Forecasting USA yearly energy mix: how to escape the small sample curse?

Julien Hambuckers
(with Z. Csereklyei, T. Kneib and P. Thurner)

University of Göttingen, Chair of Statistics

October 18, 2017

"Not all those who wander are lost."

- *J.R.R. Tolkien, The Fellowship of the Ring.*

"The artist's job is not to succumb to despair but to find an antidote for the emptiness (of existence)."

- *Woody Allen.*

Introduction

Energy mix: energy use (of a country), split into categories based on energy source (coal, gas, oil, nuclear, renewable).

E.g. with Germany: 25.2%, 32.05%, 21.5%, 7.6% and 13.65% (respectively)

Introduction

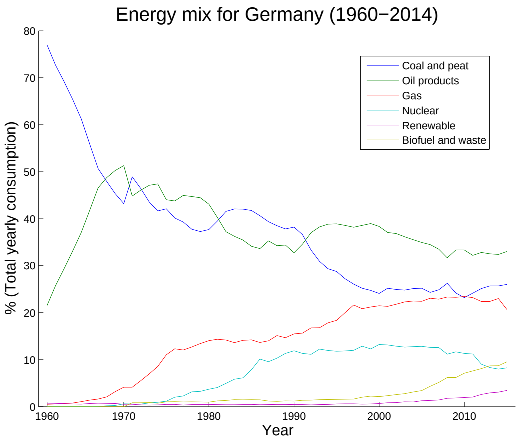

Figure: Energy mix of Germany (yearly), over the period 1960-2014, split into 6 categories.

Introduction

Energy mix: energy consumption (of a country), split into categories based on energy source (coal, gas, oil, nuclear and renewable).

E.g. with Germany: 25.2%, 32.05%, 21.5%, 7.6% and 13.65% (respectively)

Predicting the mix is relevant for energy policy:

- Allows anticipating energy transition.
- Assess the economic drivers of changes in the mix (growth, prices).
- At unchanged policy, how does the mix evolve (scenario perspective)?
- Economic approach to global warming/CO_2 emissions.

Introduction

Several factors have been found to drive changes in the mix:

- Resource availability (endowment) (Fouquet and Pearson, 2012; Burke, 2013)
- Energy demand (population, economic activity) (Gales et al., 2007; Burke, 2013).
- Income level (Burke, 2013; Csereklyei et al., 2016).
- Trade relations (Rubio and Folchi, 2012).
- Price relationships between energy type.
- Government policies!

Introduction

However, previous studies rely on "absolute" quantities, not relative ones. Quantities might vary a lot over time and lead to important forecast errors.

Also, univariate approach: essential negative correlations between proportions in the mix have not been taken into account.

Introduction

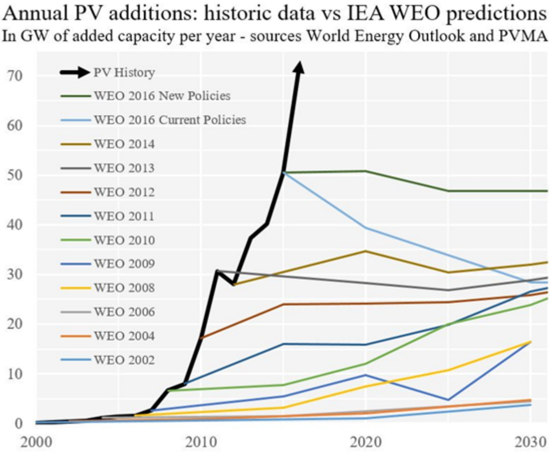

Figure: IEA "forecast" of added solar power capacity. Credit: Javier Blas.

Introduction

However, previous studies rely on "absolute" quantities, not relative ones. Quantities might vary a lot over time and lead to important forecast errors.

Also, univariate approach: essential negative correlations between proportions in the mix are not been taken into account.

Instead of hierarchically predicting the consumed amounts of each energy type, look directly at the joint distribution of the proportions.

An additional (but important) issue...

Energy mix is available at a **yearly** frequency...

We have at our disposal 45 years of yearly data...(44 points!)

We expect good (multivariate) predictions of highly unstable time series...

Multivariate model for compositional data

Energy mix has the following characteristics:

- a vector of size d, splitting proportion of energy consumption across energy types.
- Proportions sums to unity.
- ...thus the dependence between energy types is important.
- Time-dependence (AR process) and dependence with other time-varying covariates (GDP per capita, resources endowment, policy variable) are important.

$\rightarrow$ Compositional data (Aitchison, 1982; Murteira and Ramalho, 2016).

Multivariate model for compositional data

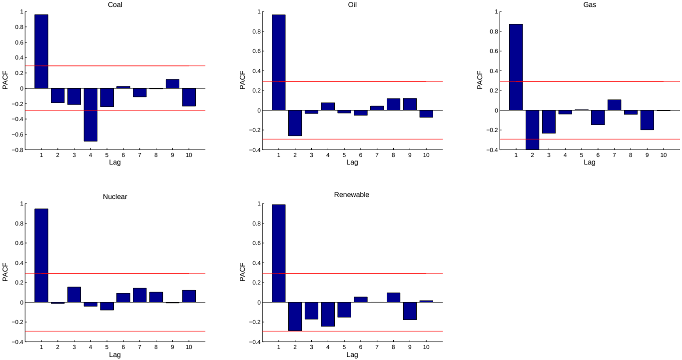

Figure: Partial auto-correlation functions for USA energy mix (1971 - 2015).

Multivariate model for compositional data

Mills (2010), Snyders et al. (2017): work on transformed data via log-ratio or hyper-spherical transformations $(\log(x_{ti}/(1-\sum_{j=1} x_{tj})))$.

Then use traditional VAR-X model.

Issue: the dependence is costly and increase exponentially with d, a dependence between energy types i and j occurs if cross AR terms are included.

Interpretation and inference are not easy tasks.

Parsimonious model for compositional data

Assume a multivariate Dirichlet regression model:

$$\begin{aligned} \mathbf{y}_t &\sim Dir(\theta_t(\mathbf{x}_t)), & (1)\\ f(\mathbf{y}_t;\theta_t) &= \prod_{j=1}^{D} y_{tj}^{(\theta_{tj}-1)} \left(\Gamma(\theta_{t0}) / \prod_{j=1}^{D} \Gamma(\theta_{tj}) \right), & (2)\\ \log(\theta_{tj}) &= \eta_{tj} = \mathbf{x}_{tj}^T \beta_j, & (3) \end{aligned}$$

$\beta = (\beta_1, \dots, \beta_D)$ being the (stacked) vector of regression parameters, $\theta_t = (\theta_{tj}, \dots, \theta_{tD}) > 0$ being the vector of distribution parameters at time t, $\theta_{t0} = \sum_{j=1}^{D} \theta_{tj}$.

Parsimonious model for compositional data

Since we are in a time series context and focus on prediction, we are interested in the distribution $\mathbf{y}_t$ conditional on past information only.

$$\eta_{tj} = (\omega_j, \mathbf{y}_{t-1}, \mathbf{x}^*_{t-1,j})^T \beta_j. \tag{4}$$

We are back in the distribution regression framework (Rigby and Stanisopoulos, 2005; Klein et al., 2015).

Conditional expectation for energy type j at time t is given by

$$\mathbb{E}(y_{tj}|\mathbf{x}_{tj}) = \theta_{tj}/\theta_{t0}. \tag{5}$$

Conditional covariance is given by

$$Cov(y_{ti}, y_{tj}) = -\theta_{ti}\theta_{tj}/(\theta_{t0}^2(\theta_{t0}+1)). \tag{6}$$

for $j \neq i$.

Parsimonious model for compositional data

- We shift from $D + DxD + \sum_{j=1}^{D} P_j$ to $\sum_{j=1}^{D} P_j + 2 * D$ parameters.
- Estimation is easily done with maximum likelihood procedures.
- Confidence and prediction intervals for $\hat{\theta}_{tj}/\hat{\theta}_{t0}$ are obtained with parametric bootstrap procedures.
- Zeros are still problematic.

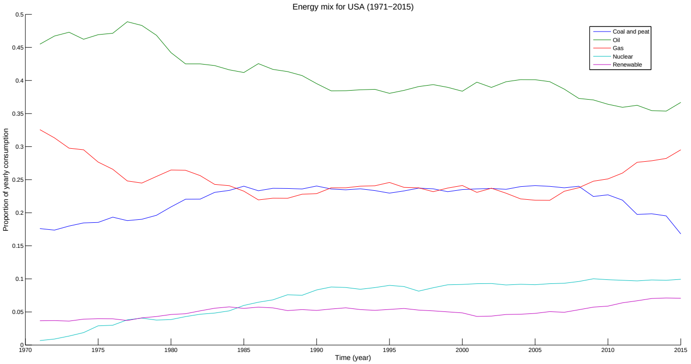

Figure: Energy mix of USA over the period (1971-2015).

Application to USA energy mix (1971 - 2015)

We assume an additive structure with

$$X_{tj} = \{\mathrm{Y_{t-1,j}}, \mathrm{Rain_{t-1}}, \mathrm{Temp_{t-1}}, \mathrm{GDPPC_{t-1}}, \acute{}\mathrm{GDP_{t-1}}, \mathrm{PRICE_{t-1,j}}\},$$

as well as $PRICE_{t-1,i}$ for some i. Other covariates tested: intensity of international trade, resources endowment over time, $Y_{t-1,i}$...

Particular challenge:

- 1977: Clear air act for a decrease in carbon emissions.
- 1985: Stabilizing of gas consumption decline.
- 2008: American Recovery and Reinvestment Act for renewable energy.

Application to USA energy mix (1971 - 2015)

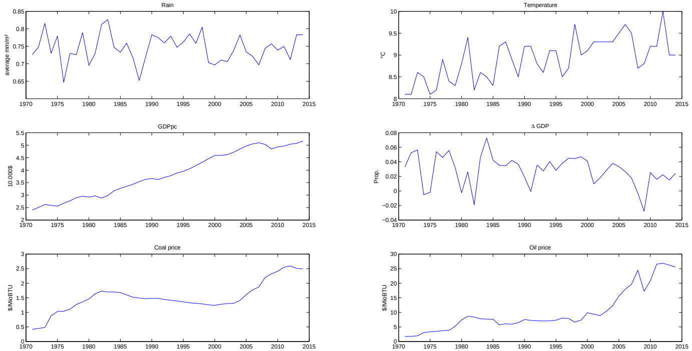

Figure: Rain level, temperature, GDP per capita and growth rate, coal and oil prices over the considered period.

Application to USA energy mix (1971 - 2015)

	USA energy mix				
Variable	Fossil	Oil	Gas	Nuclear	Renew.
cst.	0.144	0.76	0.628	-0.017	-0.004
	(10.47)	(10.91)	(10.4)	(101.31)	(144.11)
GDPpc	0.107	0.056	0.015	0.385	-0.011
	(1.27)	(1.32)	(1.4)	(1.37)	(1.29)
Δ GDP	-0.231	0.434	0.128	-0.165	0.531
	(19.22)	(18.94)	(19.13)	(17.54)	(17.09)
Rain	0.25	0.333	0.634	0.041	0.131
	(10.02)	(10.01)	(10)	(9.18)	(9.06)
Temp.	0.236	0.191	0.208	0.213	0.155
	(1.31)	(1.3)	(1.29)	(1.31)	(1.25)
Coal price	-0.017	-	-0.241	-	-
	(0.26)	-	(0.41)	-	-
Oil price	0.001	0.008	0.036	-	0.0299
	(0.03)	(0.02)	(0.05)	-	(0.03)
Gas price	-	-	-0.0207		
	-	-	(0.05)	-	-
Nuc. price	-	-	-	-0.001	0.000
	-	-	-	(0.05)	(0.08)
Fossil(t-1)	1.673	-	-	-	-
	(3.98)	-	-	-	-
Oil(t-1)	-	1.926	-	-	-
	-	(2.19)	-	-	-
Gas(t-1)	-	-	1.585	-	-
	-	-	(5.23)	-	-
Nuc(t-1)	-	-	-	2.439	-
	-	-	-	(4.81)	-
Renew.(t-1)	-	-	-	-	0.728
	-	-	-	-	(17.2)
MAE	0.60%	1.12%	0.65 %	0.82 %	0.49%
MAPE	2.79%	2.65%	2.60%	25.15%	10.99%

Application to USA energy mix (1971 - 2015)

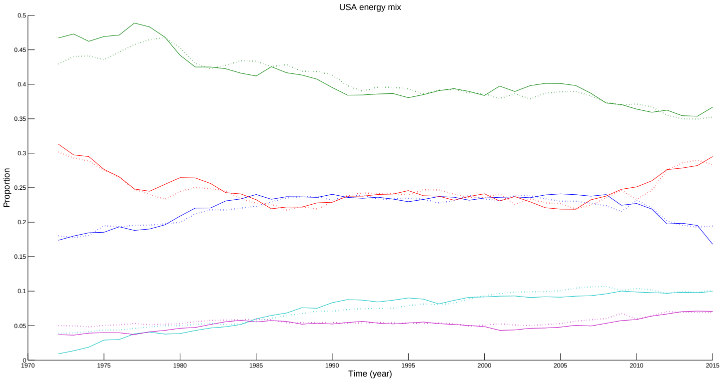

Figure: Solid: observed energy mix over the period 1972-2015 for USA. Dashed: one-year-ahead (in-sample) forecast of the expected mix for fossil (blue), oil (green), gas (red), nuclear (cyan) and renewable (purple).

Bootstrap algorithm:

- Use asymptotic normality of $\hat{\Theta}$: Generate B realization $\Theta^*_{(b)}$ from $N(0, \hat{\Sigma})$ where $\hat{\Sigma}$ is obtained from the variance-covariance matrix of the MLE.
- Build θ^*_{tj}, conditional on the historical covariates.
- Build $\theta^*_{tj}/\theta^*_{t0}$.
- CI is obtained from the empirical quantiles.

Application to USA energy mix (1971 - 2015)

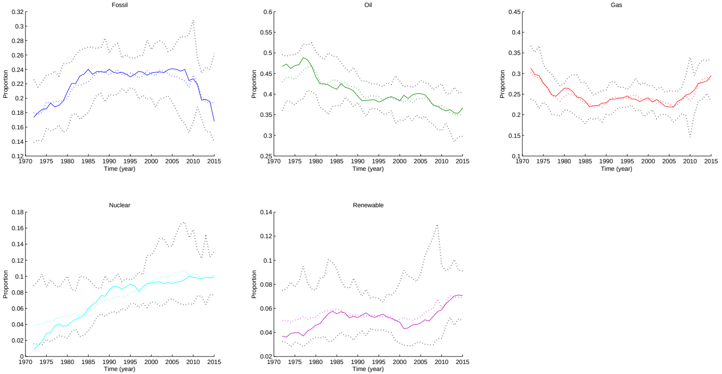

Figure: Confidence intervals for the estimated expected mix, for the different energy types.

Application to USA energy mix (2007 - 2015)

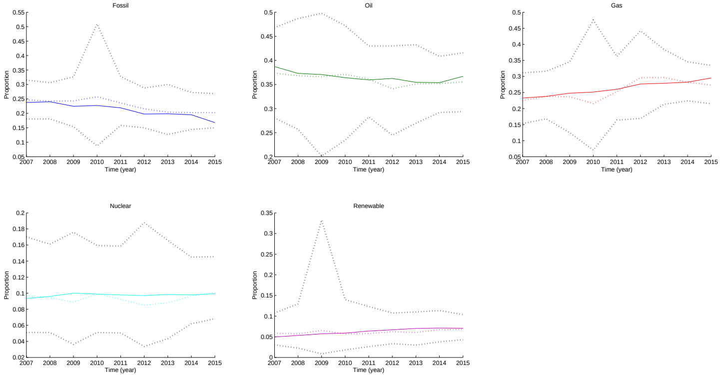

Figure: Solid: observed energy mix over the period 2007-2015. Dashed colored: out-of-sample forecasts obtained with a rolling window procedure. Dashed black: bootstrap prediction intervals (5000 resamples).

Application to USA energy mix (1971 - 2015)

USA energy mix						
Sample	Crit.	Fossil	Oil	Gas	Nuclear	Renew.
IS (rolling)	MAE	0.47%	0.99%	0.56%	0.75%	0.46%
	MAPE	2.09 %	2.36 %	2.30 %	15.41%	9.89%
OOS (rolling)	MAE	1.58 %	0.78 %	1.36 %	0.53 %	0.54%
	MAPE	7.81 %	2.10 %	5.12 %	5.37 %	9.00%
OOS (VAR-X, rolling)	MAE	1.21%	2.75%	1.48%	2.10%	0.69%
	MAPE	6.08%	7.63%	5.78%	21.58%	10.73%

Table: MAE and MAPE for out-of-sample forecasts over 2007-2015 with rolling estimation samples (35 years).

Conclusion and perspective

- Does it make sense to predict so complex data with so little information?
- How do we include policies in the regression framework ?
- There are thousands of them.
- It is difficult to say what they target exactly.
- How to stay parsimonious ?
- Comparison point (random walk, VAR-X, EIA).