Joint Modeling of Sectoral Inflation Dynamics in Eurozone Countries[1].

by Julia Schaumburg and Quint Wiersma

Vrije Universiteit Amsterdam, the Netherlands

Tinbergen Institute, Amsterdam, the Netherlands

February, 2025

We introduce a new time-varying parameter tensor autoregressive (TAR) framework to model multi-dimensional time series. Our method is able to deal with high-dimensional time series with a natural tensor representation in a structured manner by exploiting the multi-dimensional structure of data using tensor decompositions. This allows us to combine flexibility with interpretability in our modeling approach. Time-variation is build into the tensor decomposition and modeled using a latent dynamic factor model. Estimation of the model parameters is done using a combination of the Expectation-Maximization algorithm and alternating least squares. We derive stationarity restrictions on the model parameters and impulse response functions. A Monte Carlo simulation study shows the good small sample performance of our method, even in high-dimensional settings that would have been infeasible in a standard vector autoregressive framework. We apply our method to model sectoral inflation for the five biggest economies in the eurozone, and estimate a sectoral level Phillips curve. We find that there is significant time-variation in the inflation dynamics and that the dynamic responses of the system are heterogeneous across countries, sectors, and variables. Our findings highlight heterogeneous local Phillips curves, where some sectors of the economy show a flatter curve than others. In response to the energy crisis following the Russian invasion of Ukraine in the first quarter of 2022 we observe muted responses of output growth to inflationary shocks.

Keywords: tensor autoregressive models, tensor decompositions, time varying parameters, dynamic factor model, Kalman filter

JEL Classification: C32, C51, E31

[1]Corresponding author: Quint Wiersma, Vrije Universiteit Amsterdam, School of Business and Economics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Phone: +31 20 598 10 00, Email: q.wiersma@vu.nl

1 || Introduction

Multi-dimensional array, or tensor, data is prominent in economics, finance, climate science, and many other fields. Statistical tools to analyze these high-dimensional data in a parsimonious and interpretable framework are of importance. In economics one such an example would be the modeling of a system of macroeconomic variables over time across a variety of geographical units (e.g. regions, countries, states), where the data exhibits a matrix structure. Traditionally these macroeconomic systems have been modeled using panel vector autoregressions (VARs), see e.g. Canova and Ciccarelli (2004, 2009). A downside of the panel VAR approach is that the number of parameters in the model, for each lag, is quadratic in the product of the number of variables and geographical units. Hence, the VAR model suffers from the curse of dimensionality, especially when matrix-valued data is modeled in vectorized form as is the case in the panel VAR model. Currently panel VARs have been limited to very small systems and mainly using computationally intensive Bayesian approaches. This issue was already noted by Chen et al. (2021a) who model the matrix structure of the data directly using their matrix autoregression (MAR).

Recently, several new approaches have been proposed to model the tensor-valued time series in their natural tensor representation, see Li and Xiao (2021), Wang et al. (2021), and Billio et al. (2023). However, it is well documented that when dealing with macroeconomic time series one has to consider some form time-variation in the parameters of the model. This is due to the fact that the underlying economic relationships driving the data are potentially not constant over time, e.g. crisis vs non-crisis periods, different monetary policy regimes, etc.. See Cogley and Sargent (2001), Canova and Ciccarelli (2004), Cogley and Sargent (2005), Primiceri (2005), and Canova and Ciccarelli (2009) for discussions and evidence on the importance of allowing for changing dynamic responses over time. The current approaches in the literature cannot capture this instability in the parameters of the model over time.

In this paper we extend the approach of Chen et al. (2021a), in spirit of Billio et al. (2023), to handle multi-dimensional data of arbitrary dimensions, but on top of that allow for time-variation in the autoregressive parameters using the approach of Gorgi et al. (2018). Our time-varying parameter tensor autoregressive (TAR) model can both model the tensor structure of the data parsimoniously through the use of a CANDECOMP/PARAFAC (CP) decomposition, a higher-order singular value decomposition, and capture time-variation in the autoregressive coefficient tensor through a dynamic factor specification. The dynamic factor can be interpreted as a time-varying scale in our autoregressive coefficient tensor, while the CP decomposition elements capture the normalized transition from the variables in the past to the present. This specification allows for local instabilities, but is guaranteed to generate a stationary system under some standard conditions on the dynamic factor process. Our tensor-based approach can be cast into a restricted time-varying VAR or dynamic factor model, but the restrictions that are automatically implied by the tensor structure are more meaningful and economically interpretable.

In contrast to the computationally intensive Bayesian approaches that have been proposed in the literature to estimate VAR and TAR models (see e.g. Canova and Ciccarelli, 2004; Primiceri, 2005; Canova and Ciccarelli, 2009; Billio et al., 2023), we propose a computationally efficient algorithm to estimate the parameters of the model. Our estimation algorithm is based on the Expectation-Conditional Maximization (ECM) algorithm of Meng and Rubin (1993) in combination with the alternating least squares (ALS) algorithm. The ALS algorithm has been used extensively estimating tensor decompositions, such as CP, see Kolda (2006), Bader and Kolda (2006), and Kolda and Bader (2009). Moreover, it has also been proposed in the recent literature for estimating tensor autoregressive models (Chen et al., 2021a; Li and Xiao, 2021; Wang et al., 2021) and tensor factor models (Chen et al., 2021b). As we show that our model can be reformulated as a dynamic factor model, we have that our TAR model admits a natural state space representation. The E-step of the ECM algorithm is then readily available from a routine application of the Kalman filter and smoother routines, see Shumway and Stoffer (1982) for an early application of the EM algorithm to state space models. In a Monte Carlo simulation study we illustrate the good small sample performance of our estimation procedure.

We derive in line with Gorgi et al. (2018) impulse response functions (IRFs) for our model which are compatible with the generalized IRFs defined by Koop et al. (1996) for nonlinear models. The IRFs rely on the overall stability of the process. We show that these time-varying IRFs are decomposable into a static and a dynamic component enabling us to have easy-to-interpret IRFs. The static component measures the initial impact of a shock on the system, while the dynamic component measures the persistence of the shock over the horizon across time.

In the empirical application we consider modeling inflation at the sectoral level. We use a system of the five biggest economies of the eurozone for a variety of sectors with three macroeconomic variables, namely output, employment, and inflation. Given the current high inflation period and the specific nature of these inflationary shocks, namely a negative energy supply shock, we try to model inflation at a disaggregated level. Using this approach we are able to find heterogenous inflation dynamics for different sectors of the economy, as well as, investigate their impact on output. Seeing the recent discussion of a flattening of the Phillips curve (see Vlekke et al., 2020; Del Negro et al., 2020; Hazell et al., 2022) and the heterogeneous results across countries (Vlekke et al., 2020), it might be warranted to look at local sectoral level Phillips curves.

There is a large body of literature concerned with modeling inflation, for an overview see Faust and Wright (2013). The main modeling approaches are Phillips curve inspired models (see e.g. Stock and Watson, 2008, 2010; Stella and Stock, 2013), time-varying parameter VARs (see e.g. Primiceri, 2005; Cogley et al., 2010), and stochastic volatility models (see e.g. Chan et al., 2013; Chan, 2017). These different strands of the literature seem to agree on the fact that the persistence and volatility of inflation have changed over time. Our approach bears most similarity to the Bayesian time-varying parameter VAR approach of Primiceri (2005) and Cogley et al. (2010). However, we differ, besides the local nature of our modeling approach, in the way we introduce time-variation in the parameters. The random walk specification of Primiceri (2005) and Cogley et al. (2010) allows for the counterintuitive implication of unbounded growth in inflation, while our model exhibits mean-reverting behavior.

Using our time-varying TAR, we find evidence for heterogeneous local Phillips curves, where some sectors of the economy show a flatter curve than others. Moreover, in line with the literature, we also find substantial time-variation in the Phillips curve, with stronger inflationary responses to employment at the end of the sample. Regarding the impact of the energy supply shock arising from the invasion of Ukraine by Russia in 2022 on the economy, we find dampened responses compared to before the start of the war.

The remainder of the paper is structured as follows. First, in section 2 our time-varying parameter tensor autoregressive model will be introduced along side with some short introduction into tensor calculus. The estimation routine will be outlined in section 3. The validity and performance of our method will be highlighted in section 4 through a Monte Carlo simulation study. The application of our method to model sectoral inflation and a local Phillips curve will be discussed in section 5. Finally, section 6 concludes.

2 ‖ Time-Varying Tensor Autoregressive Model

2.1 ‖ Model

Consider a n-th order tensor $\mathcal{Y}_t \in \mathbb{R}^{J_1 \times \cdots \times J_n}$ of time series, where J_k is a scalar denoting the dimension of mode k. We define the general time-varying parameter tensor autoregressive (TAR) model with p autoregressive lags as follows

$$\mathcal{Y}_t = \sum_{p=1}^{P} \langle \mathcal{A}_t^p, \mathcal{Y}_{t-p} \rangle + \mathcal{E}_t, \quad t \in \mathbb{Z}, \tag{1}$$

where $\mathcal{A}_t^p \in \mathbb{R}^{J_1 \times \cdots \times J_n \times J_1 \times \cdots \times J_n}$, $\mathcal{E}_t \in \mathbb{R}^{J_1 \times \cdots \times J_n}$, and $\langle \cdot, \cdot \rangle$ denotes the contracted tensor product as defined in (A.1) in Appendix A. We refer for a general introduction of tensor notation and algebra concepts can be to Appendix A and Kolda (2006), Bader and Kolda (2006), and Kolda and Bader (2009).

Regarding the innovation tensor $\mathcal{E}_t$ we specify a tensor normal distribution (see Hoff, 2011), i.e. $\mathcal{E}_t \sim \mathcal{TN}(0; \Sigma_1, \ldots, \Sigma_n)$. The tensor normal assumption of the innovation tensor implies a separable covariance structure along each mode of the tensor, this structure has been used extensively in the spatial statistics literature (see Cressie, 2015). Intuitively, this covariance structure implies that $\Sigma_k \in \mathbb{R}^{J_k \times J_k}$ captures the dependency structure along mode k. In vectorized form we get the following kronecker structure for the covariance matrix of innovations

$$\varepsilon_t = \text{vec}(\mathcal{E}_t) \sim \mathcal{N}(0, \Sigma),$$

with $\Sigma = \Sigma_n \otimes \cdots \otimes \Sigma_1$.

For each autoregressive coefficient tensor $\mathcal{A}_t^p$ in (1) we specify a time-varying rank-R_p CP decomposition as introduced in (A.3),

$$\mathcal{A}_t^p = \sum_{r=1}^{R_p} \lambda_t^{p,r} \cdot u_1^{p,r} \circ \cdots \circ u_{2n}^{p,r}, \tag{2}$$

where $\circ$ denotes the outer product. The dynamic factor $\lambda_t^{p,r}$ follows an autoregressive structure, i.e.

$$\lambda_t^{p,r} = \alpha_{p,r} + \phi_{p,r}\lambda_{t-1}^{p,r} + \eta_t^{p,r}, \quad \eta_t^{p,r} \sim \mathcal{N}(0, \sigma_{p,r}^2), \tag{3}$$

where we assume $|\phi_{p,r}| < 1$. We assume that $\eta_t^{p,r}$ and $\mathcal{E}_t$ are uncorrelated at all time points.

The model described by (1) in combination with the time-varying CP decomposition for each autoregressive tensor in (2) and (3) we refer to as the time-varying TAR$(P, R_1, \ldots, R_P)$, which highlights that each lag in the model can have a different CP rank for the autoregressive tensor.

Using the proposed CP decompositions for the autoregressive coefficient tensors we are able to rewrite our model (1) into a multi-linear structure using the mode-k products, see (A.2),

$$\begin{aligned} \mathcal{Y}_t &= \sum_{p=1}^{P} \langle \mathcal{A}_t^p, \mathcal{Y}_{t-p} \rangle + \mathcal{E}_t \\ &= \sum_{p=1}^{P} \sum_{r=1}^{R_p} \lambda_t^{p,r} \cdot \mathcal{Y}_{t-p} \times_1 U_1^{p,r} \cdots \times_n U_n^{p,r} + \mathcal{E}_t, \end{aligned} \tag{4}$$

where $U_k^{p,r} = u_{k+n}^{p,r} \circ u_k^{p,r}$. This formulation shows intuitively how the different elements of the CP decomposition play a role in the coefficient tensor. The dynamic latent factor $\lambda_t^{p,r}$ acts as a time-varying scale parameter, while the U's take the role of normalized transition matrices alongside each mode of the tensor. Within each $U_k^{p,r}$ we have that $u_k^{p,r}$ are the weights in the weighted sum of the time series p periods before for the r-th rank comoponent alongside mode k, while the elements in $u_{k+n}^{p,r}$ measure the impact of this weighted sum of past time series alongside mode k on the future values of the time series along mode k.

Our modeling framework can be cast into the VAR framework, when we vectorize our TAR model we obtain the following time-varying VAR representation of the model

$$y_t = \sum_{p=1}^{P} A_{p,t}^{(\mathcal{R})} y_{t-p} + \varepsilon_t,$$

with $y_t = \text{vec}(\mathcal{Y}_t)$ an $\prod_{k=1}^n J_k$-dimensional vector and $A_{p,t}^{(\mathcal{R})} = \lambda_t^{p,r} \cdot (u_{2n}^{p,r} \otimes \cdots \otimes u_{n+1}^{p,r}) \circ (u_n^{p,r} \otimes \cdots \otimes u_1^{p,r})$ an $\prod_{k=1}^n J_k \times \prod_{k=1}^n J_k$ matrix, where $\mathcal{R} = \{n+1, \ldots, 2n\}$. Using the multi-linear formulation of our model (4) we are able to alternatively write the matricized coefficient tensor as $A_{p,t}^{(\mathcal{R})} = \lambda_t^{p,r} \cdot (U_n^{p,r} \otimes \cdots \otimes U_1^{p,r})$. See Appendix A for a discussion of the matricization of tensors. The benefit of modeling the time series in their natural tensor representation becomes apparent from this formulation. The number of parameters related to the constant part in our autoregressive coefficient tensor is only $2\sum_{k=1}^n J_k$, while in a standard VAR framework this would be $\prod_{k=1}^n J_k^2$.

From this VAR representation of our model it becomes obvious that our model can also be viewed as a restricted time-varying dynamic factor model, i.e.

$$\begin{aligned} y_t &= L_t \lambda_t + \varepsilon_t \\ &= \sum_{p=1}^{P} \sum_{r=1}^{R_p} L_t^{p,r} \lambda_t^{p,r} + \varepsilon_t, \end{aligned}$$

where the time varying loading matrix is given by $L_t = (L_t^{1,1}, \ldots, L_t^{1,R_1}, L_t^{2,1}, \ldots, L_t^{P.R_P})$ with $L_t^{p,r} = (u_{2n}^{p,r} \otimes \cdots \otimes u_{n+1}^{p,r}) \circ (u_n^{p,r} \otimes \cdots \otimes u_1^{p,r}) y_{t-p}$ and $\lambda_t = (\lambda_t^{1,1}, \ldots, \lambda_t^{1,R_1}, \lambda_t^{2,1}, \ldots, \lambda_t^{P,R_P})'$.

Our model is similar in spirit to other proposed models in the literature regarding tensor-valued time series modeling. Chen et al. (2021a) and Li and Xiao (2021) choose to impose a multi-linear structure on the model, i.e. they model the autoregressive coefficient structure directly using a multi-linear product as introduced by Hoff (2015). This approach is intuitive, but has as a downside that the parsimonious structure imposed on the model is ad-hoc. In this paper we start from the tensor generalization of the linear structure of the VAR model, see (1), and impose dimensionality reductions through the means of a well-defined tensor decomposition. Similar in this regard is the work of Wang et al. (2021) who also consider a tensor decomposition to introduce a parsimonious representation in (1). Our approaches differ in the type of decomposition considered, Wang et al. (2021) use the Tucker decomposition. Moreover, they focus on a high-dimensional setting in which regularization on top of a parsimonious structure needs to be considered. Wang et al. (2021) consider multiple different penalized estimators and show their finite sample error bounds. The CP decomposition has been used before to create a lower rank structure in tensor models, Billio et al. (2023) use this approach in an autoregressive setting similar to ours. However, to the best of our knowledge our approach is the first to allow for time-variation in the autoregressive dynamics. The way we introduce the time-variation in the autoregressive coefficient tensor is similar to Gorgi et al. (2018), which allow for time-variation in their VAR model by means of a latent factor structure.

2.2 || Stationarity

We establish the conditions under which our time-varying TAR model is stationary. Following Gorgi et al. (2018), we note that because of the assumption $|\phi_{p,r}| < 1$ we have that $\{\lambda_t^{p,r}\}_{t\in\mathbb{Z}}$ is a strictly stationary process and, hence, the matrices $\{A_{p,t}^{(\mathcal{R})}\}_{t\in\mathbb{Z}}$ form a stationary sequence. Moreover, as the factors are normally distributed we have that all elements of $A_{p,t}^{(\mathcal{R})}$ are normally distributed $\forall t \in \mathbb{Z}$ and, hence,

$$\mathbb{E}\,\rho(A_{p,t}^{(\mathcal{R})}) \leq \mathbb{E}\|A_{p,t}^{(\mathcal{R})}\|_F \leq \sqrt{\mathbb{E}\,\mathrm{tr}(A_{p,t}^{(\mathcal{R})\prime} A_{p,t}^{(\mathcal{R})})} < \infty,$$

where $\rho(\cdot)$ denotes the spectral norm. It follows then that for the companion matrix

$$A_t = \begin{pmatrix} A_{1,t}^{(\mathcal{R})} & A_{2,t}^{(\mathcal{R})} & \cdots & A_{P-1,t}^{(\mathcal{R})} & A_{P,t}^{(\mathcal{R})} \\ I_{\prod_{k=1}^n J_k} & 0 & \cdots & 0 & 0 \\ 0 & I_{\prod_{k=1}^n J_k} & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_{\prod_{k=1}^n J_k} & 0 \end{pmatrix}, \tag{5}$$

where $I_{\prod_{k=1}^n J_k}$ denotes the identity matrix of dimensions $\prod_{k=1}^n J_k \times \prod_{k=1}^n J_k$, also holds that $\mathbb{E}\,\rho(A_t) < \infty$. Then given by Theorem 1.1 of Bougerol and Picard (1992), we have that whenever the Lyapunov exponent γ_m is strictly negative, i.e.

$$\gamma_m = \frac{1}{m}\,\mathbb{E}\log\|A_{t-1}\cdots A_{t-m}\| < 0, \tag{6}$$

for some $m \in \mathbb{N}$, the process $\{y_t\}_{t\in\mathbb{Z}}$ and, hence, $\{\mathcal{Y}_t\}_{t\in\mathbb{Z}}$ is strictly stationary.

In general, for a given set of parameters, we can verify this stationarity condition through simulations. We select a large m and then estimate the expectation in (6) by a Monte Carlo estimate using a sufficiently large numbers of draws of $\mathcal{Y}_t$ from the TAR model. In an empirical setting one can validate that the estimated model is stationary using the same approach but replacing the parameter values by their estimates. In the special case of the TAR(1, 1) model we are able derive from (6) a condition on the parameters of the factor dynamics for which the TAR(1, 1) model is stationary.

Considering $m = 1$ we obtain the following inequality for the Lyapunov exponent with the spectral norm,

$$
\begin{aligned}
\gamma_1 &= \mathbb{E}\log\rho(A^{(\mathcal{R})}_{1,t-1}) \\
&= \mathbb{E}\log|\lambda^{1,1}_{t-1}| + \log\rho((u^{1,1}_{2n}\otimes\cdots\otimes u^{1,1}_{n+1})\circ(u^{1,1}_{n}\otimes\cdots\otimes u^{1,1}_{1})) \\
&= \mathbb{E}\log|\lambda^{1,1}_{t-1}| + \log\|u^{1,1}_{2n}\otimes\cdots\otimes u^{1,1}_{n+1}\|_2 + \log\|u^{1,1}_{n}\otimes\cdots\otimes u^{1,1}_{1}\|_2 \\
&= \mathbb{E}\log|\lambda^{1,1}_{t-1}| + \sum_{k=1}^{n}\log\|u^{1,1}_{k+n}\|_2 + \sum_{k=1}^{n}\log\|u^{1,1}_{k}\|_2 \\
&\leq \log\mathbb{E}|\lambda^{1,1}_{t-1}|,
\end{aligned}
$$

where the final step uses that all u's are unit vectors and the inequality follows from Jensen's Inequality. As $\lambda^{1,1}_{t-1}$ follows unconditionally a normal distribution with mean $\mu = \alpha_{1,1}/(1-\phi_{1,1})$ and variance $\sigma^2 = \sigma^2_{1,1}/(1-\phi^2_{1,1})$, we have that $|\lambda^{1,1}_{t-1}|$ follows a folded normal distribution, hence, $\mathbb{E}|\lambda^{1,1}_{t-1}| = \sigma^2\sqrt{2/\pi}e^{(-\mu^2/2\sigma^2)} + \mu(1-2\Phi(-\mu/\sigma))$, where $\Phi(\cdot)$ is the cdf of a standard normal. This implies that $\gamma_1 < 0$ whenever $\mathbb{E}|\lambda^{1,1}_{t-1}| < 0$ and all conditions of Theorem 1.1 of Bougerol and Picard (1992) are satisfied. We conclude that our time-varying TAR(1, 1) model is stationary as long as the dynamic factors process parameters obey the restriction $\mathbb{E}|\lambda^{1,1}_{t-1}| < 0$.

2.3 || Forecasting

Next we show how our proposed time-varying TAR model can be used for forecasting. The h-step ahead forecast is defined as the conditional expectation

$$
\hat{\mathcal{Y}}_{t+h} = \mathbb{E}[\mathcal{Y}_{t+h}|\mathcal{F}_t],
$$

where $\mathcal{F}_t$ denotes the information set up to time point t, i.e $\mathcal{F}_t = \{\mathcal{Y}_t,\dots,\mathcal{Y}_1\}$. Instead of finding a closed form expression for the conditional expectation we use a Monte Carlo estimator to obtain this conditional expectation. To obtain the Monte Carlo estimate for the conditional expectation we simulate M conditional paths for $\mathcal{Y}_{t+1},\dots,\mathcal{Y}_{t+h}$ given $\mathcal{F}_t$ and a set of parameters and then estimate the conditional expectation as follows

$$
\hat{\mathbb{E}}\left[\mathcal{Y}_{t+h}|\mathcal{F}_t\right] = \frac{1}{M}\sum_{i=1}^{M}\mathcal{Y}^{(i)}_{t+h},
$$

where $\mathcal{Y}^{(i)}_{t+h}$ denotes a draw from $p(\mathcal{Y}_{t+h}|\mathcal{F}_t)$.

2.4 || Impulse response function

An important tool to analyze dynamic responses of an autoregressive model is the impulse response function (IRF). The IRF allows us to trace out the response of one variable in our multivariate system to an impulse or shock of another variable in the system. We will derive in the forecast error and orthogonal IRFs for our time-varying TAR model. In doing so we follow the approach of Gorgi et al. (2018) who derive IRFs for their time-varying VAR model, and show that their definition of the IRF coincides with the non-linear IRF definition provided by Koop et al. (1996).

Given the stationarity condition discussed above, the time-varying TAR model admits a moving average representation. More specifically, the VAR formulation of our model can be written as an infinite length vector moving average, see Lütkepohl (2005). Hence,

$$y_t = \sum_{j=0}^{\infty} \Phi_{t-j} \varepsilon_{t-j},$$

with $\Phi_t = I$, where I denotes the identity matrix of appropriate size, and $\Phi_{t-j} = \prod_{i=0}^{j-1} A_{t-i}$ with A_t being the companion matrix as defined in (5).

Given this moving average representation, the IRF of a shock at time t at horizon h is then given by

$$\Psi_h(t) = \frac{\partial y_{t+h}}{\partial \varepsilon_t} = \prod_{i=0}^{h-1} A_{t+h-i}. \tag{7}$$

However, because this definition of the IRF depends on random shocks $\eta_1^{p,r}, \ldots, \eta_{t+h}^{p,r}$, through its dependence on $\lambda_{t+1}^{p,r}, \ldots, \lambda_{t+h}^{p,r}$, the current formulation is not directly implementable. Hence, we follow Gorgi et al. (2018) and take the conditional expectation of (7) on the past observed data before the shock occurred to make the IRF formula operational, i.e.

$$\Psi_h(t) = \mathbb{E}\left[\frac{\partial y_{t+h}}{\partial \varepsilon_t} | \mathcal{F}_{t-1}\right] = \mathbb{E}\left[\prod_{i=0}^{h-1} A_{t+h-i} | \mathcal{F}_{t-1}\right]. \tag{8}$$

Similar to our approach when forecasting, we evaluate the conditional expression using a Monte Carlo estimator. We sample M draws from the conditional density $p(\lambda_{t+h}, \ldots, \lambda_{t+1} | \mathcal{F}_{t-1})$ using Bayes' rule and an application of the Kalman filter to obtain a draw from $\prod_{i=0}^{h-1} A_{t+h-i}$ conditional on $\mathcal{F}_{t-1}$.

Note that because our model can be casted in a dynamic factor model framework it can be formulated as a state space model. This allows us to directly make use of the Kalman filter and smoother routines. Using Bayes' rule we can rewrite the joint conditional density $p(\lambda_{t+h}, \ldots, \lambda_{t+1} | \mathcal{F}_t)$ as

$$p(\lambda_{t+h}, \ldots, \lambda_{t+1} | \mathcal{F}_t) = p(\lambda_{t+h} | \lambda_{t+h-1}) \cdots p(\lambda_{t+2} | \lambda_{t+1}) p(\lambda_{t+1} | \mathcal{F}_t).$$

Observe that $p(\lambda_{t+h-j}^{p,r} | \lambda_{t+h-j-1}^{p,r}) \sim \mathcal{N}(\alpha_{p,r} + \phi_{p,r} \lambda_{t+h-j-1}^{p,r}, \sigma_{p,r}^2)$ and $\lambda_t^{p,r}$ and $\lambda_t^{q,s}$ for each $(p, r) \neq (q, s)$ and that $p(\lambda_{t+1} | \mathcal{F}_t)$ also follows a Gaussian distribution, with the mean and variance readily available from an application of the Kalman filter.

In the special case of a TAR(1, 1) model we can decompose this general expression (8) for the IRF into two components, one capturing the initial impact and one capturing the decay over the horizon,

$$\Psi_h(t) = \mathbb{E}\left[\prod_{i=0}^{h-1} \lambda^{1,1}_{t+h-i} \cdot (u^{1,1}_{2n} \otimes \cdots \otimes u^{1,1}_{n+1}) \circ (u^{1,1}_{n} \otimes \cdots \otimes u^{1,1}_{1}) | \mathcal{F}_{t-1}\right]$$
$$= \underbrace{(u^{1,1}_{2n} \otimes \cdots \otimes u^{1,1}_{n+1}) \circ (u^{1,1}_{n} \otimes \cdots \otimes u^{1,1}_{1})}_{\text{initial impact}} \cdot \underbrace{\tau^{h-1} \cdot \mathbb{E}\left[\prod_{i=0}^{h-1} \lambda^{1,1}_{t+h-i} | \mathcal{F}_{t-1}\right]}_{\text{decay}},$$

with $\tau = (u^{1,1}_{n} \otimes \cdots \otimes u^{1,1}_{1})'(u^{1,1}_{2n} \otimes \cdots \otimes u^{1,1}_{n+1})$.

So far, we have discussed the forecast error IRFs. The orthogonalized counterparts are calculated by first taking a Cholesky decomposition of the covariance matrix $\Sigma = \Sigma_n \otimes \cdots \otimes \Sigma_1$, which can be calculated by a kronecker product of Cholesky decompositions of the individual mode covariance matrices as $\Sigma^{\frac{1}{2}} = \Sigma_n^{\frac{1}{2}} \otimes \cdots \otimes \Sigma_1^{\frac{1}{2}}$, where we denote the Cholesky decomposition of Σ_k as $\Sigma_k = \Sigma_k^{\frac{1}{2}} \Sigma_k^{\frac{1}{2}\prime}$. The orthogonalized IRF can then be calculated as

$$\Psi^{\star}_h(t) = \Psi_h(t)\Sigma^{\frac{1}{2}}. \tag{9}$$

The Cholesky decomposition used to orthogonalize the shocks in the IRFs imposes a recursive structure on the model along each mode, hence, the orthogonalized IRFs in (9) are not invariant to the ordering of the variables along each mode. An advantage of the tensor modeling approach over a vectorized form VAR approach is the simplified recursive structure in the model when applying a Cholesky decomposition for identification of the structural shocks. When a vectorized model is used one has to order the time series at all modes jointly, e.g. in a country, sector, and macroeconomic variable model one would need to order the time series at the country-sector-variable level. In contrast, in our TAR model we would only need to order the time series along each mode separately combined with an ordering of the modes.

A consequence of the tensor normal distribution for the innovations and, hence, the separable covariance structure along each mode, is that the recursive structure of the model is restricted. More specifically, the orthogonalized IRFs produced by the TAR model allow for contemporaneous responses between time series when the within mode contemporaneous response is allowed according to the Cholesky decomposition. As an example our model restricts contemporaneous responses of sector i in country A to sector j in country B if sectors are ordered such that sector i appears before sector j in the mode of the tensor related to the sectors of the economy. This is different from the VAR approach, where sector j in country A only has a restricted contemporaneous response to sector i in country B whenever the specific sector-country combination is appearing above the other sector-country combination in the order of the time series.

3 || Estimation

Estimating the parameters of the time-varying TAR model is done via a combination of the EM and alternating least squares algorithms. Considering a total of T time series observations, the complete data log-likelihood for the vectorized data is given by

$$\begin{aligned}
\mathcal{L}(y_{1:T}, \lambda_{P+1:T}) &= \sum_{t=P+1}^{T} \log p(y_t | y_{t-1}, \dots, y_{t-p}, \lambda_t) + \sum_{t=P+2}^{T} \log p(\lambda_t | \lambda_{t-1}) \\
&= \sum_{t=P+1}^{T} \log p(y_t | y_{t-1}, \dots, y_{t-p}, \lambda_t) + \sum_{t=P+2}^{T} \sum_{p=1}^{P} \sum_{r=1}^{R_p} \log p(\lambda_t^{p,r} | \lambda_{t-1}^{p,r}) \\
&= -\frac{T-P}{2} \prod_{k=1}^{n} J_k \log(2\pi) - \frac{T-P}{2} \sum_{k=1}^{n} J_{-k} \log\det(\Sigma_k) \\
&\quad - \frac{1}{2} \sum_{t=P+1}^{T} \left\| \Sigma^{-\frac{1}{2}} (y_t - L_t \lambda_t) \right\|_2^2 \\
&\quad - \frac{(T-P-1)R}{2} \log(2\pi) - \frac{T-P-1}{2} \log(\sigma_{p,r}^2) \\
&\quad - \frac{1}{2} \sum_{t=P+2}^{T} \sum_{p=1}^{P} \sum_{r=1}^{R_p} \frac{(\lambda_t^{p,r} - \alpha_{p,r} - \phi_{p,r} \lambda_{t-1}^{p,r})^2}{\sigma_{p,r}^2},
\end{aligned} \tag{10}$$

with $y_{1:T} = (y_1', \dots, y_T')'$, $\lambda_{P+1:T} = (\lambda_{P+1}', \dots, \lambda_T')'$, $U_{p,r}^{(\mathcal{R})} = \mathrm{mat}_{\mathcal{R}}(\mathcal{U}_{p,r})$, see Appendix A for the definition of the matricization operator $\mathrm{mat}(\cdot)$, and $\mathcal{U}_{p,r} = u_1^{p,r} \circ \cdots \circ u_{2n}^{p,r}$, and $R = \sum_{p=1}^{P} R_p$.

Taking the conditional expectation of the complete data log-likelihood in (10) w.r.t. the λ's conditional on the data for given parameter values yields the following

$$\begin{aligned}
\mathbb{E}\left[\mathcal{L}(y_{1:T}, \lambda_{P+1:T}) | \mathcal{F}_T\right] &= -\frac{T-P}{2} \prod_{k=1}^{n} J_k \log(2\pi) - \frac{T-P}{2} \sum_{k=1}^{n} J_{-k} \log\det(\Sigma_k) \\
&\quad - \frac{1}{2} \sum_{t=P+1}^{T} \left\| \Sigma^{-\frac{1}{2}} \left(y_t - L_t \hat{\lambda}_t\right) \right\|_2^2 - \frac{1}{2} \sum_{t=P+1}^{T} \left\| \Sigma^{-\frac{1}{2}} L_t \hat{V}_t^{\frac{1}{2}} \right\|_2^2 \\
&\quad - \frac{(T-P-1)R}{2} \log(2\pi) - \frac{T-P-1}{2} \sum_{p=1}^{P} \sum_{r=1}^{R_p} \log(\sigma_{p,r}^2) \\
&\quad - \frac{1}{2} \sum_{t=P+2}^{T} \sum_{p=1}^{P} \sum_{r=1}^{R_p} \frac{\mathbb{E}[(\lambda_t^{p,r} - \alpha_{p,r} - \phi_{p,r} \lambda_{t-1}^{p,r})^2 | \mathcal{F}_T]}{\sigma_{p,r}^2},
\end{aligned} \tag{11}$$

where $\hat{\lambda}_t = \mathbb{E}[\lambda_t | \mathcal{F}_T]$ and $\hat{V}_t = \mathrm{Var}(\lambda_t | \mathcal{F}_T)$ are obtained from the Kalman smoother, as well as, $\mathbb{E}[\lambda_t \lambda_{t-1} | \mathcal{F}_T]$ which is a straightforward application of the covariance smoother, see Durbin and Koopman (2012).

The expectation conditional maximization (ECM) algorithm of Meng and Rubin (1993) then revolves around iteratively updating the parameters of the model, i.e. $u_1^{p,r}, \dots, u_{2n}^{p,r}$, $\alpha_{p,r}$, $\phi_{p,r}$ $\sigma_{p,r}^2$ for all (p,r) and $\Sigma_1, \dots, \Sigma_n$, keeping the other parameters fixed at their current estimates, by (conditionally) maximizing (11).

Starting with the optimization of $\alpha_{p,r}$, $\phi_{p,r}$, and $\sigma^2_{p,r}$ we observe that our update step becomes equivalent to the standard complete data ordinary least squares solution using maximum likelihood estimation where the sufficient statistics are replaced by their expectations, i.e.

$$\phi_{p,r} \leftarrow \frac{\sum_{t=p+2}^{T} \mathbb{E}[\lambda_t^{p,r}\lambda_{t-1}^{p,r}|\mathcal{F}_T] - \hat{\lambda}_t\hat{\lambda}_{t-1}}{\sum_{t=P+2}^{T} \mathbb{E}[(\lambda_{t-1}^{p,r})^2|\mathcal{F}_T] - \mathbb{E}[\lambda_{t-1}^{p,r}|\mathcal{F}_T]^2}, \tag{12}$$

$$\alpha_{p,r} \leftarrow \frac{1}{T-P-1} \sum_{t=P+2}^{T} \hat{\lambda}_t^{p,r} - \phi_{p,r}\hat{\lambda}_{t-1}^{p,r}, \tag{13}$$

and

$$\sigma^2_{p,r} \leftarrow \frac{1}{T-P-1} \sum_{t=P+2}^{T} \mathbb{E}[(\lambda_t^{p,r} - \alpha_{p,r} - \phi_{p,r}\lambda_{t-1}^{p,r})^2|\mathcal{F}_T]. \tag{14}$$

The mean squared error part of (11) related to $u_1^{p,r}, \ldots, u_{2n}^{p,r}$ for all (p,r) and $\Sigma_1, \ldots, \Sigma_n$ can be rewritten in tensor representation. This tensor representation will help us separate efficiently the optimization of the different parameters acting on the different modes. To ensure that the notation stays tractable we replace the double superscripts p and r, that indicate the lag and CP-rank component within each lag, by a single superscript $r_* = 1, \ldots, R$. The relationship between r_* and (p,r) is given by $p = \sum_{q=1}^{P} \mathbb{I}(r_* > \sum_{j=1}^{q} R_j)$ and $r = r_* - \sum_{q=1}^{p} R_q$. Using this single superscript notation we obtain

$$\begin{aligned} &-\frac{1}{2}\sum_{t=P+1}^{T} \left\|\Sigma^{-\frac{1}{2}}\left(y_t - L_t\hat{\lambda}_t\right)\right\|_2^2 - \frac{1}{2}\sum_{t=2}^{T}\left\|\Sigma^{-\frac{1}{2}}L_t\hat{V}_t^{\frac{1}{2}}\right\|_2^2 \\ &\qquad = -\frac{1}{2}\sum_{t=P+1}^{T}\left\|\mathcal{Z}_t - \sum_{r_*=1}^{R}\hat{\lambda}_t^{r_*}\cdot\mathcal{X}_t^{r_*}\right\|^2 - \frac{1}{2}\sum_{t=P+1}^{T}\sum_{r_*=1}^{R}\left\|\sum_{s_*=r_*}^{R}\hat{v}_t^{s_*,r_*}\cdot\mathcal{X}_t^{s_*}\right\|^2, \end{aligned}$$

where $\hat{v}_t^{s_*,r_*}$ is the element on position (s_*, r_*) in lower triangular matrix $\hat{V}_t^{\frac{1}{2}}$,

$$\mathcal{Z}_t = \mathcal{Y}_t \times_1 \Sigma_1^{-\frac{1}{2}} \cdots \times_n \Sigma_n^{-\frac{1}{2}},$$

and

$$\mathcal{X}_t^{r_*} = \mathcal{Y}_{t-p} \times_1 \Sigma_1^{-\frac{1}{2}} U_1^{p,r} \cdots \times_n \Sigma_n^{-\frac{1}{2}} U_n^{p,r}.$$

When we then matricize along mode k, see Appendix A, we get the following relevant part of the objective function (11) with regard to mode-k parameters $u_k^{r_*}$, $u_{k+n}^{r_*}$, and Σ_k, where r_* is linked to (p,r) in the above mentioned way,

$$-\frac{T-P}{2}\prod_{l=1}^{n}J_l\log(2\pi)-\frac{T-P}{2}J_{-k}\log\det(\Sigma_k)$$
$$-\frac{1}{2}\sum_{t=P+1}^{T}\left\|\Sigma_k^{-\frac{1}{2}}\left(Z_t^{(k)}-\sum_{r_*=1}^{R}\hat{\lambda}_t^{r_*}U_k^{r_*}X_t^{(k,r_*)}\right)\right\|_F^2$$
$$-\frac{1}{2}\sum_{t=P+1}^{T}\sum_{r_*=1}^{R}\left\|\sum_{s_*=r_*}^{R}\hat{v}_t^{s_*,r_*}\Sigma_k^{-\frac{1}{2}}U_k^{s_*}X_t^{(k,s_*)}\right\|_F^2, \quad (15)$$

where

$$Z_t^{(k)}=\operatorname{mat}_k\left(\mathcal{Y}_t\times_1\Sigma_1^{-\frac{1}{2}}\cdots\times_{k-1}\Sigma_{k-1}^{-\frac{1}{2}}\times_k I\times_{k+1}\Sigma_{k+1}^{-\frac{1}{2}}\cdots\times_n\Sigma_n^{-\frac{1}{2}}\right)$$

and

$$X_t^{(k,r_*)}=\operatorname{mat}_k\left(\mathcal{Y}_{t-p}\times_1\Sigma_1^{-\frac{1}{2}}U_1^{p,r}\cdots\times_{k-1}\Sigma_{k-1}^{-\frac{1}{2}}U_{k-1}^{p,r}\times_k I\times_{k+1}\Sigma_{k+1}^{-\frac{1}{2}}U_{k+1}^{p,r}\cdots\times_n\Sigma_n^{-\frac{1}{2}}U_n^{p,r}\right).$$

For each mode k optimizing sequentially (15) w.r.t. $u_k^{r_*}$, $u_{k+n}^{r_*}$, and Σ_k we obtain

$$u_k^{r_*}\leftarrow\left(u_{k+n}^{r_*\prime}\Sigma_k^{-1}u_{k+n}^{r_*}\right)^{-1}\left(\sum_{t=P+1}^{T}\left((\hat{\lambda}_t^{r_*})^2+\sum_{s_*=1}^{r_*}(\hat{v}_t^{r_*,s_*})^2\right)X_t^{(k,r_*)}X_t^{(k,r_*)\prime}\right)^{-1}$$
$$\cdot\sum_{t=P+1}^{T}X_{t-1}^{(k,r_*)}Z_t^{(k,r_*)\prime}\Sigma_k^{-1}u_{k+n}^{r_*}, \quad (16)$$

$$u_{k+n}^{r_*}\leftarrow\left(\sum_{t=P+1}^{T}\left((\hat{\lambda}_t^{r_*})^2+\sum_{s_*=1}^{r_*}(\hat{v}_t^{r_*,s_*})^2\right)u_k^{r_*\prime}X_t^{(k,r_*)}X_t^{(k,r_*)\prime}u_k^{r_*}\right)^{-1}\sum_{t=P+1}^{T}Z_t^{(k,r_*)}X_t^{(k,r_*)\prime}u_k^{r_*}, \quad (17)$$

and

$$\Sigma_k\leftarrow\frac{1}{(T-P-1)J_{-k}}\sum_{t=P+1}^{T}\left(E_t^{(k)}E_t^{(k)\prime}\right.$$
$$\left.+\sum_{r_*=1}^{R}\sum_{s_*=r_*}^{R}\sum_{j_*=r_*}^{R}\hat{v}_t^{s_*,r_*}\hat{v}_t^{j_*,r_*}U_k^{s_*}X_t^{(k,s_*)}X_t^{(k,j_*)\prime}U_k^{j_*\prime}\right), \quad (18)$$

with $E_t^{(k)}=Z_t^{(k)}-\hat{\lambda}_tU_kX_{t-1}^{(k)}$ and

$$Z_t^{(k,r_*)}=\hat{\lambda}_t^{r_*}Z_t^{(k)}-\sum_{s_*\neq r_*}\left(\hat{\lambda}_t^{r_*}\hat{\lambda}_t^{s_*}+\sum_{j_*=1}^{\min\{s_*,r_*\}}\hat{v}_t^{r_*,j_*}\hat{v}_t^{s_*,j_*}U_k^{s_*}X_t^{(k,s_*)}\right)$$

After each update of (16) and (17) we normalize the vectors $u_k^{r_*}$ and $u_{k+n}^{r_*}$ to have unit length.

As we have an identification problem regarding the scale of each individual Σ_k, which is apparent from $\Sigma = \Sigma_1 \otimes \cdots \otimes \Sigma_n$, we follow Chen et al. (2021a) and Li and Xiao (2021) and normalize each Σ_k for $k \neq n$ after each update step such that $\|\Sigma_k\|_F = 1$.

The complete algorithm to estimate all parameters is summarized in Algorithm 1[2].

Algorithm 1: Alternating least squares ECM algorithm for time-varying TARs

Input: Initial estimates of $u_1^{r_*,(0)}, \dots, u_{2n}^{r_*,(0)}$ and $\alpha_{r_*}^{(0)}$, $\phi_{r_*}^{(0)}$, and $\sigma_{r_*}^{2,(0)}$ for all $r_* = 1, \dots, R$, and $\Sigma_1^{(0)}, \dots, \Sigma_n^{(0)}$

Iterate until convergence: For iteration m.

The E-step

- Run the Kalman filter and smoother recursions to obtain $\hat{\lambda}_t$, $\hat{V}_t$, and $\mathbb{E}[\lambda_t \lambda_{t-1} | \mathcal{F}_T]$ (based on the estimates of iteration $m-1$).

The CM-step

- Update $\phi_{r_*}^{(m)}$ according to (12).
- Update $\alpha_{r_*}^{(m)}$ according to (13).
- Update $\sigma_{r_*}^{2,(m)}$ according to (14).
- For each mode k
  - Update $u_k^{r_*,(m)}$ according to (16) and normalize such that $\|u_k^{r_*,(m)}\|_2 = 1$.
  - Update $u_{k+n}^{r_*,(m)}$ according to (17) and normalize such that $\|u_{k+n}^{r_*,(m)}\|_2 = 1$.
  - Update $\Sigma_k^{r_*,(m)}$ according to (18) and normalize such that $\|\Sigma_k^{r_*,(m)}\|_F = 1$ when $k \neq n$.

3.1 ‖ Initialization

As is highlighted in Algorithm 1 we need a proper initialization of the parameters. We propose the following initialization scheme for the parameters. First, we estimate a full autoregressive coefficient matrix using a ridge regression on the vectorized data, where a ridge regularization term is needed due to the high-dimensionality of the data, i.e.

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=P+1}^{T} \|y_t - \beta y_{t-1:t-P}\|_2^2 + \tau \|\beta\|_2^2,$$

where $y_{t-1:t-P} = (y_{t-1}, \dots, y_{t-p})$. The regularization parameter τ is chosen using an exhaustive grid search.

With this estimated autoregressive coefficient matrix at hand we initialize the u's using a static rank R_p CP decomposition for each lag p, see Kolda (2006), Bader and Kolda (2006), and

[2] Julia code for working with tensor autoregressions, both static and dynamic, such as fitting of the model to the data, IRF analysis, and forecasting can be found on `https://github.com/qntwrsm/TensorAutoregressions.jl`

Kolda and Bader (2009). In order to estimate the parameters governing the transition dynamics of the dynamic factors we first estimate the λ_t's for each time point separately using a regression

$$\hat{\lambda}_t = \arg\min_{\lambda_t} \|y_t - L_t \lambda_t\|_2^2.$$

Based on these estimated factors the autoregressive dependence parameter $\phi_{p,r}$'s, intercepts $\alpha_{p,r}$'s, and variances $\sigma^2_{p,r}$'s are estimated using a simple linear regression. Finally, the covariance matrices are initialized by their sample equivalents for each specific mode.

3.2 ‖ Log-likelihood function

The goodness-of-fit of the model can be evaluated using the log-likelihood and likelihood related goodness-of-fit measures, i.e. information criteria. As the model can be formulated as a restricted time-varying dynamic factor model we can cast the model into state space framework and, hence, the log-likelihood is readily available by a routine application of the Kalman filter (see Durbin and Koopman, 2012). The log-likelihood can also form the basis for an information criteria based selection procedure for the number of autoregressive lags and/or the rank of the CP decomposition.

4 ‖ Simulation Study

To verify the small sample performance of our estimator as outlined in Algorithm 1 we conduct two Monte Carlo studies. One study focuses on a small dimensional setting that mimics the setting of the empirical application in the next section where we model inflation in a three variable macroeconomic system considering the five biggest economies of the eurozone split-up over three sectors. In this setting the number of time series is equal to 45, which would imply that a standard panel VAR(1) modeling approach would yield 45^2 autoregressive coefficients. The second setting we consider is a high-dimensional setting in which we expand the system to being a $10 \times 3 \times 3$ tensor. In both setups we include 2 lags, where the first lag is of rank 2 and the second lag of rank 1, i.e. TAR(2,2,1) model. In the small-dimensional setup we have in 66 autoregressive CP decomposition parameters, 27 variances and covariances, and 9 parameters controlling the latent factor dependence, leading to a total of 112 parameters. In the large-dimensional setup we have a total of 208 parameters. It is immediately obvious how modeling the tensor structure directly leads to an enormous reduction in the number of parameters.

In both settings we set the true model parameters in a similar way. The factor autoregressive coefficients $\phi_{p,r}$ are generated from $\mathcal{U}(0.4, 0.8)$, the intercepts $\alpha_{p,r}$ are set to zero, and the variances are drawn from $\mathcal{U}(0, 1)$ in both settings. Each $u_j^{p,r}$ is generated as a $\mathcal{N}(0, I)$ random vector, which is normalized to have unit length. Similarly for each mode k the free entries of

Table I

Simulation results static parameters of CP decomposition

	$U_k^{1,1}$			$U_k^{1,2}$			$U_k^{2,1}$		
	mode 1	mode 2	mode 3	mode 1	mode 2	mode 3	mode 1	mode 2	mode 3
					small setup				
$T = 250$	0.0519	0.0226	0.0281	0.2214	0.1375	0.1475	0.7232	0.4400	0.4193
$T = 500$	0.0314	0.0141	0.0180	0.1515	0.0880	0.0935	0.4819	0.2766	0.2508
$T = 1000$	0.0197	0.0090	0.0120	0.1004	0.0579	0.0618	0.3203	0.1807	0.1567
					large setup				
$T = 250$	0.5013	0.1266	0.1024	0.2645	0.0809	0.0876	0.4043	0.1895	0.1880
$T = 500$	0.3376	0.0790	0.0646	0.1720	0.0527	0.0552	0.2862	0.1245	0.1268
$T = 1000$	0.2341	0.0538	0.0443	0.1163	0.0352	0.0367	0.1996	0.0868	0.0874

Table II

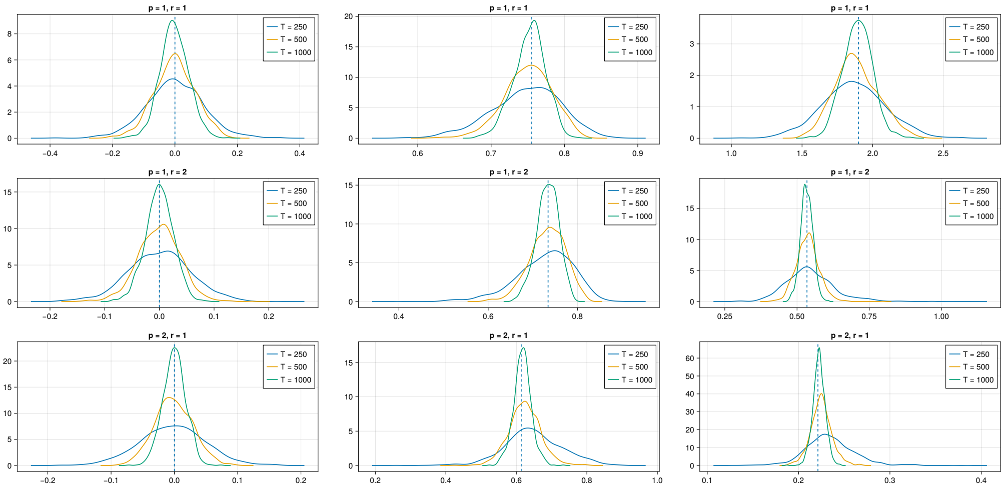
Simulation results covariance

	Σ_k					
	small setup			large setup		
	mode 1	mode 2	mode 3	mode 1	mode 2	mode 3
$T = 250$	0.0382	0.0297	0.0892	0.0618	0.0173	0.0545
$T = 500$	0.0268	0.0210	0.0628	0.0436	0.0120	0.0386
$T = 1000$	0.0191	0.0144	0.0449	0.0308	0.0084	0.0262

the covariance matrix operating on that mode, Σ_k, are generated from independent $\mathcal{U}(0, 1)$. To overcome the scaling identification issue we normalize all these individual covariance matrices to have a Frobenius norm equal to one.

The Monte Carlo study then samples a 1000 simulated tensor time series according to the true model for different sample sizes of 250, 500, and 1000 time series observations. For each simulation we estimate the time varying TAR model, where we initialize the parameters in the way outlined in the previous section.

The results for the static parameters of the model, that is the CP decomposition components $u_j^{p,r}$ and covariance matrices Σ_k, are shown in Table I and Table II. Tables I and II show the average Frobenius norm of the difference between the true parameters of the model and the estimated ones over the simulations. For the CP decomposition components we show the Frobenius norm of the $U_k^{p,r} = u_{k+n}^{p,r} \circ u_k^{p,r}$ for each mode k to have a more concise representation of the results. We observe that the small sample performance of the estimator is very good in general and that the Frobenius norm goes down to zero relatively fast when the sample size



(A) $\alpha_{p,r}$ (B) $\phi_{p,r}$ (C) $\sigma^2_{p,r}$

FIGURE I

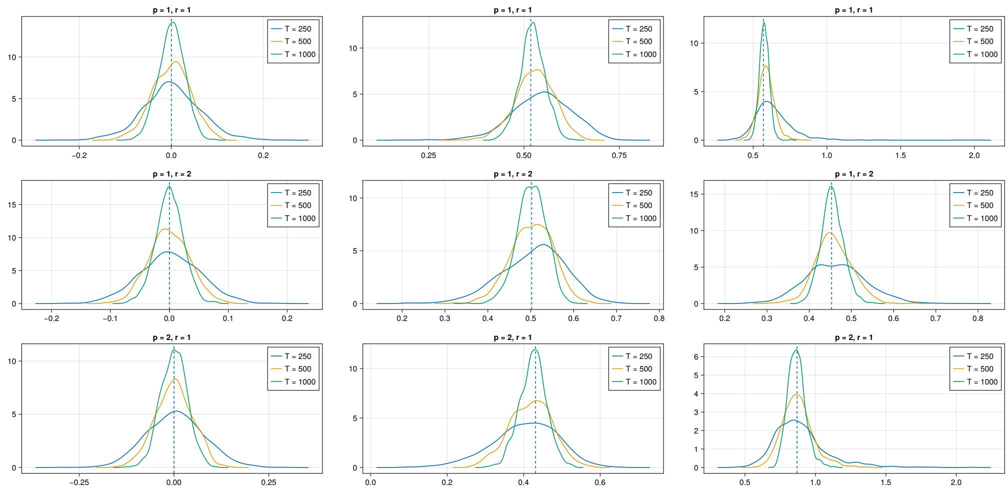
SIMULATION RESULTS FOR FACTOR DYNAMICS IN SMALL SETUP

increases. It turns out that the CP decomposition components are a bit more difficult to estimate compared to the covariance matrices.

For the parameters governing the dynamics of the dynamic factors, i.e. $\alpha_{p,r}$, $\phi_{p,r}$, and $\sigma^2_{p,r}$ we visualize the Monte Carlo simulation study through the means of kernel density plots as shown in Figures I and Figure II. From Figures I and II it is apparent that as the sample grows all three parameters governing the time dynamics of the factors converge to its true value irrespective of the setup considered. For smaller sample sizes a slight left-skewed distribution is observed for the autoregressive persistence parameter $\phi_{p,r}$, which is quite common for estimators of the autoregressive parameters when the true parameter is close to unity. Across the board we observe that when the sample size increases the distribution of the estimators seems to converge to a normal distribution centered around the true value. Irrespective of the setup the simulation study seems to indicate that the dynamic process of the factors governing the time-variation in the second lag of the model are more challenging to identify accuretly, which can be seen from the spread in the density plots of the last row in Figures I and II. Overall Figures I and II seem to indicate good small sample performance of our approach even for large-dimensional tensors.

5 || EMPIRICAL APPLICATION: SECTORAL INFLATION

Given the current high inflation period and the very specific origin of these inflationary shocks, namely the energy supply side shocks, we try to model inflation at a disaggregated level, namely that of different sectors of the economy in eurozone member states. Using this approach we are able to find heterogenous inflation dynamics for different sectors of the economy, as well as,



(A) $\alpha_{p,r}$ (B) $\phi_{p,r}$ (C) $\sigma^2_{p,r}$

Figure II

Simulation results for factor dynamics in large setup

investigating their impact on output. Seeing the recent discussion of a flattening of the Phillips curve (see Vlekke et al., 2020; Del Negro et al., 2020; Hazell et al., 2022) and the heterogeneous results across countries (Vlekke et al., 2020), it might be warranted to look at local sectoral level Phillips curves. While the wage Phillips curve has traditionally been stronger, recent evidence also shows a declining relationship there (Galí and Gambetti, 2019).

5.1 ‖ Sectoral data

In order to model inflation at the sectoral level and be able to produce local Phillips curve estimates we study the five biggest economies of the eurozone, i.e. Germany, France, Italy, Spain, and the Netherlands. For these countries we obtain data for three different sectors, as classified by NACE, of the economy, namely industry, manufacturing, and construction. These sectors are selected due to their data availability, unfortunately data on the services sector was unavailable for most eurozone member states. However, the industry sector comprises the energy sector which is the driving factor behind the strong inflation of the previous year. For these three sectors we collect three macroeconomic variables, namely gross value added, employment levels, and PPI. All data was obtained from Eurostat and is seasonally and calendar adjusted using the TRAMO-SEATS method. The sample spans the period from 2000-Q1–2022-Q4 for a total of 92 time series observations and covers the covid pandemic, as well as, the invasion of Ukraine. In order to assure stationarity we transform all variables into growth rates.

The data is visualized in Figure III. An interesting fact is the heterogeneous impact of the covid pandemic on the different countries, we see a much stronger response in e.g. output for the

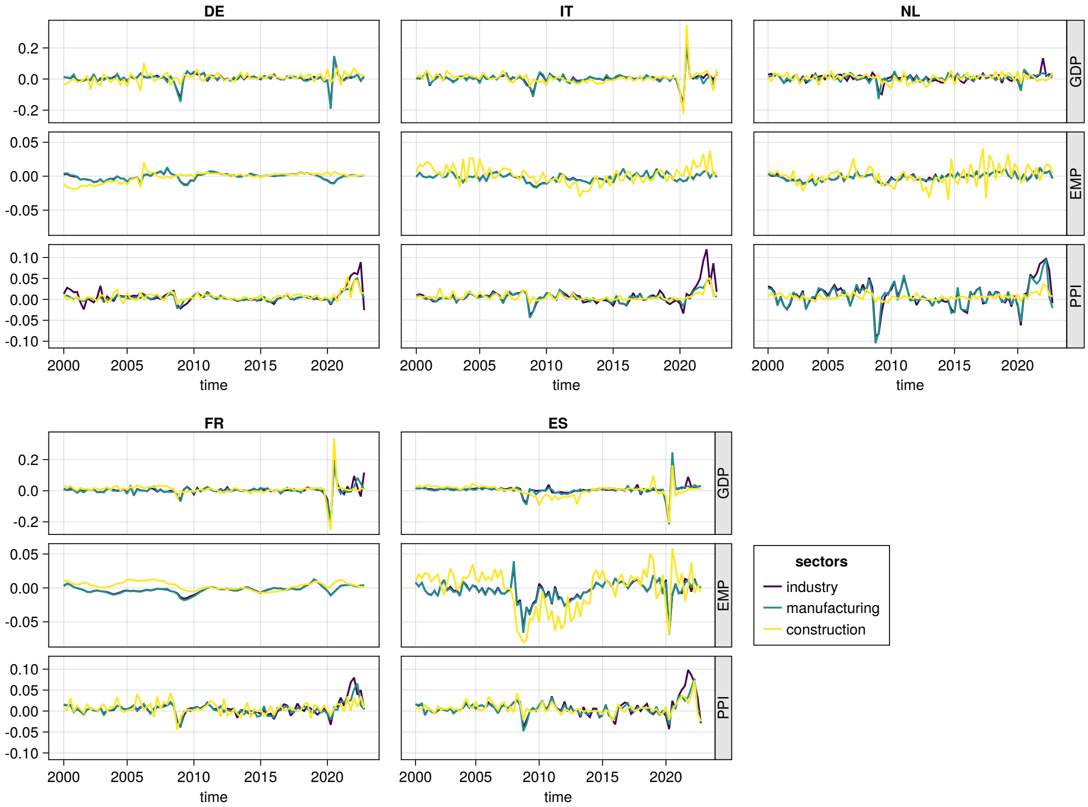


Figure III

Growth rates of output, employment and PPI for all countries and sectors southern European countries. On the other hand, employment dynamics are much less volatile in Germany and France compared to the other countries. Regarding inflation dynamics, it seems that there is heterogeneity across the sectors for all countries, especially since 2020. Construction is the sector which had the least impact of the covid pandemic and the current energy crisis. Observe also that the inflation dynamics since 2020 have been different across countries. These observations support our approach of modeling sectoral inflation.

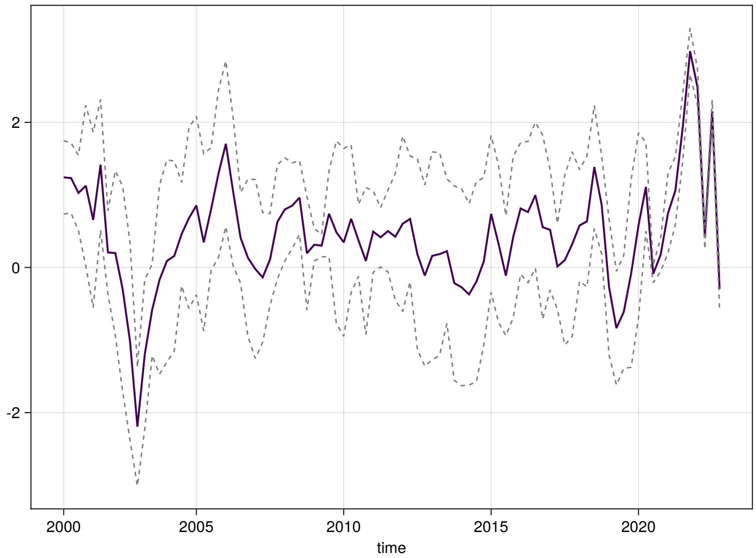


Figure IV

Smoother estimate of λ_t

5.2 || Estimation results

We fit our time-varying TAR(1,1) model to the $(5 \times 3 \times 3)$-dimensional tensor system with 91 time series observations, where we initialize the parameters of the model in a data-driven manner as discussed in Section 3. The dynamic scale of the autoregressive coefficient tensor, λ_t, is estimated to be only weakly persistent with a $\hat{\phi}$ of roughly 0.6. The smoothed estimate of λ_t is shown in Figure IV. We observe at the end of the sample, a strong upward trend in the dynamic scale of the time-varying autoregressive coefficient tensor, with clear spikes. Note that in both spikes λ_t is larger than unity, which implies that the system was locally explosive. However, as the autoregressive parameter of the factor is inside the unit circle the system as a whole is itself stationary. With this locally explosive behavior at the end of the sample the model is able to capture the sudden rise in inflation and strong reactions of output. The weak persistence of the dynamic scale and the close to zero smoothed dynamic scale throughout the most part of the sample is in line with the general findings in the literature that the persistence of inflation declined from the 1990s onwards, see Cogley et al. (2010) and Chan et al. (2013).

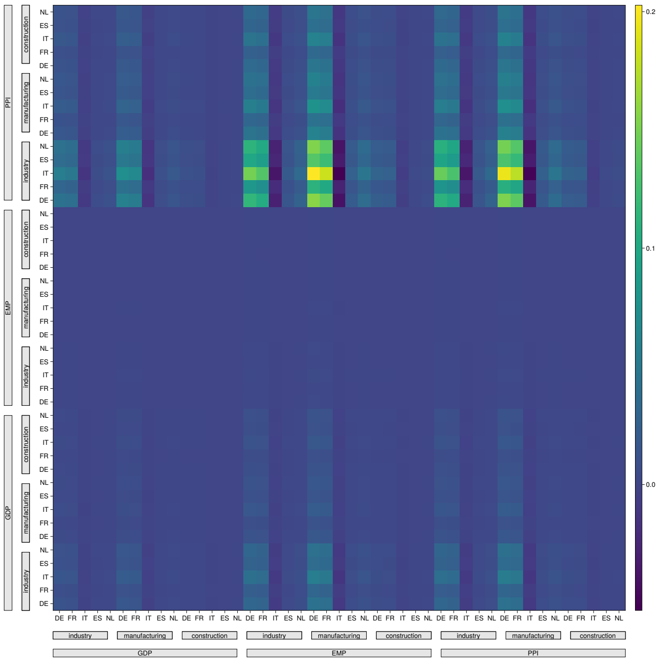


FIGURE V

MATRICIZED STATIC COMPONENT OF THE AUTOREGRESSIVE COEFFICIENT TENSOR, $\hat{U}^{(4,5,6)}$

Figure V shows the estimated matricized static component of the autoregressive coefficient tensor $\mathcal{A}_t$, i.e. $U^{(4,5,6)}$. From the heatmap it is clear that employment is not affected by the other variables in the system, as seen from the entire rows showing values close to zero. However, we do observe responses of the inflation and output to lagged employment, where the former shows the strongest responses. An interesting fact is that Italy seems to have a negative relationship to the other countries, apparent from the darker blue columns in Figure V, but also shows the strongest positive relationship for the industry sector when it comes to inflation. In general we observe that for inflation in the industry sector across the different countries the relationship is strongest to the two biggest economies of the eurozone, Germany and France.

An interesting observation from Figure V is that the dependence of output on its lagged values is muted across all countries and sectors, as can be seen from the lower left corner of the heatmap. This can also be seen from Figure VI which shows the estimated covariance matrix $\hat{\Sigma} = \hat{\Sigma}_3 \otimes \hat{\Sigma}_2 \otimes \hat{\Sigma}_1$. From Figure VI we observe the strongest idiosyncratic variance components for the output corner of the heatmap. Especially for the construction sector of Spain and the Netherlands we observe strong unexplained variation.

5.3 || LOCAL PHILLIPS CURVE

In order to assess the heterogeneity across Phillips curves estimated at the sectoral level we turn to the IRFs, where we investigate the response of inflation to a shock in employment. In section

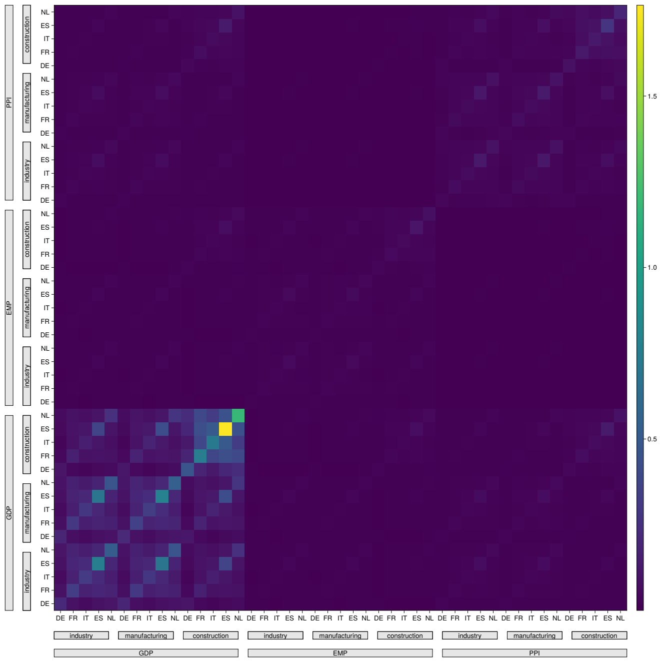


Figure VI

Estimated covariance matrix $\hat{\Sigma}$

2 we outlined how the IRF of our time-varying TAR model can be decomposed into two parts. One that is related to the initial impact of the shock and one related to the decay over the considered horizon. The decay component of the time-varying IRF is shown in Figure VII, while the initial impact component is shown in Figure V. Note that for the orthogonalized IRF analysis the initial impact shown in Figure V needs to be multiplied by the Cholesky decomposition of the covariance matrix shown in Figure VI.

We observe from Figure VII a fast decay to zero as the horizon grows, which is expected from a one-lag stationary autoregressive model. Interestingly, we observe strong heterogeneity over time when it comes to the initial level of the decay. In the middle of the sample we observe calm and muted initial decay responses, even during the financial crisis period, which might come as a surprise. The initial decay component of the IRF is more extreme in the beginning and end of the sample. We observe a strong negative response in the beginning of the sample in line with what we saw for the smoothed factors in Figure IV. While, at the end of the sample we see more volatile decay responses, with a peak for the covid pandemic and the following quarters. This indicates that the relationship between inflation and employment was more flat in the middle of the sample, but became more steep near the end of the sample and was even inverted in the early 2000s.

When focusing on the initial impact response as highlighted in Figure V, we find evidence for strong sectoral and country specific heterogeneity when it comes to the response of inflation

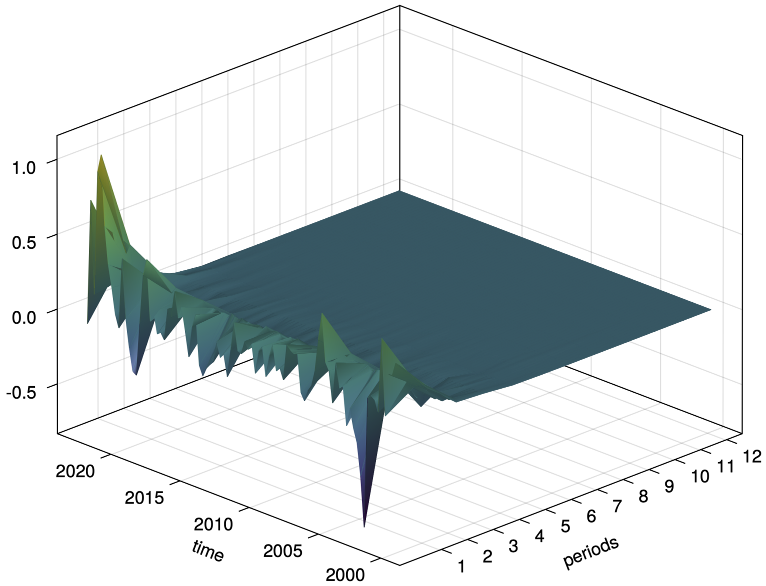


Figure VII

Decay component of the time-varying IRFs

on a shock in employment, as can be seen from the upper middle part of the heatmap. Inflation in some sectors, such as manufacturing and construction, shows a more flat response to an employment shock. This is inline with the general tendency to find a flat Phillips curve, see Del Negro et al. (2020) and Hazell et al. (2022). On the other hand, we observe a stronger response for the industry sector. Moreover, in line with Vlekke et al. (2020) we observe a large heterogeneity across countries for the strength of the employment-inflation link.

All in all, these results indicate heterogeneity of the relation between inflation and employment over time, sector, and country in the eurozone. This advocates that the Phillips curve should be studied locally and dynamically.

5.4 || Energy supply shock

Finally, we turn to the local effects of the energy supply shock due to the invasion of Ukraine by Russia in the first quarter of 2022. In order to analyze the impact of the war on the macroeconomy we compare the orthogonalized IRFs of our time-varying TAR model at the quarter before the invasion, 2021-Q4, and the quarter of the start of the war, 2022-Q1. The shock we analyze with our orthogonalized IRFs is a shock to inflation in the industry sector of Germany, as Germany had the largest energy exposure to Russia of all the countries considered and the energy sector is part of the industry sector. This analysis allows us to investigate how the different countries and sectors of the economy are reacting to the energy supply shock. As the ordering of the time series along each mode matter for the Cholesky decomposition, used to identify structural shocks, we ordered the countries in descending order of size of the economy, the sectors as industry, manufacturing, and construction, and, finally, the variables as output, employment, and inflation.

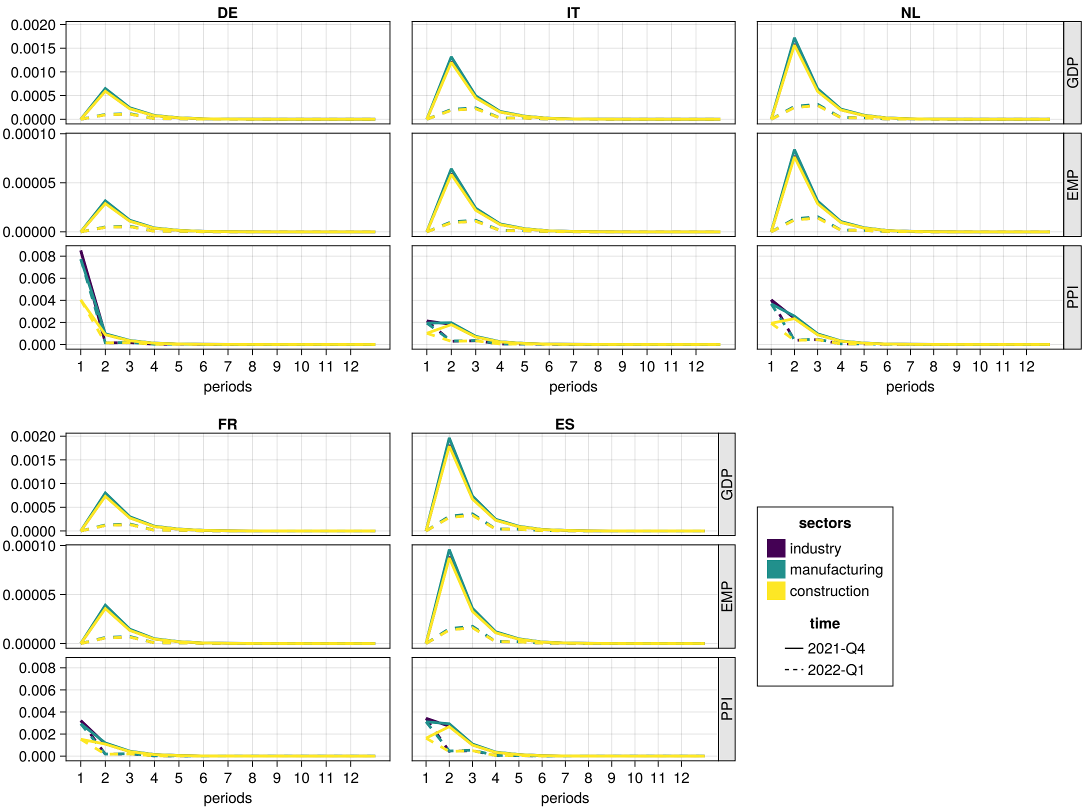


Figure VIII

Orthogonalized IRFs of a shock to inflation in the industry sector of Germany

The IRFs are reported in Figure VIII. We observe that for all sectors and countries the reaction of output is more muted in the quarter of the war compared to the quarter before. We observe very homogeneous responses across the sectors for output and employment. This could be related to the similarity of the considered sectors. The propagation of the inflationary shock does show some heterogeneous responses across sectors, where the manufacturing sector seems to be least affected. This provides some support for potentially modeling inflation at the sectoral level as sectors react differently to the shock.

6 || Conclusion

In this paper we have introduced a novel dynamic autoregressive model to handle tensor-valued time series data efficiently and parsimoniously. We propose the time-varying TAR model that models directly the tensor representation of the data using a time-varying CP decomposition with a dynamic factor structure. The time-varying TAR model can be cast into the restricted time-varying VAR and dynamic factor model frameworks. We derive stationarity conditions as well as impulse response functions for our model. Estimation of the parameters of the TAR model is done efficiently by a tailor-made algorithm based on EM and alternating least squares. In the Monte Carlo study we show the good small sample performance of our estimator. In our application we apply our method to model sectoral inflation in the eurozone for the five biggest economies across a variety of sectors from 2000–2022. We find evidence for heterogeneous local Phillips curves, where some sectors of the economy show a more flat curve than others. Moreover, we also find substantial time-variation in the Phillips curve, with stronger inflationary responses to employment at the end of the sample. We also investigate the impact of the energy supply shock arising from the invasion of Ukraine by Russia in 2022 on the economy and find in general dampened responses compared to before the start of the war.

References

Bader, B. W. and T. G. Kolda (2006): "Algorithm 862: MATLAB Tensor Classes for Fast Algorithm Prototyping", *ACM Trans. Math. Softw.*, 32(4), 635–653.

Billio, M., R. Casarin, M. Iacopini, and S. Kaufmann (2023): "Bayesian Dynamic Tensor Regression", *Journal of Business & Economic Statistics*, 41(2), 429–439.

Bougerol, P. and N. Picard (1992): "Strict Stationarity of Generalized Autoregressive Processes", *The Annals of Probability*, 20(4), 1714–1730.

Canova, F. and M. Ciccarelli (2004): "Forecasting and turning point predictions in a Bayesian panel VAR model", *Journal of Econometrics*, 120(2), 327–359.

— (2009): "Estimating Multicountry VAR Models", *International Economic Review*, 50(3), 929–959.

Chan, J. C. C. (2017): "The Stochastic Volatility in Mean Model With Time-Varying Parameters: An Application to Inflation Modeling", *Journal of Business & Economic Statistics*, 35(1), 17–28.

Chan, J. C. C., G. Koop, and S. M. Potter (2013): "A New Model of Trend Inflation", *Journal of Business & Economic Statistics*, 31(1), 94–106.

Chen, R., H. Xiao, and D. Yang (2021a): "Autoregressive models for matrix-valued time series", *Journal of Econometrics*, 222(1, Part B), Annals Issue:Financial Econometrics in the Age of the Digital Economy, 539–560.

Chen, R., D. Yang, and C.-H. Zhang (2021b): "Factor Models for High-Dimensional Tensor Time Series", *Journal of the American Statistical Association*, 0(0), 1–23.

Cogley, T., G. E. Primiceri, and T. J. Sargent (2010): "Inflation-Gap Persistence in the US", *American Economic Journal: Macroeconomics*, 2(1), 43–69.

Cogley, T. and T. J. Sargent (2001): "Evolving Post-World War II U.S. Inflation Dynamics", *NBER Macroeconomics Annual*, 16, 331–373.

— (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US", *Review of Economic Dynamics*, 8(2), Monetary Policy and Learning, 262–302.

Cressie, N. (2015): *Statistics for spatial data*, John Wiley & Sons.

Del Negro, M., M. Lenza, G. E. Primiceri, and A. Tambalotti (2020): "What's up with the Phillips Curve?", Working Paper 27003, National Bureau of Economic Research.

Durbin, J. and S. J. Koopman (2012): *Time series analysis by state space methods*, Oxford university press.

Faust, J. and J. H. Wright (2013): "Chapter 1 - Forecasting Inflation", in: *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, vol. 2, Handbook of Economic Forecasting, Elsevier, 2–56.

Galí, J. and L. Gambetti (2019): "Has the U.S. Wage Phillips Curve Flattened? A Semi-Structural Exploration", Working Paper 25476, National Bureau of Economic Research.

Gorgi, P., S. J. Koopman, and J. Schaumburg (2018): "Time-varying VAR Models With Structural Dynamic Factors", mimeo.

Hazell, J., J. Herreño, E. Nakamura, and J. Steinsson (2022): "The Slope of the Phillips Curve: Evidence from U.S. States*", *The Quarterly Journal of Economics*, 137(3), 1299–1344.

Hoff, P. D. (2011): "Separable covariance arrays via the Tucker product, with applications to multivariate relational data", *Bayesian Analysis*, 6(2), 179–196.

— (2015): "Multilinear tensor regression for longitudinal relational data", *The Annals of Applied Statistics*, 9(3), 1169–1193.

Kolda, T. G. and B. W. Bader (2009): "Tensor Decompositions and Applications", *SIAM Review*, 51(3), 455–500.

Kolda, T. G. (2006): "Multilinear operators for higher-order decompositions."

Koop, G., M. Pesaran, and S. M. Potter (1996): "Impulse response analysis in nonlinear multivariate models", *Journal of Econometrics*, 74(1), 119–147.

Li, Z. and H. Xiao (2021): "Multi-linear tensor autoregressive models", mimeo.

Lütkepohl, H. (2005): *New introduction to multiple time series analysis*, Springer Science & Business Media.

Meng, X.-L. and D. B. Rubin (1993): "Maximum likelihood estimation via the ECM algorithm: A general framework", *Biometrika*, 80(2), 267–278.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy", *The Review of Economic Studies*, 72(3), 821–852.

Shumway, R. H. and D. S. Stoffer (1982): "An Approach To Time Series Smoothing And Forecasting Using The Em Algorithm", *Journal of Time Series Analysis*, 3(4), 253–264.

Stella, A. and J. H. Stock (2013): "A state-dependent model for inflation forecasting", Discussion Paper 1062, FRB.

Stock, J. H. and M. W. Watson (2008): "Phillips Curve Inflation Forecasts", Working Paper 14322, National Bureau of Economic Research.

— (2010): "Modeling Inflation After the Crisis", Working Paper 16488, National Bureau of Economic Research.

Vlekke, M., S. J. Koopman, and M. C. Mellens (2020): "An assessment of the Phillips curve over time: evidence for the United States and the euro area", Discussion Paper, CPB Netherlands Bureau for Economic Policy Analysis.

Wang, D., Y. Zheng, and G. Li (2021): "High-Dimensional Low-Rank Tensor Autoregressive Time Series Modeling", mimeo.

Appendix

A || Tensor Notation

In this appendix we will discuss some tensor notation and algebra concepts related to our TAR model. For a general introduction and more in-depth treatment of the concepts that will be outlined below see e.g. Kolda (2006), Bader and Kolda (2006), and Kolda and Bader (2009).

We will denote tensors, a multi-dimensional array, in a calligraphic font, e.g. a n-th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_n}$ with entries $\mathcal{X}(i_1, \ldots, i_n)$. The order of a tensor indicates the number of dimensions, also called modes. Analogous to the concept of rows and columns for matrices we define fibers for a tensor, the mode-k fiber is the vector obtained by fixing all but the k-th mode.

A tensor can be unfolded into a matrix, this process is called matricization. Let $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_n}$. The mode-$k$ matricization is denoted by $\text{mat}_k(\mathcal{X}) = X^{(k)} \in \mathbb{R}^{I_k \times I_{-k}}$ where $I_{-k} = \prod_{j \neq k} I_j$. In other words the matricization operator maps the indices $(i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_n)$ to a single index j that indexes the columns of the resulting matrix. The exact mapping of the indices can be found in e.g. Kolda (2006), Bader and Kolda (2006), and Kolda and Bader (2009). More generally we could also consider matricization of a tensor on a subset of modes, i.e. $\text{mat}_{\mathcal{R}}(\mathcal{X}) = X^{(\mathcal{R})} \in \mathbb{R}^{\prod_{j \in \mathcal{R}} I_j \times \prod_{j \in \mathcal{C}} I_j}$ where $\mathcal{R}$ denotes the subset of modes that will be mapped to the rows of the resulting matrix and similarly $\mathcal{C}$ denotes the subset of modes that will be mapped to the columns of $X^{(\mathcal{R})}$, with $\mathcal{R}$ and $\mathcal{C}$ partitioning the set of modes $\{1, \ldots, n\}$. This operation is a similar to vectorizing a matrix, moreover vectorization of a tensor is a special case of matricization, namely $\text{vec}(\mathcal{X}) = \text{mat}_{\{1,\ldots,n\}}(\mathcal{X})$.

Next, we introduce tensor multiplications. Assume that we have two tensors $\mathcal{X} \in \mathbb{R}^{J_1 \times \cdots \times J_m \times I_1 \times \cdots \times I_n}$ and $\mathcal{Y} \in \mathbb{R}^{J_1 \times \cdots \times J_m \times K_1 \times \cdots \times K_p}$, then the (contracted) tensor product is given by

$$\langle \mathcal{X}, \mathcal{Y} \rangle_{i_1,\ldots,i_n,k_1,\ldots,k_p} = \sum_{j_1=1}^{J_1} \cdots \sum_{j_m=1}^{J_m} \mathcal{X}(j_1, \ldots, j_m, i_1, \ldots, i_n) \cdot \mathcal{Y}(j_1, \ldots, j_m, k_1, \ldots, k_p). \tag{A.1}$$

When we have a n-th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_n}$ and a matrix $U \in R^{J_k \times I_k}$, the mode-k product is given by

$$(\mathcal{X} \times_k U)_{i_1,\ldots,i_{k-1},j_k,i_{k+1},\ldots,i_n} = \sum_{i_k=1}^{I_k} \mathcal{X}(i_1, \ldots, i_n) U(j_k, i_k). \tag{A.2}$$

Moreover, when $V \in \mathbb{R}^{L_k \times J_k}$, the repeated mode-$k$ product is given by

$$\mathcal{X} \times_k U \times_k V = \mathcal{X} \times_k VU.$$

The norm of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_n}$ is given by

$$\|\mathcal{X}\|^2 = \sum_{i_1=1}^{I_1} \cdots \sum_{i_n=1}^{I_n} \mathcal{X}^2(i_1, \ldots, i_n).$$

The norm of a tensor and the respective norms of the matricized and vectorized version of the tensor are equivalent, i.e. when e.g. considering mode-k matricization

$$\|\mathcal{X}\| = \|X^{(k)}\|_F = \|\text{vec}(\mathcal{X})\|_2.$$

Finally, similarly to matrix decompositions we can decompose a tensor into components of a lower dimensional space. The two most common decompositions are Tucker and CP, which can be thought of as higher-order generalizations of the singular value decomposition for matrices.

The Tucker decomposition of a n-th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_n}$ is given by

$$\mathcal{X} = \lambda \times_1 U_1 \cdots \times_n U_n,$$

where $U_k \in \mathbb{R}^{I_k \times J_k}$ and $\lambda \in \mathbb{R}^{J_1 \times \cdots \times J_n}$.

The CP decomposition of rank R of a n-th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_n}$ is given by

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r \cdot u_{r,1} \circ \cdots \circ u_{r,n}, \tag{A.3}$$

where $u_{r,k} \in \mathbb{R}^{I_k}$ with $\|u_{r,k}\| = 1$, λ_r is a scalar, and $\circ$ denoting the outer product.