Second Thoughts on Double Marginalization

BY JOHN KWOKA AND MARGARET SLADE

VERTICAL INTEGRATION BETWEEN companies can lead to numerous cost‐lowering efficiencies that are due, for example, to economies of scope, eliminating contracting costs, aligning incentives, and enhancing relationship‐specific investments. Nevertheless, the elimination of double marginalization (EDM) has received the most attention in antitrust literature and practice.

Economic textbooks describe a setting in which an upstream monopoly faces a downstream monopoly in a fixed proportion production process. In this case, the unintegrated downstream firm “over‐prices” by marking up the already marked‐up upstream product. By eliminating this double margin, or successive monopoly markups, vertical integration lowers prices and leads to benefits for both producers and consumers.

This scenario has formed a key element of competition policy for decades. To illustrate, Erik Hovenkamp and Neel Sukhatme’s analysis of vertical mergers concluded: “Almost all vertical mergers have potential procompetitive effects, the most robust of which is the elimination of double marginalization.”1 A statement by a then‐senior Federal Trade Commission enforcer described the prospect of eliminating double marginalization as an “intrinsic” efficiency justification of vertical mergers.2 That view is echoed in the FTC’s recent decision to approve the vertical United Health/DaVita merger, where the majority asserted that “[a] major source of these [claimed efficiency] benefits is the elimination of double‐marginalization.”3 And perhaps most recently, and famously, in the antitrust case against the AT&T/Time Warner merger, the government conceded benefits from EDM on the order of $350 million without qualification.4

In some cases the conclusions reached might be the result of correct applications of the textbook model of vertical integration. Antitrust and regulatory policy, however, has tended to apply the EDM argument uncritically, ignoring several key assumptions and issues. The result, we think, has been the widespread risk of misuse of the economic proposition and potentially erroneous policy decisions. Moreover, it has distracted attention from traditional supply‐side efficiencies from vertical integration, such as economies of scope, elimination of duplicate facilities such as a head office, the ability to coordinate other aspects of the vertical chain, and the expectation of productivity growth due to knowledge transfers.

In this article we review the underlying model of double marginalization and then address its strong assumptions and other possible limitations on its use as a policy prescription.

Double Marginalization: The Simple Theory

The textbook model of double marginalization is appealingly simple. Let us suppose there is a single upstream manufacturer of a product sold to a single downstream retailer. The downstream firm in turn sells the product to final consumers. Each firm is a monopolist at its stage and, as do all monopolists, each charges a markup on the product that is above its own marginal cost, where the markups are determined by the elasticity of demand that each faces. Assuming constant marginal costs at each stage and linear demand, Figure 1 depicts a simplified demonstration of this theory.

Consider first the case of the fully integrated monopolist facing downstream demand $D_d$ and constant costs at both
stages given by \( d \) at the downstream stage and \( c \) for the upstream stage. That monopolist determines its profit-maximizing point of operation by setting the marginal revenue that is associated with its demand curve—MR\(d\)—equal to \( c + d \), yielding \( Q^* \) as the quantity to be sold and \( P^* \) as the corresponding price.

Next, we analyze pricing when the successive monopolies at the two stages independently seek to maximize their own profits. The final market “downstream” monopolist maximizes its profit against the same final demand curve and the same final stage marginal cost \( d \), but now faces a price for the necessary input that is higher than its marginal cost \( c \). That is because the upstream stage is itself a monopoly that, in maximizing its own profit, raises the wholesale price to its buyer above its own marginal cost. As shown in Figure 1, when the independent downstream firm now maximizes its profit, it equates the same marginal revenue curve as before to the total of its own (and unchanged) marginal cost \( d \) plus the upstream price that it now faces. The latter is greater than the upstream marginal cost, leading to a new lower output point \( Q' \) and a corresponding higher price \( P' \).

The efficiency implications of this demonstration are straightforward. Vertical integration clearly makes consumers better off since output is greater and price is lower than under successive monopoly. It is also the case that producers collectively are better off. Formal analysis proves this, but the following logic also suffices: since the integrated firm could choose any price on the final demand curve, by selecting \( P^* \) rather than the \( P' \) (or any other price), it follows that \( P^* \) yields greatest total profit. Integration, in short, results in benefits to both producers and consumers.

This straightforward proposition has represented a powerful argument favoring vertical integration in countless industries over a considerable period of time. And where it is applicable, it is correct as it stands. But as we will discuss below, there is a series of caveats, conditions, and qualifications that affect its applicability.

**Double Marginalization: The Crucial Assumptions**

There are a number of important caveats and concerns about the core proposition that favors vertical integration. This section discusses three that are part of the core theory and must be examined in all cases.

**Fixed vs. Variable Proportions.** The simple model described in Figure 1 is crucially dependent on the assumption of a fixed proportion production process at the downstream stage. This means that the downstream firm sells exactly what it buys, a common example of which is a retailer buying a specified product from its manufacturer and then selling it to final customers. When both have market power, integration in principle does indeed avoid double marginalization.

The above description of EDM does not hold, however, when the downstream stage is subject to variable proportions, for in this case, the unintegrated downstream firm can avoid some of the adverse effects of the inflated wholesale price by substituting away from use of that product. For example, if the upstream product is steel, a downstream firm may be able to use some aluminum or plastic in the production of the same final product. Price will be lower and quantity will be greater than if no substitution were possible, and indeed, both will vary with the degree of substitutability. In any case, substitution diminishes the penalty of unintegrated operation and the gain from integration, leaving the overall impact on total surplus from integration ambiguous.

As formally demonstrated by Fred Westfield, the penalty in lost profit and surplus from unintegrated operation depends on the elasticity of substitution between inputs. That penalty increases as that elasticity decreases and is maximized in the case of fixed proportions when the elasticity is zero. Estimating elasticities of substitution between inputs is similar to estimating the demand elasticities for merger simulations, which is an economic exercise often performed in the analysis of horizontal mergers.

Even in cases where fixed proportions would seem to apply, the favorable assessment of EDM might require a bit more thought since in actual practice “substitution” might not be so straightforward. Consider the above example of replacing steel with plastic in the production of some products. Although such substitution might result in the same apparent product (e.g., a “clothes washer”), the use of plastic might also result in a less durable product. That is, substitution produces a somewhat different product and not—as the textbook case assumes—the exact identical product with an alternative input.

In reality, this may be more common than the fixed-proportion and identical-product scenarios. Consider an auto manufacturer and a dealer, where the transaction seems necessarily to involve fixed proportions. Yet to the extent that the final customer also values point of sale service as part of the transaction, the downstream retailer can offer more helpful advice or greater variety of inventory in a manner that alters the relevant transaction so that it involves not simply the identical vehicle.

Whenever this more general type of substitution is possible, the determination of the efficiency implications of vertical integration is more difficult than the simple case of fixed proportions in production. Here the full and correct analysis would also need to consider the efficiency effects of subtle changes in the product itself as the producer makes strategic business decisions resulting in products with different durability, quality, or services, and with correspondingly different consumer demand. Applying the simple fixed-proportions theory where its assumption of an unchanged product does not hold cannot yield the correct answer.

**Merger Specificity.** A second important qualification to the standard EDM model is the implicit assumption that the actual cost savings require vertical integration for their realization. In theory, this is not correct, and, in practice, it is often not the case. If vertical integration were strictly neces-
sary, then we might expect nothing but a single vertically integrated structure in most or all production processes. Instead, businesses routinely develop contractual arrangements that yield the same cost benefits as integration. The mechanism for this involves nonlinear prices, quantity minimums, or other contractual provisions that move the parties to the same equilibrium as shown in Figure 1.

To illustrate, suppose—not implausibly—that the two monopolists described above realize that the key impediment to increasing both their profit streams—not just one at the expense of the other—is the distorted wholesale price. All that would be required to reach the overall profit-maximizing point is to ensure that the wholesale price is the upstream marginal cost, since the downstream firm would then be acting on the real economic cost in making its output decision. The two firms could enter into a contract that (1) specifies the wholesale price to be the upstream marginal cost, thereby maximizing total profit, and then (2) provides for a division of the now larger overall profits in a manner that increases the profits to both parties. For example, the downstream firm could pay a fixed fee—a two-part tariff—to the upstream supplier.

Such nonlinear pricing contracts are to be expected whenever the two parties recognize their interest in avoiding the pricing externality associated with independent pricing, and certainly when there is greater profit at stake. As a policy matter, therefore, vertical integration is not necessary to achieve the benefits of EDM. For this reason, in competition analysis the parties can be required to explain why contracting is impossible, since the default assumption could reasonably be that a simple contract would achieve the cost reduction without any of the practical difficulties and competition concerns associated with full integration. Indeed, the Department of Justice explained its position in Comcast/NBCU by declaring that “much, if not all, of any potential double marginalization is reduced, if not completely eliminated, through the course of contract negotiations between programmers and distributors over quantity and penetration discounts, tiering requirements, and other explicit and verifiable conditions.”

It might be argued that the contracting alternative might itself be costly and cannot fully achieve the benefits of EDM. Phillip Gayle, for example, finds that codeshare agreements among airlines that price different segments in a route independently do not fully eliminate double margins when carriers also offer competing single-product alternatives. Where incomplete internalization can be documented, the proper antitrust calculus would be to credit integration with no more than the incremental cost savings over and above what contracting can achieve. Although determining the exact incremental cost savings may not be empirically possible, crediting integration with the full benefits would be incorrect. Moreover, this calculation needs to factor in any costs incurred in the process of integration, including the negotiations and the operational integration.

Real vs. Pecuniary Economies. The third broad issue with EDM concerns cognizable efficiencies. In antitrust analysis these are limited to “real” economies, that is, those that save on actual resources. By contrast, if an input is simply repriced but no fewer units of the input are used in the production process, the lower costs are termed “pecuniary” and not credited as a “real” efficiency to merging firms. It is worth noting that EDM by itself is only a pecuniary economy: in the model described in Figure 1 there is no unit of output that is produced with fewer resources under vertical integration than with independent operation of the two stages. By itself, therefore, EDM does not result in a “real” cognizable economic efficiency.

That said, as is also evident in Figure 1, EDM does alter the market equilibrium in a direction that improves economic efficiency. The reason is that, as a direct result of the decrease in pecuniary cost, output has increased. This output increase reduces deadweight loss while increasing total profit and surplus to consumers. But the true economic benefit derives entirely from output expansion, not from a decrease in unit input usage. In this respect, labeling EDM as a cost-reducing benefit is misleading, if not incorrect, and more importantly, it overlooks significant differences in the economic benefits that may arise. A reduction in unit input usage directly results in real cost savings that accrue to some combination of firms and consumers. By contrast, a reduction in pecuniary costs per se results in no real cost savings, although in this case it induces an output increase, a lower price, and a deadweight loss reduction. It is, however, noteworthy that the first effect works through supply whereas the second works through demand.

Does this distinction make any difference? In both cases, the new industry equilibrium entails greater efficiency, so viewed in that way it appears largely a matter of semantics. But as we discuss later, it is possible that in practice the internal transfer price to the downstream division might not equal marginal cost, and thus, the efficiency gains from the EDM in terms of the reduction in pecuniary costs may not be fully realized.

Finally, as a practical matter, antitrust authorities normally deal with horizontal pricing externalities, such as changes in markups, which could mean that they, not the parties, would deal with vertical pricing “efficiencies.”

Double Marginalization: Variations in Specific Settings

The above qualifications are central to the theory of EDM in the sense that they apply uniformly and should be considered in all cases where vertical integration is claimed to be justified by such cost savings. Next discussed are other more specific circumstances in which the simple model and its implications may not hold.

Non-Monopoly Markets. The classic model of double marginalization outlined above rests on the assumption of two-stage monopoly: an upstream monopolist sells its output
to a downstream monopoly. While there may be such actual cases, most of the commonly cited examples and applications involve markets with imperfect competition at either or both stages, and the implications for double marginalization are quite sensitive to the market structures at both stages. Michael Salinger’s model serves to illustrate one possibility. That model posits successive Cournot competitors, so that a vertical merger reduces competition in the upstream market even as it creates a more aggressive downstream firm. The net competitive effect depends on several parameters, such as the integrated firm’s elasticity of supply of the final good and the elasticity of the residual demand facing the unintegrated final good producers. Often, those parameters are not readily available.

Other more subtle scenarios may also disrupt the conclusions of the classic model when markets are not monopolies. One such scenario involves the case of an upstream monopolist that integrates with a downstream firm that has rivals. If its margin on sales to the independent rivals exceeds the margin on its supply to its integrated downstream affiliate, the integrated firm will face an opportunity cost to supplying its affiliate. That, in turn, will reduce its internal usage of the input, limiting the gain in the total surplus that would otherwise be expected from EDM. And in the limit, of course, there will clearly be no gain from EDM if the upstream industry is competitive.

Concerns over the confounding effects of competition at either stage led Paul Joskow, in his review of the economics of vertical integration, to conclude that when vertical integration affects the intensity of competition at either stage, "This will in turn affect the incentive to vertically integrate, the distribution of profit between firms from vertical integration, and consumer prices. Moreover, the welfare effects are now more likely to be ambiguous." The implications from the simple two-stage monopolist model are not robust to these generalizations of market competitiveness.

**Multiproduct Firms.** The basic model described above involves single-product monopolies at each stage and, subject to the other qualifications, predicts that product price will fall after vertical integration. However, Salinger has shown that this implication is in fact crucially dependent on the assumption that each firm is not only a monopoly but also a single-product firm. The case of multi-product firms, which is quite common, is considerably more ambiguous in its overall efficiency effects. Salinger cites mergers involving soft drink bottlers and cable television as examples of multi-product firms where the implications are in fact likely to differ.

Salinger’s demonstration is an application of Edgeworth’s Paradox of Taxation. Edgeworth had shown that a tax on one of two substitute goods sold by a monopolist can in fact lead to lower prices for both. In the present context, Salinger shows that the effect on the price of a good from vertical integration can be quite different when that good is a substitute in final demand for a non-integrated product of the downstream firm (as with a vertical merger between Coca-Cola and its bottlers when the bottlers also bottle the unintegrated Dr Pepper brands). Indeed, as he demonstrates, both prices can fall, both prices can rise, or the price of the now-integrated product can fall while the other one rises. In particular, the fact that the downstream demand facing the integrated firm is joint is crucial. Indeed, this fact implies that, by raising the prices of the unintegrated brands (e.g., Dr Pepper brands), demand is shifted towards the integrated substitutes (e.g., Coca-Cola brands) and, if the shift is sufficiently large, all prices can rise.

Salinger analyzes the efficiency effects of a vertical merger involving such a multi-product firm and cautions that the overall efficiency effect is an empirical question that can be difficult to answer, but it is not answered by appeal to theory that is designed for a different setting.

**Strategic Behavior and Opportunity Cost.** As noted previously, pure monopoly in both stages is rare, and one particular alternative market structure has been especially important in recent vertical mergers. This involves the case where the merger combines an upstream monopolist with one of its downstream customers, which means that other downstream players must now acquire a crucial input from a firm that competes with them. Having both the ability and incentive to disadvantage its direct downstream rivals, the integrated firm will predictably raise the input price in order to weaken the competitive force of its downstream rivals.

This strategy of raising rivals’ costs is well understood, but its implications for double marginalization are not. The reason is that in raising the input price to its rivals, the integrated firm—apart from any efficiencies from integration—stands to lose some revenues and profits on forgone sales to those rivals. The opportunity cost of vertical integration in this setting was recognized as an offset to the merging parties’ argument before the FCC in the proposed Comcast/NBCU merger. Steven Salop and Daniel Culley have shown how the magnitude of this opportunity cost can be factored into a broader calculation of the price effects of such a vertical merger.
Gopal Das Varma and Martino De Stefano have recently formalized this offsetting effect.\textsuperscript{20} They note that standard analysis of vertical mergers between oligopolists has involved a two-stage process: first, the competitive risk that the integrated firm will engage in a strategy of raising its rivals’ costs (RRC) is assessed, and second, if that risk is credible and the effect substantial, the cost savings from EDM is measured as an offset. They make the fundamental point that these effects are not separable. Rather, the magnitude of EDM affects the shift in demand away from the integrated firm’s downstream rivals, while the shape of their reduced demand curves in turn affects the integrated firm’s incentive to engage in a strategy of raising rivals’ costs. Indeed, depending on the shape of the demand curve for the unintegrated downstream product as a function of the upstream input price, the gains from RRC can increase or decrease and, with it, the magnitude and even the direction of the change in the upstream price charged to rivals.

In addition, they show that EDM not only affects the size of RRC, it can also change its sign. With take-it-or-leave-it price offers upstream, a sign reversal—lowering rivals’ costs—is more apt to occur when the price elasticity of rival demand is high and the diversion rate from the rival to the integrated downstream division is low. The former occurs because the monopolist loses more sales when it raises its price to its competitor and the latter occurs because the customers that leave the rival are less apt to purchase from the integrated firm.

Das Varma and De Stefano’s model is sufficiently flexible to accommodate differences in bargaining power of the upstream vis-à-vis the downstream unintegrated firm, with take-it-or-leave-it offers a special case. This flexibility, in conjunction with differences in demand curves—for example, linear vs. logit—leads to a variety of possible outcomes. They note that the price increase inflicted on rivals falls, often by a substantial amount, as the bargaining power of the upstream firm rises, with the limiting case being take-it-or-leave-it offers upstream. Moreover, increases in the price charged to a downstream rival are smaller under linear demand compared to logit. Finally, retail prices also depend on equilibrium cost pass-through. These substantial ambiguities lead Das Varma and De Stefano to conclude that “the standard technique—that does not account for the link between EDM and RRC—can significantly miss the mark when it comes to predicting price effects.”\textsuperscript{21}

\textbf{Internal Transfer Pricing.} The efficiency gains from EDM depend critically on the assumption that an integrated firm will set an internal transfer price—the implicit price at which a unit is transferred from upstream to downstream—equal to marginal cost. In other words, full information and frictionless profit maximization is assumed. However, there are reasons why this might not always be the case.

There is a theoretical literature that argues that operating an integrated firm as a set of separate divisions can be efficient because the divisions will compete fiercely among themselves to lower costs.\textsuperscript{22} On the other hand, when separate divisions make separate pricing decisions, they might not fully internalize the cost of their pricing decisions to other divisions. Principal-agent models highlight the differences between the objectives of owners and managers of the firm that can imply deviations from cost minimization. Moreover, a related organizational literature cautions that vertical integration is responsive to verifiable costs but may impose a variety of private, non-verifiable costs that are overlooked.\textsuperscript{23}

There is also some empirical evidence of these transfer pricing issues. In an empirical study of vertical integration in multi-channel television markets, Gregory Crawford and others find that divisions internalize most but not all of the profits of other units—specifically, divisions internalize $0.79 of each dollar of profit realized by other integrated units.\textsuperscript{24} In addition, Christian Michel and Stefan Weiergraeber find that, after a merger in the ready-to-eat breakfast cereal industry, it takes time for internal profit maximization to occur, but it is eventually realized.\textsuperscript{25} These studies imply that the assumption of efficient internal pricing might not be realistic or that it will not occur immediately, both of which factors lessen the size of the EDM effect.

Finally, there will be no internal transfer price, and thus no EDM, if the downstream firm does not use the upstream product. This might seem like an unusual situation. However, Enghin Atalay, Ali Hortacsu, and Chad Syverson show that no shipments between vertically integrated firms is the rule, not the exception.\textsuperscript{26}

As with the other factors discussed above, it is difficult to forecast how the integrated firm will price post-merger. At a minimum, however, a sensitivity analysis that varies the degree of internalization of cost minimization is warranted.

\textbf{Conclusion}

In recent years, competition and regulatory policies have finally begun to assess the possible adverse effects that can be associated with vertical integration and mergers. That concern is well founded and long overdue. But policy analysis has continued to treat the claimed benefits from EDM relatively uncritically, too often automatically crediting vertical mergers with the cost saving benefits predicted by the classic economic model.

This can be a critical policy error since the classic EDM model is based on a long list of assumptions that do not necessarily hold, including that production is characterized by fixed proportions, contracting cannot achieve similar results, pecuniary economies cause real output changes, pure monopoly exists both upstream and downstream, upstream marginal cost becomes the operative transfer price, each firm sells only a single product, and raising rivals’ costs does not offset the benefit. Although some of these are well understood, the full array of necessary assumptions tends not to be recognized.

To illustrate, in a recent speech, the Assistant Attorney General for Antitrust stated that vertical integration would eliminate double marginalization, and hence get antitrust
credit, when both firms mark-up price over marginal cost, when contracting is not possible, and when consumer demand makes the effect sizeable. This list of conditions overlooks several additional crucial assumptions. The Draft Vertical Merger Guidelines recently released by the Department of Justice and Federal Trade Commission similarly elevate the elimination of double marginalization as an efficiency defense for a vertical merger, again with an incomplete listing of the limitations of the proposition.

There is no dispute that empirical studies show that vertical integration can reduce costs and yield benefits to both consumers and producers. However, most of these studies involve competitive or monopolistically competitive settings rather than the double monopoly structure that is the foundation of the EDM implication, and there is in fact little empirical evidence about the outcome of vertical integration in imperfect market settings where policy is most concerned. As a result, we believe that some vertical mergers do indeed warrant close scrutiny. Our concern is that, in evaluating such mergers, the calculation of EDM and its consequences is often simplistic and flawed. Moreover, we also believe that the heavy focus on EDM as the principal source of efficiency is misplaced. Rather, the efficiencies that should matter most are traditional supply-side efficiencies, such as economies of scope, elimination of duplicate facilities, and the ability to coordinate other aspects of the vertical chain.

4 Joe Palazzolo et al., Decoding Judge Leon’s AT&T-Time Warner Decision, WALL ST. J. (June 12, 2018).
6 See, e.g., id. at 688 (concluding: “Only in the case where [price] falls can we be sure that total surplus is increasing. In the opposite case, total surplus can increase or decrease.”) For a more complete analysis, see Fred M. Westfield, Vertical Integration: Does Product Price Rise or Fall?, 71 AM. ECON. REV. 334 (1981).
7 Westfield, supra note 6.
8 For this reason, industrial organization textbooks typically discuss vertical integration as one method, along with spot markets and contracting, for mediating the transaction between upstream and downstream firms. See, e.g., Dennis Carlton & Jeffrey Perloff, MODERN INDUSTRIAL ORGANIZATION ch. 12 (4th ed. 2000).
14 Paul Joskow, Vertical Integration, 55 ANTITRUST BULL. 545, 554 (2010).
17 Salinger, supra note 15, at 545.
21 Both Das Varma and De Stefano, and Domnenco and Sibley, show how simulations can be used to forecast EDM and RRC when they are determined simultaneously. In practice, such simulation models would have to incorporate assumptions on demand and bargaining that are consistent with the structure and practices in the industry that is studied. See id. at 19; Gleb B. Domnenco & David S. Sibley, Simulating Vertical Mergers and the Vertical GUPPI Approach (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3447687.
29 Francine Lafontaine & Margaret Slade, Vertical Integration and Firm Boundaries: The Evidence, 45 J. Econ. Literature 629 (2007). See also Margaret Slade, Vertical Mergers: Ex Post Evidence and Ex Ante Evaluation Methods (Microeconomics.ca Working Papers, Vancouver School of Economics, June 2019). In addition, the existing studies generally do not consider possible quality effects or the contracting alternative.