1. Background

In a recent paper we investigate the problem of incentivising flexibility in electricity markets. As the share of intermittent renewable energy increases in the generation mix, power systems are exposed to greater levels of uncertainty and risk. This requires planners, policy makers and industry to incentivise flexibility in these systems, that is: their adaptability to unforeseen variations in generation and demand. The growing need for flexibility, along with the fact that it is costly to provide, highlights the importance of efficient procurement.

A distinctive feature of power systems is that they require instantaneous equilibrium between supply and demand. Traditionally, utilities have operated with fairly predictable and mature technologies. To deal with the challenges of uncertainty and variability, which aren’t new, a stock of balancing services and reserves have been available to system operators to ensure that the system remains in balance at all times. However, as decarbonisation climbs up the policy agenda and renewable generation becomes more prevalent in power systems throughout the world, increased uncertainty and variability represent greater challenges for system operation. With substantial shares of renewables, the system operator’s problem is to predict fluctuations in the net load, which is the difference between total demand (load) and variable generation. In other words, it is the demand that must be met by other sources if all renewable generation is utilised. This magnitude is harder to predict accurately – i.e. contains greater uncertainty – as it depends on two random variables, namely demand and renewable generation.

As a commodity, flexibility has multiple attributes such as capacity, ramp rate, duration and lead time among which there are complementarities. Furthermore, users of flexibility differ in the way in which they combine the flexibility attributes, creating heterogeneity. Additionally, along with traditional sources, which already enable flexibility, a number of business models, such as thermostat-based demand response, aggregators and small storage providers, are emerging in electricity markets and are expected to constitute important sources of flexibility in future decentralised power systems. However, due to high transaction costs relative to the size of resource, the emerging small resources cannot directly participate in an organised electricity market and compete. Therefore, we ask the fundamental question of how should the provision of flexibility, as a multi-dimensional and heterogeneous commodity, be incentivised in this context? A topic which despite its importance has largely remained unaddressed by the economic literature.

2. Flexibility in the power system

With a greater reliance on variable renewable energy (VRE) and its projected increase in years to come, stakeholders, electricity industry analysts and academics have become interested in the question of efficiently integrating these sources of generation into the grid. Operating systems with substantial shares of VRE increases the uncertainty and variability of power systems, particularly in the short term, which requires operators, business, and policy makers to incentivize the provision of flexibility, if higher levels of VRE penetration are to be achieved.

Traditionally, system operators have focused on forecasting system demand (load), which varies randomly. In contrast, with large shares of VRE, it is sensible for operators to focus on the net load - electricity demand minus VRE generation - which represents the demand that must be met by non-renewable sources.

1 The term “sources” is employed here in its widest possible sense: it could refer to generation, conventional or not, but it could also involve any change in demand that helps to keep the system in balance.
The impact of uncertainty can be readily perceived in the net load, given that it depends on not one but two random variables (i.e. load and generation), reducing the accuracy of forecasts. The effect of variability, on the other hand, becomes evident in the shorter peaks, steeper ramps and lower turn-downs required from the non-renewable sources of generation. In consequence, operators require flexibility-enabling assets (resources) that can adapt to these patterns.

But because “sometimes examples of inflexibility are easier to document than flexibility” (Cochran et al., 2014), it is interesting to describe what happens when flexibility is unavailable. On the technical side, the following impacts may be perceived:

1. **Difficulty balancing demand and supply** which result in frequency excursions and dropped load.\(^2\)
2. **Significant VRE curtailments**, as a result of excess supply or transmission constraints
3. **Area balance violations** which reflects that a system cannot meet its balancing responsibility

In relation to markets, the following may happen:

(i) **Negative market prices** which might reflect that conventional generators are unable to reduce output, demand that cannot absorb excess supply, surplus of renewable energy or limited transmission capacity

(ii) **Price volatility** which can reflect insufficient transmission capacity, limited ramping availability, insufficiently fast response or limited ability to reduce demand

Illustrating the need for power system flexibility can be understood better by way of examples. The first of these comes from the Danish power system. In the early hours of the 10th of July, 2015, wind blew in Denmark so strongly that 140% of domestic electricity demand was met by the output from windfarms. The excess power generated was exported through the interconnections with Sweden, Norway and Germany.\(^3\)

Figure 1 illustrates the evolution of the Danish load, wind power output, net load (calculated as load minus wind power output) and the net exchange through all interconnections with neighbouring Germany, Sweden and Norway for both Danish bidding areas (DK1, i.e. western Denmark and DK2, i.e. eastern Denmark), over the course of four days (08/07/2015 – 11/07/2015). The typical short-term features of systems operating with large shares of renewables become evident in the figure.

Ideally, to make good use of VRE, generation and demand would have a positive correlation, that is: demand is high when renewables are available or, conversely, demand is low when renewables become scarce. However, a consequence of this situation is that non-renewable, dispatchable generators face shorter peaks and, consequently, are compensated for less operating hours, which adversely affects their cost recovery. As an example of this situation, EURELECTRIC (2011) cites the Spanish case – where renewables are dispatched first and at zero variable cost – which has led to plummeting utilisation rates for thermal units. Comparing utilisation rates between 2005-2007 with the 2009-2010 period, the publication reports a 47% decrease for coal-fired power plants and a 30% decrease for combined cycle gas turbines (CCGT) plants.

\(^2\) “The two fundamental characteristics of power delivered to a customer are frequency and voltage. As long as these remain correct the customer will have access to the needed power, and it will have the required characteristics” (Stoft, 2002).

\(^3\) Many media outlets rejoiced with the record figure, as it proved that relying entirely on VRE was indeed feasible if there is sufficient flexibility in the system. For example, see the article in The Guardian: http://www.theguardian.com/environment/2015/jul/10/denmark-wind-windfarm-power-exceed-electricity-demand
Figure 1: Net load (load minus wind power generation) and net power exchange in the Danish power system 08/07/2015 – 11/07/2015

Source: Boscán (2016a)

Back to the Danish example, as can be observed in figure 1, during the daytime hours of Wednesday 8th, Thursday 9th and Friday 10th, the net load showed (relative to load) shorter peaks which coincided with high levels of demand and generation of wind energy. For example, between 10:00 and 11:00 on the 9th, net load peaked when 84% of total demand was being satisfied by wind generation. On the same day, net load peaked again between 17:00 and 18:00 when 90% of load was being met by wind.

However, VRE supply and load are not always positively correlated. In an empirical study of the Nordic countries, Holttinen (2005) finds that wind power output and load were slightly positively (31%) correlated in the period 2000-2002, but when the sample is restricted to the winter months, correlation is very close to zero.

In contrast to positive correlation between generation and demand, whenever these two variables are negatively – and thus unfavourably – correlated, two different ramping effects on the remaining generation base are induced. If renewable supply decreases at the same time as increased demand, system operators must dispatch generation that is able to ramp up quickly. In figure 1, this can be readily observed from the afternoon hours of Friday 10th onwards, when wind generation began to decrease at the same time that load increased. Specifically, between 17:00 and 22:00 on that date, net load exhibited a steep ramp rate at the same time that load declined and this happened because wind power output was falling faster than load.

In contrast, if renewable supply is high when demand is low, dispatchable generators face deeper turn downs as they must give way to renewables. In figure 1, this becomes evident in the early hours of Wednesday 8th when net load declined considerably more steeply than load, as wind power
generation increased. In fact, the achievement of the 140% record happened as a coincidence of low demand (during early morning hours), high wind power output and the existence of interconnections with neighbouring countries, which explains the negative values for the net load. Instead of curtailing generation from wind turbines – a sign of inflexibility – power was exported. Specifically, in relation to Denmark and its approach to managing the intermittency of wind power, Green and Vasilakos (2011) find that on windy days, Denmark uses exports as a kind of electricity storage. They state that “…short-term fluctuations in wind output are highly correlated with short-term fluctuations in net exports of electricity, which is exactly the efficient pattern of operation dictated by (their) model”, net load and net exchanges are almost perfectly correlated (98%) in figure 1. Beyond the lesson learnt about Danish system operation with high shares of wind power their finding illustrates, more generally, the role of interconnections as an asset to enable power system flexibility.

A second figure (see figure 2) completes the illustration of the Danish power system during the record wind power output days in July 2015. In it, the net loads of each Danish bidding area (which differs from the aggregated Danish net load shown in figure 1) are shown together with the day-ahead (Elspot) system price and Danish price (note that prices coincided in both Danish bidding areas). Observe (red dot) the price that cleared the day-ahead market when the record output was achieved. It is easy to see that western (DK1) net load exhibits greater variability than its eastern (DK2) counterpart. Steeper ramps and lower turn-downs (including negative values, associated to exports) in the west than in the east can be explained by the greater concentration of windfarms in this area of the country and to differences in weather conditions.

Regarding prices, it is interesting to note that, relative to the Nordpool system price, prices in the two DK bidding areas were always slightly lower in the period 08/07/2015-10/07/2015, coinciding with the build-up of wind power output. Sufficient capacity in the interconnections allowed exporting without curtailing when there was excess supply. However, as soon as wind power output decreased and net load (particularly in DK1) exhibited a steep ramp, the DK price increased. On average, 92% of load was covered by wind power on the 10th of July, whereas only 8% was covered on the 11th. The lower power output had to be compensated with imports on the 11th: on average, 62% of the load was covered by trade with neighbouring areas on that day. However, the remainder had to be covered with ramping capability. The fact that the price average between the 10th and the 11th spiked so markedly4 shows insufficient flexibility. Although it is difficult to know exactly what happened with the data considered in this example, it is likely that relatively inefficient dispatchable generators had to ramp up.

4 The average DK price quadrupled, from 5.41 EUR/MWh on 10th of July to 21.83 EUR/MWh on 11th of July
Figure 2: Net load (load minus wind power generation) in the DK1 (west), DK2 (east) bidding areas together with system and DK Elspot (day-ahead) prices 08/07/2015 – 11/07/2015

Another example comes from power system operation in California, the most populous state in the US, which has one of the most ambitious energy and environmental goals in the whole country, including a 50% of generation from renewable sources by 2030. Consequently – in an analysis that included every single day of operation between 2012 and 2020 – the California Independent System Operator (CAISO, 2016) has identified a number of prospective operational challenges, including:

1. **Short, steep ramps** in both upward and downward directions, which requires either dispatching or shutting down generation resources quickly and for short periods of time.

2. **Risk of oversupply** when excess generation, including renewables, exceeds real-time demand.

3. **Reduced frequency response** when less operating resources are available to automatically adjust electricity output for grid reliability purposes.

The first two elements of the previous list can be observed in figures 3 and 4, known in energy circles as “duck curves”. In the first of these, actual and projected net load for the 11th of January (i.e. a winter day) is shown. Note three ramps: the first (known as the duck’s tail, estimated by CAISO at 8000 MW) happens during the early morning hours, builds up from around 4 AM and lasts until 7 AM.

---

5 According to the Energy Information Administration (EIA): “In 2014, California ranked fourth in the nation in conventional hydroelectric generation, second in net electricity generation from other renewable energy resources, and first as a producer of electricity from both solar energy and geothermal energy” [http://www.eia.gov/state/?sid=CA](http://www.eia.gov/state/?sid=CA)

---

The contents of this paper are the author’s sole responsibility. They do not necessarily represent the views of the Oxford Institute for Energy Studies or any of its Members.
sunrise time. The second (known as the duck’s belly) is downwards and reflects the increasing renewable supply (particularly solar) displacing conventional generation. At around 4 PM, sunset time, an upward ramp (known as the duck’s neck, estimated at around 11000 MW) appears again as demand increases and the output from solar PVs disappears in the night hours.

On spring days, as shown in figure 4, the risk of oversupply heightens as the sun rises earlier and sets later, inducing a more pronounced “duck belly”, where the afternoon ramp is estimated at 13000 MW in approximately three hours. Consequently, (CAISO, 2016) requires flexible resources that are able to:

1. Sustain an upward/downward ramp,
2. Respond for a defined period of time,
3. Change ramp directions quickly,
4. Store energy or modify use,
5. React quickly and meet expected operating levels,
6. Start with short notice from a zero or low-electricity operating level,
7. Start and stop multiple times per day, and
8. Accurately forecast operating capability

Figure 3: California’s “Duck curve” on a winter day (net load on January 11th 2012-2020)

To summarise: with high shares of VRE, system operators require flexibility-enabling resources in order to increase the flexibility in the short-term operations of power systems. It is worth noting that the economic properties of flexibility refer both to the flexibility-enabling assets and to the agents who own the assets. However, when it comes to incentivising the provision of flexibility, there are significant distinctions between both. When considering supply-side resources like generation, owners of power plants have technical constraints to supply flexibility but will typically behave in an economically rational way, i.e. as profit maximisers. Owners of flexibility-enabling assets participating in, say, a demand response program, can be assumed to be relatively uniform regarding the assets they possess (e.g., air conditioners), but their cost of provision (disutilities) should not be straightforwardly assumed to be uniform. The latter are consumers and may react to behavioural elements beyond utility maximisation.
3. Trading flexibility services

Although flexibility has always been needed and indeed used in the operation of power systems, it has rarely been traded as a distinct commodity. Usually, conventional generators have been able to provide the flexibility that the system requires and have been compensated on the basis of its energy and/or capacity component. However, as the requirement for flexibility increases in power systems, there is a need for designing products that serve specific technical requirements. These can be bought and sold, along with other components of electricity services such as energy and capacity.\(^6\)

Mainly, final users of flexibility are Transmission System Operators (TSOs) or Independent System Operators (ISOs) -- who are responsible for balancing the high-voltage grid -- Distribution System Operators (DSOs) -- who are responsible for the reliable operation of distribution networks and quality of service -- together with other market players who may use flexibility to meet their energy and balancing obligations. Access to flexible resources to manage network constraints leads to an effective integration of distributed energy resources, allows network companies to optimise their network reinforcement capital investments, and improve the reliability and quality of service. It also enables other market players to optimise their energy portfolio in order to meet their energy market and balancing obligations at minimum cost, for example, arbitrating between generation and demand response (Boscán and Poudineh, 2016).

Depending on the specific situation, the incentive to supply flexibility can be provided through bilateral contracts or auctions specifically designed to account for its economic properties. We focus on the latter method, undimining the possibility of introducing competition. Indeed, bilateral procurement contracts without competition among sellers exist across the supply chain of flexibility services for various reasons, including the presence of transaction costs or insufficient prospects of achieving market thickness.

For example, due to high transaction costs relative to the size of the resource, small providers, such as households, cannot participate directly in an organised market and compete against each other. In such situations, aggregation provides an opportunity for small generation and demand resources to offer their flexibility in existing electricity markets (Eurelectric, 2014). According to this emerging business model, retailers or third parties act as intermediaries between providers and buyers of flexibility. This situation creates an environment in which “efficient” bilateral contracts are the most sensible method of procurement.

The role of aggregation is expected to rise in the transition towards a greater reliance on renewables. Currently, in many European countries intermittent renewables are being treated as conventional generation in the sense that they have the same obligations for their imbalance position (limited to full balance responsibility depending on the country)\(^7\) and entitlement to participate in balancing market as any other source of generation. This has the effect of encouraging improved forecasting on the bidders’ side and the entry of competitive aggregators to minimise balancing risks and offer ancillary services from renewable resources. Furthermore, this provides incentives for renewables to be firmed up by, for example, entering into separate contracts with owners of flexible resources such as residential demand response and storage facilities.

The Nest Learning Thermostat is another good real-world example of how bilateral contracts for flexibility services can be utilised in the integration of renewable resources. Nest, as the manufacturer of smart thermostat technology, partners with utility companies to provide a residential demand response programmes. Under the so-called “Rush hour scheme”, an attribute of the contracted consumer’s consumption (e.g. air conditioner temperature) is adjusted automatically by a utility company to manage fluctuations of demand and supply. Consumers are offered a menu of contracts

\(^6\) Note that we refer to “products”, in plural because flexibility services can take different forms.

\(^7\) For example, in Denmark, Finland, Estonia, Netherlands, Spain and Sweden renewable resources have full balancing responsibility (EC, 2013).
with different lead times (e.g. from on-demand to 24-hour notice in advance), duration of adjustments in consumption (e.g. 30 minutes to 4 hours) and payments.

The contract design problem arises because there are different classes of suppliers. Specifically, in relation to the thermostat example, different consumers experience different disutilities for the various dimensions of flexibility they provide. Such disutilities are privately held pieces of information by each resource provider. For example, one household may incur a high disutility for the short lead time and another household for long duration of load control. Logically, the former household prefers a contract with higher lead time but can sacrifice on load control duration, whereas the latter values more a contract with shorter load control duration. Therefore, the contracts should be (but currently are not necessarily) designed in a way that each participating agent truthfully self-selects its own contract, given the presence of multidimensional information asymmetry between the buyer and sellers. The bilateral contract model of this paper focuses on this category of contractual settings in electricity markets.

A key message of the paper is that flexibility is an inherently multi-dimensional and heterogeneous commodity as both the cost of producing it and the utility that is derived from utilising it depends on more than one factor. Capacity, ramp rate, duration, and lead time are among the many elements that describe flexibility. Because of this, different flexibility-enabling resources possess differing levels of efficiency, implying that flexibility is not a homogenous commodity. Additionally, an equally relevant economic property of power system flexibility is that its composing elements (e.g., capacity, ramp rate, duration) are best understood as imperfect complements: the cost of production and the utility derived from it are always non-separable. Not only do buyers and sellers of flexibility value it according to the different elements that compose it, but the overall valuation cannot be additively decomposed. In other words, the elements that compose flexibility are always valued simultaneously.

Therefore, as flexibility differs significantly from commodities traditionally traded in existing electricity markets, such as energy or capacity, correctly accounting for its economic properties is essential to create the incentives to enable it in electricity systems.

For example, consider the design of markets for flexibility services, a topic that has remained largely unaddressed in the existing literature of power system economics. For example, if competition among flexibility providers is possible, a multi-attribute auction may be useful to procure flexibility in an efficient manner. The multi-attribute auction is an allocation mechanism in which more than one feature of the commodity is valued (e.g., MW, MW/min and emission performance). Therefore, it allows the principal to incentivise, for example, capacity, flexibility and emission performance simultaneously in a single auction. In situations where competition is not feasible, multi-dimensional bilateral contracts are an alternative method of procurement, which ensures economic efficiency.

4. Conclusions

In the analysis of bilateral contracts for power system flexibility, we found inspiration in existing thermostat-based demand response programmes in which utilities incentivise their customers to modify their consumption during peak hours or when the system reliability is at stake. Relying on automation, customers allow utilities to automatically reduce their air conditioning (during summer) or electric heating (during winter) consumption in exchange for payments. Some companies pay customers for each season in which the customer enrols or give a rebate on the device. Others give a flat credit on the customer’s electricity bill, while others pay per peak hour. If smart metering technology is available, companies will compare actual vs. typical consumption and reward them accordingly.

While an important feature of this approach is that it reduces the customers’ transaction cost to act flexibly – a relevant barrier to successfully achieving price responsiveness – a demand response programme cannot be based on a “representative agent” approach in which customers do not differ

---

The contents of this paper are the author’s sole responsibility. They do not necessarily represent the views of the Oxford Institute for Energy Studies or any of its Members.

---

*Recently some works have been started in this area (see for example Boscán (2016b), who proposes Product-Mix Exchanges, extensions to the Product-Mix Auction proposed by Klemperer (2010) to trade with flexibility services).*
from one another. We claim that designing “efficient” contracts based on this kind of approach is not possible given that suppliers naturally differ in their cost (or disutility) of provision across the different dimensions of flexibility. This is true even when, for example, consumers have identical flexibility enabling assets (e.g., similar air conditioners).

In contrast, we have taken a normative approach to illustrate how the multi-dimensional adverse selection model discussed in our paper can be employed to design bilateral flexibility-enabling contracts that ensure economic efficiency. The main ingredient required to apply our proposed contract design approach is information regarding the distribution of types (i.e., the probabilities that show whether a randomly chosen resource provider is efficient, relatively inefficient or fully inefficient in proving multiple dimensions of flexibility) and the hidden unit cost parameters in the suppliers’ cost functions (i.e., the marginal cost of providing flexibility at each dimension). A key question that follows is how to elicit the relevant information from flexibility suppliers. In practice, acquiring this information can be challenging, especially if the principal does not interact with end-users. However, if the principal is a DSO operating in a smart grid environment, this would be much easier because it allows the customer base to anonymously reveal their type.

In summary, flexibility is an increasingly relevant quality for power systems operating with greater shares of VRE, as these require greater resilience to respond to the challenges of uncertainty and variability. Engineers have successfully developed a rich literature filled with models that describe what flexibility is and how to measure it but have, so far, left the question of how to incentivize its provision largely unaddressed. However, the question of creating incentives is a topic beyond the realm of engineering studies. Instead, economists – armed with an engineer’s perspective – are better suited for the task of designing products, contracts and markets – in a word: institutions – that enable flexibility in a power system. Informed by the technical features of flexibility, this paper has taken a normative perspective and has contributed to the analysis of flexibility from an economic perspective to inform the design of bilateral flexibility contracts. According to this view, flexibility is a heterogeneous product, which has multiple attributes among which buyers and sellers cannot perfectly substitute.

The results of our study provide important insights on designing efficient contracts for flexibility services which can be utilised in a, for example, demand response programme. First, the results show that optimal flexibility contracts depend on the way that providers are distributed. For example, if we assume flexibility with two dimensions (capacity and duration) which is procured from a group of flexibility providers then distribution of type is a set of probabilities that show what percentage of the group is efficient (low cost) in both dimensions or inefficient (high cost) in both dimensions or efficient in one dimension and inefficient in the other dimension. In the model presented in the paper, the information about distribution of types is concentrated in one parameter, such as covariance of type. The information about distribution of types can be obtained through indirectly observing the consumers’ behavior.

The second important result of the model in this paper is that designing the optimal contract for flexibility services is complicated not only because of multidimensional information asymmetry but also because of the fact that the composing elements of flexibility are non-separable – capacity, ramp rate and duration, cannot be produced or consumed separately. This “non-separable externality” leads to further distortion of inefficient flexibility providers beyond the fundamental rent efficiency trade off prevailing under the traditional models of contract designs with information asymmetry. The non-separability distortion is unique to non-conventional commodities such as flexibility and it does not exist when the cost and utility function of an agent and a principal is separable – a condition that has almost always been assumed in the contract theory literature.

To read the entire paper please click here.
References:


Boscán, L. and Poudineh, R. “Procuring power system flexibility in competitive environments”. In preparation.


