ECONOMETRIC MODELLING OF PRICES OF SELECTED PRODUCTS FROM THE CONSUMER BASKET

Ondřej Šimpach

Abstract

The aim of the paper is to select particular products from the consumer basket and to model their prices in mutual relation. Multivariate time series model is used to describe the development of prices of two products in three groups and try to examine the relation between them. Vector Autoregressive Model was chosen where variables $x_{1,t}$ and $x_{2,t}$ (t = 1, 2, ..., T) are both endogenous and exogenous at the same time. The model was applied on time series of two dairy products, two meat products, and two alcohol beverages consumer prices. Observed period was from 01/2020 to 03/2025 (183 observations).

First, the seasonality of the time series was checked, and the series were seasonally adjusted where necessary (butter, both alcohol beverages). Then, it was examined by the Granger test of spurious regression whether there was a regression relationship between time series. VAR(6) for dairy products, VAR(4) for meat products, and VAR(3) for alcohol models were estimated. Finally, the selected models were diagnosed for autocorrelation, heteroscedasticity and normality of the residuals. These findings provide valuable insights for price forecasting, market analysis, and understanding consumer goods pricing mechanisms in the Czech market.

Key words: consumer prices, econometric modelling, multivariate time series models, VAR model

JEL Code: C32, Q11

Introduction

Consumer basket was created to be used for the calculation of the Consumer Price Index by Statistical office. It contains certain number of the representative products and services that are consumed by the households. We will focus on food and drink items that tend to have relations among themselves. Usually, expensive goods are substituted by cheaper. For example, beef meat is substituted by pork or non-meat alternatives. Caputo, Lusk and Blaustein-Rejto (2025) used an equilibrium displacement model, which links shifts in demand for plant-based meat alternatives to livestock and poultry supplies.

There can be a relation also between prices of products. 1) Whether consumers are willing to pay a premium for food products carrying local and organic claims; and 2) whether local and organic claims are complements or substitutes was examined by Gracia (2014) by a choice experiment in Spain for eggs.

In practice, time series are characterized by the fact that their development is linked to certain relationships between them. Also, food market develops based on the relation between various prices. For example, Aké (2017) analysed the non-linear relationship between the production of agricultural foods (cereals and vegetable oils) and the production of energy by using food. Girardi et al. (2015) found out that the correlation between agricultural prices and stock market returns tends to increase during periods of financial turmoil.

Modelling of the development of consumer prices of beef, pork and poultry meat using Box-Jenkins methodology (ARIMA) or VAR model was done by Šimpach and Šimpachová Pechrová (2018a) who used monthly data for 01/2006-02/2018. Slaughter pigs' prices from 01/1998 to 06/2016 were modelled and projected by SARIMA and VAR by Šimpach and Šimpachová Pechrová (2018b). ARIMA and VAR was used also by Šimpach and Šimpachová Pechrová (2018c) for modelling of monthly prices of milk from 01/1998 to 06/2016. Pechrová and Šimpach (2017a) modelled sugar prices in the period 09/2000–08/2016 also by ARIMA model. Pechrová and Šimpach (2017b) found out that the best model for the consumer prices of eggs was ARIMA(1,0,0) with constant and unit impulses in crisis months (03, 05, 07/2012). Nevertheless, VAR models might be preferred to ARIMA models because of the possibility to include relations between time series.

1 Methodology

The aim of the paper is to select particular products from the consumer basket and to model their prices in mutual relation. Multivariate time series model is used to describe the development of prices of both products and to examine the relation between them.

In practice, time series development is linked to certain relationships between them. This relationship must be identified and specified and therefore multivariate time series modelling is preferred. Forecasts that are constructed based on models characterizing the relationships between multiple time series can be more accurate than forecasts based on a single time series model. m-dimensional time series can be defined as a sequence of chronologically ordered values of m various random variables, i.e. m-dimensional random vector $\mathbf{X}_t = (X_{1t}, X_{2t}, ..., X_{mt})$. An m-dimensional time series represents a particular realization of an m-dimensional random process. The length of the time series X_t is equal to n, where $t = 1, 2, ... T$. There are various types of multivariate models based on whether there is a cointegration relation between them. One of them is Vector autoregressive (VAR) model that is commonly used to assess dynamic relationships when multiple variables are observed over a sufficiently long time period. It is a generalization of AR model to multivariate time series. Variables $x_{1,t}$ and $x_{2,t}$ ($t = 1, 2, ..., T$) are both endogenous and exogenous at the same time in VAR.

The variables in all evaluated time series are random and simultaneously dependent, i.e. they have an endogenous character, while their maximum lag length is the same. All VAR components (i.e. univariate AR processes) are assumed to be stationary.

If the reduced VAR form is used, on the right-hand side of which there are only lagged values (they are therefore predetermined at a given time), the method of least squares can be used for estimation. Otherwise, method of maximum likelihood must be used.

Before the modelling itself, the seasonality of the time series must be checked by F-test for seasonality. The series were seasonally adjusted where it was necessary. Then, it was examined by the Granger test of spurious regression whether there was a true regression relationship between time series. Then VAR(p) model can be constructed in a form:

$$\mathbf{X}_t = \mathbf{c} + \boldsymbol{\phi}_1\mathbf{X}_{t-1} + \boldsymbol{\phi}_2\mathbf{X}_{t-2} + \cdots + \boldsymbol{\phi}_p\mathbf{X}_{t-p} + \mathbf{a}_t \qquad (1)$$

where $\mathbf{c}$ is vector of constants, $\boldsymbol{\phi}_i$ is a matrix of parameters, $\mathbf{X}_{t-p}$ is a matrix of lagged explanatory variables (t is time, p is lag) and $\mathbf{a}_t$ is a matrix white noise process (of a Gaussian type). Model VAR is of order p, if the matrix of parameters $\boldsymbol{\phi}_p$ is non-zero (i.e. at least one of its elements is non-zero) and matrix $\boldsymbol{\phi}_i$, $i = p + 1, p + 2, ...$ are zero.

Estimated model parameters are verified with standard univariate and multivariate diagnostic tools for autocorrelation, heteroskedasticity and normality of the non-systematic component.

Autocorrelation of the non-systematic component was tested by modified portmanteau test (Hosking, 1980 and other authors) with null hypothesis H_0: $\rho_{a1} = \rho_{a2} = \cdots = \rho_{ak} = \rho_{a2_{mxm}} = 0$ (all correlation coefficients ρ are equal to zero) and alternative hypothesis H_1: non H_0. If the test criterion Q_m^T is higher than critical value of χ^2 distribution for VAR(p) than the H_0 is rejected at the selected significance level.

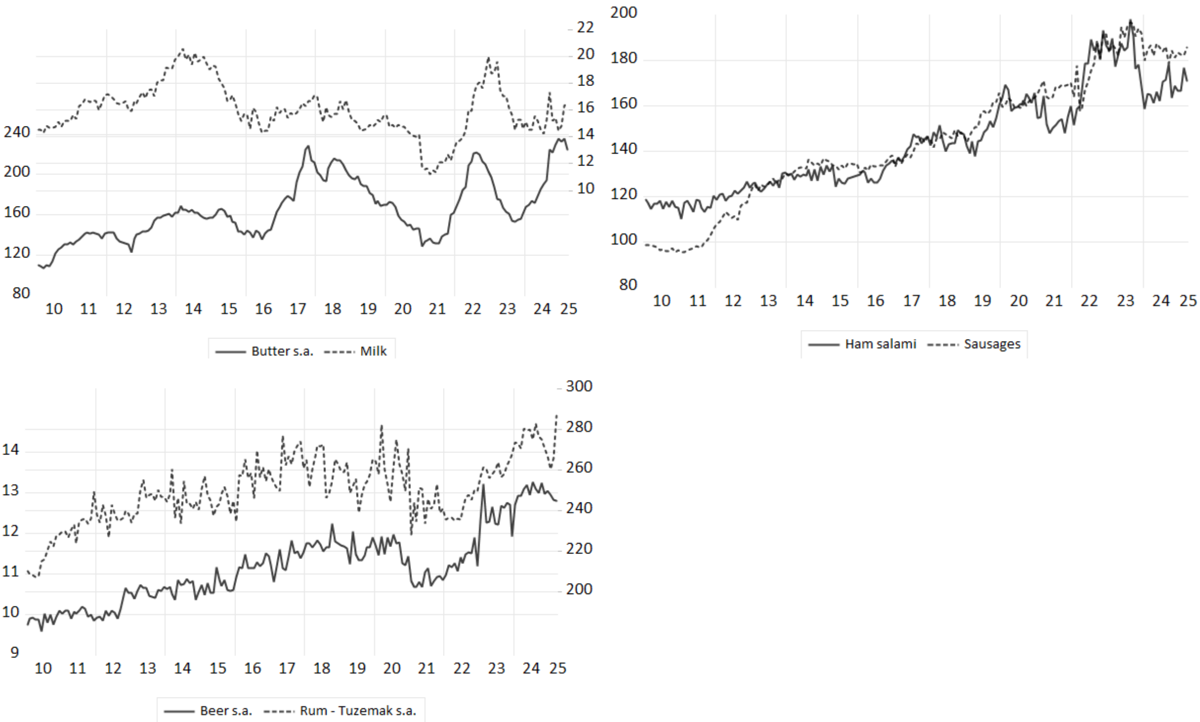
Normality of the non-systematic component is important for maximum likelihood estimation of VAR parameters, parameter testing and interval estimation. It is tested whether the process $\mathbf{a}_t$ is of Gaussian type, i.e. variables $\mathbf{a}_t$ have m-dimensional normal distribution.

There are two statistics: MSK^2 and MKU^2. Statistics MSK^2 (multivariate skewness) is a test criterion for skewness testing. Statistics MKU^2 (multivariate kurtosis) is a test criterion for kurtosis testing. Joint criterion (multivariate Jarque–Bera): $MJB = (MSK^2 + MKU^2)$. If the null hypothesis holds, it asymptotically follows distribution $\chi^2[2\ d.f.]$ at the selected significance level.

White test adjusted for multivariate time series is used for heteroskedasticity testing. Null hypothesis assumes that H_0: residuals are multivariate homoscedastic and H_1: non H_0. Test is based on auxiliary regression of squared residues on original independent variables, and the test statistic is derived from its coefficients. If the null hypothesis holds, it asymptotically follows distribution χ^2 at the selected significance level.

The model was applied on seasonally adjusted time series (when necessary) – of two dairy products (butter and milk), two meat product (ham salami and sausages), and two alcohol beverages (beer and rum–tuzemák) consumer prices. Observed period was from 01/2020 to 03/2025 (183 observations). The empirical development is described at Fig. 1. The data was taken from Czech Statistical Office from public database.

Fig. 1: Development of the consumer prices



Source: Czech Statistical Office (2025), own elaboration

2 Results

First, the seasonality of the time series was checked, and the series were seasonally adjusted where necessary. Then, it was examined by the Granger test of spurious regression whether there was a regression relationship between time series. Finally, VAR models were estimated and diagnosed for autocorrelation, heteroscedasticity and normality of the residuals.

2.1 Seasonality

If relationships are sought between two or more time series, it is necessary to ensure a certain degree of compatibility between them. In our case, this means that it is not desirable for one analysed time series to show seasonality and the other not.

In the case of the selected time series of dairy products, i.e. butter and milk, the F-test of seasonality found that the time series butter contains seasonality at the 5% significance level, milk does not contain it. To ensure the degree of compatibility, the time series butter after seasonal adjustment will be used, the time series milk will be the original. In the case of the selected time series of meat products, i.e. ham salami and sausages, the F-test of seasonality found that seasonality is not present at the 5% significance level. It is therefore possible to work with the original time series. In the case of the selected time series of alcohol, i.e. beer and rum–tuzemák, it was found by the F-test of seasonality that the time series of beer and rum–tuzemák contain seasonality at the 5% significance level. Both time series were therefore used seasonally adjusted.

2.2 Granger test of spurious regression

Using the Granger regression test, it was evaluated whether the regression relationship is true or only spurious for the time series of dairy products, meat products and alcohol. After deep examination of the residuals of the so-called static models, it was found that the residuals are finally stationary at the 5% significance level and the true regression relationships can therefore exist.

2.3 Vector Autoregressive model

Therefore, we do not work with seasonality in any time series (either there was no seasonality in the time series or there was, but it was removed) and the periodicity of the time series is monthly. During the construction of VAR models and the choice of the appropriate lag order, we cannot rely only on the economic knowledge and the production processes of individual products. It is necessary to use the information criteria FPE, AIC, BIC and H–Q. Based on the calculated results, it was recommended to choose the length VAR(6) for dairy products, the length VAR(4) for meat products and the length VAR(3) for alcohol (see Fig. 2).

Fig. 2: VAR(6) model for diary products, VAR(4) for meat products, VAR(3) for alcohol

	BUTTER_SA	MILK
BUTTER_SA(-1)	1.051268	-0.002390
	(0.07993)	(0.00917)
	[13.1527]	[-0.26054]
BUTTER_SA(-2)	0.269064	0.020734
	(0.11640)	(0.01336)
	[2.31147]	[1.55175]
BUTTER_SA(-3)	-0.114366	-0.007092
	(0.11497)	(0.01320)
	[-0.99476]	[-0.53743]
BUTTER_SA(-4)	-0.317932	0.000439
	(0.11425)	(0.01311)
	[-2.78283]	[0.03349]
BUTTER_SA(-5)	-0.086436	-0.011617
	(0.11485)	(0.01318)
	[-0.75261]	[-0.88122]
BUTTER_SA(-6)	0.163246	-0.000656
	(0.07991)	(0.00917)
	[2.04280]	[-0.07147]
MILK(-1)	1.002393	0.776469
	(0.69125)	(0.07935)
	[1.45012]	[9.78584]
MILK(-2)	-1.871341	0.041617
	(0.86643)	(0.09945)
	[-2.15983]	[0.41846]
MILK(-3)	-0.481720	0.238870
	(0.87650)	(0.10061)
	[-0.54959]	[2.37420]
MILK(-4)	1.529698	-0.085214
	(0.89408)	(0.10263)
	[1.71092]	[-0.83032]
MILK(-5)	1.084411	0.148476
	(0.89558)	(0.10280)
	[1.21085]	[1.44431]
MILK(-6)	-1.720562	-0.171191
	(0.71761)	(0.08237)
	[-2.39763]	[-2.07827]
C	13.44885	0.895296
	(4.21368)	(0.48367)
	[3.19171]	[1.85104]

	HAM_SALAMI	SAUSAGES
HAM_SALAMI(-1)	0.643402	0.149792
	(0.08001)	(0.05617)
	[8.04136]	[2.66663]
HAM_SALAMI(-2)	0.311904	-0.064811
	(0.09524)	(0.06686)
	[3.27503]	[-0.96932]
HAM_SALAMI(-3)	0.072108	0.139187
	(0.09652)	(0.06776)
	[0.74712]	[2.05413]
HAM_SALAMI(-4)	-0.149275	-0.147327
	(0.08352)	(0.05863)
	[-1.78739]	[-2.51270]
SAUSAGES(-1)	-0.002403	0.492386
	(0.11235)	(0.07888)
	[-0.02139]	[6.24230]
SAUSAGES(-2)	0.062703	0.346698
	(0.12329)	(0.08656)
	[0.50858]	[4.00539]
SAUSAGES(-3)	0.133164	0.185253
	(0.12128)	(0.08515)
	[1.09795]	[2.17565]
SAUSAGES(-4)	-0.110123	-0.088256
	(0.10765)	(0.07558)
	[-1.02295]	[-1.16774]
C	5.771690	-1.159530
	(2.68473)	(1.88484)
	[2.14982]	[-0.61519]

	BEER_SA	RUM_TUZEMAK_SA
BEER_SA(-1)	0.564344	5.927239
	(0.07400)	(2.25498)
	[7.62658]	[2.62851]
BEER_SA(-2)	0.125734	-1.089122
	(0.08503)	(2.59121)
	[1.47869]	[-0.42031]
BEER_SA(-3)	0.303056	0.430302
	(0.07673)	(2.33816)
	[3.94982]	[0.18403]
RUM_TUZEMAK_SA(-1)	-0.002743	0.329127
	(0.00258)	(0.07863)
	[-1.06326]	[4.18580]
RUM_TUZEMAK_SA(-2)	0.002034	0.301862
	(0.00261)	(0.07965)
	[0.77826]	[3.78970]
RUM_TUZEMAK_SA(-3)	-2.97E-06	-0.013450
	(0.00251)	(0.07651)
	[-0.00118]	[-0.17580]
C	0.283556	37.15037
	(0.32947)	(10.0402)
	[0.86065]	[3.70017]

Source: Czech Statistical Office (2025), own elaboration

From the estimated models and their statistically significant parameters, it can be deduced that the consumer price of butter at time *t* directly depends on what this price was 1 and 2 months ago. This consumer price indirectly depends on how much butter cost 4 months ago and again directly depends on what this price was 6 months ago. The consumer price of butter at time *t* should also depend on the consumer price of milk. It can be deduced that the consumer price of butter at time *t* is indirectly related to what the consumer price of milk was 2 and 6 months ago, respectively. The estimated constant in this equation is statistically significant at the 5% significance level. The evaluation of the second equation of the VAR model for dairy products is as follows: the consumer price of milk at time *t* does not depend at all on the consumer price of butter, probably because milk is an input for the production of butter. However, the consumer price of milk at time *t* certainly depends on the consumer prices of milk in the lag. It can be seen that there is a direct dependence on the consumer price recorded 1 and 3 months ago. In addition, the consumer price of milk at time *t* also indirectly depends on its price 6 months ago. The estimated constant in this equation is not statistically significant at the 5% significance level.

The results of the model for meat products can be commented in a similar way. The consumer price of ham salami at time *t* directly depends on the consumer price of this salami, which was recorded 1 and 2 months ago. The consumer price of ham salami at time *t* does not depend on any consumer price of sausages at all. The estimated constant in this equation is statistically significant at the 5% significance level. The evaluation of the second equation of the VAR model for meat products is as follows: the consumer price of sausages at time *t* directly depends on the consumer price of ham salami recorded 1 and 3 months ago, and in addition, it also indirectly depends on the consumer price of this salami 4 months ago. The consumer price of sausages at time *t* also depends on itself, specifically directly on the consumer price that was recorded 1, 2 and 3 months ago. The estimated constant in this equation is statistically insignificant at the 5% significance level.

The results of the model describing the relationships for alcohol beverages are also remarkable. The estimated model and its statistically significant parameters show that the consumer price of beer at time *t* directly depends on the consumer price of beer recorded 1 and 3 months ago. However, the consumer price of beer does not depend on the consumer price of rum–tuzemák. The estimated constant in this equation is not statistically significant at the 5% significance level. The consumer price of rum–tuzemák depends on the consumer price, and directly. There is a direct connection with the price of beer 1 month ago. The consumer price of domestic rum (rum–tuzemák) at time *t* directly depends on its own consumer price recorded 1 and 2 months ago. The estimated constant in this equation is statistically significant at the 5% significance level.

Conclusion

Based on the VAR model analysis of consumer prices for dairy products (VAR(6)), meat products (VAR(4)), and alcoholic beverages (VAR(3)), several conclusions can be drawn. The analysis reveals distinct patterns of price relationships within and across product categories. All analysed products demonstrate autoregressive behaviour, meaning current prices are significantly influenced by their own historical values. The lag structures vary by product category: dairy products show the longest memory effects (up to 6 months), while alcoholic beverages have shorter adjustment periods (up to 3 months). This suggests different market adjustment speeds and inventory cycles across product categories.

Dairy products show asymmetric interdependence – while butter prices are influenced by historical milk prices (2 and 6 months ago), milk prices are independent on butter prices, likely reflecting the production relationship where milk serves as an input for butter manufacturing.

Meat products demonstrate limited cross-product influence – ham salami prices show no dependence on sausage prices, while sausage prices are significantly affected by historical ham salami prices across multiple time periods. This suggests a unidirectional relationship.

Alcoholic beverages exhibit product-specific pricing patterns – beer and rum–tuzemák prices operate largely independently of each other, with each product primarily influenced by its own historical pricing. However, rum–tuzemák prices show some sensitivity to beer prices from 1 month ago, indicating potential weak substitution effects.

The models reflect the different complexity levels of price dynamics across product categories. These findings provide valuable insights for price forecasting, market analysis, and understanding consumer goods pricing mechanisms in the Czech market.

Acknowledgment

The paper was supported by long-term institutional support of research activities by the Faculty of Informatics and Statistics of Prague University of Economics and Business, research project funded by VŠE IP 400040.

References

Aké, S. C. (2017). The nonlinear relation between biofuels and food prices. *Investigación Económica*, *76*(299), 3-26.

Caputo, V., Lusk, J. L., & Blaustein-Rejto, D. (2025). Plant-based versus conventional meat in food away from home settings: Substitution, complementarity, and market impacts. *Agricultural Economics*. Advance online publication. https://doi.org/10.1111/agec.70002

Girardi, D. (2015). Financialization of food. Modelling the time-varying relation between agricultural prices and stock market dynamics. *International Review of Applied Economics*, *29*(4), 482-505. https://doi.org/10.1080/02692171.2015.1016406

Gracia, A., Barreiro-Hurlé, J., & López-Galán, B. (2014). Are local and organic claims complements or substitutes? A consumer preferences study for eggs. *Journal of Agricultural Economics*, *65*(1), 49-67. https://doi.org/10.1111/1477-9552.12036

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, *75*(371), 602-608. https://doi.org/10.2307/2287656

Pechrová, M., & Šimpach, O. (2017a). Modelling the development of the consumer price of sugar. In P. Pražák (Ed.), *Mathematical methods in economics (MME 2017)*, University of Hradec Králové (pp. 527-531)

Pechrová, M., & Šimpach, O. (2017). Modelling the development of eggs prices in the CR. In P. Jedlička, P. Marešová, & I. Soukal (Eds.), *Hradec Economic Days*, University of Hradec Králové (Vol. 7, pp. 682-688)

Šimpach, O., & Šimpachová Pechrová, M. (2018a). Food availability and the development of the consumer prices of meat. In K. Tomšík (Ed.), *Agrarian perspectives XXVII - Food safety - Food security*, Czech University of Life Sciences (pp. 401-407).

Šimpach, O., & Šimpachová Pechrová, M. (2018b). Modelling of slaughter pigs' prices in the Czech Republic using stochastic models. In L. Váchová & V. Kratochvíl (Eds.), *Mathematical methods in economics (MME 2018)*, Prague University of Economics and Business (pp. 539-544).

Šimpach, O., & Šimpachová Pechrová, M. (2018c). Development of milk prices in the Czech Republic. In P. Jedlička, P. Marešová, & I. Soukal (Eds.), *Hradec Economic Days*, Pt II, 2018 University of Hradec Králové (Vol. 8, pp. 378-388)

Contact

Ondřej Šimpach

Prague University of Economics and Business

Winston Churchill Sq. 1968/4, 130 67 Prague 3, Czech Republic

ondrej.simpach@vse.cz