Supplemental Appendix: Cleansing Oil

Rui Sousa[*]

Department of Economics,
Northwestern University

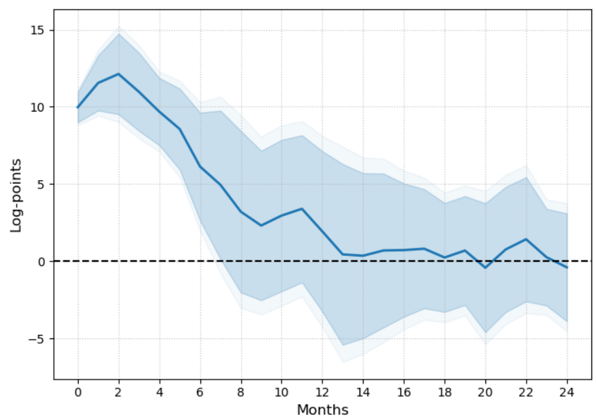
June 13, 2026

A Empirical Validation of Oil Shock Identification

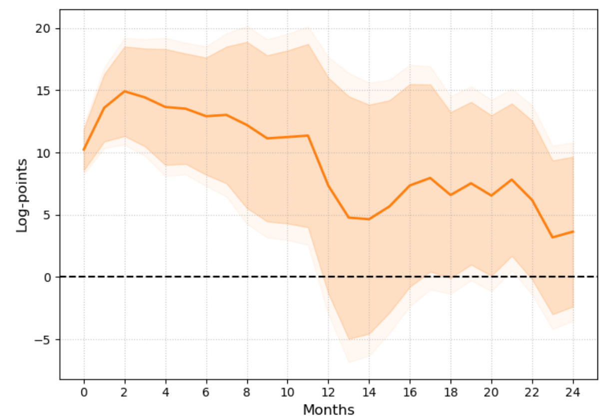
Figure A.1 compares the impulse responses functions generated from estimating equation (1) for the real WTI oil price and OECD+6 industrial production index using both the oil news shock of Känzig (2021) and the structural oil supply shock of Baumeister and Hamilton (2019), both normalised to a 10% increase in oil prices on impact. The two series generate virtually identical price dynamics: oil prices rise sharply on impact and decay to zero within eight months, so agents face the same price signal regardless of which shock is used. The two series diverge, however, in their implied economic activity effects. The news shock delivers the expected contractionary response: industrial production falls on impact and remains depressed for the full two-year horizon. The structural supply shock, by contrast, yields a positive and statistically significant increase in industrial production over the 2002–2024 sample — the opposite of what theory and the broader literature predict.

[*]Contact: *rui.agm.sousa@u.northwestern.edu*.

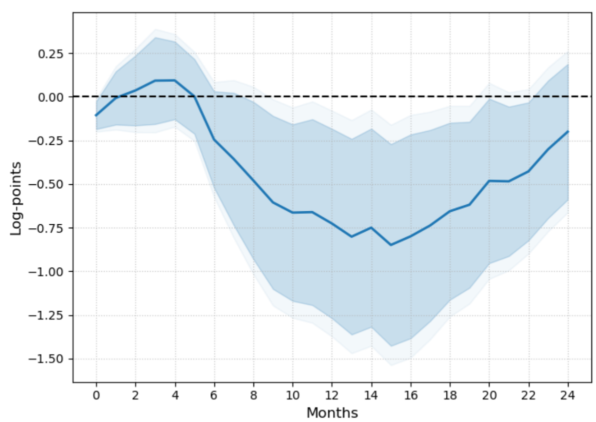
Figure A.1: Oil News Shock vs. Structural Supply Shock: Price and Activity Responses



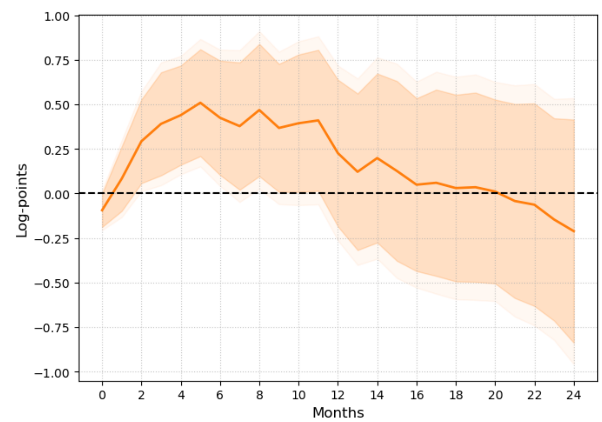
(a) Real WTI — Oil News Shock



(b) Real WTI — Structural Supply Shock



(c) Industrial Production — Oil News Shock

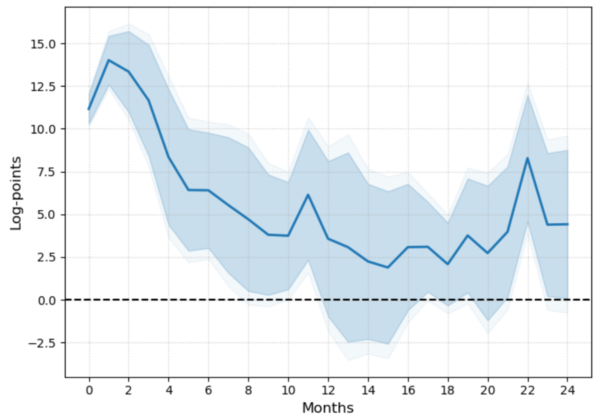


(d) Industrial Production — Structural Supply Shock

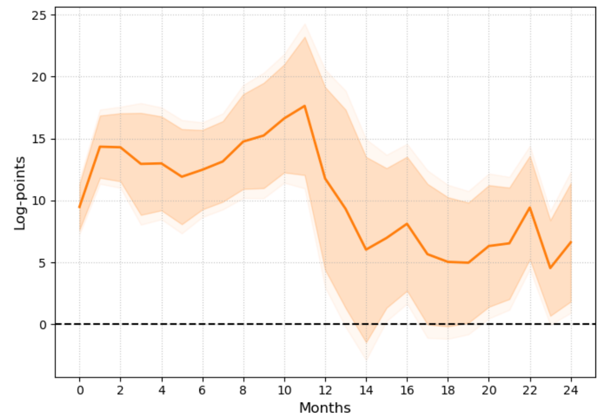
Notes: Impulse responses to each shock series, normalised to a 10% increase in oil prices on impact. Left column: oil supply news shock of Känzig (2021). Right column: structural oil supply shock of Baumeister and Hamilton (2019). Top row: real WTI oil price (log-points). Bottom row: OECD+6 industrial production (log-points). Shaded bands correspond to 90% and 95% confidence intervals based on HAC standard errors. COVID-19 (2020) is excluded. Shocks: January 2002 – December 2024; outcome data extends to December 2025.

There are many reasons why this could be happening. One is the contamination of the shocks, as argued by Mori and Peersman (2024). Nevertheless, including more financial controls does not solve the problem. When using a shorter sample starting in 2010, consistent with the panel data, the problem becomes less severe, though the response is more delayed than what has been documented in the literature. The results are plotted in figure A.2. The oil news shock, by contrast, generates coherent dynamics across the samples used in the main analysis: starting in 2002, when most time-series begin, or in 2010, when panel coverage broadens significantly, the shock induces the same response in industrial production and oil prices, albeit with different persistence. I therefore use the oil news shock throughout to ensure consistency.

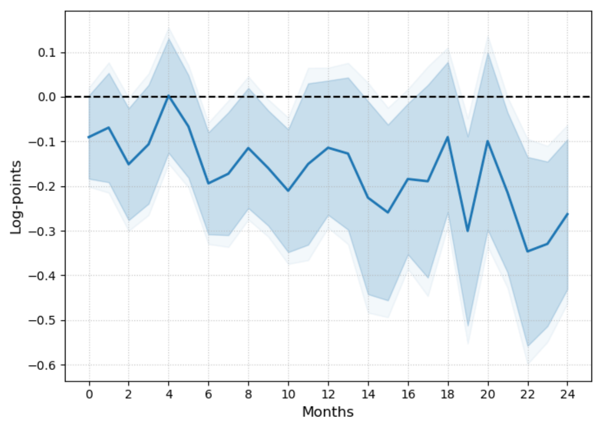
Figure A.2: Oil News Shock vs. Structural Supply Shock: Price and Activity Responses — 2010 Sample



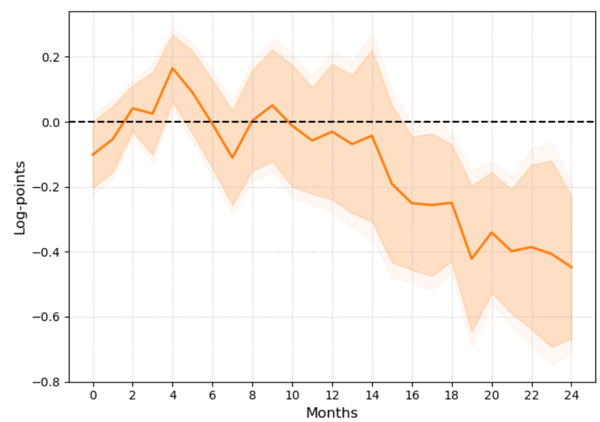
(a) Real WTI — Oil News Shock



(b) Real WTI — Structural Supply Shock



(c) Industrial Production — Oil News Shock



(d) Industrial Production — Structural Supply Shock

Notes: Impulse responses to each shock series, normalised to a 10% increase in oil prices on impact. Left column: oil supply news shock of Känzig (2021). Right column: structural oil supply shock of Baumeister and Hamilton (2019). Top row: real WTI oil price (log-points). Bottom row: OECD+6 industrial production (log-points). Shaded bands correspond to 90% and 95% confidence intervals based on HAC standard errors. COVID-19 (2020) is excluded. Shocks: January 2010 – December 2024; outcome data extends to December 2025.

B Construction of the CO_2-Equivalent Concentration Index

The time-series analysis uses a CO_2-equivalent (CO_2e) concentration index that aggregates the atmospheric burden of four major greenhouse gases—CO_2, CH_4, N_2O, and SF_6—into a single, consistent metric. Monthly global-mean concentrations of CO_2 (ppm), CH_4 (ppb), N_2O (ppb), and SF_6 (ppt) are obtained from NOAA.

Radiative forcing. Each gas contributes to total radiative forcing (RF) relative to its pre-industrial (1750) reference concentration. I use the simplified stratospheric-temperature-adjusted RF (SARF) expressions of Etminan et al. (2016) as adopted in Intergovernmental Panel on Climate Change (IPCC) (2023). Let C, M, N, S denote current concentrations of CO_2 (ppm), CH_4 (ppb), N_2O (ppb), and SF_6 (ppt), with pre-industrial reference values $C_0 = 278.3$, $M_0 = 729.2$, $N_0 = 270.1$, $S_0 = 0$. The individual SARF terms ($\mathrm{W\,m^{-2}}$) are:

$$\Delta F_{\mathrm{CO_2}}(C,N) = \left[\alpha(C) + c_1\sqrt{N}\right]\ln\left(\frac{C}{C_0}\right)\cdot e_1, \qquad \alpha(C) = d_1 + a_1(C-C_0)^2 + b_1(C-C_0) \tag{1}$$

$$\Delta F_{\mathrm{CH_4}}(M,N) = \left[a_2\sqrt{M} + b_2\sqrt{N} + d_2\right]\left(\sqrt{M} - \sqrt{M_0}\right)\cdot e_2 \tag{2}$$

$$\Delta F_{\mathrm{N_2O}}(C,M,N) = \left[a_3\sqrt{C} + b_3\sqrt{N} + c_3\sqrt{M} + d_3\right]\left(\sqrt{N} - \sqrt{N_0}\right)\cdot e_3 \tag{3}$$

$$\Delta F_{\mathrm{SF_6}}(S) = \omega\,(S - S_0) \tag{4}$$

The scalars e_1, e_2, e_3 are the stratospheric temperature adjustment factors that convert instantaneous RF to SARF. Total forcing is $\Delta F_{\mathrm{total}} = \sum_g \Delta F_g$. Parameter values are listed in table B.1.

Table B.1: SARF parameter values (Etminan et al. 2016; Intergovernmental Panel on Climate Change (IPCC) 2023). †Units of $\mathrm{W\,m^{-2}\,ppt^{-1}}$.

Gas	Parameter	Value
CO_2	a_1	-2.48×10^{-7}
	b_1	7.59×10^{-4}
	c_1	-2.15×10^{-3}
	d_1	5.2488
	e_1 (adj. factor)	1.05
CH_4	a_2	-8.96×10^{-5}
	b_2	-1.25×10^{-4}
	d_2	0.045194
	e_2 (adj. factor)	0.86
N_2O	a_3	-3.42×10^{-4}
	b_3	2.546×10^{-4}
	c_3	-2.44×10^{-4}
	d_3	0.12173
	e_3 (adj. factor)	1.07
SF_6	$\omega^\dagger$	5.67×10^{-4}

CO_2-equivalent concentration. Total forcing ΔF_{total} is inverted to a CO_2e concentration C_{eq} defined as the level of CO_2 that, holding N_2O at its pre-industrial value N_0, would produce the same forcing as all four gases combined:

$$\Delta F_{CO_2}(C_{\text{eq}},\, N_0) = \Delta F_{\text{total}}. \tag{5}$$

Since equation (1) is not analytically invertible in C, C_{eq} is solved numerically each period using Brent's method over the interval $[C_0,\, 5000]$ ppm.

C Panel Data Description

The empirical analysis draws on six main data sources. Table C.1 summarises the key characteristics of each dataset. For time-series variables, the table reports the sample span and number of non-missing monthly observations. For panel variables it additionally reports the number of countries contributing at least one non-missing observation over the sample period. All samples are restricted to January 2002 onwards to coincide with the availability of the oil supply news shock series, which runs through December 2024. Panel outcome data extends to December 2025; the additional year contributes only as forward outcomes y_{t+h} for shocks realised in 2024, leaving the shock identification sample unchanged. Table C.2 reports start and end dates and summary statistics for the shock series and global control variables.

Electricity generation and CO_2 emissions come from Ember, a monthly panel covering over eighty countries from the early 2000s onwards. Wind and solar installed capacity, also from Ember, is available for a narrower sample of 25 countries starting in 2016. Industrial production at the country level is sourced from the IMF, covering the OECD and selected emerging economies. Oil product consumption data (gasoline, diesel, jet fuel, LPG, and residual fuel) are from the Joint Organisations Data Initiative (JODI), as is natural gas consumption by the power sector. Coal production at the country level comes from the International Energy Agency (IEA). Coverage across these datasets is unbalanced: the number of countries reporting data varies across variables and over time, as documented in Figures C.1 to C.5.

Table C.1: Data Sources and Summary Statistics

Dataset	Source	Countries	Start	End	Obs	Min	Mean	Median	Max	Std
CO_2e Concentration	NOAA	—	2002	2025	288	416.8	451.6	449.6	491.6	21.0
Real WTI Oil Price	FRED	—	2002	2025	287	0.2	0.8	0.7	1.7	0.3
Oil News Shock	KZ	—	2002	2024	276	-3.1	0.0	-0.0	1.9	0.6
Oil Supply Shock	BH	—	2002	2024	276	-1.9	0.1	0.1	5.3	0.6
Ind Prod	BH	—	2002	2025	288	60.5	86.0	86.4	110.7	13.6
Global Fin Cycle	MR	—	2002	2024	276	-2.9	0.2	0.1	2.6	1.0
Financial Uncertainty	JLN	—	2002	2025	288	0.7	1.0	0.9	1.4	0.1
Electricity Gen	Ember	87	2002	2025	10301	0.0	29.8	5.1	1015.2	93.3
Wind & Solar Capacity	Ember	25	2016	2025	2874	0.2	67.4	18.1	2082.1	181.5
(IMF) Ind Prod	IMF	65	2002	2025	16548	32.3	107.1	103.8	450.7	25.4
Oil Products	JODI	103	2002	2025	23762	0.0	7792.8	1242.2	306022.0	31206.1
Natural Gas	JODI	77	2009	2025	12058	0.0	2962.0	628.2	111046.0	8886.8
Coal Production	IEA	33	2016	2025	1188	0.0	65121.0	7768.0	1283191.0	183687.0

Notes: Sample restricted to years $\geq$ 2002. COVID-19 (2020) is excluded from regressions but included in the statistics above. Obs. counts reflect non-missing observations for the key variable in each row. For panel variables, *Countries* reports the number of countries with at least one non-missing observation. KZ: Känzig (2021); BH: Baumeister and Hamilton (2019); MR: Miranda-Agrippino and Rey (2020); JLN:Jurado, Ludvigson, and Ng (2015).

Table C.2: Global Variables: Shocks and Controls

Variable	Source	Series start	End	N	Mean	Std	Min	Median	Max
Oil Supply News Shock	Kaenzig (2021)	1975	2024	288	0.0	0.6	-3.1	-0.0	1.9
Structural Oil Supply Shock	Baumeister–Hamilton (2019)	1975	2024	288	0.1	0.6	-1.9	0.1	5.3
Global Financial Cycle	Miranda-Agrippino & Rey (2020)	1980	2024	288	0.2	1.0	-2.9	0.1	2.6
Financial Uncertainty (3M-ahead)	Jurado et al. (2015)	1970	2025	300	1.0	0.1	0.7	1.0	1.4
World Industrial Production (OECD+6)	Baumeister–Hamilton (2019)	1970	2025	300	85.0	14.2	60.0	85.1	110.7
Real WTI Oil Price	FRED	1970	2025	299	0.8	0.3	0.2	0.7	1.7

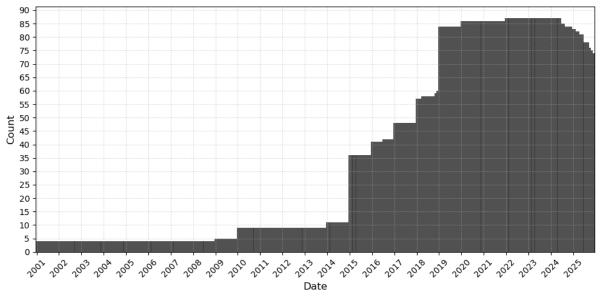
Notes: Series start reports the first non-missing year for each variable; control variables may have data predating 2001, but the oil supply shock series begin in January 2002. N, Mean, and all distributional statistics are computed from January 2001 onwards, matching the 12-month lag pre-period retained in the data (analysis sample starts in 2002). Oil supply shocks run through December 2024; panel outcome data extends to December 2025, with 2025 observations entering only as forward differences $y_{t+h} - y_{t-1}$ for shocks realised in 2024. KZ: Känzig (2021); BH: Baumeister and Hamilton (2019); MR: Miranda-Agrippino and Rey (2020); JLN: Jurado, Ludvigson, and Ng (2015).

Table C.3: Descriptive Statistics — Panel Variables

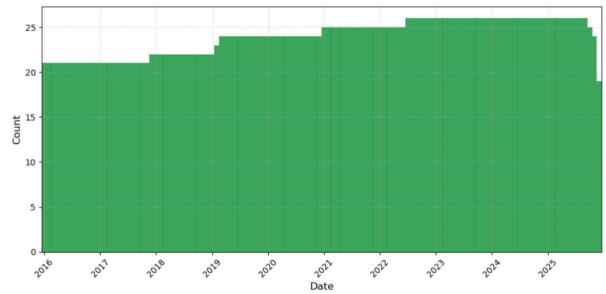
	Elec. Gen. (TWh)	W&S Cap. (GW)	IMF Ind. Prod.	ZEV Reg.	Alt+Hyb. Reg.	ICE Reg.	JODI Diesel	JODI Gasoline	JODI Jet Fuel	JODI LPG	JODI Res. Fuel	JODI Gas Power	JODI Gas Total	IEA Coal	IEA Cok. Coal	IEA Lignite
Mean	30.02	67.44	106.27	7,333.19	10,838.27	116,369.50	8,210.12	7,792.84	1,698.68	2,557.72	2,133.20	1,098.16	2,962.03	65,121.02	59,999.47	5,121.55
Std. Dev.	93.58	181.47	25.55	44,667.25	33,639.45	290,744.53	17,753.97	31,206.10	5,702.88	7,029.54	4,145.64	3,475.62	8,886.85	183,686.96	183,953.45	7,851.97
Min	0.02	0.20	32.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p25	1.56	6.46	93.00	16.00	103.25	7,108.00	649.02	354.90	15.76	84.00	56.46	53.00	250.00	620.00	0.00	0.00
Median	5.07	18.10	103.38	240.00	1,255.50	21,318.00	2,551.32	1,242.15	324.54	406.00	412.80	274.87	628.25	7,768.00	644.00	816.00
p75	17.61	48.08	115.18	2,207.00	6,243.09	94,134.75	7,324.38	4,495.23	1,183.84	2,108.00	2,100.16	800.00	2,996.75	31,914.50	24,431.75	8,992.50
Max	1,015.19	2,082.09	450.70	829,100.00	555,317.00	2,665,499.00	140,150.00	306,022.00	58,116.00	89,989.88	43,249.69	43,258.00	111,046.00	1,283,191.00	1,283,191.00	45,747.00
Observations	10,349	2,874	17,196	7,206	7,206	7,206	23,967	23,762	14,556	23,760	23,856	10,452	12,058	1,188	1,188	1,188
Countries	87	25	65	52	52	52	103	103	91	103	103	78	77	33	33	33
Months	300	120	300	300	300	300	288	288	204	288	288	204	204	39	39	39
Max ctry./month	86	25	64	52	52	52	94	94	84	94	94	63	71	33	33	33
Min ctry./month	4	18	37	6	6	6	58	58	49	57	56	7	5	27	27	27
Start date	2001-01	2016-01	2001-01	2001-01	2001-01	2001-01	2002-01	2002-01	2009-01	2002-01	2002-01	2009-01	2009-01	2016-03	2016-03	2016-03
End date	2025-12	2025-12	2025-12	2025-12	2025-12	2025-12	2025-12	2025-12	2025-12	2025-12	2025-12	2025-12	2025-12	2025-09	2025-09	2025-09

Notes: Elec. = Electricity; W&S = Wind & Solar; IMF Ind. Prod. = IMF Industrial Production Index; ZEV = Zero-Emission Vehicle; ICE = Internal Combustion Engine; Alt+Hyb. = Alternative Fuel + Hybrid; Cok. = Coking; Res. = Residual; JODI = Joint Organisations Data Initiative (quantities in KBBL = thousands of barrels, except gas power and total in M3 = cubic meters); IEA = International Energy Agency (quantities in KT = kilotons); Ctry. = Countries.

Figure C.1: Country Coverage — Electricity Generation & Capacity



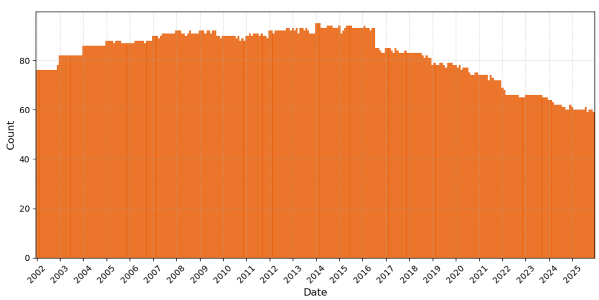
(a) Total Generation (TWh)



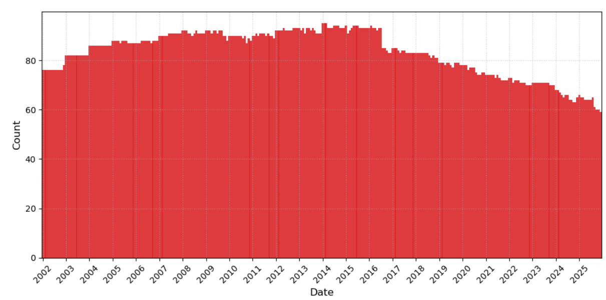
(b) Wind & Solar Capacity (GW)

Notes: Each bar shows the number of countries with non-missing data in that month. Source: Ember.

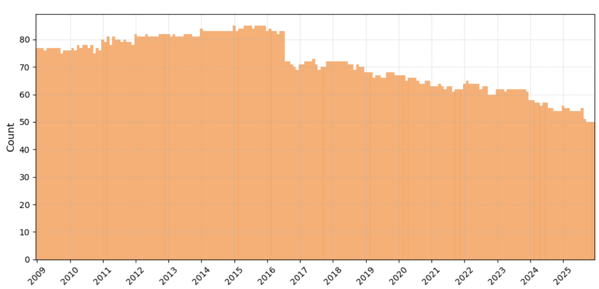
Figure C.2: Country Coverage — JODI Oil Products



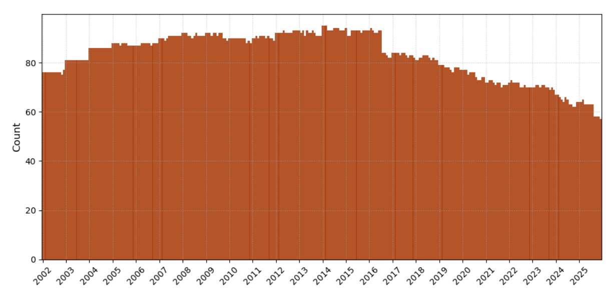
(a) Gasoline



(b) Diesel



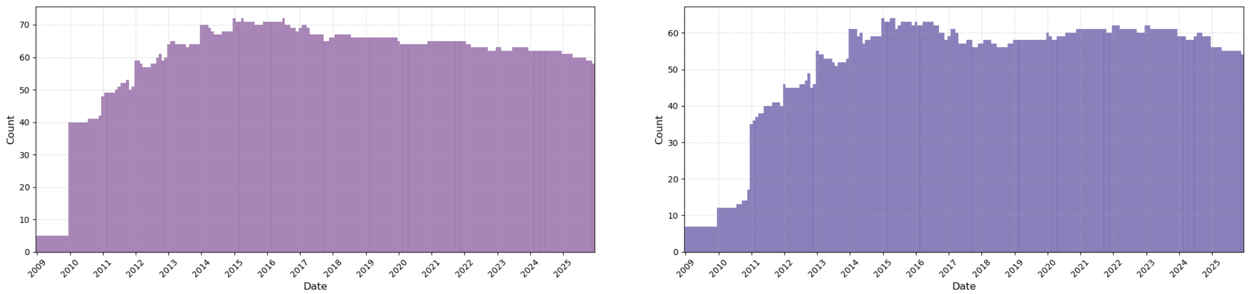
(c) Jet Fuel / Kerosene



(d) Residual Fuel

Notes: Each bar shows the number of countries with non-missing data in that month. Source: JODI.

Figure C.3: Country Coverage — Natural Gas & Coal

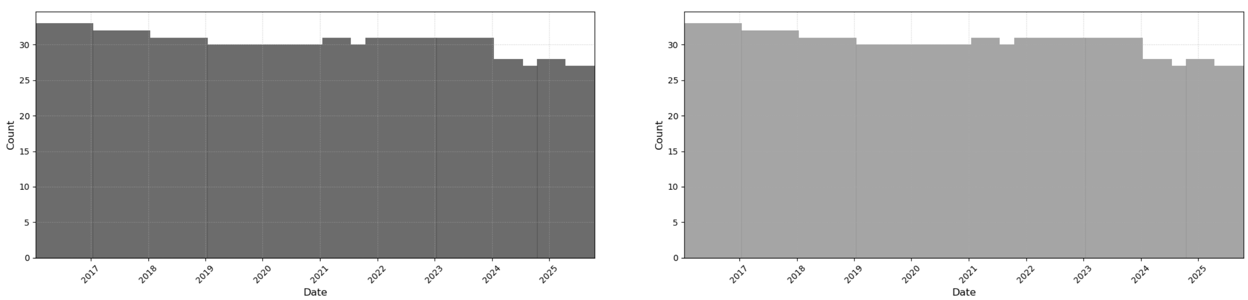


(a) Natural Gas — Total (JODI)

(b) Natural Gas — Power Sector (JODI)

Notes: Each bar shows the number of countries with non-missing data in that month. Source: JODI.

Figure C.4: Country Coverage — Coal Production

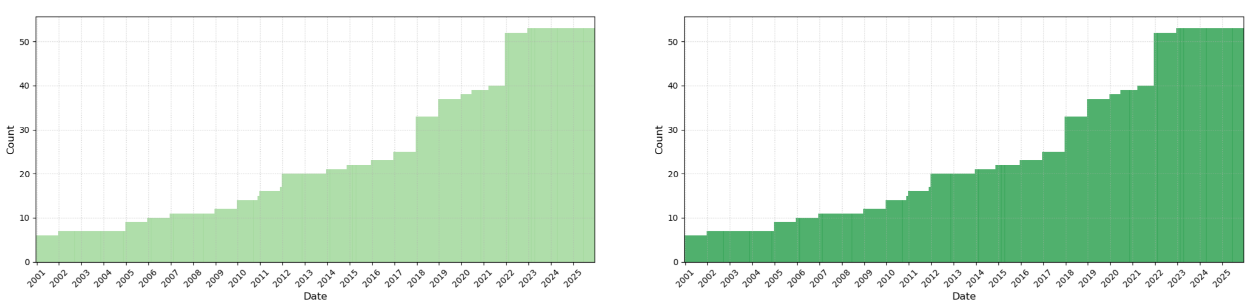


(a) Total Coal

(b) Lignite

Notes: Each bar shows the number of countries with non-missing data in that month. Source: IEA.

Figure C.5: Country Coverage — Vehicle Registrations



(a) Non-ICE Vehicles

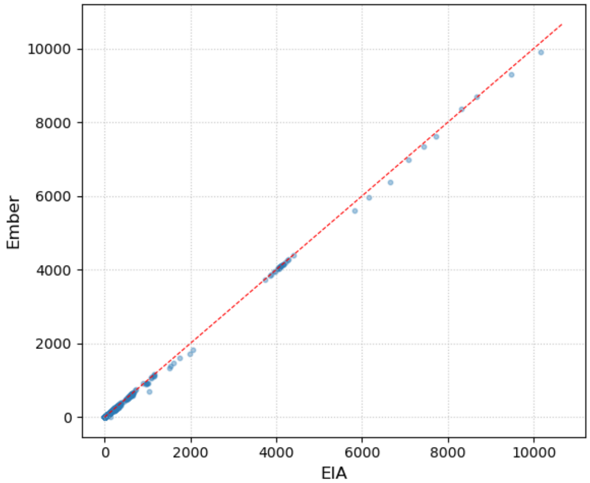
(b) Total Registrations

Notes: Each bar shows the number of countries with non-missing data in that month.

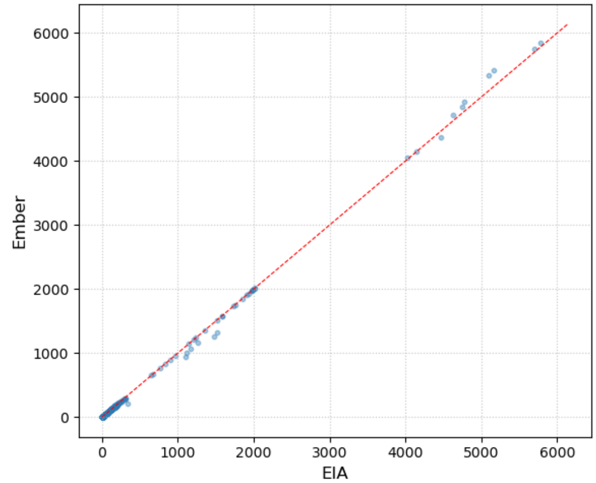
C.1 Ember–EIA Comparison

To validate the Ember dataset, I compare it against the annual international statistics published by the US Energy Information Administration (EIA). For each country and variable present in both datasets, I aggregate Ember's monthly data to an annual total and compare it against the corresponding EIA annual figure. Figure C.6 plots this comparison for electricity generation by fuel type, and figure C.7 extends the comparison to installed wind and solar capacity.

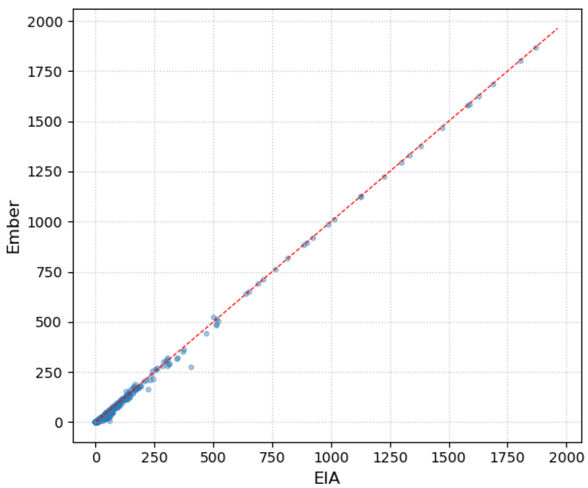
Figure C.6: Ember vs. EIA Annual Electricity Generation



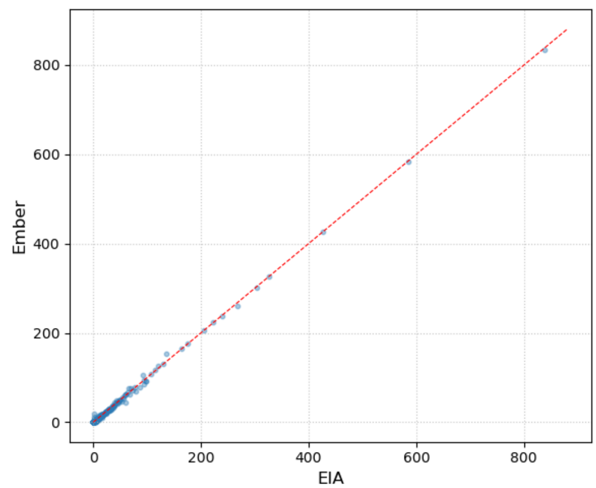
(a) Total Generation



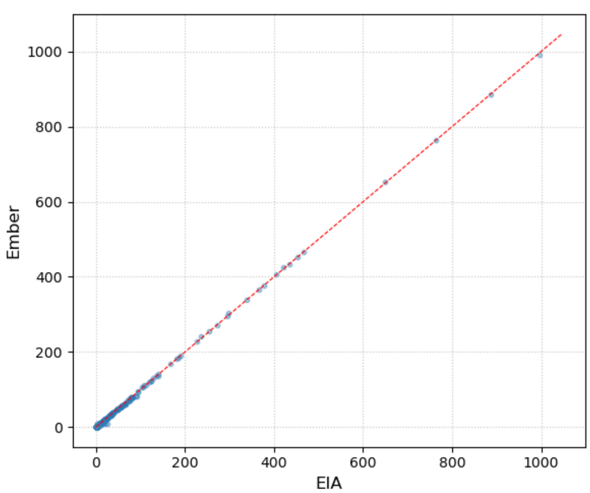
(b) Coal



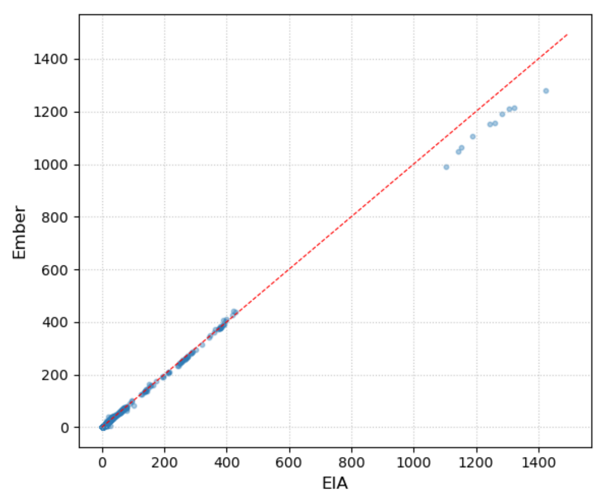
(c) Gas



(d) Solar



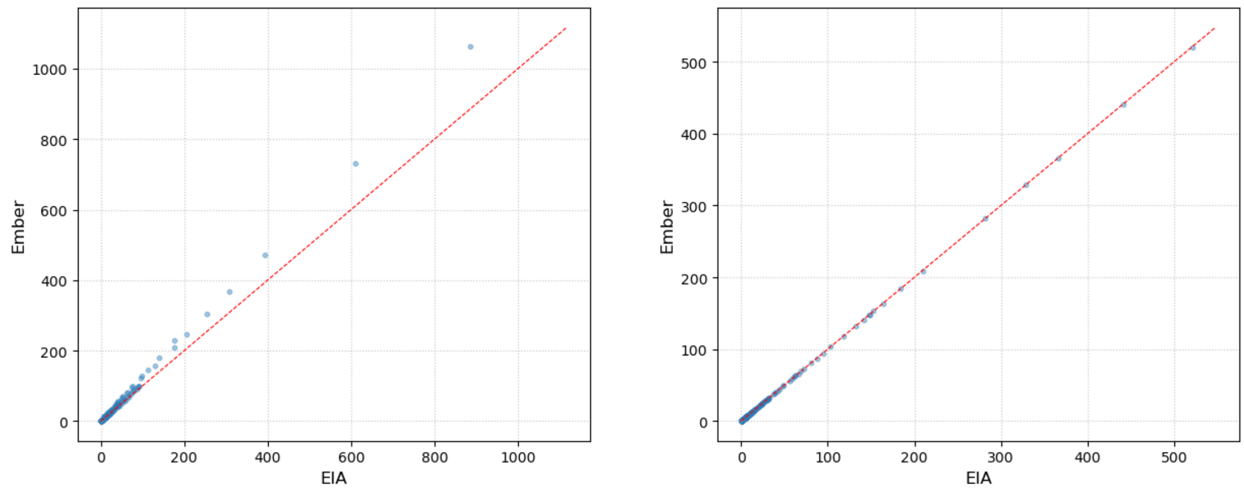
(e) Wind



(f) Hydro

Notes: Each point represents one country-year observation. Ember monthly generation is aggregated to annual totals and compared against EIA annual reporting. Both axes in TWh.

Figure C.7: Ember vs. EIA Annual Installed Capacity



(a) Solar Installed Capacity (GW) (b) Wind Installed Capacity (GW)

Notes: Each point represents one country-year observation. Ember annual capacity is compared against EIA annual reporting. Both axes in GW.

D Panel Analysis Robustness Checks

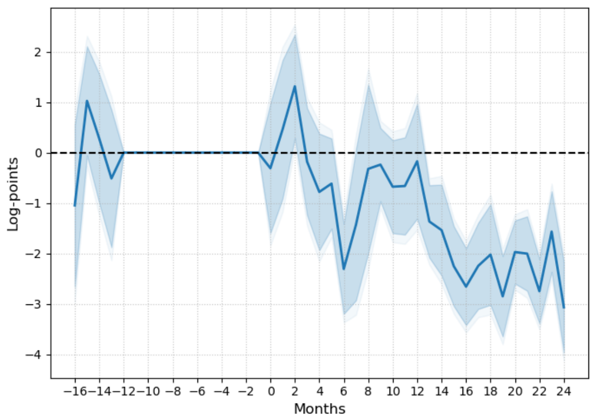
D.1 Renewable Generation on the Capacity Sample

The capacity panel covers 25 countries starting in January 2016 — a sample that is both geographically narrower and temporally shorter than the generation panel, which draws on more than 80 countries from 2002. Any direct comparison between generation and capacity impulse responses therefore risks conflating genuine causal effects with artefacts of sample composition: the 25 countries with capacity data may have systematically different generation dynamics, independent of any link between capacity and generation.

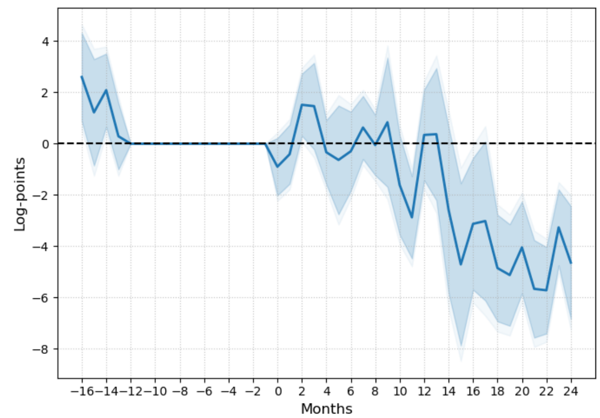
To address this concern, I re-estimate the generation local projections restricting the sample to those country–month observations for which the corresponding installed capacity variable is non-missing. Specifically, solar generation and renewables-excluding-nuclear are estimated on the subsample defined by non-missing solar capacity, and wind generation on the subsample defined by non-missing wind capacity. This yields impulse responses on an estimation sample identical to those used for the capacity results in figure 10, permitting a like-for-like comparison. The results are presented in figure D.1.

For solar, the dynamics closely replicate those from the full sample: the response follows the same trajectory stabilizing at log-points at the two-year horizon, compared to log-points in the full sample. The solar generation decline is therefore not driven by differences in country coverage between the generation and capacity panels. For wind, the restricted sample produces a volatile response: generation initially declines but subsequently rises, with wide confidence bands throughout. This noisy pattern probably reflects the substantially smaller sample and the inability to capture the underlying signal.

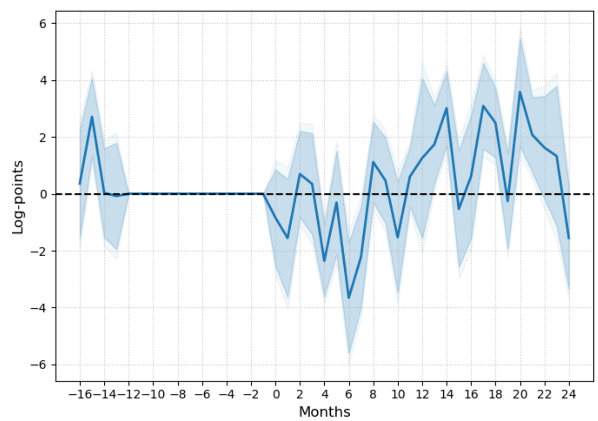
Figure D.1: Response of Renewable Electricity Generation to an Oil Supply News Shock — Capacity Sample



(a) Renewables x-Nuclear



(b) Solar



(c) Wind

Notes: Impulse responses to the oil supply news shock of Känzig (2021), normalised to a 10% increase in oil prices on impact. The dependent variable is the $h + 1$ change in log monthly electricity generation (TWh) from the respective source. Each regression is restricted to country–month observations for which the corresponding installed capacity variable is non-missing: solar generation and renewables-excluding-nuclear on the solar capacity sample, wind generation on the wind capacity sample. This matches the estimation sample to the capacity regressions reported in figure 10. Shaded bands correspond to 90% and 95% confidence intervals based on Driscoll–Kraay standard errors (Driscoll and Kraay 1998) with 12 lags. Regressions are weighted by the lagged 6-month moving average of the respective generation series (TWh). COVID-19 (2020) is excluded from the sample. Shocks: January 2016 – December 2024; outcome data extends to December 2025.

References

Baumeister, Christiane, and James D. Hamilton. 2019. Structural interpretation of vector autoregressions with incomplete identification: revisiting the role of oil supply and demand shocks. *American Economic Review* 109, no. 5 (May 1, 2019): 1873–1910. ISSN: 0002-8282. https://doi.org/10.1257/aer.20151569.

Driscoll, John C., and Aart C. Kraay. 1998. Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics* 80 (4): 549–560. ISSN: 0034-6535.

Etminan, M., G. Myhre, E. J. Highwood, and K. P. Shine. 2016. Radiative forcing of carbon dioxide, methane, and nitrous oxide: a significant revision of the methane radiative forcing. *Geophysical Research Letters* 43 (24): 12, 614–12, 623. ISSN: 1944-8007. https://doi.org/10.1002/2016GL071930.

Intergovernmental Panel on Climate Change (IPCC), ed. 2023. The earth's energy budget, climate feedbacks and climate sensitivity. In *Climate change 2021 – the physical science basis: working group i contribution to the sixth assessment report of the intergovernmental panel on climate change,* 923–1054. Cambridge: Cambridge University Press. ISBN: 978-1-009-15788-9. https://doi.org/10.1017/9781009157896.009.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng. 2015. Measuring uncertainty. *American Economic Review* 105, no. 3 (March 1, 2015): 1177–1216. ISSN: 0002-8282. https://doi.org/10.1257/aer.20131193.

Känzig, Diego R. 2021. The macroeconomic effects of oil supply news: evidence from OPEC announcements. *American Economic Review* 111 (4).

Miranda-Agrippino, Silvia, and Hélène Rey. 2020. U.s. monetary policy and the global financial cycle. *The Review of Economic Studies* 87, no. 6 (November 1, 2020): 2754–2776. ISSN: 0034-6527. https://doi.org/10.1093/restud/rdaa019.

Mori, Lorenzo, and Gert Peersman. 2024. *Estimating the macroeconomic effects of oil supply news.* SSRN Scholarly Paper 5054291. Rochester, NY: Social Science Research Network, November 1, 2024. https://doi.org/10.2139/ssrn.5054291.