Taxation under Learning-by-doing

Miltos Makris  Alessandro Pavan

Kent  Northwestern

November 2019
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)
  - first 2 years of employment, wages grow, on average, by 8.5% in 1st year and 7.5% in 2nd

- Thompson (2012), Levitt et. al. (2013)
  - reduction in unit costs from production, particularly strong in early years, “bounded learning”
This Paper

- Dynamic Mirrleesian economy in which agents’ productivity
  - own private information
  - **stochastic**
  - evolves **endogenously** over lifecycle (due to LBD)

- Novel effects on (labor) wedges

- Quantitatively significant impact on optimal tax codes
  - level
  - progressivity
  - dynamics
Related Literature

- **Optimal Labor Income Taxation:** Mirrlees (1971), Diamond (1998), Saez (2001)...
  - static & exogenous productivity

- **New Dynamic Public Finance:** Albanesi and Sleet (2006), Golosov, Tsyvinski and Werning (2006), Kocherlakota (2005, 2010), Kapicka (2013), Farhi and Werning (2013), Golosov, Tsyvinski and Troschkin (2016) ...
  - dynamic & exogenous productivity

  - future productivity is private information, stochastic and side-product of labor
Summary of Main Results

- LBD leads to higher distortions (wedges)

- SB allocations can be (approximately) implemented by simple age-dependent taxes, invariant in past incomes

- Higher and less progressive tax rates than under current US tax code

- ... but lower and more progressive than without LBD
Road Map

- Qualitative Analysis
  - Model
  - Labor distortions (wedges)

- Quantitative Analysis
  - Optimal reform of calibrated economy
  - Approximate implementation
  - Role of stochasticity
  - Counterfactual analysis: role of LBD on proposed reforms

- Conclusions
Qualitative Analysis
Environment

- Two working periods/blocks $t = 1, 2$: “young” and “old”
- Linear labor production
- Period-1 productivity: $\theta_1$
  - privately observed at beginning of $t = 1$
  - drawn from cdf $F_1$ (density $f_1$)
- Period-2 productivity: $\theta_2$
  - privately observed at beginning of $t = 2$
  - drawn from cdf $F_2(\cdot|\theta_1, y_1)$ (FOSD)
  - dependence on $y_1$: LBD

Example: $\theta_2 = \theta_1^{\xi} l_1^{\zeta} \varepsilon_2 = \theta_1^{\xi} \left( \frac{y_1}{\theta_1} \right)^{\zeta} \varepsilon_2 = \theta_1^{\rho} y_1^{\zeta} \varepsilon_2$
Environment

- Period-\(t\) flow utility:
  \[ v(c_t) - \psi(y_t, \theta_t) \]

- e.g. \(\psi(y_t, \theta_t) = \frac{1}{1+\phi} \left( \frac{y_t}{\theta_t} \right)^{1+\phi} \) where \(1/\phi\) is Frisch elasticity

- Discount factor (for both workers and planner): \(\delta\)
Setting up the problem

- Let $\theta^2 \equiv (\theta_1, \theta_2)$ and $\theta^1 \equiv (\theta_1)$

- Worker expected life-time utility:

$$V_1(\theta_1) = \mathbb{E} \left[ \sum_t \delta^{t-1} \left( \nu(c_t(\tilde{\theta}^t)) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) \right) \bigg| \theta_1, y_1(\theta_1) \right]$$

- Worker expected life-time tax bill:

$$R_1(\theta_1) = \mathbb{E} \left[ \sum_t \delta^{t-1} \left( y_t(\tilde{\theta}^t) - c_t(\tilde{\theta}^t) \right) \bigg| \theta_1, y_1(\theta_1) \right]$$
Setting up the dual Utilitarian problem

- **Dual**: planner maximizes tax revenues

\[
\int R_1(\theta_1) dF_1(\theta_1)
\]

s.t. participation/redistribution constraint

\[
\int V_1(\theta_1) dF_1(\theta_1) \geq \kappa
\]

and incentive-compatibility constraints
First Best: period-1 output

For any $\theta_1$

$$\frac{\psi(y_1(\theta_1), \theta_1)}{\nu'(c_1(\theta_1))} = 1 + LD_1(\theta_1)$$

where

$$LD_1(\theta_1) \equiv \delta \frac{\partial}{\partial y_1} \mathbb{E} \left[ y_2(\tilde{\theta}) - c_2(\tilde{\theta}) + \frac{\nu(c_2(\tilde{\theta})) - \psi(y_2(\tilde{\theta}), \tilde{\theta})}{\nu'(c_2(\tilde{\theta}))} \bigg| \theta_1, y_1(\theta_1) \right]$$

- output driven by marginal production cost expressed in terms of tax revenues (consumption)
- output driven also by LBD impact on future tax revenues, and workers continuation utility
- LBD effect via change in conditional distribution

$\Rightarrow$ Higher period-1 output under LBD, for any given $\theta_1$ (due to FOSD and increasing period-2 net surplus)
Second Best

- When productivity is workers’ private information, FB not incentive compatible

- Higher productivity workers would mimic lower types to take advantage of cost differentials and skill persistence

- Need to give high types “rents”: higher consumption (lower taxes) than under FB

- Value of distorting output: smaller rents to highly productive workers

- Under LBD: extra value in distorting period-1 output: smaller expected rents thanks to shift in period-2 distribution
Labor Wedges

Definition

Period-1 Labor wedge:

\[ W_1(\theta_1) \equiv 1 + LD_1(\theta_1) - \frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))}. \]

Relative wedge:

\[ \hat{W}_1 \equiv W_1/ \frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))} \]
Relative Wedges (FOA)

- Period-1 wedges under SB allocations:

\[ \hat{W}_1(\theta_1) = [RA_1(\theta_1) - D_1(\theta_1)] [\hat{W}_1^{RRN}(\theta_1) + \Omega_1(\theta_1)] \]

where

- \( \hat{W}_1^{RRN} \): wedge under Rawlsian objective, RN agents, no LBD
- \( \Omega_1 \): LBD effect
- \( RA_1 \): correction due to higher costs of non-transferable utility
- \( D_1 \): correction due to higher Pareto weights given to types above \( \theta_1 \)
LBD Effect

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

- "handicap" \( h_2(\theta, y) \equiv -\frac{1-F_1(\theta_1)}{\theta_1 f_1(\theta_1)} \rho_2 \psi_\theta(y_2(\theta_2), \theta_2) \): cost of rents associated with compensation to type \((\theta_1, \theta_2)\)

- LBD contributes to higher expected period-2 handicaps
  \[ \Rightarrow \text{extra benefit of lowering } y_1(\theta_1) \Rightarrow \text{higher wedges in early years} \]

- \( \Omega_1(\theta_1) \) increasing in \( \theta_1 \), if \( \theta_1 \) and \( y_1 \) strong complements and \( \frac{1-F_1(\theta_1)}{\theta_1 f_1(\theta_1)} / \psi_y(y_1(\theta_1), \theta_1) \) not very decreasing
  \[ \Rightarrow \text{benefit of distorting } y_1 \text{ downwards stronger for higher } \theta_1 \]
  \[ \Rightarrow \text{more progressivity} \]
Quantitative Analysis
Calibrated Economy

- \( T = 40 \)
- \( v(c) = \log(c) \)
- \( \psi(y_t, \theta_t) = \frac{1}{1+\phi} \left( \frac{y_t}{\theta_t} \right)^{1+\phi} \) with \( \phi = 2 \) (Frisch elasticity = 0.5)
- \( r = 1 - \frac{1}{\beta} = 4\% \) with \( \delta = \beta^{20} \)
- \( \theta_1 = h_1 \varepsilon_1 \)
- \( \theta_2 = \theta_1^\rho y_1^\zeta \varepsilon_2 \)
- \( \varepsilon_t \) iid Pareto-Lognormal \( (\lambda, \sigma) \) with mean 1
- U.S. income tax estimation in Heathcote et. al. (2017)
  \[
  T(y) = y - e^{\tau_0} y^{1-0.181}
  \]
Using estimated moments in Huggett et. al. (2011)

<table>
<thead>
<tr>
<th>Param</th>
<th>Value</th>
<th>Target Moment</th>
<th>Data</th>
<th>Abs % Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρ</td>
<td>0.4505</td>
<td>mean earn’s ratio</td>
<td>0.868</td>
<td>0.0015%</td>
</tr>
<tr>
<td>ζ</td>
<td>0.2175</td>
<td>Var. log-earn’s young</td>
<td>0.335</td>
<td>1%</td>
</tr>
<tr>
<td>h₁</td>
<td>0.4795</td>
<td>Var. log-earn’s old</td>
<td>0.435</td>
<td>0.009%</td>
</tr>
<tr>
<td>σ</td>
<td>0.5573</td>
<td>Gini earn’s young</td>
<td>0.3175</td>
<td>1.7%</td>
</tr>
<tr>
<td>λ</td>
<td>5.9907</td>
<td>mean/median earn’s young</td>
<td>1.335</td>
<td>1.25%</td>
</tr>
</tbody>
</table>

Table: Calibrated Parameters
Second Best: Quantitative Analysis

- Optimal reform: **4.0409%** increase in consumption at all histories

- Inverse U-shape wedges as functions of (conditional) income percentile
  - low-end LBD factor
  - moderate skill persistence
  - shock distribution close to Lognormal

- Increasing (conditional average) wedges over time
  - high stochasticity **and** risk aversion / low-end LBD factor
Approximate Implementation

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

\[ T_1(y_1) = -B + y_1 - e^{τ_{0.1}} y^{1-τ_1} \]

and

\[ T_2(y_2) = y_2 - e^{τ_{0.2}} y^{1-τ_2} \]

- Loss in consumption (relative to SB): 0.155%

- Optimal **linear** age-dependent taxes \( τ_1 = 38\% \) and \( τ_2 = 46\% \)

- Loss in consumption (relative to SB): 0.1567\%
Reform: Revenue-neutral Tax Rates

Figure: Tax rates as functions of income percentile
Comparative Statics: Stochasticity

Figure: Variations in stochasticity
Counterfactual: Importance of LBD

- Similar calibration but with **exogenous productivity**
  \[ \theta_2 = h_2 \theta_1 \hat{\rho} \varepsilon_2 \]

- Calibrated (conditional) distributions very close to those under LBD

- Higher persistence: \( \hat{\rho} = 0.6 \) (with LBD, \( \rho = 0.4 \))

- Ignoring LBD: 15% overestimation of benefits of reforming US tax code

- SB allocations: implemented arbitrarily well by age-dependent taxes invariant in past incomes
  - but with **higher period-1 rates**
Importance of LBD

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

- LBD: important qualitative and quantitative implications for
  - level
  - progressivity
  - dynamics
  - benefits
  in reforming US tax code

- Future work:
  - sector-specific LBD heterogeneity
  - hidden savings
  - political economy constraints
  - spillovers
  - partial commitment
  - …
THANKS!
Motivation

Qualitative Analysis

Quantitative Analysis

Conclusions

Appx: T

V & R

IC

RA term

SB

Handicaps

Multi-period Environment

- $T$: length of working life in years (even number)
- Linear labour production
- Labour productivity in period $\tau$: $\vartheta_\tau$
- $\vartheta_\tau = \theta_1$ for $\tau = 1, ..., T/2$
- $\theta_1$ privately observed by worker at beginning of “young age”
- $F_1$: cdf of initial distribution (density $f_1$)
- $\vartheta_\tau = \theta_2$ for $\tau = T/2 + 1, ..., T$
- $\theta_2$ privately observed by worker at beginning of “old age”
- $F_2(\cdot | \theta_1, \bar{y})$: cdf of $\theta_2$ - satisfies FOSD
- LBD: dependence on weighed average of output as young $\bar{y}$

Example: $\theta_2 = \theta_1^{\rho+\zeta} \bar{y}^\zeta \varepsilon_2 = \theta_1^{\rho+\zeta} \left( \frac{\bar{y}}{\theta_1} \right)^\zeta \varepsilon_2 = \theta_1^{\rho} \bar{y}^\zeta \varepsilon_2$
Environment

- Period-$\tau$ worker’s payoff:
  \[ v(c_\tau) - \psi(y_\tau, \vartheta_\tau) \]

- Discount factor $\beta$

- Planner maximizes average expected life-time utility s.t. exogenous expected tax revenue requirements

- $\beta = 1/(1 + r)$
Environment

- LBD active in each of first $T/2$ years with **declining** weights

\[ \hat{\beta}_\tau / \sum_{\tau=1}^{T/2} \hat{\beta}_\tau \]

- When \((\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_{T/2})\) is proportional to \((1, \beta, ..., \beta^{T/2-1})\), consumption and earnings constant over each block of $T/2$ years

- Isomorphic to 2-period model
  - discount factor $\delta = \beta^{T/2}$
  - period-$t$ history: $\theta^t$ with $\theta^1 = \theta_1$, $\theta^2 = (\theta_1, \theta_2)$
  - allocation: \((y_t(\theta^t), c_t(\theta^t))_{t=1,2}\)

- LBD:

\[ \bar{y} = y_1(\theta_1) \]
Worker expected life-time utility given allocation:

$$V_1(\theta_1) = v(c_1(\theta_1)) - \psi(y_1(\theta_1), \theta_1) +$$

$$\delta \int \left( v(c_t(\theta_1, \tilde{\theta}_2)) - \psi(y_t(\theta_1, \tilde{\theta}_2), \tilde{\theta}_2) \right) F_2(\tilde{\theta}_2|\theta_1, y_1(\theta_1))$$

Expected life-time tax bill:

$$R(\theta_1) = y_1(\theta_1) - c_1(\theta_1) +$$

$$\delta \int \left( y_t(\theta_1, \tilde{\theta}_2) - c_t(\theta_1, \tilde{\theta}_2) \right) F_2(\tilde{\theta}_2|\theta_1, y_1(\theta_1))$$
Incentive Compatibility

\[ v(c_1(\theta_1)) - \psi(y_1(\theta_1), \theta_1) + \delta \int (v(c_2(\theta_1, \theta_2)) - \psi(y_2(\theta_1, \theta_2), \theta_2)) \, dF_2(\theta_2|\theta_1, y_1(\theta_1)) \geq \]

\[ v(c_1(\hat{\theta}_1)) - \psi(y_1(\hat{\theta}_1), \theta_1) + \]

\[ \delta \int \left( v(c_2(\hat{\theta}_1, \hat{\theta}_2(\theta_2))) - \psi(y_2(\hat{\theta}_1, \hat{\theta}_2(\theta_2)), \theta_2) \right) \, dF_2(\theta_2|\theta_1, y_1(\hat{\theta}_1)) \]

and

\[ v(c_2(\theta_1, \theta_2)) - \psi(y_2(\theta_1, \theta_2), \theta_2) \geq v(c_2(\theta_1, \hat{\theta}_2)) - \psi(y_2(\theta_1, \hat{\theta}_2), \theta_2) \]
Second Best: Impulse Responses

Impulse Responses

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

$$I^2_1(\theta, y_1) = \frac{\partial Z_2(\theta_1, y_1, \varepsilon_2)}{\partial \theta_1} = \rho \frac{\theta_2}{\theta_1}$$

where $\theta = (\theta_1, \theta_2)$ and $\varepsilon_2 = Z_2^{-1}(\theta_2; \theta_1, y_1) = \frac{\theta_2}{\theta_1^\rho y_1^\zeta}$

and

$$I^1_1(\theta, y_1) = 1$$
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 $\theta_1$, IC-2 requires that:
  - $V_2(\theta_1, \cdot)$ Lipschitz continuous and s.t. (e.g., Mirrlees)
    \[ V_2(\theta_1, \theta_2) = V_2(\theta_1, \underline{\theta}_2) - \int_{\underline{\theta}_2}^{\theta_2} \psi(y_2(\theta_1, s), s)ds, \]
  - $y_2(\theta_1, \cdot)$ nondecreasing
Second Best: Incentive Compatibility

IC-1 requires that

\[ V_1(\theta_1) = V_1(\bar{\theta}_1) \]

\[ -\int_{\bar{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s) ds + \delta \mathbb{E} \left[ l_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(\tilde{\theta}, \bar{\theta}_2)|s, y_1(s)) \right] \right\} ds \]

and

\[ \int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s) + \delta \mathbb{E} \left[ l_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(s, \bar{\theta}_2), \bar{\theta}_2)|s, y_1(s)) \right] \right\} ds \leq \]

\[ \int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(\hat{\theta}_1), s) + \delta \mathbb{E} \left[ l_1^2(\tilde{\theta}, y_1(\hat{\theta}_1)) \psi_\theta(y_2(\hat{\theta}_1, \bar{\theta}_2), \bar{\theta}_2)|s, y_1(\hat{\theta}_1)) \right] \right\} ds \]
Risk Aversion Effect

- Increasing lifetime utility by $v'(c_1(\theta_1))\Omega_1(\theta_1)$ increases rents to all higher types.

- One util compensation requires $1/v'(c_t)$ units of consumption.

- Risk aversion increases cost of increasing expected future information rents: $RA_1(\theta_1) > 1$.

- Risk aversion contributes to \textit{amplification} of LBD \textit{level} effect.
  - risk aversion increases benefit of shifting future distribution towards lower types.

- BUT, risk aversion leads also to an \textit{alleviation} of LBD effects.
  - higher cost of future rents $\rightarrow$ lower future rents $\rightarrow$ lower need to shift future distribution.
Redistribution Effect

- Increasing lifetime utility by \( v'(c_1(\theta_1))\Omega_1(\theta_1) \) is valued by an Utilitarian planner

- One social util costs \( \int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s) \) in revenue terms

- This effect counteracts amplification effect of risk aversion
Recursive Problem

- “Promised continuation payoff”:

\[ \Pi_{t+1}(\theta^t) \equiv \int V_{t+1}(\theta^t, \theta_{t+1}) dF_{t+1}(\theta_{t+1} \mid \theta_t, y_t(\theta^t)) \]

- “Marginal promise”:

\[ Z_{t+1}(\theta^t) \equiv -E^\lambda[x] \theta^t \left[ \sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} l_{t}^{\tau}(\theta^\tau, y^{\tau-1}) \psi_{\theta}(y^\tau, \theta^\tau) \right] \]
Towards Recursive Problem

With these definitions we have:

\[ V_t(\theta^t) = v(c_t(\theta^t)) - \psi(y_t(\theta^t), \theta_t) + \delta \Pi_{t+1}(\theta^t) \]

and

\[ \frac{\partial V_t(\theta^t)}{\partial \theta_t} = -\psi_\theta(y_t(\theta^t), \theta_t) + \delta Z_{t+1}(\theta^t) \]
The Recursive Problem

\[ Q_t(\theta^{t-1}, y_{t-1}(\theta^{t-1}), \Pi_t(\theta^{t-1}), Z_t(\theta^{t-1})) \equiv \max_{y_t(\cdot), V_t(\cdot), \Pi_{t+1}(\cdot), Z_{t+1}(\cdot)} \]

\[ y_t(\theta^t) - v^{-1}(V_t(\theta^t) + \psi(y_t(\theta^t), \theta_t) - \delta \Pi_{t+1}(\theta^t)) + \]

\[ \delta \mathbb{E} \left[ Q_{t+1}(\theta^t, y_t(\theta^t), \Pi_{t+1}(\theta^t), Z_{t+1}(\theta^t)) \mid \theta^t, y_t(\theta^t) \right] \]

subject to

\[ \frac{\partial V_t(\theta^t)}{\partial \theta_t} = -\psi(\theta^t, y_t(\theta^t), \theta_t) + \delta Z_{t+1}(\theta^t) \]

\[ \Pi_t(\theta^{t-1}) = \int V_t(\theta_t) dF(\theta_t \mid \theta_{t-1}, y_{t-1}(\theta^{t-1})) \]

and for \( t > 1 \)

\[ Z_t(\theta^{t-1}) = \int \left[ -\psi(y_t(\theta^t), \theta_t) + \delta Z_{t+1}(\theta^t) \right] \times \]

\[ \times l_{t-1}^t(\theta^t, y_{t-1}(\theta^{t-1})) dF_t(\theta_t \mid \theta_{t-1}, y_{t-1}(\theta^{t-1})) \]
Given $\theta = (\theta_1, \theta_2)$,

$$1 = \psi_y(y_2(\theta), \theta_2) - \frac{f_1(\theta_1)}{1 - F_1(\theta_1)} l_1^2(\theta, y_1(\theta_1)) \psi_{y\theta}(y_2(\theta), \theta_2)$$
Second Best under RN and Rawls: period-1 output

Given $\theta_1$,

$$1 + LD_1(\theta_1)$$

$$= \psi_y(y_1(\theta_1), \theta_1) - \frac{f_1(\theta_1)}{1 - F_1(\theta_1)} \psi_{y\theta}(y_1(\theta_1), \theta_1)$$

$$+ \delta \frac{\partial}{\partial y_1} \mathbb{E} \left[ \frac{f_1(\theta_1)}{1 - F_1(\theta_1)} l_1^2(\tilde{\theta}, y_1(\theta_1)) \psi(\theta_2(\tilde{\theta}), \tilde{\theta}_2) | \theta_1, y_1(\theta_1) \right]$$
Second Best: Handicaps

- Expected tax revenues **under risk neutrality** (using IC):

\[
\mathbb{E} \left[ \sum_t \delta^{t-1} \left( y_t(\tilde{\theta}^t) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) - h_t(\tilde{\theta}^t, y^t(\tilde{\theta}^t)) \right) \right] - V_1(\theta_1),
\]

- First-period "handicap":

\[
h_1(\theta_1, y_1) \equiv - \frac{f_1(\theta_1)}{1-F_1(\theta_1)} \psi_\theta(y_1, \theta_1)
\]

- Second-period "handicap":

\[
h_2(\theta, y) \equiv - \frac{f_1(\theta_1)}{1-F_1(\theta_1)} l_1^2(\theta, y_1) \psi_\theta(y_2, \theta_2)
\]

- Handicaps: costs to planner from asymmetric information