Does Renewable Energy Renew Energy Efficiency?
Online Appendices

A. Additional Tables

Table A.1: Renewable Energy Procurement Approaches of Selected Companies (Adapted from IRENA 2018)

Company	Electricity consumption (TWh)	Renewable electricity consumption (TWh)	Share of renewable electricity (%)	of Which		
				PPA	RECs	Generation
Alphabet, Inc.	6.21	2.82	45%	100%		
Apple Inc.	1.45	1.45	100%	41%	32%	24%
Deutsche Bahn AG	10.5	3.44	33%		98%	2%
H&M Hennes & Mauritz AB	1.39	1.15	83%		100%	
Intel Corporation	5.46	4.3	79%		99%	1%
Kohl's Corporation	1.24	1.24	100%		100%	
Microsoft Corporation	4.85	4.79	99%	17%	83%	
Volkswagen AG	12.36	4.03	33%	92%		8%

Table A.2: Summary Statistics: Monthly Observations from 2015-2020 with Instrumental Variable ($Clean\%$)

	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Energy consumption (GJ)	7,278	9,500	10,039	2,665	6,961	11,941
Fossil (GJ)	7,278	5,305	7,719	847	2,944	6,859
Renewable (GJ)	7,278	4,195	5,145	731	2,686	5,303
$RE\%$	7,278	0.49	0.32	0.24	0.51	0.74
$RE\%^{PPA}$	7,278	0.38	0.32	0.002	0.35	0.63
$RE\%^{REC}$	7,278	0.06	0.18	0.00	0.00	0.00
$RE\%^{OnSite}$	7,278	0.06	0.18	0.00	0.00	0.00
Clean%	7,278	0.41	0.24	0.24	0.35	0.53
$Production$ (tons)	7,278	7,810	8,400	2,399	5,141	10,048
$EPTOP$ (GJ/ton)	7,278	2.80	16.46	0.66	1.19	1.94
$\overline{Cost}^F$ (\$/GJ)	7,278	10.91	10.28	4.92	9.23	14.76
$\overline{Cost}^R$ (\$/GJ)	7,278	21.86	21.17	12.96	20.97	28.01

Table A.3: Summary Statistics for Policy Variables

	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
IEAc	13,176	2.61	3.13	0.00	1.00	4.00
IEAc-GPT	13,176	2.56	3.12	0.00	1.00	4.00
GPTc	13,176	6.58	8.54	1.00	3.00	7.67

Table A.4: Main Regression with Energy Consumption as the Dependent Variable

	Dependent variable:			
	$\log(Energy)$			
	OLS			IV
	(1)	(2)	(3)	(4)
$RE\%$	-0.403^{*}	-0.320^{***}	-0.695^{***}	-0.805^{*}
	(0.218)	(0.101)	(0.224)	(0.477)
$\log(Prod)$		0.467^{***}	0.968^{***}	0.967^{***}
		(0.057)	(0.040)	(0.029)
$\overline{Cost}^{F}$		0.0005	-0.003^{*}	-0.004
		(0.0005)	(0.002)	(0.002)
$\overline{Cost}^{G}$		0.0001	0.001^{**}	0.001^{*}
		(0.0002)	(0.0003)	(0.001)
Site/Mgr FE	No	Yes	Yes	Yes
Month/Year FE	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes
Observations	13,176	13,176	7,278	7,278
Adjusted R^2	0.010	0.933	0.946	0.946

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

This table shows the impact of $RE\%$ on energy consumption. Columns (1) and (2) are based on the full sample; Column (2) incorporates both manager and site fixed effects; Column (3) uses the subsample for which data on $Clean\%$ is available. Column (4) shows the estimates of the IV model using $Clean\%$ as the instrument.

Table A.5: First-stage Regressions of $RE\%$ on $Clean\%$, and of $OnSite$ and $OnSite \times RE\%$ on Three Policy Instruments

	Dependent variable:						
		Policy IV: IEAc		Policy IV: IEAc-GPT		Policy IV: GPTc	
	$RE\%$	$OnSite$	$OnSite \times RE\%$	$OnSite$	$OnSite \times RE\%$	$OnSite$	$OnSite \times RE\%$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Clean\%$	0.002***						
	(0.001)						
$Policy$		−0.052***	−0.042***	−0.054***	−0.043***	−0.037***	−0.016***
		(0.015)	(0.012)	(0.016)	(0.013)	(0.012)	(0.006)
$RE\%$		0.063	0.261***	0.062	0.259***	0.054	0.225***
		(0.052)	(0.069)	(0.052)	(0.069)	(0.047)	(0.065)
$Policy \times RE\%$		0.012	0.040***	0.013	0.041***	0.006*	0.020***
		(0.008)	(0.014)	(0.008)	(0.015)	(0.003)	(0.005)
Site/Mgr/Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SW-test (p-val)	-	0.0039	0.0016	0.0051	0.0014	0.0130	0.0001
Cond. F-stat	23.6	68.24	594.92	63.08	598.72	30.55	812.32
Observations	7,278	13,176	13,176	13,176	13,176	13,176	13,176
Adjusted R^2	0.866	0.948	0.936	0.948	0.937	0.948	0.937

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.6: Regressions of Site Production and Energy Costs on Instrumental Variables

	Dependent variable:							
	log($Prod$)				$\overline{Cost}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Clean\%$	0.108				−370.490			
	(0.150)				(315.177)			
IEAc		0.116				−209.991		
		(0.088)				(181.249)		
IEAc-GPT			0.061				−108.998	
			(0.063)				(185.930)	
GPTc				0.295				−206.988
				(0.188)				(193.788)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,278	13,176	13,176	13,176	7,278	13,176	13,176	13,176
Adjusted R^2	0.667	0.716	0.716	0.716	0.023	0.025	0.025	0.025

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.7: EE Responses to Threshold-Parameterized RE Changes

	Dependent Variable				
	$\log(EPTOP)$				
	20%	30%	40%	50%	60%
Green	-0.164^{***}	-0.176^{***}	-0.185^{***}	-0.205^{***}	-0.244^{**}
	(0.057)	(0.055)	(0.059)	(0.074)	(0.095)
$\log(Prod)$	0.091^{*}	0.083	0.046	0.043	0.063
	(0.049)	(0.052)	(0.053)	(0.053)	(0.044)
$\overline{Cost}^F$	-0.002	-0.001	-0.001	-0.002	-0.002^{*}
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
$\overline{Cost}^R$	-0.0004	-0.0003	-0.0005	-0.0001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Site/Mgr, Mon, Yr FE	Yes	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes	Yes
Observations	6,590	8,229	9,766	10,296	10,831
Adjusted R^2	0.755	0.751	0.756	0.768	0.785
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Table A.8: Hierarchical Specification

Dependent Variable: log($EPTOP$)

	$RE\%$	$RE\%^{Buy}$	$RE\%^{OnSite}$	$RE\%^{PPA}$	$RE\%^{REC}$	$RE\%^{OnSite}$
(S)	-0.247^{***}					
	(0.013)					
		-0.151^{***}	0.162^{***}			
		(0.013)	(0.040)			
				-0.154^{***}	-0.141^{***}	0.163^{***}
				(0.015)	(0.023)	(0.040)
(M)	-0.264^{***}					
	(0.016)					
		-0.299^{***}	0.110^{***}			
		(0.017)	(0.047)			
				-0.307^{***}	-0.252^{***}	0.119^{***}
				(0.018)	(0.029)	(0.046)
(R)	-0.180^{***}					
	(0.014)					
		-0.205^{***}	0.114^{***}			
		(0.015)	(0.041)			
				-0.211^{***}	-0.188^{***}	0.114^{***}
				(0.016)	(0.024)	(0.041)
(P)	-0.141^{***}					
	(0.013)					
		-0.172^{***}	0.183^{***}			
		(0.013)	(0.041)			
				-0.183^{***}	-0.142^{***}	0.186^{***}
				(0.015)	(0.023)	(0.041)
(R/P)	-0.197^{***}					
	(0.014)					
		-0.232^{***}	0.152^{***}			
		(0.015)	(0.042)			
				-0.240^{***}	-0.207^{***}	0.152^{***}
				(0.016)	(0.025)	(0.042)

Note: *p<0.1; **p<0.05; ***p<0.01

(S) and (M) capture site and manager random effects, respectively. (R) and (P) capture regional and product category effects, respectively, both with nested sites. (R/P) includes both regional and product category cross effects. All specifications include region and product category time trends, as well as month/year fixed effects and controls.

Table A.9: Sample Robustness Tests – Extended to 2011/Excluding 2020

	Dependent variable:					
	$\log(EPTOP)$					
	(1)	(2)	(3)	(4)	(5)	(6)
$RE\%$	-0.215^{**}			-0.188^{***}		
	(0.094)			(0.064)		
$RE\%^{Buy}$		-0.293^{***}			-0.210^{***}	
		(0.104)			(0.066)	
$RE\%^{PPA}$			-0.314^{***}			-0.225^{***}
			(0.108)			(0.065)
$RE\%^{REC}$			-0.211^{*}			-0.177^{**}
			(0.111)			(0.084)
$RE\%^{OnSite}$		0.121	0.119		0.080	0.080
		(0.081)	(0.081)		(0.089)	(0.089)
Site/Mgr FE	Yes	Yes	Yes	Yes	Yes	Yes
Month/Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Time Trends	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26,844	26,844	26,844	10,980	10,980	10,980
Adjusted R^2	0.844	0.845	0.845	0.833	0.833	0.833
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

B. Managerial Survey Results

1. What were your motivations for investing in renewable energy? (select primary)
   - (i) Be more environmentally friendly - 2%
   - (ii) Achieve emissions targets - 6%
   - (iii) Improve cost efficiency - 71%
   - (iv) Following corporate mandate - 21%
2. What were your motivations for investing in renewable energy? (select multiple)
   - (i) Be more environmentally friendly - 31%
   - (ii) Achieve emissions targets - 43%
   - (iii) Improve cost efficiency - 94%
   - (iv) Following corporate mandate - 85%
3. As a manager, after you began sourcing renewable energy, how did your attitude towards energy efficiency change?
   - (i) No Change - 3%
   - (ii) Increase focus on energy efficiency - 74%
   - (iii) Decrease focus on energy efficiency - 23%
4. How did the attitude of the plant employees toward energy efficiency change after investing in renewable energy?
   - (i) No Change - 5%
   - (ii) Increase focus on energy efficiency - 77%
   - (iii) Decrease focus on energy efficiency - 18%

C. Policy Instrumental Variable

In this section, we discuss the construction of our instrumental variable and its validation. We first describe the data and context upon which the measure is based, then detail the construction method, and finally, demonstrate the validity of the instruments across various diagnostic tests.

C.1 Data and Context

Our data is drawn from the IEA. The Policies and Measures database (https://www.iea.org/policies/about) is a compilation of past, current, and planned policies related to greenhouse gases, energy efficiency, and renewable and clean energy technologies. It integrates data from several specific databases including the IEA/IRENA Renewable Energy Policies and Measures database, the IEA Energy Efficiency database, the Addressing Climate Change database, the Building Energy Efficiency Policies database, as well as information on carbon capture utilization and storage (CCUS) and methane abatement policies. It contains policies implemented since 1999, with governments having the opportunity to review the information periodically.

As of October 2023, the database contains 7,460 policies. Each entry details the country, region, year of enactment, status, and jurisdiction of the policy, along with a list of associated topics, policy types, technologies, and sectors. The policies are classified into 15 topics as shown in Table C.1 below, and tagged with various policy types from a list of 180 possible values, such as grants, taxes, and prohibitions. This classification facilitates granular analysis of each policy's relevance to either energy efficiency or renewable energy, or both.

Table C.1: List of Policy Topics in the IEA Policies and Measures Database

Energy Efficiency	Critical Minerals	Technology R&D and Innovation
Energy Poverty	Energy Security	Carbon Capture Utilisation and Storage
Energy Access	**Renewable Energy**	Air Quality
Electrification	Methane Abatement	Digitalisation
Just Transition	Energy Water Nexus	Cities

Each policy may carry multiple labels, indicating its relevance to various areas in the energy sector. Among all policies, there are 2,910 categorized under renewable energy and 3,632 under energy efficiency. While some policies carry both EE and RE labels, many are categorized under one label, reflecting their distinct legislative focus. Notably, of the 2,910 RE policies, 80% (2,320) do not carry an EE topic label. These 2,320 RE policies span 142 different countries, encompassing nearly all nations in our dataset, with only two exceptions, both of which are major fossil fuel exporters. For these two countries, the policy variable (constructed in Section C.3) takes value 0 for all periods.

C.2 Policy Data Curation

To construct the sample for our $RE\%$ instruments, we target policies that promote on-site renewable energy, while carefully excluding those that support energy efficiency. We curate the policy data based on three different approaches. First, we utilize the categorizations provided by IEA following their review of each policy (IEAc). Second, we employ GPT to refine these categorizations by identifying and removing policies that may affect EE, even if not explicitly labeled as such (IEAc-GPT). Third, we use GPT to directly assess and label all policies, targeting those that impact on-site renewable installation (GPTc).

C.2.1 IEA Categorization

Our first approach uses IEA's policy categorizations. Starting with the topic of the policy, we isolate only those RE policies that do not also address EE. Specifically, we filter for policies tagged under the 'Renewable Energy' topic within the Policies and Measures database, while removing any policies that are also tagged under the topic of 'Energy Efficiency'. This initial filter ensures our sample exclusively targets policies promoting RE.

Next, we refine our selection by examining the types of policies, filtering only for those types that directly affect the financial incentives for installing renewable technologies. Not all policy types have a direct impact. For example, strategic plans may outline a country's energy-related goals without allocating necessary resources to achieve those goals. Recognizing the diverse policy types, we further refine our sample to include only those policies that provided direct financial incentives. Specifically, we filter for the following policy types: (i) Payments, finance and taxation, (ii) Payments and transfers, (iii) Grants, (iv) Performance-based payments, (v) Feed-in tariffs/premiums, (vi) Finance, and (vii) Taxes, fees and charges. This excludes policies that are (i) Strategic plans, (ii) Targets, Plans and Framework legislation, and (iii) Regulations. With these filters, we target only policies that directly affect the financial incentives for installing on-site generation capacity within each country.

Finally, we focus only on those policies relevant to the commercial and industrial sectors. Not all policies are expected to impact manufacturing firms. For instance, policies promoting electric buses may influence RE but are unlikely to directly affect RE installation for factory operations. To ensure our included policies are pertinent to manufacturing, we filter for those policies related to power generation, industry, and buildings (27 categories in total). This curated dataset forms the basis of our instrumental variable.

C.2.2 Refining IEA Categorizations using GPT

GPT offers distinct advantages for policy assessment, particularly when dealing with large volumes of data, such as our dataset of over 7,400 policies. Stechemesser et al. (2024) use a similar approach in assessing the impact of climate related policies. It can rapidly and uniformly assess policy descriptions, maintaining consistency that human evaluators may lack due to inherent biases and variability. However, it is essential to acknowledge GPT's limitations in contextual understanding, particularly regarding nuanced implications that may not be explicitly stated. While GPT can identify patterns and keywords, it may struggle to assess strategic or long-term implications comprehensively. For a deeper discussion of the advantages and limitations of GPT, we refer readers to a nascent but rapidly growing literature (Bommasani et al. 2021, Bubeck et al. 2023).

The accuracy of the policy classifications is crucial to the validity of our instrumental variable. To enhance the accuracy, we implement a hybrid approach that combines outputs from IEA, GPT, and human assessments, leveraging the strengths of each source of categorization. In short, we use GPT to assess polices, then compare the GPT outputs to the IEA categorizations to identify any discrepancies, and for policies with categorization disagreements, we employ human analysts to conduct independent assessments to resolve discrepancies. We detail the procedure below.

We start with the sample constructed from the IEA categorizations. Using the OpenAI API and the GPT-4o-mini-2024-07-18 model, we define the GPT system role using the following prompt: *"You are to provide ratings based on direct support for Energy Efficiency as an integer 0, 1, or 2, with 0 being low direct support, 1 being medium direct support, 2 being high direct support. Do NOT rate based on the support for renewable energy or anything else. Focus on direct support for energy efficiency. Return a table with ID, rating, and justification."*

We provide the ID and Description of each policy as the user content *individually* to the system.[10] Given that GPT models utilize transformer architecture for handling text sequences and rely on probabilities and sampling to generate outputs, we repeat this process three times on different days for robustness. Each policy, therefore, receives a combined rating ranging from 0 to 6 (sum of the three ratings). Policies with a total rating exceeding 1 – indicating any iteration found some support for EE – are flagged for review. We adopt this conservative threshold to ensure that any policies with potential confounders are reviewed. The flagged policies, 22 in total, are reviewed by three humans individually, each assigning a rating between 0 to 6. Policies that receive a human rating greater than 1 are subsequently removed from the final policy set. This occurred for seven policies.

C.2.3 GPT Classification

In the first and second approaches, we utilized the IEA labels to establish the initial policy set. However, this may inadvertently exclude relevant policies that support renewable generation but were not captured by IEA's categorization. Therefore, in our third approach, we examine the 5,940 policies

[10] We also experimented with batch processing, providing policies in groups of up to 100. However, this sometimes led to errors in the output, including justifications referencing different policies or omissions altogether.

implemented after 2014, using GPT to determine their relevance to on-site renewable generation.

We implement this approach in three stages. In the first stage, we ask GPT to rate all policies based on their support for EE using the same prompt as above. Applying the same conservative threshold, we remove any policies that receive a rating exceeding 1 in any of the three iterations. Second, we ask GPT to evaluate the remaining policies based on their support for RE using an analogous prompt. Specifically, we define the system role as: *"You are to provide ratings based on direct support for Renewable Energy as an integer 0, 1, or 2, with 0 being low direct support, 1 being medium direct support, 2 being high direct support. Do NOT rate based on the support for energy efficiency or anything else. Focus on direct support for renewable energy. Return a table with ID, rating, and justification."* We again take a conservative approach and retain only policies receiving at least 5 out of 6 rating points, indicating strong relevance to RE.

In the final stage, we concentrate on policies that support commercial and industrial on-site energy production. We ask GPT to rate the remaining policies based on the following prompt: *"You are to provide ratings based on direct support for commercial on-site renewable energy as an integer 0, 1, or 2, with 0 being low to no direct support, 1 being medium direct support, 2 being high direct support. On-site renewable energy is defined as energy generated at the point of consumption. Do NOT rate based on the support for general renewable energy, secondary effects or anything else. Focus on direct support for commercial on-site renewable energy. Return a table with ID, rating, and justification."* Note that this prompt specifically defines on-site energy to distinguish our target policies—energy generated specifically for the manufacturer—from those supporting large-scale installations, such as solar farms that would typically be operated by utility companies or external power providers.

This multi-step approach is decidedly more complicated than a one-shot method that combines the three criteria into a single prompt. While a single prompt may offer efficiency, separate prompts facilitate detailed analysis and justification for each criterion. This method enables easier identification of inaccuracies and misunderstandings during human reviews of the outputs. Recent literature (Reynolds and McDonell 2021, Khot et al. 2022, Kojima et al. 2022) also suggests that a step-by-step approach may yield better outputs.

We compare the final GPT ratings to the IEA categorizations. For policies where there is disagreement between the IEA categorizations and GPT ratings, defined as policies that are either excluded using IEA categorizations while receiving at least 5 out of 6 GPT rating points or included under IEA categorization but receiving less than 6 out of 6 rating points, three humans review each policy individually and create a consensus rating. The final policy ratings $\{0, 1, 2\}$ are formed using the human consensus for any policies with disagreements and the GPT rating for the remainder. Ratings are averaged and rounded to the nearest integer.

C.3 Instrumental Variable Construction

We construct three alternative measures of our policy instrument: IEAc, IEAc-GPT, GPTc, corresponding to the three approaches in Section C.2. Summary statistics are provided in Table A.3.

Most policies are enacted at the country level, with exceptions for the European Union (EU). To account for EU-wide policies, which apply across all member states, we attribute them to each individual country within the EU. Starting in 2014, one year before the start of our data to account for policy implementation lag, we track the intensity of on-site renewable energy policy support over time as follows.

IEAc and IEAc-GPT are defined as the number of policies that have been enacted up to the focal year. Note that we do not reduce the number of policies after a policy has been phased out, as policy impacts can be persistent. For example, policies subsidizing solar panel installation will result in more solar energy production even after subsidies are withdrawn when the policy expires.

GPTc is defined as the cumulative sum of the RE support ratings 0-2 of policies that have been enacted up to the focal year. This measure aims to capture not only the number of policies but also the differences in the level of support between them.

We compute these measures at the country-year level. All manufacturing sites within the country will have the same instrument value for a given year, but the value varies across countries and over time. Summary statistics are shown in Table A.3.

C.4 Validation

We validate the consistency and quality of our ratings approach, as well as the validity of instruments calculated using the resulting variables. We measure the consistency of the GPT and human ratings, and we spot-check the ratings and justifications for a pre-determined set of policies that have been identified as potentially challenging to rate as well as for a randomly selected subset of policies. We report the consistency measures, as well as the weak instrument tests below. We then argue that the exclusion criteria is satisfied for the instrument.

Inter-rater reliability For the IEAc-GPT and GPTc measures that involve multiple ratings for each policy, we calculate the inter-rater reliability (IRR), also known as reliability of agreement, between the different raters: (i) GPT amongst itself, (ii) Humans amongst themselves, (iii) GPT and Humans. We use Fleiss' Kappa as our statistical measure (Fleiss 1971), which extends Cohen's Kappa for use with more than two raters and for categorical data. It measures the degree of agreement among independent reviewers, accounting for agreements that might occur by chance and has been used to evaluate machine learning techniques for natural language processing tasks. It ranges from 0 to 1 where 0 indicates agreement equivalent to chance, and 1 indicates perfect agreement. (See

Artstein and Poesio 2008 and Gwet 2014 for more details.)

For IEAc-GPT, the IRR between the three GPT-generated measures of policy support for EE is 0.895 (p-val < 0.001), suggesting a very high level of agreement across the GPT iterations. For GPTc, we use GPT ratings over the three stages to identify EE, RE, and OnSite relevance of each policy. We find IRRs of 0.906 (p-val < 0.001), 0.863 (p-val < 0.001), and 0.799 (p-val < 0.001) respectively.

For those policies that underwent human review, we find an IRR of 0.556 (p-val < 0.001) among human ratings, indicating fair to good consensus among the raters. Given that human raters only reviewed 113 policies that are more difficult to assess, a corresponding comparison of the GPT IRR for those policies is 0.685 (p-val < 0.001).

Policy Review To ensure the quality of outputs, the research team selected policies with varying degrees of support for RE and EE as test cases, including policies that both enhance and reduce support. These included policies that (i) had EE but no RE, (ii) EE and RE, (iii) no EE but RE. The team evaluated the quality of the output for this subset before proceeding with rating the full set of policies.

An example of (i) is the following policy:

> The Industrial Energy Transformation Fund (IETF) aims to transform the way the industry sector consumes energy by helping companies with high energy use cut their bills and reduce carbon emissions. It was announced in Budget 2018 as a 315 million GBP fund and is delivered in 2 phases: • Phase 1, now closed, offered 70 million GBP of funding for companies in England, Wales and Northern Ireland. • Phase 2, with the remainder of the funding, launches in 2021. The Phase 2 budget is allocated across 4 competition windows, worth a combined 220 million GBP.

GPT rated this policy a "6" in terms of affecting EE with the justification that "The IETF provides significant funding specifically targeted at improving energy efficiency in high energy-consuming industries, with multiple rounds of substantial financial investment and a clear commitment to reduce carbon emissions through energy efficiency measures." Therefore, this policy is excluded.

The following is an example of (ii):

> The programmes use EU and Hungarian government funds to provide subsidies for energy efficiency and renewable energy investments. Various sectors are targeted, including energy production, transport, distribution and end-use, with a special focus on public buildings. The programme aims to improve energy efficiency across sectors, including energy production, transport, distribution and end-use, with special focus on public

> buildings. It also aims to increase the share of renewable energy to 13% by 2020. Small- and medium-sized enterprises, public institutions and non-profit organisations can apply for the subsidy. The programmes support the following planned activities in the field of renewable energy: -Supporting biomass-utilization, waste utilization, biological waste based biogas production and utilization, geothermal electric and/or thermal energy generation and utilisation, heat pump systems, solar power -Improving existing water power plants' life span, efficiency, and energy conversion efficiency, as well as supporting the establishment of smaller water power plants; -Electric energy production via wind power; -The introduction and application of tools and regulatory measures that support the integration of renewable energy sources into the electricity and other energy networks.

GPT rated this policy a "6" in terms of supporting RE with the justification that "The programme provides substantial support for a variety of renewable energy investments, including biomass, geothermal, solar, wind, and hydro, directly aiming to increase the share of renewable energy." It also rated it a "6" in terms of affecting EE with the justification that "The programmes provide direct subsidies specifically for energy efficiency investments across various sectors, actively supporting activities aimed at improving efficiency and increasing the share of renewable energy. The focus on multiple sectors and specific measures for public buildings and renewable energy utilization denotes a high level of direct support for energy efficiency." Therefore, this policy is excluded, despite it being potentially relevant.

Finally, the following is an example of (iii):

> On 11 of May 2018, the government of India adopted the Scheme to Support Promotion Of Biomass-Based Cogeneration In Sugar Mills And Other Industries In The Country (Up To March 2020). The programme will provide a Central financial Assistance (CFA) for projects utilizing biomass like bagasse, agro-based industrial residue, crop residues, wood produced through energy plantations, weeds, wood waste produced in industrial operations, etc. Municipal Solid Waste is not covered under the programme.Central financial Assistance (CFA) will be provided at the rate of Rs.25 Lakh/MW (for bagasse cogeneration projects) and Rs.50 Lakh/MW (Nonbagasse Cogeneration projects) under the scheme. The CFA will be back-ended and will be released in one instalment after successful commissioning and commencement of commercial generation and performance testing of the plant. The CFA will be released to the term loan account to reduce the loan component of the promoter.

GPT rated this policy a "6" in terms of supporting RE with the justification that "The government of India's scheme specifically promotes biomass-based cogeneration, providing substantial financial assistance for projects that support the development of renewable energy sources directly. The targeted financial support indicates a strong commitment to fostering renewable energy initiatives.". It also rated it a "1" in terms of affecting EE with the dissenting rating being that "The scheme provides Central Financial Assistance for biomass-based cogeneration, indicating medium direct support for energy efficiency through financial incentives, though it focuses specifically on biomass and does not have a broader application for other forms of energy efficiency." Therefore, this policy is potentially relevant depending on its support of on-site RE specifically.

In terms of the policy's support for on-site RE, GPT determined that "The scheme provides financial assistance for biomass-based cogeneration projects, which directly supports the generation of energy at the point of consumption in commercial settings. However, the focus is specifically on biomass and may not encompass a wider range of on-site renewable energy solutions, hence a medium level of support is appropriate." Across multiple GPT iterations, it received an average rating of "1.33." After human review, the consensus rating was set at "2."

Instrument Relevance The instruments constructed in Section C.3 are relevant to the *OnSite* indicator as more financially favorable policies increase the likelihood of installing on-site generation.

To demonstrate this quantitatively, Table A.5 shows the results from estimating the first-stage regressions of our instrumental variable and its interaction term on the potentially endogenous *OnSite* variable and its interaction term. We find that the coefficients of the instruments are economically and statistically relevant. Furthermore, the Cragg-Donald F-statistic, which tests for weak instruments, are 68.6, 69.1, and 73.1, all exceeding the accepted threshold of 10. Finally, we also report the results of the Sanderson-Windmeijer (S-W) multivariate F-test of excluded instruments, which assesses the strength of instruments when there are multiple endogenous variables and clustered observations (Sanderson and Windmeijer 2016). All tests are statistically significant, indicating that the excluded instruments are jointly significant, (i.e., they are relevant and strong instruments).

The instrument satisfies the exclusion restriction as the scope of the RE policies would not directly influence EE, except through on-site RE adoption. The policies were explicitly selected and screened for their impact on EE, and the first-stage regression controls for any impact through renewable energy from other sources.

Appendix References

Artstein, R., M. Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics* **34**(4) 555–596.

Bommasani, R., D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint* arXiv:2108.07258.

Bubeck, S., V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint* arXiv:2303.12712.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* **76**(5) 378.

Gwet, K. L. 2014. *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.

IRENA. 2018. Corporate sourcing of renewables: Market and industry trends. International Renewable Energy Agency (IRENA), Abu Dhabi.

Khot, T., H. Trivedi, M. Finlayson, Y. Fu, K. Richardson, P. Clark, A. Sabharwal. 2022. Decomposed prompting: A modular approach for solving complex tasks. *arXiv preprint* arXiv:2210.02406.

Kojima, T., S. S. Gu, M. Reid, Y. Matsuo, Y. Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in Neural Information Processing Systems* **35** 22199–22213.

Reynolds, L., K. McDonell. 2021. Prompt programming for large language models: Beyond the few-shot paradigm. *2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

Sanderson, E., F. Windmeijer. 2016. A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics* **190**(2) 212–221.

Stechemesser, A., N. Koch, E. Mark, E. Dilger, P. Klösel, L. Menicacci, D. Nachtigall, F. Pretis, N. Ritter, M. Schwarz, et al. 2024. Climate policies that achieved major emission reductions: Global evidence from two decades. *Science* **385**(6711) 884–892.