WORKING PAPER · NO. 2022-38

Personal Increasing Returns: Analytics and Applications

Casey B. Mulligan

MARCH 2022

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

PERSONAL INCREASING RETURNS:
ANALYTICS AND APPLICATIONS

Casey B. Mulligan

March 2022

I appreciate financial support from the University of Chicago's Initiative on Enabling Choice and Competition in Healthcare. This paper is part of a project of adding new chapters to Chicago Price Theory, which contains Kevin M. Murphy's Economics 301 lectures. Those lectures and further recommendations from Murphy strongly influenced the preparation of this paper.

© 2022 by Casey B. Mulligan. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Personal Increasing Returns: Analytics and Applications
Casey B. Mulligan
March 2022
JEL No. D11,I12,J24

ABSTRACT

The human capital investment model with endogenous labor supply is generalized to consumer and health behaviors while retaining the tractability of comparative-static analysis of a single first-order condition. Accounting for the endogenous specialization responses is essential to properly distinguish supply and demand factors and to understand how the magnitude of their effects vary across time and circumstances. Even signing effects of policy interventions can hinge on the existence and extent of personal increasing returns. Applications include the gender gap in earnings, the dynamics of substance abuse, effects of taxes on human capital, the tradeoff between product quality and quantity, and unintended consequences of energy regulation. Metrics are provided for assessing the extent of personal increasing returns.

Casey B. Mulligan
University of Chicago
Department of Economics
1126 East 59th Street
Chicago, IL 60637
and NBER
c-mulligan@uchicago.edu

I. Introduction

Although excluded from the prototype supply and demand analysis, increasing returns at the individual level are evident in many domains. The average consumer purchases only a narrow selection of available products while the average worker performs only a narrow range of tasks. Accounting for the endogenous specialization responses that loom large in many applications is essential to properly distinguish supply and demand factors and to understand how the magnitudes of their effects vary across time and circumstances. Even signing effects of policy interventions can hinge on the existence and extent of personal increasing returns.

The purpose of this paper is to generalize the familiar human capital investment model to consumer and health behaviors while retaining the tractability of comparative-static analysis of a single first-order condition. Applications include the gender gap in earnings, the dynamics of substance abuse, effects of taxes on human capital, the tradeoff between product quality and quantity, and unintended consequences of energy regulation.

Section II models the consumer of a specific good who can access lower prices at a cost. The more he intends to consume, the greater the incentive to obtain a low price. The low price itself encourages additional consumption, resulting in larger behavioral responses than would occur under constant returns. By considering decision making in two stages, I show how the model is isomorphic with the prototype supply and demand model (taken as a model of individual optimization), except that supply slopes down.

Section III shows how the consumer model is also isomorphic with the familiar model of labor supply in which the worker has the opportunity to make human investments that increase her wage. Here the supply (opportunity cost) of leisure slopes down in the amount of leisure, which amounts to a demand for labor that slopes up. Labor supply shifts thereby generate a positive correlation between wages and work, much like labor demand shifts do.

Several applications, and the metrics they offer for assessing the degree of increasing returns, are cited in Section IV. Section V focuses on the case where supply and demand cross multiple times, where more may be consumed in response to an upward shift in supply. This appears to violate the law of demand, until we recognize that endogenous choices can move marginal prices in the opposite direction as the changes that would occur with behavior fixed.

Consumers encounter many pairs of goods and services that are complements.[1] Usually each member of a complementary pair of goods is modeled as a separate argument in a utility function that is quasiconcave in the relevant range, which rules out nonlocal kinds of substitution behavior emphasized in Section V.[2] The household-production setup in this paper, and human capital models generally, are special in that one of the complementary pair is a nonrival input, which generates the increasing returns.[3] Our approach also has additional analytical tractability in that (local and nonlocal) comparative statics are derived from a single "supply-demand" first order condition. A common flavor to the two approaches, though, is that substitution and income effects are amplified by adjusting the purchases of complements.[4]

II. Increasing Consumption Returns

The model features a consumer with income y and quasiconcave utility $u(c,q)$ depending on the consumption q of a specific good and all other goods c. The price of c is normalized to one. The marginal rate of substitution in u, in units of the price of q, is denoted $M(c,q)$. The price of q is $p(h) > 0$, with $p'(h) < 0$ and $p''(h) > 0$. h denotes a costly action – hereafter "investment" – taken by the consumer to reduce the price of q, albeit at a diminishing rate. The endpoints of the pricing function satisfy $p(0) > p(y) > 0$.

[1] Becker and Murphy (1993, 2003).
[2] Becker and Murphy's (1988) rational addiction model is one example of complements generating increasing returns; they conclude that addiction behaviors have multiple steady states.
[3] Even without increasing returns, the household production approach to complements has additional predictive power in that the effect of one good on the value of the other can be measured in principle.
[4] Jaffe, et al. (2019, Chapter 6).

Consumer expenditures on c, h, and q are limited by income y. The demand system is the solution to (1):

$$\max_{h,q} u(y - p(h)q - h, q) \tag{1}$$

As discussed further below, h is a nonrival input into delivering the consumers services associated with q because h has the same opportunity cost and obtains the same price reduction regardless of how many units of q are purchased. In some applications, h is literally an out-of-pocket cost fungible with income as modeled in (1). Other times h indicates foregone "quality" attributes to the extent that part of their opportunity cost is independent of the quantity consumed.

II.A. A Two-Stage Solution

The demand system (1) can be described in two stages. Let $E(q)$ denote the "first stage" minimum cost to obtain a given level of q and $h(q)$ the amount of investment attaining that minimum:

$$E(q) \equiv \min_{h \geq 0}\{h + qp(h)\} \tag{2}$$

The shape of $E(q)$ is essential for describing the second stage. Its first and second derivatives come from the envelope theorem and the first order condition for minimization:

$$E'(q) = p\big(h(q)\big) > 0 \tag{3}$$

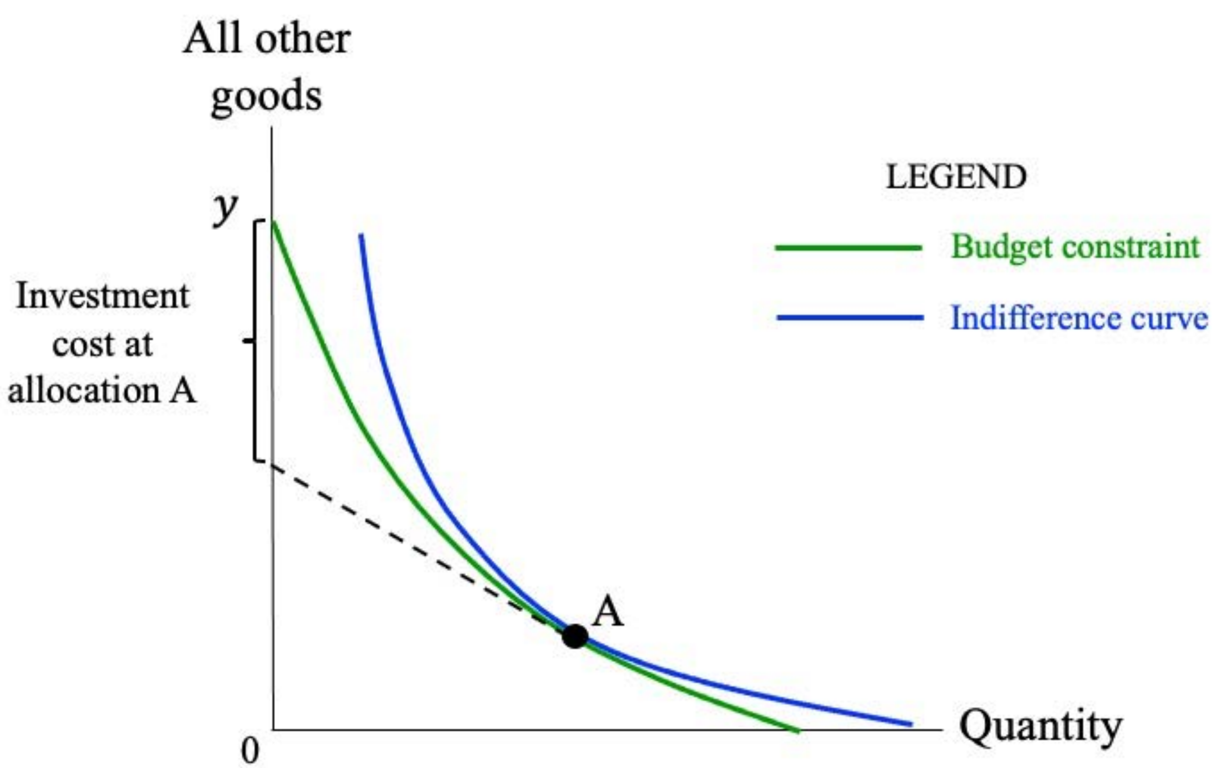
$$E''(q) = p'\big(h(q)\big)h'(q) \leq 0 \tag{4}$$

where the inequality is strict if nonnegative investment constraint does not bind. It may bind for q sufficiently close to zero, in which case $E(q)$ is linear increasing. $h(q)$ is increasing everywhere that its nonnegativity constraint does not bind. That is, h and q are complements.

The remaining results of the paper follow from the strict concavity of the first-stage expenditure function $E(q)$ over some range. The non-rival input specification (2) is a compact and often relevant way of obtaining strict concavity, but is not required. We could replace in (2) the investment cost h with $hg(q)$, with $g(q) > 0$ and $g'(q) > 0$ reflecting the alternative assumption that h costs more if it is applied to more units of q. As long as $g(q)$ is strictly concave, $E(q)$ will be strictly concave too. With that noted, the investment cost throughout this paper is taken to be just h.

The budget constraint $c + E(q) = y$ can be drawn in the $[q,c]$ plane beginning at $q = h = 0$, $c = y$, and $E(0) = 0$. The rest can be drawn from there by moving away from $q = 0$ following the mirror of E, as shown by the green budget constraint in Figure 1. The choice set is not convex, reflecting increasing returns in the consumption of q. The amount of investment achieving a given allocation $\{q,c\}$ can be calculated algebraically as $y - c - qE'(q)$ or geometrically from the intercept of the tangent line going through the allocation.

Figure 1. Personal increasing returns in the commodity space; consumer version



The second stage of the consumer's problem (5) describes the consumer's optimal choice of an allocation on the convex budget constraint shown in Figure 1:

$$\max_q u(y - E(q), q) \tag{5}$$

This second stage resembles the canonical two-good choice subject to a downward-sloping budget constraint, except that the constraint is convex. Therefore, the second stage maximization (5) may not be concave or have a unique solution. Assume for the moment that the first-order condition describes a unique optimum:[5]

$$M(y - E(q), q) = E'(q) \tag{6}$$

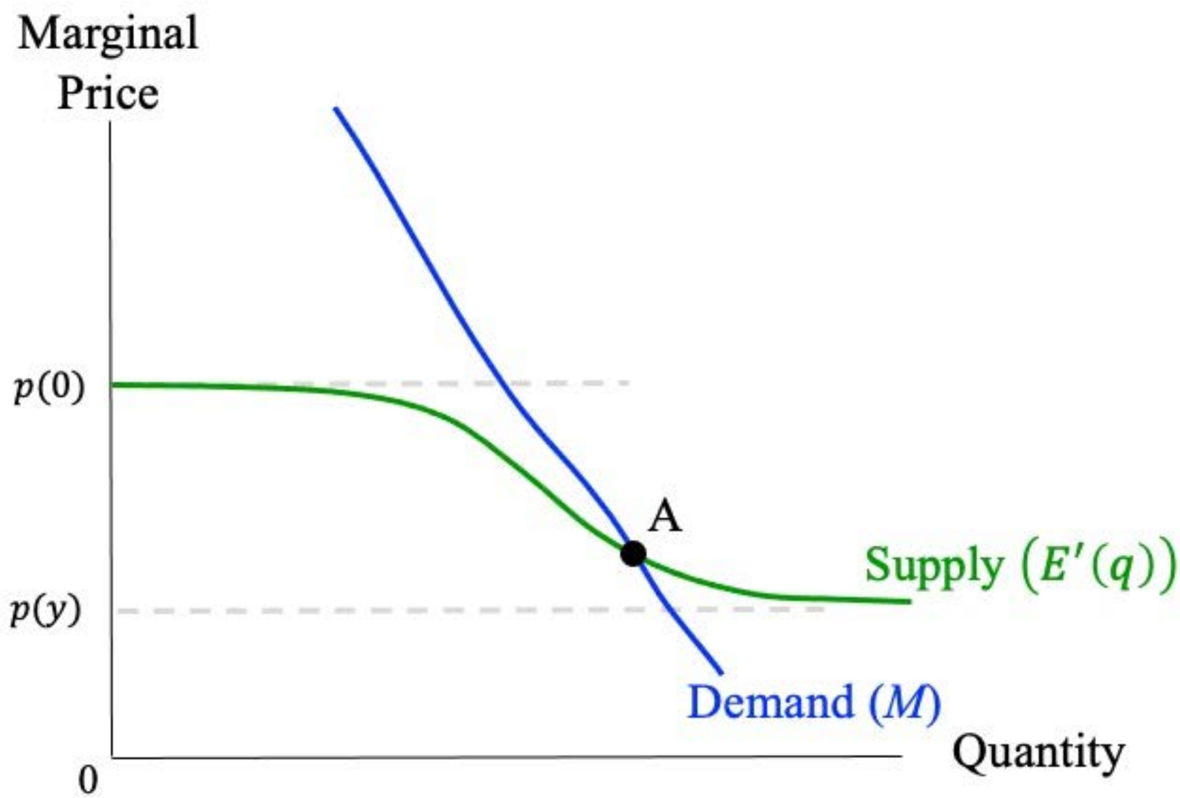
Equation (6)'s LHS "demand" slopes down in q because u is quasiconcave. The RHS "supply" also slopes down, but the second-order condition requires that it cut demand from below.[6]

II.C. Comparative Statics

Figure 2 displays the optimum in the case that the second-order conditions are satisfied, as witnessed by the fact that the demand curve is cut by supply from below. An *increase* in demand would reduce price, which increases quantity by more than it would if price where held constant. The amounts of price reduction and additional quantity are greater with more price-elastic demand or supply that slopes down more.

[5] The uniqueness assumption is dropped in Section V.
[6] The local second-order condition is $dM(y - E(q), q)/dq < E''(q)$.

Figure 2. Personal increasing returns
in quantity-price space



To formalize the comparative statics, let *dm* denote the magnitude of an addition to the marginal rate of substitution function (an additive wedge added to the FOC (6)):

$$d \ln q = \frac{\varepsilon_d \varepsilon_s}{\varepsilon_d - \varepsilon_s} \frac{dm}{p} \tag{7}$$

$$d \ln E'(q) = \frac{\varepsilon_d}{\varepsilon_d - \varepsilon_s} \frac{dm}{p} \tag{8}$$

where $\varepsilon_d < 0$ and $\varepsilon_s < 0$ are the inverse of the local elasticities of the LHS and RHS of (7) with respect to q, respectively.[7] The second order conditions of the consumer's problem require $\varepsilon_d - \varepsilon_s > 0$.

Equation (7) is the usual equilibrium incidence formula except that the supply elasticity has the opposite sign from the prototype case.[8] The coefficient on dm/p is positive as it is in the prototype case but has greater magnitude than ε_d, rather than less. In words, supply amplifies the quantity impact of a demand shift rather than choking part of it off. The amplification can be arbitrarily large by having a supply curve that slopes down enough to nearly violate the second order condition (that is, slope down almost as much as demand does). Put yet another way, the conventional supply and demand model bounds the magnitudes of quantity and price changes from above, whereas the increasing returns model bounds them from below. The increasing-returns model is supply and demand on steroids.

Table 1 shows equation (7)'s coefficient on dm/p as a ratio to the linear-budget case (where the coefficient would be $-\varepsilon_d > 0$) for various values of $\varepsilon_s/\varepsilon_d$. The first row shows how the amplification can be arbitrarily large for ε_s close enough to ε_d. When supply is twice as price elastic as demand, quantity responses are twice as large as they are in the linear-budget case. The linear case is approached as ε_s becomes infinitely elastic.

---

[7] ε_d is also the local price elasticity of the Hicksian demand for q. Although, viewed as a function of q, M is not the inverse of Hicksian demand, the two curves both coincide and are tangent at a local optimum.

[8] Also, supply and demand are used here to describe an individual's optimum rather than a market equilibrium although in some cases equilibrium and individual optimum have a lot in common.

Table 1. Quantity-change amplification in the personal increasing returns model

$\varepsilon_s/\varepsilon_d$	$\varepsilon_s/(\varepsilon_s - \varepsilon_d)$
1	∞
1.25	5
1.5	3
2	2
5	1.25
∞	1

Note: ε_d and ε_s are the price elasticities of demand and supply, respectively, with $\varepsilon_s < \varepsilon_d < 0$.
This formula applies to local comparative statics.

Increasing returns also have sharp qualitative predictions. Quantity and marginal price cannot change in the same direction without both curves shifting. Even in that case, the vertical shifts in supply and demand must be in the same direction and in similar magnitudes but with somewhat more demand shift. In this sense, quantity and marginal price changes in the same direction are relatively rare. If this prediction is overlooked, it can be easy to confuse supply and demand, as illustrated in Section III.

The effects of supply and demand impulses are unlikely to be constant over time or across markets because small-scale consumption does not justify investments in price reductions and there are technological limits on how much even the highest-quantity consumer can reduce his price. Recall Figure 2's mirror-S-shaped supply curve. As incomes and other demand factors trend over time, a market may emerge from the flat part of the supply curve into the steeper part and then later proceed to the high-quantity flat part. Without accounting for increasing returns, the magnitude of effects estimated in one phase are not good indicators of effects to be observed in the next. The changes would resemble the "cascades" from models of information or social interactions, even though neither of those forces are present in this model.

III. The Supplies of Labor and Human Capital

The prototype human-capital investment model has exactly the form (5) and (6). To obtain this result, interpret q as leisure time, $1 - q$ as work time, y as the available income when

q = 1, and $p(h)$ the wage rate, with $p'(h) > 0$ and $p''(h) < 0$.[9] The budget constraint, demand system, and expenditure function are (9), (10), and (11), respectively:

$$c + h = y + p(h)(1 - q) \tag{9}$$

$$\max_{h,q} u(y - p(h)(q - 1) - h, q) \tag{10}$$

$$E(q) \equiv \min_{h \geq 0}\{h + (q - 1)p(h)\} = -\max_{h \geq 0}\{(1 - q)p(h) - h\} < 0 \tag{11}$$

This human-capital version of the model has an expenditure function that embeds an extra term, a function $p(h)$ with a different shape, and a cost-minimizing investment $h(q)$ that decreases with q. h is a complement with a bad rather than a good. The expenditure function $E(q)$ is negative rather than positive. Nevertheless, $E(q)$ still has the same shape as before: increasing everywhere and strictly concave whenever the nonnegative investment constraint does not bind. Therefore, the second stage problem is (5) regardless of whether we are interested in the human capital version or the consumer version.[10] The first order condition (6) describes leisure demand $M()$ and the marginal opportunity cost of leisure E'.[11]

Either interpretation of personal increasing returns can be extended to include tax wedges and other public policy variables. Here I consider a flat-rate τ labor-income tax whose base excludes a fraction $\phi \geq 0$ of the investment expense h from the tax base. The first-stage expenditure function changes from (11) to (12):

$$E(q) \equiv -\max_{h \geq 0}\{(1 - \tau)(1 - q)p(h) - (1 - \phi\tau)h\} \tag{12}$$

---

[9] See Becker and Ghez (1975) and Heckman (1974). Heckman (1976) also emphasizes effects of income taxation on human capital through the equilibrium interest rate.

[10] Drawing the human-capital budget constraint in the [q, c] plane begins at $q = 1$ (zero labor), where the $h \geq 0$ constraint binds, $E(1) = 0$, and $c = y$. The investment amount supporting an allocation is calculated algebraically as $y + (1-q)E'(q) - c$, which geometrically is the vertical distance between (i) the intersection of Figure 1's tangent line with $q = 1$ and (ii) the point $\{1,y\}$.

[11] An equivalent alternative perspective on (6) is to treat M as an upward sloping supply of labor and E' as upward sloping labor demand.

Clearly, conditional on the time worked $1-q$, investment h and pre-tax wages $p(h)$ are invariant to the tax rate as long as the full investment costs is deductible from wages for the purpose of taxation ($\phi = 1$) and the same marginal tax rate applies at the time of the investment as at the time of earnings. This is the "conventional wisdom" about taxes and human capital dating back at least to Goode (1962), Becker (1964), Boskin (1975).[12]

The conventional wisdom is just a first step because work time is also a choice variable. The second step is sometimes dismissed based on the assertion that work time is approximately invariant to taxes due to offsetting income and substitution effects. But this is a result of the framework and does not change the fact that the budget constraint is nonlinear, not to mention that the assertion fails in important instances. Modeling tax changes as multiplicative shifter A of the expenditure function $AE(q)$ in the neighborhood of $A = 1$ as well as possible changes in nonlabor income, comparative statics derived from the first order condition (6) are (13):

$$E(q)\frac{M_c}{M}\frac{dy}{E(q)} + \frac{\varepsilon_s - \varepsilon_d}{\varepsilon_d \varepsilon_s} d\ln q = \left[1 + \frac{E(q)}{M} M_c\right] d\ln A \tag{13}$$

where the M subscripts indicate partial derivatives of the marginal rate of substitution function. The ε subscripts refer to the same elasticities from equation (7), which in this case refer to the supply and demand for leisure.

The RHS of (13) has a positive substitution term followed by an income term $E(q)M_c/M$, which is negative when leisure is a normal good. These two terms may exactly cancel, as they do with (i) $y = 0$ and (ii) a unit elasticity of the marginal rate of substitution function with respect to c. But when they do not cancel, or the dy term is not zero, the comparative statics are amplified relative to what they would be with a linear budget constraint because equation (13) has the same amplification term multiplying $dlnq$ as in equation (7).[13]

Marginal tax rates implicit in many public assistance programs are an important case where elevated tax rates are associated with an income effect encouraging leisure rather than discouraging it (Mulligan 2012). Here we not only expect marginal tax rates to reduce labor supply (increase leisure) at a given wage, but that wages would fall as human capital is not

---

[12] The "conventional wisdom" quote is from Heckman (1976).

[13] The linear budget constraint case is the limit as ε_s approaches $-\infty$.

maintained, which further discourages work especially to the extent that part of the additional lost earnings are replaced by additional assistance from public programs or even private sources. Sargent and Ljungqvist (1998) argue that the mutual feedback between work and human capital was an important driver of postwar European unemployment. More recently, Mulligan (2022) finds increased consumption of dangerous drugs among working aged adults in the U.S., especially during the two episodes of elevated unemployment benefits.

For the same reasons, shifts in labor supply may not change wages in the opposite direction. This is often recognized in analysis of female labor markets, where increased labor supply of women over time is believed to have contributed to the accumulation of work experience, educational credentials, and other work-related skills that raise women's wages.[14] However, this is sometimes forgotten in the context of male labor markets where, for example, both wages and employment rates fell over time for less educated men. Anne Case and Angus Deaton (2020, p. 70) dismiss the role of labor supply "[i]f people are withdrawing their labor, wages should rise; but in the late part of the twentieth century and into the twenty-first, wages fell along with employment, a clear indication that the problem lies with falling demand, not falling supply."[15] In the personal increasing returns view, drug addictions would not increase wages as the consumers withdraw their labor but rather would decrease them as "leisure" depletes health and other human capital.

IV. The Prevalence of Quantity Discounts

Adding the possibility of downward-sloping supply without attention to its origins would reduce the predictive power of economic theory by adding degrees of freedom to *ex poste* rationalize observed behavior and prices. The derivations in Section II provide additional

[14] Becker (1985); O'Neill and Polachek (1993); Goldin and Katz (2002); Greenwood, Seshadri and Yorukoglu (2005); Mulligan and Rubinstein (2008).
[15] Their supply-demand logic is flawed even in the conventional model (Mulligan, Bradford, et al. 2018). For example, a market with a nearly vertical supply curve has essentially all of its quantity changes explained by supply regardless of whether prices rise or fall.

empirical implications as well as providing metrics for assessing whether and how much supply slopes down.

Scarce factor supplies can render the supply of q upward sloping at a market level even if it is downward sloping at an individual level.[16] This is one reason why personal increasing returns might be more visible when comparing market participants to each other – such as men and women in the same labor market – than comparing across markets. Industries also differ in the degree to which their activity feeds back on input prices, often with inputs supplied more elastically to consumer-goods manufacturers than to agricultural or mining industries.

Because the supply curve $E'(q)$ varies between $p(y)$ and $p(0)$ in the price dimension, $p(0) - p(y) > 0$ is an index of the amount of personal increasing returns in the model and the degree to which Figure 2's supply curve slopes down. Human capital applications have a high potential for increasing returns because individuals have considerable scope to affect their wage in either direction. By some estimates, migration from Mexico to the U.S. quadruples hourly wages (Aguayo-Tellez and Rivera-Mendoza 2011). The median holder of a professional degree earns 60 percent more than the median college graduate, who earns 75 percent more than the median high school graduate.[17] Five years of job tenure increase hourly wages about 30 percent. Failure to maintain health – regular consumption of dangerous illicit drugs is an example – can significantly reduce wages.

Quantity discounts are common for manufactured goods, with high-volume consumers sometimes cutting their price in half.[18] Mulligan (2020, 2022) applies the personal increasing returns model to recreational consumers of drugs and alcohol, where prices can vary by a factor of three or more. Motorists can reduce their per-mile fuel expenses by a factor of two by purchasing a smaller and more fuel-efficient car. Some of the foregone attributes of the larger car, such as the option of moving additional passengers or luggage or maintaining traction during inclement weather, may be at least partly valued independent of the number of miles to be

[16] In the notation of this paper, scarce factor supplies could be included by allowing the pricing function $p(h)$ to depend on q as well as h.
[17] These are earnings and therefore a combination of hours and wage effects. With a wage-elasticity of hours equal to one, the change in log earnings is twice the change in log hourly wages.
[18] Schotanus, Telgen and de Boer (2009). See also Murphy, Snyder and Topel (2014).

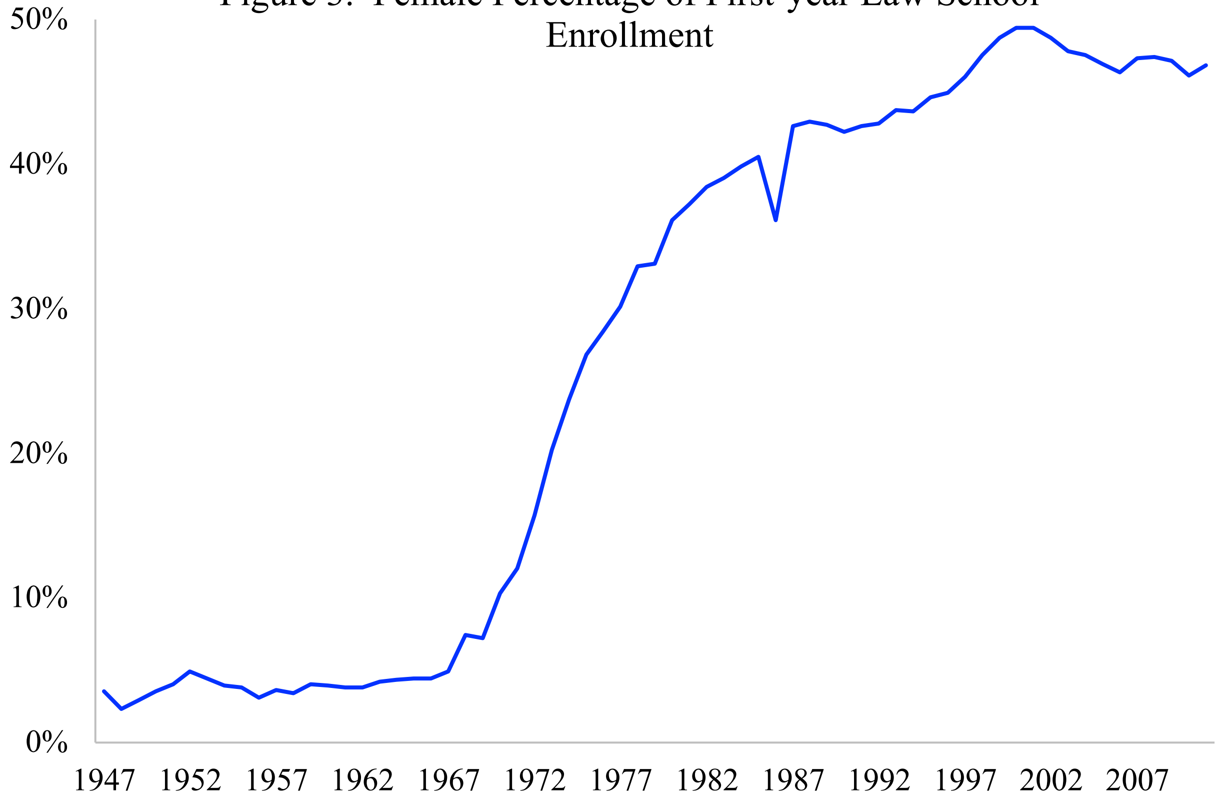
driven. Indeed, many goods have a quality dimension whose value to consumers is partly independent of the quantity consumed; consumers can "purchase" lower prices by foregoing quality.

With downward-sloping supply, a vertical supply shift and any demand shift has a price effect that reinforces the quantity change. The amount of the reinforcement is more when demand is price elastic and would be zero if demand were vertical. Among the human capital applications, the reinforcement would be greater when leisure demand is wage elastic (that is, labor supply is wage elastic), as it is for married women. Personal increasing returns are Becker's (1985) explanation of why quantities and wages increased so much over a short period of time in the married-female labor market, where labor supply is known to be more wage elastic than it is in other labor markets.

Figure 3 shows an example from U.S. law schools.[19] The percentage of women enrolling remained below five percent for decades. Beginning in 1968, the percentage increased six fold within ten years. After 1980, the changes were much less. Complementary behaviors may also help explain why time allocation and consumer spending changes so rapidly as male and female workers pass through "retirement ages" (Kopecky 2011).

[19] Statistics refer to schools approved by the American Bar Association (https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/statistics/jd_enrollment_1yr_total_gender.authcheckdam.pdf).

Figure 3. Female Percentage of First-year Law School Enrollment



Source: American Bar Association

V. Substitution Behavior in the "Wrong" Direction

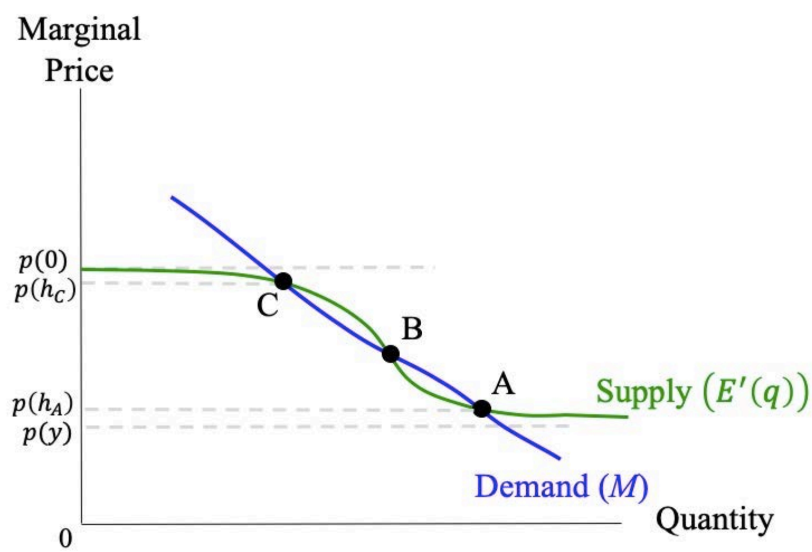
The first-order condition (6) may not be sufficient to describe the optimum to the second stage maximization (5), which may not be unique. Figure 4 shows three allocations where (6) is satisfied, as witnessed by the fact that supply and demand cross three times. The second-order condition fails at *B* because supply crosses demand from above. Whether local optimum *A* is preferred to local optimum *C* depends on the comparison of the area to the left of the Hicksian demand curve and between $p(h_A)$ and $p(h_C)$ to the additional investment $h_A - h_C$ required to obtain the lower price.[20]

[20] Figure 3 does not show the Hicksian demand curve, but it is tangent to *M* at both allocations A and C. See the Appendix for a numerical example.

Now consider shifting the pricing function *p*(*h*) – the supply curve in Figure 4 – to increase *p* at each *h*, although not necessarily the same amount. That shift would move both allocations *A* and *C* up the demand curve. If either of them was the unique optimum, then the consumer would respond to higher prices by consuming less, at least if the shift of the pricing function were small enough. However, a consumer indifferent between *A* and *C* without the shift may find allocation *A* to be the unique optimum as a result of the upward shift in supply, especially to the degree that supply shifts up more at *C* than it does at *A*. An increase in the marginal cost schedule everywhere may thereby result in the consumer switching from *C* to *A* – consuming substantially more – to further reduce his marginal cost.[21]

Figure 4. Personal increasing returns

Allocations A and C are on the same indifference curve, and dominate B



Local optima *A* and *C* are compared based on the area to the left of the Hicksian demand curve (not shown, but closely related to *M* as discussed in the text) and the additional investment $h_A - h_C$ required to obtain the lower price.

Perhaps the market primarily includes consumers for whom *A* or *C* was their unique optimum and therefore primarily consumers who respond to the supply shift by reducing consumption. Nevertheless, the consumption increases much more for each of those (few)

[21] The consumer who switches from *C* to *A* as a result of the adverse supply shift is still worse off even though the he consumes more *q*. This odd situation is related to the result from price index theory where cost of living indices put nonnegative weight on all prices whereas the price index for describing behavior puts negative weight on the prices of inferior inputs.

consumers who were indifferent between A and C. The latter may dominate so that the aggregate of both types of consumers responds to an upward shift in supply by consuming more on average.

Mortality from prescription drugs reached alarming levels in the U.S. by 2010, which motivated policymakers to (in effect) increase the cost of opioid prescriptions for recreational users. According to the personal increasing returns model, this policy might increase recreational opioid consumption if it induced consumers to switch to opioid sources with significantly lower marginal cost. Indeed, several studies found large numbers of prescription consumers switching to heroin and later illicitly-manufactured fentanyl (Alpert, Powell and Pacula 2018, Evans, Lieber and Power 2019, Mallatt 2018, Powell, Alpert and Pacula 2019, Ruhm 2019, Schnell 2018). Mulligan (2020) notes the large and growing price gap between prescription opioids and their illicit competitors and uses the personal increasing returns model as a quantitative description of that market. In that model, q denotes the quantity of opioids regardless of source and h denotes the cost of accessing illicitly manufactured fentanyl, which is cheaper at the margin than prescription opioids.

Mulligan (2022) observes that alcoholic beverages consumed at bars and restaurants have an element of quality as revealed by the fact that consumers pay three times as much per gallon at such venues than they would drinking at home. Personal increasing returns may be a mechanism by which the 2020-21 pandemic increase alcohol-related deaths around the world, because the pandemic increased the cost of drinking at bars and restaurants more than it increased the cost of drinking at home.

Gasoline taxes are sometimes levied with the intention of reducing traffic congestion by raising the cost per hour driven (Parry and Small 2005). The well-known "rebound effect" says that the gas tax encourages switches to fuel efficient cars, which may encourage more driving due to the lower hourly costs of such cars (Kahn 2020). In our model, let q denote hours driven. Consumers can pay a cost h to reduce the cost per hour driven, such as selecting a more expensive but more fuel efficient model or driving at different times of day. Even though the gas increases the hourly cost of driving at any given h, we have the possibility of nonlocal changes in behavior that result in more hours driven. Moreover, our model predicts that the size of the rebound effect varies over time as consumers have a wider selection of vehicles: a larger gap between $p(0)$ and $p(y)$ over time means that $1/\varepsilon_S$ is getting more negative.

The design of excise taxes can also have these surprising effects. A tax levied per unit quantity shifts Figure 4's supply curve up uniformly at each quantity (by the amount of the tax) and thereby be expected to reduce quantity consumed. An *ad valorem* tax, on the other hand, rotates the supply curve and may thereby raise quantity consumed by encouraging a shift to low quality. A similar distinction exists between implicit employment taxes, such as unemployment benefits, and the marginal tax rates on income created as public assistance is withheld on the basis of beneficiary income (Mulligan 2012).

VI. Conclusions

Many important consumer behaviors involve complementary goods and potential increasing returns. This paper shows how individual-level behaviors can nonetheless be examined in a tractable single-equation supply-demand framework with downward-sloping supply. It indicates when downward-sloping supply is applicable and provide metrics for assessing the magnitude of the negative slope. Even if supply and demand cross only once, the quantity effects of cost and demand changes are amplified relative to what they would be with horizontal supply because they are reinforced by marginal price changes. For example, subsidizing leisure with an unemployment benefit system reduces the net-of-benefit wage, which reduces work, which reduces human capital investment, which further reduces the net-of-benefit wage.

Although the paper focuses on individual-level behaviors, many of the tools in this paper could be applied at the industry level too. However, the activity in some industries feeds back on input prices, which may push toward diminishing returns at the industry level even while individuals might face increasing returns. To the extent that industry supply slopes down too, positive price-quantity correlations over time or across markets are expected to be rare because both supply and demand slope in the same direction. Indeed, if the possibility of increasing returns is overlooked, movements along a supply curve can be confused with movements along the demand curve.[22]

I cite several examples – such as gas taxes or alcoholism – where supply and demand may cross more than once. Multiple crossings raise the possibility that behavior depends on non-local properties of the supply curve. In such cases, an upward shift in supply at each quantity may nonetheless increase the equilibrium quantity because consumers' best response to an unwelcome situation is to invest in complementary behaviors that lower marginal costs. The law of demand appears to be violated, until we recognize that endogenous choices can move marginal prices in the opposite direction as the changes that would occur with behavior fixed. The metrics provided in this paper help predict when tax and regulatory policies may have the opposite of their intended effects.

.

[22] As shown in Section IV, this refers to supply and demand for a good. The roles are reversed in a market for a "bad" such as labor.

VII. Appendix: A Numerical Example of Nonlocal Substitution

.

In this example, $u(c,q)$ is Cobb-Douglas with equal budget shares. The consumer has income $y = 2$. The second-stage Hicksian demand for q is $u/\sqrt{p(h)}$.

The price function $p(h)$ is has two phases. In the first phase $h \in [0,h_2]$, price is a strictly increasing and strictly convex polynomial in h. In the second phase, p is constant.

$$p(h) \equiv \frac{67}{12} - \frac{47}{4}f(h) + \frac{59}{8}f(h)^2 - \frac{5}{6}f(h)^4 \tag{14}$$

$$f(h) \equiv \begin{cases} 3rd\ root\ of\ 16x^5 - 118x^3 + 141x^2 - 24h - 9 & h \in [0, h_2] \\ q_2 & h > h_2 \end{cases} \tag{15}$$

$$h_2 \equiv \frac{2}{3}q_2^5 - \frac{59}{12}q_2^3 - \frac{47}{8}q_2^2 - \frac{3}{8} \tag{16}$$

$$q_2 \equiv 2nd\ root\ of\ 40x^3 - 177x + 141 \tag{17}$$

where h_2 and q_2 are known constants in the space of algebraic numbers. With (18) the optimal policy function, we confirm that h_2 is the cost-minimizing investment corresponding to the quantity q_2:

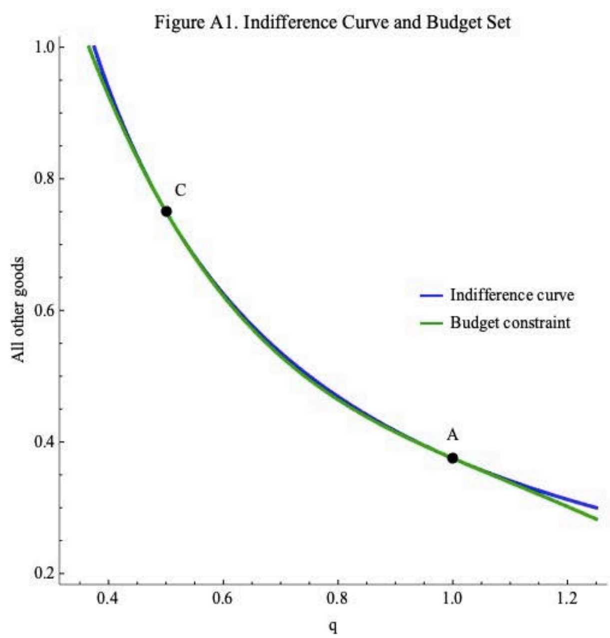
$$h(q) = \begin{cases} 0 & q \in [0, q_1) \\ \frac{2}{3}q^5 - \frac{59}{12}q^3 - \frac{47}{8}q^2 - \frac{3}{8} & q \in [q_1, q_2] \\ h_2 & q > q_2 \end{cases} \tag{18}$$

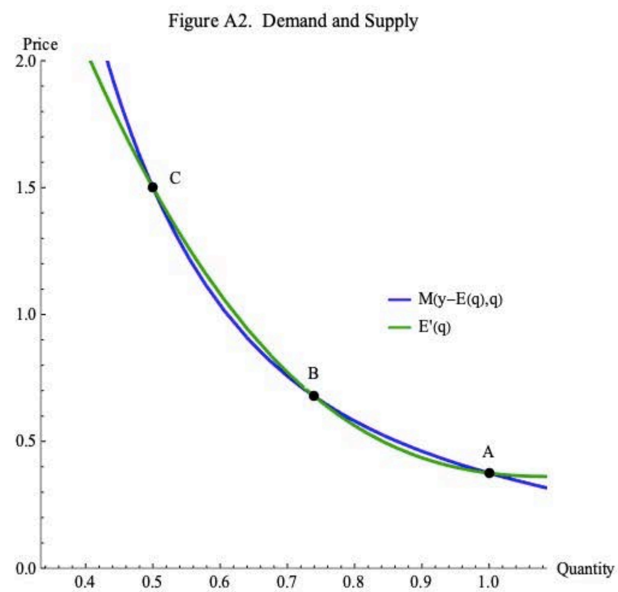
$$q_1 \equiv 3rd\ root\ of\ \frac{2}{3}x^5 - \frac{59}{12}x^3 - \frac{47}{8}x^2 - \frac{3}{8} \tag{19}$$

For $q \in (q_1,q_2)$, $h(q)$ is positive and strictly increasing. The first-stage expenditure function is expenditure evaluated at the optimal investment:

$$E(q) = h(q) + p\big(h(q)\big)q = \begin{cases} p(0)q & q \in [0, q_1) \\ -\frac{1}{6}q^5 - \frac{59}{24}q^3 - \frac{47}{8}q^2 + \frac{67}{12}q - \frac{3}{8} & q \in [q_1, q_2] \\ h_2 + p(h_2)q & q > q_2 \end{cases} \tag{20}$$

For $q \in (q_1,q_2)$, $E(q)$ is positive, strictly increasing, and strictly concave. It is linear outside that range. Figure A1 is a plot of the budget constraint (green) together with the indifference curve (blue) corresponding to the optimal choices of q. Figure A2 is the corresponding plot of M and $E'(q)$.



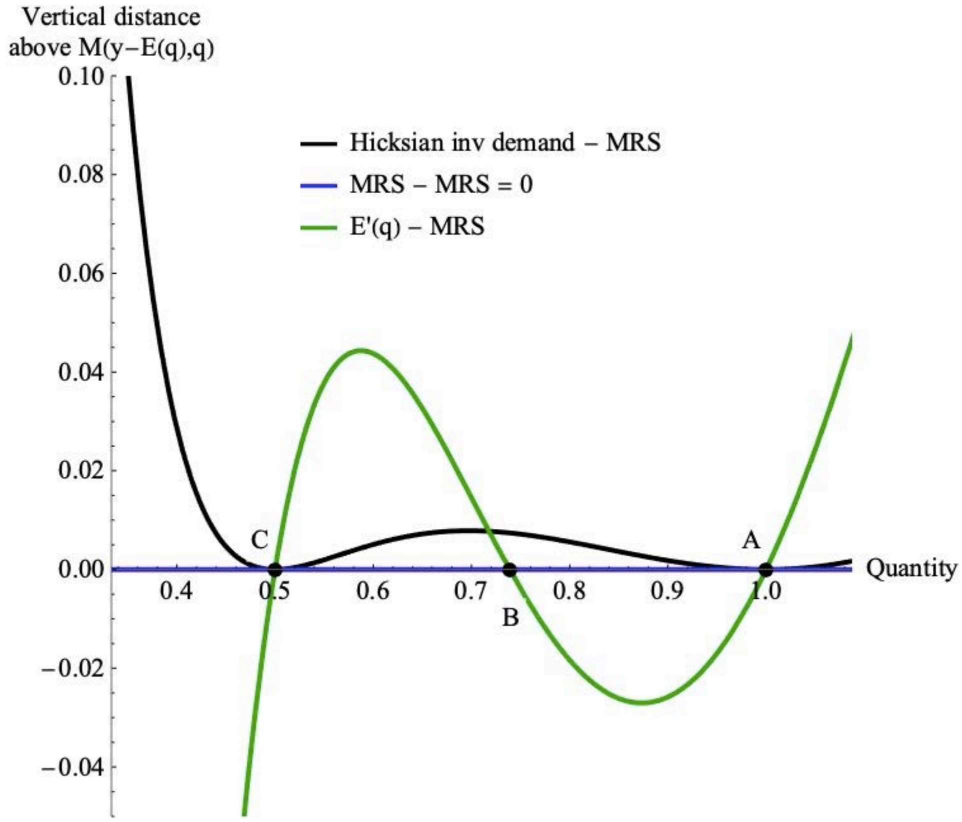


The points labeled *A* and *C* are analogs to the same-labeled points in Figure 4. Their quantities are 1 and ½, respectively, and prices 3/8 and 3/2. They are on the same Hicksian demand curve (not shown) because the consumption of other goods is 3/8 and 3/4, respectively.

In order to focus on the differences between Figure A2's two curves as well as showing their relation to Hicksian demand, Figure A3 shows the vertical (price-dimension) distances of $E'(q)$ and inverse Hicksian demand from M.[23] At allocations *A* and *C*, $E'(q)$ cuts both M and Hicksian demand from below. $E'(q)$ crosses each of them a third time, with the M crossing labeled as point *B* in Figures 4, A2, and A3.

[23] Each tick on Figure A3's vertical axis is 1/25th of that in Figure A2.

Figure A3. Vertical gaps between demand and supply



This numerical example, and symbolic derivations in the main text, was prepared in a Wolfram Language Notebook available at nber.org.

XI. Bibliography

Aguayo-Tellez, Ernesto, and Christian I. Rivera-Mendoza. "Migration from Mexico to the United States: wage benefits of crossing the border and going to the US interior." *Politics & Policy* 39 (2011): 119–140.

Alpert, Abby, David Powell, and Rosalie Liccardo Pacula. "Supply-side drug policy in the presence of substitutes: Evidence from the introduction of abuse-deterrent opioids." *American Economic Journal: Economic Policy* 10 (2018): 1–35.

Becker, Gary S. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics* 3 (1985): S33–S58.

—. *Human capital: A theoretical and empirical analysis, with special reference to education.* Chicago: University of Chicago Press, 1964.

Becker, Gary S., and Gilbert R. Ghez. *The allocation of time and goods over the life cycle.* Columbia University Press (for NBER), 1975.

Becker, Gary S., and Kevin M. Murphy. "A Simple Theory of Advertising as a Good or Bad." *Quarterly Journal of Economics* 108, no. 4 (November 1993): 941-964.

Becker, Gary S., and Kevin M. Murphy. "A Theory of Rational Addiction." *Journal of Political Economy* 96, no. 4 (August 1988): 675-700.

—. *Social Economics.* Cambridge, MA: Harvard University Press, 2003.

Boskin, Michael J. "Notes on the tax treatment of human capital." *NBER working paper*, no. 0116 (November 1975).

Case, Anne, and Angus Deaton. *Deaths of Despair and the Future of Capitalism.* Princeton University Press, 2020.

Evans, William N., Ethan M. J. Lieber, and Patrick Power. "How the reformulation of OxyContin ignited the heroin epidemic." *Review of Economics and Statistics* 101 (2019): 1–15.

Goldin, Claudia, and Lawrence F. Katz. "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy* 110, no. 4 (August 2002): 730–70.

Goode, Richard. "Income, Consumption, and Property as Bases of Taxation." *The American Economic Review* 52 (1962): 327–334.

Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu. "Engines of liberation." *The Review of Economic Studies* 72 (2005): 109–133.

Heckman, James J. "A Life-Cycle Model of Earnings, Learning, and Consumption." *Journal of Political Economy* 84, no. 4 (August 1976): S11-44.

Heckman, James. "Life cycle consumption and labor supply: An explanation of the relationship between income and consumption over the life cycle." *The American Economic Review* 64 (1974): 188–194.

Jaffe, Sonia, Robert Minton, Casey B. Mulligan, and Kevin M. Murphy. *Chicago Price Theory.* Princeton University Press (ChicagoPriceTheory.com), 2019.

Kahn, Matthew E. "Fundamentals of Environmental and Urban Economics." *ME Kahn*, 2020.

Kopecky, Karen A. "The trend in retirement." *International Economic Review* 52 (2011): 287–316.

Mallatt, Justine. "The effect of prescription drug monitoring programs on opioid prescriptions and heroin crime rates." *Available at SSRN 3050692*, 2018.

Mulligan, Casey B. "Lethal Unemployment Bonuses? Substitution and Income Effects on Substance Abuse, 2020-21." *NBER working paper*, no. 29719 (February 2022).

Mulligan, Casey B. "Prices and Federal Policies in Opioid Markets." *NBER working paper*, no. 26182 (February 2020).

—. *The Redistribution Recession.* New York: Oxford University Press (redistributionrecession.com), 2012.

Mulligan, Casey B., and Yona Rubinstein. "Selection, Investment, and Women's Relative Wages over Time." *Quarterly Journal of Economics* 123, no. 3 (August 2008): 1061–110.

Mulligan, Casey B., Russell Bradford, James H. Davenport, Matthew England, and Zak Tonks. "Non-linear Real Arithmetic Benchmarks derived from Automated Reasoning in Economics." Edited by Anna M. Bigatti and Martin Brain. *Proceedings of the 3rd Workshop on Satisfiability Checking and Symbolic Computation.* Oxford, UK, 2018. 48-60.

Murphy, Kevin M., Edward A. Snyder, and Robert H. Topel. "Competitive discounts and antitrust policy." *The Oxford Handbook of International Antitrust Economics, Volume 2*, 2014: 89.

O'Neill, June, and Solomon Polachek. "Why the gender gap in wages narrowed in the 1980s." *Journal of Labor economics* 11 (1993): 205–228.

Parry, Ian W. H., and Kenneth A. Small. "Does Britain or the United States have the right gasoline tax?" *American Economic Review* 95 (2005): 1276–1289.

Powell, David, Abby Alpert, and Rosalie L. Pacula. "A transitioning epidemic: how the opioid crisis is driving the rise in hepatitis C." *Health Affairs* 38 (2019): 287–294.

Ruhm, Christopher J. "Drivers of the fatal drug epidemic." *Journal of health economics* 64 (2019): 25–42.

Sargent, Thomas, and Lars Ljungqvist. "The European Unemployment Dilemma." *Journal of Political Economy* 106, no. 3 (June 1998): 514-50.

Schnell, Molly. "The Economics of Physician Behavior." *Princeton University PhD Dissertation*, 2018.

Schotanus, Fredo, Jan Telgen, and Luitzen de Boer. "Unraveling quantity discounts." *Omega* 37 (2009): 510–521.