developed an original partial-correlation based estimation method for Panel-VAR model. Together with other existing methods, the method is used to forecast European economic variables. Results of out-of-sample forecasting show that the estimation method provides a parsimonious yet competitive estimation method for PVAR models.

# Contents

# I. Introduction

With the globalization of investment and supply chain, countries and livelihoods in different corners of the world have been more closely dependent on each other than ever. To measure economic activity in an age of globalization, it is crucial to account for the economic linkages between different economies. Economists recognize such changes and have begun to take global macroeconomic conditions into consideration when forecasting macroeconomic conditions. The first step of such consideration is to identify the existence of correlations between macroeconomic variables in different countries. Many economic theories have already shed lights on some correlations between macroeconomic variables. Such implications are often drawn from economists' intuitions and logical deductions. The famous "Golden Ratio" from the Solow Model, which illustrates the correlation between production and consumption, is one of the examples of such kind. Yet many other relations between macroeconomic variables are remained ambiguous. With the void of theories and the complications of the real-world economies, it has been difficult to identify relations between macroeconomic variables across different countries. It is therefore more favorable for researchers to seek statistical evidence between macroeconomic variables. However, there are two main issues that hinder researchers to identify such relations[Müller and Watson(2018)]. One the one hand, macroeconomic variables are typically released in a monthly, quarterly or annually basis. Within a normal period of investigation, such as a ten-year period, the size of available macroeconomic data is often relatively small. On the other hand, statistical conclusions regarding relations between macroeconomic variables depend crucially on the unclear persistent-patterns of macroeconomic time series. When trying to forecast economic variables of one country, these difficulties sometimes hinder researchers to determine a concise yet informative set of explanatory variables. Literature has come up with two main solutions to such issues. The first approach, Dynamic Stochastic General Equilibrium(DSGE), makes use of economic theories and enables researchers to address important policy questions when properly impose parameter restrictions[Canova and Ciccarelli(2013)]. The main issue for DSGE is the reliability of the imposed restrictions, which are determined by researchers' intuition and might not be in line with empirical findings[Canova and Ciccarelli(2013)].

The other popular approach, Panel Vector Autoregressive model(PVAR), uses the well-known Vector Autoregressive model(VAR) to account for the heterogeneity and interdependence between macroeconomic variables of different countries. Intuitively, the Panel Vector Autoregressive model is no more than a normal VAR model with an additional cross-section dimension[Canova and Ciccarelli(2013)]. In theory, the model could successfully capture the international spillover effects. However, there could be an explosive number of parameters needed to be estimated due to the large number of countries, macroeconomic variables and the number of lags that might be incorporated. The size of the long-run macroeconomic observation, one the other hand, is normally relatively small within a given time period. For example, when forecasting European countries' monthly GDP growth rate, if a model uses data from 50 countries with 10 economic explanatory variables and 6 lags for the forecasting, the total number of parameters in one single regression equation will be as large as 3000. At the same time, the sample size for a single equation is as small as 120 in a ten-year span. Clearly, the estimation of such models would be impossible without a proper mechanism to reduce the number of dimensions. Besides from selecting variables with economic intuitions, most common variables reduction approaches make use of penalty-based techniques, such as the Lasso penalizing estimation, Bayesian class methods and many others[Uematsu and Tanaka(2019)]. The main contribution of this paper is to develop an alternative variables reduction mechanism for Panel Vector Autoregressive model. Results show that the newly invented mechanism is particularly suitable for high-dimensional models and performs spectacularly well in out-of-sample forecasting in comparison with almost all of the existing methods.

To develop a powerful mechanism for variables reduction in macroeconomic setting, it is crucial to identify variables that are causally related with the interested variable. However, the causal relation between different macroeconomic variables remains difficult to identify. One country economic variable could affect another country either directly or indirectly, both of which would be hard to verify. For instance, the Inflation rate in Iceland might not have directly influence to the GDP of South Korea. Yet, these variables could be highly correlated via some unexpected channels. For example, it is possible that the popularity of an American series in South Korea drastically increases the number of South Korean visitors in Iceland, which boosts Iceland's GDP in a certain quarter. The difficulty of identifying causal relation leads us to an easier approach, identifying correlation. As shown in the methodology, correlations can be a good indicator of the existence of causal relations under certain conditions. For the purpose of models shrinkage, it is desirable to have statistical tools that test correlations between long-run macroeconomic variables. The aforementioned issues told us that testing correlations of Long-Run macroeconomic data suffer from two main issues, uncertain persistence pattern, and the paucity of data. The former is especially problematic if the assumed persistence pattern of the method is wrong. For example, a statistical inference would be unreliable if the data is assumed to be stationary($I(0)$) when it is in fact an $I(1)$ process. Yet the second issue, the limited amount of data, make it difficult for researchers to determine the persistent pattern of a macroeconomic time series. Such characteristics of low-frequency data add difficulties to statistical test of macroeconomic variables' correlations.

As such, a powerful and robust inference method for macroeconomic variables correlation is crucial for a partial correlation based variable shrinkage method. In 2018, Müller and Watson have brought out an inference method that is robust to issues mentioned[Müller and Watson(2018)]. The newly introduced inference method lay the foundation for my partial correlation based variables selection method and is essential to this paper. Therefore, I will introduce the intuition and specification of the method in the first part of this paper. I will then introduce some innovative insight of the partial correlations causal relation brought forth by Bühlmann[Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis], which provide a component for my method. The final two sections of the paper will present results of real-world applications of the method.

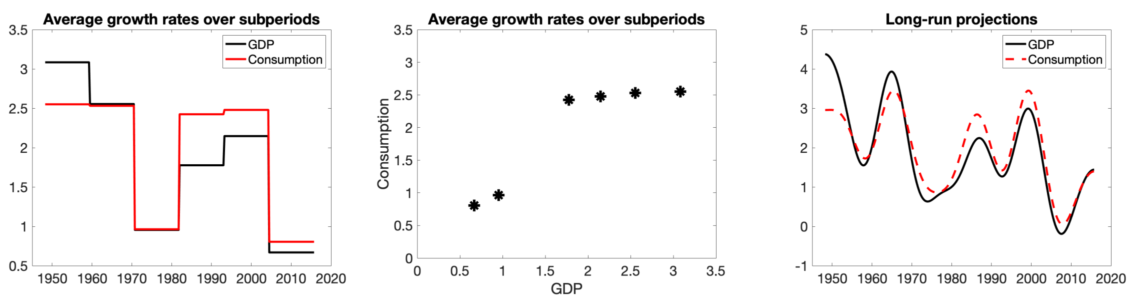
# II. Long-Run Covariability

Many of the most successful macroeconomic theories have predicted that long-run economic indicators such as GDP, consumption, money supply, inflation, are highly correlated with each other. Such findings are in line with real-world phenomena and are often used for economic forecasting. This section aims to introduce a newly purposed robust inference construction method[Müller and Watson(2018)] to formally test these relations. I will use the paired relations between quarterly GDP and consumption, short-term interest-rate and long-term interest rate, to exemplify this inference construction method. Codes and data, provided by Ulrich K. Müller[1], are used to replicate Figure 1, 2, Table I, II, in [Müller and Watson(2018)]. Figure 1, 2 in [Müller and Watson(2018)] correspond to Figure I, II, III below, and Table I, II in [Müller and Watson(2018)] are named as the same in this paper. The codes and data are only used in this section except for one Matlab® function, which is also used to construct the partial-faithfulness PVAR model shown in the following sections[2].

## *A. Cosine Transformation and Spectral Analysis*

According to arguably the most famous finding of macroeconomics, consumption are believed to be proportional to GDP. This relation has been backed by data of different countries on different time spans. Panels 1 and 2 of figure 1 shows the relation between the long-term average of U.S. GDP growth rate and U.S consumption growth rate. I divided the period of 1948-2015 into six sub-periods and calculated the average over individual sub-period. Panel 1 plots the average against the sample period and it shows a clear correlation between the long-term average of GDP and consumption – the upward and downward co-movements are apparent for the examined periods. Panel 2 plots the six growth rates of GDP and consumption against each other. The panel revealed a seemingly proportional relation between GDP and consumption.

***Figure 1***




To investigate the correlations of multiple small sample time series, an immediate thought is to calculate the correlation matrix. However, a simple covariance matrix might fail to account for a stochastic trend, cointegration and other interdependent characteristics with the existence of unknown persistent patents. On the other hand, the stochastic nature of time series would

[1] source: https://www.princeton.edu/ umueller/
[2] Specifically, the lrcov.m function provided by Ulrich K. Müller

drastically increase the variation of the estimated correlation, especially with small sample settings. We therefore need to resort to other methods for the purpose of identifying correlation. A time series, in essence, can be expressed either in time dimension or frequency dimension. This thought leads us to the idea of Cosine Transformation, a method that transforms the original time-dimensional function to the frequency dimension without losing any information.

This idea gives us a completely different, yet compelling angle of viewing time series. Intuitively, the Cosine Transformation of a time series implies that the fluctuation of macroeconomic conditions could be fully explained by "economic cycles" of different lengths. Denote $x_t$ as the value of the underlying time series at time $t$, $x_{1:T}$ as $[x_1, x_2, ...x_T]$‘, the vector containing information of the entire period. Further denote $T$ as the total number of observations. A full Cosine Transformation outputs a series of the same length as the original one, with each element representing the projection of the original series on a cosine basis in different frequencies. The following equation corresponds to such transformation basis with frequency $2T/n$.

$$[\cos(\frac{n\pi}{T}(1-1/2)), \cos(\frac{n\pi}{T}(2-1/2)), ...., \cos(\frac{n\pi}{T}(T-1/2))]$$

The original time series will be decomposed into "economic cycles" with different periods, ranging from the maximum length of the time series, T, to the minimum length, 2. Denote $X_n$ as the $n^{th} \in [1, 2, ..., T]$ transformed data, $X_n$ serves as the information indicator of the series with economic cycles length $2T/n$. Since we are interested in the long-run correlation between variables, short-run fluctuations would be redundant. Therefore, we could solely investigate the transformed data with longer-term economic cycles. In the U.S., for example, the minimum length of a long-run economic cycle is assumed to be 11 years. If this assumption holds, Economic cycles with a length smaller than 11 years would contain little long-run information about the time series. Defines $q$ as the highest value of the cosine functions' multiple. As our data contains 271 observations, we define $q = 12$ to filter out the short-run variations. In a general setting of the Cosine Transformation filtering, the minimum periods of the cosine function is $T/2q$. The Cosine Transformed data correspond to $X_n$, $n \in [1, 2, ..., q]$ below.

$$x_t = X_1\sqrt{2}\cos(\frac{1\pi}{T}(t-1/2)) + X_2\sqrt{2}\cos(\frac{2\pi}{T}(t-1/2)) + .... + X_q\sqrt{2}\cos(\frac{q\pi}{T}(t-1/2)) + \epsilon, \quad (1)$$

With $\epsilon$ defined as the error term that corresponds to the short-run variations of a time series. For empirical macroeconomic data, we could obtain the Cosine Transformed data by using the above equation as a linear regression model and its corresponding estimated coefficients as the transformed data. For notation ease, let's denote the matrix representation of explanatory variables as a $T \times q$ matrix $\Psi_T$, as defined in [Müller and Watson(2018)], with

$$\Psi((t-1/2)/T) = [\sqrt{2}\cos(1(t-1/2)/T), \sqrt{2}\cos(2(t-1/2)/T), ..., \sqrt{2}\cos(q(t-1/2)/T)]'$$

as the $t^{th}$ row among $T$ total rows. Since equation (2) and (3) shows that the coefficient of the regression corresponds to the transformed time series, I equate the regression coefficients with the transformed data, $X_T = [X_1, X_2, ...X_q]'$. From the property of a cosine function, we can deduce that $\Psi_T' I_q = 0$, $\Psi_T' \Psi_T = I_q$, with $I_q$ being the $q \times q$ identity matrix. In the following paragraphs, I will deduct some useful properties between the filtered data and the transformed coefficients, as being shown on [Müller and Watson(2018)]. Using the above properties, the filtered data can be represented as,

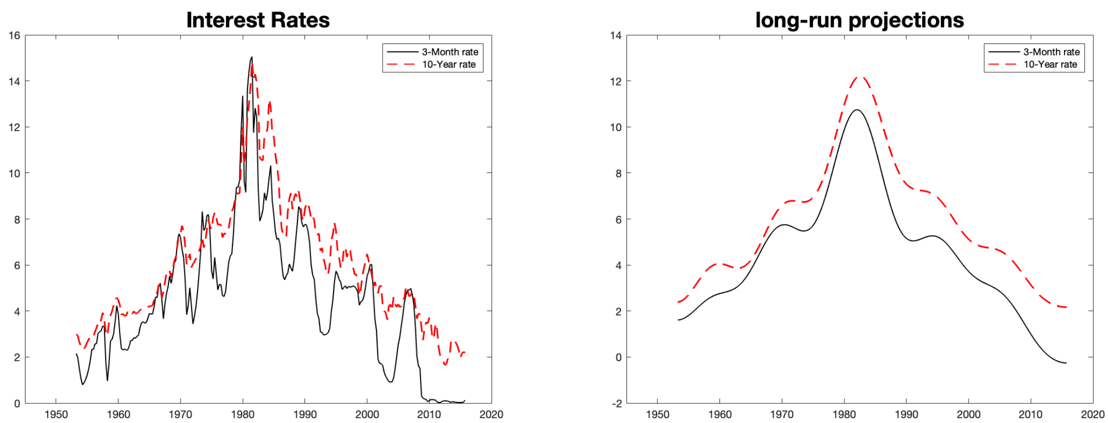
$$\hat{x}_t = X_T' \Psi_T \tag{2}$$

Since the regression is essentially an Ordinary Least Square regression, we have

$$X_T = T^{-1} \Psi_T' x_{1:T} \tag{3}$$

Panel 3 of Figure 1 above plots out the continuous predicted values of GDP and consumption at each time slot. It shows that by using the predicted values, we effectively filtered out the short-run variation of the time series and leaves only the long-run characteristics. The co-movement of the "transformed" GDP and consumption shown on this panel is even more apparent than the simple average shown on Panel 1 of figure 1, which is the common approach to obtain long-run information. These evidence shows that the Cosine Transformation filtered out the short-run information of GDP and consumption data. Example of short-term and long-term interest rate also ratify the use of such filtering method. It is commonly known that the short-term interest rate and the long-term interest rate of U.S. Treasury Bill are closely related to each other. Panel 1 of figure 2 shows the raw data of 3-month and 10-year interest rate on the period of 1953-2015, revealing a seemingly co-move but noisy pattern, whereas Panel 2 shows the filtered series with smoother transitions. The Cosine Transformation does effectively eliminate the short-run variations of these two series.

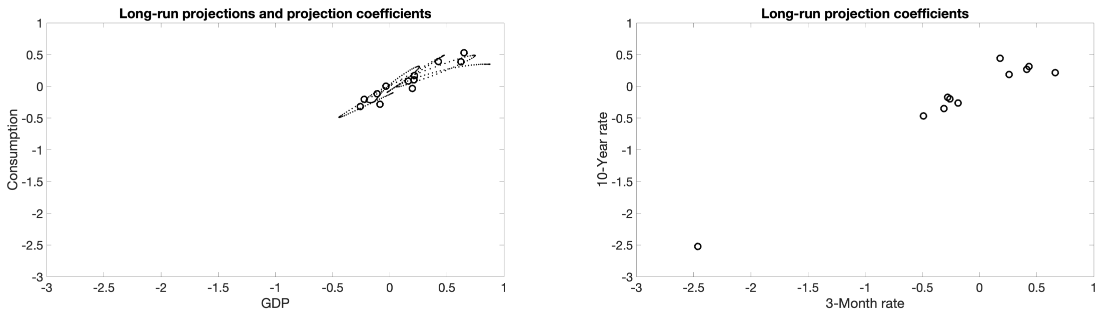
***Figure 2.*** *Raw Interest Rate Data and Filtered Data*




There is another appealing character of using the Cosine Transformation. Since the transformed data contains long-run information of the time series, it also captures the long-run characteristics of a time series with a relatively small amount of data. The first panel of Figure 3 below plots the coefficients of GDP's transformation(or the transformed data) against consumption's counterpart in larger circles. This panel also plots the filtered quarterly values of GDP against filtered values of consumption in small dots. As shown on the panel, there is a strong correlated between the transformed data and the filtered data in the example of GDP and Consumption – both reveal approximately the same proportional relationship between GDP and Consumption. Second panel of Figure 3 plots the coefficient of short term interest rate against its counterpart of the long-term interest rate. This plot also indicate an proportional relation between the two variables. It shows that the transformed data contain seemingly the same long-run information as the filtered data. In fact, the covariance between predicted values of two time series are directed linked to that of the transformed data. For two transformed time series with same length $\hat{x}_t$ and $\hat{y}_t$, we have[3]

$$\begin{aligned} Cov\begin{pmatrix}\hat{x}_t \\ \hat{y}_t\end{pmatrix} = T^{-1}\sum_{t=1}^{T}\begin{pmatrix}\hat{x}_t \\ \hat{y}_t\end{pmatrix}\begin{pmatrix}\hat{x}_t & \hat{y}_t\end{pmatrix} &= T^{-1}\begin{pmatrix}X_t' \\ Y_t'\end{pmatrix}\Psi_T'\Psi_T\begin{pmatrix}X_t & Y_t\end{pmatrix} \\ &= T^{-1}\begin{pmatrix}X_t' \\ Y_t'\end{pmatrix}\begin{pmatrix}X_t & Y_t\end{pmatrix} \\ &= T^{-1}\begin{pmatrix}X_t'X_t & X_t'Y_t \\ Y_t'X_t & Y_t'Y_t\end{pmatrix} \end{aligned} \tag{4}$$

***Figure 3***




Put it differently, the transformed data summarizes the long-run information of the time series. If the assumption that the transformation indeed captures the long-run information, as being proven by the graphs shown, we could investigate time series' relations by solely investigating the properties of the parsimonious transformed data. In fact, if we denote the covariance matrix of $\begin{pmatrix}\hat{x}_t & \hat{y}_t\end{pmatrix}$ as $\Omega_T$, covariance matrix of $\begin{pmatrix}X_t & Y_t\end{pmatrix}$ as $\Sigma_T$ and $\Sigma_{XX,t}$, $tr(\Sigma_{XY,t})$ as the partition of $\Sigma_T$ matrix, the following equation holds[4],

[3][Müller and Watson(2018)]
[4][Müller and Watson(2018)]

$$\Omega_T = T^{-1}\sum_{t=1}^{T}\left[\begin{pmatrix}\hat{x}_t\\ \hat{y}_t\end{pmatrix}\begin{pmatrix}\hat{x}_t & \hat{y}_t\end{pmatrix}\right] = \sum_{j=1}^{q}\left[\begin{pmatrix}X_{jT}\\ Y_{jT}\end{pmatrix}\begin{pmatrix}X_{jT} & Y_{jT}\end{pmatrix}\right] = \begin{pmatrix}tr(\Sigma_{XX,t}) & tr(\Sigma_{XY,t})\\ tr(\Sigma_{YX,t}) & tr(\Sigma_{YY,t})\end{pmatrix}$$

Let $x$ and $y$ be two time series of interest and denote $\beta$ as the coefficient of regression $y = \beta x + \epsilon$. To fully explore the information of the covariability between $x$ and $y$, I denote $\rho_T$ as the $R^2$ of the regression, $\sigma^2_{y|x,T}$ as the average prediction error of the regression, and $\beta_T$ as the estimated regression coefficient. The values of these variables are calculated as follows,

$$\begin{aligned}
\beta_T &= T^{-1}\frac{\sum \hat{x}_t\hat{y}_t}{\sum \hat{x}_t^{\,2}} = \frac{\Omega_{xy,t}}{\Omega_{xx,t}}\\
\rho_T &= T^{-1}\frac{\sum \hat{x}_t\hat{y}_t}{\sqrt{\sum \hat{x}_t^{\,2}\hat{y}_t^{\,2}}} = \frac{\Omega_{xy,t}}{\sqrt{\Omega_{xx,t}\Omega_{yy,t}}}\\
\sigma^2_{y|x,T} &= T^{-1}\sum \hat{y}_t^{\,2} - T^{-1}\frac{\sum(\hat{x}_t\hat{y}_t)^2}{\sum \hat{x}_t^{\,2}} = \Omega_{yy,t} - \frac{\Omega^2_{xy,t}}{\Omega_{xx,t}}
\end{aligned}$$

If the persistent patterns of two time series are known and the aim is to test the two series' covariability, it is sufficient to have efficient inference construction methods for the aforementioned three variables. [Müller and Watson(2015)] has developed such inference methods for $I(0)$ and $I(1)$ series, of which the capacity of testing covariability was proven sufficient. Table 1 shows the resulting confidence intervals of $\beta_T$ using both $I(0)$ and $I(1)$ as the assumed persistent patterns. However, if there are misspecification issues, the inferences might lead to incorrect and undesirable conclusions. Their sequential paper [Müller and Watson(2018)] developed a robust inference model to resolve such issue. I will illustrate the intuition of the inference model in detail.

| | | $\rho$ | $\beta$ | $\sigma$ |
|---|---|---|---|---|
| **GDP and consumption** | | | | |
| $I(0)$ | Estimated Value | 0.93 | 0.76 | 0.36 |
| | 67%CI | (0.87, 0.96) | (0.67, 0.85) | (0.30, 0.46) |
| | 90%CI | (0.80, 0.97) | (0.60, 0.92) | (0.27, 0.55) |
| $I(1)$ | Estimated Value | 0.93 | 0.84 | 0.35 |
| | 67%CI | (0.88, 0.96) | (0.74, 0.94) | (0.29, 0.45) |
| | 90%CI | (0.82, 0.97) | (0.66, 1.01) | (0.26, 0.54) |
| Highest period | 135.5 | Lowest period | 11.3 | |
| **Long-term and short-term interest rate** | | | | |
| $I(0)$ | Estimated Value | 0.98 | 0.96 | 0.60 |
| | 67%CI | (0.96, 0.98) | (0.90, 1.03) | (0.50, 0.79) |
| | 90%CI | (0.93, 0.99) | (0.84, 1.08) | (0.44, 0.96) |
| $I(1)$ | Estimated Value | 0.97 | 0.93 | 0.38 |
| | 67%CI | (0.93, 0.98) | (0.85, 1.01) | (0.32, 0.50) |
| | 90%CI | (0.90, 0.98) | (0.78, 1.07) | (0.28, 0.61) |
| Highest period | 125.5 | Lowest period | 11.4 | |

***Table I Estimates and Confidence Interval Under*** $I(0)$ ***and*** $I(1)$

### B. *Inference with Spectral Density*

The aforementioned filtering method essentially transforms data from time dimension to frequency dimension using the following equation,

$$x_\omega = \sum_{t=1}^{T} x_t \cos\left(\frac{\omega\pi}{N}(n + \frac{1}{2})\right), \tag{5}$$

with $\omega$ as the frequency variable of the time series. Since this paper aims to investigate the long-run behavior of time series, we will only use data that contains long-run information. That is, transformed data with longer length of economic cycles, or equivalently, have $\omega$ close to zero. Similar to an ordinary density function, a spectral density $S_x(\omega)$ is a function of frequency $\omega$.

The one-to-one relation in equation (5) implies an one-to-one relation between an ordinary density function and a spectral density function, which makes them effectively equivalent. The following analysis will deduct useful properties from frequency dimension to construct inferences. From frequency's point of view, the distribution of the transformation's data is a spectrum density function. [Müller and Watson(2018)] shows that asymptotically, the covariance matrix of the transformation coefficients, or transformed data, depends solely on the spectral density of a time series. Therefore, as long as one could successfully determine the spectral density of the transformed data, inferences of different parameters in the original time series will follow directly. In our context, define $z_t = \begin{pmatrix} x_t & y_t \end{pmatrix}'$, $\Delta z_t$ as its first differentiation, $\omega$ as spectrum's frequency and $S_z(\omega)$ as its asymptotic spectral density function. Further define asymptotic covariance matrix of $Z_t$ as $\Omega$, with

$$\sqrt{T}\text{Var}\begin{pmatrix} X_T \\ Y_T \end{pmatrix} T\Omega_T \to \Omega \tag{6}$$

If some conditions of a spectral density [Müller and Watson(2018)] holds. There is an asymptotic one-to-one relation between spectral density and $\Omega$. Muller and Watson have also validated the use of asymptotic properties of the spectral density function when the size of the time series is as large as 250[Müller and Watson(2018)]. Therefore, the asymptotic relation could be used to construct inferences for parameters of interest. The asymptotic one-to-one transformation between $\Omega$ and spectral density is as follows[Müller and Watson(2015)][5],

$$\Omega = \int_{-\infty}^{\infty} (I_2 \otimes \int_0^1 e^{i\omega s}\Psi(s)ds)' S_z(\omega)(I_2 \otimes \int_0^1 e^{-i\omega s}\Psi(s)ds)d\omega \tag{7}$$

To overcome the misspecification issue, Müller and Watson developed a more general specification of the persistence patterns. The generalized specification is based on the distribution of time series' joint spectral density. As illustrated before, there is an one-to-one relationship between a spectral density and time series. [Müller and Watson(2018)] introduced a inference construction method adaptable to more generalized specification of persistence pattern. The authors used a more generalized asymptotic spectral density to achieve this generalization. This newly created spectral density specification, named as $(A, B, c, d)$ model, extends the univariate $(b, c, d)$ model[Müller and Watson(2016)] and has the following form,

$$S_z \propto A \begin{pmatrix} (\omega^2 + c_1^2)^{-d_1} & 0 \\ 0 & (\omega^2 + c_2^2)^{-d_2} \end{pmatrix} A' + BB', \tag{8}$$

where $A$ is an unrestricted $2 \times 2$ matrix and $B$ is an lower triangular $2 \times 2$ matrix. The $(A, B, c, d)$ model provides a parsimonious yet general specification for time series' spectral density. Therefore, it overcomes the misspecification issue that might happen frequently in long-run macroeconomic time series analysis. In this specification, asymptotic spectral density of $I(0)$ time series corresponds to a $(A, B, c, d)$ model with $A = 0$ while $I(1)$ time series is represented by the model with $d_1 = d_2 = 1$, $B = 0$ and $C = 0$. Other specification of parameters in the $(A, B, c, d)$ model covers a wide range of time series specifications, including mean-reverting, cointegrated models and models with combination of different persistence elements. Detailed range that the $(A, B, c, d)$ model represents can be found on section 3 of [Müller and Watson(2018)].

To construct confidence intervals for the the interested parameters, I minimized the length of confidence intervals for any given confidence level. It has been shown that the estimated parameters of interest, denoted as $\gamma = (\beta,\ \rho,\ \sigma_{y|x})$ , is a function of the estimated spectral density. If denotes estimated parameters in $(A, B, c, d)$ as $\theta$ over a parameter space $\Theta$, $(X, Y)$ as the two input data, $H(X, Y)$ as the confidence interval of $\gamma$, we are minimizing $E[lgth(H(X, Y))]$

[5][Müller and Watson(2018)] presumably have mistakenly written $\omega$ as $\lambda$ in the paper. Such notation also differs from their previous work[Müller and Watson(2015)]

over the parameter space with $lgth(H(X,Y))$ the length of the confidence interval. A crucial question in this minimization problem is which parameter $\theta$ across the parameter space should we emphasize more heavily on when minimizing the confidence intervals. If denote $W$ as the weighting distribution for different value of $\theta$, the problem of interest can be written as,

Given,

$$\underset{\theta \in \Theta}{Inf} P(\gamma \in H(X,Y)) \geq 1 - \alpha \tag{9}$$

Minimizing the following equation[6],

$$\underset{H}{min} \int E[lgth(H(X,Y))]dW(\theta) \tag{10}$$

Using the above specification and $W$ functions specified in appendix of [Müller and Watson(2018)], table 2 shows the estimations of interested parameters and confidence intervals of different levels. Since the potential confidence interval, under some realization of (X,Y), could be empty, I used the method specified in [Müller and Watson(2016)] to construct Bayes Credible Sets of different confidence levels. Results are shown below.

| | | $\rho$ | $\beta$ | $\sigma$ |
|---|---|---|---|---|
| **GDP and consumption** | | | | |
| $(A, B, c, d)$ | Estimated Value | 0.91 | 0.77 | 0.41 |
| | 67%CI | (0.84, 0.96) | (0.66, 0.87) | (0.33, 0.53) |
| | 90%CI | (0.71, 0.97) | (0.48, 0.96) | (0.27, 0.55) |
| | 67%Bayes CS | (0.84, 0.96) | (0.66, 0.87) | (0.33, 0.53) |
| | 90%Bayes CS | (0.71, 0.97) | (0.58, 0.96) | (0.29, 0.66) |
| **Long-term and short-term interest rate** | | | | |
| $(A, B, c, d)$ | Estimated Value | 0.96 | 0.95 | 0.63 |
| | 67%CI | (0.92, 0.98) | (0.87, 1.07) | (0.49, 0.97) |
| | 90%CI | (0.89, 0.99) | (0.76, 1.16) | (0.42, 1.27) |
| | 67%Bayes CS | (0.92, 0.98) | (0.87, 1.03) | (0.49, 0.82) |
| | 90%Bayes CS | (0.89, 0.99) | (0.81, 1.09) | (0.42, 1.02) |

***Table II Estimates and Confidence Interval Using*** $(A, B, c, d)$ ***model***

# III. Partial Faithfulness with Long Run Macroeconomic Data

The second part of the methodology introduces an original partial-correlation based PVAR estimation method tailored specifically to long-run low-frequency macroeconomic data. This partial – correlation approach makes use of the concept partial faithfulness introduced by Bühlmann[Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis] as well as Müller and Watson's robust inference construction method. This methodology section will begin with introducing the motivation of using a partial-correlated based method for variable selection. It continues with explanation of some key concepts, few conditions to establish the algorithm . The section

[6] [Müller and Watson(2018)]

will conclude with the concrete algorithm of the method.

### *A. Correlation, Partial Correlation and Casual Relation*

The ultimate goal of a variable selection method in a linear model such as the PVAR, is to identify the causal relation between a variable and the dependent variable. In an idealized linear model under the condition that the potential explanatory variables are not inter-correlated, a causal relation between one explanatory variable and the dependent variable necessarily implies a statistically significant correlation. To choose variables that are might be causally affecting the dependent variable, it is legitimate to only retain explanatory variables that are significantly correlated with the dependent variable. However, such condition holds rarely in most of the real-world cases which most of the explanatory variables are correlated. In fact, one of the main motivation of using the VAR model is to account for such inter-dependencies between variables. With the failure of the uncorrelated premise, simply correlation could no longer be used to shrink the potential explanatory variable set. Yet a stronger concept, partial correlation would be a useful tool for variable selection. Consider a simple linear data generating process(DGP),

$$Y = \beta_1 x_1 + \beta_2 x_2 + ...\beta_n x_n + \epsilon_Y$$

With some $\text{Cov}(x_i, x_j) \neq 0$ for some $i, j \in [1, 2, ..., n]$. The covariance, or equivalently the correlation, between $Y$ and explanatory variable $x_k, k \in [1, 2, ..., n]$, is not necessarily zero even if $\beta_k = 0$ due to the existence of covariances between explanatory variables.

Assume that a set of variables $X = [x_i |\ i \in [1, 2, ..., n], i \neq k]$ simultaneously affects both $Y$ and $x_k$. Denote $\text{Corr}(Y, x_k|X)$ as the partial correlation of $Y$ and $x_k$, the partial correlation accounts for the impact of $X$ on both $Y$ and $x_k$ and leaves out the "pure" correlation between $Y$ and $x_k$. For the ease of expression, I will use common variables, denoted as $X$, to refer to the set of variables that simultaneously affects the investigating variables. Under a linear setting, it could be computed as the correlation between $e_Y$ and $e_{x_k}$ defined in equation (11). Assume that $x_k$ is partially linearly explained by $x_k = \sum_{i \neq k} w_i x_i + a + \epsilon_k$, where $K$ is the idiosyncratic and unexplained part of $x_k$, we have

$$\begin{aligned} e_Y &= Y - \sum_{i \neq k} \beta_i x_i = \beta_k x_k + \epsilon_Y \\ e_{x_k} &= x_k - \sum_{i \neq k} w_i x_i = K + \epsilon_k \end{aligned} \tag{11}$$

Since the idiosyncratic component of $x_k$ is by definition correlated with $K$, a nonzero $\beta_k$ is equivalent to $\text{Corr}(e_Y, e_{x_k}) \neq 0$. In other words, zero coefficient $\beta_k$ in a linear setting implies a zero partial correlation. Despite the fact that the direction of this argument could not be inverse, the above argument points an direction for variable selection in a linear model. So long as the common variable, such as $X$ in our example, could be successful identified, variables that are significant partial correlated with the dependent variable will be a superset of the set of the causally related variables. Finally, the above deduction also provides a formula to compute partial correlation in a linear setting and will be later used in the algorithm.

### *B. Partial Faithfulness: Correlation and Variable Selection*

It has been shown that partial correlation is a potentially useful approach to shrink the size of a high-dimensional linear models. Unlike penalty–based approaches such as Lasso, this approach aims to find the superset of the exact explanatory variables. However, it is difficult to identify a set of variables that are simultaneously affecting some other variables. Common approaches to determine partial correlations is to recursively incorporate potential common variables. Here, I will use the reverse process of such approach to shrink the potential explanatory variable set. The proposing algorithm bases mainly on the findings of [Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis] and makes use of the $(A, B, c, d)$ inference construction method so that it is robust to persistence pattern misspecification issue. Before formally introducing the algorithm, I will first introduce the concept and related theorems of partial faithfulness purposed by Bühlmann.

Consider a simple linear model, with $X = [x_1, x_2, ..., x_p]$ as a $[P \times 1]$ vector contents the potential explanatory variables of $Y$, $\mu_x$ as the mean vector and $\Sigma_x$ as the covariance matrix of $X$,

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + ...\beta_p x_p + \epsilon_Y \tag{12}$$

The goal of a variable selection algorithm is to identify the $\beta$s which have true value equal to 0. For notation ease, I refer a set of indices of that represent non-zero $\beta$s as the active set $A$: $A = [i = 1, 2, ..., p|\, \beta_i \neq 0]$ and the cardinal of $|A|$ as $n$. I further denote $S$ as a subset of the indices, $S \subseteq [1, 2, ..., p]$, $x^S$ as $[x^i|\, i \in S]$, $S^C$ as the complement set of $S$, $p(a, b|\, W)$ as the partial correlation of $a$ and $b$ given a common variables set is $W$. With these notations, the concept partial faithfulness is proposed by Buhlmann with following definition:

*Theorem 1: If $X \in \mathbb{R}^P$ represents a random vector and $Y \in \mathbb{R}$ a random variable, the distribution of $(X, Y)$ is defined as partial faithfulness distribution if for every $i \in [1, 2, ..., p]$,*[7]

$$p(Y, x_i|x^S) = 0 \text{ for some } S \subseteq [i]^C \implies \beta_i = 0$$

Bühlmann further specified two conditions under which it is almost certain that the distributions of nonzero coefficient generating process are partial faithfulness distributions. I will refer the proof of this theorem to the appendix of [Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis].

*Condition 1: $\Sigma_x$ is strictly positive definite.*[8]

*Condition 2: coefficients of $X$ satisfies: $[\beta_i|\, i \in A] \sim f(a)da$, where $f(\cdot)$ is a density function of an absolutely continuous distribution on a subset of $\mathbb{R}^n$ in terms of Lebesgue measure density.*[9]

[7] [Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis]
[8] [Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis]
[9] [Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis]

The first condition, which researchers could test conveniently with eigenvalues, implies that the linear model should be clearly identifiable. The implication of the second condition is slightly more complicated and requires some extra effort. However, a multivariate normal distribution is itself an absolutely continuous distribution in terms of Lebesgue measure density. As such, the conditions generally holds for PVAR model with its multivariate normality assumption.

Conventionally, we need the complete set of common variables to determine the partial correlation between an explanatory variable and the dependent variable. It is only possible to discard the variable from the explanatory variable set $X$ if the such partial correlation is insignificant. The partial-correlation is determined by using the conventional recursive partial correlation calculation method that updates partial correlation by adding one variable to the common variable set at a time. With the given theorem, the idea of the partial faithful algorithm is to develop a variable selection process that excludes an explanatory variable as soon as it discovers an insignificance partial correlation given few common variables(or even one variable). Formally, the following corollary justify this idea,

*Corollary: Given that a linear model satisfies the two conditions given above, for every $i \in [1, 2, ..., n]$, $\beta_i \neq 0$ is equivalent to $p(Y, x_i|\ x^S) \neq 0$ for every $S \subseteq [j]^C$.*[10]

This corollary enables us to conduct an efficient iterative selection algorithm which results in a variable set that at least contains the active set, as long as the computations of partial correlation inferences are powerful and the amount of iterations is sufficient. Using the corollary, we set the active set $A$ to be empty in the first iteration and add all potential explanatory variables that is significantly correlated with the dependent variable into the first active set. The resulting new active set $A^1$ would necessarily be a superset of the true active set $A$,

$$A^1 = [i = 1, 2, ..., p|\ \text{Cov}(Y, x_i) \neq 0] \supseteq A \tag{13}$$

The next iteration removes variables that are not significantly partial correlated with $Y$ given any subset of $A^1$,

$$A^2 = [i \in A^1|\ p(Y, x_i|x^S) \neq 0]\ \ \forall S \subseteq (A^1 \setminus \{i\}) \tag{14}$$

Using the same reasoning, $A \subseteq A^2 \subseteq A^1$. The process reduces the number of elements in the active set but also guarantees that the resulting active set is a superset of the true active set. This idea can be formulated into the following algorithm. Given P is the number of potential explanatory variables in a linear model, [Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis] has porpused the following algorithm,

[10] [Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis]

**Algorithm 1** Partial Faithfulness Algorithm

**Step 1** Set the first active set, set m=1:

$A^1 = [i = 1, 2, ..., p|\ \mathrm{Cov}(Y, x_i \neq 0)]$

**Step 2** Repeat m=m+1

$A^m = [i \in A^{m-1}|\ p(Y, x_i|x^S) \neq 0]\ \ \forall S \subseteq (A^{m-1} \setminus \{i\})$ with $|S| = m - 1$

Until $|A^m| \leqslant m$

**End**

[Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis] has proven in its appendix that with condition 1, the partial faithfulness algorithm is sufficient to identify the true active set $A$. In mathematical notation,

$$m_{\text{stop}} = \min\{m|\ |A^m| \leqslant m\} \implies A^{m_{\text{stop}}} = A, \tag{15}$$

### *C. Estimating PVAR with Partial Faithfulness Algorithm*

This section will introduce the specification of a PVAR model and its estimation method with the partial faithfulness algorithm. As mentioned before, Panel Vector Autoregressive Model(PVAR) is an application of a normal VAR model with an additional cross-sectional dimension. In the context of this research, PVAR models different macroeconomic variables of multiple countries in one system of equations. Theoretically, the model accounts for the heterogeneity and international spillover effects of macroeconomic variables by including all domestic and foreign macroeconomic variables in one equation. Assume that there are $Q$ country and $V$ macroeconomic variables in a PVAR application, the PVAR model consists of $V$ equations. For country $i \in [1, 2, ..., V]$, individual country $i$ has the following specification,

$$y_{it} = \beta_{i1}Y_{t-1} + \beta_{i2}Y_{t-2} + ... + \beta_{ip}Y_{t-p}, \tag{16}$$

where $y_{it}$ is a $[V \times 1]$ matrix with $V$ series of macroeconomic data for country $i$ at time $t$ and the complete dependent variable set $Y_t$ is defined as $[y_{1t}, y_{2t}, ...y_{Qt}]'$. The main problem of estimating such model is the large number of parameters that need to be estimated. That is, the number of parameters, $\beta_{ij,t}$, often becomes excessively large. To resolve this issue, we will iteratively select explanatory variables in one equation using the partial faithfulness algorithm.

As long as a model satisfies the two conditions, Bühlmann, Kalisch, and Maathuis have proven that the partial faithfulness algorithm is theoretically feasible, efficient and consistent. However, the performance of the partial faithfulness algorithm is subject to the inference construction method of partial correlations. Faulty inferences of the partial correlation are destructive to the selection algorithm. Macroeconomic data distinguishes itself from other ordinary explanatory variables, such as bacterium's productions, from several perspectives. On the one hand, macroeconomic data often has apparent time series characteristics such as seasonality and trend. On the other hands, short-run macroeconomic data are often affected by aberrant events, such as fluctuation of oil prices and stock market performance. Such events commonly affect almost all types of macroeconomic variable in a short period of time. Such events produce seemingly correlated relations between macroeconomic variables but do not imply any "real" correlations and causal relations between variables. Despite the possibility of using the first – order difference to clean the data, the paucity of sample size makes it difficult for researchers to intuitively identify all characteristics of a time series. A direct solution to the issues is to filter out the short term variation of time series. Yet mere filtering method itself does not account for the misspecification issue. Different persistent patterns, such as $I(0)$, $I(1)$, and different interactions between time series, such as cointegration, can lead to completely different outcomes if an inference construction method is not robust to misspecification. As such, using ordinary correlation testing methods, such as the Fisher's $Z$-transformation, to determine macroeconomic variables' correlation, could be dangerous.

In light of the issues mentioned, I make use of the Cosine Transformation as well as the $(A, B, c, d)$ model to test partial correlation relations. To resolve the first categories of issues, such as the seasonality, trend, I apply the aforementioned Cosine Transformation on the original data and only retain data corresponds to the long-run information. I then utilize the $(A, B, c, d)$ model to account for the unclear persistent-pattern issue. Denote $x_k$ as the parameter of interest and $x^S$ as the current common variable set. Further Denote $\hat{B_k}$ and $\hat{B_Y}$ as the estimated coefficients of $x_k$ and $Y$ on common variable set $x^S$, as shown on equation (17).

$$\begin{aligned} e_{x_k} &= x_k - \hat{B}_k x^S \\ e_Y &= Y - \hat{B}_Y x^S \end{aligned} \tag{17}$$

The significance of $p(Y, x_k|\, x^S)$ is equivalent to the significance of $\text{Cov}(e_{x_k}, e_Y)$. I will first filter series $x_k$ and $Y$ using equation (1). I will then use equation (7) and (8) to construct confidence interval for $e_{x_k}$ and $e_Y$. Specific algorithm is shown below.

**Algorithm 2** PVAR Variable Selectiono Using Partial Faithfulness

**Begin Selection:** Each Equation in PVAR at a time
$q = 1$, $v = 1$
**for** $q \leqslant Q$ **do**
  **for** $v \leqslant V$ **do**
    $X^{\text{final}}_{v,q} \leftarrow$ **Partial–Faithfulness Algorithm** $(y_{v,q}, [Y_1, Y_2, ..., Y_p])$
    $v = v + 1$
  **end for**
  $q = q + 1$
**end for**
**Estimate the System of Equations:**
$[y_{v,q}, X^{\text{final}}_{v,q}|\ \forall v \in [1, 2, ..., V]\ \ \forall q \in [1, 2, ..., q]]$

After the selection of variables for every individual equations, the PVAR estimate the model using standard VAR estimation method. There is an additional benefit of using partial faithfulness PVAR estimation. In the setting of PVAR, the variables that the partial faithfulness algorithm discard are not correlation with the underlying dependent variable in any possible means. The partial faithfulness algorithm, therefore, provides a picture of the linkages of domestic and foreign macroeconomic conditions.

# IV. Real World Applications

This section will present the performances of partial faithfulness PVAR(faithful-PVAR) in both real-world cases and simulation. With the assumption that European countries have close connections in economic activity and USA has an influential power over the global economy, I use real economic data of four European countries and the USA to present the explaining power of the faithful-PVAR model. These real-world applications use data from two sources. The OECD releases inflation, industrial output, and Total Factor Production data of its member states on a monthly basis. And the Global Economic Monitor(GEM) dataset from World Bank releases the monthly Real Effective Exchange Rate data. I used these five variables for the empirical application. Data from 2001.1 to 2011.6 and data from 2011.7 to 2016.6 are used as the training period of the partial-faithfulness algorithm and the forecasting period respectively. In 2017, [Schnücker(2017)] compared the forecasting result of few PVAR estimation methods. Since the forecasting setting of this section is nearly identical as [Schnücker(2017)]'s[11], I compared the estimation result of [Schnücker(2017)] with the faithful-PVAR. Regarding the toolbox used in the applications, I built the Partial Faithfulness algorithm in Matlab® and constructed the Panel Vector Autoregressive model on the basis of the Matlab®'s Econometrics ToolBox toolbox[12].

Results showed that the faithful-PVAR model performs extraordinary well in comparison with other PVAR models in out-of-sample forecasting. The below section analyzed the performances of faithful-PVAR against two other benchmark PVAR models in three real-world cases:

**First Case** The first case aims to conduct macroeconomic forecasting for five countries (Germany, France, USA, UK, and Italy) and two macroeconomic variables (growth rate of both CPI and Industrial Production) within the period of 2001.1 to 2016.6.

**Second Case** On the basis of the first case, the second case adds two more variables (growth rate of both Unemployment Rate and Effective Exchange Rate) to increase the dimension of the PVAR model.

**Third Case** On top of the second case, the third case adds one final country(Spain) to further increase the dimension of the model.

The following table summarizes the information of three real world cases and settings of the simulation. Due to the page limite, simulation results are placed in the appendix.

[11][Schnücker(2017)] used the REER data published by the OECD whereas this paper used the World Bank's version

[12]Link: https://www.mathworks.com/products/econometrics.html

| | Forecasting Case (1) | Forecasting Case (2) | Forecasting Case (3) | Simulation |
|---|---|---|---|---|
| Country | DEU, FRA, ITA, USA, GBR | DEU, FRA, ITA, USA, GBR | DEU, FRA, ITA, USA, GBR, ESP | Two Countries |
| Macro Variables | CPI, IP | CPI, IP, REER, UR | CPI, IP, REER, UR | Three Variables |
| # Lags | 6 | 6 | 6 | 4 |
| # Parameter (per Equation) | 60 | 120 | 144 | 24 |
| # Observations (per Equation) | 131 | 131 | 131 | 100 |
| Periods | 2001.1-2016.6 | 2001.1-2016.6 | 2001.1-2016.6 | |

***Table III Summary of PVAR Models corresponding to Three Real World Cases and Simulation***

The result showed that faithful-PVAR performs consistently better than the benchmark models. The Simulation section in appendix shows that the faithful-PVAR model has more precise parameter estimation and has imposed better parameter restrictions in comparison with one of the benchmark models. Unfortunately, since it is unfeasible to estimation a PVAR model without imposing any constraints, it is not possible to include performance of such model in our comparison. Before unfolding the details of models' performance, I will first introduce the specifications of two benchmark models.

**restricted-PVAR** For each macroeconomic variable of each individual country, the restricted-PVAR model incorporates macroeconomic variables of the same country and discards the others. The model, therefore, does not account for international linkages between variables. The model does allow for an unrestricted covariance matrix for innovation process.

**country-PVAR** The setting is similar to the restricted-PVAR model. That is, the country-PVAR model accounts for the influences of macroeconomic conditions within the same country but does not allow international influences. Unlike restricted-PVAR, the model does not allow for an unrestricted covariance matrix and imposes a diagonal restriction for the innovation covariance matrix. The restricted-PVAR model has a block diagonal restriction for all autoregressive coefficient matrices and a diagonal covariance matrix.

The following paragraphs presents the out-of-sample forecasting results of faithful-PVAR and other benchmarks models. The partial faithfulness variable selection algorithm used data from 2001.1 to 2011.6 to determine significance of partial correlation between variables. Due to time constraint, the partial faithfulness algorithm ran only once, over the mentioned period, and the variable selection results were used repeatedly for moving window coefficient estimations as well as h-step ahead forecasting over the period of 2011.7-2016.6. As shown in the simulation selection in appendix, the increase of iterations in partial faithfulness algorithm drastically increased the number of zero constraints on parameters and leads to different yet interesting results. I will present the out-of-sample forecasting result for both 1 iteration and 4 iterations cases. The evaluation of the forecasting power is mainly base on Mean Squared Forecasting Error(MSFE). As mentioned, [Schnücker(2017)] compared the forecasting performances of the PVAR models using different estimation methods, such as Lasso, Bayesian, and others. Thanks to the work of Schnücker, it is possible to compare performances of different dimension shrinkage approaches in the setting of PVAR. Since the first and third forecasting cases in this paper used the almost the identical sources of data[13], forecast on the same prediction period, and both used country-PVAR and restricted-PVAR model as benchmarks model, I compared the forecasting MSFE results of Lasso-PVAR, Bayesian-CC-PVAR and Bayesian-SSSS-PVAR from [Schnücker(2017)] in words.

| | country-PVAR | restricted-PVAR | faithful-PVAR | |
|---|---|---|---|---|
| | **Baseline** | | (1 Iteration) | (4 Iteration) |
| **MSFE h=1** | | | | |
| (1) | 1.00 | 0.86 | 0.88 | **0.84** |
| (2) | 1.00 | 0.95 | 0.80 | **0.78** |
| (3) | 1.00 | - | 0.85 | **0.83** |
| **MSFE h=6** | | | | |
| (1) | 1.00 | 0.94 | **0.89** | **0.89** |
| (2) | 1.00 | 0.95 | **0.87** | **0.87** |
| (3) | 1.00 | - | **0.89** | **0.89** |
| **MSFE h=12** | 1.00 | | | |
| (1) | 1.00 | **0.95** | 1.01 | 1.02 |
| (2) | 1.00 | 1.01 | **0.99** | **0.99** |
| (3) | **1.00** | - | 1.01 | 1.01 |

***Table IV Summary of Forecasting Results (MSFE of different models are expressed as percentage of country-PVAR's counterpart)***

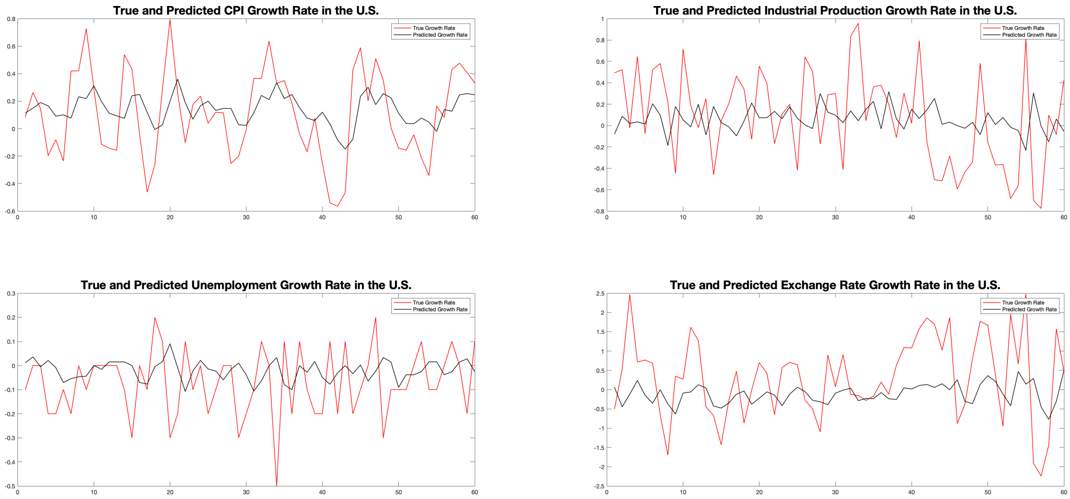
The forecasting results shown on Table III and Table IV are surprising and encouraging – not much that the faithful-PVAR performs better than other models in nearly all cases, but rather the fact that it achieves so using merely around 15% as many variables as other methods. Table IV shows that the faithful-PVAR model performs exceedingly better than country-PVAR and restricted-PVAR model in almost all forecasting scenarios, with one-year-ahead forecasting the only exception – twelve-step-ahead forecasting for models possibly converges to unconditional and do not differ as much. Comparing the forecasting results of faithful-PVAR to Lasso-PVAR, Bayesian-CC-PVAR and Bayesian-SSSS-PVAR for forecasting cases (1) and (3), the faithful-PVAR performs at least equally good as Lasso-PVAR in almost all cases, at least equally good as Bayesian-CC-PVAR in all cases, and better than Bayesian-SSSS-PVAR in all cases. The only exception happens in the case of $h = 12$, case (3), where the Lasso-PVAR performs 5% better than faithful-PVAR.

The following figures plot the true growth rates of U.S. macroeconomic indicators, represented by the red line, against their one-step-ahead faithful-PVAR predicated values in case (3). Figures for all four variables shows that the predicted values capture the latent tendency of the true value. It is also clear that forecasts of the four variables are not perfect, as the examples are more of toy models to compare different PVAR models. For macroeconomic variables as complex as CPI, Industrial Production, and the others, their precise moving patterns are unlikely to be perfectly captured with only four variables. In fact, [Stock and Watson(2005)] uses 131

[13] As mentioned, [Schnücker(2017)] used the OECD data for REER whereas this paper used the World Bank's version

macroeconomic variables in PVAR models to achieve close-to-decent results. There is another potential improvement of the PVAR models. As I set the assumption that the innovation process is multivariate normally distributed, the predicted values of four variables are not as volatile as true values. This assumption can further relax and might lead to better forecasting results.

***Figure 4.*** *One-Step-Ahead forecasting for U.S. Economy using four-iterations Faithful-PVAR*




The other noteworthy result is the parsimonious characteristics of the partial faithfulness algorithm. The following table shows the number of parameter included in each PVAR models. With only one and four iteration, the partial faithfulness selection has shrunk the parameter size to about 12% in the first case and 4% in the second and third cases. The number of parameter included in faithful-PVAR model is merely one fifth of that in country-PVAR, yet the prediction power increases by nearly 20% for one step ahead forecasting. As I increased the iteration of partial faithfulness process, the MSFE converged due to the 'small-dimension' characteristics of our forecasting cases. Due to time constraint, I have not conducted faithful-PVAR forecasting for higher dimension PVAR with more iterations in the variable selection process.

| **Number\|Percentage** | country-PVAR | | restricted-PVAR | | faithful-PVAR 1 iteration | | faithful-PVAR 4 iteration | |
|---|---|---|---|---|---|---|---|---|
| (1) | 12 | 20% | 12 | 20% | 6.2 | 12.4% | 3.30 | 6.7% |
| (2) | 24 | 20% | 24 | 20% | 5.55 | 4.6% | 3.25 | 3.3% |
| (3) | 24 | 16.7% | 24 | 16.7% | 5.91 | 4.1% | 3.25 | 2.7% |

***Table V Number and Percentage of Included Parameters Per Equation***

# V. Conclusion

This paper proposed a partial-correlation based dimension reduction method for high-dimensional Panel Vector Autoregressive model. With the use of a correct control variable set, explanatory variables that are partially correlated with the dependent variable could potentially be causally related to the dependent variable. The newly created approach uses such idea to shrink the dimension of a PVAR model. Due to the unclear persistent pattern as well as the small sample size issues commonly exist for macroeconomic time series, this paper first makes use of a robust inference construction method introduced by [Müller and Watson(2018)] to identify the long-run correlations between potential explanatory variables. With the aim of pinpointing the long-run correlation, [Müller and Watson(2018)]'s method identifies correlations by studying the asymptotic distribution of time series' long-run component from frequency dimension's viewpoint. Thanks to the one-to-one relation between spectral density and covariance matrix, this newly introduced inference construction method uses a generalized specification of spectral density to construct inferences that are adaptable to different persistent patterns. With the robust inference construction method, it is possible for a partial correlation based variable selection method to construct robust inference for correlations, given that the set of common variables is known. However, it is generally difficult to pinpoint the exact set of control variables. In this paper, I used the concept of partial faithfulness and the algorithm purposed by Bühlmann[Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis] for variable selection. The algorithm allows a selection process to proceed without knowing the exact set of common variables. The out-of-sample forecasting result showed that the method results in parsimonious models and competitive capacity in comparison with other PVAR estimation methods, such as the Bayesian-PVAR, the restricted-PVAR, and other penalty-based PVAR. With the number of explanatory variables being around 15% numbers as much of the restricted-models, the method performs consistently better.

However, due to time constraint, there are two categories of relevant questions remained unanswered. The first category lies in the forecasting potential of the model. With only four iterations, the model has shown its capacity to perform competitive out-of-sample forecasting with a small number of explanatory variables. However, the prediction of U.S. economy on the above cases is not yet perfect, presumably due to the "small number" of explanatory variables. The faithful-PVAR model is yet to release its full potential. [Stock and Watson(2005)] used 131 macroeconomic variables in different VAR models to explain the movement of business cycles which brought decent results. With the use of faithful-PVAR, it is feasible to include 131 macroeconomic monthly data with 100 countries in one PVAR model. Questions of the second category address theoretical properties and limitations of the model. Müller and Watson have derived the asymptotic properties and evaluated its finite sample performance for $(A, B, c, d)$ model's inferences. Despite the similarity between correlation and autocorrelation, using the $(A, B, c, d)$ model to construct auto-correlation inference needs justifications and possibly adjustments. Such research would also shed light on the limitations of the faithful-PVAR model. There is another assumption that might be challenged: the partial-faithfulness variable selection process makes the assumption that the coefficients of the PVAR models follows an absolute continuous distribution in terms of Lebegue measure density. The assumption holds under the common assumption that the innovation terms of the PVAR are normally distribution. However, this assumption might be violated for some real-world applications. Research regarding the robustness of the faithful-PVAR model would help researchers to evaluate the accuracy of the estimation.

# References


[Bühlmann et al.(2010)Bühlmann, Kalisch, and Maathuis] Bühlmann, Peter, Markus Kalisch, and Marloes H Maathuis, 2010, Variable selection in high-dimensional linear models: partially faithful distributions and the pc-simple algorithm, *Biometrika* 97, 261–278.

[Canova and Ciccarelli(2013)] Canova, Fabio, and Matteo Ciccarelli, 2013, Panel vector autoregressive models: A survey the views expressed in this article are those of the authors and do not necessarily reflect those of the ecb or the eurosystem., in *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*, 205–246 (Emerald Group Publishing Limited).

[Müller and Watson(2015)] Müller, Ulrich K, and Mark W Watson, 2015, Low-frequency econometrics, Technical report, National Bureau of Economic Research.

[Müller and Watson(2016)] Müller, Ulrich K, and Mark W Watson, 2016, Measuring uncertainty about long-run predictions, *Review of Economic Studies* 83, 1711–1740.

[Müller and Watson(2018)] Müller, Ulrich K, and Mark W Watson, 2018, Long-run covariability, *Econometrica* 86, 775–804.

[Schnücker(2017)] Schnücker, Annika, 2017, Penalized estimation of panel vector autoregressive models: a lasso approach, Technical report, Mimeo.

[Stock and Watson(2005)] Stock, James H, and Mark W Watson, 2005, An empirical comparison of methods for forecasting using many predictors, *Manuscript, Princeton University* .

[Uematsu and Tanaka(2019)] Uematsu, Yoshimasa, and Shinya Tanaka, 2019, High-dimensional macroeconomic forecasting and variable selection via penalized regression, *The Econometrics Journal* 22, 34–56.

# Appendix A. Simulation Result

As illustrated above, the simulation section uses 3 countries, 2 macroeconomic indicators, and 4 lags. The autoregressive coefficients of the data generating process(DGP) are set as follows,

$$\text{AR}(1) = \begin{pmatrix} 0.3 & 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & 0 & 0 \\ 0.1 & 0.2 & 0.3 & 0.25 & 0 & 0.2 \\ 0 & 0.2 & 0 & 0.3 & 0 & 0 \\ 0.2 & 0.2 & 0.2 & 0 & 0.3 & 0 \\ 0 & 0.2 & 0 & 0.2 & 0 & 0.3 \end{pmatrix} \quad \text{AR}(2) = \mathbf{0}$$

$$\text{AR}(3) = \begin{pmatrix} 0.15 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0.15 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.15 & 0.1 & 0 & 0.1 \\ 0 & 0.1 & 0 & 0.15 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0 & 0.15 & 0 \\ 0 & 0.1 & 0 & 0.1 & 0 & 0.15 \end{pmatrix} \quad \text{AR}(4) = \mathbf{0}$$

The innovation matrix, that is, the covariance matrix of error terms between equations, are set as the following,

$$\Sigma^{DGP} = \begin{pmatrix} 0.2 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.2 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.2 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.2 \end{pmatrix}$$

The evaluation of the simulation mainly focuses on two aspects, the mean squared error(MSE) between the true AR coefficients and estimated counterpart, the correctness of setting constraints on AR coefficents. Specifically,

**MSE**: the presented MSE are the average of squared distance between individual DGP's coefficients and corresponding estimated coefficients. Denote $N$ as the number of equations and $n$ the $n^{th}$ equation in the PVAR system,

$$MSE = \frac{1}{N^2 P} \sum_{n=1}^{N} \sum_{i=1}^{NP} (\hat{\beta}_{ni} - \beta_{ni}) \tag{A1}$$

**Correct Sparsity**: This indicator calculates the ratio of: The number of the underlying PVAR model correctly set a zero constraint on a zero coefficient and correctly relaxes a constraint on a nonzero coefficient over the number of entire coefficients.

The simulation runs 100 times and the results are shown on the below table take the average of individual simulations. As illustrated in the methodology section, with the increase of the iterations in the partial-faithfulness algorithm, the precision of the algorithm will increase. It implies that the increase of iteration will render a no less precise set of explanatory variables. The simulation will investigate the performances of the faithful-PVAR model with both 1 iteration and 4 iterations settings. Denote the number of iteration as $m$, the following table presents the simulation result.

| | country-PVAR | restricted-PVAR | Faithful-PVAR |
|---|---|---|---|
| **(1) Faithfulness One Iteration** | | | |
| Correct Sparsity Rate | 37.5% | 37.5% | 65.8% |
| MSE | 0.0182<br>(100%) | 0.0131<br>(72.0%) | 0.0164<br>(90.1%) |
| Number of Incorporated Variables | 48 | 48 | 19.6 |
| **(2) Faithfulness Four Iteration** | | | |
| Correct Sparsity Rate | 37.5% | 37.5% | 67.8% |
| MSE | 0.0187<br>(100%) | 0.0130<br>(69.6%) | 0.0151<br>(80.7%) |
| Number of Incorporated Variables | 48 | 48 | 9.0 |

***Table VI Summary Simulation Result***

The simulation result shows that the Mean Squared Error(MSE) of faithful-PVAR locates between country-PVAR and restricted-PVAR. The increase of iteration will also contribute to an increase in MSE. However, the variable selection process is by no mean in perfect accordance with the data generating process(DGP), presumably due to the small sample size of the data. I have performed large sample size simulations with a larger number of iterations, the result shows a clear latent of convergence in variable selection. Due to time constraint, I did not incorporate the large sample simulation result in this section.

# Appendix B. Matlab Functions

This section present Matlab functions that are used to implement the partial-faithfulness PVAR model. There are two categories of functions, one category of functions are provided by [Müller and Watson(2018)] and was used to construct confidence intervals. The other categories of functions are programmed by myself, including the Partial-Faithfulness selection procedure, PVAR estimation, $h$ step forecasting evaluation, simulation, and so on. To replication the result of partial-faithfulness PVAR, please first import data from **WorkingDataBig.mat**, add matlab_lrcov_programs to the path and enter **UltimateRunner.m**. Detailed use of the functions are shown below.

**partial.m**: This function examines the partial correlation between y and individual variable in variable set x, controlling on variable set z. Confidence only takes value 1, 2, 3, corresponding to 67%, 90%, 95% respectively. The first row of x and z are indices of the variable and should, therefore, be removed when doing operations.

**faithfulness.m**: This function runs the different procedures of the PC-aglorihtm and returns variables accordingly. (Designed to be embeded in **partial.m**)

**pvarEstimation.m**: Aggregate the entire processes of faithfulness-PVAR. First: Create a PVAR object. Second: Identify partial correlations between y an xs in equations i. Constraints in "varm" object are set by "AR" specification. The input of "AR" should be a $P \times 1$ cell array, with cell i being a $num \times 1$ array, containing values of all lag coefficients for lag i

**forecasting.m**: Return h-step ahead forecasting result for the entire forecasting periods

**lagconvert.m**: This function serves as a panel data matcher that convert the indices which are sorted variable-wise to country-wise (e.g. from [CPIUK; CPIUSA; GDPUK; GDPUSA] to from [CPIUK; GDPUK; GDPUSA; CPIUSA], if no lags)

**ARrestricted.m**: This function creates the restricted AR coefficient matrices for the country-PVAR and restricted-PVAR model

**CountZero.m**: This function counts the number of parameter restriction ina PVAR model. It is mainly used to count the number of zero-restrictions in Partial-faithfulness PVAR

**MSEcal.m**: Calculate the Mean Squared Error for the simulation

**MSEsingle.m**: Calculate the Mean Squared Error for the simulation, country-PVAR's verison

**MSFE.m**: Calculate the Mean Squared Forecasting Error for forecasting

**MSFEindividual.m**: Calculate the Mean Squared Forecasting Error for forecasting, country-PVAR's verison

**MSFEstat.m**: Calculate the Mean Squared Forecasting Error for forecasting. Static verison: used only 2001.1 to 2011.7 for parameter estimation and used the result statically for forecasting. Even thought the faithful-PVAR performs spectacularly well, this method might not be used in reality and thus not shown on this paper

**Panelindex.m**: This function linked the none-Panel sorted data to Panel data. e.g. linked indices of [CPIUK; CPIUSA; GDPUK; GDPUSA] their corresponding countries

**percentgae.m**: Calculate percentages of the MSFE using country-PVAR as baseline

**RunningProgram.m**: Interface of the all forecasting programs, take raw data as input and return all results from forecasting

**UltimateRunner.m**: This function aggregates all final commands of the program. One could obtain final results using this function

**simulation.m**: Return simulation results for the simulation section

**simulationCI.m**: Used for furthure research to investigate asymptotic properties of faithful-PVAR, results are not included in the paper since it has not been fully explored

**simulationCI.m**: Used for furthure research to investigate asymptotic properties of faithful-PVAR, results are not included in the paper since it has not been fully explored (country-PVAR version)

**Sparsity.m**: Calculate Sparsity as defined in simulation methodology

### *Appendix B. Functions Provided by Müller*

Comments of key sub-functions in each of the following functions are added.

**lrcov.m**: This function computes longrun covariability statistics for (x,y) (Used in **partial.m** for partial-faithful PVAR)

**Figure_1.m**: Plots Figure 1 in this paper. Made adjustments and added comments for unclear points

**Figure_2.m**: Plots Figure 2 in this paper. Made adjustments and added comments for unclear points

**Table_1_a.m**: Return GDP and consumption result of table 1 in this paper. Added comments for unclear points

**Table_1_b.m**: Return interested rates result of table 1 in this paper. Added comments for unclear points

**Table_2_a.m**: Return GDP and consumption result of table 2 in this paper. Added comments for unclear points

**Table_2_b.m**: Return interested rates result of table 2 in this paper. Added comments for unclear points