v1.1 (minor revision to first version)
32nd International Input-Output Association Conference, Seville, Spain June 22-26, 2026

Reconciling Data Years in EEIO Models and Resulting Emission Factors: Alternatives and Best Practices

Wesley Ingwersen, Ph.D.[1,*]
Mo Li, Ph.D.[2]
Fatemeh Kiani Salmi [3]
Reed Miller, Ph.D. [3]
Scott Matthews, Ph.D. [4]
Yi Yang, Ph.D.[5]
Sangwon Suh, Ph.D. [2, 6]

[1] Stanford University
[2] Watershed Technology, Inc.
[3] University of Maine
[4] Keylogic, Inc.
[5] Chongqing University
[6] Tsinghua University

* corresponding author: ingwersw@stanford.edu

Introduction

Developing environmentally-extended input-output (EEIO) models requires pairing economic input-output tables (IOTs) and environmental data for the same regions, sectors, and years. IOTs are often not available for the same year(s) as the environmental data. And furthermore, for applications like the development of Emissions Factors (EFs) from these models that are used along with organizational spend data to compute organizational footprints, neither environmental data nor IOTs may be available to match the year of the spend data.These differences in years of data presents an obvious conundrum for model developers who wish to provide final factors that are temporally relevant to users as well as a challenge to communicate the year(s) that the EFs represent. While creation of models with mixed year data is very common in EEIO models due to data scarcity, the specific adjustment procedures, their interpretation, and clear best practices in reconciling mixed year data are missing.

Objective and Novelty

The objective of this paper is to describe common adjustment procedures used to integrate data from different years into a EEIO model, provide interpretations in understandable language, demonstrate their effects with real data, and to provide some best practices for consideration to the broader IO and EEIO community.

Approach

We approach this paper in the style of a problem-solution essay, using a situation-problem-solution-evaluation (SPSE) structure. We provide an example situation of differences in data years in EEIO data along with the expectations of the model, and we break this down into a series of problems, for which we present alternative solutions both descriptively and quantitatively, and evaluate the merits of each.

Situation

We intend to estimate annual embodied GHG emission factors (EFs) for products in the US in the format of kg CO2e/USD in producer price for the time series of 2019 to 2024. To develop these estimates we will construct single-region environmentally-extended input-output models, in commodity x commodity format, with greenhouse gas emissions (GHG) extensions, using the most detailed level IOT in Producer price (PRO) and, optionally, associated industry output and industry price indices. We assume we have correspondences in place between industry and commodity definitions between the IOT and related datasets so wrangling the datasets into a common commodity classification is not an issue. Although we acknowledge that single-region models have scope limitations and do not recommend them for developing EFs, they provide a sufficient context for this analysis and the take-aways can be extended to MRIOs.

Key data sets available for use in construction of the model are given in Table 1 by region along with the most recent year they are available.

Table 1. Key data available for use in MRIO construction.

Data	**Data Years**[1]	**Frequency/Delay in Release**
IOT, detailed industries and commodities (PRO)	2017	Every 5 yrs/ Delay 5 yrs
Producer price index and gross output (PRO), benchmark industries	2017-2024	Every 1 yr/ Delay 0.8 yrs
IOT, aggregate industries and commodities	2017-2024	Every 1 yr/ Delay 0.8 yrs
National GHG emissions inventory	2017-2023	Every 1 yr/Delay 2 yrs

[1] Data availability as of Apr 14, 2026

Problem

None of the critical data used to construct the EEIO model are available for the most recent year that we are targeting, 2024. Neither IOT nor GHG data are available for the same years over the time period.

Solutions

We propose general solutions indicated by the primary number. For each solution we calculate **A**, **L**, **q**, **B**, **d** and **n** matrices and vectors. All solutions share the same **E**.

The matrices created as defined in Table 2.
Table 2. Nomenclature Used in This Study

Matrix / vector	**Definition**	**Basic Derivation**
A	Direct requirements matrix (\$/\$)	$Ux^{-1}\,Vq^{-1}$
α	Vector of scaling ratios for A matrix derived using aggregate data	AA'^{-1}
B_i	Direct emissions matrix by industry (kg CO2e/\$)	$E\,\hat{x}^{-1}$
d	Direct emissions vector by commodity (kg CO2e/\$)	$B_i Vq^{-1}$
E	Environmental (GHG) totals by industry matrix in kgCO2e	Attributed from GHGI to sectors using sector attribution model
e (c)	Environmental totals by commodity in CO2e	$d\,\hat{q}$
L	Leontief total requirements matrix	$(I - A)^{-1}$
n	Emission factor vector in PUR	BL
q	commodity output vector	$i\,V$
ρ	Inflation adjustment vector (by commodity for given year)	
Π	Price index matrix (industry x year)	given

U	Use table	given
V	Supply (Make) table	given
x	Industry output vector	given
y	Final demand vector	given or calculated as $q - Aq$

Solution 1 - Use data from last common reporting year (2017)

In this option only 2017 data are used, resulting in one set of EFs used for the entire time period. This scenario represents a choice of practitioners that are uncomfortable with mixing data years and choose to only develop emission factors from a year of common data.

Solution 2 - Use all data without adjustment

In this option the data are used as is with the latest data of each type used for each year without adjustment. For the B matrix the annual GHG data are used with the same year annual industry output and they are used along with the 2017 IOT as given. The scenario represents a choice to use available data but not to adjust them.

Solution 3 - Inflate only. Use all detailed deflators to a) adjust **B** to match original **A** year or b) adjust **A** to match **B** year

a) Use IOT as is to make the **A** matrix in 2017 dollars and use annual GHG data with annual output to make the **B** matrix; then deflate the **B** matrix to IOT year (2017)
b) For each year, deflate the **A** matrix to current year dollars and use annual GHG data with annual output to create the **B** matrix in current year dollars.

Solution 4 - Scale **A** and **q** and **x** using aggregate data

This solution uses the more aggregate IOT to estimate an **A**, **q**, **x** and a normalized **V** that can be used). This approach scales these matrices and vectors using ratios of the related aggregate data for the target year using ratios of the respective aggregate data over the time period, in the general pattern for example of (2022 aggregate/2017 aggregate) * 2017 detail.

Solution Treatments

An assumption across all scenarios is that the final factors used by practitioners are converted into the monetary year matching their spend data, the so-called reporting year. We assume here that the reporting year is 2024 and therefore the EFs are converted into kg CO2e per 2024 USD if the final result is not already in that currency year. For example the adjustment to a 2024 USD for 2017 USD EFs is performed with the following equation:

$$n\ (kg\ CO2e/USD\ 2024)\ =\ n(kg\ CO2e/USD\ 2017)\ \rho\ (USD\ 2024/USD\ 2017)$$

where ρ values are unique to every commodity in the **n** vector.

Evaluation

As part of the evaluation we consider both the theoretical and practical ramifications of the solutions. Then we evaluate the quantitative effects of the solutions on the time series of emissions factors and related variables that are produced by each solution. The major questions that we attempt to answer are:
1) What does an approach theoretically mean?
2) Is the approach theoretically and mathematically correct?
3) What are the results of each approach on the direct and indirect emission factors over the time series?
4) Do the different approaches have different implications on the stability of emission factors?

Theoretical and Practical Considerations

Solution 1 - The use of a common data year avoids the problems associated with the data year reconciliation, and therefore the integration of data from the same year, assuming the geographical and technological scopes of the economic and environmental data are the same, is sound. However, practically, the data provide no real measure of change in the direct or indirect emissions intensities of commodities, skipping out on the real changes in emissions or production input intensity occurring over the time period. The change in EFs as a result of a conversion to a current year dollar for use in emission reporting may mislead the user into perceiving a change in real emission intensity.

Solution 2 - This solution makes use of annual GHG emissions data and annual industry output data. The production inputs are held constant.
Theoretically the implicit assumption in the **A** and **L** matrices are that the direct and indirect production inputs are constant across years in price but not in quantity. This is an implicit assumption of *perfect price inelasticity* in the input structure because the input dollar values per output dollar values are constant.

However an issue results in a mixing of units in the calculation of **n**, for instance for year 2023

$$n\ (kg\ CO2e\ 2023/USD\ 2023)\ \neq\ d\ (kg\ CO2e\ 2023/USD\ 2023)\ \ L\ (USD\ 2017/USD\ 2017)$$

This invalidates the solution on purely mathematical grounds because non-equivalent units are used in the calculation. But it also is theoretically invalid, because the physical quantity of the commodity input per output in the **A** matrix is not consistent with the **d** matrix emission factor quantity for the emission value. For instance, if $1 of commodity **A** in $2017 has a value of

\$1.25 in \$2023, then the **d** emission factors will be 1/1.25 or 80% lower than they must be to properly account for the input quantity.

Solution 3a - Like solution 2, this solution makes use of the annual GHG emissions data and annual industry output data. Also like solution 2, the **A** matrix is constant, applying the assumption of *perfect price inelasticity* in the input requirements in **A** over the time span.

$$n\ (kg\ CO2e\ 2023/USD\ 2017)\ =\ d\ (kg\ CO2e\ 2023/USD\ 2017)\ \ L\ (USD\ 2017/USD\ 2017)$$

Before preparing **d** In the above equation, industry output in 2023 is deflated to \$2017 using the following equation:

$$x = x_{2023} \frac{\Pi_{\$\ 2017}}{\Pi_{\$\ 2023}}$$

and then used to create B_i, then **B** and finally **d** using the equations in Table 2.

However, unlike in Solution 2, there are no mathematical or practical problems with the calculation of n because the units of the calculation are consistent.

Solution 3b - Like with solution 2 and 3, the annual GHG and industry output data are used. But in this case, **A**, and not **B**, is deflated to match the target year using the following equations, using again the example year of 2023:

$$A_{2023} = \hat{\rho}^{-1} A_{2017}\, \hat{\rho}$$

$$\rho = \frac{\Pi_{2017}}{\Pi_{2023}} V \hat{q}^{-1}$$

The implicit assumption in the deflation of **A** is of *perfect price elasticity*, meaning that producers will maintain the same quantitative input structure regardless of the price of the inputs.

$$n\ (kg\ CO2e\ 2023/USD\ 2023)\ =\ d\ (kg\ CO2e\ 2023/USD\ 2023)\ \ L\ (USD\ 2023/USD\ 2023)$$

Mathematically and theoretically this solution is also sound. Solution 3b, by assuming a constant physical input quantity across the time span, is more consistent with what practitioners in life cycle assessment (LCA) use in life cycle inventory data, which is to reuse an data for a production process with set ratios of physical inputs to outputs over a study period, or otherwise are incognizant of the temporal matching of the data used and the study target year.

Solution 4

The scaling of the **A** matrix using the aggregate tables is done as follows.
The aggregate table ($\mathbf{A_s}$) for the target year (e.g. 2023) is inflated to be in the base year (2017) USD. A matrix of ratios (**α**) between values in the aggregate $\mathbf{A_s}$ for the target year to the aggregate $\mathbf{A_s}$ in the base year is created and then transformed in the detail level schema using a detail-to-summary commodity correspondence matrix $\boldsymbol{O_{ds,\,c}}$. Finally the **α** matching the original base year detailed **A** is multiplied element-wise.

$$A_{s2023,\,2017\,\$} = \hat{\rho}^{-1} A_{s2023}\, \hat{\rho}$$

$$\alpha_{s2023\,/\,s2017} = A_{s2023,\,s2017\,\$} \circ A^{-1}_{s2017,\,s2017\,\$}$$

$$\alpha_{2023\,/\,2017} = O_{ds_c} \alpha_{s2023\,/\,s2017} O'_{ds_c}$$

$$A_{2023,\,2017\,\$} = \alpha \circ A_{2017}$$

Similar scaling is performed for the **V** using this method from which **q** is derived as the column sums of the scaled **V**.

Quantitative Evaluation

We calculated direct and direct (***d***) + indirect emissions factors (***n***) for 405 US consumed commodities based on the schema used in the US 2017 benchmark IOT for each solution, using a unique EEIO model for each, hence solution ≈ model.
We inflate the final **d** and **n** EF into a common dollar year of 2024 to show comparable results across the time series.

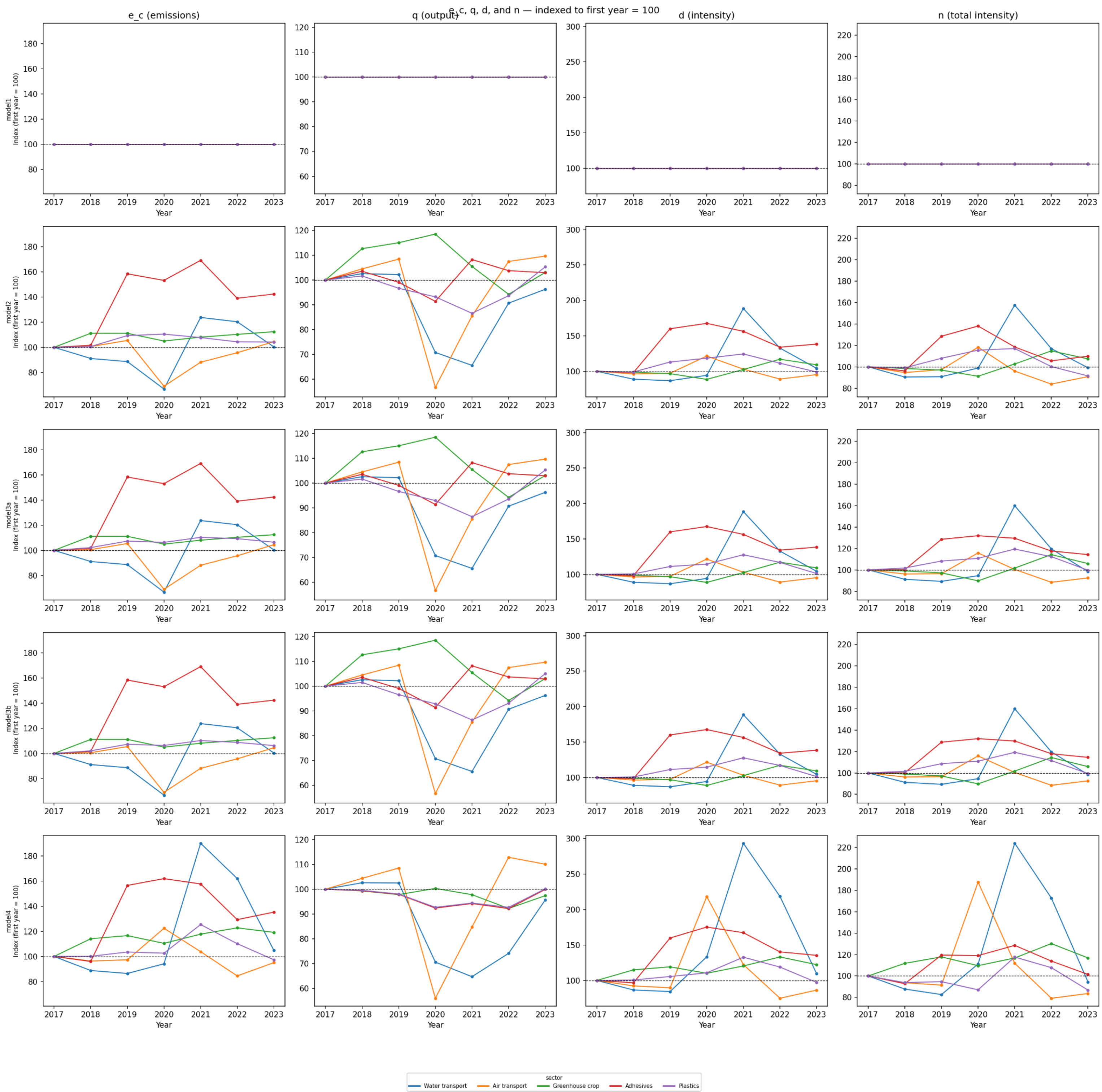

Figure 1. **e**, **q**, **d**, and **n** relative change over time series (columns) for 5 commodities with greatest year-over-year change for all solutions (models).2017 values for each variable are indexed to 100 for 2017. Blue = Water transport; Orange = Air transport; Green = Greenhouse crops; Red= Plastics; Purple = Adhesives.

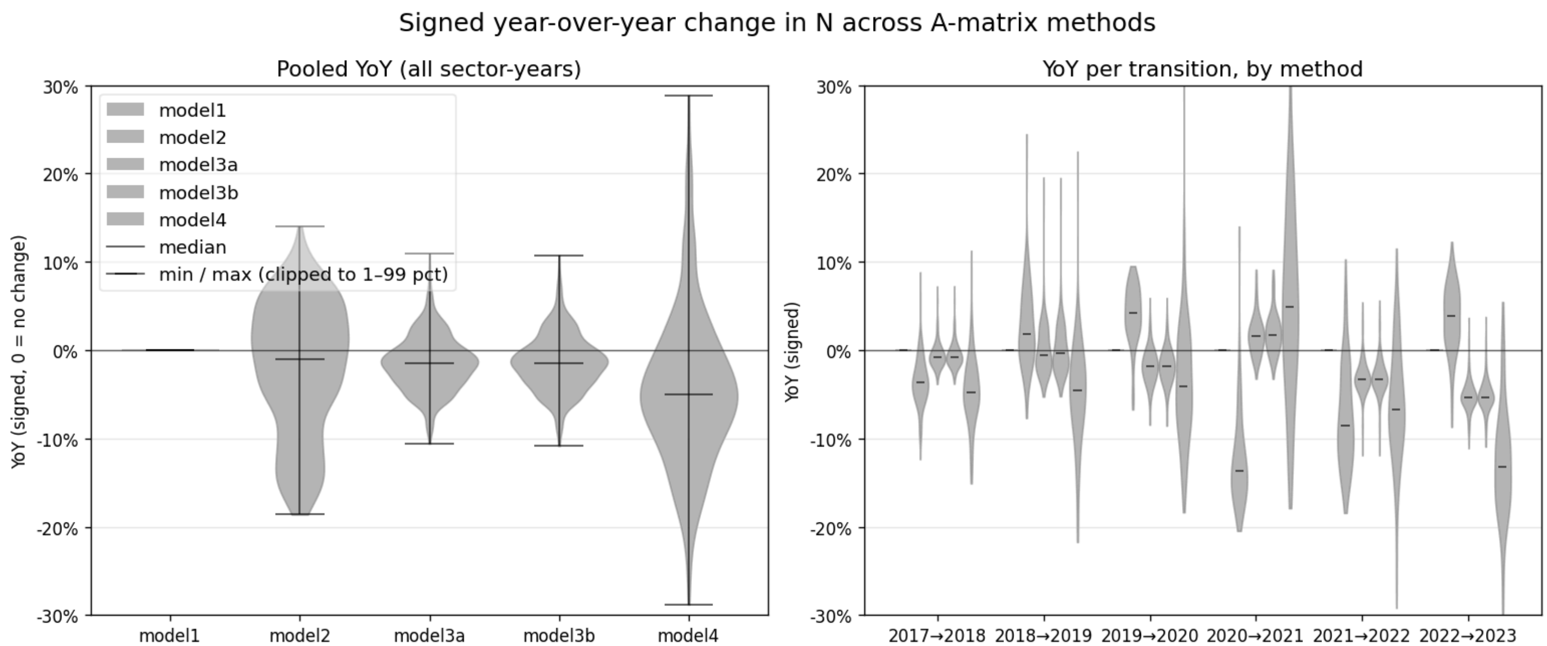

Figure 2. Mean total year-over-year change in **n** (a) by model. Year-over-year total change in **n** in models (b).

Figure 1 shows results of sectors for solutions (models) 1, 2, 3a, 3b, and 4 from 2017 to 2023 in USD 2024 that show the greatest variation across the time series across the models. For each data series in the columns (**e, q, d** and **n**), 2017 values for each sector are pegged to 100 to enable easier comparison of trends. Trends for the sectors with the greatest cumulative interannual change in direct + indirect emission factors **n** are shown. The progression from left to right moves from the key model intermediate data products to the final result: Emissions by commodity (**e**) and commodity output (**q**) to their relationship result for the sectors and the relationship between the two in the direct emissions factors (in **d** in GHG per dollars) and then finally to the direct + indirect emission factor result (**n**).

E is identical for all models but the indexing of ***e*** by commodity is based on ***q**,* so while **E** totals are the same, the differences in **q**, visible in the second columns, are reflected as differences in **e**. As we progress to the **q** series, we see that **q** trends appear identical across these models for the 5 commodities over the time series. This can be explained by the fact that the annual detailed gross output is the starting point for **q** estimation for model 2, 3a and 3b. Model 4 instead uses the scaling approach based on data derived from the aggregate IOT. Solution/model 4 appears to mute the effects on **q** for some sectors (Plastics, Adhesives, Greenhouse Crops), while it has little/no effect for others (Air transport, Water transport).

The effects of the models 2, 3a and 3b on **d** and **n** for the selected sectors is very similar. Model 4 **d** and **n** has departures from the other models, especially for the Air and Water transport sectors. Model 4 shows Air transport double in EFs in one year and then fall back toward the starting EF in 1-2 years. Water transports triples in one year before falling back to the baseline.

Figure 2a shows the average year-over-year fluctuation in the **n** values for models 2, 3a, 3b and 4 (model 1 has no change across time) and Figure 2b the change on a year by year basis. Models 2 and 4 have a greater spread of change in **n** over the full time series and on an annual basis. Models 4’s changes are more concentrated over the full time period (Figure 2a) are more concentrated around a ~5% decrease while model 2’s changes appear to be more random although some data points are less extreme than in model 4.

In order to explore the trends in individual EFs over the time period, we did a year-over-year (YoY) correlation of *Δ**e*** (change in e) and *Δ***q** (change in **q**). The expected outcome would be that they would move in the same direction, although if carbon intensity is decreasing, that it would not necessarily be at the same rate (if there is a change in carbon intensity). The result is in Table 2. The models show a weak correlation between the two variables over the full time series, generally with a weak negative correlation for model 4.

Table 2. Year-over-year Pearson correlation coefficient between *Δ**q*** and *Δ**e***

model	**transition**	**value**
model2	2017-2018	0.253369
model2	2018-2019	0.064674

model	transition	value
model2	2019-2020	0.183832
model2	2020-2021	0.042434
model2	2021-2022	0.162197
model2	2022-2023	-0.00487
model3a	2017-2018	0.248762
model3a	2018-2019	0.063782
model3a	2019-2020	0.178799
model3a	2020-2021	0.041303
model3a	2021-2022	0.149997
model3a	2022-2023	-0.00562
model3b	2017-2018	0.248891
model3b	2018-2019	0.063653
model3b	2019-2020	0.179124
model3b	2020-2021	0.041489
model3b	2021-2022	0.151004
model3b	2022-2023	-0.00548
model4	2017-2018	-0.21623
model4	2018-2019	-0.04229
model4	2019-2020	-0.14517
model4	2020-2021	-0.08735
model4	2021-2022	-0.16231
model4	2022-2023	-0.06503

Conclusion

EEIO models may leverage datasets from different years, but the data must be appropriately treated, particularly to link the environmental data year with the economic output year, and also to treat and link the requirements tables with the environmental intensity matrix. To fail to integrate more current data risks inaccurate results due to real change. To fail to harmonize the data years (to use more current data for a component but not to acknowledge or treat that) gives models with incompatible data.

The key findings from our trial application of different solutions to a US-based EEIO model time series as are follows:

1. Where data from different years are mixed, the initial **B** matrix relationship must use environmental and economic data from a common year and the pairing of the **A** with the **B** matrix must be deflated to common monetary years before combining.
2. Whether deflating the **B** matrix back in time to match the original **A** matrix year, or inflating the **A** matrix forward in time to match the **B** year, which imply different theoretical assumptions, the results are quantitatively the same.
3. The use of aggregate data of a more recent vintage ideally provides more contemporary data for building detailed level EFs, but can lead to EFs with increased instability when aggregate data show significant interannual change.
4. These approaches to reconcile data of more recent years will not help create more explainability or stability in EF trends, especially when there is unexplainable interannual increase followed by decrease in EFs.

Source Code Availability

The EEIO models were constructed using the Cornerstone bedrock repository. The source code that builds the model for all years and prepares the figures and statistics described in this paper is available at
https://github.com/cornerstone-data/bedrock/tree/reconciling_data_years_final/bedrock/analysis/reconciling_data_years