KAPSARC
مركز الملك عبدالله للدراسات والبحوث البترولية
King Abdullah Petroleum Studies and Research Center

Carbon dioxide emission baseline forecasts for Saudi Arabia using the Structural Time Series Model and Autometrics

- Jeyhun I. Mikayilov
- Coauthored with Anwar Gasim and Lester C. Hunt

- 7th AIEE Energy Symposium on Energy Security, December 16, 2022

Motivation

- As a party to the Paris Agreement, Saudi Arabia submitted a baseline emissions reduction target as part of its nationally determined contribution.
- However, Saudi Arabia's quantitative baseline scenario is not publicly available.
- Therefore, we use two different econometric methods within a univariate framework to generate baseline emissions forecasts for Saudi Arabia.
- In doing so, we extend current drivers, trends and policies into the future without making assumptions about certain factors, such as economic growth, in the coming decades.

Objective

- To inform climate policymaking by projecting Saudi Arabia's baseline CO_2 emissions.
- We use two different econometric methods: the STSM and Autometrics
- Using two different methods provides a robustness check, as each method has strengths and weaknesses
- Make forecasts until 2060
- Forecast combination as an "insurance policy."

Literature Review

- Köne and Büke (2010), 25 top emitters, including KSA. Makes forecasts until 2030.
- Alshammari (2020). Makes forecasts until 2050.
- Energy Policy Simulator (EPS; KAPSARC 2022), 2050.
- Climate Action Tracker (2022). It provides projections up to 2030.

Methodology

General to Specific (*Gets*) Modeling Approach

Autometrics - Multipath block-search Machine-learning algorithm

Structural Time Series Modeling (STSM) Approach

Model Specification

$$co_{2_t} = intercept + \alpha_1 co_{2_{t-1}} + \alpha_2 co_{2_{t-2}} + \alpha_3 co_{2_{t-3}} + \alpha_4 co_{2_{t-4}} + random\ error\ term. \qquad (1)$$

- ❑ **How is the trend modeled?**
- ❑ **How are interventions modeled?**

Autometrics

$$co_{2_t} = \alpha_0 + \alpha_1 co_{2_{t-1}} + \alpha_2 co_{2_{t-2}} + \alpha_3 co_{2_{t-3}} + \alpha_4 co_{2_{t-4}} + \sum_1^T \vartheta_i \, IIS_t + \sum_1^T \tau_i \, SIS_t + \sum_1^T \varphi_i \, DIIS_t + \sum_1^T \omega_i \, TIS_t + \varepsilon_t. \quad \text{(A1)}$$

Here,

IIS_t = impulse-indicator saturation, which takes a value of one at time *t* and a value of zero otherwise;

SIS_t = step-indicator saturation, which takes a value of one until time *t* and a value of zero otherwise;

$DIIS_t$ = differenced impulse-indicator saturation, which takes a value of one at time *t*, a value of -1 at time *t+1* and a value of zero otherwise;

TIS_t = trend-indicator saturation, which takes a value of zero after time *t+1*, a value of -1 at time *t*, a value of -2 at time *t-1*, a value of -3 at time *t-2* and so forth.

$\alpha_i, \vartheta_i, \tau_i, \varphi_i, \omega_i$ are regression coefficients to be estimated, and ε_t is a random error term with $\varepsilon_t \sim NID\ (0, \sigma_\varepsilon^2)$.

Structural Time Series Model

$$co_{2_t} = \gamma_t + \alpha_1 co_{2_{t-1}} + \alpha_2 co_{2_{t-2}} + \alpha_3 co_{2_{t-3}} + \alpha_4 co_{2_{t-4}} + \varepsilon_t. \qquad \text{(A2)}$$

Here, γ_t is the stochastic trend, which is also interpreted as a time-varying intercept, and ε_t is a random error term with $\varepsilon_t \sim NID\ (0, \sigma_\varepsilon^2)$.

The stochastic trend consists of a level μ_t and a slope β_t, which are defined as follows:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \qquad \text{(A3)}$$

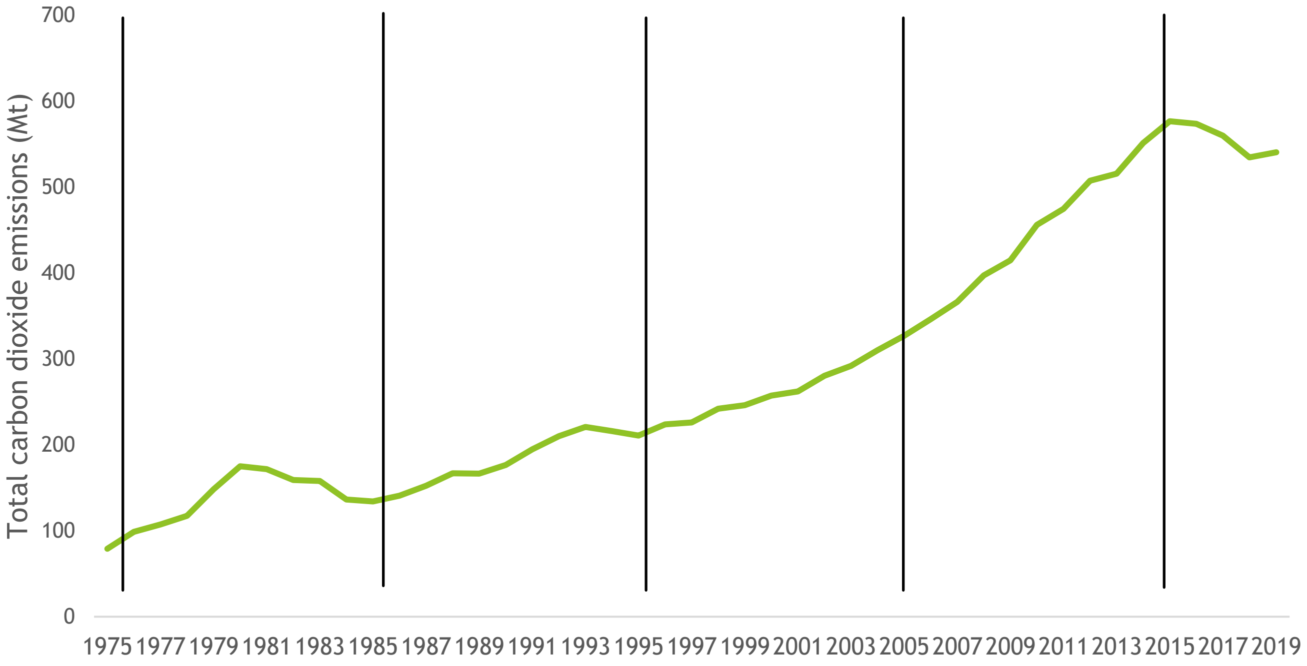
$$\beta_t = \beta_{t-1} + \xi_t. \qquad \text{(A4)}$$

Here, $\eta_t \sim NID\ (0, \sigma_\eta^2)$ and $\xi_t \sim NID\ (0, \sigma_\xi^2)$ are mutually uncorrelated random disturbance terms. If the variance of either η_t or ξ_t is zero, then that component of the trend is deterministic. If both hyperparameters are zero, then the stochastic trend collapses into a deterministic trend.

Data, CO2 Emissions, Mt

The time series data on emissions for Saudi Arabia is from Enerdata (2022). Sample 1984-2019.

We model total CO_2 emissions, excluding those associated with land use, land use change and forestry (LULUCF).



Source: Enerdata 2022.

Estimation Results

Automatrics Results

$$\widehat{co_2}_t = 0.1421^{*} + 1.2878^{***} co_{2_{t-1}} - 0.3079^{**} co_{2_{t-2}} - 0.1857^{***} IIS_{1984}, \quad (2)$$

STSM Results

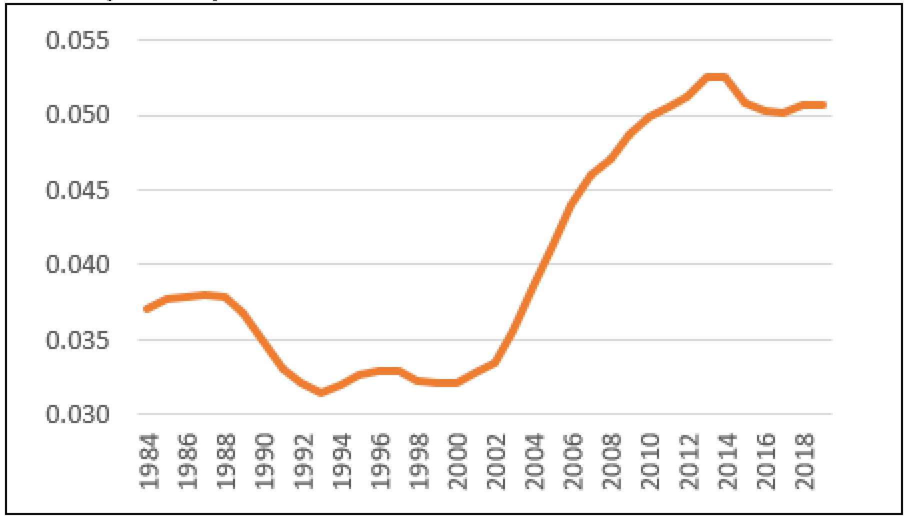
$$\widehat{co_2}_t = \hat{\gamma}_t + 0.5803^{***} co_{2_{t-1}} - 0.6591^{***} co_{3_{t-3}}. \quad (3.1)$$

The estimated trend ($\hat{\gamma}_t$) is given by

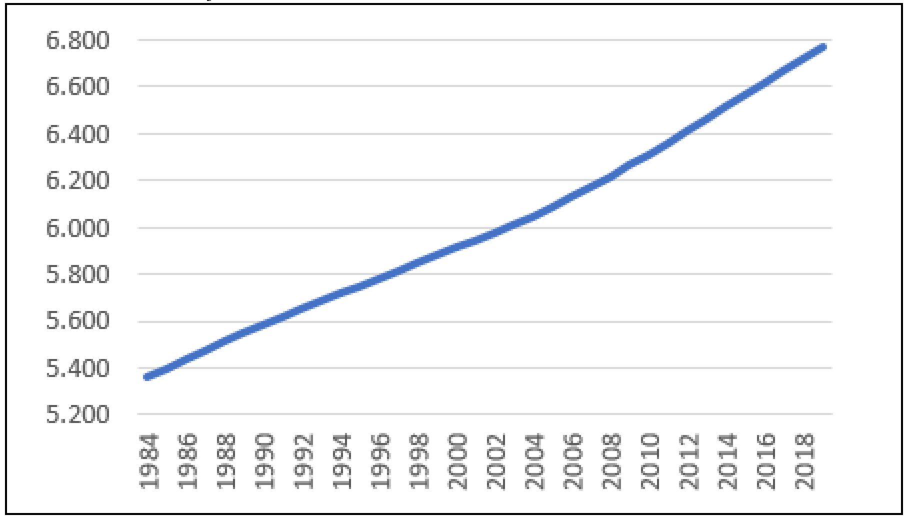
$$\hat{\gamma}_t = \hat{\mu}_t - 0.0787^{***} Lvl_{1987} - 0.0638^{***} Irr_{1989} + 0.0882^{***} Lvl_{1991} + 0.0711^{***} Irr_{1996} + 0.0163^{*} Slp_{2001} + 0.0300^{**} Irr_{2010} - 0.0525^{***} Slp_{2015}. \quad (3.2)$$

Estimated slope (β) and level (μ) of the preferred STSM specification

A: Slope component of the estimated trend



B: Level component of the estimated trend

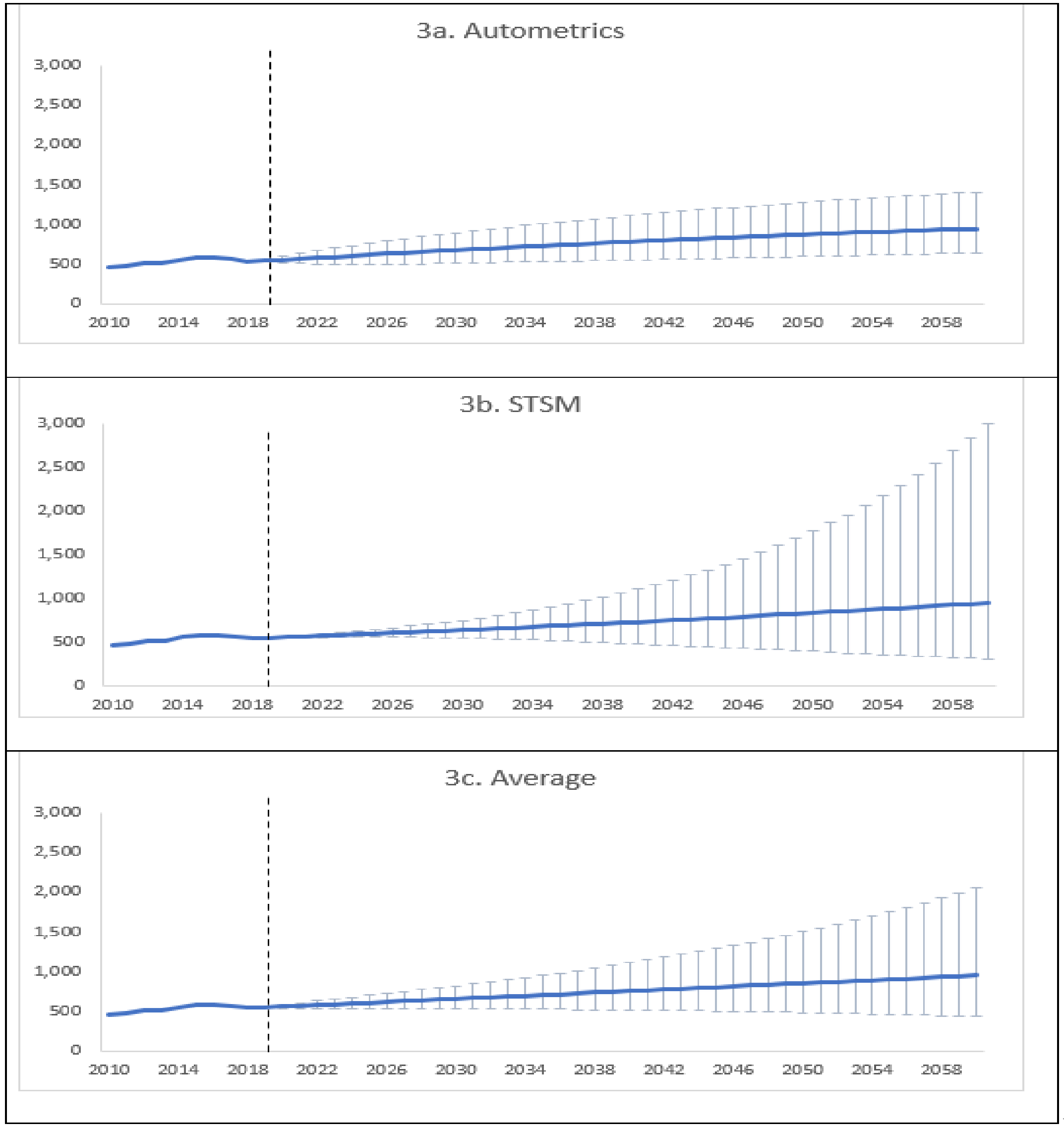


Forecast Results
Total CO2 emissions (in Mt)

Year	Autometrics	STSM	Average
1984	136.2	136.2	136.2
1990	176.5	176.5	176.5
2000	257.3	257.3	257.3
2010	455.7	455.7	455.7
2019	540.4	540.4	540.4
2025	611.8	595.1	603.4
2030	670.8	632.2	651.2
2035	726.1	680.1	702.8
2040	777.5	726.5	751.6
2050	867.5	831.0	849.0
2060	940.9	947.9	944.4

- Source: Authors' calculations.
- Note: Values reflect historical data up to 2019 and projections after that.

Forecast Results
Total historical CO2 emissions (2010-2019) and projections (2020-2060) in Mt



- Source: Authors' calculations

Conclusion

This study provides a robust baseline projection for Saudi Arabia's CO2 emissions.

651 Mt (522 to 817) by 2030, the target year of Saudi Arabia's NDC, in a baseline scenario.

944 Mt by (585 to) 2060, the year in which Saudi Arabia aims to achieve net-zero emissions.

KAPSARC

مركز الملك عبدالله للدراسات والبحوث البترولية

King Abdullah Petroleum Studies and Research Center

Preferred Autometrics specification

Time period	**1984-2019**		
Estimated coefficients		***Residual diagnostics***	
α_0	0.1421*	*AR(1-2)*	0.76
α_1	1.2878***	*ARCH (1-1)*	0.23
α_2	-0.3079**	*Normality*	2.75
α_3	-	*Hetero*	0.69
α_4	-	*Hetero-X*	0.54
Interventions/indicator		*RESET*	2.00
Impulse 1984	-0.1857***		
Goodness of fit			
R^2	0.995		
$\bar{R}^2$	0.994		
F	*2107*		

Source:Notes: *, ** and *** represent significance at the 10%, 5% and 1% levels, respectively. R^2 is the coefficient of determination, and $\bar{R}^2$ is the adjusted coefficient of determination. *F* is the overall goodness-of-fit statistic and follows an $F_{(3, 23)}$ distribution. *AR(1-2)* is the second-order autocorrelation statistic, which follows an $F_{(2, 30)}$ distribution. *ARCH (1-1)* is the first-order autoregressive conditional heteroskedasticity statistic, which follows an $F_{(1, 34)}$ distribution.

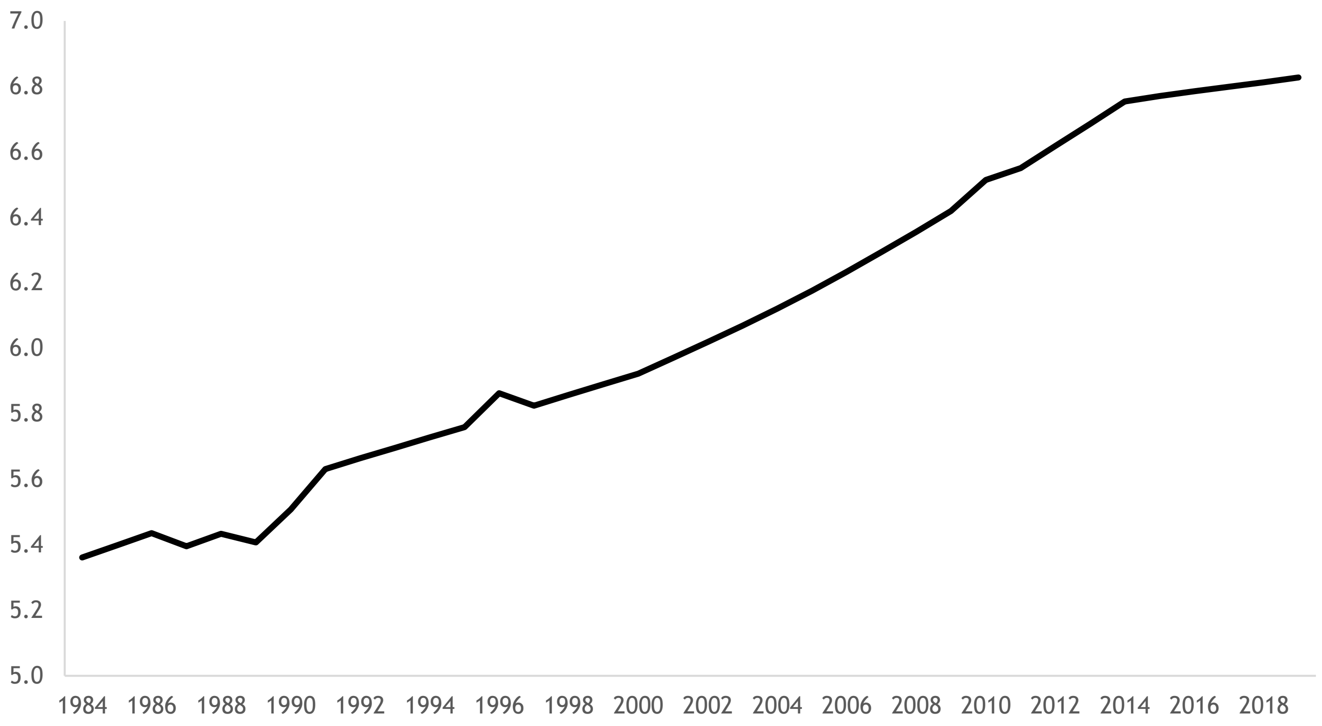
Normality is reflected by the Doornik and Hansen statistic, which follows an approximate $\chi^2_{(2)}$ distribution. Hetero and hetero-x are heteroscedastic statistics that follow $F_{(4, 30)}$ and $F_{(5, 29)}$ distributions, respectively. *RESET* is the Ramsey RESET statistic, which follows an $F_{(2, 30)}$ distribution.

Preferred STSM specification

Time period	**1984 to 2019**		
Estimated coefficients		***Residual diagnostics***	
α_1	0.5803***	*Normality*	0.66
α_2	-	*H(8)*	4.46
α_3	-0.6591***	*r(1)*	-0.12
α_4	-	*r(2)*	-0.02
Interventions		*r(3)*	0.13
Level 1987	-0.0787***	*r(6)*	0.03
Irregular 1989	-0.0638***	*DW*	2.16
Level 1991	0.0882***	*Q(6,4)*	1.85
Irregular 1996	0.0711***	***Auxiliary residuals:***	
Slope 2001	0.0163*	*Normality - Irregular*	0.64
Irregular 2010	0.0300**	*Normality - Level*	0.84
Slope 2015	-0.0525***	*Normality - Slope*	1.03
Trend components	Fixed level	***Prediction failure***	11.08
	Stochastic slope		
Goodness of fit			
p.e.v.	0.00023634		
AIC	-7.6836		
BIC	-7.1557		
R^2	0.999		
R_d^2	0.881		

Source:Notes: *, ** and *** represent significance at the 10%, 5% and 1% levels, respectively. R^2 is the coefficient of determination, and R_d^2 is the coefficient of determination based on differences. *p.e.v.* is the prediction error variance. *AIC* is the Akaike information criterion, and *BIC* is the Bayesian information criterion. Normality is determined based on the Bowman-Shenton statistic, which follows an approximate $\chi^2_{(2)}$ distribution. *H(8)* is a heteroscedasticity statistic that follows an $F_{(8,\,8)}$ distribution. *r(1)*, *r(2)*, *r(3)* and *r(6)* are the serial correlation coefficients at the equivalent residual lags and are approximately normally distributed. *DW* is the Durbin-Watson statistic. *Q(6,4)* is the Box-Ljung statistic, which follows a $\chi^2_{(4)}$ distribution. The prediction failure is a predictive failure statistic that follows a $\chi^2_{(7)}$ distribution.

Estimated trend (γ) from the STSM preferred specification



Source: Authors' calculations.