Taxation under Learning by Doing

Miltos Makris Alessandro Pavan

March 2, 2018

Learning by Doing

- **Learning-by-doing (LBD)** :
  - positive effect of time spent at work on productivity
  - human capital investment side-product of labor supply

- LBD: significant source of productivity growth
  - Dustmann and Meghir (2005)
    - in first 2 years of employment, wages grow, on average, by 8.5% in 1th year and 7.5% in 2nd
  - Blundell and MaCurdy (1999) and Farber (1999)
    - overviews of effects of work experience on wage dynamics

This Paper

- Effects of LBD on optimal tax codes
- Dynamic Mirrleesian economy in which agents' productivity
  - their own private information
  - stochastic
  - evolves endogenously over lifecycle (due to LBD)
- Novel effects contributing to higher labor wedges
- Quantitatively significant impact on optimal codes
  - level
  - progressivity
  - dynamics
- Dynamic mechanism design with endogenous types

Related literature

- **Optimal taxation**: Mirrlees (1971), Diamond (1998), Saez (2001)...
  - static
  - exogenous productivity
- **New Dynamic Public Finance:** Albanesi and Sleet (2006), Kocherlakota (2010), Gorry and Oberfield (2012), Kapicka (2013), Farhi and Werning (2013), and Golosov et al. (2016)...
  - dynamic
  - exogenous productivity
- **Taxation w. Human Capital Accumulation:** Krause (2009), Best and Kleven (2013), Kapicka (2006, 2015a,b), Kapicka and Neira (2016), and Stantcheva (2016)
  - LBD: side-product of labor supply (cannot be controlled separately)
  - stochastic effect on future productivity
  - time-evolving private information

Road Map

- Simple texbook environment
- Risk aversion
- Utilitarian objective
- Quantitative analysis
- Conclusions

Simple Environment

- T=2 (case $T=\infty$: "Incentives for Endogenous Types")
- θ_t: productivity
  - privately observed by worker at beginning of period t
- F_1: cdf of initial distribution (density f_1)
- $F_2(\cdot|\theta_1, y_1)$: cdf of θ_2
  - dependence on y_1: LBD
- Example:

$$\theta_2 = Z_2(\theta_1, y_1, \varepsilon_2) = \theta_1^{\rho} y_1^{\zeta} \varepsilon_2$$

  ζ: **intensity of LBD**

- **Impulse Response**

$$\boxed{I_1^2(\theta, y_1) = \left.\frac{\partial Z_2(\theta_1, y_1, \varepsilon_2)}{\partial \theta_1}\right|_{\varepsilon_2 : Z_2(\theta_1, y_1, \varepsilon_2) = \theta_2} = \rho \frac{\theta_2}{\theta_1}}$$

where $\theta \equiv (\theta_1, \theta_2)$

Simple Environment

- Worker's payoff:

$$U^A = \sum_t \delta^{t-1} (c_t - \psi(y_t, \theta_t))$$

- $\psi(y_t, \theta_t) = \frac{1}{1+\phi} \left(\frac{y_t}{\theta_t} \right)^{1+\phi}$
- Allocation rule $\chi(\theta) = (y_t(\theta^t), c_t(\theta^t))_{t=1,2}$
- Worker expected life-time utility

$$V_1(\theta_1) = \mathbb{E}^{\lambda[\chi]|\theta_1} \left[\sum_t \delta^{t-1} \left(c_t(\tilde{\theta}^t) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) \right) \right]$$

where $\lambda[\chi]$ is endogenous distribution over $\Theta = \Theta_1 \times \Theta_2$ under χ

Principal's (dual) problem:

- Principal's Rawlsian problem

- Maximizing expected tax revenues

$$R = \mathbb{E}^{\lambda[\chi]}\left[\sum_t \delta^{t-1}\left(y_t(\tilde{\theta}^t) - c_t(\tilde{\theta}^t)\right)\right]$$

subject to Rawlsian constraint

$$\min_{\theta_1} V_1(\theta_1) \geq \kappa$$

First Best: period-2 output

- For any $\theta = (\theta_1, \theta_2)$:

$$\max_{\mathbf{y_2}} [y_2 - \psi(y_2, \theta_2)]$$

- FOC:

$$\psi_y(y_2^*(\theta), \theta_2) = 1,$$

$\Rightarrow$ output driven by marginal production cost

First Best: period-1 output

- For any θ_1:

$$\max_{\mathbf{y_1}} \left\{ y_1 - \psi(y_1, \theta_1) + \delta \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - \psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right] \right\}$$

- FOC:

$$1 + LD_1^{\chi}(\theta_1) = \psi_y(y_1(\theta_1), \theta_1)$$

where

$$LD_1^{\chi}(\theta_1) \equiv \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - \psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]$$

$\Rightarrow$ output driven **also** by LBD impact on future expected surplus via its effect on future conditional distribution
$\Rightarrow$ Higher output under LBD for any given θ_1 (due to FOSD and increasing period-2 surplus)

Second Best: Incentive Compatibility

- Continuation utility (history $\theta = (\theta_1, \theta_2)$):

$$V_2(\theta) \equiv c_2(\theta) - \psi(y_2(\theta), \theta_2)$$

- IC-2: for any θ_1, $V_2(\theta_1, \cdot)$ Lipschitz continuous and s.t. (Mirrlees)

$$V_2(\theta_1, \theta_2) = V_2(\theta_1, \underline{\theta}_2) - \int_{\underline{\theta}_2}^{\theta_2} \psi_\theta(y_2(\theta_1, s), s) ds$$

- IC-1: $V_1(\cdot)$ Lipschitz continuous and s.t. (Pavan, Segal, Toikka)

$$V_1(\theta_1) = V_1(\underline{\theta}_1)$$

$$- \int_{\underline{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s) ds + \delta \mathbb{E}^{\lambda[\chi]|s} \left[I_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right] \right\} ds$$

- In addition, IC also requires $y_1(\cdot)$ and $y_2(\cdot)$ satisfy integrability constraints (ignored and checked ex post)

Second Best: Handicaps

- Expected tax revenues equal:

$$R = \mathbb{E}^{\lambda[\chi]}\left[\sum_t \delta^{t-1}\left(y_t(\tilde{\theta}^t) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) - h_t(\tilde{\theta}^t, y^t(\tilde{\theta}^t))\right)\right] - V_1(\underline{\theta}_1),$$

  - first-period "handicap":

$$h_1(\theta_1, y_1) \equiv -\tfrac{1}{\gamma_1(\theta_1)}\psi_\theta(y_1, \theta_1) \qquad \textit{where} \;\; \gamma_1(\theta_1) \equiv \frac{f_1(\theta_1)}{1 - F_1(\theta_1)}$$

  - second-period "handicap":

$$h_2(\theta, y) \equiv -\tfrac{I_1^2(\theta, y_1)}{\gamma_1(\theta_1)}\psi_\theta(y_2, \theta_2)$$

- Handicaps: costs to planner due to asymmetric information

Second Best: period-2 output

- Given $\theta = (\theta_1, \theta_2)$,

$$1 = \psi_y(y_2, \theta_2) - \frac{1}{\gamma_1(\theta_1)} I_1^2(\theta, y_1(\theta_1)) \psi_{y\theta}(y_2, \theta_2)$$

- Value of distorting period-2 output at $\theta = (\theta_1, \theta_2)$: smaller rents to (period-1!) types $\theta_1' > \theta_1$
    - **impulse responses**

Second Best: period-1 output

- Given θ_1,

$$1 + LD_1^{\chi}(\theta_1)$$

$$= \psi_y(y_1(\theta_1), \theta_1) - \frac{1}{\gamma_1(\theta_1)} \psi_{y\theta}(y_1(\theta_1), \theta_1)$$

$$+\delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)} \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]$$

- Value of distorting period-1 output: smaller rents to higher (period-1) types
  - smaller rents in future periods
- Two channels through which LBD affects cost of future rents:
  - change in distribution of θ_2
  - change in impulse response of θ_2 to θ_1 (hence handicaps)

Second Best: Labor Wedges

Definition

Labor wedges:

$$W_1(\theta_1) \equiv 1 - \frac{\psi_y(y_1(\theta_1), \theta_1)}{1 + LD_1^{\chi}(\theta_1)} \text{ and } W_2(\theta) \equiv 1 - \psi_y(y_2(\theta), \theta_2).$$

- Relative wedges:

$$\widehat{W_t} \equiv \frac{W_t}{1 - W_t}$$

Second Best: Wedges

Proposition

Under risk neutrality and Rawlsian objective,

$$\boxed{\widehat{W}_t \equiv \hat{W}_t^{RRN} + \Omega_t}$$

where

$$\hat{W}_t^{RRN} \equiv -\frac{I_1^t(\theta^t, y^{t-1}(\theta^{t-1}))}{\gamma_1(\theta_1)} \frac{\psi_{y\theta}(y_t(\theta^t), \theta_t)}{\psi_y(y_t(\theta^t), \theta_t)}$$

are wedges without LBD and

$$\Omega_1 \equiv \delta \frac{\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[h_2(\tilde{\theta}, y^t(\tilde{\theta})) \right]}{\psi_y(y_1(\theta_1), \theta_1)}$$

and

$$\Omega_2 \equiv 0$$

are corrections due to LBD

Effects of LBD on wedges

- Suppose $\psi(y_t, \theta_t) = \frac{1}{1+\phi}\left(\frac{y_t}{\theta_t}\right)^{1+\phi}$ and $\theta_2 = \theta_1^{\rho} y_1^{\zeta} \varepsilon_2$
- Then
  - $\hat{W}_1(\theta_1) > \hat{W}_1^{RRN}(\theta_1)$
  - $\hat{W}_1(\theta_1) - \hat{W}_2(\theta) > \hat{W}_1^{RRN}(\theta_1) - \hat{W}_2^{RRN}(\theta)$
- When, in addition F_1 has a Pareto tail progressivity of $\hat{W}_1(\theta_1)$ higher than progressivity of $\hat{W}_1^{RRN}(\theta_1)$ at tail

Effects of LBD on wedges – Intuition

- LBD contributes to higher expected period-2 handicaps

  $\Rightarrow$ extra benefit of lowering y_1

  $\Rightarrow$ higher wedges

- Expected period-2 rents increasing in θ_1

  $\Rightarrow$ benefit of distorting y_1 downwards stronger for higher θ_1

  $\Rightarrow$ more progressivity

- Effects of LBD declining with t

  $\Rightarrow$ wedges declining over life-cycle

Pareto a-la Kapicka (2013)

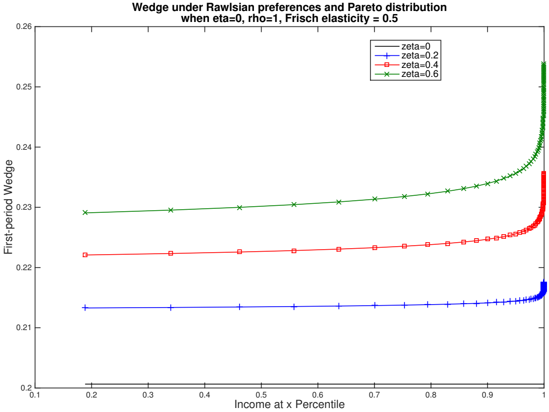

Figure: Period-1 wedges: risk-neutral Rawlsian Pareto case

Pareto-log-normal a-la Diamond (1998)

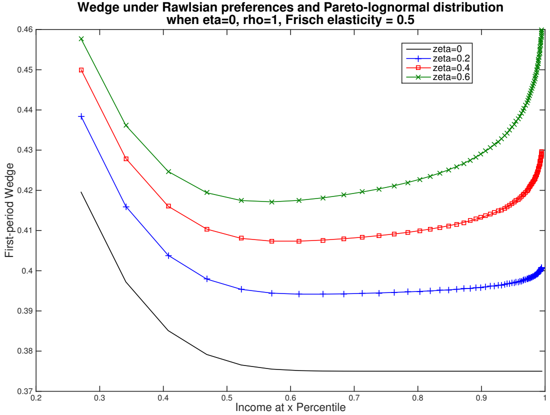

Figure: Period-1 wedges: risk-neutral Rawlsian Pareto-lognormal case

Risk Aversion

- Agent's utility from consumption: $v(c_t)$, with $v''(\cdot) < 0$
- One util compensation requires $1/v'(c_t)$ units of consumption
- Risk aversion increases cost of future information rents
- Effect of LBD: $RA(\theta_1)\Omega(\theta)$ where

$$RA(\theta_1) \equiv v'(c_1(\theta_1)) \int_{\theta_1}^{\overline{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1 - F_1(\theta_1)}$$

  is correction due to risk-aversion
- Risk aversion contributes to **amplification** of LBD **level** effect
  - risk aversion increases benefit of shifting future distribution towards lower types
- Risk aversion contributes to **amplification** of LBD **progressivity** effect
  - benefit more pronounced for high period-1 types: their expected future rents are higher
- BUT, risk aversion leads also to an **alleviation** of LBD level and progressivity effects
  - higher cost of future rents $\rightarrow$ lower future incomes (hence lower Ω_1)

Risk Aversion

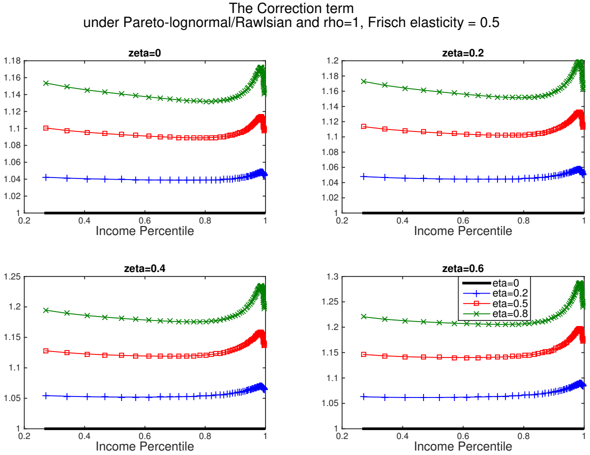

Figure: *RA* correction term: Rawlsian Pareto-lognormal case

Risk Aversion

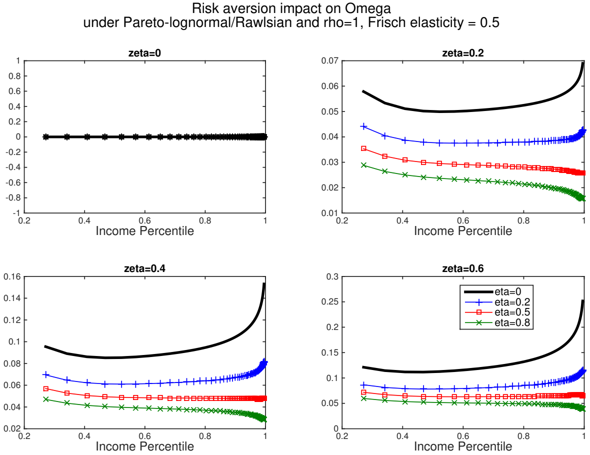

Figure: LBD term Ω: Rawlsian Pareto-lognormal case

Risk Aversion

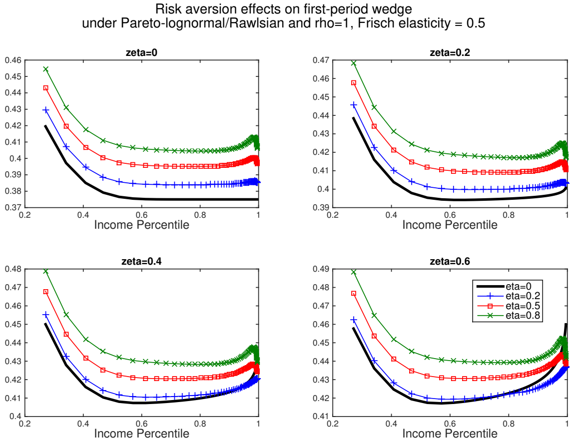

Figure: Period-1 wedges: risk-averse Rawlsian Pareto-lognormal case

Utilitarian

- Redistribution constraint:

$$\int_{\underline{\theta}_1}^{\overline{\theta}_1} V_1(\theta_1) dF_1(\theta_1) \geq \kappa$$

- Increasing lifetime utility by $v'(c_1(\theta_1))\Omega_1(\theta_1)$ now relaxes redistribution constraint

$$\boxed{\widehat{W_1}(\theta_1) = \hat{W}_1^{URA}(\theta_1) + [RA(\theta_1) - D(\theta_1)]\,\Omega(\theta_1),}$$

where

$$D(\theta_1) \equiv v'(c_1(\theta_1)) \int_{\underline{\theta}_1}^{\overline{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s)$$

is novel correction term reflecting higher Pareto weights assigned to types above $\underline{\theta}_1$

- Novel effect reduces amplification effect of risk aversion

Utilitarian

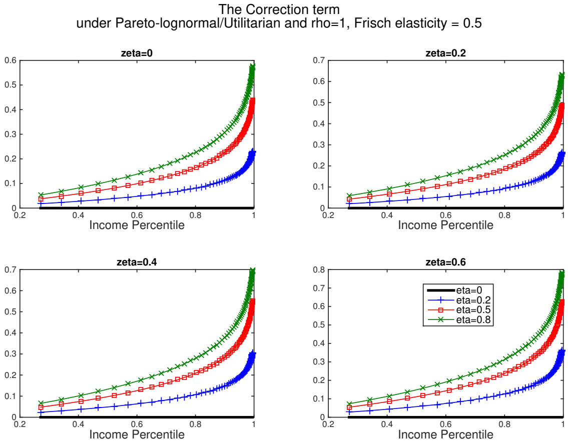

Figure: $RA - D$ correction term: Utilitarian Pareto-lognormal case

Utilitarian

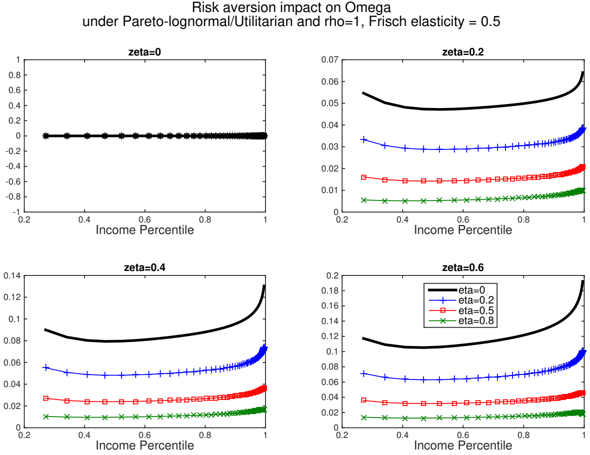

Figure: LBD term Ω: Utilitarian Pareto-lognormal case

Utilitarian

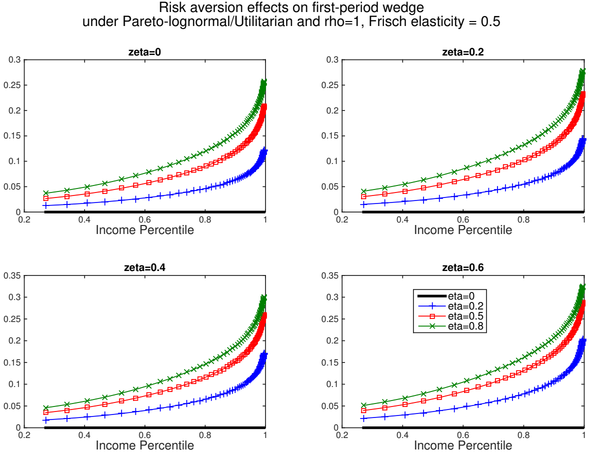

Figure: Period-1 wedges: Utilitarian Pareto-lognormal case

Lognormal

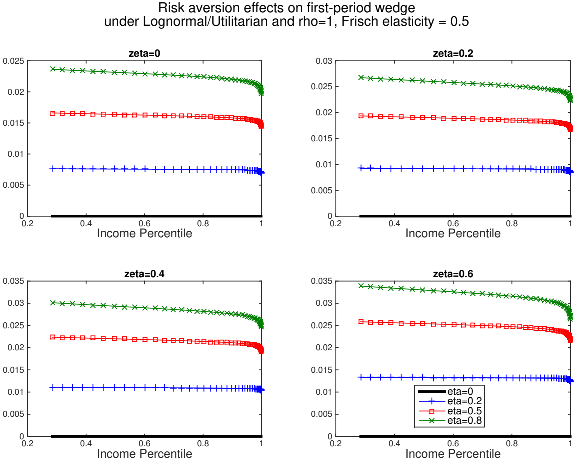

Figure: Period-1 wedges: Utilitarian Lognormal case

Quantitative Analysis

- Calibrate 40-working-years model with productivity changing at year 21
- Annual discount factor β
- LBD active in each of first 20 years (weights β^{s-1})
- $\beta = 1/(1+r)$
- Isomorphic to 2-period model with $\delta = \beta^{20}$ and $V_t = \frac{\hat{V}_{\mathbf{20}(t-\mathbf{1})+\mathbf{1}}}{\sum_{s=\mathbf{1}}^{\mathbf{20}} \beta^{s-\mathbf{1}}}$, $t = 1, 2.$
- U.S. income tax estimation from Heathcote et. al. (2016)

$$T(y) = y - e^{\tau_0} y^{1-0.181}$$

- $r = 0.04$, $v = log$, $\phi = 2$ $\theta_1 = h_1 \varepsilon_1$, and ε_t iid Pareto-Lognormal (λ, σ) with mean 1

Quantitative Analysis

Definition	Symbol	Value	As in
CRRA parameter	η	1	FW, K, GTT, S, KN
Frisch elasticity of labor	$1/\phi$	0.5	FW, GTT, S, BK
Annual interest rate	r	4%	KN
Annual discount factor	β	$1/(1+r)$	FW, K, GTT, S, BK
Working years per period	–	20	BK, KN
Cutoff year	–	21	BK

Table: Exogenous parameters

Quantitative Analysis

Using estimated moments in (Huggett et. al. 2011)

Symbol	Value	Target Moment	Data	Abs Perc. Deviation
ρ	0.4505	mean earnings ratio	0.868	0.0015%
ζ	0.2175	Var. log-earnings young	0.335	1%
h_1	0.4795	Var. log-earnings old	0.435	0.009%
σ	0.5573	Gini earnings young	0.3175	1.7%
λ	5.9907	mean-to-median earnings young	1.335	1.25%

Table: Calibrated Parameters

Quantitative Analysis

- Optimal reform: 4.0348% increase in consumption at all histories
- For some histories, wedges decreasing over time
- For other histories, wedges are increasing over time
- Conditional average period-2 wedge higher than period-1 wedge
- Unconditional average period-2 wedge (0.4854) higher than unconditional period-1 wedge (0.3733)
- Inverse U-shape wedges as functions of (conditional) income percentile
  - shock distribution close to Lognormal
  - very high risk aversion
  - low-end LBD factor
  - moderate skill persistence

Quantitative Analysis

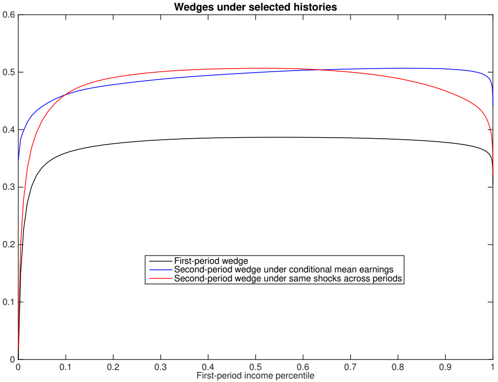

Figure: Optimal wedges for selected histories

Quantitative Analysis

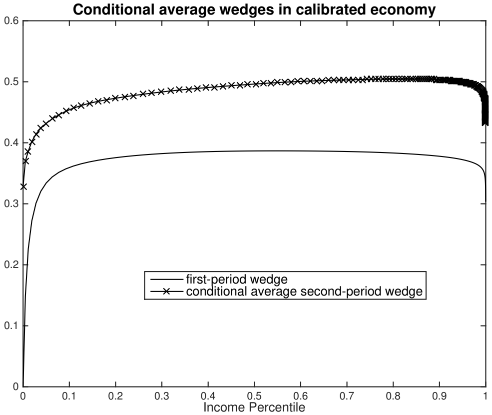

Figure: Period-1 wedges and conditional period-2 wedges as a function of period-1 income percentile.

Quantitative Analysis

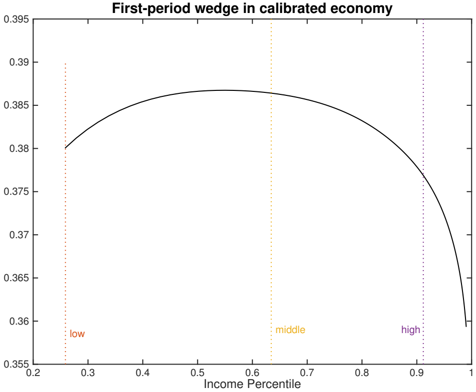

Figure: Period-1 optimal wedge. Vertical lines indicate period-1 income percentiles corresponding to low, middle, and high earnings.

Quantitative Analysis

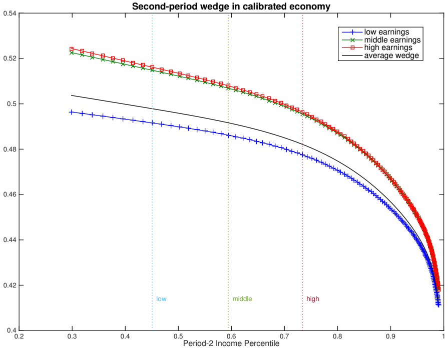

Figure: Period-2 wedges as function of period-2 earnings percentiles for (a) low, middle and high period-1 earnings, and (b) weighted average of period-1 productivities.

Taxes

- Optimal allocations implemented arbitrarily well by **age-dependent taxes invariant in past incomes**:

$$T_1(y_1) = -B + y_1 - e^{\tau_{0,1}} y^{1-\tau_1}$$

  and

$$T_2(y_2) = y_2 - e^{\tau_{0,2}} y^{1-\tau_2}$$

- Loss in consumption (relative to SB): 0.1489%
- Optimal **linear** age-dependent taxes $t_1 = 38\%$ and $t_2 = 46\%$
  - loss in consumption (relative to SB): 0.1506%
- Optimal **linear age-independent** linear tax rate: 41.25%
  - loss in consumption (relative to SB): 0.2361%

Taxes

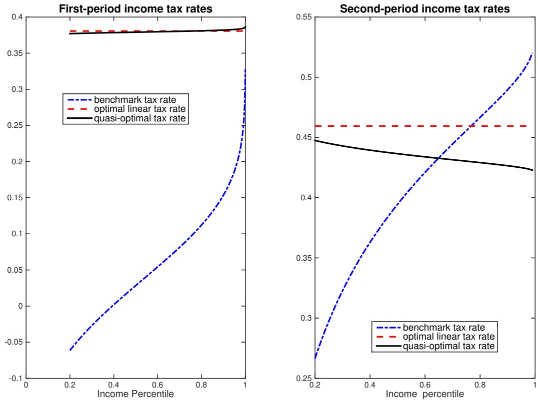

Figure: Tax rates as functions of income

Importance of LBD

- Similar calibration but with **exogenous productivity**

$$\theta_2 = h_2 \theta_1^{\widehat{\rho}} \varepsilon_2$$

- Calibrated (conditional) distributions very close to those under LBD
- Optimal allocations implemented arbitrarily well by age-dependent taxes invariant in past incomes
- Ignoring LBD: 15% overestimation of benefits of reforming US tax code

Importance of LBD

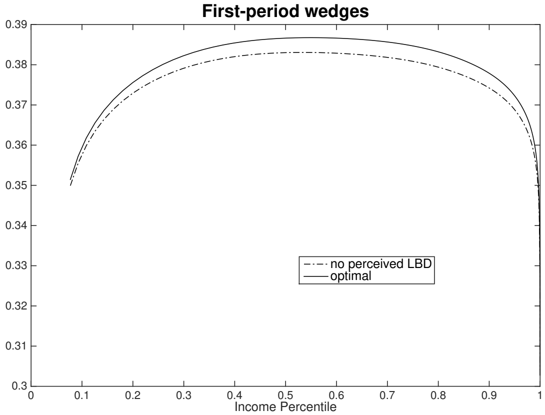

Figure: First-period wedges with and without LBD

Importance of LBD

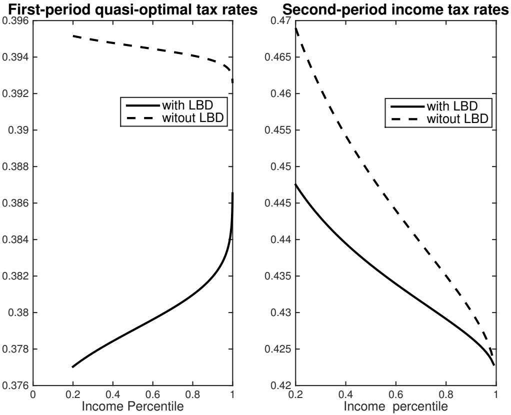

Figure: Quasi-optimal income tax rates with and without LBD

Conclusions

- LBD: important qualitative and quantitative implications
  - level
  - progressivity
  - dynamics
  - benefits of reforming US tax code
- Ongoing work
  - arbitrary horizons (recursive approach)
  - general wedge decomposition
- Future work:
  - hidden savings
  - political economy constraints
  - partial commitment
  - . . .

THANKS!

Second Best: Incentive Compatibility

- Continuation utility (history $\theta = (\theta_1, \theta_2)$):

$$V_2(\theta) \equiv c_2(\theta) - \psi(y_2(\theta), \theta_2)$$

- For any θ_1, IC-2 requires that
  - $V_2(\theta_1, \cdot)$ Lipschitz continuous and s.t. (e.g., Mirrlees)

$$V_2(\theta_1, \theta_2) = V_2(\theta_1, \underline{\theta}_2) - \int_{\underline{\theta}_2}^{\theta_2} \psi_\theta(y_2(\theta_1, s), s)ds,$$

  - $y_2(\theta_1, \cdot)$ nondecreasing

Second Best: Incentive Compatibility

- IC-1 requires that

$$V_1(\theta_1) = V_1(\underline{\theta}_1)$$

$$-\int_{\underline{\theta}_1}^{\theta_1}\left\{\psi_\theta(y_1(s),s)ds + \delta\mathbb{E}^{\lambda[\chi]|s}\left[I_1^2(\tilde{\theta},y_1(s))\psi_\theta(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]\right\}ds$$

and

$$\int_{\hat{\theta}_1}^{\theta_1}\left\{\psi_\theta(y_1(s),s) + \delta\mathbb{E}^{\lambda[\chi]|s,y_1(s)}\left[I_1^2(\tilde{\theta},y_1(s))\psi_\theta(y_2(s,\tilde{\theta}_2),\tilde{\theta}_2)\right]\right\}ds$$

$$\leq \int_{\hat{\theta}_1}^{\theta_1}\left\{\psi_\theta(y_1(\hat{\theta}_1),s) + \delta\mathbb{E}^{\lambda[\chi]|s,y_1(\hat{\theta}_1)}\left[I_1^2(\tilde{\theta},y_1(\hat{\theta}_1))\psi_\theta(y_2(\hat{\theta}_1,\tilde{\theta}_2),\tilde{\theta}_2)\right]\right\}ds$$

Sufficient Statistics

- Let

$$\hat{E}_1(y_1) \equiv \frac{1-\tau_1(y_1)}{y_1} \frac{\partial \widehat{y}_1(1-\tau_1(y_1), \theta_1(y_1))}{\partial(1-\tau_1)}$$

$$e_{\mathscr{T}_2|y_1} \equiv \frac{\partial \mathbb{E}[\tilde{\mathscr{T}}_2|y_1]}{\partial y_1} \frac{y_1}{\mathbb{E}[\tilde{\mathscr{T}}_2|y_1]}$$

Proposition

Under the optimal tax code

$$\frac{\tau_1(y_1)}{1-\tau_1(y_1)} = \frac{1-H_Y(y_1)}{y_1 \hat{h}_Y(y_1)} \frac{1}{\hat{E}_1(y_1)} \left[\frac{1}{1+\delta e_{\mathscr{T}_2|y_1} \frac{\mathbb{E}[\tilde{\mathscr{T}}_2|y_1]}{\tau_1(y_1) y_1}} \right].$$

Sufficient Statistics

- Let

$$\hat{E}_2(y_1, y_2) \equiv \frac{1-\tau_2(y_1,y_2)}{y_2} \frac{\partial \widehat{y}_2(1-\tau_2(y_1,y_2), \theta_2(y_1,y_2))}{\partial(1-\tau_2)}$$

Proposition

Under optimal tax code

$$\frac{\tau_2(y_1,y_2)}{1-\tau_2(y_1,y_2)} = \left[\frac{\partial \widetilde{H}_O(y_2|y_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1}\right] \frac{1-H_Y(y_1)}{h_Y(y_1) y_2 \hat{h}_O(y_2|y_1)} \frac{1}{\hat{E}_2(y_1,y_2)}$$

- Results established with novel perturbations (reforms) accounting for endogeneity of period-2 productivity