Energy Economics 126 (2023) 107011

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco

High renewable electricity penetration: Marginal curtailment and market failure under "subsidy-free" entry

David M. Newbery[1]

Faculty of Economics, EPRG Cambridge, Sidgwick Ave, Cambridge, CB3 9DE, UK

ARTICLE INFO

JEL classification:
H23
L94
Q28
Q42
Q48

Keywords:
Renewable electricity
Marginal wind curtailment
Integration costs
market failures
Inertia chargesbreak

ABSTRACT

Ambitious plans to decarbonize electricity will require high levels of variable renewable electricity (VRE). At high VRE penetration, the surplus that cannot be exported must be curtailed (spilled). The last MW of wind capacity will be curtailed 3+times more hours than the average, but even in efficiently designed markets, price signals for VRE investment are given by average, not marginal, curtailment, creating a "tragedy of the commons" that requires a corrective charge to restore efficiency. The paper sets out an analytical model calibrated to Ireland in 2026, showing the source of this distortion and estimates of its magnitude.

1. Introduction

Ambitious plans to decarbonize electricity will require very high levels of variable renewable electricity (VRE) generation, specifically from on- and off-shore wind and solar PV. Fortunately, the cost of VRE is now competitive with conventional generation (at least with a suitable carbon price) even under normal market conditions — and the 2022 energy crisis has dramatically emphasized this cost advantage. This holds out the attractive prospect of "subsidy-free" VRE entry, although there are sound market design principles for providing suitable long-term contracts to reduce risk and hence lower the cost of finance. Such contracts may well be cheaper than the expected future wholesale prices but even so attractive to most developers. Other VRE may still choose to enter on a merchant basis (Gohdes et al., 2022; Flottmann et al., 2022). The average capacity factor for onshore wind between 2013–17 was below 30% in 16 of the 18 countries surveyed by Gönül et al. (in press), with the UK at 25% and the world average 23%. Offshore wind has a higher capacity factor, with the world-wide average from 2010–21 at 40% (Fernández, 2023), also the UK average from 2017–22. The UK average of all wind from 2017–2021 was 32%, with 2009–21 average for on-shore wind 26.4% (DUKES, 2022, table 6.3). UK solar PV averaged 10.8% (2014–21), while the global average has been rising as PV is increasingly located in lower latitudes, rising to about 17.5% (Fernández, 2023). In the US the range is from 28% in the SW states to 16% in the NE (EIA data). Northern Europe (above 50^0 N) averages 15% or less (Wu et al., 2022).

The share of VRE in total annual generation will be determined by its average capacity factor, but the peak to average ratio is its inverse, thus 3–4:1 for on-shore wind, 2–2.5:1 off-shore, and 7–10:1 for PV in N Europe, lower in the south. At high VRE penetration (e.g. above 50%) peak output will exceed total demand (including exports and storage), and the resulting surplus must be curtailed (i.e. spilled or discarded). Marginal curtailment, that is the number of hours the last MW of capacity is prevented from delivering, is typically more than three times average curtailment, but price signals for VRE investment are given by average, not marginal, curtailment, leading to potentially excess entry by merchant VRE, even if normal marginal system integration costs are efficiently priced.

The contribution of this article is, first, to emphasize the implication of far higher marginal than average curtailment of each VRE technology for assessing the benefit of further additions of that technology. The second and main contribution is to identify a largely overlooked integration cost that in conventional systems is costless, but becomes costly once curtailment becomes unavoidable. The third contribution of this article is to cast doubt on claims (Ueckerdt et al., 2013; Korpås and

E-mail address: dmgn@cam.ac.uk.

[1] I am indebted to very helpful comments from Par Holmberg, Robert Ritz, Richard Green, Stan Zachary and Chris Dent with the usual disclaimer and to the presentations at the EERA Expert Workshop In Leuven in September, 2022 (EERA, 2022). I am also indebted to the two journal referees.

[2] provided they experience constant returns to scale, which is an acceptable and the default assumption in power system modelling.

https://doi.org/10.1016/j.eneco.2023.107011
Received 24 January 2023; Received in revised form 21 July 2023; Accepted 28 August 2023
Available online 9 September 2023

Botterud, 2020) derived from residual load curve analytical models that efficient pricing will in long-run equilibrium result in revenue adequacy for all types of generation and storage batteries.[2]

The overlooked system integration cost is the need to provide adequate inertia. System inertia is costless with adequate controllable generation synchronously connected to the network, as spinning turbines provide inertia whenever they are exporting electricity. Non-synchronous (or inverter-based) generation like wind, solar PV and DC links that connect to the network by power electronics (inverters) lack the natural inertia of a spinning turbine.

Inertia is needed to even out fluctuations in demand and supply — if there is a sudden shortfall, the system frequency (50 or 60 hertz — cycles per second) starts to fall, and its rate of fall depends on how much power can be drawn from the inertia connected to the system. Systems are designed to ensure that frequency is kept within a narrow range (e.g. 49.5 to 50.5 hertz in GB) and primary reserve capacity is kept ready to increase output before that limit is reached. The lower the system inertia, the more challenging it is to activate this replacement inertia in time. In addition, if the rate of change of frequency exceeds a critical level then generation and load will automatically disconnect to protect equipment, potentially leading to a system-wide black-out or requiring load shedding.

As the share of simultaneous non-synchronous (inverter-based) penetration (SNSP) reaches a critical level, the share of inertial generation will fall to a dangerous level at which frequency can no longer be reliably maintained within its safe limits. The System Operator will intervene to keep enough inertia on the system by curtailing non-synchronous generation. As VRE penetration increases, new ways of providing synthetic inertia without the need for the spinning mass of turbines are likely to be introduced. Already synchronous condensers and grid-forming inverters are increasingly deployed to manage shortages of conventional inertia, notably in Australia where interconnections are weak. In South Australia, VRE penetration reached 68.4% in 2022. In a notable recent event when the grid was disrupted by broken transmission lines, South Australia was able to accept a peak VRE penetration of 91.5% on 19 November 2022 because of the presence of four new synchronous condensers and the Hornsdale battery (150 MW, 193 MWh). “In order to manage the power system, AEMO directed some synchronous generators online for FCAS (Frequency Control Ancillary Service) provision, and instructed the local network provide (sic) to curtail distributed PV for four to 10 h each day from November 13 to 17, and on November 19. *Renew Economy* understands this affected up to 400 MW of capacity.[3] Even with these synthetic sources of inertia, system stability still required synchronous generation and limits on VRE penetration. This article considers two stability requirements — for inertia, already discussed, and the need to provide fast-acting reserves to maintain demand–supply balance in the event of outages or line failures.

VRE at high penetration levels will inevitably lead to curtailment, either for system stability reasons, or more obviously because potential VRE supply exceeds total system demand (including export and storage options). Curtailment can be reduced but only by costly actions, keeping more expensive flexible generation operating, and by investing in synthetic inertia, transmission investment (including interconnectors to other regions or countries) and/or storage. The contention of this article is that analytical and systems models have overlooked additional inertia costs, which can be significant as the quantification developed below argues. Thus in the case of the Single Electricity Market (SEM) of the island of Ireland this cost of inertia can add 10%–20% to the cost of wind investment under 2022 system operating codes. In future the SEM aims to raise SNSP levels by a combination of ancillary services, synthetic inertia and the requirement for new wind farms to provide a fast frequency response (FFR) service (IEA, 2021, p81).

The next section surveys the literature and identifies the various integration costs of VRE. Section 3 sets out the analytic model and shows cost recovery in liberalized electricity markets for dispatchable generation. Section 4 introduces VRE, marginal curtailment and the need for a corrective charge to account for required inertia before quantifying this curtailment cost for the SEM. Section 5 concludes with some policy implications.

2. Literature review and integration costs

The literature on the costs and benefits of high VRE penetration can be found in the rather separate electrical engineering and energy economics literature, with recent attempts to reconcile them (Ueckerdt et al., 2013; EERA, 2022). The engineering cost-based approach starts from the levelized cost of electricity (LCoE) of a particular technology, such as wind, and then adds a number of system integration costs to determine the marginal social cost of that technology, often described as the System LCoE or the market value levelized cost, MLCoE. The philosophy of cost-based approaches is to provide estimates for policy makers to assess the costs of decarbonizing or otherwise changing their power systems (National Grid ESO, 2022). The UK Government periodically updates the cost elements of different technologies and their LCoEs (BEIS, 2020), which provide only part of the costs needed in constructing suitable energy portfolios. The more comprehensive energy system optimization models attempt to include the ancillary service and storage options but typically only model long-run equilibrium configurations, often with a limited range of system stability constraints. Zerrahn and Schill (2017) provide an excellent review of such models, but even their comprehensive model does not address inertia requirements.

In contrast, the economic approach either uses observed market prices to measure the cost of the ancillary services needed for VRE (e.g. Savelli et al., 2022), or examines whether their prices are at efficient levels. If not, the observed distortions may be removed by a change in charges, prices or contract design (Newbery, 2023a). The economic argument is that an efficient set of market prices for all the various ancillary services and grid charges will guide generators to make efficient entry and operating choices (Joskow, 2011). The different time scales of forward-looking investment and hourly operating decisions may require quite complex contracts for network access. That requires a forward or system view to determine future prices, amounts of ancillary services needed and resulting network charges, so there is an obvious practical overlap with the engineering/systems modelling and economic approaches.

The three main categories of the system integration costs of VRE normally identified are (i) *balancing* costs arising from the unpredictability of output and load combined with the difficulty of costlessly varying the output of controllable generation, including additional reserve costs for system security, (ii) the merit-order or *profile* costs caused by the lack of correlation of VRE with load, thus adversely impacting the residual demand to be met by controllable generation, and (iii) a catchall category of *grid* costs that depend on network constraints, the location of VRE and resulting congestion costs (Heptonstall and Gross, 2020; Savelli et al., 2022).

The economic approach is top-down (from the market to cost) compared to the engineering bottom-up cost-based approach. It starts by recognizing that generation assets provide a wide range of services with differing values. For an efficient choice of technology all these characteristics need to be properly rewarded. Energy Systems Catapult (2019) (fig. 7) lists these as the five C’s: Commodity (i.e. energy, MWh), Capacity (MW), Capability (the ability to provide a range of ancillary services), Congestion (its management, the value of reducing congestion), and Carbon (the cost of CO_2 emissions). Depending on the category these will vary with, and thus need to be differentiated by,

[3] https://reneweconomy.com.au/glimpse-of-the-future-south-australia-peaked-at-91-5-pct-wind-and-solar-when-links-were-down/

time and location. One obvious problem in looking at the value of any technology is that many of these value streams are jointly produced but variously constrained (thus energy cannot exceed capacity, and energy at one location may add to or reduce congestion depending on its capability). Capability can become critical at high VRE penetration as system stability will require a minimum level of dispatchable plant running to deliver other ancillary services than inertia (such as primary reserves).

The reconciliation of the two approaches in long-run equilibrium can either start with costs (the LCoE) and add the (marginal) system integration costs to reach the correct (social) market value (MLCoE) (Ueckerdt et al., 2013) or start with the observed wholesale market value of energy and deduct the technology cost (LCoE) to derive the market value of the integration services (which in equilibrium should be the same as the marginal integration costs). The technology-specific integration costs are the additional services that have to be procured from the rest of the system to allow the unit to be efficiently accommodated. Efficient decentralization then requires targeting marginal additional integration costs on those who cause them (Milligan et al., 2011). There is usually a considerable difference between the short-run integration costs and the costs after reaching long-run equilibrium. That difference can be called the transition cost (Ueckerdt et al., 2013). Short-run costs can be reduced by installing battery storage at critical points, reducing fossil capacity, reinforcing networks and building interconnectors (Newbery, 2018b). Short-run integration costs are highly system dependent and unlikely to be similar across different systems, even controlling for the level of VRE penetration (Hirth et al., 2015), p 927, (Heptonstall and Gross, 2020)). Even in long-run equilibrium, integration costs will depend on the resource base (wind, sun, available flexible plant options), as well as market size and network constraints that are not cost-effective to remove.

Mowers and Mai (2021) provide an extensive survey to demonstrate and explain the wide range of differences in integration costs. They point to the limitations of using observed market prices for various ancillary services and the difficulty of disentangling integration costs derived from system simulation studies. Instead they suggest a theoretical framework for determining the relative value of each technology (e.g. wind) compared to a reference technology. They suggest the reference could be the hypothetical "technology that contributes to all system requirements (at all locations) in proportion to the requirement levels themselves" (Mowers and Mai, 2021, p3). The integration cost is then the shortfall in the value offered by the technology. This approach is similar to that suggested for measuring a plant's contribution to system adequacy by Effective Load Carrying Capacity, discussed next.

2.1. De-rating capacity for system adequacy

Effective Load Carrying Capacity (ELCC) or Equivalent Firm Capacity (EFC) attempts to measure the contribution any plant makes to system adequacy, a concept much debated in the literature.[4] "ELCC measures the amount of load that can be added to a system given the addition of a resource, while maintaining the same level of reliability . . . " (ESIG (2021, p18). However, "Numerous studies suggest that the ELCC of a resource type is highly dependent on the underlying resource mix and the load profile — both of which change continuously." ((ESIG, 2021, p19.) ESIG (2021), fig. 4) shows that as the number of dispatchable independent units increases, so the system probability of normal failure falls asymptotically.[5] In consequence, in reasonably large systems each technology type can be equally de-rated. The same is not true for VRE, where the resource (wind or sun) is typically highly correlated over a region, implying that it is a better approximation to consider them as single large units, whose loss has a material and uncertain impact. Thus the average capacity factor of wind in winter, when scarcities are more likely, may be well above its annual average, but despite this, wind cannot be sufficiently relied upon to deliver at times of system stress to count as firm. If VRE is paid the Value of Lost Load in scarcity hours, it may earn considerably more than its de-rated capacity suggests, so that unless its unpredictability is properly taken into account, it will be overpaid, inducing excess entry.

Bothwell and Hobbs (2017, p174) note that "the marginal contribution of wind and solar often decreases as the installed amount increases (Keane et al., 2011)." Part of the reason is curtailment, discussed below, but a more important reason is the correlation impact discussed above. Keane et al. (2011) is particularly relevant in underlining that the EFC of wind not only depends on the amount of wind capacity, but on the strength of the wind in any year, illustrating this for Ireland between 1999 and 2008. Storage can raise the EFC of VRE by reducing the consequences of short-term variability.

This dependency and its implication for the measurement of EFC has been brought more up to date in Zachary et al. (2019). That article also provides a useful discussion of the relationship between two different reliability metrics, the Loss of Load Expectation (LoLE, number of hours on average per year when load may be shed) and Expected Energy Unserved (EEU). For many but not all purposes there is a direct mapping between them, justifying the choice of LoLE as a suitable metric (but not for the evaluation of storage). They note that VRE can be treated in the same way as conventional plant only if "the process of variable generation is statistically independent of that of demand, in which case the de-rated level of variable generation is close to its mean value" — which only holds if in addition the Value of Lost Load remains constant regardless of the length of disconnection, and that people are risk neutral as between a high probability of a short disconnection and a lower probability of a long disconnection, each with the same number of EEU. Söder et al. (2020) show that there is no uniform solution in their review of de-rating VRE in 14 different counties.

2.2. Other integration costs

Most studies of the impact of VRE concentrate on their price impact — the merit order or profile effect in which low variable cost renewables push out the supply curve and lower prices, impacting the profit of incumbents. The static merit order impact of renewables capacity in displacing fossil plant is well-understood (Clò et al., 2015; Cludius et al., 2014; Deane et al., 2017; Green and Vasilakos, 2010; Ketterer, 2014; Csereklyei et al., 2019). The long-run equilibrium effect is quite different, depending on the entry and exit decisions of conventional plant. Green and Léautier (2015) provide the most sophisticated analytical model. Korpås and Botterud (2020) show that in long-run equilibrium all plant, including VRE and battery electrical storage, breaks even under constant returns to scale — in stark contrast to the short-run profile profit impacts, and a more reasonable approximation, albeit a claim that is questioned in the model below.

The literature on learning effects is mainly concerned to estimate its rate, summarized in Rubin et al. (2015). Gambhir et al. (2021) discuss how it should be treated in system models. Green and Léautier (2015) include learning-by-doing in their model of optimal support for renewables, and calibrate the model for GB. Newbery (2018a) develops

[4] There is a related debate about whether to measure system reliability by the Loss of Load Expectation (LoLE hrs/yr), as in this article, or Expected Energy Unserved (EEU, MWh/yr), not to mention measuring resilience to rare but high impact weather events, where different technologies may be more or less adversely impacted, with different contributions to reliability (Wolak, 2022).

[5] excluding design or generic faults that may require outages of a number of similar plants, such as the 2022 French nuclear outages.

an algebraic model to estimate the global benefits of additional investment and the justified subsidy, which varies between technologies and thus justifies different subsidies to each type (apart from charging different grid connection charges reflecting their time-different impacts on network costs).

The literature on curtailment concentrates on either local curtailment and congestion management, as discussed by Joos and Staffell (2018) for Britain and Germany, or the need to increase storage for a lower-cost future system (Pudjianto et al., 2014; Weiss and Wänn, 2013). Henriot (2015) looks at inflexibilities caused by priority dispatch of VRE and intertemporal constraints where inflexible generation takes many hours from cold to be available. These costs depend on the system flexibility and are ignored in the model below. At past rather low levels of penetration, Heptonstall and Gross (2020) find that “the median values for the share of VRE output curtailed across all penetration levels is consistently low, not exceeding 5%” but as this article shows, because the marginal curtailment is many times the average level, this can rapidly rise without a very flexible system. The SEM, where curtailment was already above 8% by 2020, therefore provides a foretaste of the future. To the best of our knowledge there are no studies on the implications of the difference between marginal and average curtailment for market failure and the need for a corrective charge.

2.3. The requirements of efficient pricing

Efficient pricing requires identifying and correcting all externalities. Both decarbonization and supporting VRE are global public goods, VRE through its learning externalities that lower the cost of future investment. To solve the problem of financing such public goods, the EU requires (in its *Clean Energy Package*) member states to agree to targets for emissions reduction and VRE penetration — an excellent example of turning these into club goods (Buchanan, 1965). The UNFCCC Paris Agreement and *Mission Innovation*[6] are examples of widening the club, ideally to the whole world. The EU’s targets are set out in the *2030 Climate and Energy Framework*.[7] For these to be delivered in liberalized electricity markets, a number of market failures and distortions will have to be addressed.

The first and most obvious is that the external costs of fossil generation, and particularly CO_2 emissions, will need to be properly charged.[8] The EU’s chosen instrument is the Emissions Trading System, but until its reform in 2018, the resulting carbon prices were well below the social cost of carbon. From December 2020 to January 2022 the EU and GB carbon prices were above €80($86)/tonne CO_2 and so above the Paris target-compliant level. In countries with inadequate carbon prices, VRE integration costs can be credited with the value of CO_2 displaced, which will be fuel-mix dependent (Savelli et al., 2022; Chyong et al., 2020).

The second is that the external learning benefits of deploying VRE should be appropriately rewarded (see (Newbery, 2018a) and references therein; (Gambhir et al., 2021)). The EU’s policy here has been to set targets for renewables share in total energy, and to encourage innovation through its European Strategic Energy Technology Plan (which, however, is aspirational rather than requiring binding commitments). As learning depends on developing, designing and installing reliable capacity, the learning benefits are a function of cumulative installed capacity, not subsequent output (when the electrons are the same as those from fossil generation). That implies the subsidy should be paid to reliable capacity (e.g. for the first 30,000 MWh/MW)[9] and not to output (as with the EU’s assigned target shares of output), which would distort the market (Newbery et al., 2018; Newbery, 2023a). Unfortunately, most subsidy systems create considerable distortion costs (Peng and Poudineh, 2019).

The third implication of high VRE penetration is to threaten the efficiency of investment decisions in the flexible plant required for capacity reliability. There is growing consensus that, while an energy-only market with prices capped at the Value of Lost Load might, in ideal circumstances, deliver the right level of reliability, a capacity auction, perhaps for Reliability Options, reduces the risk (particularly of future policy uncertainty) and hence the cost of delivering reliability (Batlle et al., 2007; Grubb and Newbery, 2018; Newbery, 2016a, 2017). Holmberg and Ritz (2020) investigate the case for capacity payments with price caps for systems with high renewables penetration in a model complementary to that developed here. Joskow and Tirole (2007) set out the stringent conditions under which well-designed markets could deliver the specified level of reliability in markets with price caps and capacity obligations, where only a fraction of customers respond to real-time scarcity prices. Working back from a derivation of the Value of Lost Load (VoLL, which they point out is unlikely to be independent of the nature of the load-shedding event), they show in their benchmark case that all generators and Load Serving Entities should face the VoLL in cases of load shedding. They conclude that the unusual physical characteristics of electricity and networks “makes achieving an efficient allocation of resources with competitive wholesale and retail market mechanisms a very challenging task.” (Joskow and Tirole, 2007, p83). This article highlights the way in which VRE should be treated in de-rating for procuring adequate capacity.

The fourth implication is that marginal system costs should be charged to the source of these costs (Milligan et al., 2011). Kaffine et al. (2020) point to an additional cost of VRE, in that intermittency over short time periods raises CO_2 emissions from flexible fossil generation, but that would be addressed if CO_2 were correctly priced. There are empirical estimates of the short-run systems costs of variable renewables at varying levels of penetration (Heptonstall and Gross, 2020) discussed above, and simulations of possible future costs at high VRE levels (e.g. the extensive list of references in Committee on Climate Change (2019)), but little by way of simple modelling that can give better insights and simpler estimates.[10]

All of these are widely recognized in the literature discussed above, but there is an additional cost that has not been recognized or quantified. Beyond some level of penetration, excess VRE must be curtailed. If wholesale prices are efficiently set, this will cause the price to fall to the avoidable cost of the (very low, possibly zero) marginal VRE, encouraging self-curtailment. The contribution of this article is to argue that the *marginal* curtailment is many times higher (3–4+) than the *average* level of curtailment. For *efficient* investment decisions it is marginal curtailment that is relevant, while the market only values (or penalizes) average curtailment, resulting in a “tragedy of the commons” (Hardin, 1968). This is similar to the distortion that is claimed to arise in some models of collective ownership (e.g. Meade (1972) in which n workers share in total profit, but may have only $1/n$ incentive to add to that profit — a theory that has spawned an immense literature.

This article models and quantifies this implication of curtailment for the specific case of wind using residual demand analysis, a widely used (Korpås and Botterud, 2020) but necessarily limited approach to full system integration analysis. Residual demand analysis assumes that decisions in any hour have no dependence on decisions in any other hour, and thus ignores start-up costs and ramping constraints, as well as short-term uncertainties that may impact costs. A number of integration costs that might appear to be missing from this approach can be included. Thus congestion and grid expansion costs can be included by assuming nodal pricing and efficient connection contracts,

[6] see http://mission-innovation.net/

[7] at https://ec.europa.eu/clima/policies/strategies/2030_en

[8] Other air pollutants, particularly from coal, can also be costly — see Holland et al. (2020).

[9] Steinhilber (2016) notes that this is the support used for wind in some parts of China.

[10] But see Ueckerdt et al. (2013) and Korpås and Botterud (2020).

with the market area suitably redefined. Capacity or adequacy costs can be included by suitable de-rating of nominal capacity to Equivalent Firm Capacity (EFC), with all the problems discussed above.

The simple model below shows first, the source and magnitude of this curtailment cost and second, the source and size of the shortfall between *social* value (that depend on *marginal* curtailment) and *market* revenue that is only reduced by *average* curtailment. It derives the corrective inertia charge needed in a liberalized market to allow "subsidy-free" entry of VRE and estimates its magnitude. The data for the empirical estimates are taken from the Single Electricity Market (SEM) of the island of Ireland (downloadable from Newbery (2020)), where the value of the global learning externality (derived from Newbery (2018a)) is shown to be comparable to and offsets the inertia charge. The SEM is a particularly important market to study, as it is widely recognized as being at the forefront of addressing the challenge of high VRE penetration in a small, isolated system (Newbery, 2021).

3. The model

The model is the simplest version to illustrate the problem. A fuller model is available in the earlier working paper (Newbery, 2020) but the results are essentially the same. In this version all obvious market failures are assumed away or internalized, so that VRE faces market prices when deciding whether or not to generate, the carbon price (relevant for fossil generation) is at the efficient level and all learning spill-overs are either zero or properly remunerated. As with all models using residual demand curves, it is assumed that ramping constraints do not bind, start-up costs can be ignored, and that output in any hour is independent of that in any other hour. Congestion and balancing costs are internalized by nodal pricing. There are constant returns to scale, required to show that revenues can be equal to costs in equilibrium.

The electricity market is isolated and has determined the Value of Lost Load (VoLL), V, which sets the reliability standard of L hours Loss of Load Expectation (LoLE) per year. To show that the problem arises because of curtailment and not because prices can fall to zero, initially there are three conventional (i.e. controllable or dispatchable) types of generator: peaking plant (e.g. open-cycle gas turbines), conventional fossil base-load plant (e.g. combined cycle gas turbines), and nuclear plant with zero variable cost. In even modest-sized regions, most individual generating units are small relative to total demand, so can be considered smoothly expansible at constant cost, giving constant returns to scale. Their Equivalent Firm (or de-rated) Capacities (EFCs)[11] are P, F, and N, with annual unit fixed costs (to recover the capital and other fixed costs) r_J and unit variable operating costs v_J ($J = P, F$, and N for the plant types), with $v_P > v_F > v_N = 0$, and $r_N > r_F > r_P$.

Let $D(t)$ be demand in calendar hour t with the Load Duration curve $D(h)$ with $D' < 0$, so that load is re-ordered with the highest load in hour 1, where h is the number of hours that demand is higher than $D(h)$. Thus $t = 1$ might be hour 1 of Jan 1, $h = 1$ is the hour of highest demand. Then $D(L)$ is the required firm (de-rated) capacity required to meet the reliability standard. The market is in long-run equilibrium with all future costs and $D(h)$ known and constant. The ability of all plant to generate up to its EFC is assumed independent of its output in the preceding hour, so that in the ordered hours, h, the sole determinant of plant outputs are $D(h)$ and their relative cost. The total cost of meeting demand (except for the L hours of lost load) and the cost of the amount of lost load (in MWh), valued at the VoLL, V, is

$$C = Nr_N + Fr_F + Pr_P + V\int_0^L (D(h) - D(L))dh + v_P PL$$
$$+ v_P \int_L^{h_P} (D(h) - N - F)dh$$

[11] National Grid ESO (2019) explains the concept and resulting de-rating factors for GB. De-rating correctly implies that de-rated capacity can run 8760 h/year.

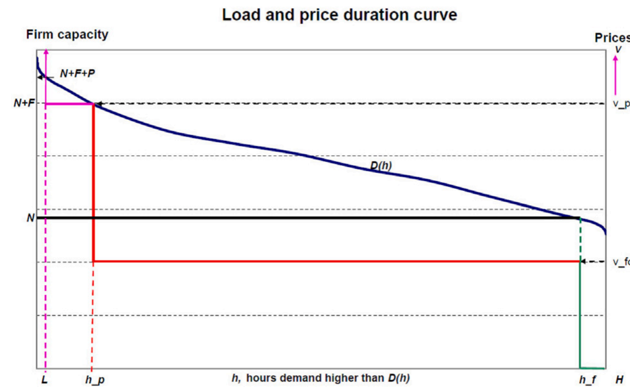

Fig. 1. Load and efficient price duration curves.

$$+ v_F F h_p + v_F \int_{h_P}^{h_F} (D(h) - N)dh + v_N N h_F + v_N \int_{h_F}^{H} D(h)dh, \quad (1)$$

(where the last two terms can be ignored as $v_N = 0$). Here h_p is the number of hours peak generators run, h_F is the number of hours fossil baseload plant runs, and H is the number of hours in the year (8760). Peaking generation only runs when $D(h) \geq N + F$ and fossil generation will run all the hours for which $D(h) \geq N$. The first integral is the cost of lost load, the following term ($v_P PL$) is the cost of running peaking plant for the hours that it runs at full capacity, followed by the second integral when it runs at less than full capacity, and similarly for the other technologies. Fig. 1 shows the load duration curve (using GB load data to illustrate its likely shape) and the amounts of firm capacity and the aligned efficient price duration curve on the right hand axis (truncated as $V \gg v_P$).

The LoLE, L, satisfies $D(L) = N + F + P$, fixed by the VoLL, V in ((3)) below. It follows that $\partial P/\partial F = \partial P/\partial N = -1$, allowing us to ignore the choice of P in the optimization. The first order conditions for cost minimization of (1) are

$$\begin{aligned} 0 &= \frac{\partial C}{\partial F} = (r_F - r_P) - v_P(L + h_P - L) + v_F h_P, \\ 0 &= \frac{\partial C}{\partial N} = (r_N - r_P) - v_P h_P - v_F(h_F - h_P) + v_N h_F \\ &= (r_N - r_P) - h_F v_F - h_P(v_P - v_F), \\ h_P &= \frac{r_F - r_P}{v_P - v_F}, \quad h_F = \frac{r_N - r_F}{v_F}, \end{aligned} \quad (2)$$

which is a well-established result from screening curve analysis (e.g. see Stoft, 2002).[12] Once the critical hours are fixed, the required efficient EFCs for different plant can be deduced from $N = D^{-1}(h_F)$, $F = D^{-1}(h_P) - N$, $P = D(L) - N - F$. While the critical hours h_J depend only on costs, the capacities and amount generated by each type of plant depend on demand.

3.1. Decentralizing the efficient solution

Efficient prices when there is adequate capacity to meet demand, p_J, will be equal to the short-run system marginal cost (SMC) set by the most expensive plant required to meet demand. In this constant returns case $p_J = v_J$, $J = P, F, N$ unless load must be shed. When load is shed to balance demand and supply (in the L shortage hours) the price will be set by the demand side at the Value of Lost Load, (VoLL), $p_L = V$. Peaking generation only runs for h_P hours and must cover its cost.

[12] Screening curves are the plot of total cost of any technology against hours run. The least-cost plant mix is their lower envelope, so that peak capacity has the lowest intercept but the steepest slope and meets the base-load fossil where $r_P + h_P v_P = r_F + h_P v_F$. Similarly the fossil total cost line meets the nuclear total cost line where $r_F + h_F v_F = r_N$.

It only makes profits when prices are higher than its avoidable cost, which, except for lost load hours, will also be the price. This leads to a very simple (and widely recognized) relationship between the VoLL, V, the LoLE, L, and the net cost of new entry (net CoNE, i.e. net of any revenues earned in the market):

$$r_p = L(p_L - v_p) + (h_P - L)(p_P - v_p) = L(V - v_p). \quad (3)$$

The net unit profit of base-load fossil plant will be, after cancelling zero terms where $p_F = v_F$ and substituting for V from (3) and h_P from (2):

$$\begin{aligned}\Pi_F &= L(V - v_F) + (h_p - L)(v_P - v_F) - r_F, \\ &= r_p + h_p(v_P - v_F) - r_F = 0.\end{aligned}$$

In other words, free entry that drives net profit to zero delivers the efficient volume of this capacity (as shown earlier in Newbery (2016b)). The same is readily shown to be true of nuclear net profits (where free entry is interpreted as "subsidy-free" contracted entry). Indeed, the reason for including nuclear power with, in this case, an assumed variable cost of zero, is to show that it is not the zero variable cost of wind that is the cause of the problem, merely the symptom. Nor is the source of the problem the sudden discontinuity in the market price when moving from one marginal technology to another (including nuclear power or wind) as Fig. 1 shows several sharp discontinuities.

4. High penetration of variable renewable electricity

In many markets (e.g. Queensland, see Gohdes et al. (2022)) the cost of solar PV has fallen so far that it is very competitive against fossil generation, and households are willing to install solar panels without subsidy. Similarly on-shore wind and increasingly off-shore wind are becoming competitive (Jansen et al., 2020). Thus in the UK, the July 2022 renewables auction cleared at a strike price[13] £$_{2022}$46 (\$70)/MWh for solar PV and at £$_{2022}$42.5(\$65)/MWh for off-shore wind, considerably below the pre-crisis carbon inclusive cost of gas-fired generation. As such they often seek "subsidy-free" long-term contracts.[14] The model is one of long-run equilibrium in which capacity has time to adjust, and thus rules out short-run disequilibrium reasons for new plant entry.

Consider wind as the exemplar VRE whose output in any hour per MW of capacity is $\tilde{\phi}$, a random variable independently drawn from its distribution for each hour, and thus with a constant expectation, ϕ. To simplify, its variable cost is taken as $v_W = 0$ (both assumptions are relaxed in Newbery (2020)). If name-plate installed capacity of wind is W, its EFC is taken at δW, determined according to best practice by the System Operator when procuring required capacity (and discussed further below). The model above is now modified by dropping nuclear power, and replacing the reliability constraint with $D(L) = P + F + \delta W$. Conventional plant now needs to deliver the residual demand–demand net of VRE, $R(t)$. The potential supply to meet total demand $D(t)$ in hour t is $G(t)+\phi(t)W$ (where $G(t)$ is output from conventional generation) but his may be excessive, in which case the wind will need to be curtailed.

In addition, there are a set of requirements to ensure system stability, explained in more detail in Newbery (2021). Initially the relevant constraint is that the share of non-synchronous generation (specifically VRE, as it is connected to the grid asynchronously and cannot normally deliver inertia) must be kept below a specified fraction of demand to reduce the rate at which frequency drops with a supply loss or a sudden increase in demand. The Grid Codes specify the allowable Rate of Change of Frequency (RoCoF) that determines the amount of inertia to avoid breaching the RoCoF standard. This is normally specified by the maximum acceptable System Non-Synchronous Penetration (SNSP). Thus in the Single Electricity Market (SEM) of the island of Ireland studied in Newbery (2021) the target 2020 SNSP is 75% (since achieved). The 2030 SNSP target is 90%, presumably requiring substantial synthetic inertia and other frequency responsive services. At higher levels of SNSP the requirement to have other flexible plant running to deliver primary reserves (in particular to manage the loss of the largest plant or network link on the system) will require a minimum of m MW of dispatchable plant immediately available.

The level of SNSP will initially be critical in determining the amount of curtailment and hence the size of the resulting market distortion, and to that end define the required share of conventional generation as β = 1 - SNSP (so β = 25% in the SEM case). Thus $G(t) \geq \beta D(t)$ and curtailment will be needed in amount $k(\phi(t)W, t) = \max(0, \phi(t)W - \max\{(1-\beta)D(t), m\})$. In this section it is assumed that $Min(\phi(t)W - (1-\beta)D(t)) > m$, so the inertia constraint binds first. This condition will be relaxed after discussing the inertia constraint. While $\phi(t)W$ is potential wind output, actual or useful wind output will be $w(t) = \phi(t)W - k(\phi(t)W, t)$. Residual demand is then $R(t) = D(t) - w(t)$ and can be ordered for the set of hours with and without curtailment. Define h as hours without curtailment, ordered so that $R(h), R' < 0$, over $[0, H - h^*)$. For the remaining h^* hours wind is curtailed. It is convenient to define y as hours with curtailment with the curtailment function $k(\phi(y)W, y) \equiv k(y, W)$ separately ranked with $k' < 0$, over the range $[0, y^*)$, where $y^* = h^*$ is the solution to

$$k(h^*, W) = \phi W - (1-\beta)D(H - h^*) = 0. \quad (4)$$

An example may be helpful. Suppose, quite plausibly,[15] that $D(h) = M - (M - m)h/H$ on $[0, H]$ and (implausibly) that $\phi(h)$ is linear and perfectly negatively correlated with demand: $\phi(h) = h/H$. Then $h^* = y^*$ solves

$$\begin{aligned}W(1 - h^*/H) &= (1-\beta)\{M - (M-m)(1 - h^*/H)\}, \\ h^*/H &= \frac{W - (1-\beta)m}{W + (M-m)(1-\beta)}.\end{aligned} \quad (5)$$

Fig. 2 illustrates the residual demand curve, curtailment and actual wind output, plotted as functions of h as curtailment increases monotonically beyond $H - h^*$ because of its perfect negative correlation with demand.

Fig. 3 gives an illustrative (but still stylized) example using GB demand and actual wind data for 2018, but scaling wind up every hour by a factor of three, and then considering an increase in wind capacity of 1000 MW (from the assumed start level of 39,100 MW).[16] It shows the residual demand ranked in descending order over all hours, h, with the volume of wind curtailed in the same hour, h. The curtailment function is then graphed as $k(y, W)$ on $[0, y^*)$ with $k' < 0$. In this more realistic case where wind has little correlation with demand there is no simple relation between h and y.

The normal way to measure curtailment is the volume of wind curtailed, $\int_0^{h^*} k(W, h)dh$, which in general will be higher than $h^* W \phi$ as curtailment hours are likely to be hours of above average capacity factors. Existing wind farms experience average curtailment per MW of installed capacity (and the associated hours of zero profit) of $\int_0^{h^*} k(W, h)dh/W$ per MW of capacity. Marginal curtailment caused by the entry of 1 MW of extra wind capacity is

$$\begin{aligned}\frac{\partial}{\partial W}\int_0^{h^*} k(W, h)dh &= k(W, h^*)\frac{\partial h^*}{\partial W} + \int_0^{h^*}\frac{\partial k(W, h)}{\partial W}dh, \\ &= \int_0^{h^*}\frac{\partial k}{\partial W}dh.\end{aligned} \quad (6)$$

[13] The auctions were for a CfD with FiT, which pays the strike price on metred output for 15 years, regardless of the market price.

[14] In this perfect foresight stationary world long-term contracts are redundant but in an uncertain world with missing long-term futures markets they are essential for reducing finance costs.

[15] The demand duration curve of Fig. 1 was taken from GB data and is nearly linear over much of its length.

[16] GB demand is as measured, PV is ignored, actual wind in each hour is trebled, all storage and exports/imports are ignored.

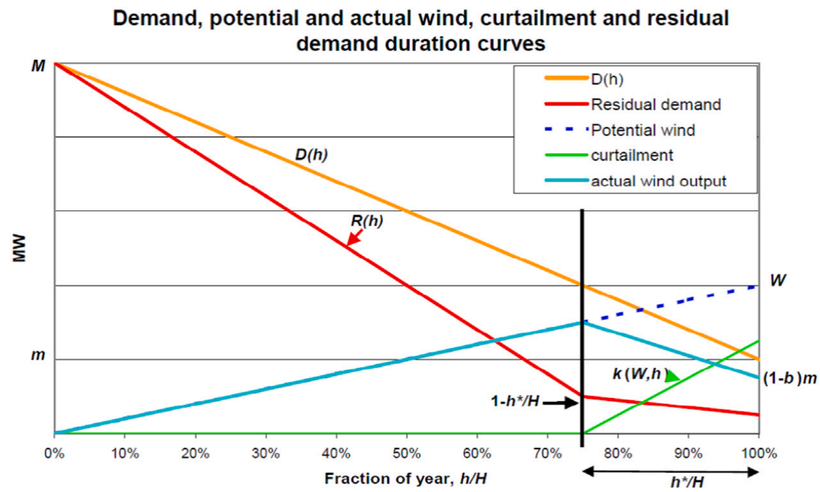

Fig. 2. Hypothetical duration curves for perfectly negatively correlated wind.

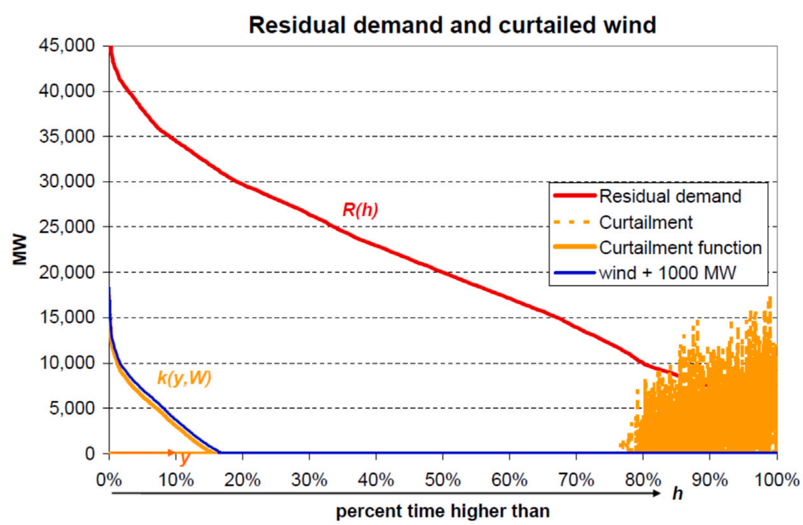

Fig. 3. GB residual demand and curtailment function, scaled 2018 wind.

The ratio of the marginal to average curtailment is $W \int_0^{h*} \frac{\partial k}{\partial W} dh / \int_0^{h*} k dh$. In Fig. 3, the curtailment function is roughly linear in y over much of its range and can be approximated by

$$k = \alpha(W - W_0)(1 - y/y^*), \tag{7}$$

with W_0 the level of wind at which curtailment first appears. In this GB case W_0 = 20,551 MW, W = 35,928 MW),[17] $\alpha = 0.318$, and $y^* = h^* = 1361$ h. Appendix shows that the ratio of the marginal to average curtailment from (13) is just $2W/(W - W_0)$ or 4.7, considerably greater than 2. This is a startling result and has a number of implications, some of which are pursued elsewhere (Newbery, 2023b). Perhaps the most obvious is that the cost of delivering a flow of output rises almost inversely with marginal curtailment. Thus if average curtailment is 15%, and as a result marginal curtailment is 50%, the last MW of installed VRE capacity will be almost twice the cost per MWh compared to low penetrations with no system-wide curtailment. Newbery (2021) gives more realistic estimates for island of Ireland in 2026 taking account of storage and exports and finds the ratio 3.66.

[17] The peak capacity factor for the whole of the UK is 93.4% (as wind farms in different locations are not perfectly correlated) so the peak wind output is divided by 0.934 to derive the implied capacity. The increase in 1000 MW gives an increase in peak wind output of 934 MW.

Total system costs with wind but without nuclear and replacing $P = D(L) - F - \delta W$, where δW is the System Operator's reliability EFC of wind capacity W, will be

$$\begin{aligned} C = {} & W r_W + F r_F + (D(L) - F - \delta W)(r_P + v_P L) \\ & + V \int_0^L (D(h) - D(L)) dh \\ & + v_P \int_L^{h_P} (D(h) - F - \phi W) dh + v_F F h_p \\ & + v_F \int_{h_P}^{H-h^*} (D(h) - \phi W) dh, \end{aligned} \tag{8}$$

as the variable cost of wind is zero. The first-order condition for minimizing fossil generation cost are unchanged:

$$0 = \frac{\partial C}{\partial F} = (r_F - r_P) - (v_P - v_F) h_P,$$
$$h_P = \Delta r / \Delta v, \quad \Delta r \equiv r_F - r_P, \quad \Delta v \equiv v_P - v_F,$$

and the length of time the peaking plant is needed is invariant to installed capacities (although capacities do depend on demand).

The total surplus (consumer surplus less generation cost) is $S = V \int_0^H D(h) dh - C$, which, after noting that the envelope condition allows us to remove all terms in F from (8), becomes

$$S = V \int_L^H D(h) dh - W r_W - (D(L) - \delta W)(r_P + v_P L) \tag{9}$$

$$+v_P\int_L^{h_P}(\phi W - D(h))dh + v_F\int_{h_P}^{H-h^*}(\phi W - D(h))dh.$$

4.1. The marginal social value of VRE

Curtailment implies that the efficient price during curtailed periods will be the avoidable cost of wind (or VRE more generally), taken as zero, and as $\partial h^*/\partial W > 0$, additional wind will cannibalize the revenue from existing wind, as the number of profitable hours will decrease. However, new entrants enjoy the average, not the marginal curtailment that is relevant for assessing the benefits of additional wind investment. The social benefit of an extra MW of wind capacity will be, from (3) and (9):

$$\begin{aligned}\frac{\partial S}{\partial W} &= \delta(r_P + v_P L) + \phi v_P(h_P - L) + \phi v_F(H - h^* - h_P)\\ &\quad - v_F(D(H-h^*) - \phi W)\partial h^*/\partial W - r_W,\\ &= (\delta V - \phi v_P)L + \phi\{r_F - r_P + v_F(H-h^*)\}\\ &\quad - v_F\beta D(H-h^*)\partial h^*/\partial W - r_W. \qquad (10)\end{aligned}$$

In the first line note that $h_P = \Delta r/\Delta v$ while in the second line $\phi W = (1-\beta)D(H-h^*)$ from (4), and so simplifies to $v_F\beta D(H-h^*)\partial h^*/\partial W$. The last line gives the surplus from 1 MW of extra wind, which can be compared to the market revenue below.

4.2. Decentralizing the efficient solution: charging for inertia

As before, efficient prices, $p(h)$, are equal to the System Marginal Cost (if not load shedding) or the VoLL (when shedding load). For $0 \le h \le L$, $p(h) = V$, for $L < h \le h_P$, $p(h) = v_P$, for $h_P < h < H - h^*$, $p(h) = v_F$, and for $H - h^* \le h \le H$, $p(h) = 0$ (the avoidable cost of wind). As in Section 3, free entry with consistent choices of V, L ensures conventional plant just covers its costs. The expected market unit net surplus (per MW of wind) (as $E\tilde{\phi} = \phi$) will be

$$\begin{aligned}M_W &= \phi\{VL + v_P(h_P - L) + (H - h^* - h_P)v_F\} - r_W,\\ &= \phi\{r_P + h_P(v_P - v_F) + (H-h^*)v_F\} - r_W,\\ &= \phi\{r_F + (H-h^*)v_F\} - r_W.\end{aligned}$$

If it is left to wind producers to decide whether or not to enter,[18] then efficient entry requires that marginal surplus/MW, $\partial S/\partial W$, of Eq. (10) is equal to the expected net market surplus/MW. Normally one might expect that if all externalities (emissions pricing, learning spill-overs, congestion and balancing via nodal pricing) are internalized, then the efficient equilibrium ought to be supported in a competitive market, but that is not the case here without recognizing the cost of inertia. The annualized value of this cost including any excess capacity credit is τ/MWyear (if negative, a subsidy) to restore equality and hence efficient entry, with $\tau = M_W - \partial S/\partial W$:

$$\begin{aligned}\tau(W) &= \phi\{r_F + (H-h^*)v_F\} - r_W - (\delta V - \phi v_P)L\\ &\quad - \phi\{r_F - r_P + v_F(H-h^*)\}\\ &\quad + v_F\beta D(H-h^*)\partial h^*/\partial W + r_W,\\ &= (\phi - \delta)VL + v_F\beta D(H-h^*)\partial h^*/\partial W > 0, \qquad (11)\end{aligned}$$

substituting for $\phi r_P = \phi(V - v_P)L$ in the top line. If there is no need to curtail wind, $h^* = \partial h^*/\partial W = 0$, and (11) can be interpreted as a method of de-rating wind to achieve efficient entry, $\delta = \phi$, consistent with the claim that in the absence of any correlation of wind with demand, wind should be de-rated by its average capacity factor, ϕ (Zachary et al., 2019). Allowing for such correlations gives a different result (see (Newbery, 2020)).[19]

If $\partial h^*/\partial W > 0$ then there is an additional cost proportional to the cost, v_F, of the volume β of inertial services required for the additional number of hours, $\partial h^*/\partial W$ caused by the addition of wind capacity. It is difficult to imagine how this can be charged as an ancillary inertia service charge in hours of surplus wind, and in any case the primary signal needed is not short-run dispatch (that is given by the efficient wholesale price in surplus hours of zero) but the entry signal, the time when the capacity value is typically determined for new entrants in the capacity auction.

It may be possible to devise a suitable contract to address the inertia externality. Periodic auctions for VRE could clear at a strike price that would be payable on non-curtailed day-ahead forecast output for a fixed number of forecast full-output hours (e.g. 30,000 MWh/MW). Curtailment would be determined on a last-in, first to be curtailed rule, and curtailed hours would not count to the full-output contacted hours. Later entrants would face increasing delays in earning their strike-priced output, raising the cost of entry.

It would appear that if the inertia constraint is relaxed, in the extreme in which $\beta = 0$, and wind is suitably de-rated as above, then $\tau = 0$, and there is no distortion. This is misleading, as there are other system stability constraints, which might be most simply addressed by assuming the need for a minimum level of flexible capacity, m, able to almost instantly respond to sudden demand–supply imbalances (e.g. caused by the loss of the largest generation unit or transmission link, the N-1 constraint). At some point $Min\phi(t)W - (1-\beta)D(t) < m$, and at that point the inertia constraint, the last term in Eq. (9), will be replaced by the cost of running the required reserves that displace (curtail) wind: $mv_F(H - h^*)$.[20] Its derivative is $-mv_F\partial h^*/\partial W$ and so Eq. (11) can be replaced by

$$\tau(W) = (\phi - \delta)VL + v_F m\partial h^*/\partial W > 0. \qquad (12)$$

This is consistent with (11), as the point at which the inertia constraint is replaced by the reserve constraint, $m = D(H - h^*)$.

4.3. Numerical estimates

The inertia charge is best measured as a percent of the annual fixed cost, τ/r_W:

$$\tau(W)/r_W = (v_F/r_W)\beta D(H-h^*)\partial h^*/\partial W.$$

For the illustrative example of Fig. 3, $\beta = 25\%$, $\partial h^*/\partial W = 0.108$ and $\beta D(H - h^*) = 8812$ h. Table 1, taken from Newbery (2020, Table 2) for the Single Electricity Market (SEM) of the island of Ireland, gives values for v_F (using 2019 prices) and r_F, giving $\tau/r_F = 48\%$, high because it ignored important relevant features of storage and export to avoid curtailment. More soundly based data from Newbery (2020) for the SEM considered two cases for β (25% and in the ambitious case, 15%) and for the export and storage opportunities in 2026. The costs are converted (at 2018 exchange rates of €1.13 = £1) to € and shown in Table 1. The projected median gas price ws then €21.4/MWh (FES, 2019) while the CO_2 price was taken as €40 /tonne. The corrective charge in the first case ($\beta = 25\%$) is $\tau/r_W = 20\%$ and in the ambitious case just under 10% of annual fixed costs.

Table 1 Cost estimates[21]

[18] Greve and Rocha (2020, p91) note that a 2019 Dutch off-shore wind tender "introduced a no subsidy requirement."

[19] If reliability is interpreted as no black-outs in 90% of years, then it would be more appropriate to look at the distribution of wind output in stress hours, and take derating as the lower 10th percentile. See Wilton et al. (2014) and for the importance of rare events, Wolak (2022).

[20] If such plant has a different variable cost then v_F can be replaced by the appropriate value.

[21] BEIS (2020) gives 2025 (medium) capital costs for base and peaking plant (open-cycle gas turbine) and on-shore wind, as well as the fixed and variable operating costs and fuel efficiencies. The capital cost figures for base and peaking are per *derated* MW, and so the cost per installed MW needs to be inflated to allow for this.

Table 1
Cost estimates.[a]

r_F	€85,218/MWyr	v_F	€61/MWh
r_P	€37,012/MWyr	v_P	€91/MWh
r_W	€120,132/MWyr	v_W	€7/MWh
Δr	€48,206/MWyr	Δv	€30/MWh
h_P	1607 h	L	8 hours

[a] BEIS (2020) gives 2025 (medium) capital costs for base and peaking plant (open-cycle gas turbine) and on-shore wind, as well as the fixed and variable operating costs and fuel efficiencies. The capital cost figures for base and peaking are per *derated* MW, and so the cost per installed MW needs to be inflated to allow for this.

4.4. Learning externalities

Learning externalities were discussed above in 2.2 and 2.3 above and arise when each installation reduces the cost of successive installations. As such it is a club good in which members of the club (the beneficiaries of the future cost reductions) can collectively agree to share the burden of the current higher cost investments. The EU achieves this by agreeing national targets for the volume of renewable energy each Member State should deliver. The assumption above was that the learning spillovers were already corrected, but the empirical figures for the annualized capital costs were not so corrected. Newbery (2018a) shows how to calculate the globally desirable level of subsidy and Newbery (2020) derives the values, with a central estimate for the SEM in 2026 of 10% of the capital cost. This is comparable to, and offsets, the ambitious scenario corrective charge and therefore roughly cancels it out. Taking the uncorrected IRENA (2019) learning rate estimates at face value, the learning subsidy might be 16% of the capital cost, again, not far short of the corrective charge in the base case, at least for the SEM.

5. Conclusions

Once a wind turbine is commissioned and connected, it will generate so long as it is not constrained or off-line. Some constraints are local, caused by transmission limits, and are best handled either by nodal pricing, or in its absence by offering non-firm connections in such locations (with the option of paying a fair share of any grid reinforcement costs needed to provide firm access to the rest of the system, as described by the LCNF project *Plug and Play*).[22] The constraints considered here are system wide, and need a system-wide solution. The first part of good system design is to ensure that carbon costs are properly charged, innovative technologies are compensated for their external learning benefits, and electricity pricing into the grid reflects the social marginal cost of generation, cleansed of distortionary subsidies (except where, as a second best, carbon taxes are below their social cost, and zero carbon generation can be compensated per MWh for the underpricing of any carbon displaced).

The remaining element of good market design is to ensure the efficient entry (and type) of new generation. With an efficient energy-only market, or suitably auctioned capacity payments, fossil entry can be left to market signals. The capacity credit for wind is rather more complicated to calculate and very sensitive to demand and wind conditions in winter months, as well as reliability and resilience requirements. The tentative conclusion from rough calculations (in the fuller model set out in (Newbery, 2020)) is that its capacity credit seems lower than those currently used. The key new factor considered here is that once wind penetration is high enough to cause system-wide curtailment, additional wind imposes an additional cost that is not reflected in market prices, as the marginal curtailment is many times higher than the average curtailment that sets prices. This can be viewed as a "tragedy of the commons" — freeloading on the inertial services that are caused by the marginal entrant but experienced by all VRE. These extra costs reflect the cost of the share of dispatchable generation, β, that is needed to supply the inertia, and is simplest to consider as an annual charge to levy at the time of new entry of VRE. The two cases considered above give rise to material annual charges of 10%–20% of annual fixed costs, roughly proportional to β. Offsetting this corrective charge, the global learning externality (mostly reaped abroad, but internalized if other countries offer similar subsidies as a club payment, e.g. under the EU *Clean Energy Package*) might be 11%–17% of annual fixed costs and therefore of comparable magnitude.

The conclusion is that the capacity credit offered to VRE may differ (and be lower than) the capacity credit used to determine the amount of de-rated capacity needed to meet the reliability standard. Most other ancillary or system services can be handled through markets, auctions, and proper network charging contracts. The curtailment effect or inertia charge will depend very much on system characteristics (penetration and SNSP most directly) and in the SEM is comparable to the likely justified global learning subsidy. Whether this would be true in other systems or with higher VRE penetration should be explored as part of wider study of the appropriate way to support VRE, and the extent to and manner in which to grant capacity payments to wind. The simpler alternative is to set the renewables target and run auctions for the amount of renewables by allowing them to bid for the strike price in a Contract for Difference (CfD) for the first 30,000 full operating hours (i.e. MWh/MW), which would provide a revenue stream for about 12–15 years, perhaps with the variant of not paying for nor counting the plant's constrained hours. The proper design of the CfD to avoid other distortions has been set out in Newbery (2023a), and includes an efficient long-term contract for network use. Savelli et al. (2022) proposed another form of CfD to correct for locational differences, but do not include any technological differentiation for the inertial charge, which will differ as between wind and PV, and their correlation with system curtailment.

CRediT authorship contribution statement

David M. Newbery: Conceptualization, Methodology, Modelling, Writing.

Appendix. Linearizing curtailment functions

The curtailment function in Fig. 3 is roughly linear and can be approximated by (now replacing y by h for convenience)

$$k = A(1 - h/h_r^*) + \alpha(W - W_r),$$

where subscript r refers to a reference level of wind, W_r. If this holds over a wide enough range then there will be no curtailment until $W = W_0$, in which case $A = \alpha(W_r - W_0)$, and

$$\begin{aligned} h^* &= h_r^* \frac{W - W_0}{W_r - W_0}, \\ k &= \alpha(W - W_0)(1 - h/h^*), \\ \frac{\partial h^*}{\partial W} &= \frac{h_r^*}{W_r - W_0} = \frac{h^*}{W - W_0}. \end{aligned}$$

It follows that

$$\begin{aligned} \int_0^{h*} k dh &= \frac{1}{2}\alpha(W - W_0)h^*, \\ \int_0^{h*} \frac{\partial k}{\partial W} dh &= \alpha \int_0^{h*} (1 - h/h^*)dh + \alpha(W - W_0) \int_0^{h*} \frac{h}{h^{*2}} \frac{\partial h^*}{\partial W}, \\ &= \alpha h^*. \end{aligned}$$

[22] at https://www.ofgem.gov.uk/sites/default/files/docs/2015/05/fpp_sdrc_reward_application_v2.0_pxm_2015-05-01_final_0.pdf

The ratio of the marginal to the average curtailment is then

$$\frac{W\int_0^{h*}\frac{\partial k}{\partial W}dh}{\int_0^{h*}kdh}=\frac{2W}{W-W_0}>2, \tag{13}$$

which will be high if $W-W_0$ is small, and tend to 2 as $W\gg W_0$.

References

Batlle, C., Vazquez, C., Rivier, M., Perez-Arriaga, I.J., 2007. Enhancing power supply adequacy in Spain: Migrating from capacity payments to reliability options. Energy Policy 35 (9), 4545–4554. http://dx.doi.org/10.1016/j.enpol.2007.04.002.

BEIS, 2020. Electricity generation costs. at https://www.gov.uk/government/publications/beis-electricity-generation-costs-2020.

Bothwell, C., Hobbs, B., 2017. Crediting wind and solar renewables in electricity capacity markets: The effects of alternative definitions upon market efficiency. Energy J. 38 (01), http://dx.doi.org/10.5547/01956574.38.SI1.cbot.

Buchanan, J.M., 1965. An economic theory of clubs. Economica 32 (125), 1–14, At https://www.jstor.org/stable/2552442.

Chyong, C.K., Guo, B., Newbery, D., 2020. The impact of a Carbon Tax on the CO_2 emissions reduction of wind. Energy J. 41 (1), 1–31. http://dx.doi.org/10.5547/01956574.41.1.cchy.

Clò, S., Cataldi, P., Zoppoli, A., 2015. The merit-order effect in the Italian power market: The impact of solar and wind generation on national wholesale electricity prices. Energy Policy 77, 79–88. http://dx.doi.org/10.1016/j.enpol.2014.11.038.

Cludius, J., Hermann, H., Matthes, F. Chr., Graichen, V., 2014. The merit order effect of wind and photovoltaic electricity generation in Germany 2008–2016: Estimation and distributional implications. Energy Econ. 44, 302–313. http://dx.doi.org/10.1016/j.eneco.2014.04.020.

Committee on Climate Change, 2019. Technical Annex: Integrating variable renewables into the UK electricity system. at https://www.theccc.org.uk/wp-content/uploads/2019/04/Technical-Annex-Integrating-variable-renewables-into-the-UK-electricity-system.pdf.

Csereklyei, Z., Qu, S., Ancev, T., 2019. The effect of wind and solar power generation on wholesale electricity prices in Australia. Energy Policy 131, 358–369. http://dx.doi.org/10.1016/j.enpol.2019.04.007.

Deane, P.S. Collins, Gallachóir, B.Ó., Eid, C., Hartel, R., Keles, D., Fichtner, W., 2017. Impact on electricity markets: Merit order effect of renewable energies. In: Ch 16 in Europe's Energy Transition - Insights for Policy Making. Academic Press, http://dx.doi.org/10.1016/B978-0-12-809806-6.00016-X.

DUKES, 2022. Digest of UK Energy Statistics. BEIS, at https://www.gov.uk/government/collections/digest-of-uk-energy-statistics-dukes.

EERA, 2022. Joint Program on Energy Systems Integration Expert Workshop Towards a Common Understanding of Energy System Costs. Leuven, (Belgium), pp. 19–20.

Energy Systems Catapult, 2019. Towards a new framework for electricity markets. at https://es.catapult.org.uk/reports/towards-a-new-framework-for-electricity-markets/. (Accessed 9 January 2023).

ESIG, 2021. Refining resource adequacy for modern power systems. at https://www.esig.energy/reports-briefs.

Fernández, L., 2023. Global Offshore Wind Capacity Factors. pp. 2010–2021, at https://www.statista.com/statistics/1368679/global-offshore-wind-capacity-factor/.

FES, 2019. National grid ESO future energy scenarios 2019 data workbook. at https://www.nationalgrideso.com/future-energy/future-energy-scenarios/fes-2019-documents.

Flottmann, J., Akimov, A., Simshauser, P., 2022. Firming merchant renewable generators in Australia's National Electricity Market. Econ. Anal. Policy 74, http://dx.doi.org/10.1016/j.eap.2022.02.005.

Gambhir, A., Green, R., Grubb, M., Heptonstall, P., Wilson, C., R., 2021. How are future energy technology costs estimated? Can we do better? Int. Rev. Environ. Resour. Econ. 15 (4), 271–318. http://dx.doi.org/10.1561/101.00000128.

Gohdes, N., Simshauser, P., Wilson, C., 2022. Renewable entry costs, project finance and the role of revenue quality in Australia's National Electricity Market. Energy Econ. 114, 106312. http://dx.doi.org/10.1016/j.eneco.2022.106312.

Green, R.J., Léautier, T.-O., 2015. Do costs fall faster than revenues? In: Dynamics of Renewables Entry into Electricity Markets. Toulouse School of Economics TSE-591 at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.697.4750&rep=rep1&type=pdf.

Green, R., Vasilakos, N., 2010. Market behaviour with large amounts of intermittent generation. Energy Policy 38, 3211–3220, https://www.sciencedirect.com/science/article/pii/S0301421509005461.

Greve, T., Rocha, M., 2020. Policy and theoretical implications of zero-subsidy bids in the German offshore wind tenders. Energy J. 41 (4), 89–104. http://dx.doi.org/10.5547/01956574.41.4.tgre.

Grubb, M., Newbery, D., 2018. UK electricity market reform and the energy transition: Emerging lessons. Energy J. 39 (6), 1–24. http://dx.doi.org/10.5547/01956574.39.6.mgru.

Hardin, G., 1968. The tragedy of the commons. Science 162 (3859), 1243–1248. http://dx.doi.org/10.1126/science.162.3859.1243.

Henriot, A., 2015. Economic curtailment of intermittent renewable energy sources. Energy Econ. 49, 370–379. http://dx.doi.org/10.1016/j.eneco.2015.03.002.

Heptonstall, P.J., Gross, R.J.K., 2020. A systematic review of the costs and impacts of integrating variable renewables into power grids. Nat. Energy http://dx.doi.org/10.1038/s41560-020-00695-4.

Hirth, L., Ueckerdt, F., Edenhofer, O., 2015. Integration costs revisited – An economic framework for wind and solar variability. Renew. Energy 74, 925–939. http://dx.doi.org/10.1016/j.renene.2014.08.065.

Holland, S.P., Mansur, E.T., Muller, N.Z., Yates, A.J., 2020. Decompositions and policy consequences of an extraordinary decline in air pollution from electricity generation. Am. Econ. J: Econ. Policy 12 (4), 244–274. http://dx.doi.org/10.1257/pol.20190390.

Holmberg, P., Ritz, R., 2020. Optimal capacity mechanisms for competitive electricity markets. 41, (S1), http://dx.doi.org/10.5547/01956574.42.S12.phol,

IEA, 2021. Conditions and Requirements for the Technical Feasibility of a Power System with a High Share of Renewables in France Towards 2050. Paris.

IRENA, 2019. Future of Wind: Deployment, Investment, Technology, Grid Integration and Socio-Economic Aspects (A Global Energy Transformation Paper). International Renewable Energy Agency, Abu Dhabi, At https://www.irena.org/wind.

Jansen, M., Staffell, I., Kitzing, L., et al., 2020. Offshore wind competitiveness in mature markets without subsidy. Nat. Energy 5, 614–622. http://dx.doi.org/10.1038/s41560-020-0661-2.

Joos, M., Staffell, I., 2018. Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany. Renew. Sustain. Energy Rev. 86, 45–65. http://dx.doi.org/10.1016/j.rser.2018.01.009.

Joskow, P., 2011. Comparing the costs of intermittent and dispatchable electricity generating technologies. Am. Econ. Rev 101, 238–241, https://www.aeaweb.org/articles?id=10.1257/aer.101.3.238.

Joskow, P.L., Tirole, J., 2007. Reliability and competitive electricity markets. Rand J. Econ. 38 (1), 68–84. http://dx.doi.org/10.1111/j.1756-2171.2007.tb00044.x.

Kaffine, D.T., McBee, B.J., Ericson, S.J., 2020. Intermittency and CO2 reductions from wind energy. Energy J. 41 (5), 49–80. http://dx.doi.org/10.5547/01956574.41.5.dkaf.

Keane, A., Milligan, M., Dent, C.J., Hasche, B., D'Annunzio, C., Dragoon, K., Holttinen, H., Samaan, N., Soder, L., O'Malley, M., 2011. Capacity value of wind power. IEEE Trans. Power Syst. 26 (2), 564–572. http://dx.doi.org/10.1109/TPWRS.2010.2062543.

Ketterer, J.C., 2014. The impact of wind power generation on the electricity price in Germany. Energy Econ. 44, 270–280. http://dx.doi.org/10.1016/j.eneco.2014.04.003.

Korpås, M., Botterud, A., 2020. Optimality Conditions and Cost Recovery in Electricity Markets with Variable Renewable Energy and Energy Storage. MIT CEEPR Working Paper 2020-005, At https://ceepr.mit.edu/workingpaper/optimality-conditions-and-cost-recovery-in-electricity-markets-with-variable-renewable-energy-and-energy-storage/.

Meade, J., 1972. The theory of labour-managed firms and of profit-sharing. Econom. J. 82, 402–428, https://www.jstor.org/stable/2229945.

Milligan, M., Ela, E., Hodge, B.-M., Kirby, B., Lew, D., Clark, C., DeCesaro, J., Lynn, K., 2011. Integration of variable generation, cost-causation, and integration costs. Electr. J. 24 (9), 51–63, https://www.sciencedirect.com/science/article/pii/S1040619011002636.

Mowers, M., Mai, T., 2021. A relative value framework: Why do different electricity system technologies have different economic value? Electr. J. 34, 1–9. http://dx.doi.org/10.1016/j.tej.2021.107007, 107007.

National Grid ESO, 2019. De-rating factor methodology for renewables participation in the capacity market. At https://www.emrdeliverybody.com/Prequalification/EMR%20DB%20Consultation%20response%20-%20De-rating%20Factor%20Methodology%20for%20Renewables%20Participation%20in%20the%20CM.pdf.

National Grid ESO, 2022. Future energy scenarios 2022 report. at https://www.nationalgrideso.com/future-energy/future-energy-scenarios.

Newbery, D.M., 2016a. Missing money and missing markets: Reliability, capacity auctions and interconnectors. Energy Policy 94, 401–410. http://dx.doi.org/10.1016/j.enpol.2015.10.028.

Newbery, D., 2016b. The long-run equilibrium impact of intermittent renewables on wholesale electricity prices. EPRG WP 1601 at https://www.eprg.group.cam.ac.uk/wp-content/uploads/2016/01/1601-Text.pdf.

Newbery, D.M., 2017. Tales of two islands - lessons for EU energy policy from electricity market reforms in britain and Ireland. Energy Policy 105, 597–607. http://dx.doi.org/10.1016/j.enpol.2016.10.015.

Newbery, D., 2018a. Evaluating the case for supporting renewable electricity. Energy Policy 120, 684–696. http://dx.doi.org/10.1016/j.enpol.2018.05.029.

Newbery, D.M., 2018b. Shifting demand and supply over time and space to manage intermittent generation: the economics of electrical storage. Energy Policy 113, 711–720. http://dx.doi.org/10.1016/jenpol.2017.11.044.

Newbery, D., 2020. Club goods and a tragedy of the commons: the Clean Energy Package and wind curtailment. EPRG WP 2036 at https://www.eprg.group.cam.ac.uk/eprg-working-paper-2036/.

Newbery, D., 2021. National Energy and Climate Plans for the island of Ireland: wind curtailment, interconnectors and storage. Energy Policy 158, 1–11. http://dx.doi.org/10.1016/j.enpol.2021.112513, 112513.

Newbery, D., 2023a. Designing efficient renewable electricity support schemes. Energy J. 44 (3), 1–22. http://dx.doi.org/10.5547/01956574.44.3.dnew.

Newbery, D., 2023b. Estimating the target-consistent carbon price for the GB electricity sector. EPRG WP 23XX.

Newbery, D., Pollitt, M.G., Ritz, R.A., Strielkowski, W., 2018. Market design for a high-renewables European electricity system. Renew. Sustain. Energy Rev. 91, 695–707. http://dx.doi.org/10.1016/j.rser.2018.04.025.

Peng, D., Poudineh, R., 2019. Electricity market design under increasing renewables penetration: Misalignments observed in the European Union. Util. Policy 61, 1–13. http://dx.doi.org/10.1016/j.jup.2019.100970.

Pudjianto, Danny, Aunedi, Marko, Djapic, Predrag, Strbac, Goran, 2014. Whole-systems assessment of the value of energy storage in low-carbon electricity systems. IEEE Transactions on Smart Grid 5 (2), 1098–1109. http://dx.doi.org/10.1109/TSG.2013.2282039.

Rubin, E.S., Azevedo, I.M.L., Jaramilloa, P., Yeh, S., 2015. A review of learning rates for electricity supply technologies. Energy Policy 86, 198–218. http://dx.doi.org/10.1016/j.enpol.2015.06.011.

Savelli, I., Hardy, J., Hepburn, C., Morstyn, T., 2022. Putting wind and solar in their place: Internalising congestion and other system-wide costs with enhanced contracts for difference in Great Britain. Energy Econ. 113, 106218. http://dx.doi.org/10.1016/j.eneco.2022.106218.

Söder, L., Tomasson, E., Estanqueiro, A., Flynn, D., Hodge, B., Kiviluoma, J., Korpås, M., Neau, E., Couto, A., Pudjianto, D., Strbac, G., Burke, D., Gómez, T., Das, K., Cutululis, N.A., Hertem, D. Van, Höschle, H., Matevosyan, J., von Roon, S., Carlini, E., Caprabianca, M., de Vries, L., 2020. Review of wind generation within adequacy calculations and capacity markets for different power systems. Renew. Sustain. Energy Rev. 119, 109540. http://dx.doi.org/10.1016/j.rser.2019.109540.

Steinhilber, S., 2016. Onshore Wind Concession Auctions in China: Instruments and Lessons Learnt. AURES Report D4.1-CN, March. http://aures2project.eu/wp-content/uploads/2021/07/pdf_china.pdf.

Stoft, S., 2002. Power System Economics: Designing Markets for Electricity. Wiley.

Ueckerdt, F., Hirth, L., Luderer, G., Edenhofer, O., 2013. System LCOE: What are the costs of variable renewables? Energy 63 (C), 61–75, At https://www.pik-potsdam.de/members/edenh/publications-1/SystemLCOE.pdf.

Weiss, T., Wänn, A., 2013. Does Ireland need more storage? In: Power System Overview and RES Integration. at https://www.store-project.eu/documents/target-country-results/en_GB/energy-storage-needs-in-ireland-executive-summary.

Wilton, E., Delarue, E., D'haeseleer, W., van Sark, W., 2014. Reconsidering the capacity credit of wind power: Application of cumulative prospect theory. Renew. Energy 68, 752–760. http://dx.doi.org/10.1016/j.renene.2014.02.051.

Wolak, F., 2022. Long-term resource adequacy in wholesale electricity markets with significant intermittent renewables. Environ. Energy Policy Econ. 3 (2022).

Wu, C., Zhang, XP., Sterling, M., 2022. Solar power generation intermittency and aggregation. Sci. Rep. 12, 1363. http://dx.doi.org/10.1038/s41598-022-05247-2.

Zachary, S., Wilson, A., Dent, C., 2019. The integration of variable generation and storage into electricity capacity markets. at https://arxiv.org/abs/1907.05973.

Zerrahn, A., Schill, W.P., 2017. Long-run power storage requirements for high shares of renewables: Review and a new model. Renew. Sustain. Energy Rev. 79, 1518–1534. http://dx.doi.org/10.1016/j.rser.2016.11.098.