ADJUSTMENT COSTS

This article surveys the use of adjustment frictions in macroeconomic research, exploring the consequences of convex and non-convex adjustment costs for firm-level decisions and the dynamics of macroeconomic aggregates. The mechanics of these frictions are illustrated using several prominent examples including the partial adjustment model of employment, the q-theoretic investment model, and lumpy adjustment models of investment and employment. We also review the (S,s) inventory model, where stock accumulation is explained as the result of fixed delivery costs, and briefly discuss (S,s) decision rules arising from piecewise-linear costs in the context of capital irreversibility and firing taxes.

Across a wide body of macroeconomic research, the interest in adjustment costs has been largely utilitarian. In designing theoretical models to organize our understanding of patterns observed in the data, we make hard choices about which of the many elements affecting the decisions of actual firms and households and the outcomes of their market interactions to include. Given their necessary simplicity, we often find that the predictions of the theoretical economies we are able to analyse are too stark relative to the behaviour observed in actual economies. Thus, in a variety of settings we have adopted adjustment costs in our economic laboratories to summarize omitted frictional elements that reduce, delay or protract changes in the demand and supply of final goods and their factor inputs in response to changes in economic conditions.

In these few pages, we describe the mechanics of commonly used adjustment costs and briefly discuss their role in several leading macroeconomic applications. Since a comprehensive survey is beyond the scope of this article, many important applications have been excluded. However, where possible we direct the reader to influential research on these topics.

1. Convex costs

Until relatively recently, most macroeconomic research involving adjustment costs emphasized the use of convex cost functions to penalize swift changes in aggregate variables and thereby induce gradual movements over time. Historically, models with convex adjustment costs were developed as a theoretical foundation to explain why the inclusion of lagged dependent variables in empirical models of factor demand led to sharp improvements in their econometric performance. While early researchers had found decision-theoretic models based on static demand theory unable to account for the serial correlation observed in aggregate employment and investment, these same models performed relatively well when they were augmented with ad hoc distributed lags of the dependent variable or its theoretical determinants (as in the flexible accelerator model of Koyck, 1954, or the flexible user-cost model of Hall and Jorgenson, 1967). These lags were broadly motivated by the idea that certain frictions prevent firms from immediately attaining their chosen employment or capital levels, instead engendering gradual, partial adjustment towards these target levels over time.
For example, by assuming that firms adjusted their workforces at constant rate $\lambda \in (0,1)$ towards the target implied by static demand theory, $N^*_t$, current employment could be written as a distributed lag of previous target employments:

$$N_t = \lambda N^*_t + (1-\lambda)N_{t-1} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j N^*_{t-j}. \quad (1)$$

To implement such partial adjustment models, researchers replaced the distributed lag of unobservable targets with distributed lags of each observable series the theory suggested should influence them – for instance, real wages. In this way, lags of the determinants of demand were introduced into the estimation equation, thus introducing the empirically desirable serial correlation.

Without some theoretical basis to explain their empirical success, partial adjustment models might have been abandoned quickly. A partial resolution arrived in the mid- to late 1960s with the application of capital adjustment costs in models of investment (see Eisner and Strotz, 1963; Lucas, 1967; Gould, 1968; Treadway, 1971). There, gradual aggregate adjustment broadly consistent with the analogue to (1) was obtained by assuming that, beyond other costs associated with the acquisition of capital (for example, user costs), the very act of adjusting the capital stock incurred real output costs. These costs, $\Phi(k',k)$, were strictly increasing and convex in the distance between the chosen new level of capital and the current level, $|k' - k|$, thereby implying a smoothly rising marginal adjustment cost in the size of the current adjustment. As such, they introduced dynamic elements into the firm’s previously static decision problem and led it to smooth its investment activities over time. Nonetheless, so long as the treatment of expectations was incomplete, the mapping to a partial adjustment equation could not be robustly established.

The work of Sargent (1978) extended the theory in the context of employment adjustment by showing how, under rational expectations, the partial adjustment model could be derived from the profit maximization problem of a firm facing quadratic adjustment costs. To simplify the problem somewhat, consider a firm that enters any period with employment $n_{t-1}$ and incurs costs, $\Phi(n_t, n_{t-1}) \equiv \frac{\psi}{2}(n_t - n_{t-1})^2$, in altering its workforce for production. Next, assume that the firm’s production function is quadratic, $f(n_t, z_t) \equiv (f_0 + z_t)n_t - \frac{f_1}{2}n_t^2$, where $f_0 > 0$, $f_1 > 0$, and $z_t$ is a serially correlated exogenous productivity process, as is the real wage, $w_t$. Discounting its future earnings by $\beta \in (0,1)$ and given initial employment $n_{t-1}$, the firm selects $\{n_t\}_{t=0}^{\infty}$ to maximize its expected present discounted value, $E\left[\sum_{t=0}^{\infty} \beta^t (f(n_t - z_t) - w_t n_t - \Phi(n_t, n_{t-1})) \mid z_0, w_0\right]$, arriving at a sequence of Euler equations:

$$\beta E_n n_{t+1} - \left(1 + \beta + \frac{f_1}{\psi}\right)n_t + n_{t-1} = \frac{w_t - a_t - f_0}{\psi}.$$  

If we isolate the two real roots of this second-order stochastic difference equation, the solution is precisely (1) above, with target employment in each date given by

$$N^*_t = \left[E_n \sum_{j=0}^{\infty} (\beta/\lambda)^j (x z_{t+j} - \chi w_{t+j}) \right]$$  

(2)
and the parameters $\lambda$, $\chi_z$, and $\chi_w$ determined by the adjustment cost parameter $\phi$, the discount factor $\beta$, and the parameters of the production function.

For researchers implementing equations like (1), an important contribution of Sargent’s model was in illustrating how the very features that linked current employment to its lagged determinants also necessarily divorced each date’s target, $N_t^*$, from the statically derived optima assumed in early partial adjustment estimations. Notice that the firm’s target in (2) involves expectations of each variable affecting the future value marginal product of labour, because, given adjustment costs, this current choice influences its future level of employment. Moreover, as an increase in the adjustment cost parameter, $\phi$, shifts the marginal adjustment cost schedule upward at all dates, it not only implies a slower adjustment rate (lower $\lambda$) but also increases the influence of these expectations of future variables in the determination of the current target.

Across the many models including convex adjustment costs, quadratic cost functions have been by far the most common specification, essentially for sake of tractability. Note that, given the quadratic form of $\Phi(n_t, n_{t-1})$ above, firms’ decision rules described by (1) and (2) are linear. As such, they aggregate conveniently to represent economy-wide factor demand in partial adjustment models. (Hamermesh, 1989, and Hamermesh and Pfann, 1996, discuss the role of these costs in partial adjustment models of employment demand. Chirinko, 1993, Hassett and Hubbard, 1997, and Caballero, 1999, survey their use in empirical investment equations. Hall, 2004, estimates an industry-level model of production with quadratic adjustment costs applied to both labour and capital.)

A similar cost function appears in the history of q-theoretic investment models, unifying neoclassical investment theory with the theory of Brainard and Tobin (1968) and Tobin (1969), which holds that investment should be positively related to average Q, the ratio of the value of the firm relative to its capital stock. Appending the neoclassical model with a general convex adjustment cost function, Abel (1979) moved to reconcile the two theories by showing that the expected discounted marginal value of capital for a firm, marginal q, is sufficient to determine its investment rate. The reconciliation was complete when Hayashi (1982) showed that average Q is identical to marginal q if firms are perfectly competitive and both the production function and $\Phi(k', k)$ are linearly homogenous (for example, $\Phi(k', k) = \frac{\phi}{2} \frac{(k' - k)^2}{k}$).

Since the mid-1980s, macroeconomic analysis has become firmly grounded in dynamic stochastic equilibrium analysis. Nonetheless, the gradual movements implied by equilibrium relative price changes have often proven inadequate in reconciling models to data; thus, convex costs have continued to appear. A famous early application to capital adjustment is the industry equilibrium study of investment by Lucas and Prescott (1971). More recently, examples of general equilibrium models adopting these frictions may be found in almost every field of macroeconomics.

2. Non-convex costs

Despite their relative success in reproducing the persistence of aggregate series, empirical models based on convex adjustment costs have fared poorly along other dimensions. For example, estimations of the neoclassical investment model attribute very low explanatory power to average Q and assign large coefficients to adjustment cost parameters in explaining changes in investment
Large estimates of adjustment costs, which in turn imply implausibly slow adjustment speeds, are also a recurring problem for linear quadratic inventory models (Ramey and West, 1999). Elsewhere, the sharp difference between rates of employment adjustment estimated from high-frequency firm-level data and those estimated from low-frequency aggregate data suggests spatial and temporal bias inconsistent with the common assumption of symmetric quadratic adjustment costs (Hamermesh and Pfann, 1996). Moreover, there is mounting microeconomic evidence suggesting that the predominant adjustment frictions confronting firms in actual economies may be non-convex, rather than convex, in nature.

Contrary to the smooth, continual adjustments implied by convex cost models, recent microeconomic studies reveal that firm-level factor adjustment exhibits long periods of relative inactivity punctuated by infrequent and large, or lumpy, changes in stocks. Examining capital adjustment in a 17-year sample of large, continuing US manufacturing plants, Doms and Dunne (1998) find that roughly 25 per cent of the typical plant’s cumulative investment occurs in a single year, and more than half of plants exhibit capital adjustment of at least 37 per cent within one year. Using a similar dataset, Cooper, Haltiwanger and Power (1999) provide additional evidence of lumpy investment, and they show that the conditional probability of a large investment episode rises in the time since the last such episode. Microeconomic evidence of non-smooth employment adjustment is abundant (see Hamermesh and Pfann, 1996). For example, examining monthly data on employment and output across seven US manufacturing plants between 1983 and 1987, Hamermesh (1989) finds that plant-level employment remains roughly constant over long periods while production fluctuates. These long episodes of constancy are broken by infrequent but large jumps, at times roughly coinciding with the largest output fluctuations. (Interestingly, while the convex cost model is inconsistent with the lumpy employment adjustments at each plant, Hamermesh finds that it represents the aggregate of employment – and production – across plants reasonably well.) Beginning with Scarf (1960), a number of theoretical studies have shown that precisely this variety of nonlinear microeconomic adjustment can arise when firms are confronted with non-convex adjustment technologies.

2.1 (S,s) stock adjustment

Scarf (1960) provided the earliest formal analysis of microeconomic adjustment behaviour in the presence of non-convex adjustment costs. There, the adjustment cost was a simple fixed cost, \( \varphi > 0 \), incurred at any time a firm wished to adjust its stock of inventories. (Beginning with the work of Barro, 1972, and Sheshinski and Weiss, 1977, fixed costs have also been used to develop models of (S,s) firm-level price adjustment. Early studies examining the potential for monetary non-neutralities in such settings include Sheshinski and Weiss, 1983; Caplin and Spulber, 1987; and Caplin and Leahy, 1991. More recent general equilibrium analyses include Caplin and Leahy, 1997; Dotsey, King and Wolman, 1999; Gertler and Leahy, 2006; and Golosov and Lucas, forthcoming.) We briefly review the model below.

Consider a retail firm entering any period with inventories, \( y > 0 \), of a homogenous good available for sale. The firm faces stochastic demand, \( \xi \), drawn from a time-invariant distribution \( F(\xi) \), and the value of its sales is \( p \min \{ y, \xi \} \). At the end of the period, it may place an order \( x > 0 \) to increase its available stock for the next period; \( y' = y - \min \{ y, \xi \} + x \). The cost of any such order
is $\varphi + cx$, where $c > 0$ represents the unit cost of the good held in inventory. By proving K-concavity of the value function, Scarf was able to establish that the firm’s optimal decision rule takes the following one-sided $(S, s)$ form. (Scarf, 2005, shows this decision rule generalizes to a setting where the firm selectively sells its inventories with the option of leaving some demand unsatisfied. See Dixit, 1993, for a characterization of two-sided $(S, s)$ policies arising in continuous time settings involving fixed and piecewise linear adjustment costs.)

$$x = \begin{cases} 0 & \text{for } y \in (s, S] \\ S - y & \text{for } y \leq s \end{cases}$$

To avoid repeatedly incurring fixed costs, the firm places no orders so long as its sales do not move its stock outside the interval $(s, S]$. Only when its inventories have fallen to the lower threshold, $s$, does it take action, resetting its stock to $S$. Thus, the increasing returns adjustment technology implied by fixed order costs induces infrequent and relatively large, or lumpy, orders.

Just as firm-level data indicates lumpiness in microeconomic capital and employment adjustment, there are a number of studies suggesting that firms in both manufacturing and trade manage their inventories according to $(S, s)$ policies resembling that obtained in Scarf’s path-breaking analysis (for example, Mosser, 1991; Hall and Rust, 2000). Nonetheless, despite the empirical difficulties associated with convex cost inventory models (Blinder and Maccini, 1991; Ramey and West, 1999), the implications of firm-level inventory policies under non-convex adjustment costs have been left relatively unexplored by macroeconomists. To reproduce the relatively smooth changes observed in the aggregate, such models necessarily involve a distribution of firms over inventory levels. As this distribution becomes part of the economy’s aggregate state vector, the resulting high dimensionality makes it difficult to determine equilibrium prices, including real wages and interest rates. It is this basic problem that has generally dissuaded researchers from undertaking dynamic stochastic general equilibrium analyses of environments involving non-convexities, among them the $(S, s)$ inventory model.

One exception to this is found in Fisher and Hornstein (2000). Building on the work of Caplin (1985) and Caballero and Engel (1991), who study the aggregate implications of exogenous $(S, s)$ policies across firms, Fisher and Hornstein construct an environment that endogenously yields time-invariant one-sided $(S, s)$ adjustment rules and a constant order size per adjusting firm. This allows them to tractably study $(S, s)$ inventory policies in general equilibrium without confronting substantial heterogeneity across firms. More generally, in models involving time-varying two-sided $(S, s)$ policies, the heterogeneity becomes more cumbersome, as in Khan and Thomas’ (2006a) general equilibrium business cycle study. There, at the start of any period, each firm observes the current state and then chooses whether to order intermediate goods for use in production. Given this timing, alongside positive real interest rates, inventories would never be held in the absence of some friction. However, by confronting firms with idiosyncratic order costs independent of their chosen order sizes, continual orders are deterred, and $(S, s)$ inventory adjustment adopted. Based on the results of their calibrated model, Khan and Thomas conclude that such non-convex costs can be quite successful in explaining not only the existence of aggregate inventories but also their cyclical dynamics.
2.2 Implications for aggregate investment

Non-convex adjustment costs imply distributed lags in aggregate series similar to those generated by convex costs, because they stagger the lumpy adjustments undertaken by individual firms in response to shocks (King and Thomas, 2006). However, they are distinguished by their potential for aggregate nonlinearity, which has generated particular interest within investment theory. A number of influential partial equilibrium studies (Caballero and Engel, 1999; Cooper, Haltiwanger and Power, 1999; Caballero, Engel and Haltiwanger, 1995) have argued that investment models with non-convex costs empirically outperform convex cost models because they can deliver disproportionately sharp changes in aggregate investment demand following large aggregate shocks. (Caballero and Engel, 1993, and Caballero, Engel, and Haltiwanger, 1997, arrive at similar conclusions in the context of employment adjustment.)

Caballero and Engel (1999) examine generalized \((S, s)\) policies rationalized by stochastic fixed adjustment costs, \(\phi\), distributed i.i.d. across firms and over time. In this environment, a firm’s capital, \(k\), becomes part of its state vector alongside its total factor productivity, \(z\). Moreover, microeconomic adjustment becomes probabilistic; firms with the same current gap between actual and target capital do not necessarily behave identically; rather, those with relatively low \(\phi\) draws are more likely to alter their capital than those drawing high costs. If we transform Caballero and Engel’s gap-based analysis to reflect the firm-level state, \((k, z)\), the implication is an adjustment hazard, \(\Lambda(k, z)\), indicating what fraction of each group of firms sharing a common current state will choose to adjust their capital to a common target, \(k^*(z)\). The resulting generalized \((S, s)\) adjustment model allows convenient aggregation and has been studied in a variety of settings. (Dotsey, King and Wolman, 1999, apply this basic framework to price adjustment, Thomas, 2002, adopts it in an equilibrium business cycle model with lumpy investment, and King and Thomas, 2006, use it to examine employment adjustment.)

To understand how this mechanism can affect the dynamics of aggregate investment, consider the following simple partial equilibrium example described by Khan and Thomas (2003). Assume that total factor productivity, \(z\), is a Markov process common to all firms. If there have been no aggregate shocks for many periods, the distribution of firms will have support at \(k^*(z)\), \((1 - \delta)k^*(z)\), \((1 - \delta)^2k^*(z)\), and so on. As a firm’s capital stock depreciates further below the target, \(k^*(z)\), the maximum adjustment cost it will accept to reset its capital stock to that target, \(\phi(k, z)\), rises. Thus, the adjustment hazard, \(\Lambda(k, z)\), is increasing in the distance \(|k^*(z)k|\). Finally, the total measure of adjusting firms is \(\int \Lambda(k, z)\mu(\text{dk})\), and aggregate investment is \(I = \int \Lambda(k, z)(k^*(z) - (1 - \delta)k)\mu(\text{dk})\).

Suppose that a negative aggregate shock reduces \(z\) to \(z_L\), thereby reducing expected future marginal productivity of capital. This causes a downward shift in the target stock, placing it strictly within the existing range of capital held by firms. Thus, \(\Lambda(k, z)\) falls for many firms, rising only for those with the highest levels of capital. As a result, the total adjustment rate can actually fall, thereby dampening the fall in aggregate investment demand implied by the reduced target. By contrast, when a positive technology shock raises \(z\) to \(z_H\), the target capital rises above that currently held by any firm. This increases the total adjustment rate, compounding the effect of the raised target to which firms adjust.
More generally, this example illustrates that, when there is an aggregate shock, and thus a change in the target, higher moments of the distribution of capital across firms matter in determining movements in aggregate investment, because the adjustment hazard is a non-trivial function of capital. (This is an important distinction relative to the convex cost/partial adjustment model. Rotemberg, 1987, shows its aggregate dynamics are reproduced by a model where individual firms adjust infrequently, but all face a common probability of undertaking adjustment, independent of their individual states. Given this constant hazard, only the first moment of the distribution is relevant in determining aggregate changes.) Alternatively, in the language of Caballero (1999, p. 841), microeconomic non-convexities can generate an important ‘time-varying/history-dependent aggregate elasticity’ of investment to shocks by allowing changes in the synchronization of firms’ capital adjustments.

Although findings like those above echo throughout partial equilibrium studies involving lumpy adjustments, the omission of market-clearing relative prices (for example, equilibrium interest rates) may be critical to the inferred macroeconomic importance of non-convex factor adjustment costs. Significant aggregate nonlinearities can only occur if adjustment hazards exhibit large changes in response to shocks. Clearly, from the example above, such changes depend entirely on the extent to which \( k^*(z) \) responds to changes in \( z \). However, just as the capital adopted by a representative firm facing no adjustment costs varies far less when prices adjust to clear all markets, Thomas (2002) and Khan and Thomas (2003; 2006b) show that the target capital(s) selected by firms facing non-convex costs exhibit changes an order of magnitude smaller in general equilibrium. Because large movements in target capital, and hence in aggregate investment demand, would imply intolerable consumption volatility for households (at least in the closed-economy settings examined in these studies), they do not occur in equilibrium. Instead, small changes in relative prices serve to discourage sharp changes in \( k^*(z) \), thereby preventing large synchronizations in firms’ investment timing and leaving the aggregate series largely unaffected by the microeconomic lumpiness caused by non-convex adjustment costs.

### 3. Piecewise-linear costs

Among the adjustment frictions commonly applied in macroeconomic research, we have thus far omitted an important type of convex costs, namely, piecewise-linear adjustment costs, which are often associated with partial irreversibilities in investment and employment. As these costs have quite different implications from those described in section 1, we briefly discuss them here. Like non-convex costs, piecewise-linear costs lead to \((S,s)\) decision rules. However, as they yield no increasing returns in the adjustment technology, they do not in themselves cause lumpiness. Rather, when the firm’s relevant state variable reaches the lower or upper bound of its tolerated region of inaction, the firm undertakes small adjustments to maintain it at that bound. (To explore the extreme case of complete irreversibility, see Pindyck, 1988, for an analysis that emphasizes the option value of waiting to invest, or Bertola, 1998, for a characterization of firm decision rules using standard dynamic programming. Dixit and Pindyck (1994) provide a comprehensive survey of this literature.)

Partial irreversibilities have been widely examined in investment theory as an explanation for the common empirical finding that investment is insensitive to Tobin’s q. Abel and Eberly
(1994) characterize firm-level investment when the purchase price of capital, \( p^+_K \), exceeds its sale price, \( p^-_K \) (and there are flow-fixed and convex adjustment costs). They show that this cost structure makes investment a nonlinear function of marginal \( q \), implying a range of values over which the firm does not invest. (Veracierto, 2002, solves a general equilibrium business cycle model where the resale price of capital goods is a constant fraction of the purchase price. Examining a wide range of values for this irreversibility parameter, he concludes that such frictions have no quantitatively significant effects for business cycle dynamics.) Elsewhere, in the context of employment adjustment, a simple example of piecewise-linear costs is an environment where firms incur no adjustment costs in increasing their employment, but pay a tax of \( \varphi > 0 \) per worker fired. The implications of such firing costs for aggregate employment are theoretically ambiguous. While their direct effect is to discourage firing, they also induce a reluctance to hire. Bentolila and Bertola (1990) provide an early analysis suggesting that the direct effect dominates, while Hopenhayn and Rogerson (1993) find the converse.

4. Conclusion
Throughout the history of their use, the primary purpose of adjustment costs has been to reduce the distance between model-generated and actual economic time series. Because they largely represent implicit costs of forgone output, we have little ability to directly measure adjustment frictions. Thus, when we adopt them to enhance the empirical performance of our models, the resulting improvements are, in some sense, a measure of our ignorance.

As suggested by the discussion above, the existence and size of particular adjustment frictions has typically been inferred from the extent to which they modify dynamic behaviour within a specific model to more closely resemble that in the data. This raises an obvious, but sometimes forgotten, point. Adjustment costs derived within a given class of model may be quite inappropriate in a second, distinct class of model. For example, the relative sizes of various types of adjustment frictions needed to reconcile theoretical and actual microeconomic data can differ sharply depending on the specification of equilibrium and firm-level shocks.

Aubhik Khan and Julia K. Thomas

See also inventory theory; irreversible investment; s-S models

Bibliography


*Index terms*

adjustment costs
adjustment hazards
business cycles
convex cost functions
distributed lags
dynamic stochastic equilibrium analysis
Euler equations
frictions
intermediate goods
inventory policies
inventory theory
investment theory
linear quadratic inventory models
lumpy investment
market-clearing relative prices
neoclassical investment theory
neutrality of money
partial adjustment
partial equilibrium
piecewise-linear adjustment costs
production functions
quadratic cost functions
rational expectations
s-S models
Tobin’s q
total factor productivity