PROJECTION OF AIR EMISSIONS ACCOUNTS FOR GREENHOUSE GASES USING ARDL MODEL

Ondřej Šimpach

Abstract

The aim of the paper is to model and predict the air emissions for greenhouse gases (from transportation and storage in thousand tonnes) using Autoregressive distributed lag (ARDL) model. Explanatory variable is road freight transport in thousand tonnes which is predicted by Seasonal autoregressive integrated moving average (SARIMA) model. The results of ARDL model are compared to deterministic additive decomposition model using mean square error (MSE). The time series from Eurostat include 63 observations of quarterly data from Q1/2010 to Q3/2025, prediction horizon is up to Q4/2027. ARDL(6,2) model revealed that emissions are statistically significantly dependent on itself up to 6 lags (indirectly at lags 2 and 5), and on road freight volume in actual volume and up to second lag (indirectly at lag 1). Also, the unit impulse (+1) for covid pandemic at Q2/2020 and Q4/2020 was significant. Transport was projected by SARIMA(2,0,0)(1,0,0)$_4$. The most probable way of development was the middle variant which was used for predicting emissions. Middle projection of emissions expects their slight decrease. Prediction by deterministic additive decomposition had narrower confidence intervals and expected emissions decrease. Its MSE was ten times higher than those of ARDL model. Stochastic projection is, therefore, preferred for the prediction.

Key words: ARDL, freight transport, prediction, time series, transportation emission

JEL Code: C22, Q56

Introduction

Transportation is seen as a major contributor to greenhouse gas (GHG) emissions worldwide. “With increasing population growth and food demand, the spatial decoupling of production and consumption has intensified, driving a surge in transnational agricultural products transportation” (Gao et al., 2026). The European Union (EU) is legally bound to achieve climate neutrality by 2050, but current greenhouse gas (GHG) trajectories indicate that this target may be missed. (Schau et al., 2026). “Significant emission reductions are needed in the transport sector to reach climate change mitigation objectives” (Camilleri et al., 2024). The emissions

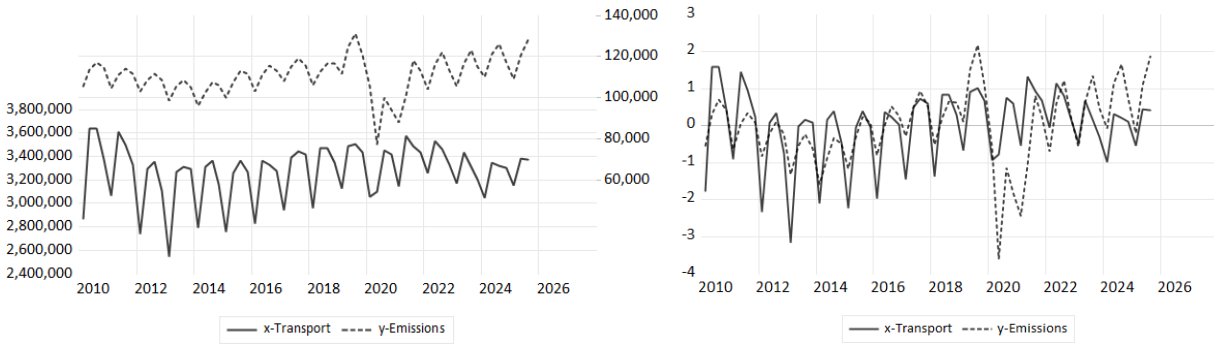
are monitored, because the aim of the EU is to cut emissions by at least 50% by 2030, and achieve the climate neutrality by 2050. Air emissions accounts assembled by the Eurostat record flows of gaseous and particulate materials (7 greenhouse gases and 7 air pollutants) emitted into the atmosphere as a result of economic activity. The emissions can be reduced with help of modern technologies – e. g. by AI that can optimize vehicle energy management, traffic flow, and alternative fuel technologies, such as hydrogen fuel cells and biofuels. (Miller et al., 2024).

Petrovic et al. (2020) studied the relationship between transport GHG emissions and the economic activity within the EU during 2000–2014. They found out that an inverse-U relationship (Kuznets curve) exists between transport GHG emissions and GDP per capita. Andrés and Padilla (2018) modelled the influence of population, economic activity, transport volume and structural characteristics of transport activity in terms of transport energy intensity, of transport modes' share and of energy sources' mix on changes in greenhouse gas emissions from transport in the EU-28 during 1990–2014. Charly and Calufield (2025) linked traffic volume, fleet, and emission factors and developed a model for link-based transport emissions and showed that emissions generated by a vehicle were much higher in rural than urban areas.

1 Material and Methods

The aim of the paper is to model and project for period Q4/2025 to Q4/2027 the emissions from transportation and storage in thousand tonnes (y_t – *emissions*). Explanatory variable is road freight transport in thousand tonnes (x_t – *transport*). The time series were obtained from Eurostat from Air emissions accounts and included 63 observations of quarterly data from Q1/2010 to Q3/2025. The development of both time series is in Fig 1.

Fig. 1: Road freight transport (x_t) and emissions from transportation and storage (y_t) (in thousand tonnes) in Q1/2010–Q3/2025 – original values (left), standardized values (right)



Source: own elaboration based on data from Eurostat (2026a) and Eurostat (2026b)

There is clear economic shock visible (decline of road freight transport during Covid-19 pandemic) that shall be modelled by unit impulse. The seasonality of both time series was tested by seasonality F-test and Kruskal-Wallis test.

Then, the ARDL model that expressed the relationship between variables in a single-equation time-series was setup. The dependent and independent variables were related not only contemporaneously, but across historical (lagged) values. A general ARDL(p, q) model was given by formula (1) – see e.g. Šimpach & Šimpachová Pechrová (2025):

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=0}^{q} \delta_j x_{t-j} + \varepsilon_t, \quad (1)$$

where β_0 is a constant, β_i is a coefficient associated with lags of y_t, and δ_t with lags of x_t. ε_t is a white noise, p is lag for dependent and q for independent variable. The optimal lag orders were obtained by minimizing a model selection criterion (AIC). Diagnostics tests of the residuals were performed next – autocorrelation was tested by Breusch-Godfrey serial correlation LM test, homoskedasticity by ARCH test and normality by Jarque-Bera test.

The projection of emissions was done for 9 quarters – from 4Q/2025 to 4Q/2027. Explanatory variable (x_t – transport) was projected before using seasonal SARIMA(P,D,Q)(p,d,q)$_4$ model, where P is lag for seasonal dependent and Q for seasonal independent variable, and D and d are orders of integration.

The ARDL model was compared by mean square error (MSE) to the deterministic additive decomposition model of y_t – emissions (2). The modelling and projection of x_t – transport was done also by additive decomposition.

$$y_t = \beta_0 + \beta_2 x_t + \beta_1 t + \sum_{i=1}^{3} \delta_i d_{i,t} + \varepsilon_t, \quad (2)$$

where β_0 is constant, β_1 coefficient for transport x_t. and β_3 coefficient for time variable t (ranging from 1 to $T = 63$), d_1 to d_3 are dummy variables for seasons with coefficients δ_1 to δ_3, and ε_t is white noise. Calculated model with 3 dummy variables is then recalculated for all 4 seasonal coefficients. Projection was also done for 9 quarters – from 4Q/2025 to 4Q/2027.

Calculations were done in EViews 13 and in MS Excel.

2 Results and Discussion

First, the transportation and storage in thousand tonnes and road freight transport in thousand tonnes time series was tested for seasonality (see Tab. 1.). Almost all tests found statistically significant seasonality at least at 0.05 level of significance. Therefore, it is necessary to use the SARIMA(P,D,Q)(p,d,q) model to model and project the time series of x_t transport.

Tab. 1: Seasonality tests for x_t transport and y_t emissions

Tests for seasonality	x_t transport		y_t emissions	
	Statistics	Significance	Statistics	Significance
Test for the presence of seasonality assuming stability (F-value)	104.479	**	47.844	**
Nonparametric test for the presence of seasonality assuming stability (Kruskal-Wallis statistics)	46.806	***	50.805	***
Moving seasonality test (F-value)	4.481	**	1.220	

Note: *** α = 0.01, ** α = 0.05, * α = 0.1

Source: own elaboration

2.1 ARDL model

ARDL model without any adjustments was constructed and tested, but the results and diagnostics were not optimal. There were an autocorrelation, a heteroscedasticity and a non-normality of the residuals present. The residual plot showed outliers and the need to add exogenous shocks due to Covid pandemic. Particularly, the impulses (+1) were added to 2 observations: Q2/2020 and Q4/2020. The results of new ARDL(6,2) model with 6 lags for *y* and 2 lags for *x* are presented in Tab. 2. All coefficients are statistically significant at minimally 0.05 level of significance. The emissions directly dependent on themself – one, three, four and six quarters before. Indirect relation was with two and five quarters before. Emissions also are directly dependent on road freight volume in time *t* (level) and two quarters before Indirect relation was found with one quarter before. The exogenous shocks were set correctly, because unit impulse was statistically significant too.

Tab. 2: ARDL(6,2) model of y_t emissions explained by x_t transport

Variable	Coefficient	Std. error	t-statistic	Significance
β_0 (constant)	-24648.670	12173.250	-2.025	**
β_1 (y_{t-1})	0.866	0.065	13.339	***
β_2 (y_{t-2})	-0.768	0.094	-8.206	***
β_3 (y_{t-3})	0.472	0.072	6.561	***
β_4 (y_{t-4})	0.233	0.066	3.526	***
β_5 (y_{t-5})	-0.225	0.067	-3.363	***
β_6 (y_{t-6})	0.152	0.058	2.626	**
δ_0 (x_t)	0.014	0.002	7.248	***
δ_1 (x_{t-1})	-0.005	0.002	-2.127	**
δ_2 (x_{t-2})	0.008	0.002	3.164	***
Impulse	-33746.800	2422.069	-13.933	***

Source: own elaboration

The autocorrelation of the residuals was still present in the model as same as in unadjusted ARDL model, but in contrast, ARDL(6,2) had homoscedastic and normal residuals (see results of diagnostics tests in Tab. 3).

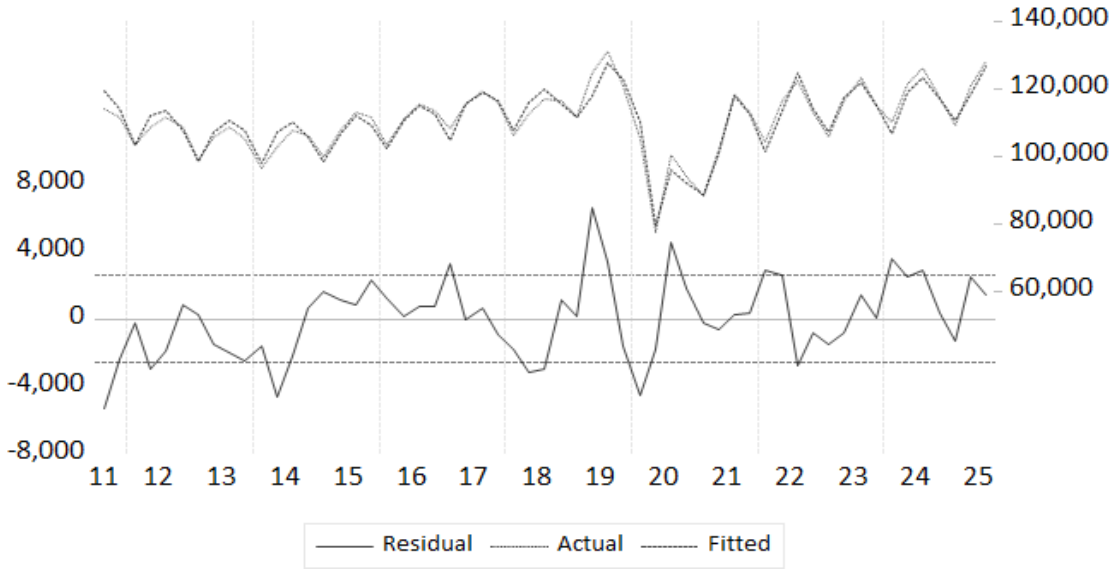
Tab. 3: Diagnostic tests of the ARDL(6,2) model

Null hypothesis H_0	Test	Statistics	Significance
No autocorrelation	Breusch-Godfrey Serial Correlation LM test (F-value)	7.395	***
Homoscedasticity	ARCH heteroscedasticity test (F-value)	0.039	
Normality	Jarque-Bera normality test (JB statistics)	0.209	

Source: own elaboration

From Fig. 2 can be seen that the model fits the empirical values very well – the development is almost similar. Hence, we can use the model for predictions to the future. MSE was calculated as the mean squared difference between actual and fitted values and its value was 5 398 068.37.

Fig. 2: Residuals and actual and fitted values of ARDL(6,2) model



Source: own elaboration

Before predicting the emissions y_t, the explanatory variable transport x_t had to be predicted. Because of the seasonality of the transport, SARIMA model was used. Final SARIMA(2,0,0)(1;0;0)$_4$ with unit impulses (+1) at observation 2020/Q2 and 2020/Q4 is presented in Tab. 4.

Tab. 4: SARIMA(2,0,0)(1,0,0)$_4$ model for x_t transport

Variable	Coefficient	Std. error	t-statistic	Significance
IMP	-166334.200	48197.970	-3.451	***
AR(1)	0.609	0.113	5.410	***
SAR(4)	0.655	0.127	5.145	***
SAR(8)	0.342	0.126	2.707	***

Source: own elaboration

The results of the diagnostics are presented in Tab 5. There was no autocorrelation of residuals found in the model. The residuals are also homoscedastic and normally distributed.

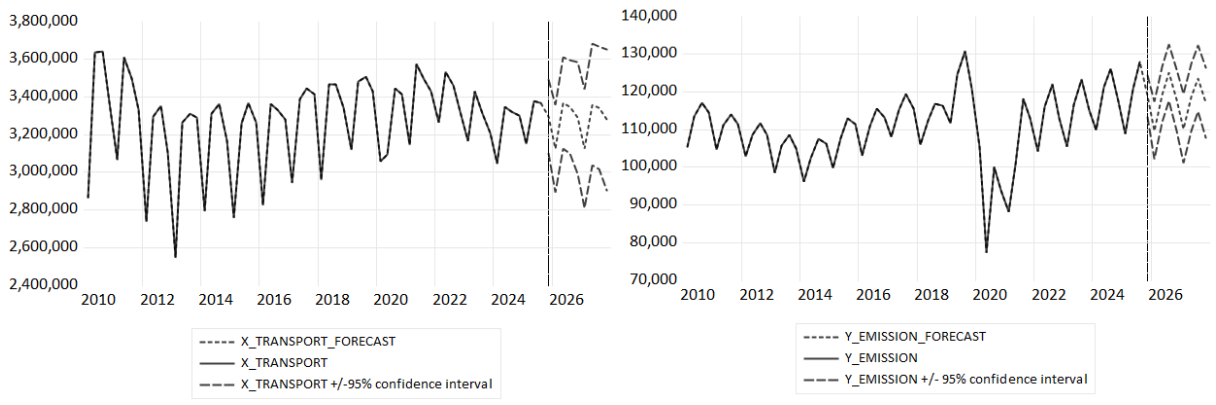
Tab. 5: Diagnostic tests of the SARIMA(2,0,0)(1;0;0)$_4$ model

Null hypothesis H_0	Test	Statistics	Significance
No autocorrelation	Breusch-Godfrey Serial Correlation LM test (F-value)	1.543	
Homoscedasticity	ARCH heteroscedasticity test (F-value)	0.301	
Normality	Jarque-Bera normality test (JB statistics)	0.487	

Source: own elaboration

SARIMA(2,0,0)(1,0,0)$_4$ model was used for prediction of x_t transport. Finally, the prediction of y_t emissions for 9 quarters ahead was done by ARDL(6,2) model. Empirical and predicted values together with 95% confidence intervals for x_t transport and y_t emissions are presented in Fig. 3. The most probable way of development of x_t transport was the middle variant which was used for predicting y_t emissions. Middle projection of emissions expects their slight decrease. However, the increase of emissions (upper 95% confidence interval) might be also probable, because it follows the growing trend since 2020 after the pandemic.

Fig. 3: Empirical and predicted values for x_t transport (left) and y_t emissions (right)



Source: own elaboration

2.2 Deterministic additive decomposition

For comparison, the deterministic additive decomposition of the y_t emissions was done with explanatory variable x_t transport. The model was used for prediction, where x_t transport was also predicted by additive decomposition. Estimated and recalculated coefficients for x_t and y_t models are displayed at Tab. 6.

Tab. 6: Additive decomposition of x_t transport s (up) and y_t emission (down)

x_t – transport	Coeff. original	t-statistics	Significance	Coeff. recalculated	
β_0 (constant)	3257608.398	69.786	***	β_0^* (constant)	3220753.910
β_2 (time)	1693.979	1.782	*	β_2^* (time)	1693.979
δ_1 (d_1)	-343647.629	-6.972	***	δ_1^* (d_1)	-380502.117
δ_2 (d_2)	91336.267	1.853	*	δ_2^* (d_2)	54481.779
δ_3 (d_3)	104893.412	2.128	**	δ_3^* (d_3)	68038.925
				δ_4^* (d_4)	36854.488
y_t – emissions	**Coeff. original**	**t-statistics**	**Significance**	**Coeff. recalculated**	
β_0 (constant)	62275.769	2.590	**	β_0^* (constant)	61774.212
β_1 (transport)	0.014	1.904	*	β_1^* (transport)	0.014
β_2 (time)	98.513	1.805	*	β_2^* (time)	98.513
δ_1 (d_1)	-3104.631	-0.831		δ_1^* (d_1)	-3606.189
δ_2 (d_2)	-2624.543	-0.926		δ_2^* (d_2)	-3126.101
δ_3 (d_3)	3722.945	1.302		δ_3^* (d_3)	3221.388
				δ_4^* (d_4)	501.557

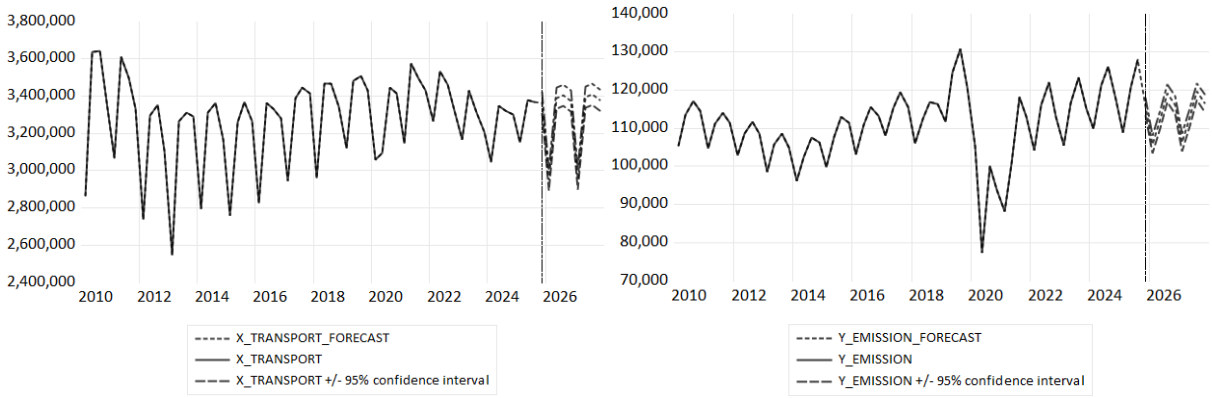
Source: own elaboration

Additive decomposition cannot capture external shocks as good as SARIMA and ARDL models which can be seen from Fig. 4. The MSE of the model was tenth times higher than in case of ARDL model and accounted to 53 887 720. The stochastic model better captures the development than deterministic model as expected. In contrast to stochastic model, the confidence intervals in deterministic models are narrower. Only lower 95% interval expect decrease of the transport volume, upper interval and medium prediction expect increase (that is probable because of the previous trend). Projected decrease of emissions does not follow the increasing trend from previous years but rather expect decline.

This is in line with Rečka, Máca & Scasny (2023) who calculated on the case study of the Czech Republic that "given the high level of GHG emissions in 1990 and the significant reduction in GHG emissions in the 1990s, Czechia could achieve a 55% reduction by 2030".

So, they also expect a decreasing trend in emissions. Janhuaton, Ratanavaraha & Jomnonkwao (2024) did a scenario analysis by artificial neural network model and also expected a reduction in CO_2 emissions from fuel combustion in the transportation sector by 2030.

Fig. 4: Empirical and predicted values for x_t transport (left) and y_t emissions (right)



Source: own elaboration

Conclusion

The aim of the paper was to model and project for period Q4/2025 to Q4/2027 the emissions from transportation and storage in thousand tonnes (y_t – *emissions*) by ARDL model with explanatory variable road freight transport in thousand tonnes (x_t – *transport*). The time-series data, obtained from Eurostat, spanned 63 quarterly observations from Q1/2010 to Q3/2025.

The empirical findings based on the optimal ARDL(6,2) model demonstrated that air emissions are significantly influenced by their own historical values up to six lags, as well as current and lagged values of road freight transport (up to the second lag). Accounting for exogenous structural shocks – specifically the COVID-19 pandemic in Q2/2020 and Q4/2020 via unit impulses – proved to be essential for capturing non-systemic variations. Road freight transport was predicted using an optimal SARIMA(2,0,0)(1,0,0) model. The middle projection scenario indicated a slight overall decrease in transport emissions over the predicted horizon, aligning with broader policy goals for climate mitigation, although upper 95% confidence interval suggests possible increase of the emissions.

When evaluated against a traditional deterministic additive decomposition model, the stochastic ARDL approach proved significantly more robust. The deterministic model yielded a Mean Squared Error (MSE = 53 887 720) nearly ten times higher than that of the ARDL model (MSE = 5 398 068), primarily due to its inability to account for unexpected external shocks effectively. Nevertheless, the confidence intervals in additive decomposition model

were narrower and the prediction of the emissions expected their decrease in all variants (middle and ±95% confidence intervals).

We may conclude that stochastic time-series modelling (specifically ARDL model complemented with SARIMA) offers a suitable framework for projecting transport-related emissions. While the predicted slight decline in emissions aligns with broader regional decarbonization expectations, continuous tracking and active intervention (such as integrating technological innovations and transport optimization) remain crucial to achieving the European Union's long-term climate neutrality targets.

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

Andrés, L., & Padilla, E. (2018). Driving factors of GHG emissions in the EU transport activity. *Transport Policy*, *61*, 60–74. https://doi.org/10.1016/j.tranpol.2017.10.008

Gao, Z., Chen, X., Yin, Y., Hou, P., Yin, T., Long, Y., Han, S., & Li, Y. (2026). Spatiotemporal dynamics of global agricultural transport greenhouse gas emissions from trade and the climate implications of localization strategies. *Environmental Science & Technology*, *60*(1), 376–387. https://doi.org/10.1021/acs.est.5c05787

Camilleri, R., Attard, M., & Hickman, R. (2024). Participatory policy packaging for transport backcasting: A pathway for reducing CO2 emissions from transport in Malta. *Sustainability*, *16*(1), Article 430. https://doi.org/10.3390/su16010430

Charly, A., & Caulfield, B. (2025). Transport emission modelling based on a bottom-up approach to facilitate sustainable transport planning. *Sustainable Futures*, *9*, Article 100435. https://doi.org/10.1016/j.sftr.2025.100435

Eurostat (2026a). Road freight transport by type of operation and type of transport - (t, tkm, vehicle-km) - quarterly data [road_go_tq]. https://doi.org/10.2908/ROAD_GO_TQ_TOTT

Eurostat (2026b). Air emissions accounts for greenhouse gases by NACE Rev. 2 activity – quarterly data [env_ac_aigg_q]. https://doi.org/10.2908/ENV_AC_AIGG_Q

Janhuaton, T., Ratanavaraha, V., & Jomnonkwao, S. (2024). Forecasting Thailand's transportation CO2 emissions: A comparison among artificial intelligent models. *Forecasting*, *6*(2), 462–484. https://doi.org/10.3390/forecast6020026

Miller, T., Durlik, I., Kostecka, E., Łobodzińska, A., & Matuszak, M. (2024). The emerging role of artificial intelligence in enhancing energy efficiency and reducing GHG emissions in transport systems. *Energies*, *17*(24), Article 6271. https://doi.org/10.3390/en17246271

Petrović, N., Bojović, N., Petrović, M., & Jovanović, V. (2020). A study of the environmental Kuznets curve for transport greenhouse gas emissions in the European Union. *Facta Universitatis, Series: Mechanical Engineering*, *18*(3), 513–524. https://doi.org/10.22190/FUME171212010P

Rečka, L., Máca, V., & Ščasný, M. (2023). Green Deal and carbon neutrality assessment of Czechia. *Energies*, *16*(5), Article 2152. https://doi.org/10.3390/en16052152

Schau, E. M., Golinucci, N., Beltrami, F., Rocco, M. V., Prina, M. G., Rinaldi, L., & Sparber, W. (2026). Impact of sufficiency on achieving the EU's 2030 and 2050 greenhouse gas emissions targets. *Sustainable Futures*, *11*, Article 101791. https://doi.org/10.1016/j.sftr.2026.101791

Šimpach, O. & Šimpachová Pechrová, M. (2025). Modely ekonomických a finančních časových řad. VŠE, FIS: Prague, 182 p. ISBN 978-80-245-2579-2

Contact

Ondřej Šimpach

Prague University of Economics and Business

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

ondrej.simpach@vse.cz