Quantifying inefficiency in incomplete asset markets*

Mario Tirelli† and Sergio Turner‡

September 2007

Abstract

It is known that the incompleteness of asset markets causes inefficiency in almost every equilibrium. Yet unexplored is the "size" of this inefficiency.

The size of a Pareto improvement is the total willingness to pay for it, out of current consumption. Inefficiency is the maximum size of any Pareto improving reallocation.

Inefficiency of US consumption in middle age is computed to be 10-11% of total consumption in youth, for CRRA parameters 1.5-3.25, in a calibrated economy.

The inefficiency of a general economy is approximated. A natural approximation, based on marginal rates of substitution (MRS), is preposterously crude in the calibrated economy, owing to a law of diminishing willingness to pay.

Alternative approximations end up being functions of a classical notion, weighted social welfare maximized subject to resource constraints. They are simple, sharper in general and accurate in the calibrated economy.

Keywords: incomplete markets, Pareto improvement, inefficiency, willingness to pay, income mobility, income distribution, social welfare function

JEL Classification: D52, D61, H11, H20

*For their comments, we thank the particpants in the 2^{nd} CARESS-Cowles Conference on General Equilibrium, Colloquia on Mathematical Economics at the University of Rome 3, micro theory seminars at Rice and UPenn, European Workshop on General Equilibrium XVI. We thank particularly Gaetano Bloise, Geoffroy de Clippell, Peter Hartley, Dirk Krueger, Felix Kubler, Víctor Ríos-Rull, and Norovsambuu Tumennasan.

†Department of Economics, University of Rome 3

‡Department of Economics, Brown University

1 Introduction

Suppose a policy maker wants to Pareto improve on an allocation parameterized by future uncertainty, and carrying out any reallocation involves a present cost. A basic question is whether *any* Pareto improving reallocation is "worth" this cost. Quantitatively, "how inefficient" is an allocation parameterized by future uncertainty?

Consider households with preferences for consumption $x^h = (x_0^h, x_{\mathbf{1}}^h)$ of a current amount and an uncertain future amount. The current *willingness to pay* for a stochastic change z^h in future consumption is by definition the maximum value w^h making household h weakly prefer $x^h + (-w^h, z^h)$ to x^h. In this sense $(-w, z)$ defines a weak Pareto improvement, whose **size** is the total willingness to pay Σw^h. The **inefficiency** of an allocation x is the maximum size of any weakly Pareto improving reallocation:

$$\rho_x := \sup_{x_{\mathbf{1}}+z \gg 0, \Sigma z^h = 0} \Sigma w^h \tag{1}$$

This measure is for ordinal preferences and in real terms, lying between 0 and Σx_0^h. It is 0 exactly at Pareto efficiency. Debreu's (1951) coefficient of resource utilization is similar, the main distinction being that ours requires payment in the present only–when there is willingness–whereas his in the future as well–when there is moral hazard in place of willingness.[1]

The measure is not the welfare cost of the "business cycle," the focus of a large literature[2] since Lucas (1987), nor of a "permanent shock," a fitter interpretation of the future here. A single, representative household illustrates: Pareto efficiency and $\rho_x = 0$ are automatic, whereas the welfare cost of a large permanent shock is hardly 0. The randomness of total income is fixed in our question, but smoothable in Lucas'; our payment is for allocating, Lucas' is for eliminating, this randomness.

There is a literature on inefficiency of equilibria involving asset markets, when these incompletely insure against the future. It detects the generic existence of Pareto improvements, but does not quantify their significance–owing to the underlying technique, as explained below. This generic existence is robust. For economies with multiple goods, it survives various limits on reallocations: rebalances of portfolios, Geanakoplos and Polemarchakis (1986), Geanakoplos et al. (1990), and Stiglitz (1982); lump sum changes in current income plus a mild instrument, Citanna, Kajii, and Villanacci (1998); taxation of asset trades, Citanna, Polemarchakis, and Tirelli (2006); anonymous income taxes, Tirelli (2003); excise taxes or capital gains taxes, Turner (2005). For economies with a single good[3], it survives behavioral agents, Nagata (2005). We focus on quantifying inefficiency, not on detecting it.

Equipped with a measure of inefficiency, we ask two questions:

- How inefficient is an equilibrium allocation, calibrated to capture many people facing a major risk against which insurance markets are incomplete?
- Is there a closed formula for inefficiency, to invite research about the relation between the size of inefficiency and the underlying economy?

[1] Only in Debreu (1959) is time explicitly recognized.

[2] E.g. Barro (2007), İmrohoroğlu (1989), Krebs (2003), Krusell and Smith (1999), Kurz (2005), Levine and Zame (2002), Ríos-Rull (1994).

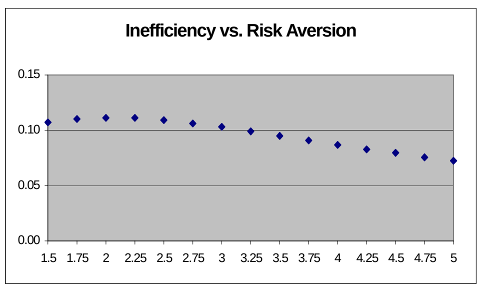
[3] Magill and Quinzii (1996) is the reference here.

1.1 Estimation of the inefficiency of US middle age consumption

In applying this measure of inefficiency, it is fundamental to specify first the underlying uncertainty. We highlight a major risk faced by US youth, income in middle age. This risk is underscored by the high income mobility and inequality. Income belongs to a different quintile in middle age than it did in youth with high probability, at least $.46$ according to Gottschalk and Danzinger (1997).[4] At the time of the middle aged, any quintile's median income exceeds the preceding quintile's by 45% or more, McNeill (1999).

We must estimate the consumption distribution and preferences. To estimate the former, we adjust data on income distribution by data on savings rates. To estimate the latter, we fix them to be time additive, von Neumann-Morgenstern, $v(x_0)+\delta^h\Sigma\pi_s v(x_s)$, with felicity $v=\frac{c^{1-\beta}}{1-\beta}$ that has CRRA $\beta>1$, and a probability space capturing Gottschalk and Danzinger's probabilities of transitioning between quintiles. Preferences differ only in the patience parameter δ^h, calibrated to make the estimated consumption distribution optimal given the interest rate.

This calibrated economy's inefficiency is numerically computable. Youth are willing to pay a fraction $\frac{\rho}{\Sigma x_0^h}=.10-.11$ of their total current consumption for a weakly Pareto improving reallocation of their total future consumption, if the CRRA is $1.5 \leftrightarrow 3.25$:[5]



These estimates, though rooted in data, are limited in their reliability by the crudeness of the state-preference model, which ignores how the reallocations underlying (1) affect incentives for creating income. Still, for many models inefficiency is quantifiable in the same spirit, as the supremum total willingness to pay for a Pareto improving feasible change.

1.2 Formulas for *approximate* inefficiency of a general economy

A formula for the inefficiency of a general economy is elusive, even if a formula for willingness to pay exists. Nonetheless, we

- note a formula, if preferences are representable and quasilinear in current consumption
- derive a natural formula L_x for approximate inefficiency of a general economy, *linear* in marginal rates of substitution
- note how crude L_x is in the calibrated economy, and why by a law of diminishing willingness to pay

[4]Incomes are from a panel. This mobility is beyond that associated with income being monotone in age.

[5]For Kocherlakota (1996), the plotted range $1.5-5$ is empirically wide.

- derive *discrete* and *quadratic* formulas D_x, Q_x for approximate inefficiency of a general economy
- prove both alternatives are generally sharper, and note their accuracy in the calibrated economy
- relate D_x, Q_x and a classical notion, social welfare maximized subject to resource constraints
- extend the results to economies with multiple goods

We stop short of theoretical applications of these approximations, a future topic. Throughout, we assume preferences admit time separable representations, $u_0 + u_{\mathbf{1}}$.

A two-step **method** drives all approximations of inefficiency. First, we derive an upper bound $w \leq \hat{w}$ on the willingness to pay for a given change z in future consumption. Second, we compute $\hat{\rho}_x$, the value of problem (1) relaxed to $\sup_{x_{\mathbf{1}}+z\gg 0, \Sigma z^h=0} \Sigma \hat{w}^h$ or an upper bound thereof–this $\hat{\rho}_x$ is the method's approximation.

1.2.1 Marginal rates of substitution are a poor foundation

Marginal rates of substitution are appealing. They detect inefficiency both theoretically–that households' marginal rates of substitution are not all equal–and practically–that small Pareto improving reallocations solve some linear system. Being so good at detecting inefficiency, they should be good at quantifying it.

Specifically, a natural approximation of the willingness to pay for a change z in future consumption is ∇z, where $\nabla \in \mathbb{R}^S$ is the vector of marginal rates of substitution of current income for future income. The two-step method then yields the **linear approximation** of inefficiency (theorem 1):

$$\rho_x \leq L_x := \Sigma \left(\triangledown^* - \triangledown^h\right) x_{\mathbf{1}}^h$$

where $\triangledown^* := (\max_h \triangledown_s^h)_s$ is the stochastic maximum MRS over all households, an index of the most deprived. In this approximation, all households fully donate their future consumption to the most deprived.

Intuitive and appealing as it is, this linear approximation is preposterously crude in the calibrated economy (CRRA $= 2.5$). While inefficiency is $\rho_x = .109 \cdot \Sigma x_0^h$, the approximation is over fourfold, $L_x = .457 \cdot \Sigma x_0^h$.

The linear approximation is crude because of a law of diminishing willingness to pay. Given a direction $z \in \mathbb{R}^S$ of change in future consumption, change $z(t) := tz$ is parameterized by its "size" $0 \leq t \leq 1$. How does the willingness to pay $w(z(t))$ for a change depends on its size? It turns out that $w(z(t))$ is concave–the willingness to pay for a marginal change is diminishing.

In sum, marginal rates of substitution are useful to detect inefficiency, but inept to quantify it.

1.2.2 A classical social program is a better foundation

Sharper approximations of inefficiency turn out to be functions of a classical social program,

$$F^* := \sup_{y_{\mathbf{1}}\gg 0, \Sigma y_{\mathbf{1}}^h = r_{\mathbf{1}}} \Sigma \mu^h u_{\mathbf{1}}^h(y_{\mathbf{1}}^h)$$

namely, maximizing **future social welfare** $F : \mathbb{R}_+^{SH} \to \mathbb{R}, F(y_{\mathbf{1}}) := \Sigma \mu^h u_{\mathbf{1}}^h(y_{\mathbf{1}}^h)$ with weights $\mu^h := \frac{1}{u_0^{h\prime}(x_0)}$ constrained by future resources $r_{\mathbf{1}} := \Sigma x_{\mathbf{1}}^h$.

The **discrete approximation** (theorem 2) is intuitive,

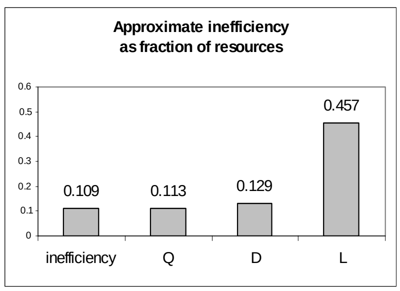
$$\rho_x \le D_x := F^* - F(x_{\mathbf{1}}) \tag{2}$$

just the failure to solve the social program. The **quadratic approximation** (theorem 3) is less intuitive, itself a function of the discrete one,

$$\rho_x \le Q_x := \sqrt{\overline{T}_0^2 + 2\overline{T}_0 \cdot D_x} - \overline{T}_0 \tag{3}$$

where $\overline{T}_0 := \Sigma T_0^h$ is total risk tolerance, and $T_0^h = -u_0'(x_0)/u_0''(x_0)$.

These approximations are increasingly sharper, in that $\rho_x \le Q_x \le D_x \le L_x$ (proposition 3). The sharpening is dramatic in the calibrated economy (CRRA $= 2.5$):



relative errors $\frac{\hat{\rho}-\rho}{\rho}$

Q_x	D_x	L_x
$+3.7\%$	$+18.3\%$	$+319.3\%$

The quadratic is excellent, the discrete very good, the linear preposterous.

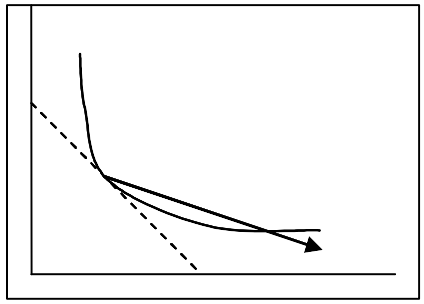
Both are precise at Pareto efficiency, where $\rho_x = 0$. To see why, we recall the classical result that a Pareto efficient interior allocation solves $\sup_{y \gg 0, \Sigma y^h = r} \Sigma \mu^h u^h(y^h)$, which thanks to time separable utilities $u = u_0 + u_{\mathbf{1}}$ in turn implies it solves F^*, i.e. $F^* = F(x_{\mathbf{1}})$. Thus at a Pareto efficient interior allocation, both approximations (2), (3) take value 0, precisely ρ_x.

Lastly we note how all results extend to economies with multiple goods. The key is to let current income I_0 play the role of the numeraire, with current utility $u_0(x_0)$ replaced by its indirect utility $v_0(p_0, I_0)$. Now inefficiency and all bounds on it are defined in terms of the allocation $(I_0, x_{\mathbf{1}})$ as before, as well as current prices p_0. Although multiple goods underlie much of the literature on the generic *presence* of equilibrium inefficiency, they seem immaterial for the *size* of inefficiency. After all, with state separable preferences, in each state the equilibrium allocation of goods already is Pareto efficient; only the interstate allocation of income might not be–and income is a numeraire as above.

1.3 Related literature

The closest literature is purely qualitative, on the Pareto inefficiency of asset market equilibria, following Stiglitz (1982) as cited above. Adding a quantitative dimension is what motivates the measure of inefficiency, estimating it in a calibrated economy and approximating it is a general economy.

This literature lacks a quantitative dimension because of its very technique, linearization. Linearization is limited to infinitesimal reallocations. It tells whether there is a direction of reallocation (in the sense of directional derivative) that Pareto improves the equilibrium allocation. Although conclusions at the infinitesimal scale extend to conclusions at the neighborhood scale, they do not beyond; indeed, a direction that is improving if followed infinitesimally may become impairing if followed finitely:



Accordingly, linearization is mute on the size of the Pareto improvements it proves to exist. In contrast, our measure of inefficiency does quantify Pareto improvements, and is unconstrained by any neighborhood.

An apparently close literature, which is quantitative, is on the welfare cost of business cycles, following Lucas (1987) as cited above. In truth, it deals with payments for smoothing (changing) total future income, whereas we focus on payments for allocating (fixing) total future income.

The paper's organization follows. **Section 2** quantifies inefficiency in terms of willingness to pay. **Section 3** numerically computes the inefficiency of an economy calibrated to US income mobility, income distribution, and savings rates. **Section** 4 describes the method for deriving formulas for approximate inefficiency of a general economy. **Section 5** shows how marginal rates of substitution approximate inefficiency, but only crudely in the calibrated economy, because of a law of diminishing willingness to pay. **Section 6** derives alternative approximations, sharper in general and accurate in the calibrated economy. Finally, **Section 7** is an extension to multiple goods. An appendix contains proofs.

2 Quantifying inefficiency

Inefficiency is a qualitative notion–we seek to quantify it. The idea is in the spirit of Debreu's (1951) coefficient of resource utilization, but emphasizes willingness to pay and the timing of payment.

Let $\succeq$ be a preference on $\mathbb{R}^{1+S}_{++}$, whose points $(x_0, \boldsymbol{x_1})$ specify consumption of a sole good in the present 0 and in the future states of nature $1, ..., S$. Let x be a status quo consumption, and $z \gg -\boldsymbol{x_1}$ a change in future consumption. The **willingness to pay** for this change, in terms of current consumption, is by definition the supremum $w \in \mathbb{R}$ such that

$$x + (-w, z) \succsim x \tag{4}$$

Fixing the status quo, this defines a function $w = w(z)$ bounded above by x_0, whose argument we occasionally omit.[6] It may take negative values, interpretable as compensation.

[6] The set of w such that (4) is bounded above by x_0; if nonempty, its supremum exists by completeness of $\mathbb{R}$. Nonemptiness holds if the preference obeys **0-desirability**: $x \in \mathbb{R}^{1+S}_{++}, \boldsymbol{x_1} + z \in \mathbb{R}^{S}_{++} \Rightarrow (x_0 - w, \boldsymbol{x_1} + z) \succsim x$ for some $w \in \mathbb{R}$, possibly negative. So $w(z) \leq x_0$ is defined.

Remark 1 *If preference is continuous and Inada in current consumption, then (4) holds with indifference at the willingness to pay* $w(z)$*. Further, if the preference is increasing in current consumption, a solution* w *of* $x+(-w,z)\sim x$ *is unique.*[7]

The timing of payment–out of current consumption, not future–matches the interpretation of preferences as being *ex ante* the realization of the state of nature. *Ex post*, the willingness to pay in the realized state s may be different, if naturally defined, such as with state-separable preferences.

We quantify inefficiency in terms of willingness to pay. Let the **economy** $(\succeq, x)$ specify for each household $h=1,...,H$ a preference, as in remark 1, and a status quo consumption; let w^h be the implied willingness to pay functions.

Definition 1 *The **inefficiency** of* $(\succeq, x)$ *is the value* ρ_x *of*

$$\sup \Sigma w^h(z^h) \qquad s.t. \qquad x_{\mathbf{1}}+z\gg 0, \Sigma z^h=0 \tag{5}$$

The measure is for ordinal preferences, denominated in current resources, and lies in $[0, \Sigma x_0^h]$.[8] It is society's supremum willingness to pay for an allocation that, with the payment, is just weakly Pareto improving. A solution z of this problem, if it exists, is an **optimal arbitrage**. An arbitrageur could elicit Σw^h in the present, without adding to future resources.

There is a computationally useful characterization of optimal arbitrage:

Proposition 1 *Suppose in addition preferences are transitive.*[9] *Suppose* $z\in\mathbb{R}^{SH}$ *is feasible for (5). Then it is a solution iff the* $x^h+(-w^h,z^h)$ *define a Pareto optimum.*

This implies a characterization of Pareto efficiency:

Corollary 1 x *is Pareto efficient iff* $\rho_x=0$.

Occasionally, to get a measure of inefficiency at most 1 we quote the Σw^h of problem (5) as a fraction of current resources, $\Sigma w^h=\phi r_0$, where $r:=\Sigma x^h$ is notation for resources. So current consumption $c_0^h:=x_0^h-w^h$ satisfies $\Sigma c_0^h=\tilde{\phi} r_0$ with $\tilde{\phi}:=1-\phi$, which is a measure of *efficiency.* A similar notion is Debreu's (1951) coefficient of resource utilization $\tilde{\phi}$, except that he requires current *and* future consumption to satisfy $\Sigma c^h=\tilde{\phi} r$. His notion refers to the fraction of resources, *in every state and not just today,* to which society is willing to deprive itself and still be weakly Pareto better off. As noted, with ex ante preferences, a willingness to pay today may disappear in a future state, a moral hazard hidden in Debreu's (1951) timeless model. The timing of payments is the key distinction between Debreu's measure and ours.

Lucas (1987) asks about the willingness to pay for changing future resources $r_{\mathbf{1}}:=\Sigma x_{\mathbf{1}}^h$ to their expectation $E[r_{\mathbf{1}}]$, eliminating their risk. This question is unrelated to inefficiency and to our measure, which by the constraint $\Sigma z^h=0$ fixes future resources, preserving their risk.

[7] Continuity means that $(y_0, y_{\mathbf{1}})\succ x$ implies $(\tilde{y}_0, y_{\mathbf{1}})\succ x$ in some neighborhood $\tilde{y}_0\approx y_0$. Inada means that always $x_0-w(z)>0$. Increasingness means that $x\in\mathbb{R}^{1+S}_{++}, \epsilon>0\Rightarrow x+(\epsilon,0)\succ x$.

[8] $z=0$ is feasible for (4) and has $w^h(0)=0$, so $\rho_x\geq\Sigma w^h(0)=0$. Since $w^h(z^h)\leq x_0^h$ whenever $z^h\gg -x_{\mathbf{1}}^h$, the sum $\Sigma w^h(z^h)\leq\Sigma x_0^h$, showing $\rho_x\leq\Sigma x_0^h$.

[9] "Boundary averse" means $y\succeq x\gg 0$ implies $y\gg 0$.

Remark 2 (constrained inefficiency) *Reallocations in problem (5) are state contingent. Were they constrained to arise from a particular policy–fiscal, monetary, financial–then problem (5) would measure "policy constrained inefficiency." This would be no greater than our measure, as the feasible set would be no greater; our measure is an upper bound on "policy constrained inefficiency" for any policy.*

The willingness to pay has an important representation, if the preference has a time separable $u_0(x_0) + u_{\mathbf{1}}(x_{\mathbf{1}})$ representation. By remark 1, w is characterized by the indifference $u_0(x_0 - w) + u_{\mathbf{1}}(x_{\mathbf{1}} + z) = u_0(x_0) + u_{\mathbf{1}}(x_{\mathbf{1}})$, equivalent to an equation involving the **change in future welfare** $\Delta := u_{\mathbf{1}}(x_{\mathbf{1}} + z) - u_{\mathbf{1}}(x_{\mathbf{1}})$:

$$u_0(x_0) - u_0(x_0 - w) = \Delta \tag{6}$$

Standard conditions imply an implicit function $w = w(\Delta)$ with $w(0) = 0$, representing the willingness to pay in terms of changes in future welfare.

A formula for inefficiency is elusive, even when a formula $w = w(\Delta)$ for willingness to pay is available. This unfortunate state motivates two explorations: to numerically compute inefficiency arising in an asset market equilibrium, calibrated to capture important risks, and to derive formulas for approximate inefficiency.

2.1 Formula in quasilinear case

We note a formula for inefficiency, assuming **0-quasilinear utilities**, $x_0 + u_{\mathbf{1}}(x_{\mathbf{1}})$. Equation (6) reduces to $w = \Delta$. The total willingness to pay is then $\Sigma w^h = \Sigma\Delta^h = \Sigma u^h_{\mathbf{1}}(x^h_{\mathbf{1}} + z^h) - \Sigma u^h_{\mathbf{1}}(x^h_{\mathbf{1}})$, so inefficiency (5) is the value

$$\rho_x = \left[\sup_{y_{\mathbf{1}} \gg 0, \Sigma y^h_{\mathbf{1}} = r_{\mathbf{1}}} \Sigma u^h_{\mathbf{1}}(y^h_{\mathbf{1}}) \right] - \Sigma u^h_{\mathbf{1}}(x^h_{\mathbf{1}}) \tag{7}$$

using the change of variable $y^h_{\mathbf{1}} = x^h_{\mathbf{1}} + z^h$, provided the solution has $x^h_0 - \Delta^h > 0$. Inefficiency is just the failure to to maximize "future social welfare" $\Sigma u^h_{\mathbf{1}}$. This is partially generalized in section 6.

3 Estimation of the inefficiency of US middle age consumption

One of the great risks faced by youth is consumption in middle age. How inefficiently is middle age consumption allocated? We use data on income mobility, income distribution and savings rates to calibrate the economy $(\succeq, x)$, a profile of preferences and consumption distribution. For this economy we numerically compute inefficiency ρ_x in the sense of definition 1.

3.1 Data on income mobility, income levels and savings rates

There is high risk of income mobility. Whatever one's income quintile in youth, moving to a different income quintile in middle age has high probability. Estimates of the probabilities of transitioning from 1968 quintiles to 1991 quintiles are

transition probabilities

$$\begin{array}{cccccc} \Rsh & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\ \mathbf{1} & .54 & .22 & .19 & .05 & .01 \\ \mathbf{2} & .23 & .25 & .18 & .26 & .08 \\ \mathbf{3} & .11 & .21 & .24 & .28 & .15 \\ \mathbf{4} & .05 & .23 & .23 & .19 & .3 \\ \mathbf{5} & .07 & .09 & .16 & .22 & .46 \end{array} \tag{8}$$

by Gottschalk and Danzinger (1997); see their table 4.[10] The probability of income mobility is at least $.46 \leq 1 - \textit{diagonal} = .46, .75, .76, .81, .54$, for all 1968 quintiles.

This risk of income mobility is greatly consequential, because the income distribution is greatly unequal. The median incomes of 1968, 1991 quintiles are (in 1991 dollars)

$$\begin{array}{cccccc} I_{1968}: & 9,774 & 20,722 & 29,682 & 40,330 & 70,802 \\ I_{1991}: & 9,315 & 22,725 & 35,570 & 51,317 & 96,501 \end{array} \tag{9}$$

by McNeill (1999).[11] Every 1991 median exceeds the preceding median by 45% or more.

This inequality in incomes implies inequality in consumption. Estimates of the quintiles' savings rates are (in 1973)

$$\sigma^q = -.45, . -.015, .09, .175, .286 \tag{10}$$

by Bosworth, Burtless, Sabelhaus (1991); see their table 5. Thus someone whose income in young age falls in quintile q but whose income in middle age transitions to quintile Q is estimated[12] to have

$$\begin{array}{ll} \text{consumption in youth:} & x_0^q := I_{1968}^q(1 - \sigma^q) \\ \text{consumption in middle age:} & x_{\mathbf{1}}^q := I_{1991}^Q + \sigma^q I_{1968}^q R \end{array} \tag{11}$$

where R is the gross interest rate over the period, say, $R = 1.65$.[13]

3.2 Translation of data to model

We wish to define a state space consistent with income mobility (8) and income distribution (9), and to calibrate preferences so consumption distribution (11) is optimal given the interest rate.

Time Period 0 is 1968, period 1 is 1991.

Households They are five representatives, sampled one from each 1968 income quintile. We deal with this sample instead of the whole population to keep the state space manageable for numerical computation.

States They are all the independent quintile transitions of the sampled households, totaling 5^5 states. Thus 55111 is the state where the two poorest representatives become rich and the three richest become

[10] Theirs is PSID data on 1968 and 1991 incomes, adjusted for family size. Only ages 22 to 62 appear.

[11] We switch to McNeill's (1999) quintile incomes from Gottschalk and Danzinger's (1997) because theirs are unreported. His definition of income is essentially the 1995 Panel on Poverty's, also adusted for family size.

[12] This estimate takes consumption and bequests as perfect substitutes, and ignores investments other than savings.

[13] This is the cumulated real interest rate for 1968-1991, computed from nominal yields of 1-year US Treasuries in the secondary market, and the US GDP deflator. Annualized, the real interest rate is about 2.2%, and $(1 + 0.022)^{23} \approx 1.65$.

poor. In a random sample, multiple households can transition to the same 1991 quintile; in the economy, multiple quintiles cannot transition to the same 1991 quintile–quintiles are equinumerous by definition.

State probabilities They are defined by the product rule for independent events, reflecting that households are a random sample. Thus $\pi_{55111} = (.01)(.08)(.11)(.05)(.07) \approx 3 \cdot 10^{-6}$, nearly impossible.

Assets A riskless bond with a unitary return in 1991 is tradeable in 1968 at price $\frac{1}{1.65}$.

Consumption distribution Our notion of state s determines every household's transition qQ, so that (11) defines a consumption distribution, post asset trade.

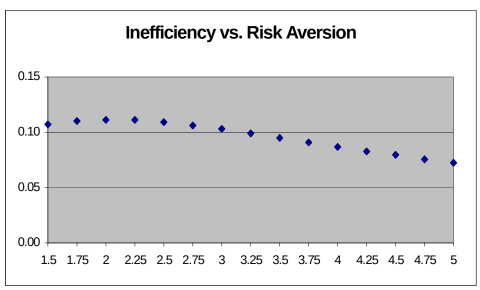
We quote inefficiency Σw^h as a fraction of total current consumption $\Sigma x_0^h =: r_0$, computed as follows. Tables (9) and (10) imply an aggregate savings rate of 0.151, so that $r_0 := .849 \cdot I_0$, where $I_0 := \Sigma I_0^q$ is total current income.

Preferences Households differ only in their patience parameters $\delta^h > 0$, otherwise having a common time separable, von Neumann-Morgenstern preference $v(x_0) + \delta^h \Sigma \pi_s v(x_s)$. The felicity $v(c) = \frac{1}{1-\beta} c^{1-\beta}$ has CRRA $\beta \in [1.5, 5]$, and π is the above state probability. We calibrate the patience parameters δ as follows.

Patience parameters Given β, we calibrate each δ^h by imposing that consumption distribution (11) is optimal, given the riskless bond's price above. For example, if $\beta = 2.5$, the annualized patience parameters $\delta^{\frac{1}{23}}$ are $.815, .957, .98, 1.006, 1.024$, increasing with income in youth. Details are in section 8.1.

3.3 Estimates

We estimate the inefficiency of US middle age consumption by having Mathematica solve problem (1) for the model economy $(\succeq, x)$ just specified.[14] Plotting $\frac{\hat{\rho}}{r_0}$,



The economy is willing to pay $10 - 11\%$ of its total current consumption for some reallocation of its future total consumption, subject to leaving everyone weakly better off, provided the CRRAs β range in $1.5 - 3.25$.

How robust are these estimates with respect to the CRRA parameter? Quite, in the range $\beta = 1.5 \leftrightarrow 3.25$.

Does the calibration pass validity tests, such as matching moments of the data? If the moment is the growth rate of per capita income, there is a match. From outside data, the growth factor of real GDP per capita during 1968-1991 is 1.5. In the calibrated economy, the stochastic growth factor in total income, $\frac{I_s + \sigma I_0}{I_0}$ with $\sigma = .151$ the aggregate savings rate, has expectation of 1.503 and standard deviation of $.34$, relative to the above state probabilities.

[14] The code is availale on request.

How are we to interpret these figures about the model economy, which is not the whole economy but merely a sample, one household from each quintile? Reallocations at the scale of the sample can be replicated to the whole economy, since quintiles are equinumerous. (Of course, many reallocations in the whole economy do not project to the sample.) Thus reallocations in our model economy represent a subset of the reallocations in the whole economy. Since definition 1 involves a maximization over all reallocations, the inefficiency of the whole economy is at least the plotted estimates, once normalized by its size $r_0 = \Sigma x_0^h$.

4 Method to approximate inefficiency of a general economy

In the absence of a formula for the inefficiency of a general economy $(\succeq, x)$, we turn to formulas for *approximate* inefficiency. A two-step method drives all approximations of inefficiency $\rho_x = \sup_{x_\mathbf{1}+z\gg 0, \Sigma z^h=0} \Sigma w^h(z^h)$:

- derive an upper bound on willingness to pay, $w^h(z^h) \leq \hat{w}^h(z^h)$
- compute the value of the relaxed problem (or an upper bound thereof)

$$\sup_{x_\mathbf{1}+z\gg 0, \Sigma z^h=0} \Sigma \hat{w}^h(z^h) \tag{12}$$

  on substituting in the definition of ρ_x the upper bound of step one.

Clearly,

Principle 1 *Inefficiency is bounded above by (12).*

The first step is essentially a Taylor approximation of equation (6) characterizing willingness to pay, $u_0(x_0) - u_0(x_0 - w) = \Delta(z)$. Care is needed that this approximation is an *upper bound.* The method yields three approximations: linear, discrete, and quadratic. Henceforth, numerical computation is absent except to probe their accuracy in the calibrated economy of section 3.2.

5 Approximation by marginal rates of substitution: crude

Marginal rates of substitution are appealing. They detect inefficiency both theoretically–that households' marginal rates of substitution are not all equal–and practically–that small Pareto improving reallocations solve some linear system. Being so good at detecting inefficiency, they should be good at quantifying it. Unfortunately, marginal rates of substitution (1) always overstate inefficiency, because they (2) ignore a law of diminishing willingness to pay, and (3) tag on a quantitatively gross error, even in reasonable cases.

A natural approximation of the willingness to pay is $w \approx \nabla z$, where ∇ is the so called **marginal rates of substitution (MRS)**, defined by the requirement that $(1, \nabla)$ be normal to the smooth indifference set through x. It turns out, this fulfills step one in the method:

Lemma 1 (linear approximation of willingness to pay) *Suppose the preference is reflexive and convex, and its indifference set through $x \gg 0$ is differentiable. The willingness to pay for z is at most*

$$w \leq \nabla z \tag{13}$$

By principle 1, inefficiency is bounded above by $* = \sup_{x_1+z\gg 0, \Sigma z^h=0} \Sigma \nabla^h z^h$, which involves the object

$$\triangledown^* := (\max_h \triangledown_s^h)_s \tag{14}$$

It is the stochastic maximum MRS over all households, an index of the most deprived. The solution of $*$ has all households fully donating their future consumption to the most deprived:

Theorem 1 (linear approximation of inefficiency) *Suppose as in lemma 1. Then inefficiency is bounded above as*

$$\rho_x \le \Sigma \left(\triangledown^* - \triangledown^h\right) x_1^h := L_x \tag{15}$$

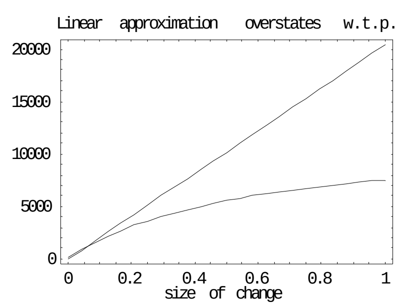
There is a reason that linearization overstates willingness to pay. Given a direction $z \in \mathbb{R}^S$ of change in future consumption, change $z(t) := tz$ is parameterized by its "size" $0 \le t \le 1$. How does the willingness to pay $w(z(t))$ for a change depends on its size?

Proposition 2 (law of diminishing willingness to pay) *Suppose the preference is increasing in current consumption and convex. Then* $w(z(t))$ *is concave. Further,* $w(z(t)) \ge tw(z)$ *for all* $0 \le t \le 1$.

5.1 Crudeness

The crudeness of marginal rates of substitution in approximating inefficiency is apparent in the calibrated economy of section 3.2 (CRRA $\beta = 2.5$). The inefficiency is $\rho_x = .109 \cdot r_0$ and the linear approximation (15) is over fourfold, $L_x = .457 \cdot r_0$.

This crudeness is present in the willingness to pay as well. Let us take the preference and status quo consumption to be that of the richest 1968 quintile, and the change z^5 in future consumption to be, say, that associated with the optimal arbitrage. The willingness to pay is $w^5 = 7514.08$ and the linear approximation (13) is nearly threefold, $\nabla^5 z^5 = 20392$. The culprit of this crudeness is the law of diminishing willingness to pay[15], as illustrated by plotting $w^5(z(t))$:



In sum, marginal rates of substitution are useful to detect inefficiency, but inept to quantify it. Granted, they are computationally much simpler than the inefficiency measure ρ_x. But so are the following approximations, sharper in general and accurate in the calibrated economy.

[15] Alvarez and Jermann (2004) linearize information about asset prices to answer Lucas' (1987) question. They find estimates far exceeding Lucas' (who does not linearize), but are mute on whether this excess owes to the linearization itself. They state an analogue of proposition 2.

6 Sharper approximations

We derive two approximations of inefficiency. Compared to the linear one based on MRS, they are sharper in general and dramatically so in the calibrated economy. Most surprisingly, inside them appears notion classically linked to efficiency, social welfare maximized subject to resource constraints:

$$F^* := \sup_{y_{\mathbf{1}} \gg 0, \Sigma y_{\mathbf{1}}^h = r_{\mathbf{1}}} \Sigma \mu^h u_{\mathbf{1}}^h(y_{\mathbf{1}}^h) \tag{16}$$

Here **future social welfare** $F : \mathbb{R}_+^{SH} \to \mathbb{R}, F(y_{\mathbf{1}}) := \Sigma \mu^h u_{\mathbf{1}}^h(y_{\mathbf{1}}^h)$ has the weights $\mu^h := \frac{1}{u_0^{h\prime}(x_0)}$, and **future resources** $r_{\mathbf{1}} := \Sigma x_{\mathbf{1}}^h$ are the status quo's. Preferences here admit a time separable representation $u_0 + u_{\mathbf{1}}$.

One application of the method results in

Theorem 2 (discrete approximation of inefficiency) *Suppose utilities for current consumption with $u_0' > 0 \geq u_0''$. Then the allocation's inefficiency is at most its failure to maximize future social welfare:*

$$\rho_x \leq F^* - F(x_{\mathbf{1}}) := D_x \tag{17}$$

Another application of the method results in something involving the **total risk tolerance** $\overline{T}_0 := \Sigma T_0^h$, where $T_0 := -u_0'(x_0)/u_0''(x_0)$.

Theorem 3 (quadratic approximation of inefficiency) *Suppose utilities for current consumption with $u_0''' \geq 0 > u_0'', -u_0'$.*[16] *Suppose also every household has a nonnegative willingness to pay for the optimal arbitrage. Then the allocation's inefficiency is at most*

$$\rho_x \leq \sqrt{\overline{T}_0^2 + 2\overline{T}_0 \cdot [F^* - F(x_{\mathbf{1}})]} - \overline{T}_0 := Q_x \tag{18}$$

This quadratic approximation Q_x is a correction of the discrete one D_x : $Q_x = \sqrt{\overline{T}_0^2 + 2\overline{T}_0 \cdot D_x} - \overline{T}_0$

A seeming weakness of theorem 3 is the high level hypothesis on the optimal arbitrage; it holds in the calibrated economy.

DISCUSSION OF APPROXIMATIONS

Both are explicit up to F^*, which to compute requires specifying future welfare $u_{\mathbf{1}}$, as illustrated below.

Both are precise at Pareto efficiency. To see why, we recall the classical result that a Pareto efficient interior allocation solves $\sup_{y \gg 0, \Sigma y^h = r} \Sigma \mu^h u^h(y^h)$, which thanks to time separable utilities $u = u_0 + u_{\mathbf{1}}$ in turn implies $x_{\mathbf{1}}$ solves (16), i.e. $F^* = F(x_{\mathbf{1}})$. Thus at a Pareto efficient interior allocation, both approximations (17), (18) take value 0, which by corollary 1 is precisely the value of ρ_x.

Approximation (17) partially generalizes formula (7).

6.1 Sharpness

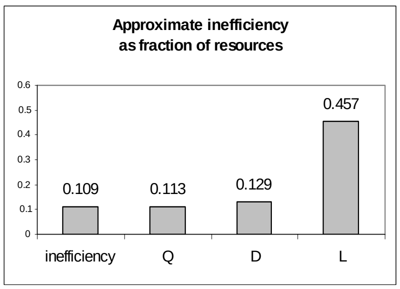
How sharp are the three approximations of inefficiency? In general,

[16] Most utilities in the linear risk tolerance class satisfy this, such as CRRA >1, CARA, log, quadratic.

Proposition 3 *Suppose as in theorems 1, 2, 3. Then the linear, discrete, and quadratic approximations of inefficiency are increasingly sharper:*

$$\rho_x \leq Q_x \leq D_x \leq L_x$$

The sharpenings are dramatic in the calibrated economy of section 3.2 with CRRA $\beta = 2.5$. As fractions of total current consumption $r_0 = \Sigma x_0^h$, they are:



relative errors $\frac{\hat{\rho}-\rho}{\rho}$

Q_x	D_x	L_x
$+3.7\%$	$+18.3\%$	$+319.3\%$

The quadratic is excellent, the discrete very good[17], the linear preposterous.

6.2 Computation with CRRA felicities

Approximations (17), (18) are explicit up to F^*, which is easy to compute in the following benchmark.

Proposition 4 *Suppose households' future utilities* $u_{\mathbf{1}}^h = \delta^h V$ *differ only in the patience parameters* $\delta^h > 0$, *having the same von Neumann-Morgenstern transform* $V := \Sigma \pi_s v(c_s)$ *of the felicity* $v(c) = \frac{1}{1-\beta} c^{1-\beta}$ *with CRRA* $\beta > 1$. *Then the value of problem (16) is*

$$F^* = \overline{\delta} V(r_{\mathbf{1}})$$

where $\overline{\delta} := \left[\Sigma \left(\mu^h \delta^h \right)^{\frac{1}{\beta}} \right]^\beta$ *and* $r_{\mathbf{1}}$ *are total future resources.*

Substituting this in expression (18) with $\mu^h = \frac{1}{u_0^{h\prime}}$ then gives a closed formula for approximate inefficiency. Further, in the quadratic approximation it is easy to compute $\overline{T}_0 = \frac{r_0}{\beta}$ if $u_0^h = v$.

In contrast, a closed formula for inefficiency is hopeless, even in this simplified setting. By remark 1, the equation defining willingness to pay is $\frac{1}{1-\beta} (x_0 - w)^{1-\beta} + \delta \Sigma \pi_s \frac{1}{1-\beta} (x_s + z_s)^{1-\beta} = \overline{u}$, the status quo utility. Solving for w shows inefficiency is hopeless indeed:

$$\sup_{x_{\mathbf{1}} + z \gg 0, \Sigma z^h = 0} \Sigma \left[(1-\beta) \overline{u}^h - \delta^h \Sigma \pi_s \left(x_s^h + z_s^h \right)^{1-\beta} \right]^{\frac{1}{1-\beta}}$$

[17] If the total risk tolerance $\overline{T}_0$ is high enough, one can conclude the quadratic is only marginally better than the discrete one, for then a first order Taylor expansion shows $Q \approx \frac{\overline{T}_0 D}{\sqrt{\overline{T}_0^2 + 2\overline{T}_0 \cdot D}} \approx \frac{\overline{T}_0 D}{\sqrt{\overline{T}_0^2 + 0}} = D$. But for practical purposes, this conclusion entails computing $\overline{T}_0$, and then one may as well compute the quadratic.

7 Extension to multiple goods

A final matter is whether the quantitative notion of inefficiency and the results extend to economies with multiple commodities. This is relevant since most contributions on the existence of inefficiency of equilibria with incomplete asset markets rely on the existence of multiple goods.

There is a simple extension of willingness to pay and inefficiency to the case of $L > 1$ commodities per state. It sacrifices generality slightly for simplicity. Thus suppose that preferences admit utility representations that are time separable, $u_0(x_0) + u_{\mathbf{1}}(x_{\mathbf{1}})$, where $(x_0, x_{\mathbf{1}}) \in R_+^{L(1+S)}$, and that $v_0 = v_0(p_0, I_0)$ denotes the indirect utility associated with u_0. Define a pseudo state space as $\{1, ..., S\} \times \{1, ..., L\}$, with $S^* := SL$ future states. Define a pseudo utility on $\mathbb{R}_+^{1+S^*}$ by $\tilde{u}(I_0, y_{\mathbf{1}}) := v_0(p_0, I_0) + u_{\mathbf{1}}(y_{\mathbf{1}})$, given period 0 prices p_0. The **willingness to pay** for a change $z \in \mathbb{R}^{S^*}$ in future consumption, in terms of current consumption, is by definition the supremum $w \in \mathbb{R}$ such that

$$\tilde{u}(I_0^* - w, x_{\mathbf{1}} + z) \geq \tilde{u}(I_0^*, x_{\mathbf{1}}) \tag{19}$$

where $I_0^* := p_0 x_0$ is the income necessary for the status quo current consumption x_0, in analogy to (4). We note that if x_0 is optimal in that $v_0(p_0, I_0^*) = u_0(x_0)$, then the right side of (19) is just the status quo welfare $u_0(x_0) + u_{\mathbf{1}}(x_{\mathbf{1}})$, so $w(0) = 0$ provided u_0 is increasing.

Definition 2 *Given a current spot price $p_0 \in R_{++}^L$, the **inefficiency** of $(\succeq, x)$ is the value ρ_{x,p_0} of*

$$\sup \Sigma w^h(z^h) \qquad s.t. \qquad x_{\mathbf{1}} + z \gg 0,\ \Sigma z^h = 0 \tag{20}$$

Proposition 1 extends:

Proposition 5 *Suppose u_0 is continuous and increasing. Suppose $z \in \mathbb{R}^{HS^*}$ is feasible for (20). Then it is a solution iff the $(I_0^{*h}, x_{\mathbf{1}}^h) + (-w^h, z^h)$ define a Pareto optimum (with respect to pseudo utilities).*

The lemmata describing the three approximations of willingness to pay extend :

Lemma 2 (linear) *Suppose the preference is reflexive and convex, and its indifference set through $x \gg 0$ is differentiable. The willingness to pay for z is at most*

$$w^h \leq \nabla^h z^h$$

Lemma 3 (discrete) *Suppose Du_0 is strictly positive and $D^2 u_0$ negative definite, let $v_0' := \frac{dv_0'(p_0, I_0)}{dI_0}$ be marginal utility. Then willingness to pay is at most*

$$w \leq \frac{\Delta}{v_0'}$$

Lemma 4 (quadratic) *Assume $v_0''' \geq 0 > v_0''$. If $\Delta \geq 0$ then willingness to pay is at most*

$$w \leq -T_0 + \sqrt{T_0^2 + 2T_0 \cdot \frac{\Delta}{v_0'}}$$

where $T_0 := -\frac{v_0'}{v_0''}$.

As before, these approximations are unambiguously ranked:

Proposition 6 *Suppose as in the lemmata. Then these approximations of willingness to pay are increasingly sharper,*

$$\textit{inefficiency} \leq \textit{quadratic} \leq \textit{discrete} \leq \textit{linear}$$

Lastly, the law of diminishing returns also holds, by an identical argument.

That these upper bounds on willingness to pay translate into upper bounds on inefficiency is merely a notational extension, omitted.

8 Appendix

8.1 Calibrating patience parameters

For bond purchases θ to maximize the welfare $u(x) = v(x_0) + \delta^h \Sigma \pi_s v(x_s)$ of the consumption $x = (I_0 - q\theta, I_{\mathbf{1}} + \theta)$ they finance, they must satisfy the FOC:

$$x_0^{-\beta} q = \delta \Sigma \pi_s x_s^{-\beta}$$

Substituting the price $q^{bond} = \frac{1}{1.65}$,

$$\delta = \frac{1}{1.65} \frac{x_0{}^{-\beta}}{\Sigma \pi_s x_s^{-\beta}} \tag{21}$$

Evaluating (21) at the consumption distribution x and the state probabilities π gives δ^h as a function of the CRRA parameter β.

8.2 Proposition 1

Proof. Necessity by contradiction. Let z be a solution where the $x^h + (-w^h, z^h)$ admit a Pareto superior reallocation $y \in \mathbb{R}_{++}^{H(S+1)}$, so that $\Sigma y_0^h = r_0 - \Sigma w^h$ and $\Sigma y_{\mathbf{1}}^h = \Sigma x_{\mathbf{1}}^h$, and $y^h \succsim^h x^h + (-w^h, z^h)$ without indifference for some i. By (4) and transitivity, $y^h \succsim^h x^h$. Reduce y_0^i to $y_0^i - \epsilon$ by some $\epsilon > 0$. By continuity in current consumption, a small enough ϵ is feasible, in that still $y^i + (-\epsilon, 0) \succ^i x^i + (-w^i, z^i) \succsim^i x^i$. But now this modified $\tilde{y}_0$ (identical to y_0 but for household i) sums to $r_0 - \Sigma w^h - \epsilon$. Set $\tilde{y} := (\tilde{y}_0, y_{\mathbf{1}})$ so $\tilde{y}^h \succsim^h x^h$. Set $a := \tilde{y} - x$ so that $\tilde{y}^h = x^h + (-(-a_0^h), a_{\mathbf{1}}^h) \succsim^h x^h$ and $\Sigma a_{\mathbf{1}}^h = 0$. This shows that $w^h(a_{\mathbf{1}}^h) \geq -a_0^h$ hence $\Sigma w^h(a_{\mathbf{1}}^h) \geq -\Sigma a_0^h = -\left[(r_0 - \Sigma w^h - \epsilon) - r_0\right] = \Sigma w^h + \epsilon$, which exceeds $\Sigma w^h = \rho_x$, the supremum total willingness to pay for some future reallocation z (which $a_{\mathbf{1}}$ is), a contradiction.

Sufficiency by contraposition. Let $\tilde{z}$ be a counterexample to z being a solution. Consider the two allocations $(x_0^h - w^h(\tilde{z}^h), x_{\mathbf{1}}^h + \tilde{z}^h)$, $(x_0^h - w^h(z^h), x_{\mathbf{1}}^h + z^h)$. By remark 1, they are indifferent to x^h, hence to each other by transitivity. By hypothesis, $\Sigma w^h(\tilde{z}^h) > \Sigma w^h(z^h)$, so that the first allocation has lower current resources than, but of course equal future resources to, the second. Thus change the first allocation by taking the aforementioned current slack and distributing it evenly over households; being increasing in current consumption, this makes the (so modified) first allocation preferred to the second one, using the same resources, showing the second one is not Pareto optimal. ■

8.3 Corollary 1

Proof. Necessity. By hypothesis, x is Pareto efficient, and $w^h(0) = 0$ in any case, so $(x_0^h - w^h(0), x_{\mathbf{1}}^h) = x^h$ is Pareto efficient. By proposition 1, $z = 0$ solves problem (5). So the problem's value is $\rho_x = \Sigma w^h(0) = \Sigma 0 = 0$.

Sufficiency. Suppose $y \in \mathbb{R}_{++}^{H(S+1)}, \Sigma y^h = \Sigma x^h$ with $y^h \succsim^h x^h$. Rewrite as $y^h = x^h + (-(x_0^h - y_0^h), z^h) \succsim^h x^h$, defining $z := y_{\mathbf{1}} - x_{\mathbf{1}}$. Then $w^h(z^h) \geq x_0^h - y_0^h$ by definition of willingness to pay, so $\Sigma w^h(z^h) \geq \Sigma(x_0^h - y_0^h) = 0$. Conversely, $\Sigma w^h(z^h) \leq \rho_x$ because z is feasible for problem (5) and ρ_x its value. So if $\rho_x = 0$, then $\Sigma w^h(z^h) = 0$. So the $a^h := x^h + (-w^h(z^h), z^h)$ satisfy $\Sigma a^h = \Sigma x^h$, are Pareto efficient by remark 1, and indifferent to x. Thus x is itself Pareto efficient. ■

8.4 Lemma 1, theorem 1

The proposition relies on a simple **global-infinitesimal principle** for smooth convex preferences: if $x + (z_0, z) \succeq x$ then $(1, \nabla(x)) \cdot (z_0, z) \geq 0$.

Proof. Let w be the willingness to pay for z, so that $x + (-w, z) \sim x$. Since also $x \sim x$, convexity implies $x + t(-w, z) \succsim x$ with $* = tz$ for, say, $t = \frac{1}{2}$. By the global-infinitesimal principle, $(1, \nabla(x)) \cdot (-w, z) \geq 0$, i.e. $w \leq \nabla(x)) \cdot z$. ■

Proof. By principle 1, it suffices to show the value of $\sup_{x_{\mathbf{1}}+z \gg 0, \Sigma z^h = 0} \Sigma \nabla^h z^h$ is $\Sigma \left(\triangledown^* - \triangledown^h\right) x_{\mathbf{1}}^h$. Now, $\Sigma z^h = 0$ implies $\Sigma \nabla^h z^h = \Sigma \left(\triangledown^* - \triangledown^h\right)(-z^h) = *$. In turn, $x_{\mathbf{1}} + z \geq 0$ and $\triangledown^* - \triangledown^h \geq 0$ imply $* \leq \Sigma \left(\triangledown^* - \triangledown^h\right) x_{\mathbf{1}}^h$. So the value is at most the claimed one, which is actually achieved by $z := -x_{\mathbf{1}}$. ■

8.5 Proposition 2

Proof. $\boxed{w(z(t)) \geq tw(z)}$ By remark 1, the willingness to pay $w = w(z)$ makes (4) hold with indifference. Thus w solves $x + (-w, z) \sim x$. Of course, $x + (-0, 0) \sim x$. By convexity of the preference, the t-convex combination of the latter left sides is weakly preferred to $x : x + (-tw, z(t)) \succsim x$, where $t' := 1 - t$. Since $w(z(t))$, the wtp for $z(t)$, is the supremum s such that $x + (-s, z(t)) \succsim x$, it follows $w(z(t)) \geq tw$.

$\boxed{\text{Concavity}}$ Fix $s, t \in [0, 1]$ and $a \in [0, 1]$; we want $w(z(as + a't)) \geq aw(z(s)) + a'w(z(t))$, where $a' := 1 - a$. The following indifferences hold: $x + (-w(z(s)), z(s)) \sim x, x + (-w(z(t)), z(t)) \sim x$. By convexity of the preference, $x + (-aw(z(s)) - a'w(z(t)), *) \succsim x$ where $* = az(s) + a'z(t) = z(as + a't)$. As above, by definition of wtp as the supremum, $w(*) \geq aw(z(s)) + a'w(z(t))$. ■

8.6 Theorem 2

We analyze equation (6), which characterizes the willingness to pay w for a change z in future consumption, repeated here:

$$u_0(x_0) - u_0(x_0 - w) = \Delta \tag{22}$$

Note, $\Delta, w = 0$ satisfy this equation; since u_0 is increasing, one is positive (negative) iff the other is, hence

Remark 3 *The signs of* w, Δ *agree.*

Lemma 5 (discrete approximation of willingness to pay) *Suppose* $u_0' > 0 \geq u_0''$. *Then willingness to pay is at most*[18]

$$w \leq \frac{\Delta}{u_0'} \tag{23}$$

Proof. By the Fundamental Theorem of Calculus, $u_0(x_0) - u_0(x_0 - w) = w \int_0^1 u_0'(x_0 - w + tw)dt$. This and equation (22) imply

$$w = \frac{\Delta}{\int_0^1 u_0'(x_0 - w + tw)dt} \tag{24}$$

Since $u_0'' \leq 0$, the integrand is bounded as $u_0'(x_0) \lesseqqgtr u_0'(x_0 - w + tw)$ according as $w \gtreqqless 0$[19], i.e. according as $\Delta \gtreqqless 0$ (by remark 3), giving (23). ■

Proof. By principle 1 and upper bound (23), it suffices to show $\sup_{x_{\mathbf{1}}+z \gg 0, \Sigma z^h = 0} \Sigma \frac{\Delta^h}{u_0^{h\prime}}$ has value $F^* - F(x_{\mathbf{1}})$. Now,

$$\Sigma \frac{\Delta^h}{u_0^{h\prime}} = \Sigma \mu^h \left[u_{\mathbf{1}}^h(x_{\mathbf{1}}^h + z^h) - u_{\mathbf{1}}^h(x_{\mathbf{1}}^h) \right] = \Sigma \mu^h u_{\mathbf{1}}^h(x_{\mathbf{1}}^h + z^h) - \Sigma \mu^h u_{\mathbf{1}}^h(x_{\mathbf{1}}^h) = F(x_{\mathbf{1}} + z) - F(x_{\mathbf{1}})$$

On changing variables as $y_{\mathbf{1}} = x_{\mathbf{1}} + z$ and recalling the definition of F^*, this is shown. ■

8.7 Theorem 3

Lemma 6 (quadratic approximation of willingness to pay) *Suppose* $u_0''' \geq 0 > u_0'', -u_0'$.[20] *If* $\Delta \geq 0$ *then willingness to pay is at most*

$$w \leq \sqrt{T_0^2 + 2T_0 \cdot \frac{\Delta}{u_0'}} - T_0 \tag{25}$$

Proof. Rewrite expression (24) as $\frac{\Delta}{w} = \int$. (If $w = 0$, then $\Delta = 0$ by remark 3 and the inequality is trivial.) By the Fundamental Theorem of Calculus, $u_0'(x_0 - tw) = u_0' - w \int_0^t u_0''(x_0 - sw)ds$, so this integral is expressible as

$$\begin{aligned} \int &= \int_0^1 u_0'(x_0 - w + tw)dt = \int_0^1 u_0'(x_0 - tw)dt = \int_0^1 \left[u_0' - w \int_0^t u_0''(x_0 - sw)ds \right] dt \\ &= u_0' - w \int_0^1 \int_0^t u_0''(x_0 - sw)dsdt = u_0' - w \int_0^1 (1-t) u_0''(x_0 - tw)dt \end{aligned}$$

the latter being the identity $\int_0^1 \int_0^t f(s)dsdt = \int_0^1 (1-t)f(t)dt$. Since $u_0''' \geq 0$, in the last integrand we have $u_0'' \gtreqqless u_0''(x_0 - tw)$ for all $t \in [0,1]$ according as $w \gtreqqless 0$. Thus

$$\int \geq u_0' - w \int_0^1 (1-t) u_0'' dt = u_0' - w \frac{u_0''}{2}$$

[18] To ease notation we omit the argument when it is current consumption x_0 at the status quo, as in u_0' for $u_0'(x_0)$.
[19] That is, $u_0'(x_0) \leq u_0'(x_0 - w + tw)$ if $w \geq 0$; and $u_0'(x_0) \geq u_0'(x_0 - w + tw)$ if $w \leq 0$.
[20] Most utilities in the linear risk tolerance class satisfy this, such as CRRA >1, CARA, log, quadratic.

regardless of $w's$ sign. This and $\frac{\Delta}{w} = \int$ imply $\frac{\Delta}{w} \geq u_0' - w\frac{u_0''}{2}$. Dividing by $-u_0'' > 0$ and using the identity $-\frac{\Delta}{u_0''} = T_0 \frac{\Delta}{u_0'}$ imply

$$0 \geq T_0 + \frac{w}{2} - T_0 \frac{\Delta}{wu_0'} \tag{26}$$

Finally, suppose $\Delta > 0$. Then multiplying (26) by w, which is positive by remark 3, implies the quadratic $0 \geq \frac{w^2}{2} + T_0 w - T_0 \frac{\Delta}{u_0'}$. This being convex, w lies between the roots, hence is at most the greater root, which is $-T_0 + \sqrt{T_0^2 + 2T_0 \frac{\Delta}{u_0'}}$. ■

Proof. By principle 1 and upper bound (25), it suffices to show $* = \sup_{x_1+z \gg 0, \Sigma z^h = 0} \Sigma \left[-T_0^h + \sqrt{T_0^{h2} + 2T_0^h \cdot D^h} \right]$ has value at most the Q_x in (18), with $D^h := \frac{\Delta^h}{u_0^{h\prime}}$. This is increasing in the D^h. The proof of theorem 2 shows that $\sup_{x_1+z \gg 0, \Sigma z^h = 0} \Sigma D^h$ equals $D_x = F^* - F(x_1)$. Thus $*$ is at most

$$\sup_{\Sigma D^h \leq D_x} \Sigma \left[-T_0^h + \sqrt{T_0^{h2} + 2T_0^h \cdot D^h} \right]$$

Note, $\Sigma D^h \leq D_x$ will be binding, as the objective is monotone in the D^h. Since the objective is concave and the constraint linear, the constraint qualification holds, and Kuhn-Tucker multipliers exist. The FOC are $\frac{T_0^h}{\sqrt{\cdot}} = \lambda$. Rearranging, $\frac{T_0^{2h}}{\lambda^2} = \sqrt{\cdot}^2 = T_0^{h2} + 2T_0^h \cdot D^h$ or $(\frac{1}{\lambda^2} - 1)T_0^h = 2 \cdot D^h$, which aggregated implies $(\frac{1}{\lambda^2} - 1) = \frac{2D}{\overline{T}_0}$ so that $D^h = D_x \frac{T_0^h}{\overline{T}_0}$. The insides of the square roots become

$$T_0^{h2} + 2T_0^h \cdot D_x \frac{T_0^h}{\overline{T}_0} = T_0^{h2} \left(1 + \frac{2D}{\overline{T}_0} \right)$$

so the objective becomes

$$\Sigma \left[-T_0^h + T_0^h \sqrt{1 + \frac{2D}{\overline{T}_0}} \right] = \left(-1 + \sqrt{1 + \frac{2D}{\overline{T}_0}} \right) \Sigma T_0^h = -\overline{T}_0 + \sqrt{\overline{T}_0^2 + 2\overline{T}_0 D_x}$$

■

8.8 Proposition 3

The proposition relies on a fact: a direction that is globally improving is necessarily locally improving, that tangents lie above the graph of a concave function:

Lemma 7 *Suppose $f : A \to R$ is C^1 and concave. Then for all $a + z \in A$, interior $a \in A$*

$$f(a + z) - f(a) \leq Df(a)z \tag{27}$$

Proof. of proposition 3. $\boxed{D_x \leq L_x}$. We recall $D_x = \sup_{x_1+z \gg 0, \Sigma z^h = 0} \Sigma \frac{\Delta^h}{u_0^{h\prime}}$ and $L_x = \max_{x_1+z \gg 0, \Sigma z^h = 0} \Sigma \nabla^h z^h$, so it suffices that $\frac{\Delta^h}{u_0^{h\prime}} \leq \nabla^h z^h$. This is immediate from inequality (27), as $\Delta = u_1(x_1 + z) - u_1(x_1)$ and $\nabla = \frac{D_{x_1} u}{u_0'}$. $\boxed{Q_x \leq D_x}$. Theorem 3 expresses one in terms of the other, $Q_x = \sqrt{\overline{T}_0^2 + 2\overline{T}_0 \cdot D_x} - \overline{T}_0$. Inequality (27) with $f(D) := \sqrt{\overline{T}_0^2 + 2\overline{T}_0 \cdot D}$ at $D = 0$ gives $Q_x = \sqrt{\overline{T}_0^2 + 2\overline{T}_0 \cdot D_x} - \sqrt{\overline{T}_0^2} \leq \frac{1}{2\sqrt{\overline{T}_0^2}} 2T_0 \cdot (D_x - 0) = D_x$. ■

8.9 Proposition 4

We rewrite the objective $F = \Sigma \mu^h u^h_{\mathbf{1}}(y^h_{\mathbf{1}}) = \Sigma \mu^h \delta^h \Sigma \pi_s v(y^h_s) = \Sigma \pi_s m^h v(y^h_s)$ where $m^h := \mu^h \delta^h$. Clearly, $F^* = \Sigma \pi_s a_s$ where

$$a_s := \max_{\Sigma y^h_s = r_s} \Sigma m^h v(y^h_s)$$

The Kuhn-Tucker method leads to the solution $y^h_s = r_s \tau^h$, where $\tau^h = \frac{\left(m^h\right)^{\frac{1}{\beta}}}{M}$ with $M := \Sigma \left(m^h\right)^{\frac{1}{\beta}}$. Thus $a_s = \Sigma m^h v(r_s \tau^h) = k r_s^{1-\beta}$ where $k := \frac{1}{1-\beta} \Sigma m^h \left(\tau^h\right)^{1-\beta}$. This simplifies to $k = \frac{M^\beta}{1-\beta}$, on substituting τ. Thus $a_s = \frac{M^\beta}{1-\beta} r_s^{1-\beta} = M^\beta v(r_s)$ and $F^* = M^\beta \cdot \Sigma \pi_s v(r_s) = M^\beta \cdot V(r_{\mathbf{1}})$.

References

[1] Alvarez F., U.J. Jermann, 2004, "Using Asset Prices to Measure the Cost of Business Cycles," *Journal of Political Economy* **112**(6), 1223-1256.

[2] Barro, R.J., 2007, "On the Welfare Costs of Consumption Uncertainty", Harvard University mimeo.

[3] Citanna, A., A. Kajii and A. Villanacci, 1998, "Constrained Suboptimality in Incomplete Markets: A General Approach and Two Applications," *Economic Theory* **11**, 495-522.

[4] Citanna, A., H. Polemarchakis and M. Tirelli, 2006, "The Taxation of Trades in Assets," *Journal of Economic Theory* **126**(1), 299-313.

[5] Debreu, G., 1959, *Theory of Value*, Yale University Press, New Haven.

[6] Debreu, G., 1951, "The Coefficient of Resource Utilization," *Econometrica* **19**(3), 273-292.

[7] Geanakoplos, J. and H. Polemarchakis, 1990, "Observability and Optimality," *Journal of Mathematical Economics* **19**, 153-165.

[8] Geanakoplos, J. and H. Polemarchakis, 1986, "Existence, Regularity and Constrained Suboptimality of Competitive Allocations when the Asset Market is Incomplete," in: Heller, W., Starr, R., Starrett, D. (Eds.) *Uncertainty, Information and Communication: Essays in Honor of Kenneth J. Arrow, Vol.III.* Cambridge University Press, Cambridge, 65 - 96.

[9] Geanakoplos J., 1990, "An Introduction to General Equilibrium with Incomplete Asset Markets", *Journal of Mathematical Economics*, **19**, 1-38.

[10] Gottschalk, P. and S. Danziger, 1997, "Family Income Mobility-How Much Is There and Has It Changed?" *PSC Research Report* No. 97-405.

[11] İmrohoroğlu, A., 1989, "The Cost of Business Cycles with Indivisibilities and Liquidity Constraints," *Journal of Political Economy* **97** (6), 1364-83.

[12] Krebs, T., 2003, "Growth and Welfare Effects of Business Cycles in Economies with Idiosyncratic Human Capital Risk," *Journal of Economic Dynamics* **6**, 846-868.

[13] Kocherlakota, N. R., 1996, "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature* **34**(1), 42-71.

[14] Krusell, P. and A. Smith, 1999, "On the Welfare Effects of Eliminating Business Cycles," *Journal of Economic Dynamics* **2**, 245-272.

[15] Kurz, M. 2005, "Measuring the Ex-Ante Social Cost of Aggregate Volatility", Stanford University, mimeo.

[16] Levine, D. K., W. R. Zame, 2002, "Does market incompleteness matter?," *Econometrica* **70**(5), 1805-1839.

[17] Lucas, R., 1987, *Models of business cycles*, Oxford: Basil Blackwell.

[18] McNeill, J., 1999, "Shares of Income Received by Quintiles When Equivalent Income Is Used as the Measure of Income", US Census Bureau, http://www.census.gov/hhes/www/income/newydata.pdf.

[19] Magill, M. and M. Quinzii, 1996, *Theory of Incomplete Markets*, MIT Press, Cambridge, Massachusetts.

[20] Nagata, R., 2005, "Inefficiency of equilibria with incomplete markets," *Journal of Mathematical Economics* 41(7), 887-897.

[21] Panel on Poverty and Family Assistance, 1995, *Measuring Poverty: A New Approach*, National Academy Press, Washington, DC.

[22] Ríos-Rull, J.V., 1994, "On the Quantitative Importance of Market Completeness," *Journal of Monetary Economics* **34**, 462-496.

[23] Stiglitz J. E. 1982, "The Inefficiency of Stock Market Equilibrium", *Review of Economic Studies* **49**, 241-261.

[24] Tirelli, M., 2003, "Income taxation when markets are incomplete," *Decisions in Economics and Finance* **26**(2), 97-128.

[25] Turner, S., 2003, "Pareto Improving Taxation in Incomplete Markets," Brown University Working Paper 2006-03.