Dynamic (Mis)allocation of Investments in Solar Energy

Nicolas Astier, Nicolas Hatem

Dynamic (Mis)allocation of Investments in Solar Energy*

Nicolas Astier† Nicolas Hatem‡

December 4, 2023

Abstract

Because they differ in terms of technology, size and location, solar photovoltaic installations exhibit very heterogeneous levelized costs of producing electricity. Therefore, the present value cost of meeting a given trajectory of annual solar energy production depends on which projects are commissioned when: the observed sequence of investment decisions need not be cost-efficient. We propose a methodology to assess dynamic misallocation by comparing the present value cost of realized investments to a counterfactual optimal sequence of investments. Applying our methodology to France between 2005 and 2021, we find that the observed trajectory of annual solar production could have been obtained at a present value cost almost 30% lower than its realized value. Our optimized counterfactual suggests that investments in residential solar should have on average been postponed by 7 years, while investments in medium and large-scale installations should have occurred 2 to 4 years earlier.

Keywords: solar photovoltaic, renewable energy, renewable policies, misallocation, cost-efficiency
JEL: H23, Q42, Q48, Q52

*The views expressed in this paper are those of the authors alone and may not represent the views of Engie or any other institution. We are grateful to seminar participants at the Paris School of Economics, the XI international academic symposium of the Institut d'Economia de Barcelona, the 1st PhD conference of the European Electricity Markets Chair (Paris Dauphine), the 9th International Symposium on Environment & Energy Finance Issues, NEOMA workshop on the decentralization of the energy system, the 28th EAERE Annual Conference, the 18th IAEE European Conference, the FSR Loyola Autumn Research School and the 9th Joint Research Workshop Public Policies and Sustainable Development (Paris) for their feedback and suggestions. Remaining errors are ours.

†Paris School of Economics, Ecole des Ponts. Email: nicolas.astier[at]psemail.eu
‡Engie and Paris School of Economics. Email: nicolas.hatem[at]psemail.eu

1 Introduction

Increasing the share of solar photovoltaic (PV) energy in electricity supply is a cornerstone of both existing and envisioned climate policies.[1] Therefore, investments in solar energy, which have exceeded $100 billion/year globally over the past decade (IRENA and CPI, 2020), are expected to remain high or even to accelerate in the near future.

Meeting the objectives of the Paris agreement, however, requires an energy transition of an unprecedented speed. In this context, improving the cost-efficiency of investments in solar energy, that is how much renewable electricity is produced per dollar invested, is of critical importance. Indeed, any significant inefficiency regarding when and where solar facilities are deployed means that more solar energy could have been generated with the same amount of private investments and public subsidies.

Following Callaway et al. (2018), the literature on the misallocation of solar investments has mainly focused on inefficiencies regarding *where* solar facilities are built. This body of work typically highlights that a social planner would have located solar facilities differently than what is observed in practice. In contrast, this paper studies inefficiencies regarding *when* investments in different types of solar installations took place. Indeed, solar PV installations differ significantly in terms of location (available resource, proximity to the existing power grid), size (e.g. residential vs. utility-scale installations) and technology. As a result, the levelized cost of electricity (LCOE), that is, the ratio of the sum of the discounted costs of an installation and the sum of its discounted energy output, is highly heterogeneous across solar facilities. Therefore, an exogenously given trajectory of annual solar energy generation can be met at very different present value costs. In particular, the observed timing of investment decisions need not be cost-efficient.

Building on this observation, we propose a methodology to quantify the magnitude of dynamic misallocation in solar investments, which we apply to France for the period 2005-2021. We find that the relative cost of dynamic misallocation may have been as high as 30%,

[1] Consistently, reports from the Intergovernmental Panel on Climate Change (e.g. Pörtner et al. (2022)) consider solar energy to be one of the main potential contributor to net emission reductions (see for example Figure SPM.7).

meaning that the exact same amount of solar electricity could have been produced each year at only 70% of the realized present value cost.

The magnitude of this dynamic misallocation is significantly larger than other estimates of (static) misallocation found in the literature (Sexton et al., 2021; Lamp and Samano, 2023). For example, Lamp and Samano (2023) estimate that reallocating residential solar facilities across space in Germany could have increased their social value by about 5% relative to their realized social value. Similarly, Colas and Saulnier (2023) estimates a misallocation of 6-11% for residential PV in the United States. In contrast to these papers, however, we study all categories of solar PV installations rather than focusing only on residential PV. When adding the constraint that each category of PV installations must meet the trajectory of annual output that they have actually produced, for example to capture a situation where learning-by-doing would be entirely category specific, we assess misallocation to be about 6%, in line with previous estimates.

Misallocation in solar investments necessarily arises from some source of heterogeneity in photovoltaic installations. Existing studies usually focus on the heterogeneity in the gross social *marginal value* of electricity. For example, Callaway et al. (2018) note that different solar facilities displace generation output from power plants with different fuel costs and environmental externalities. Lamp and Samano (2023) study the same source of heterogeneity, along with differences in solar irradiation. Focusing attention on differences in which power plants are displaced at the margin makes indeed perfect sense for countries that span across a large geographical area, such as the United States (Callaway et al., 2018; Sexton et al., 2021), and/or that experience high levels of congestion in their transmission grid, such as Germany (Lamp and Samano, 2023).

In contrast, we consider a situation where the marginal social value of 1 kWh of electricity generated by a solar facility is roughly uniform across space. This simplifying assumption is indeed realistic for a country with a relatively small surface and whose transmission grid experiences little congestion, such as France.[2] In addition, in the case of France, small-

[2]Consistently, Callaway et al. (2018) note "*We find that variation in the quantity of emissions displaced by wind, solar, and efficiency resources is significant across regions but limited across resources within a region*".

scale installations connected to the distribution grid have been found to deliver negligible grid savings relative to larger scale transmission-connected facilities (Astier et al., 2023). However, solar facilities are very heterogeneous in terms of levelized costs, notably due to economies of scale and differences in solar irradiation. For example, typical estimates suggest that the LCOE of a residential rooftop installation is two to three times higher than the LCOE of a large-scale ground-mounted solar farm. Even within a given type of installations, for example 90kW rooftop facilities, LCOEs remain very heterogeneous because they depend on average irradiation, roof orientation, how difficult it is to install the panels, etc.

Building on Asker et al. (2019), we consider the dynamic optimization problem faced by a social planner who must choose, over a period of several years, the commissioning dates of a given set of solar facilities. The optimization is made under the constraint to meet an exogenously given trajectory of total annual production, corresponding to the total solar output actually generated each year. Because we assume the marginal social value of electricity at a given point in time to be uniform across installations, any combination of commissioning dates that meets these annual production targets generates the same gross social surplus. The social planner's objective is therefore to minimize the present value of investment and operating costs under a set of annual aggregate output constraints. Misallocation may then arise in a dynamic sense: the present value cost of realized investments can be significantly larger than the present value cost of the optimal sequence of investments.

In a number of ways, our methodology provides a lower bound of the magnitude of dynamic misallocation. We indeed shut down three important channels of potential inefficiencies. First, as discussed above, we assume the social value of solar energy to be uniform across space. Second, we restrict attention to locations where solar facilities actually exist today. In other words, we do not allow for changes in the size or the location of solar installations. Therefore, our methodology does not suffer from possible measurement errors regarding the actual feasibility of installing a solar power plant at a given location, which would otherwise be confounded with misallocation. In addition, we take the installed capacity of a given unit as exogenous, and hence do not allow for inefficiencies in the size chosen for

each unit. Third, our optimized trajectory for annual solar production matches the realized trajectory. We thus do not allow for the possibility that it may have been welfare improving to produce more (or less) solar electricity in any given year.

Despite these conservative assumptions, we find dynamic misallocation to be very large when applying our methodology to the case of France. More precisely, we estimate that the exact same amount of solar electricity could have been generated each year from (a subset of) the exact same fleet of solar installations for about 70% of the realized present value cost. The comparison between the realized and optimized sequences of investments suggests that investments in residential solar should have on average been postponed by 7 years, while investments in medium and large-scale installations should have occurred 2 to 4 years earlier. These results are consistent with the observation that, over our period of interest, residential solar has benefited from significantly higher subsidies per unit of output than utility-scale installations. Arguably, this stronger policy support for residential solar may have pursued other policy objectives besides efficiency. If so, the dynamic misallocation we assess reflects the social opportunity cost of such objectives.

This work contributes to a vast literature on the design and efficiency of renewable electricity policies (e.g. Borenstein (2012); Fell and Linn (2013); Abrell et al. (2019a,b); Ambec and Crampes (2019); Abrell and Kosch (2022), among many others), as well as their distributional impacts (e.g. Reguant (2019); Liski and Vehviläinen (2020)). Within this literature, the cross-sectional heterogeneity in the marginal social value of renewable facilities plays a very prominent role (Cullen, 2013; Novan, 2015; Wolak, 2016; Callaway et al., 2018; Gillingham and Ovaere, 2020; Sexton et al., 2021; Lamp and Samano, 2023). In contrast, little to no attention has been dedicated to the significant heterogeneity in the investment costs of solar units, and its potential implications in terms of dynamic misallocation. This article aims to fill this gap.

The rest of the paper is organized as follows. Section 2 provides background information on the different sources of heterogeneity in the LCOE of solar facilities. Section 3 formalizes the problem we study and defines our concept of dynamic misallocation. Section 4 discusses how solar energy was deployed in France over the past two decades and describes the data

we use in our empirical application. Section 5 presents our main results. Section 6 discusses robustness checks and policy implications. Finally Section 7 concludes.

2 Solar energy and misallocation

2.1 Heterogeneity in solar installations

Solar facilities can differ along many dimensions, and can therefore be very heterogeneous in both their private and social net values.

First, the solar resource available at a given location depends on its average solar irradiation, which increases as one moves closer to the equator. France has relatively similar solar irradiance coefficients across its territory in comparison to many other countries. Yet, significant differences exist between the Northern and the Southern parts of the country. Figure B.14 in Appendix B illustrates this heterogeneity. In particular, it shows that the expected output per unit of installed capacity can be twice higher in the best locations than in the worst locations.

Second, solar panels of different technologies are characterized by different efficiencies in converting solar irradiation into electricity. For example, Crystalline Si solar cells have a maximum efficiency of 26% whereas Amorphous Si have a maximum efficiency of only 14% (Allouhi et al., 2022). In France, most PV panels are Crystalline Si (60% of ground PV and 90% of rooftop PV projects) and a minority are CdTe solar panels (27% of ground projects and 4% of shelters),[3] the latter having a maximum efficiency of about 22% (Allouhi et al., 2022). Differences induced by heterogeneous technologies across solar PV projects in France can thus typically reach a few percentage points. This difference can however be much higher between projects that install a tracking system and those that do not.[4]

Third, investments costs per unit of installed capacity can differ significantly across projects and over time. For example, fixed costs represent a much lower share of total costs for a 300,000 kW ground-mounted project than for a 3 kW unit. Moreover, the latter is

[3] Source: CRE (2019).

[4] A tracking system changes automatically the orientation of solar panels to follow the trajectory of the sun in order to maintain the optimum angle to receive the highest solar radiation possible.

usually installed on a roof, requiring complex tasks which may be hard to standardize or automate. In addition, installation costs have changed significantly over time. In particular, the prevailing prices of PV modules have decreased dramatically over the past couple of decades, notably thanks to economies of scale and technological progress in manufacturing.

Finally, for medium and large-scale installations, connection costs to the electricity grid scale approximately proportionally to their distance to the closest substation. In contrast, small-scale rooftop installations do not require new power lines to connect to the grid.

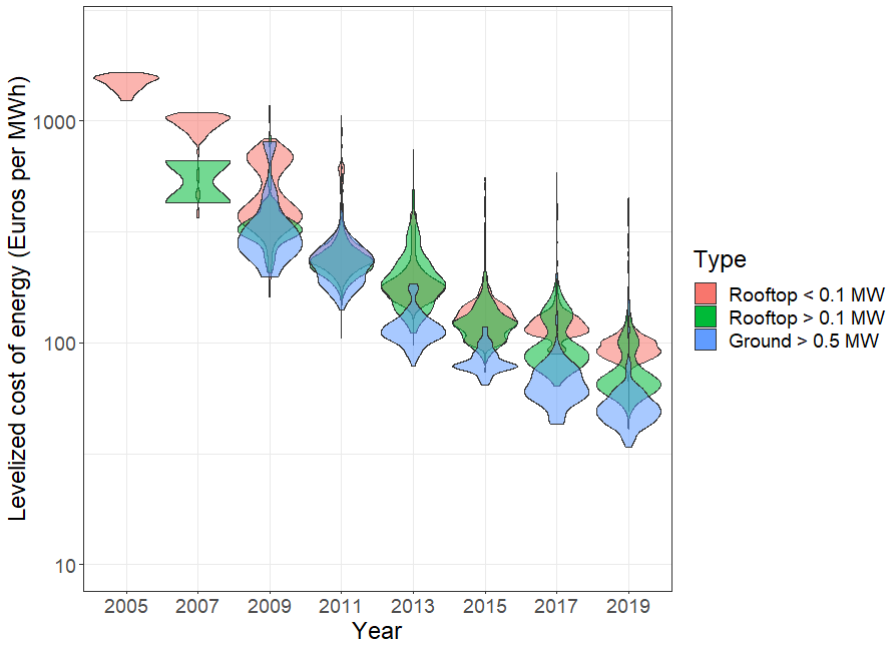


Figure 1: Distributions of LCOEs (€/MWh) for ground-mounted and rooftop PV facilities from our dataset (see Appendix B for details).

In practice, the above sources of heterogeneity are compounded. Therefore, the levelized cost of electricity (LCOE), that is, the ratio of the sum of the discounted costs of an installation and the sum of its discounted energy output, is highly heterogeneous across solar facilities. In particular, as illustrated in Figure 1, the LCOE of residential solar projects can be two to three times larger than the LCOE of utility-scale projects. Figure 1 also shows that LCOEs have decreased substantially over time across all types of solar installations.

2.2 Possible channels of misallocation

Conceptually, the net social value of the electricity produced by a solar installation is:

$$\text{Net social value} = \text{Opportunity cost of electricity} - \frac{\text{Total levelized costs}}{\text{Total levelized production}} \quad (1)$$

In this expression, the opportunity cost of electricity corresponds to the social cost of displaced power plants (in expectation, weighted by hourly PV output), that is, generators that are marginal when the solar facility generates electricity. This opportunity cost includes private costs (e.g. fuel) as well as externalities (carbon emissions, local pollutants such as PM 2.5, etc.). In what follows, we assume that this opportunity cost of electricity is uniform, or at least that its variance across solar installations is much smaller than the variance in the second term of expression (1). This simplifying assumption is motivated by the fact that we use France as a case study for our empirical application. First, France has a relatively small surface, allowing us to assume, as a first approximation, that the normalized profile of solar production is roughly similar across facilities. Second, under the prevailing market design over our period of interest, the transmission grid was seldom congested. Therefore, the power plant displaced by a solar facility in a given hour only very rarely depended on the location of the solar facility. Third, Astier et al. (2023) find that distributed solar installations in France are unlikely to deliver substantial savings in future grid investments, implying that small-scale installations connecting to the distribution grid provide negligible external benefits relative to larger transmission-connected facilities.

Total levelized costs include capital expenditures, connection costs to the power grid and operation and maintenance costs. These costs vary significantly both across types of facilities (due to economies of scale) and over time (learning-by-doing). Finally, total levelized production represents the discounted amount of electricity generated. This quantity varies by location (available resource), technology, as well as over time (technological improvements).

In most existing studies, misallocation stems from the heterogeneity in the opportunity cost of electricity and/or in total levelized production. In other words, total levelized costs are implicitly assumed to be roughly uniform across installations (per unit of capacity).

In contrast, we assume in what follows that the opportunity cost of electricity is roughly uniform across installations, but we account for the heterogeneity in both total levelized production and total levelized costs.

3 Dynamic misallocation

3.1 General framework

Assumptions and notations

Our framework builds on Asker et al. (2019), where the authors assess the magnitude of dynamic misallocation in the context of oil extraction. We adapt their approach to account for the characteristics of solar facilities that differ from those of oil fields.

Misallocation is defined relative to an "optimal" counterfactual. The choice of this counterfactual is of course critical. In particular, if the optimal counterfactual allows for outcomes that are infeasible in practice, misallocation will be partially confounded with measurement errors and, therefore, over-estimated. In order to avoid this pitfall to the largest extent possible, we make a number of conservative assumptions.

First, we keep constant the trajectory of total annual electricity generation from solar facilities $\{E_t\}_{t=1...T}$. In other words, the optimized sequence of solar investments has to produce in each year t the same number of kWh as the observed aggregate output E_t from the investments that were made in practice. We thus do not allow for the realized trajectory of solar generation being inefficient in terms of the aggregate gross social value that was created. Instead, given our assumption that the social opportunity cost of electricity is uniform across solar units, the total gross social value from solar energy is the same under both the realized and optimized scenarios.

Second, we assume that solar facilities (i) can only be installed at the location where they actually exist today; and (ii) cannot be sized differently than their actual size. We therefore do not face the risk of being mistakenly optimistic about how much solar capacity can be realistically installed in a given region, or about how large a given installation could have been. This feature is particularly important because local acceptability constraints, as

well as technical and administrative constraints, are always imperfectly observed and can prove very hard to capture accurately at a disaggregated level. In particular, restricting attention to existing solar installations alleviates the concern that land use conflicts would put constraints on how much utility-scale capacity can actually be built.

We index solar facilities by i and denote with $x_{it} \in [0,1]$ the fraction of unit i that is commissioned in year t.[5] We further denote with c_{it} the total present value cost (investment, connection to the grid, O&M discounted at commissioning date) of unit i when it is commissioned in year t (in euros). Note that this definition allows for (exogenous) learning-by-doing, that is, situations where c_{it} decreases with t. We further discuss learning-by-doing in Section 6. Finally, the output in year t of a plant i that was commissioned in year t' is denoted with $e_{it't}$ (in kWh). This formulation allows to account for both technological progress in conversion efficiency and the fact that the efficiency of solar panels decreases over time due to wear and tear.

Social planner problem

A social planner discounts future cash flows at a rate $\rho \in [0,1]$ and can build solar units from an exogenously given set of installations. He seeks to optimize the sequence of investments in solar facilities under the constraint to meet an exogenously given trajectory of aggregate production $\{E_t\}_{t=1\ldots T}$. He therefore faces the following problem:[6]

[5] We do not constrain x_{it} to be an integer because, in the case of our optimization problem, the optimal solution sets the value of x_{it} to either 0 or 1 for the vast majority of installation-year pairs. This convex relaxation greatly simplifies our numerical computations without impacting our results in any significant way.

[6] Note that the formulation of the problem implicitly assumes that the termination value does not depend on the chosen sequence of investments. This assumption does not hold exactly in our numerical application due to differential technological progress with impacts beyond year T (conversion efficiency and O&M costs). However, differences in termination values are small relative to total costs, so that neglecting differences in termination values represents a sensible simplification.

$$\min_{x_{it}} \sum_{t=1}^{T} \rho^t \left(\sum_{i=1}^{N} x_{it} c_{it} \right)$$

s.t.

$$\begin{aligned}
\forall t \in \{1, ..., T\}, \quad & \sum_{i=1}^{N} \left(\sum_{t'=1}^{t} x_{it'} e_{it't} \right) \geq E_t && (\rho^t \lambda_t) \\
\forall i \in \{1, ..., N\}, \quad & \sum_{t=1}^{T} x_{it} \leq 1 && (\bar{\mu}_i) \\
\forall i \in \{1, ..., N\}, \forall t \in \{1, ..., T\}, \quad & x_{it} \geq 0 && (\rho^t \underline{\mu}_{it})
\end{aligned}$$

The objective function is the present value total cost of solar units (capital costs, grid connection and O&M). The first set of constraints corresponds to the target trajectory of annual solar generation $\{E_t\}_{t=1...T}$. For a given unit i in a given year t, the sum $\sum_{t'=1}^{t} x_{it'} e_{it't}$ is positive if, and only if, the plant has been commissioned by year t (otherwise, $x_{it'} = 0$ for all $t' \leq t$). If the unit has indeed been commissioned by year t, we then have $\sum_{t'=1}^{t} x_{it'} e_{it't} = e_{it_c(i)t}$ where $t_c(i)$ is the commissioning year of unit i ($x_{it'} = 0$ for $t' \neq t_c(i)$). Therefore, $\sum_{t'=1}^{t} x_{it'} e_{it't}$ corresponds to the output of unit i in year t. The sum of production levels in year t across all units must be greater or equal to the target amount of solar generation E_t for that year. The second set of constraints ensures that a facility can be commissioned only once. Note that installed capacities (in kW) are not explicitly modeled as variables, but are instead indirectly captured by the variables c_{it} and $e_{it't}$. Indeed, both total cost c_{it} (in euros) and yearly output $e_{it't}$ (in kWh) will be larger for bigger units. Finally, the last set of constraints reflects the fact that solar units are physical assets. As a result, it is not possible to "short sale" generation from inefficient units to trade it off against generation from more efficient units.

A key observation is that the social planner is facing a linear optimization problem, which can be solved with a wide range of available software.[7]

Measuring misallocation

If we denote with $\{x_{it}^*\}_{i=1...N,t=1...T}$ the optimized investment decisions, we can compute

[7] We use the solver Gurobi in what follows.

the cost-efficient present value cost PV^* of meeting the trajectory of annual solar generation targets as:

$$PV^* \equiv \sum_{t=1}^{T} \rho^t \left(\sum_{i=1}^{N} x_{it}^* c_{it} \right)$$

However, the realized investments are instead $\{x_{it}^0\}_{i=1...N,t=1...T}$, with a corresponding present value cost PV^0:

$$PV^0 \equiv \sum_{t=1}^{T} \rho^t \left(\sum_{i=1}^{N} x_{it}^0 c_{it} \right)$$

We then define (relative) dynamic misallocation m as:

$$m \equiv \frac{PV^0 - PV^*}{PV^0}$$

This metric captures the fraction of the total present value cost of realized investments that may be considered as inefficient.[8]

3.2 Special case of static LCOEs

In order to build intuition, consider the simplest case where the cost and expected output of each installation do not change over time:

$$\forall i, t,\ c_{it} = c_i \text{ and } \forall i, t, t',\ e_{itt'} = e_i$$

Up to a common scaling factor, we can define the LCOE L_i of solar facility i as:[9]

$$L_i \equiv \frac{c_i}{e_i}$$

Without loss of generality we assume that solar units are indexed such that:

$$L_1 \leq L_2 \leq ... \leq L_{N-1} \leq L_N$$

[8]Note that since, by definition, $PV^* \leq PV^0$, choosing to express misallocation as a fraction of realized costs rather than of optimized costs mechanically yields lower values for relative misallocation. For example, a misallocation of 50% with our metric means the realized trajectory of investments is twice as expensive as the optimal trajectory.

[9]Since we defined c_i as total present value costs (including O&M costs) and e_i as annual electricity generation of unit i, the proper definition of the LCOE of unit i is $LCOE_i \equiv \frac{c_i}{e_i \sum_{t=1}^{\tau} \rho^t}$ where τ is the lifetime (in years) of a solar unit. We thus have $L_i = \left(\sum_{t=1}^{\tau} \rho^t\right) LCOE_i$.

Then, denoting $i_0 \equiv 0$ and $i_T \equiv N$, one can show (see Appendix D) that there exist thresholds:

$$i_0 \leq i_1 \leq ... \leq i_{T-1} \leq i_T$$

such that it is optimal to commission in year t the units whose index i is such that:

$$i_{t-1} < i \leq i_t$$

In other words, the cost-efficient investment decisions boil down to installing solar facilities in increasing order of LCOE. In the first year, the facilities with the lowest LCOE are built until total expected output reaches E_1. In the second year, the remaining facilities with the lowest LCOE are then commissioned until the total expected output of the whole generation fleet reaches E_2. And so on and so forth.

In the general case, however, the incentive to first install the facilities with the lowest LCOE has to be traded-off against the prospects of future improvements in LCOE. To see this, consider two facilities A and B with similar LCOEs today ($L_A^0 = L_B^0 - \epsilon$ with $0 < \epsilon << 1$) but different prospects in terms of future technological improvements. Specifically, assume that the social planner expects that, at a later date t, we will have $L_A^t << L_B^t$. In this situation, it is preferable to first install unit B in order to be able to benefit from the larger technological improvements that unit A will enjoy. Therefore, we cannot derive unambiguous analytical results in the general case. Instead, we turn to a realistic numerical case study.

4 Application to France

4.1 Institutional background

France ranks fifth among countries of the European Union in terms of installed solar PV capacity, with 12.5 GW installed as of January 1st, 2022 (France Territoire Solaire, 2022). France is committed to further increase its installed capacity in order to meet the renewable generation objectives set by the European Union. Consistently, the French government's

multi-annual investment plan for electricity generation targets an increase in annual installation rates.

Over the past two decades, France has relied on a number of public support mechanisms to promote solar energy. For the most part, these mechanisms have consisted of feed-in-tariffs (FiTs). These tariffs were introduced in the 2000s and represent a commitment to purchase energy at a fixed price for a period of 20 years.

Below a certain size threshold (initially 100 kW), any installation is eligible to receive an exogenously given but regularly updated FiT. The received tariff (in €/MWh), however, decreases with the size of the installation. In addition, the level of these FiTs has decreased over time, somewhat following the decrease in the costs of PV modules. For instance FiTs ranged between 300 and 550 €/MWh in 2006, and between 90 and 180 €/MWh in 2021.

Table 1: Annual PV capacity (MW) auctioned by the French Energy Regulation Commission broken down by categories.

	Auctions before 2016		Auctions after 2016	
Categories (MW)	Rooftop 0.1 - 0.25	All >0.25	Rooftop >0.1	Ground >0.5
Year	Annual volumes auctioned (MW)			
2012	240	450		
2013	300	400		
2014	120			
2015		400		
2016	240			
2017			450	1000
2018			425	1200
2019				1700
2020			450	1000
2021			450	1100

Soure: adapted from `www.photovoltaique.info`.

Larger projects (above the automatic eligibility threshold for FiTs) have to participate in technology-specific (e.g. ground-mounted versus rooftop) auctions in order to be granted a tariff. These auctions were introduced in 2011, partially as a response to the significant gap between the levels of FiTs and the declining costs of PV installations. They are organized by the French Energy Regulation Commission (CRE), using a pay-as-bid format with

technology-specific capacity targets set by the government. Table 1 shows the total volumes auctioned since 2011, broken down by year and (simplified) project categories.

The vast majority of solar units in France are distributed, that is, consist of relatively small-scale installations that connect to the distribution grid. More precisely, the French solar generation fleet may be roughly decomposed as follows (France Territoire Solaire, 2022). Almost half (46%) of the total installed capacity consists of rooftop power plants with a size lower than 250 kW.[10] Most auctions for larger installation (ground-mounted and parking shelter) were only introduced after 2016, with a size cap at 17 MW. Therefore, only 6% of the total solar capacity consists of larger installations connecting to the transmission grid.

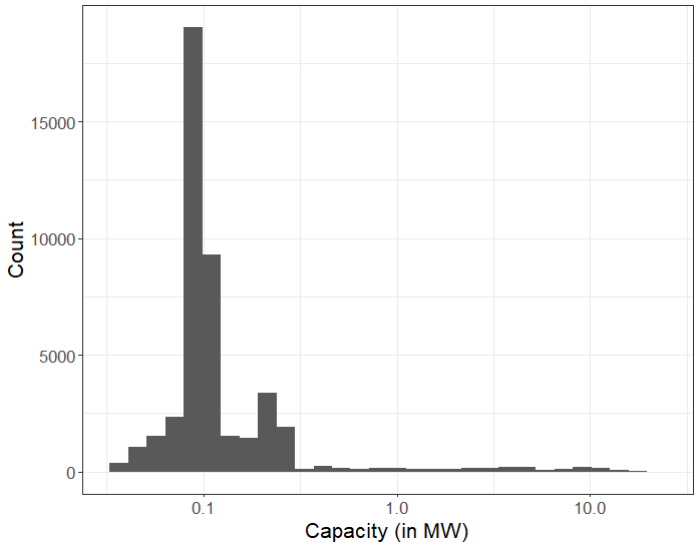


Figure 2: Histogram of PV project sizes for units bigger than 36 kW.

Figure 2 shows the distribution of the size of non-residential PV projects (installations bigger than 36 kW) as of 2022. We observe that many installations have bunched at 100 kW, which likely stems from attractive FiTs rates before 2011 and specific rooftop PV auctions before 2016.

[10]Indeed, a significant fraction of the auctions that took place before 2016 were restricted to rooftop PV projects between 100 and 250 kW (Table 1). This category also includes residential PV projects, which represent 13% of the total capacity.

4.2 Data sources

We build a dataset that keeps track of (i) when and where solar projects were commissioned; (ii) the likely costs of solar projects as a function of both their characteristics and commissioning date; and (iii) the observed or simulated yearly output of each project in a given year, also as a function of both their characteristics and commissioning date.

Our main data source is a public registry listing the universe of power plants in France.[11] We observe 50,000+ installations commissioned between 2005 and 2021.[12] The dataset reports the location of solar units at the county ("commune") or sub-county ("IRIS") level, as well as their installed capacity and, for most of them, realized annual output and upstream substation.

We define categories of installations based on size bins and on whether they are rooftop or ground-mounted. We then use several public resources to estimate how the investment cost and the conversion efficiency of the different categories of solar facilities have evolved over time (IRENA, 2020; CRE, 2014, 2019). Appendix B details how we make use of these various data sources. For ground-mounted installations, grid connection costs are assessed based on the distance to the upstream (or, when unknown, closest) grid substation.[13] Specifically, we compute the as-the-crow-flies distance between the substation to which the installation connects and the centroid of the county where it sits. This distance is then multiplied by a connection cost of 100 €/meter when the unit connects to the medium voltage grid, and by a connection cost of 1,000 €/meter when it connects to the transmission grid (see Appendix B).[14]

Finally, the public registry of power plants provides, for most medium and large-scale units, the total energy they produced in 2022. When no annual output was provided in the dataset, we retrieved expected annual capacity factors at different locations from the website

[11] https://www.data.gouv.fr/fr/datasets/registre-national-des-installations-de-production-et-de-stockage-delectricite-au-31-12-2022-2/, last accessed on 30 August 2023.

[12] The smallest installations are listed as bundles rather than as individual units (see below), so that the actual number of individual units (when counting each small rooftop unit separately) is much higher.

[13] For the vast majority of units connecting to the medium or high voltage grids, we observe the substation to which they connect. Units connecting to the low-voltage grid are rooftop installations and thus do not require dedicated power lines to connect to the closest substation.

[14] All costs are expressed in real 2019 euros.

renewable ninja (Pfenninger and Staffell, 2016; Staffell and Pfenninger, 2016).[15]

4.3 Description of the main variables

Categories of installations and observed commissioning dates (x_{it}^0)

In what follows, we assign installations to categories that we use to assess cost functions (see below and Appendix B). These categories are defined according to (i) size bins; and (ii) whether a given installation is rooftop or ground-mounted.

While the public registry of power plants reports the installed capacity of each unit, it does not explicitly specify whether a given installation is a ground-mounted or a rooftop unit. Therefore, we assess whether each observation is a rooftop or a ground-mounted unit using different assignment strategies that are detailed in Appendix C.[16]

In addition, for any installation i, we further observe in the public registry its commissioning date. However, for residential PV (rooftop $\leq$ 36 kW), the registry aggregates units at the county level. Because residential units in a given county are commissioned at different dates, detailed information on commissioning dates is lost with spatial aggregation. To address this issue, we proceed as in Astier et al. (2023) and use public information from the French Department of Energy to construct time series of aggregated residential PV capacity at the sub-regional ("departement") level (see Appendix C).

Table 2 shows the eight categories of solar installations we use in our analysis, along with relevant descriptive statistics.

Cost functions (c_{it})

For each solar unit i, we build a time series variable c_{it} that represents the total present value cost (investment and discounted O&M) of the unit, should it be commissioned in year t. The variables c_{it} are calibrated using category-specific cost functions (see Appendix B). These costs functions include the cost of PV modules (retrieved from public sources), other

[15] Given the high degree of spatial correlation in solar irradiation, we only sampled the locations of the substations which represent over 2,000 locations in France.

[16] Whether a given unit is roof- or ground-mounted is assessed based on (i) its name (when available), (ii) the prevailing size limits in technology-specific auctions, and (iii) geolocalized data on photovoltaic facilities retrieved from OpenStreetMap. Around 1,000 observations could not be assigned and are allocated proportionally to the total installed capacity for each category at national level.

Table 2: Descriptive statistics by categories of photovoltaic facilities.

Category	Obs.	Capacity Mean (MW)	Capacity Total (GW)	Years of commissioning [Perc. 20 - Perc. 80]	Capacity factor mean	Connection length mean (km)
Rooftop aggregated $< 0.036\ MW$	1,501	1.5	2.3	2008 - 2018	0.14 (0.01)	0
Rooftop $0.036 - 0.1\ MW$	25,669	0.09	2.2	2013 - 2021	0.14 (0.02)	0
Rooftop $0.1 - 0.5\ MW$	8,599	0.19	1.6	2011 - 2019	0.14 (0.02)	0.4 (1.8)
Rooftop $0.5 - 2.5\ MW$	303	1.2	0.3	2011 - 2020	0.14 (0.03)	5.1 (3.5)
Rooftop $> 2.5\ MW$	51	4.7	0.2	2012 - 2020	0.15 (0.03)	5.1 (4.1)
Ground $0.5 - 2.5\ MW$	545	1.3	0.7	2011 - 2020	0.14 (0.03)	5.9 (5.2)
Ground $2.5 - 10\ MW$	693	5.6	3.9	2014 - 2021	0.16 (0.03)	6.0 (3.9)
Ground $> 10\ MW$	181	12	2.2	2013 - 2021	0.16 (0.02)	5.9 (4.3)

Note: Perc. 20 and Perc. 80 denote the lower and upper quintiles of the commissioning year distribution. Standard deviations for capacity factors and connection lengths are enclosed in parentheses.

investment costs (that are assumed to decrease exponentially with category-specific growth rates), and grid connection costs (when relevant).[17]

Figure 3 displays the cost trajectories for the eight categories of installations. The smallest installations have cost trajectories that are (almost) always higher than the cost of other categories. In contrast, the cost trajectories for larger installations cross multiple times. For example, very large rooftop units (> 2.5 MW) start becoming more expensive than smaller rooftop units in 2019.

Energy output ($e_{it't}$)

For each solar unit i, we also compute energy output variables $e_{it't}$ which represent the energy produced by installation i in year t if it is commissioned in year t' (in particular, $e_{it't} = 0$ if $t < t'$). These installation-specific yearly outputs are built as follows. We first retrieve from the public registry of power plants the annual capacity factor of each installation for the year 2022.[18] In order to estimate capacity factors for other years and other possible

[17]The decomposition of investment costs for the project categories larger than 100 kW are obtained from CRE (2014, 2019). Investment costs for the other categories are retrieved from IRENA (2020) and extrapolated for years before 2010. The costs of modules are retrieved from IRENA (2020). For further details, see Appendix B

[18]This information is available for 93% of observations. The capacity factor for the remaining installations are estimated using the renewable ninja website (Pfenninger and Staffell, 2016; Staffell and Pfenninger,

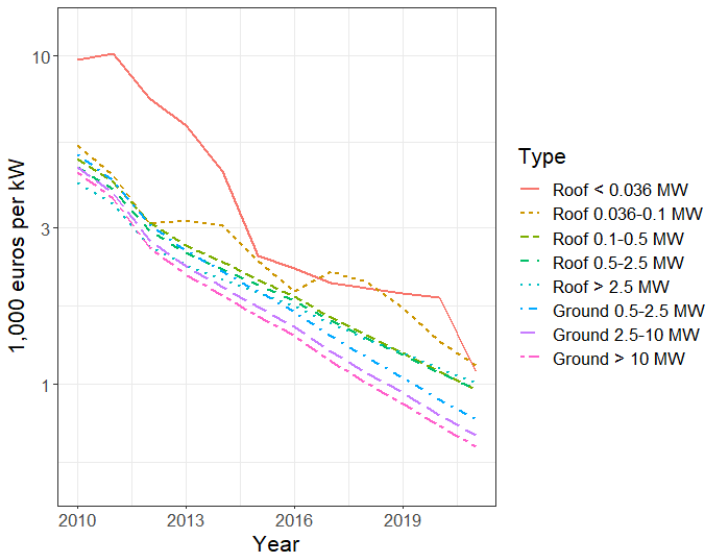


Figure 3: Cost trajectories (1,000 € per kW) by categories of photovoltaic facilities between 2010 and 2021 in logarithmic scale.

commissioning dates, we need to account for two compounding effects. On the one hand, the earlier the commissioning date, the lower the conversion efficiency of PV technology. On the other hand, the conversion efficiency of a given unit decreases over time due to wear and tear. We estimate both effects directly from our data. Specifically, because the registry of power plants has been published every year since 2017, we use previous editions to build a panel of annual energy output over 2017-2022 for a set of 16,000+ installations. Using this dataset, we estimate a -1% annual rate for depletion and 1% annual rate for technological improvement (see Appendix B.4).

The average capacity factors in 2022 are reported in Table 2, broken down by category. We notably observe that larger facilities have on average higher capacity factors. Indeed, larger installations (i) may locate in places with higher solar irradiation than average, (ii) may be equipped with tracking systems; and (iii) were installed towards the end of our study period. Unit-level capacity factors are not directly observed for the smallest installations, which represent 35% of total installed capacity. Our method to assign capacity factors to

2016). The corresponding installations are mostly residential PV systems, to which we assign capacity factors assumed to be uniform over the area supplied by a given substation (see Appendix D).

these units relies on locally uniform simulated values, and therefore under-estimates the true heterogeneity in actual capacity factors. This method for inputting missing data represents another channel through which our assessment of misallocation can be considered as conservative (i.e., a lower bound).

5 Main results

5.1 Dynamic misallocation

We use our dataset to assess the magnitude of dynamic misallocation in the case of France over the period 2005–2021. We find that the present value cost of the optimal sequence of solar investments would have been 29% cheaper than the realized present value cost (i.e., $m = 0.29$). Specifically, we estimate the optimal trajectory to have a present value cost of 18.4 billion euros, while the observed trajectory has a present value cost of 25.9 billion euros.[19] It corresponds to a decrease in the levelized cost of solar energy of about 50 euros per MWh.[20]

Figure 4 compares the realized and optimized trajectories of cumulative installed capacity, broken down by category. Our main result seems to be driven by two main effects. First, the optimization program leverages economies-of-scale by commissioning the larger solar projects much earlier than they actually were. Table 3 reports the average change in commissioning dates for each project category. The optimization program postpones on average residential PV by seven years, and installs ground-mounted and rooftop PV facilities bigger than 2.5 MW two to four years before their observed commissioning dates. Second, the optimization program anticipates differential trends in the decrease in investment costs. For example, in order to benefit from lower investment costs, it delays medium-size rooftop PV (between 0.1 and 0.5 MW) by three years compared to small-size rooftop (between 36 kW and 0.1 MW). Indeed, the investment costs of small-size rooftop PV increase between 2016

[19] Present values are computed as of 2005 with a discount rate of 4.5% (real), which is consistent with French government's guidelines for public infrastructures.

[20] The difference in LCOE is computed by taking the ratio of the present value cost of misallocation and the total levelized energy produced by the solar fleet over a 20 year period after commissioning (assuming a depletion in energy output of 1% per year and a discount rate of 4.5%).

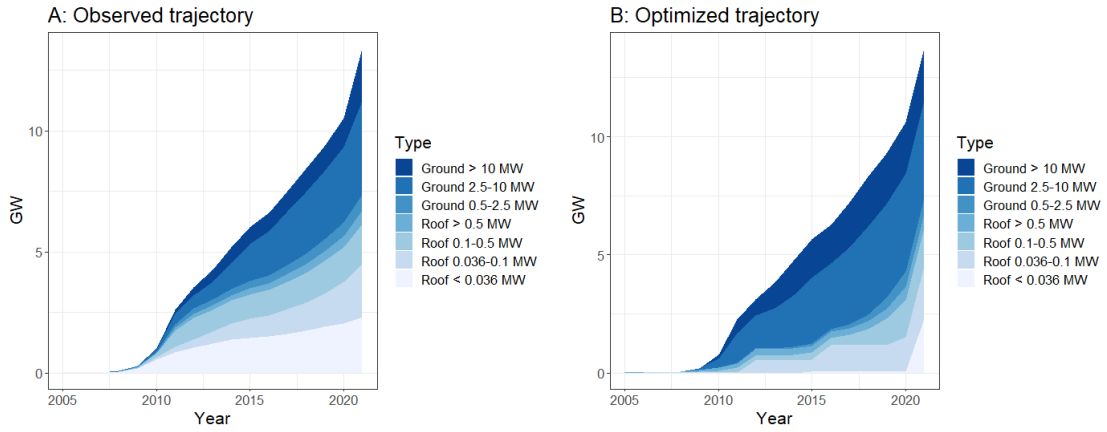


Figure 4: Realized (left panel) vs optimized (right panel) trajectories (cumulative installed capacity).
Note: Rooftop units of size ranging between 0.5-2.5 MW and greater than 2.5 MW are merged together for more visibility.

and 2019. As a result, the optimization program allocates a substantial fraction of small-size rooftop PV installations before the cost increase, with no installations of small-size PV occurring between 2016 and 2019.

Table 3: Capacity weighted average change in commissioning dates (year observed - year optimal) by category

Project Type	Ground	Rooftop
N obs.	1,419	36,123
Project Segment (MW)		
> 10	4	
2.5 - 10	2	4
0.5 - 2.5	-2	-2
0.1 - 0.5		-3
0.036 - 0.1		0
< 0.036 (Residential)		-7

5.2 Sensitivity analyses

Because they are expressed in present value terms, our results are sensitive to our choice of discount rate. Figure 5 shows, however, that the order of magnitude of misallocation remains similar for a wide range of discount rates. This counter-intuitive result stems from

the fact that we account for (exogenous) technological progress.[21] Indeed, in absolute terms, the cost of smaller installations have decreased significantly more than the cost of larger installations. A large fraction of the misallocation is thus driven by the early commissioning of a large number of small installations (see below).

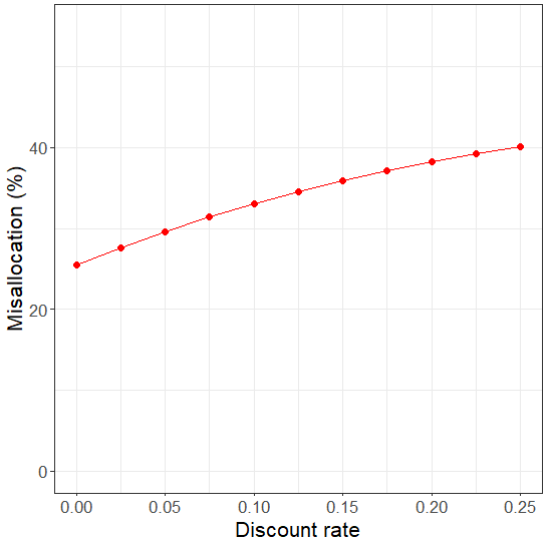


Figure 5: Effect of varying discount rates on the part of misallocation (%)

One of the main inputs of our optimization program are investment cost time series for each project category. These time series are imperfectly observed and therefore calibrated with different data sources (see Appendix B). In particular, because we do not have data before 2010, we extrapolate investment costs functions for the period 2005-2010 assuming category-specific exponential trends. We test the sensitivity of the estimated misallocation to this extrapolation by looking instead at an extreme scenario where we set investment costs for 2005-2010 to their 2010 level for each project category. Resulting cost functions thus represent a lower bound for actual investment costs. Under this assumption, we estimate that dynamic misallocation decreases to 27%, with present value costs for the realized and optimized trajectories of respectively 25.1 billion and 18.2 billion euros. Our results are therefore robust to different extrapolation strategies for investment costs occurring in the early years of our study period. Indeed, despite their early commissioning (and therefore

[21] Section 6 provides an in-depth discussion of the role of learning-by-doing.

higher weight in present value terms), the projects commissioned before 2010 only represent a small share of the total solar capacity, and therefore have a relatively small impact on the total present value cost of solar investments.

6 Discussion

In this section, we explore various reasons that may help explaining the magnitude of the estimated misallocation.

6.1 Decomposing misallocation

We decompose misallocation into various components by proceeding as follows. We first define a partition of the set of installations (e.g. small, medium and big). We then define the misallocation associated to each subset in the partition as the obtained misallocation when only optimizing the commissioning dates of all installations but the ones in the considered subset (which are frozen to their realized trajectory). In other words, we compute the misallocation associated to given subset of solar installations as the obtained misallocation when the social planner cannot change the commissioning dates of installations in this subset.

We study two such partitions of the universe of solar installations. First, we distinguish the main categories of installations (ground-mounted, large rooftop, medium rooftop and residential). Second, we partition installations according to their commissioning dates, distinguishing four distinct periods that roughly map into different phases of public support regimes.

Misallocation decomposed by PV categories

First, we consider four mutually exclusive categories of PV projects:

- Ground-mounted installations, which were mostly developed through auctions.
- Large rooftop ($>$ 100 kW), which were under FiTs until 2011 and auctions after 2011.
- Small rooftop (between 36 kW and 100 kW), which were under FiTs for the whole period.

- Residential PV (< 36 kW), which were under FiTs for the whole period.

For each category of installations, we optimize the investment trajectory while freezing the commissioning dates of the corresponding installations. The obtained misallocation can be thought of, as a first approximation, as the contribution of the considered category to the total misallocation.[22]

Table 4: Decomposition of the misallocation by project category

Frozen category	Capacity share	Present Value (Billion Euros)	Difference with PV^* (Billion Euros)	Additional LCOE (Euros per MWh)
Ground	50%	19.9	1.5	20
Rooftop>100 kW	17%	18.8	0.4	17
Rooftop 36 to 100 kW	16%	18.7	0.3	15
Residential	17%	23.9	5.5	211

Note: Additional LCOE are calculated as the difference between the present value of category-constrained optimization and unconstrained optimization, relative to the total levelized energy output of the category, assuming a lifetime of 20 years and a depletion of the energy output of 1% per year.

Table 4 reports the obtained results. The present value costs of each constrained optimization are displayed, along with their difference with the unconstrained optimal present value cost. We observe that freezing the commissioning dates of residential PV installations induces the largest misallocation, despite the fact that these units only represent 17% of total installed capacity. Expressed as a ratio of absolute misallocation cost over the total levelized energy generated, the misallocation cost of the realized trajectory for residential PV is 211 euros per MWh. In contrast, the misallocation that may be attributed to other categories is lower than 20 euros per MWh.

Misallocation decomposed by time windows

A second approach is to optimize the timing of investment decisions while taking as given the investments that took place during a given time window. Reshuffling the commissioning dates of PV installations is then only allowed for units that were commissioned outside of the considered time window. We consider four different periods, roughly corresponding to different phases in the policies supporting renewables:

[22]Such a "contribution" has to be understood in an "accounting" sense since we do not claim that our methodology causally identifies the underlying mechanisms responsible for the observed misallocation.

- 2005-2008: early FiTs.
- 2009-2012: temporary moratorium and significant subsequent changes in support regimes.
- 2013-2016: first auctions.
- 2017-2021: new auction regimes with higher volumes of large-scale projects.

Again, the difference between the present value cost of the optimal trajectory and the present value cost of a constrained trajectory is interpreted, in an accounting sense, as the contribution of the considered time window to the misallocation.

Table 5: Decomposition of the misallocation by time windows

Frozen period	Capacity share	Present Value (Billion Euros)	Difference with PV^* (Billion Euros)	Additional LCOE (Euros per MWh)
2005-2008	1%	18.7	0.3	341
2009-2012	26%	23.7	5.3	118
2013-2016	23%	19.4	1	26
2017-2021	51%	19.9	1.5	23

Note: Additional LCOE are calculated as the difference between the present value of time-constrained optimization and unconstrained optimization, relative to the total levelized energy output of plants installed in the time period, assuming a 20 years project lifetime and a depletion of the energy output of 1% per year.

Table 5 reports the obtained results. The first support mechanisms between 2005 and 2008 generated the highest misallocation per unit of energy produced (341 euros per MWh). The misallocation then gradually decreased to reach 23 euros per MWh for auctions run after 2016. In absolute value terms, the period 2009-2012 is associated with the largest misallocation.

6.2 Learning-by-doing

So far, we have assumed that the present value total cost c_{it} of installation i in year t is exogenously given. This strong assumption deserves some discussion since learning-by-doing is perceived as one of the main market failure that policies supporting PV generation were trying to address. Credible amounts of learning-by-doing may indeed rationalize relatively high initial subsidies (Van Benthem et al., 2008), assuming this learning-by-doing is not appropriable, which is itself a debatable statement (Bollinger and Gillingham, 2019).

Therefore, an important caveat of our framework is that it does not explicitly account for learning-by-doing. This concern may however be less critical than one may expect.

First, a large fraction of the investment cost of a solar unit consists of its PV modules. These modules are traded on a global market, in which France represents a negligible share of total demand. Therefore, a significant portion of the observed decrease in investment costs is truly exogenous.

Second, it seems reasonable to assume that a sizable fraction of endogenous learning, such as decreasing O&M costs or improving conversion efficiency, can be related to the total amount of electricity generated to date. Because our optimization keeps constant the trajectory of annual aggregate generation of solar units, learning along these dimensions will be identical under both the realized and optimized investment trajectories, at least to the extent that learning spills over different installations in a similar way. The extreme polar assumption would be to assume that learning is category-specific and does not spill across categories of solar installations. We discuss this scenario in the next paragraph.

Finally, a residual share of learning is likely to relate to construction costs, and therefore to scale with the total installed capacity of solar units rather than their aggregated output to date. Because we do not treat installed capacity as an explicit variable, our framework cannot account for such learning in a satisfactory manner. However, installed capacity and total generation are highly correlated, so that the discussion of the previous paragraph also applies, at least to some extent, to this type of learning-by-doing.

6.3 Cross-sectional misallocation

To study the implications of having zero learning spillovers across categories, we consider an extreme scenario where learning-by-doing is fully endogenous and category-specific. For the assumed cost trajectories to be valid in both the realized and optimized investment decisions, we therefore constrain the annual aggregate output of each category of installations to remain equal its realized value.

We believe this approach represents a lower bound estimate of dynamic misallocation under endogenous category-specific learning-by-doing. In particular, we no longer allow the

optimization to arbitrage differential learning rates across categories. Instead, inefficiencies only occur within each category, for example due to location-specific capacity factors and grid connection costs.

Under this scenario, we find that the observed sequence of investments nonetheless entails a misallocation of 6%. This value can also be interpreted as the part of the misallocation that comes exclusively from cross-sectional inefficiencies. In that regard, its order of magnitude is similar to comparable estimates found in the literature (Lamp and Samano, 2023; Colas and Saulnier, 2023).

The maps in Figures A.6 to A.9 in Appendix A show the (capacity-weighted) average difference in commissioning dates broken down by category. Ground-mounted facilities exhibit the largest variation in the changes of commissioning dates. The optimization indeed leverages the high cross-sectional heterogeneity in LCOEs for this category. Besides arbitraging differences in solar irradiation conditions (e.g. postponing the commissioning of facilities in the Northern and Eastern parts of France), changes are also driven by idiosyncratic inefficiencies, possibly tied to the relative quality of the sites on which facilities are developed. For instance, the Southwestern part of France has good solar irradiation conditions, but sees its installed capacity being postponed by 3 years on average relative to the observed commissioning dates. This might stem from a significant number of sites that have less sun exposure or are more distant from the electricity grid. In contrast, the changes in the sequence of investments in smaller units seem to be mostly driven by local solar irradiation conditions.

6.4 Imperfect information

Our framework assumes that the social planner can perfectly forecast all relevant information, including future costs of solar units. This is of course a strong assumption since most forecasts back in 2005 did not correctly anticipate the significant decrease in the costs of PV installations. One may therefore wonder whether small mistakes in forecasting future costs could rationalize the observed trajectory of investments.

We focus attention on residential PV, which we found to be associated with the bulk of the

assessed misallocation. Specifically, although smaller installations were in 2005 significantly more expensive than larger ones (on a per-MWh basis), a social planner might still have preferred to start by installing residential PV first under some beliefs about the future evolution of costs. Indeed, if larger installations were expected to benefit from much larger technological improvements relative to residential PV, it could have been rational to first install residential PV in order to wait for these improvements to materialize.

To explore this possibility, we assume that the social planner had correct beliefs about the future costs of all installation categories but residential PV. This assumption is conservative in the sense that these other categories have experienced a very strong decline in costs (about -14%/year on average), significantly larger than many forecasts from the mid-2000s. We then ask: what should have been the beliefs of the social planner regarding the future costs of residential PV in order to choose to commission these installations at the beginning of our study period?

Concretely, starting from assumed costs for 2005, we hypothesize different rates for the decrease in residential PV costs and look at when residential PV is installed in the optimized trajectory. We find that the social planner needs to believe that the cost of residential PV will *increase* by at least 3.5%/year in order to install it at the beginning of our study period. In other words, imperfect forecasts cannot rationalize the observed magnitude of dynamic misallocation. Even if the social planner was anticipating no technological progress in residential PV (and very significant technological progress for other categories), he would have installed residential PV towards the end of our study period.

6.5 Limitations

Our results suggest that most of the estimated misallocation stems from having invested early on in large amounts of residential PV. These installations have indeed been deployed at a time when small-scale rooftop PV was much more expensive than medium and large-scale units. In addition, their costs have dramatically decreased in the following years, even more so than the costs of other categories of installations (in absolute terms). Consistently, early years with high installation rates of residential PV (2005-2012) are associated to the

bulk of the assessed misallocation. Although France is not the only country to have massively deployed residential PV early on,[23] our results can of course lack external validity in other countries. In addition, a number of other caveats apply.

First, a fraction of inefficient investments may correspond to a somewhat necessary "trial and error" process when betting on a non-mature technology. In particular, our optimization assumes a perfect foresight of future cost reductions, which is of course unrealistic. Back in 2005, the forecasts for PV investment costs did not anticipate their incoming steep decrease. Although it cannot completely rationalize our results (see above), the significant gap between initial forecasts and realized costs may rationalize some amount of misallocation.

Second, we assumed learning-by-doing to be either exogenous or driven by cumulative output. Some learning-by-doing is however likely to depend on the cumulative stock of installations (e.g. decrease in balance-of-system costs). It is then an open question whether such learning is specific to each category (e.g. ground-mounted vs residential), or whether some learning spills over categories. In the latter case, the costly early residential PV installations may have actually contributed to trigger learning-by-doing effects benefiting all technologies.[24] Further research is however needed to investigate whether subsidies directed to small projects have generated significant learning effects for the overall sector. It is indeed unclear why small residential projects would generate learning that may not be obtained through pilot projects of medium or large-scale installations.

Third, the obtained misallocation may arise because the social planner was pursuing additional policy objectives beyond mere efficiency. If so, our estimate for the cost of misallocation should be interpreted as the realized opportunity cost to fulfill these other policy goals. For example, small installations may generate economic benefits that are more spread out over space (jobs, taxes, etc.) and reduce land use conflicts. Investigating the nature of these other policy goals and whether they were indeed achieved is a promising area for further research.

[23] For example, De Groote and Verboven (2019) and De Groote et al. (2022) describe a similar if not more extreme adoption pattern in Belgium.

[24] For example, the first PV developers on the French market might not have been able to implement large-scale solar power plants if they did not have prior knowledge of how to commission smaller PV installations.

7 Conclusion

This paper proposes a methodology to quantify dynamic misallocation in the deployment of solar energy. We study a dynamic optimization problem where a social planner seeks to minimize the present value cost of investing in solar power plants, under the constraint to produce a given amount of solar electricity each year. In order to derive an arguably conservative estimate, our approach purposely freezes a number of possible channels of misallocation which are particularly prone to measurement errors.

We apply our methodology to the case of France for the period 2005-2021. Our results suggest that the present value cost of investments in solar energy could have been almost 30% lower than the realized present value cost without any change in the aggregate annual production of solar energy. We then explore the mechanisms that may explain this misallocation. The early large-scale deployment of residential PV seems to be associated with the bulk of misallocation costs. This observation is consistent with the high level of the early feed-in-tariffs rates for small installations. The later introduction of auctions is in contrast associated to a significant decrease in the magnitude of misallocation.

Overall, this work shows that mechanisms supporting renewable electricity generation can be very far away from their cost-efficiency frontier. In a context where an energy transition of unprecedented speed is called for in order to meet climate objectives, improving the cost-efficiency of public spending can therefore represent a low-hanging fruit to speed up the deployment of renewables.

References

Abrell, Jan and Mirjam Kosch, "The impact of carbon prices on renewable energy support," *Journal of the Association of Environmental and Resource Economists*, 2022, *9* (3), 531–563.

_ , _ , and Sebastian Rausch, "Carbon abatement with renewables: Evaluating wind and solar subsidies in Germany and Spain," *Journal of Public Economics*, 2019, *169*, 172–202.

_ , **Sebastian Rausch, and Clemens Streitberger**, "The economics of renewable energy support," *Journal of Public Economics*, 2019, *176*, 94–117.

Allouhi, Amine, Shafiqur Rehman, Mahmut Sami Buker, and Zafar Said, "Up-to-date literature review on Solar PV systems: Technology progress, market status and R&D," *Journal of Cleaner Production*, 2022, p. 132339.

Ambec, Stefan and Claude Crampes, "Decarbonizing electricity generation with intermittent sources of energy," *Journal of the Association of Environmental and Resource Economists*, 2019, *6* (6), 1105–1134.

Asker, John, Allan Collard-Wexler, and Jan De Loecker, "(Mis)allocation, market power, and global oil extraction," *American Economic Review*, 2019, *109* (4), 1568–1615.

Astier, Nicolas, Ram Rajagopal, and Frank A Wolak, "Can Distributed Intermittent Renewable Generation Reduce Future Grid Investments? Evidence from France," *Journal of the European Economic Association*, 2023, *21* (1), 367–412.

Benthem, Arthur Van, Kenneth Gillingham, and James Sweeney, "Learning-by-doing and the optimal solar policy in California," *The Energy Journal*, 2008, *29* (3).

Bollinger, Bryan and Kenneth Gillingham, "Learning-by-doing in solar photovoltaic installations," *Available at SSRN 2342406*, 2019.

Borenstein, Severin, "The private and public economics of renewable electricity generation," *Journal of Economic Perspectives*, 2012, *26* (1), 67–92.

_ , "Private net benefits of residential solar PV: The role of electricity tariffs, tax incentives, and rebates," *Journal of the Association of Environmental and Resource Economists*, 2017, *4* (S1), S85–S122.

Callaway, Duncan S, Meredith Fowlie, and Gavin McCormick, "Location, location, location: The variable value of renewable energy and demand-side efficiency resources," *Journal of the Association of Environmental and Resource Economists*, 2018, *5* (1), 39–75.

Colas, Mark and Emmett Saulnier, "Optimal Subsidies for Residential Solar," *CESifo Working Paper*, 2023.

CRE, "Coûts et rentabilité des énergies renouvelables en France métropolitaine," Technical Report, Commission de Regulation de l'Energie, Paris 2014.

_ , "Coûts et rentabilités du grand photovoltaïque en métropole continentale," Technical Report, Commission de Regulation de l'Energie, Paris 2019.

Cullen, Joseph, "Measuring the environmental benefits of wind-generated electricity," *American Economic Journal: Economic Policy*, 2013, *5* (4), 107–133.

Enedis, "Facturation des ouvrages propres des raccordements au Réseau Public de Distribution d'électricité concédé à Enedis des installations de production d'électricité relevant d'un Schéma Régional de Raccordement au Réseau des Energies Renouvelables ou d'un volet géographique," Technical Report, Enedis Direction Technique 2021.

Feger, Fabian, Nicola Pavanini, and Doina Radulescu, "Welfare and redistribution in residential electricity markets with solar power," 2017.

Fell, Harrison and Joshua Linn, "Renewable electricity policies, heterogeneity, and cost effectiveness," *Journal of Environmental Economics and Management*, 2013, *66* (3), 688–707.

France Territoire Solaire, "Observatoire de l'énergie solaire photovoltaïque en France - 41st issue," Technical Report 2022.

_ , "Observatoire de l'énergie solaire photovoltaïque en France - 44th issue," Technical Report 2023.

Gillingham, Kenneth and Marten Ovaere, "The heterogeneous value of solar and wind energy: Empirical evidence from the United States and Europe," Technical Report, Working paper 2020.

Groote, Olivier De and Frank Verboven, "Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems," *American Economic Review*, 2019, *109* (6), 2137–72.

_ , Axel Gautier, and Frank Verboven, "The Political Economy of Financing Climate Policy? Evidence from the Solar PV Subsidy Programs," 2022.

IRENA, "Renewable Power Generation Costs in 2014," Technical Report, International Renewable Energy Agency, Abu Dhabi 2014.

_ , "Renewable Power Generation Costs in 2020," Technical Report, International Renewable Energy Agency, Abu Dhabi 2020.

_ **and CPI**, "Global Landscape of Renewable Energy Finance 2020," Technical Report, International Renewable Energy Agency and Climate Policy initiative 2020.

Lamp, Stefan and Mario Samano, "(Mis)allocation of Renewable Energy Sources," *Journal of the Association of Environmental and Resource Economists*, 2023, *10* (1), 195–229.

Liski, Matti and Iivo Vehviläinen, "Gone with the wind? An empirical analysis of the equilibrium impact of renewable energy," *Journal of the Association of Environmental and Resource Economists*, 2020, *7* (5), 873–900.

Novan, Kevin, "Valuing the wind: renewable energy policies and air pollution avoided," *American Economic Journal: Economic Policy*, 2015, *7* (3), 291–326.

Pfenninger, Stefan and Iain Staffell, "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data," *Energy*, 2016, *114*, 1251–1265.

Pörtner, Hans-Otto, Debra C Roberts, H Adams, C Adler, P Aldunce, E Ali, R Ara Begum, R Betts, R Bezner Kerr, R Biesbroek et al., "Climate change 2022: Impacts, adaptation and vulnerability," *IPCC Sixth Assessment Report*, 2022.

Reguant, Mar, "The efficiency and sectoral distributional impacts of large-scale renewable energy policies," *Journal of the Association of Environmental and Resource Economists*, 2019, *6* (S1), S129–S168.

Sexton, Steven, A Justin Kirkpatrick, Robert I Harris, and Nicholas Z Muller, "Heterogeneous solar capacity benefits, appropriability, and the costs of suboptimal siting," *Journal of the Association of Environmental and Resource Economists*, 2021, *8* (6), 1209–1244.

Staffell, Iain and Stefan Pfenninger, "Using bias-corrected reanalysis to simulate current and future wind power output," *Energy*, 2016, *114*, 1224–1239.

Wolak, Frank A, "Level versus Variability Trade-offs in Wind and Solar Generation Investments: The Case of California," *The Energy Journal*, 2016, *37* (Bollino-Madlener Special Issue).

A Supplementary Figures

Figure A.6: Ground-mounted facilities. Panel A: density of installations per sub-region as of 2021 (kW/km^2). Panel B: capacity weighted average changes in commissioning dates obtained from the optimization (year observed - year optimized).

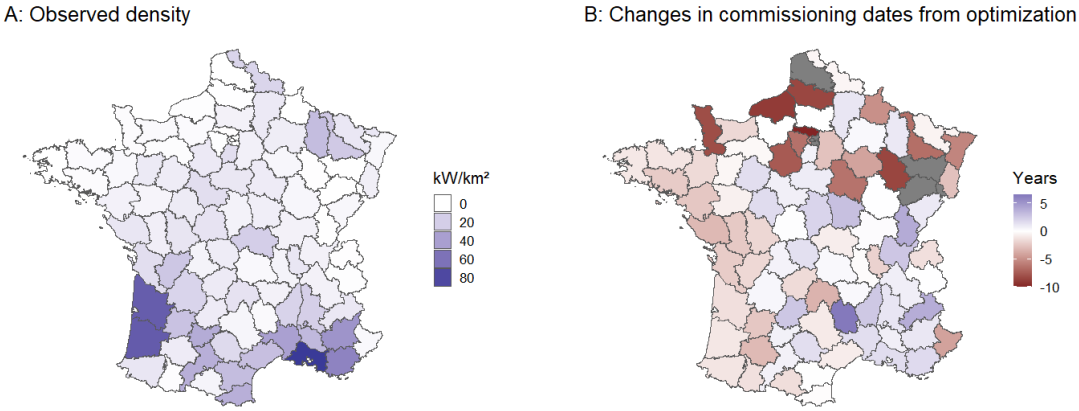


Figure A.7: Large rooftop facilities (greater than 100 kW). Panel A: density of installations per sub-region as of 2021 (kW/km^2). Panel B: capacity weighted average changes in commissioning dates obtained from the optimization (year observed - year optimized).

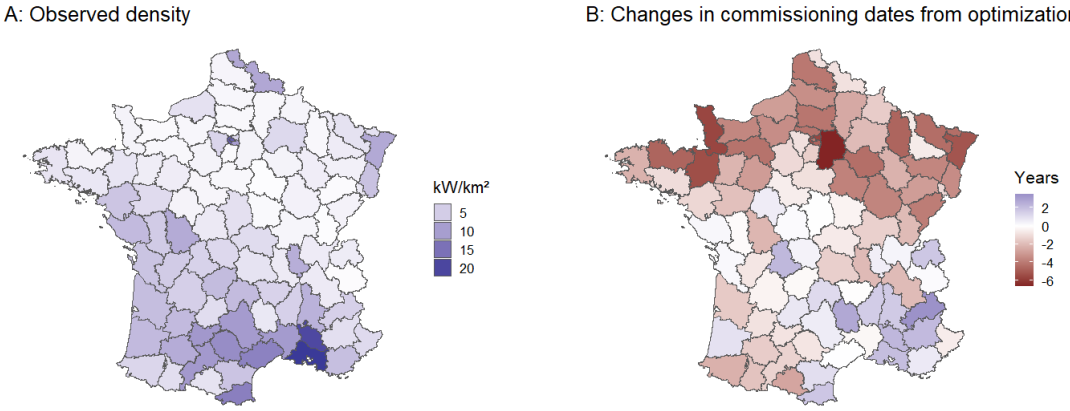


Figure A.8: Rooftop installations between 36 and 100 kW. Panel A: density of installations per sub-region as of 2021 (kW/km^2). Panel B: capacity weighted average changes in commissioning dates obtained from the optimization (year observed - year optimized).

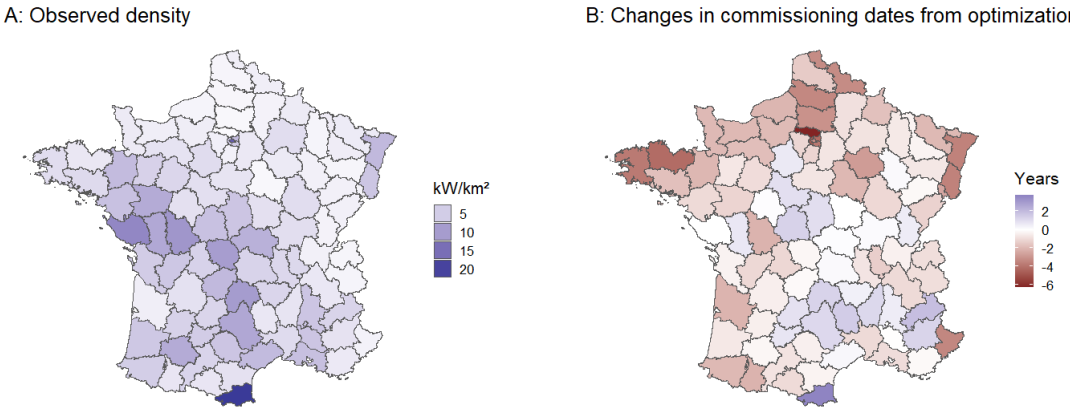
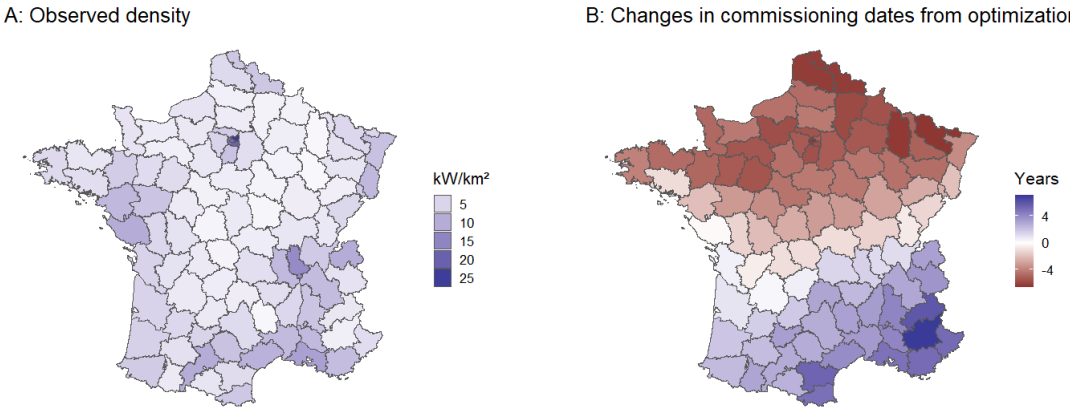


Figure A.9: Rooftop installations lower than 36 kW. Panel A: density of installations per sub-region as of 2021 (kW/km^2). Panel B: capacity weighted average changes in commissioning dates obtained from the optimization (year observed - year optimized).



B Computing solar units' costs and energy output

This Appendix details how we compute the cost and yearly energy output of each solar unit in the inventory, as a function of its commissioning date. Step 1 uses several data sources to calibrate investment cost functions for different categories of projects. Step 2 computes connection costs (for ground-mounted installations) based on the distance between solar facilities and the grid sub-stations to which they connect. Step 3 estimates the expected annual energy output as of 2022 for PV units that do not report their realized production in the inventory. Step 4 calibrates depletion and learning-by-doing factors.

B.1 Calibrating investment costs functions

We use two main data sources to calibrate different cost functions for different types of projects. First, we rely on two reports by the French Energy Regulation Commission (CRE) that provide a detailed analysis of the costs of large PV projects (greater than 100 kW) in France. The first report covers the years 2011-2015 (CRE, 2014), and the second one the years 2017-2019 (CRE, 2019). Second, for module costs and small PV projects (lower than 100 kW), we rely on reports by IRENA (IRENA, 2014, 2020).

We define different cost functions for 8 categories of projects. These categories are defined by the combination of a type of installation (ground-mounted or rooftop) and a size bucket (in MW):

- Ground-mounted >10 MW
- Ground-mounted 2.5-10 MW
- Ground-mounted 0.5-2.5 MW
- Rooftop >2.5 MW
- Rooftop 0.5-2.5 MW
- Rooftop 0.1-0.5 MW
- Rooftop 0.036 - 0.1 MW

- Rooftop <0.036 MW

We calibrate cost functions for project categories above 100 kW by extrapolating data from CRE, complemented with IRENA's data on PV modules' costs. For project categories under 100 kW, we rely on IRENA's reports only. All costs are computed in real 2019 euros.

B.1.1 Cost functions for project categories above 100 kW

To compute cost functions for large projects (>100 kW), we mainly rely on CRE's reports. CRE makes a distinction between rooftop and ground-mounted projects, and uses three size buckets that correspond to our categories. We compute total costs by summing two components: (1) PV modules' costs, which are taken from IRENA (2020) and are common to all project categories; (2) other investment costs, which are extrapolated (separately for each project category) from CRE (2014) and CRE (2019). These latter costs include construction, transformers, wires, engineering and procurement as well as the present value of future O&M costs.[25] Figure B.10 shows investment costs other than PV modules and grid connection costs as a function of the commissioning date. Figure B.11 shows the assumed evolution for the cost of PV modules.

Because we only have a few data points for the costs other than modules' cost, we extrapolate cost functions over the whole period (2005-2021) for each project category. Specifically, if we denote with $f(i,t)$ is the cost per kW of commissioning a solar unit of category i in year t, we make the following parametric assumptions:

$$f(i,t) = w_t + m_{i,t} \tag{2}$$

$$log(m_{i,t}) = a_i + b_i t \tag{3}$$

where w_t is the cost (per kW) of PV modules cost in year t (common to all categories) and $m_{i,t}$ captures all other costs, which are category-specific. We use OLS regressions to calibrate $m_{i,t}$ from cost observations in 2011, 2015, 2018, 2019 and 2020. The obtained coefficients

[25] We account for OPEX and rental costs, which are given per project category. Taxes are not included in the variable costs. Present values are computed with a discount rate of 4.5% and a project lifetime of 20 years.

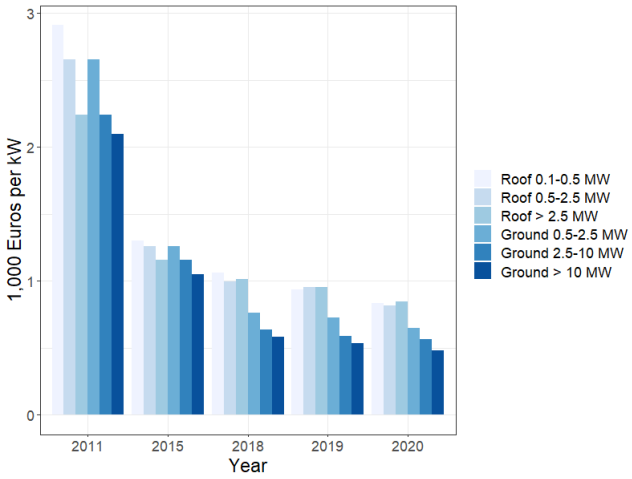


Figure B.10: Assumed investment costs (real euros 2019) other than PV modules and grid connection costs, by project category and commissioning date.

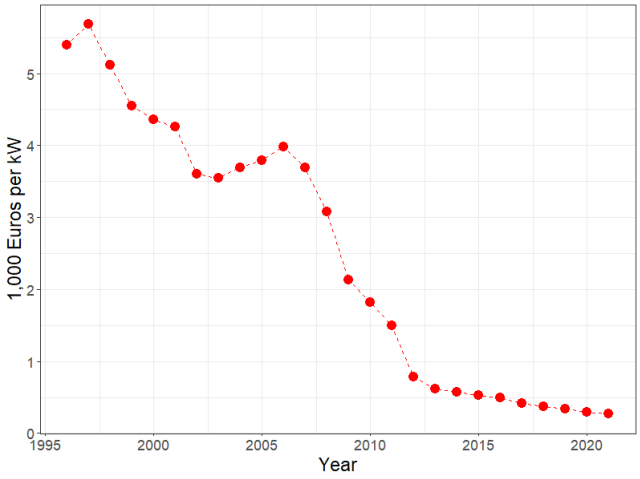


Figure B.11: Evolution of the cost of PV modules (real euros 2019)

(a_i, b_i) are reported in Table 6.

B.1.2 Cost functions for project categories below 100 kW

Cost functions for small projects ($<$ 100 kW) are calibrated using IRENA (2020), which reports time series of total installation costs for commercial and residential PV in France. Since these time series only span from 2010 to 2020, we apply the same extrapolation as

Table 6: Calibrated parameters for project categories larger than 0.1 MW

Type	Size (MW)	Coeff. a	Coeff. b
Rooftop	0.1 - 0.5	276	-0.13
Rooftop	0.5 - 2.5	259	-0.12
Rooftop	>2.5	212	-0.1
Ground-mounted	0.5 - 2.5	341	-0.17
Ground-mounted	2.5 - 10	347	-0.17
Ground-mounted	>10	360	-0.17

above outside of this time window. Coefficients obtained from the regression are reported in Table 7. IRENA (2020) only reports total installation cost, not O&M costs. We assume that the present value of O&M costs represents a fixed percentage of installation costs, equal to the average of the observed percentages for larger rooftop categories (whose range is 10-20%).

Table 7: Calibrated parameters for project categories smaller than 0.1 MW

Type	Size (MW)	Coeff. a	Coeff. b
Rooftop	< 0.036	403	-0.2
Rooftop	0.036 - 0.1	284	-0.14

B.1.3 Obtained cost functions

Figure B.12 shows the assumed cost function for each category, as well as the observed data points. Note that cost functions for the smallest categories (lower than 0.1 MW) are always higher than observed data points since the former include O&M costs while the latter do not. It is worth noting that only a negligible total capacity was installed before 2010, consisting mostly of small rooftop projects.

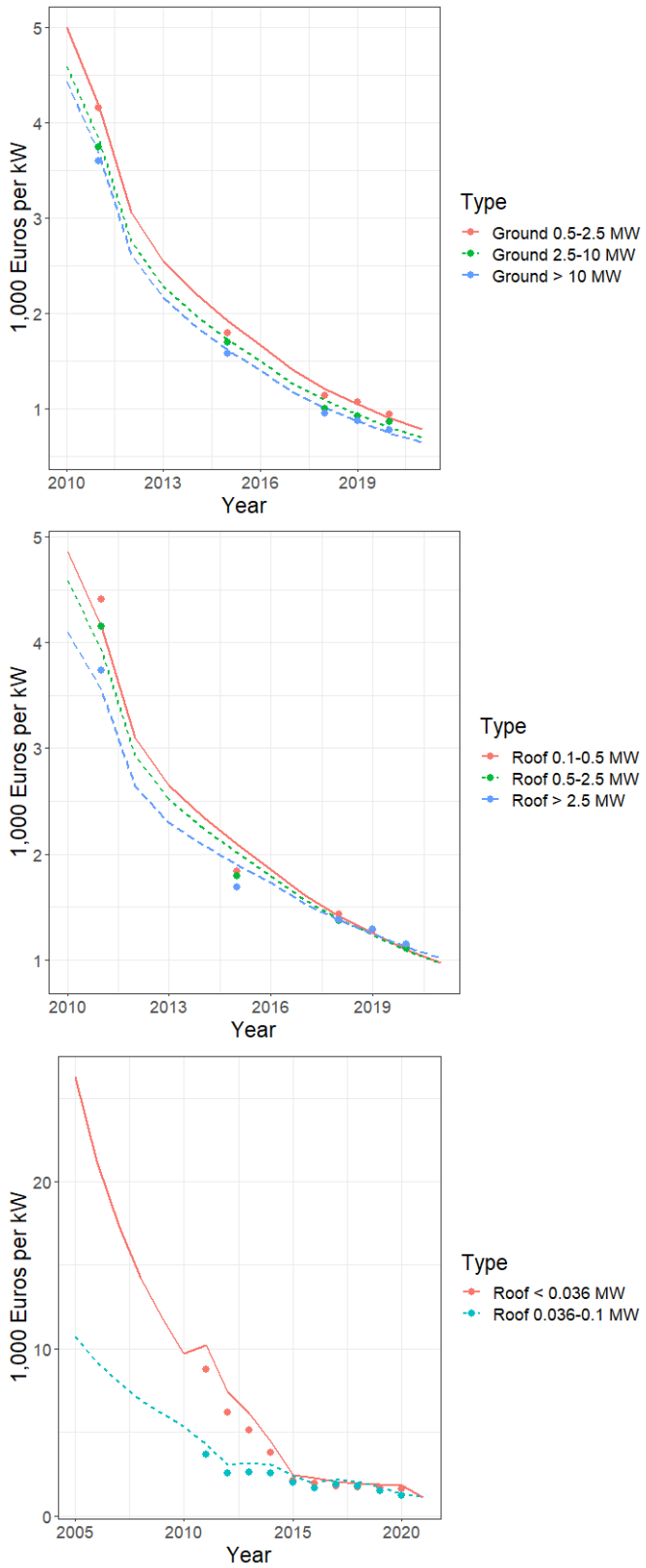


Figure B.12: Assumed cost functions vs observed data points.

B.2 Grid connection costs

Grid connection costs can represent a sizable fraction of total installation costs (up to 20% of investment costs). This paragraph describes how we estimate connection costs.

Solar installations can connect to the grid at different voltage levels. If they connect to the distribution grid, they may either connect to the low voltage level (BT) or to the medium voltage level (HTA). The size of PV installations connecting to the low voltage (resp. medium voltage) grid is capped to 250 kW (resp. 17 MW). Larger units have to connect to the grid at higher voltage levels (HTB).

Solar units connecting to the low voltage level (BT) are assumed to have zero connection costs. Indeed, these units are rooftop installations and rarely require any significant expansion or reinforcement of the power network.

Solar units connecting to either the HTA or HTB level are assumed to incur grid connection costs. These costs most often consist of building/reinforcing power lines connecting the unit to the upstream substation. Therefore, we compute connection costs as the product of the as-the-crow-flies distance between the PV installation and the substation to which it connects and a fixed connection cost per meter: 100 €/m in HTA and 1000 €/m in HTB (Enedis, 2021). The location of solar facilities is only reported at the county level. We assume they sit at the centroid of their county. GPS coordinates of substations are obtained from the TSO and DSO open data portals. For all installations but 4%, the public inventory of power plants indicates to which upstream substation each solar unit connects.[26] Finally, we correct outlier observations[27] by drawing a random distance from a log-normal distribution calibrated on the rest of the dataset.[28] Table 8 reports summary statistics for the obtained distribution of connection distances, broken down by voltage level.

[26] Installations for which this information is missing are matched to the closest substation.

[27] Our procedure only yields 1% of outliers, defined as observations whose distance of connection is either zero or larger than 50 km. These outliers likely stem from missing substations or matching errors between sub-stations and solar units.

[28] Using a log-normal distribution is motivated by the fact that distances must be non-negative, as well as by the shape of the distribution of connection distances (Figure B.13).

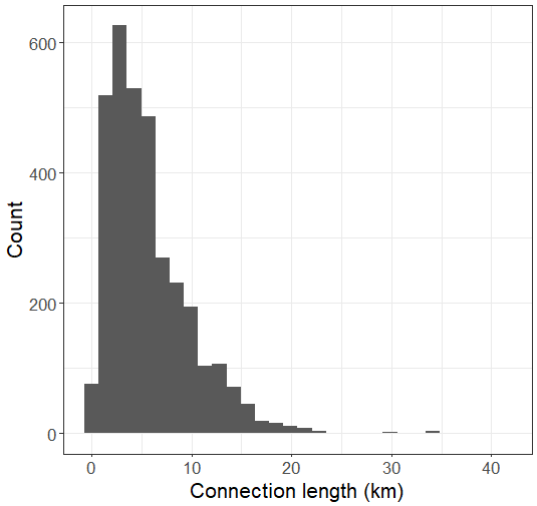


Figure B.13: Histogram of connection distances for units connecting to HTA and HTB voltage levels.

Table 8: Distribution of obtained grid connection distances (in km).

Grid Voltage	N Obs.	Min	1st Q	Median	3rd Q	Max
HTA	2,439	0.05	2.8	4.8	8.0	41
HTB	90	0.6	1.5	2.2	6.4	6.4

B.3 Yearly energy output

For most observations (∼93%) in the registry of power plants, we observe the total output they produced during the previous 12 months. As such, we observe the energy produced by each unit between December 2021 and December 2022.

For the remaining observations (7%), we first retrieve simulated capacity factors at different locations and in different years from the website renewable ninja (Pfenninger and Staffell, 2016; Staffell and Pfenninger, 2016). We sample 2,214 locations in mainland France (corresponding to the locations of grid substations) for 2010-2020 (i.e., 11 years corresponding to different weather conditions).[29] We average capacity factors over these years to obtain a single expected capacity factor per location. Figure B.14 maps these capacity factors (averaged over larger geographical units for more clarity). Finally, we match each solar installation for

[29]The API requires to set a number of parameters about the characteristics of the considered solar installation. We set these parameters to the same value for all requests: a tilt angle of 28 degrees, an azimuth angle of 178 degrees, system losses at 10%, and no tracking technology.

which no output is observed in the public registry to the closest location we sampled.

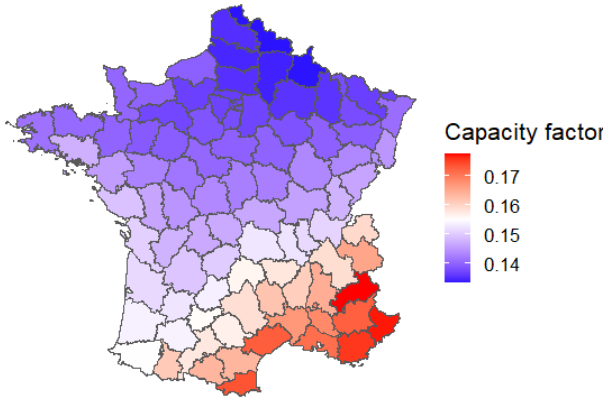


Figure B.14: Simulated capacity factors averaged over 2010-2020 at the sub-region level.

Simulated capacity factors retrieved from renewable ninja are theoretical benchmarks that do not account for outages and depletion. Table 9 compares the distribution of simulated capacity factors to the distribution of capacity factors observed in the public registry of power plants. We observe that simulated capacity factors are on average higher than observed ones. However, installations commissioned after 2018 have comparable capacity factors to simulated values. This suggests that the conversion efficiency of solar units decreases over time, a phenomenon that we account for in the next paragraph.

Table 9: Observed vs simulated capacity factors

Sample	N Obs.	Min	1st Q	Median	3rd Q	Max
Inventory	37,542	0	0.12	0.14	0.16	1.6
Inventory (after 2018)	14,454	0	0.13	0.15	0.16	1.6
Simulated	3,356	0.13	0.14	0.15	0.16	0.19

Table 9 also reveals that output data reported in the public registry is prone to mistakes. In addition, observed capacity factors can mis-represent the average performance of the unit (e.g. due to long outages). We account for outliers by setting a minimum (resp. maximum)

threshold over (resp. under) which the observed capacity factor is deemed to be inaccurate. We set the minimum threshold to 0.03 and the maximum threshold to 0.3.[30] All units reporting capacity factors above or under these thresholds are assigned energy outputs as of 2021 computed from simulated capacity factors. The obtained final distribution of capacity factors is reported in Figure B.15.

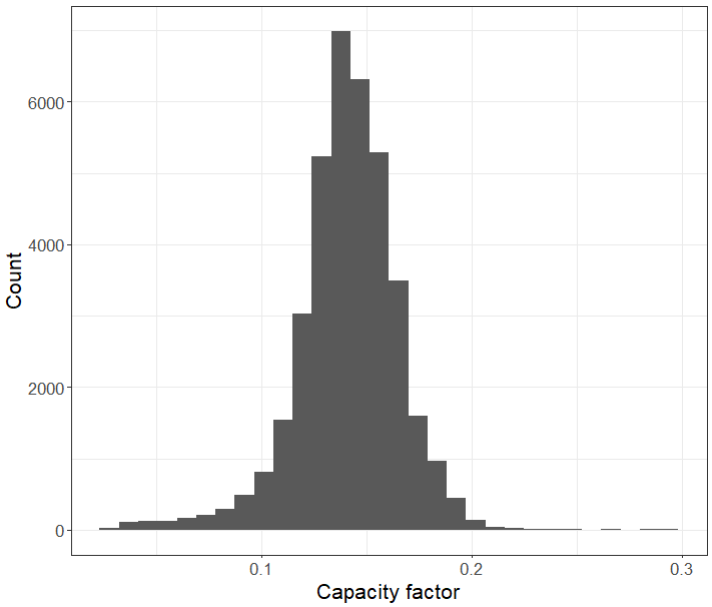


Figure B.15: Histogram of PV project capacity factors after correcting for outliers.

B.4 Learning-by-doing and depletion rates

The annual output in given year of a solar plant installed at a given location depends on its commissioning date for at least two reasons. First, conversion efficiency decreases over time due to wear and tear. Second, due to improvements in solar PV cell technologies, units commissioned later in the period tend to have higher conversion efficiencies. To account for these two effects, we assume that if a given solar unit i commissioned in year t_c would have instead been commissioned in year t, its annual output in year t' would change as follows:

[30] This maximum threshold corresponds to the highest capacity factor achievable for a plant equipped with 2-axis tracking, suffering no system loss, and installed in the most irradiated location.

$$e_{itt'} = e_{it_c}(1+\alpha)^{t-t_c}(1+\beta)^{t'-t} \tag{4}$$

Where α is the annual rate of technological improvement for solar modules and β is the annual rate of depletion. The variable e_{i,t_c} is the annual energy output following the initial year of commissioning of the plant, denoted by t_c.

We estimate depletion and technological improvement rates directly from our data. Using all public inventories of solar plants published between 2017 and 2022, we build a panel dataset consisting of the annual energy output of 16,000+ solar units for these 6 consecutive years. After applying a log-transformation and a first order approximation to equation (4), we seek to estimate the following equation:

$$ln(Y_{itt'}) = ln(e_{it_c}) + (t-t_c)\alpha + (t'-t)\beta + \epsilon_{it'} \tag{5}$$

Where $ln(Y_{itt'})$ measures the natural logarithm of the capacity factors in year t' of unit i, which was commissioned in year t. The parameters of interest are α, which measures the annual rate of increase in the output due to technological learning, and β, which captures the annual depletion rate in output.

Because we do not observe e_{it_c}, we instead estimate:

$$ln(Y_{itt'}) = c + (t-t_c)\alpha + (t'-t)\beta + ln(\Gamma_i^{TH}) + \lambda_{t'} + \epsilon_{it'} \tag{6}$$

where Γ_i^{TH} is the simulated capacity factor of unit i (see above) and $\lambda_{t'}$ are year fixed effects. Table 10 reports the obtained results. We estimate a depletion rate of -1% per year and a technological learning rate of 1% per year. Our estimates are in line with other values found in the literature. For example, De Groote and Verboven (2019) assume a yearly depletion rate of 1%, Borenstein (2017) a rate of 0.5% and Feger et al. (2017) set the depletion rate to 3% for the first year and to 0.7%/year for later years. Regarding technological learning, the literature review by Allouhi et al. (2022) suggests a rate of 0.5% per year.

Table 10: Regression results

	Capacity factor (logarithm)		
	(1)	(2)	(3)
$t - 2005$	0.011***	0.011***	0.008***
	(0.001)	(0.001)	(0.001)
$t' - t$	-0.011***	-0.012***	-0.012***
	(0.0004)	(0.0005)	(0.0005)
Γ^{TH}			0.645***
			(0.009)
Constant	-1.986***	-1.988***	-0.753***
	(0.006)	(0.007)	(0.018)
Fixed effects (Year)	No	Yes	Yes
Observations	97,554	97,554	92,412
R^2	0.048	0.051	0.106

*p<0.1; **p<0.05; ***p<0.01

Note: This table displays the results of the regression of the logarithm of capacity factors on differences in years of installation and in years of production for three specifications: column (1) does not introduce additional controls, column (2) controls for year of production fixed effects, column (3) controls for year of production fixed effects and simulated capacity factors matched from the closest location, denoted by FC^{TH}.

C Details on the public inventory of installations

This Appendix details the two steps implemented to build our final inventory of PV installations. First, we retrieve the installed capacities of small installations (<36 kW) over time by combining the list of larger installations (>36 kW) and the annual time-series of PV installations aggregated at the department level. Second, large units (>500 kW) are labeled to be either ground-mounted or rooftop projects.

C.1 Building residential PV time-series at the departement level

For most small PV units (< 36 kW), that is, residential PV, the public inventory of power plants only reports a cross-sectional view of their installed capacity. We thus use other data sources to construct annual time series for the evolution of residential capacity between 2005 and 2021. More specifically, we compute annual time series of residential PV capacity aggregated at the "departement" level, which is sufficient for our analysis.[31]

We compute these departement level time series of residential PV capacity as follows. First, for 2005 to 2016, we use a dataset from the French Department of Energy (DOE) that provides yearly panel data of total installed capacities at the departement level.[32] Because we observe their commissioning dates in the public inventory, we first build departement-level time series of non-residential installations. We then subtract these time series to the time series of total installed capacity, which leaves us with residential PV installations. Second, from 2017 onwards, we use the total capacities of residential PV, as directly reported in the inventory of the corresponding year.[33] Figures C.16 and C.17 show the obtained results for each departement. Overall, the two sources agree very well as we do not observe significant discrepancies in values between the years 2016 and 2017. As shown in Figures C.16 and C.17, we constrain time series to be weakly increasing.

[31] Given our methodology, coarser levels of spatial aggregation actually works against finding large amounts of misallocation, which strengthen the conservative nature of our estimates.

[32] https://www.statistiques.developpement-durable.gouv.fr/tableau-de-bord-solaire-photovoltaique-quatrieme-trimestre-2021. Our data consist of the publications released in the last quarter of each year.

[33] The first public inventory of power plants was released in 2017. We collect public inventories from 2017 to 2021, selecting the versions updated in December 31st of each respective year.

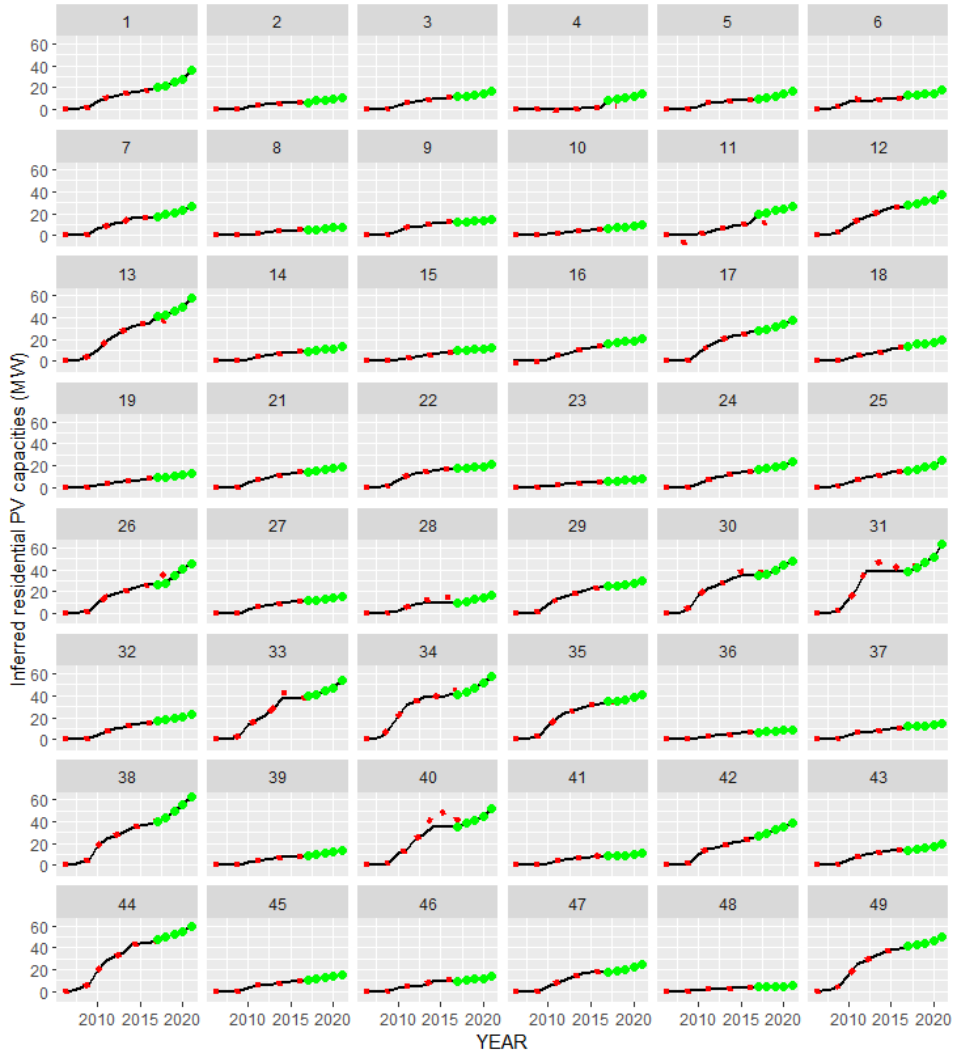


Figure C.16: Inferred departements level time series of PV capacity from aggregated units (first 49 departements). Green dots represent capacities as reported in solar plants inventories for 2017, 2018, 2019, 2020, 2021. Red dots are capacities obtained from the French Department of Energy (DOE) dataset.

Note: Green dots represent capacities as reported in solar plants inventories for 2017, 2018, 2019, 2020, 2021. Red dots are capacities deduced from the French Department of Energy (DOE) dataset.

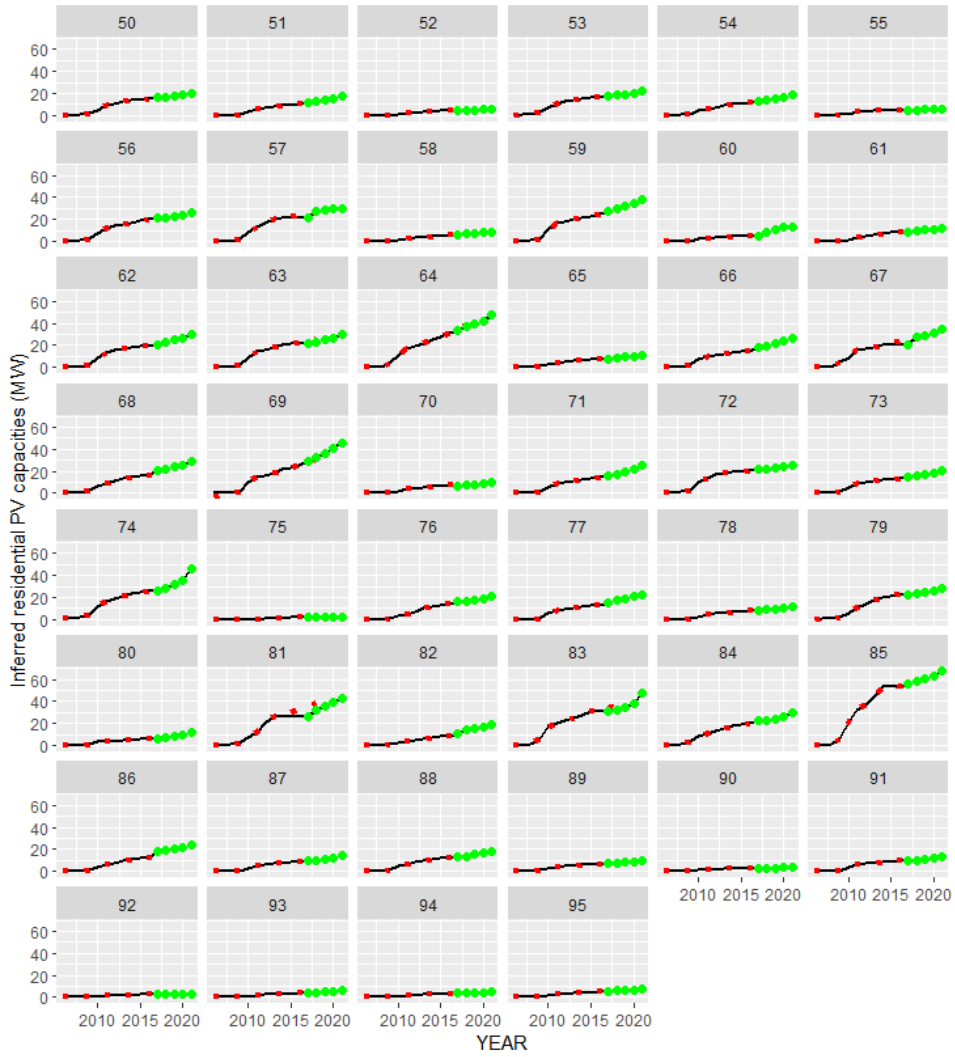


Figure C.17: Inferred departements level time series of PV capacity from aggregated units (last 45 departements).
Note: Green dots represent capacities as reported in solar plants inventories for 2017, 2018, 2019, 2020, 2021. Red dots are capacities deduced from the French Department of Energy (DOE) dataset.

C.2 Assigning technologies to large individual installations

The public inventory of power plants does not specify if a given unit is ground-mounted or rooftop. Four strategies are implemented to assign PV facilities to either rooftop or ground-mounted types, as detailed in the paragraphs below.

Dictionary of key words in installations' names

The first strategy we use to identify project types is to match installation names with a dictionary. We use the following key words:

- Words specific to rooftop installations: PARKING; PKG; OMBRIERE; TOITURE; SCI
- Words specific to buildings: SERRE; LOGISTIQUE; TECHNOPOL; LA POSTE; CENTRE COMMERCIAL; CENTRE; SAINT CHARLES; UNIVERSITE; ENTREPOT; STATION; HIPPODROME ;STADE; RESERVOIR; ARENA; OMNISPORT ;LYCEE; ETABLISSEMENT; CASERNE; HANGAR; USINE ;ZAC; SIEGE SOCIAL; BATIMENT; BAT; AEROPORT; STADE; STADIUM; CINEMA; SUPERMARCHE
- Words specific to large retailers and firms: CASINO; AUCHAN; GEANT; SANOFI; GIFI; SISLEY; IKEA; UBISOFT; LEROY MERLIN; RENAULT; LECLERC; CARREFOUR; SUPER U; SYSTEME U; HYPER U
- Words specific to ground-mounted installations: FERME SOLAIRE; CENTRALE; PARC SOLAIRE; CHAMP; AU SOL
- Known project names for ground-mounted projects: GABOTS; LAVANSOL; SOLAIREISTRES; ENFINITY; KRONOSOL; PLAINES; QUINCIEUX; TSAOS4.7; SALAUNES

This strategy allows us to assign more than 800 units to rooftop types and about 40 units to ground-mounted types.

Size thresholds

The second strategy is to define size thresholds for each type by using eligibility rules

of support mechanisms and stylized facts. From support mechanisms' rules we are able to define three thresholds:

- There are no ground-mounted PV under 500 kW. Indeed, CRE auctions are only for projects larger than 500 kW.
- Solar energy auctions before 2016 set a maximum size for rooftop projects at 4.5 MW. We therefore assume that all units above 4.5 MW and installed between 2012 and 2017 (using one year construction lag) are ground-mounted.
- Auctions after 2016 have extended the size limit for rooftop PV to 8 MW. After 2017, only units larger than 8 MW are therefore automatically assigned to being ground-mounted.

This strategy allows us to assign 33,500 additional units to rooftop types and 400 units to ground-mounted types.

Auction winners

In the third strategy, we retrieve the list of winners from ground-mounted specific auctions and match the candidates names to the installations names in the inventory. This method only identifies about 20 additional ground-mounted installations.

OpenStreetMap facilities

In the fourth strategy, we combine the list of solar units in the inventory with the list of solar installations that are reported in OpenStreetMap (OSM). OSM is an open-source database that stores geographic objects worldwide, including PV installations. OSM reports more than 1,300 PV installations that are located in mainland France[34]. As OSM focuses on spatial objects with significant land footprints, the majority of PV installations identified in the database are ground-mounted facilities. Rooftop installations listed in OSM are explicitly associated with the specific buildings on which they are installed (e.g. factory, warehouse,

[34]obtained from OpenStreetMap's API: `https://overpass-turbo.eu/`, specifying objects with label "solar" in the "plant" category and within France geographic boundaries.

stores). After being assigned to either rooftop or ground-mounted types, the OSM dataset is matched to the public inventory of solar plants using either (i) ERC codes, a unique identifier for PV installations, or (ii) projects' installed capacity and location. This allow us to match about 400 additional units to rooftop and ground-mounted types.

Finally, we proceed in two steps to attribute project types for the 1,300 remaining units with unknown types:

- Units lower than 1 MW are assigned to rooftop types until the total installed capacity of rooftop facilities within the range of 0.25 and 1 MW that is observed in mainland France is reached (France Territoire Solaire, 2023). Remaining observations lower than 1 MW are then assigned to ground-mounted PV.

- Units larger than 1 MW are randomly assigned based on the observed distribution of project types within specific size categories. For example, among units of sizes ranging between 2 and 3 MW, there are only 10% of rooftop facilities among the observations that are already assigned to a category. We thus randomly assign 10% of the unknowns within the range of 2 and 3 MW to the rooftop type and the remaining observations to ground-mounted.

At the end of our assignment procedure, we have 1,400+ ground-mounted and 36,000+ rooftop solar units.

D Special case of static LCOEs

In this Appendix, we solve the simplified case where there is neither technological progress nor wear and tear of installations:

$$\forall i, t,\ c_{it} = c_i \text{ and } \forall i, t, t',\ e_{itt'} = e_i$$

The optimization program then simplifies to:

$$\min_{x_{it}} \sum_{i=1}^{N} c_i \left(\sum_{t=1}^{T} \rho^t x_{it} \right)$$

s.t.

$$\begin{aligned}
&\forall t \in \{1, ..., T\}, && \sum_{i=1}^{N} e_i \left(\sum_{t'=1}^{t} x_{it'} \right) \geq E_t && (\rho^t \lambda_t) \\
&\forall i \in \{1, ..., N\}, && \sum_{t=1}^{T} x_{it} \leq 1 && (\bar{\mu}_i) \\
&\forall i \in \{1, ..., N\}, \forall t \in \{1, ..., T\}, && x_{it} \geq 0 && (\rho^t \underline{\mu}_{it})
\end{aligned}$$

One can then write the Lagrangian as:

$$\begin{aligned}
\mathcal{L}(x_{it}, \lambda_t, \bar{\mu}_i, \underline{\mu}_{it}) = \quad & \sum_{i=1}^{N} c_i \left(\sum_{t=1}^{T} \rho^t x_{it} \right) \\
& + \sum_{t=1}^{T} \rho^t \lambda_t \left(E_t - \sum_{i=1}^{N} e_i \left(\sum_{t'=1}^{t} x_{it'} \right) \right) \\
& + \sum_{i=1}^{N} \bar{\mu}_i \left(\sum_{t=1}^{T} x_{it} - 1 \right) \\
& - \sum_{i=1}^{N} \sum_{t=1}^{T} \rho^t \underline{\mu}_{it} x_{it}
\end{aligned}$$

Besides the complementary slackness conditions, we get (taking derivative w.r.t. x_{it}) the following first-order conditions:

$$\text{For all } i, t: \qquad \rho^t c_i - e_i \sum_{t'=t}^{T} \rho^{t'} \lambda_{t'} + \bar{\mu}_i = \rho^t \underline{\mu}_{it}$$

On the right hand-side of the equation, the multiplier $\underline{\mu}_{it}$ is non-negative and equals 0 if, and only if, $x_{it} > 0$, that is if unit i is (at least partly) commissioned in year t. If we denote $t^*(i)$ the year at which plant i is optimally commissioned, we thus have:

$$\forall t, \quad \rho^{t^*(i)} c_i - e_i \sum_{t'=t^*(i)}^{T} \rho^{t'} \lambda_{t'} \leq \rho^t c_i - e_i \sum_{t'=t}^{T} \rho^{t'} \lambda_{t'}$$

In other words:

$$t^*(i) = \underset{t}{\operatorname{argmin}} \left[\rho^t c_i - e_i \sum_{t'=t}^{T} \rho^{t'} \lambda_{t'} \right]$$

Which may be rewritten:

$$t^*(i) = \underset{t}{\operatorname{argmin}} \left[\rho^t L_i - \Lambda_t \right] \text{ where } L_i \equiv \frac{c_i}{e_i} \text{ and } \Lambda_t \equiv \sum_{t'=t}^{T} \rho^{t'} \lambda_{t'}$$

We then have the following Lemma:

Lemma 1 *For all i, j, $L_i < L_j \Rightarrow t^*(i) \leq t^*(j)$*

Proof.

By definition of $t^*(j)$, we have:

$$\forall t, \ \rho^{t^*(j)} L_j - \Lambda_{t^*(j)} \leq \rho^t L_j - \Lambda_t$$

Then:

$$\begin{aligned} \rho^{t^*(j)} L_i - \Lambda_{t^*(j)} &= \rho^{t^*(j)} (L_i - L_j) + \rho^{t^*(j)} L_j - \Lambda_{t^*(j)} \\ &\leq \rho^{t^*(j)} (L_i - L_j) + \rho^t L_j - \Lambda_t \\ &\leq (\rho^{t^*(j)} - \rho^t)(L_i - L_j) + \rho^t L_i - \Lambda_t \end{aligned}$$

For $t > t^*(j)$, we have $(\rho^{t^*(j)} - \rho^t) > 0$. In addition, $(L_i - L_j) < 0$. As a result:

$$\forall t > t^*(j), \ \rho^{t^*(j)} L_i - \Lambda_{t^*(j)} < \rho^t L_i - \Lambda_t$$

And thus:

$$t^*(i) = \operatorname{argmin}_t \left[\rho^t L_i - \Lambda_t \right] \leq t^*(j)$$

■

Assume, without loss of generality, that we have indexed units such that:

$$L_1 \leq L_2 \leq ... \leq L_{N-1} \leq L_N$$

From Lemma 1, and denoting $i_0 \equiv 0$ and $i_T \equiv N$, there exists indexes:

$$i_0 \leq i_1 \leq ... \leq i_{T-1} \leq i_T$$

such that units commissioning in year t have an index i such that:

$$i_{t-1} \leq i \leq i_t$$

Finally, we can pinpoint the thresholds i_1, i_2, ... , i_{T-1} using the target trajectory of annual solar generation. Specifically, the threshold index i_t in year t will be chosen so that unit i_t is the last one needed to meet the energy output target:

$$\sum_{i=1}^{i_t-1} e_i \leq E_t \leq \sum_{i=1}^{i_t} e_i$$