Intermittency, Flexibility and Market Design

Nils-Henrik M von der Fehr
University of Oslo

11th June 2014

Abstract

To be written. Very first draft - I apologise for any errors.

Introduction

The considerable build-up of intermittent generation in the European electricity industry has lead some observers to question the appropriateness of the current market design. Others have been less radical, but have nevertheless argued for adjustments to certain elements of the design. Here we consider one such element: the length of contract periods in the daily spot market.

In Europe, as in many other places in the world, electricity is traded on a series of energy-only markets. At the core of this set of markets is the day-ahead or spot market. Trade also takes place on a bilateral basis outside of the organised market exchanges ("over the counter" or OTC), as well as on intraday and balancing markets. However, the bulk of physical trade goes through the spot market; in the Nordic region, more than 80 percent of all trade takes place on NordPool Elspot.

On the spot market, bids and offers for the following trading day are received up to gate closure, which is generaly set at noon. The trading day is divided into a number of contract periods, typically 24 hourly periods, for which bids and offers are made separately and the market is cleared independently. In other words, each day market participants make a set of up to 24 hourly bids and offers for a period of 12 to 36 hours ahead and each of these 24 hourly markets are cleared indepently based on the received bids and offers.

The organisation of electricity markets in general and the spot market in particular cannot be understood without appealing to transaction costs. Trading is costly, for a number of reasons. First, there are the monetray costs of membership and trading fees demanded the market exchange, reflecting the underlying cost of operating such a market place. Second, there are the administrative costs of analysing the market, preparing and making bids and offers, managing contractual obligations and handling payments. Thirdly, there are costs of balancing physical consumption and generation of electricity to contractual obligations.

Transaction costs explain why bids and offers are made with considerable lead time and for discrete periods of time. In theory, trade could take place continuously in real time, with market participants constantly revising their contractual positions as well as their underlying physical consumption and generation of electricity. In practice, however, adjustments to consumption and generation must be planned in advance (cf. the start-up costs of thermal generators), while costs of transactions are reduced by limiting the number of periods for which bids and offers must be made.

On the other hand, requiring that bids and offers be made relatively far in advance of actual consumption and generation, and moreover fixing prices for extended contract periods, introduces inflexibility and results in physical imbalances. Such imbalances are handled by system operators who may request adjustments to consumption and generation, based on bids and offers made in balancing markets, but only at an additional cost, due to the limited participation in such balancing markets. There is therefore a tradeoff between transaction costs on the one hand, and costs of handling imbalances on the other, that determines the design of spot markets.

Conceivably, the introduction of intermittent energy sources, such as solar and wind power, affects this trade off and hence impacts on market design; in particular, one could imagine that the greater variation in generation over the short term requires a shortening of both lead times and contract periods.

In the next section, I set out a theoretical framework for analysing these issues which is then analysed under alternative assumptions in subsequent sections.

The General Framework

We have in mind a setting in which the market is open for trade for a certain period of time $\bar{T}$ (say, 24 hours), with length normalised to 1, i.e. $\bar{T} = 1$ (a day). The trade period is divided into N contract periods T^n, $n = 1, 2, ..., N$ (say, hours) such that $\sum_{n=1}^{N} T^n = 1$. The market is cleared independently for each contract period based on bids and offers for that period. Bids and offers for all contract periods are made at the same time, before any trade takes place. In the analysis below concentration is mostly on a generic contract period and hence the superscript n is dropped where this should not cause confusion.

Demand and supply functions at any time t are given by $q_t^d(p)$ and $q_t^s(p)$, respectively. We denote the corresponding inverse demand and supply functions by $p_t^d(q)$ and $p_t^s(q)$, respectively, where, by definition, we have $p_t^i(q_t^i(p)) = p$, $i = d, s$.

We assume that the market price p is set for the duration of the contract period T, such that expected demand equals expected supply over that period; that is,

$$E\left\{\int_{t=0}^{T} q_t^d(p)\,dt\right\} = E\left\{\int_{t=0}^{T} q_t^s(p)\,dt\right\}, \tag{1}$$

where expectations are taken with respect to the time at which bids and offers are made (ie., for some $t \leq 0$). The assumption of equilibrium in expected terms may be justified by appealing to price-taking and risk-neutral market participants who make bids and offers in a competitive market for their entire volumes over the contract period.

At any moment t, any discrepancy or imbalance between demand and supply is balanced by use of resources external to the market. More specifically, any excess demand $q_t^d(p) - q_t^s(p) > 0$ is covered at a unit cost $\bar{p} > p$, while any excess supply $q_t^d(p) - q_t^d(p) > 0$ is disposed off at a unit value $\underline{p} < p$.

The price that would ensure equilibrium between demand and supply at time t is given implictly by

$$q_t^d(p_t) = q_t^s(p_t). \tag{2}$$

It follows that there is excess demand at time t when $p_t > p$ and excess supply when $p_t < p$. We shall assume that, for any t, $\underline{p} < p_t < \overline{p}$.

The cost of excess demand at time t, compared to the hypotetical situation that the market were balanced, is given by

$$L_t^{ed} = \int_{q_t^s(p)}^{q_t^s(p_t)} (\overline{p} - p_t^s(q))\, dq + \int_{q_t^d(p_t)}^{q_t^d(p)} \left(\overline{p} - p_t^d(q)\right) dq. \tag{3}$$

The loss may interpreted as resulting partly from inefficient sourcing, costing $\bar{p}$ rather than the marginal cost of supply $p_t^s < \overline{p}$, and partly from excess consumption, costing $\bar{p}$ but only being valued at marginal willingness to pay $p_t^d < \overline{p}$.

The loss from excess demand is illustrated in Figure 1. There is excess demand at price p, while a price equal to p_t would have cleared the market. The loss from inefficient sourcing is given by the lightly shaded area, measuring the difference between $\overline{p}$ and p_t^s over the relevant range. Correspondingly, the loss from excess consumption is given by the darkly shaded area, measuring the difference between $\overline{p}$ and p_t^s over the relevant range.

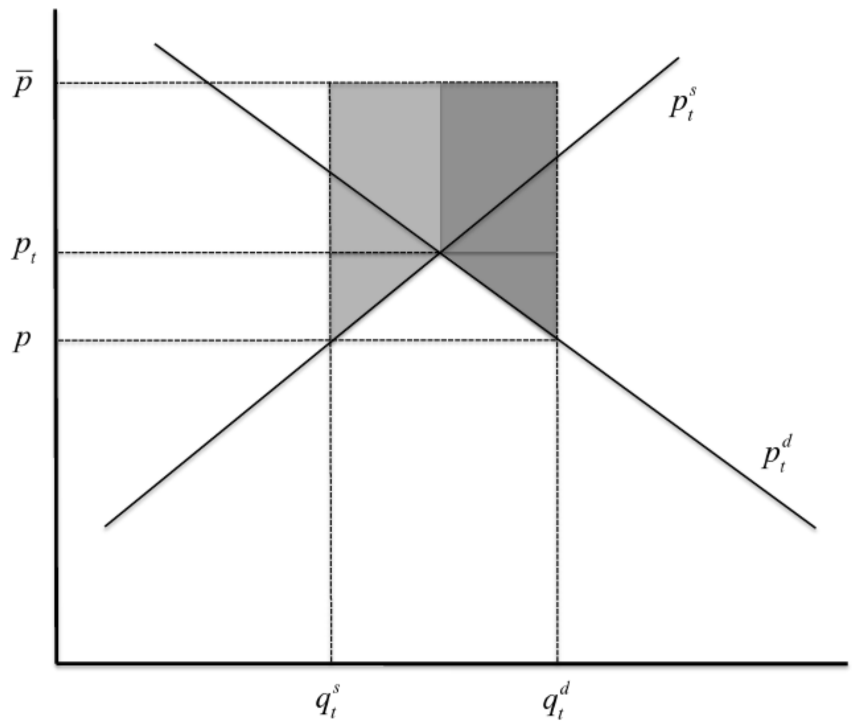

Figure 1: Loss from excess demand

The corresponding loss when there is excess supply, is given by

$$L_t^{es} = \int_{q_t^d(\underline{p})}^{q_t^d(p_t)} \left(p_t^d(q) - \underline{p}\right) dq + \int_{q_t^s(p_t)}^{q_t^s(\underline{p})} \left(p_t^s(q) - \underline{p}\right) dq. \tag{4}$$

Here the loss is resulting partly from insufficient demand, where output is disposed of at price $\underline{p}$ rather than marginal willingness to pay $p_t^d > \underline{p}$, and partly from excess supply, where output costing $p_t^s > \underline{p}$ can only be sold at $\underline{p}$.

The loss from excess supply is illustrated in Figure 2. There is excess supply at price p, but again a price equal to p_t would clear the market. The loss from inefficient demand is given by the lightly shaded area, measuring the difference between p_t^d and $\underline{p}$ over the relevant range, while the loss from excess supply is given by the darkly shaded area, measuring the difference between p_t^s and $\underline{p}$ over the relevant range.

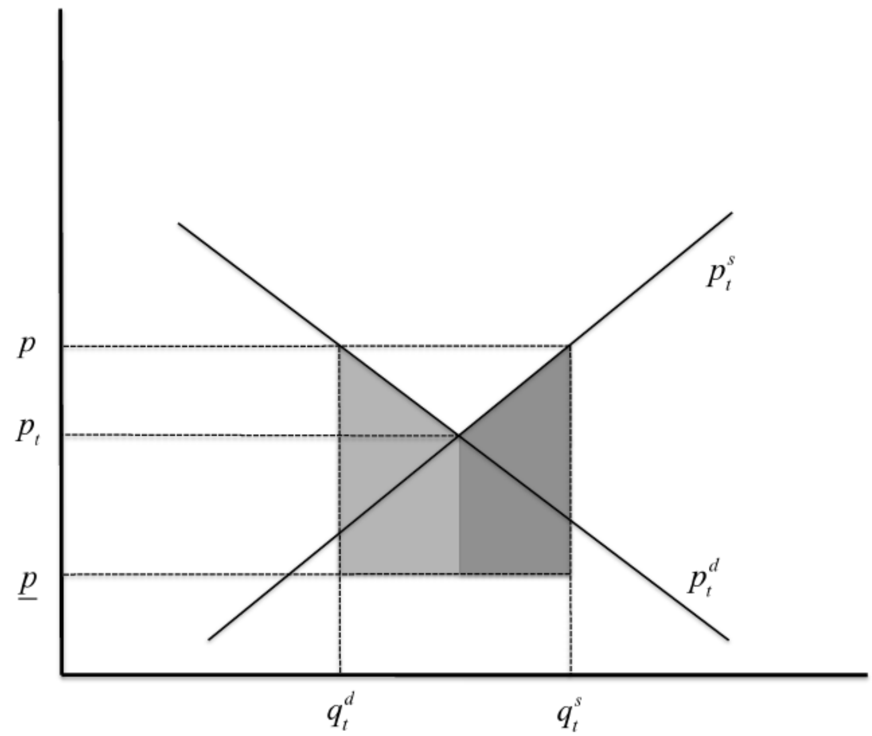

Figure 2: Loss from excess supply

The equality (2) defines p_t as a function of q_t^d and q_t^s; in particular, from the underlying distributions of q_t^d and q_t^s we may derive the distribution for p_t, summarised by the cummulative distribution function F_t. We may then express the expected cost of imbalances at time t by

$$L_t = \int_{\underline{p}}^{p} L_t^{es} dF_t(p_t) + \int_{p}^{\overline{p}} L_t^{ed} dF_t(p_t), \tag{5}$$

while the accumulated loss over the duration of the contracts is given by

$$L = \int_{t=0}^{T} L_t dt. \tag{6}$$

One result is then immediate:

Proposition 1. *If the demand-supply pair* $\left\{q_t^d, q_t^s\right\}$ *is identically and independently distributed across all* t*, then the expected loss from not clearing the market at every instant is proportional to the duration of the contract* T*.*

Proof. When $\left\{q_t^d, q_t^s\right\}$ is identically and independently distributed across all t, we have $L_t = L_0$, for all $t \in [0, T]$. It follows that $L = \int_{t=0}^{T} L_0 dt = L_0 \int_{t=0}^{T} dt = L_0 T$. □

The result demonstrates that the loss from fixing the price over time periods, and hence being unable to clear the market in all contingencies, is not due to variation in demand and supply as such; if demand and supply vary in the same manner at all times the accumulated loss over a certain time period is independent of how often price can be adjusted (so long as it cannot be adjusted continously in real time). In other words, it must be some evolution in demand and supply over time that would potentially lead to a gain from more frequent price changes.

We explore this issue in the next section by means of an example.

A Deterministic Model

In this section, we consider a model in which both demand and supply are deterministic. Demand is inelastic, but increases (or decreases) monotonically over time, while supply is linear in price, but constant over time. While clearly a gross simplification, this set up is not an entirely unreasonable representation of certain features of many electricity markets over shorter time periods: demand is not very responsible to price, but develops according to a well-known pattern; supply, on the other hand, is price responsive, but changes very little with time.

Specifically, inverse demand and supply functions are given by

$$q_t^d(p) = \alpha_0 + \alpha t \tag{7}$$

$$q_t^s(p) = 1 + \beta p \tag{8}$$

where $\alpha_0 > 1, \alpha$ and $\beta > 0$ are constants. In general, α may be both positive (increasing demand) and negative (decreasing demand); here we concentrate on the case $\alpha > 0$, but the analysis is parallell for the case $\alpha < 0$.

The price that would ensure equilibrium between demand and supply at time t (cf. (2)) is given by

$$p_t = \frac{\alpha_0 + \alpha t - 1}{\beta}. \tag{9}$$

Correspondingly, the market price p that is set for the contract period, such that in expectation (or, more appropriately, on average) demand equals supply over that period (cf. (1)), is given by

$$p = \frac{\alpha_0 + \frac{\alpha}{2}T - 1}{\beta}. \tag{10}$$

Comparing (9) and (10), we note that p is simply the average of p_t over the duration of the contract T.

The net loss from disposing of excess supply, which, since $\alpha > 0$ and demand therefore grows linearly, occurs for $t < \frac{T}{2}$, is given by

$$L_t^{es} = \alpha \left[t - \frac{T}{2}\right] \left\{\underline{p} - \frac{1}{\beta}\left[\alpha_0 - 1 + \frac{\alpha}{2}\left(t + \frac{T}{2}\right)\right]\right\}. \tag{11}$$

The net loss of covering excess demand, which occurs for $t > \frac{T}{2}$, is similarly given by

$$L_t^{ed} = \alpha \left[t - \frac{T}{2}\right] \left\{\overline{p} - \frac{1}{\beta}\left[\alpha_0 - 1 + \frac{\alpha}{2}\left(t + \frac{T}{2}\right)\right]\right\}. \tag{12}$$

Figure 3 illustrates the loss for different levels of demand at, respectively, the beginning ($t = 0$), the middle ($t = \frac{T}{2}$) and the end of the contract period ($t = T$). The loss from excess supply is at its maximum at the beginning of the contract period - illustrated by the light shaded area - but gradually falls as time approaches the middle of the contract period. At the middle of the contract period, demand equals supply, but then the loss from excess demand grows with time until the end of the contract period, where it reaches its maximum - illustrated by the dark shaded area. Note that, over time, the loss from excess supply decreases at an increasing rate ($\frac{dL_t^{es}}{dt} = \alpha\left(\underline{p} - p_t\right) < 0$, $\frac{d^2L_t^{es}}{dt^2} = -\alpha^2 < 0$), while the loss from excess demand increases at a decreasing rate ($\frac{dL_t^{ed}}{dt} = \alpha\left(\overline{p} - p_t\right) > 0$, $\frac{d^2L_t^{ed}}{dt^2} = -\alpha^2 < 0$).

The total loss over the contract period, becomes

$$L = \alpha\frac{T^2}{8}\left(\overline{p} - \underline{p} - \frac{\alpha}{\beta}\frac{T}{3}\right). \tag{13}$$

From the assumption that $\underline{p} < p_t < \overline{p}$ all $t \in [0, T]$, it follows that $\overline{p} - \underline{p} > \frac{\alpha}{\beta}T$. For a given contract length, the loss of imbalances is therefore increasing in both the growth rate of demand, α, and the steepness of the supply function, β (i.e. $\frac{dL}{d\alpha}$, $\frac{dL}{d\beta} > 0$), albeit at decreasing rates (i.e. $\frac{d^2L}{d\alpha^2}$, $\frac{d^2L}{d\beta^2} < 0$). Furthermore, we have $\frac{dL}{dT} > 0$ and $\frac{d^2L}{dT^2} > 0$; that is:

Proposition 2. *When demand is price inelastic and increasing (or decreasing) over time, while supply is price elastic but constant over time, then the loss from imbalances over the contract period is increasing and convex as a function of the length of the contract period.*

It follows that if the costs of contracting are bounded, the optimal length of the contract period will be finite.

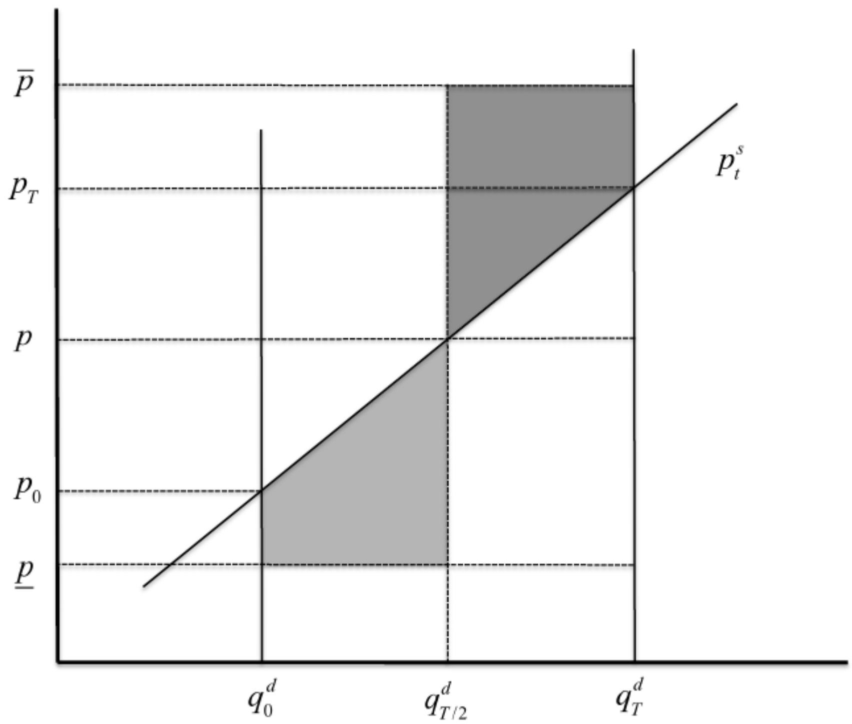

Figure 3: Loss with growing demand

Suppose the cost of contracting equals K per contracting period. Suppose further that the trading period is divided up into N contract periods of equal length (having equally long contract periods is clearly optimal given the convexity of the loss function), so that $T = \frac{1}{N}$. Then the total cost of trading and loss from imbalances over the trading period is given by

$$C = N\left(L + K\right). \tag{14}$$

If we treat N as a continous variable, we find the optimal number of contract periods, N^*, from the first-order condition $\frac{dC}{dN} = 0$, which reduces to

$$8KN^{*3} - 3\alpha\left(\overline{p} - \underline{p}\right)N^* + 2\frac{\alpha^2}{\beta} = 0. \tag{15}$$

From (15) it follows that $\frac{dN^*}{d\alpha} > 0$, $\frac{dN^*}{d\beta} > 0$, $\frac{dN^*}{dK} < 0$, $\frac{dN^*}{d\underline{p}} < 0$ and $\frac{dN^*}{d\overline{p}} > 0$. In other words:

Proposition 3. *When demand is price inelastic and increasing (or decreasing) over time, supply is price elastic but constant over time and the cost of contracting is constant, the optimal length of the conctract periods is decreasing in the growth rate of demand, the steepnes of the supply function and the cost of sourcing output to cover excess demand, but increasing in the cost of contracting and the value at which excess supply may be disposed of.*

A Stochastic Model

In this section, we extend the above deterministic model by introducing a stochastic element in the demand function. Specifically, inverse demand and supply functions are given by

$$q_t^d(p) = \alpha_0 + \alpha t + u_t \tag{16}$$

$$q_t^s(p) = 1 + \beta p \tag{17}$$

where u_t is an independently, identically and uniformly distributed variable on the interval $[-\sigma, \sigma]$. We again assume $\alpha_0 > 1$ and $\beta > 1$. Note that $\alpha > 0$ and $\sigma = 0$ corresponds to the deterministic case considered in the previous section.

The price that would ensure equilibrium between demand and supply at time t (cf. (2)) is given by

$$p_t = \frac{\alpha_0 + \alpha t + u_t - 1}{\beta}. \tag{18}$$

Since $E\{u_t\} = 0$, it follows that $E\left\{q_t^d\right\} = \alpha_0 + \alpha t$ and so the market price p that is set for the contract period such as to equalise expected demand and expected supply over that period (cf. (1)) is, as in the deterministic case, given by

$$p = \frac{\alpha_0 + \alpha\frac{T}{2} - 1}{\beta}. \tag{19}$$

For a given value of t and a given realisation of u_t, we find that the net loss from disposing of excess supply, which, comparing (18) and (19), occurs for $u_t < \alpha\left(\frac{T}{2} - t\right)$, is given by

$$L_t^{es} = \left[\alpha\left(t - \frac{T}{2}\right) + u_t\right]\left\{\underline{p} - \frac{1}{\beta}\left[\alpha_0 - 1 + \frac{\alpha}{2}\left(t + \frac{T}{2}\right) + \frac{1}{2}u_t\right]\right\}, \tag{20}$$

while the net loss of covering excess demand, which occurs for $u_t > \alpha\left(\frac{T}{2} - t\right)$, is similarly given by

$$L_t^{ed} = \left[\alpha\left(t - \frac{T}{2}\right) + u_t\right]\left\{\bar{p} - \frac{1}{\beta}\left[\alpha_0 - 1 + \frac{\alpha}{2}\left(t + \frac{T}{2}\right) + \frac{1}{2}u_t\right]\right\}. \tag{21}$$

For simplicity, we assume that variation in demand is such that both excess demand and excess supply occur with positive probability at any given point in time; in particular, we assume $\sigma > \frac{\alpha}{2}T$. Taking expectations with respect to u_t, the expected loss from imbalance at any given time t becomes

$$\begin{aligned} L_t = &\tfrac{1}{2}\bar{p}\left[\alpha\left(\tfrac{T}{2} - t\right) - \sigma\right]^2 - \tfrac{1}{2}\underline{p}\left[\alpha\left(\tfrac{T}{2} - t\right) + \sigma\right]^2 \\ &+ \tfrac{1}{\beta}\left\{2\alpha\sigma\left(\tfrac{T}{2} - t\right)\left[\alpha_0 - 1 + \tfrac{\alpha}{2}\left(\tfrac{T}{2} + t\right)\right] - \tfrac{\sigma^3}{3}\right\} \end{aligned} \tag{22}$$

Next, integrating over t, we find the total expect loss over the contract period to be

$$L = \left[\bar{p} - \underline{p}\right]\left[\frac{\alpha^2}{12}T^2 + \sigma^2\right]\frac{T}{2} - \frac{\sigma}{\beta}\left[\frac{\alpha^2}{4}T^2 + \sigma^2\right]\frac{T}{3} \tag{23}$$

Note, as follows from Proposition 1, the total expected loss,L, is proportional to the duration of the contract period, T, when the deterministic part of demand is constant over time, i.e. when $\alpha = 0$. However, in the more general case, we have $\frac{\partial L}{\partial \sigma} > 0$, $\frac{\partial^2 L}{\partial \sigma^2} > 0$ at $\sigma = \frac{\alpha}{2}T$, meaning that the total expect loss is increasing and convex in the stochastic variation in demand. Moreover, we have $\frac{\partial^2 L}{\partial T \partial \sigma} > 0$ at $\sigma = \frac{\alpha}{2}T$, meaning that the marginal loss from extending the duration of the contract period, T, is increasing in the variation in demand.

If, as in the previous section we suppose that the cost of contracting equals K per contracting period, and that the trading period is divided up into N contract periods of equal length, so that $T = \frac{1}{N}$, we find the optimal number of contract periods, N^*, from the condition

$$KN^{*3} = \frac{\alpha^2}{12}\left[\overline{p} - \underline{p} - 2\frac{\sigma}{\beta}\right]. \tag{24}$$

From (24) it follows that $\frac{dN^*}{d\sigma} < 0$.

Non-Monotonicity

To be written (demand as an inverse U over the entire trading period; asymmetric contract periods).

Demand and Supply Conditions

To be written (empirical evidence on actual development of demand and supply; implications for market design).

Conclusion

To be written.

References

[1] x