White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/114606/

Version: Accepted Version

Article:
Krause, Alan (2017) On Redistributive Taxation under the Threat of High-Skill Emigration. Social Choice and Welfare. 845–856. ISSN: 0176-1714

https://doi.org/10.1007/s00355-017-1038-8

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

On Redistributive Taxation under the Threat of High-Skill Emigration

Alan Krause*
University of York

2 February 2017

Abstract

The increasing international mobility of high-skill individuals is often seen as posing a threat to domestic social welfare, by limiting the ability of governments to tax these individuals and redistribute to the poor. In this paper, we examine a simple dynamic nonlinear income tax model without commitment. In this setting, it is shown that the threat of emigration by high-skill individuals facilitates redistribution and increases social welfare in the short-run, and has no effect on social welfare over the long-run.

Keywords: nonlinear taxation; migration; commitment.

JEL Classifications: H21; H24; F22.

*Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, U.K. E-mail: alan.krause@york.ac.uk. I thank an associate editor and two anonymous referees for comments and suggestions which led to a much improved paper. Any errors remain my responsibility.

1 Introduction

Optimal tax analyses typically assume that individuals cannot emigrate to avoid domestic taxation. Such an assumption is, however, increasingly viewed as being unrealistic, especially as it relates to high-skill individuals. When high-skill individuals are immobile, redistributive taxation must take into consideration that these individuals may change their labour supply along the intensive margin. That is, high-skill individuals may work less, thus reducing the amount of income available for redistribution. When high-skill individuals are internationally mobile, they have the additional option of changing their labour supply along the extensive margin, i.e., they may emigrate. An often-employed method to capture the threat of emigration is to introduce type-dependent participation constraints into the optimal tax problem; see, e.g., Osmundsen (1999), Krause (2009a), Simula and Trannoy (2010, 2012), and Lehmann et al. (2014). Naturally, these additional constraints reduce the level of social welfare attainable.

In this paper, we introduce the threat of emigration by high-skill individuals into a two-period nonlinear income tax model without commitment. As in the related literature, this threat is captured by introducing a participation constraint for high-skill individuals. However, we show that the introduction of the high-skill type's participation constraint facilitates redistribution and increases social welfare in the short-run, and has no effect on social welfare over the long-run. One may wonder how it is possible that the introduction of a binding constraint into an optimisation problem does not reduce the maximised value of the objective function. The participation constraint appears in period 2 and it does reduce the maximised value of the objective function (social welfare) in that period, but the specific manner in which social welfare is reduced — by limiting redistribution — enables a higher level of social welfare to be obtained in period 1. Thus social welfare summed over both periods remains unchanged. The intuition can be summarised as follows. In period 1 the government does not know each individual's skill type, and therefore implements standard (incentive-compatible) nonlinear income taxation which induces individuals to reveal their type. However, high-skill individuals know that if they reveal their type in period 1, they lose their information advantage and

will be subjected to first-best taxation in period 2. This means that high-skill individuals must be offered a very attractive tax treatment in period 1 to reveal their type, to compensate them for having to face first-best taxation in period 2. Accordingly, the ability of the government to redistribute in period 1 is severely limited. However, the threat of emigration reduces the extent of redistribution possible in period 2, meaning that high-skill individuals require less compensation in period 1 to reveal their type. This enables the government to implement more redistribution in period 1, which correspondingly increases first-period social welfare. The threat of emigration limits redistribution and reduces social welfare in period 2, but optimal taxation balances the short-run benefits against the long-run costs. Thus the threat of emigration has no effect on the level of social welfare summed over both periods.

In terms of previous results, the paper most closely related to ours is that by Leite-Monteiro (1997). He also finds that increased international mobility may enhance redistribution, but his model is entirely different to ours. In Leite-Monterio's model, the government can always implement first-best personalised lump-sum taxes, and changes in the country's skill composition *after* migration takes place is what makes enhanced redistribution a possibility. By contrast, in our model the participation constraint ensures that no one migrates, so the skill composition remains unchanged. Our paper is also related to the literature on dynamic Mirrlees (1971) nonlinear income taxation without commitment, e.g., Roberts (1984), Apps and Rees (2006), Brett and Weymark (2008a), Krause (2009b), Berliant and Ledyard (2014), and Guo and Krause (2011, 2013, 2014, 2015a, 2015b). In dynamic Mirrlees models, the question arises as to whether the government can or cannot commit to not using skill-type information revealed by individuals in earlier periods when it implements taxation in latter periods. A common theme in the literature is the highlighting of how different and counter-intuitive optimal policy can be when the commitment assumption is relaxed. The present paper provides another example of a result that, at first glance, appears quite counter-intuitive.

The remainder of the paper is organised as follows. Section 2 outlines the model and the structure of optimal taxation, while Section 3 presents and discusses our result. Section 4 discusses other possible solutions to the optimal tax problem, and what these

would imply for our result. Section 5 concludes, and the proof of our result is contained in an appendix.

2 A Simple Model

We consider a two-period model with a unit measure of individuals and with the following timing. In period 1 the government knows that there are $\phi \in (0,1)$ high-skill individuals and $(1-\phi)$ low-skill individuals in the economy, but it does not know any individual's skill type. The government therefore implements standard second-best (incentive-compatible) nonlinear income taxation, under which each individual is willing to reveal their type. Then, in period 2, the government knows each individual's skill type, and is tempted to use this information to implement first-best redistributive taxation. However, high-skill individuals have the option of emigrating, so redistribution is limited by the high-skill type's participation constraint. We assume that emigration is only possible in period 2. The implications of this assumption are discussed in Section 4, though it is consistent with the observation that individuals tend to be immobile in the short-run (period 1), but mobile in the long-run (period 2). For simplicity we assume that individuals do not save or borrow, so the only link between the periods is the revelation and use of skill-type information.

Specifically, the government in period 2 solves the following problem. Choose tax treatments $\langle c_L^2, y_L^2 \rangle$ and $\langle c_H^2, y_H^2 \rangle$ for the low-skill and high-skill individuals, respectively, to maximise:

$$(1-\phi)\left\{u(c_L^2) - \frac{y_L^2}{w_L}\right\} + \phi\left\{u(c_H^2) - \frac{y_H^2}{w_H}\right\} \tag{2.1}$$

subject to:

$$(1-\phi)\left[y_L^2 - c_L^2\right] + \phi\left[y_H^2 - c_H^2\right] \geq 0 \tag{2.2}$$

$$u(c_H^2) - \frac{y_H^2}{w_H} \geq V_H \tag{2.3}$$

where c_i^t is type i's consumption (or post-tax income) in period t, $y_i^t = w_i l_i^t$ is type i's pre-tax income in period t, with w_i denoting type i's wage rate and l_i^t denoting type i's labour supply in period t. It is assumed that $w_H > w_L > 0$ and that wages remain

constant over time. Equation (2.1) is a utilitarian social welfare function, where $u(\cdot)$ is increasing and strictly concave and the individuals' utility function is quasi-linear in labour.[1] Equation (2.2) is the government's budget constraint,[2] and equation (2.3) is the high-skill type's participation constraint. High-skill individuals can obtain a utility level of V_H by emigrating, which is their reservation utility. We assume that V_H is sufficiently high such that equation (2.3) is binding, i.e., the threat of emigration is effective. The solution to programme (2.1) – (2.3) yields the value function $W^2(\phi, w_L, w_H, V_H)$, which is the level of social welfare attainable in period 2.

In period 1 the government does not know each individual's skill type, and therefore implements second-best (incentive-compatible) nonlinear income taxation. It chooses tax treatments $\langle c_L^1, y_L^1 \rangle$ and $\langle c_H^1, y_H^1 \rangle$ for the low-skill and high-skill individuals, respectively, to maximise:

$$(1-\phi)\left\{u(c_L^1) - \frac{y_L^1}{w_L}\right\} + \phi\left\{u(c_H^1) - \frac{y_H^1}{w_H}\right\} \tag{2.4}$$

subject to:

$$(1-\phi)\left[y_L^1 - c_L^1\right] + \phi\left[y_H^1 - c_H^1\right] \geq 0 \tag{2.5}$$

$$u(c_H^1) - \frac{y_H^1}{w_H} + \delta V_H \geq u(c_L^1) - \frac{y_L^1}{w_H} + \delta \max\left[\widehat{U}_H^2, V_H\right] \tag{2.6}$$

where $\widehat{U}_H^2 = u(c_L^2) - \frac{y_L^2}{w_H}$ is the utility that a high-skill individual would obtain in period 2 from the low-skill type's tax treatment. That is, $c_L^2 = c_L^2(\phi, w_L, w_H, V_H)$ and $y_L^2 = y_L^2(\phi, w_L, w_H, V_H)$ come from the solution to programme (2.1) – (2.3). We use $\delta \in (0,1)$ to denote the discount factor.

Equation (2.4) is the first-period utilitarian social welfare function. The government in period 1 might care about the level of social welfare in period 2, but its choice of $\langle c_L^1, y_L^1 \rangle$ and $\langle c_H^1, y_H^1 \rangle$ cannot affect the level of social welfare attainable in period 2,

[1] The literature on the comparative statics of optimal nonlinear income taxes (e.g., Weymark (1987), Brett and Weymark (2008b, 2011), and Simula (2010)) has shown that results are generally obtainable only when the utility function is quasi-linear. As we make use of comparative statics methods, we also assume that the utility function is quasi-linear.

[2] We assume that the government cannot save or borrow, and that its revenue requirement is zero. Thus taxation is implemented only for redistributive purposes.

which is $W^2(\phi, w_L, w_H, V_H)$. Therefore, the government in period 1 acts as if maximising only first-period social welfare. Equation (2.5) is the government's first-period budget constraint, while equation (2.6) is the high-skill type's incentive-compatibility constraint.

In order for a high-skill individual to be willing to choose $\langle c_H^1, y_H^1 \rangle$ in period 1 and thus reveal their type, the utility they obtain from this choice in period 1 plus the utility they obtain in period 2 (which will be their reservation utility, V_H), must be greater than or equal to the utility they could obtain by mimicking. A high-skill individual has two mimicking strategies. First, a high-skill individual could choose the low-skill type's tax treatment in period 1 — effectively announcing to the government that they are low-skill — and then not emigrate in period 2. In this case, because in period 1 they announced that they are low-skill, the government in period 2 will treat them as low-skill and give them the low-skill type's tax treatment. The mimicking high-skill individual would therefore obtain utility level $\widehat{U}_H^2$ in period 2. Second, a high-skill individual could choose the low-skill type's tax treatment in period 1 and then emigrate in period 2, to avoid being treated as low-skill in period 2. In this case, they obtain utility level V_H in period 2. Which of these mimicking strategies is best depends upon $\widehat{U}_H^2$ and V_H. The first is better when $\widehat{U}_H^2 > V_H$, and the second is better when $\widehat{U}_H^2 < V_H$. Note that if $\widehat{U}_H^2 \leq V_H$, the incentive-compatibility constraint (2.6) reduces to $u(c_H^1) - \frac{y_H^1}{w_H} \geq u(c_L^1) - \frac{y_L^1}{w_H}$, which is the same as the condition required for incentive compatibility in static models. However, as we explain below (in Subsection 3.1), $\widehat{U}_H^2 > V_H$ will hold unless V_H is so high as to undo a key feature of dynamic nonlinear taxation without commitment.

We make the standard assumption that any individual who is indifferent between truthfully revealing their type and mimicking will tell the truth. Therefore, no mimicking occurs in the solution to the optimal tax problem. We also omit the low-skill type's incentive-compatibility constraint, because we make the common assumption that the redistributive goals of the government create an incentive for high-skill individuals to mimic low-skill individuals, but not vice versa.[3] As discussed earlier, emigration is only

[3]In fact, high-skill individuals generally have a stronger incentive to mimic when the government cannot commit than when it can commit. This makes it theoretically possible that 'pooling' taxation may be optimal; see, e.g., Brett and Weymark (2008a), Krause (2009b), and Guo and Krause (2015a, 2015b). In this paper, however, we focus only on 'separating' taxation, under which high-skill and low-

possible in period 2, so a participation constraint does not appear in period 1. The solution to programme $(2.4) - (2.6)$ yields the value function $W^1(\phi, w_L, w_H, \widehat{U}_H^2, V_H, \delta)$, which represents the level of social welfare attainable in the first period.

3 The Threat of Emigration and Social Welfare

It is shown in the appendix that:

Proposition *Suppose* $\widehat{U}_H^2 > V_H$. *In our dynamic nonlinear income tax model without commitment, the threat of high-skill emigration increases social welfare in period 1* $(\partial W^1(\cdot)/\partial V_H > 0)$, *decreases social welfare in period 2* $(\partial W^2(\cdot)/\partial V_H < 0)$, *and has no effect on social welfare summed over the two periods* $(\partial W^1(\cdot)/\partial V_H + \delta \partial W^2(\cdot)/\partial V_H = 0)$.

The proposition implies that the threat of high-skill emigration facilitates redistribution and increases social welfare in the short-run, and has no effect on the level of social welfare attainable over the long-run. The intuition follows from a key feature of dynamic nonlinear income taxation without commitment. In the absence of a participation constraint, high-skill individuals know that if they reveal their type in period 1, they will be subjected to first-best redistributive taxation in period 2. Therefore, high-skill individuals must be offered an attractive tax treatment in period 1 if they are to be willing to reveal their type, to compensate them for the unattractive tax treatment they will face in period 2. This restricts the government's ability to redistribute and raise social welfare in period 1. Now consider the effects of the participation constraint. The participation constraint means that high-skill individuals obtain more utility in period 2, because they will emigrate unless they receive at least their reservation utility. Therefore, high-skill individuals now have less incentive to conceal their type in period 1. Accordingly, the government can now offer high-skill individuals a less attractive tax treatment in period 1 and still obtain skill-type information. Taken together, this implies that social welfare is lower in period 2 (due to the participation constraint), but higher in period 1 (as the incentive-compatibility constraint is relaxed). Moreover, social welfare summed over the

skill individuals always receive different tax treatments. The effects of pooling taxation are discussed further in Section 4.

two periods remains unchanged, because optimal taxation balances the short-run benefits *vis-a-vis* relaxation of the incentive-compatibility constraint against the long-run costs of the participation constraint.

3.1 A Numerical Example

As our proposition assumes that $\widehat{U}_H^2 > V_H$, in this subsection we provide a numerical example of our result. We also explain why our assumption, and hence our result, will be valid unless V_H is so high as to undo a key feature of dynamic nonlinear taxation without commitment.

In the numerical example we assume that the utility function takes the form $\ln(c_i^t) - l_i^t$. This is based on Chetty (2006), who concludes that a reasonable estimate of the coefficient of relative risk aversion is one (log utility). The OECD (2014) reports that approximately one-third of adults have attained tertiary level education. We assume that these individuals are high-skill and the remainder are low-skill, i.e., we set $\phi = 1/3$. Fang (2006) and Goldin and Katz (2007) estimate that the college wage premium is approximately 60%. We therefore normalise the low-skill type's wage rate to unity ($w_L = 1$), and set the high-skill type's wage rate at $w_H = 1.6$. Following common practice, we assume an annual discount rate of 4%. However, as most individuals work for around 40 years of their lives, we take each period to be 20 years in length. An annual discount rate of 4% then implies a 20-year discount factor of $\delta = 0.456$. The table reports the parameter values used in the numerical example, as well as the results for the cases of $V_H = -1.0$ and $V_H = -0.9$.

TABLE ABOUT HERE

It can be seen that as the value of the participation constraint is increased from $V_H = -1.0$ to $V_H = -0.9$, the discounted sum of social welfare over the two periods remains unchanged (at -1.208). Social welfare increases in period 1 and decreases in period 2. These findings are an example of our proposition, since the assumption $\widehat{U}_H^2 > V_H$ is satisfied for the parameters chosen.

For the case of $V_H = -1.0$, it can be seen that the utility a high-skill individual

obtains in period 1 (-0.432 by choosing $\langle c_H^1, y_H^1\rangle$) is greater than what they could obtain by mimicking (-0.710 by choosing $\langle c_L^1, y_L^1\rangle$). Likewise, for $V_H = -0.9$, the utility a high-skill individual obtains in period 1 (-0.478 by choosing $\langle c_H^1, y_H^1\rangle$) is greater than what they could obtain by mimicking (-0.687 by choosing $\langle c_L^1, y_L^1\rangle$). As discussed earlier, a key feature of dynamic nonlinear taxation without commitment is that high-skill individuals know that they will be subjected to first-best taxation in period 2 if they reveal their type in period 1. Accordingly, high-skill individuals must be offered an attractive tax treatment in period 1 to reveal their type, as compensation for the unattractive tax treatment they will receive in period 2. Therefore, compared to static settings, $\langle c_H^1, y_H^1\rangle$ will be favourable and $\langle c_L^1, y_L^1\rangle$ will be unfavourable. That is, $u(c_H^1) - \frac{y_H^1}{w_H} > u(c_L^1) - \frac{y_L^1}{w_H}$ is a general feature of dynamic nonlinear taxation without commitment, while $u(c_H^1) - \frac{y_H^1}{w_H} = u(c_L^1) - \frac{y_L^1}{w_H}$ would hold in static settings. It is only if V_H was so high as to undo this general feature of dynamic nonlinear taxation without commitment that our assumption would no longer be valid.

3.2 A General Condition

In this subsection, we provide a general condition under which $\widehat{U}_H^2 > V_H$, assuming log utility. It is shown in the appendix that:

$$\widehat{U}_H^2 > V_H \qquad \textit{if and only if} \qquad (1-\phi)\left[\ln(w_L) - \frac{w_L}{w_H}\right] + \phi\left[\ln(w_H) - 1\right] > V_H \tag{3.1}$$

For any particular value of the participation constraint, equation (3.1) can be used to check if $\widehat{U}_H^2 > V_H$. For example, one can check that (3.1) is satisfied for the parameters used in our numerical example.

4 Discussion and Some Caveats

Our proposition is driven by an important feature of dynamic nonlinear income taxation without commitment, namely, that high-skill individuals must be offered an attractive tax treatment in period 1 to reveal their type. This restricts redistribution in period 1 and reduces social welfare. In fact, it is possible that the welfare cost is sufficiently

severe that it would be better if the government did not induce skill-type revelation. That is, it may be optimal for the government to implement pooling taxation. This involves the government offering a single tax treatment to all individuals, and therefore skill-type information is not revealed.

Our analysis implicitly assumes that pooling taxation is not optimal. If pooling taxation is optimal, our result no longer holds. However, the question arises as to how likely it is that pooling becomes optimal. Pooling taxation is quite severe in that it imposes the same tax treatment on everyone. Guo and Krause (2015b) show, for a range of empirically-plausible parameter values, that pooling taxation is not optimal in either two-period or infinite-horizon settings. Indeed, they show that even the autarkic equilibrium yields a higher level of social welfare than does pooling taxation. Therefore, while it is possible that pooling taxation is optimal, it is less likely than regular separating taxation (under which our result holds).

Perhaps a more serious caveat follows from our assumption that the participation constraint appears only in period 2. If one assumes that high-skill individuals are immobile in the short-run and mobile in the long-run, then our assumption seems reasonable. However, if a participation constraint for high-skill individuals was introduced into period 1, and it rendered the incentive-compatibility constraint slack, then our result would no longer hold. For this to occur, the value of the first-period participation constraint would have to be quite high, since incentive-compatibility requires that high-skill individuals be offered a generous first-period tax treatment (as discussed earlier). Nevertheless, it is worth noting that our result would not hold in that case. It is also worth noting that for sufficiently high values of the participation constraint, it might be better to simply let high-skill individuals emigrate. But we do not consider that possibility here.

Finally, nonlinear income taxation *a la* Mirrlees (1971) can be viewed as an application of principal-agent theory, so one may wonder if an analogue of our result would apply in other applications of principal-agent theory in dynamic settings. For example, repeated contracting *vis-a-vis* insurance, regulation, or employment agreements.[4] An

[4]Laffont and Martimort (2002) provide a textbook treatment of principal-agent theory and its applications.

important difference is that in nonlinear income taxation, the principal (the government) seeks to maximise the aggregate welfare of the agents (the individuals), rather than having its own separate objective. That said, in all dynamic applications without commitment, the agent-type with the information advantage must be compensated to give-up that advantage. So it seems possible that an analogue of our result may apply in these other applications as well.

5 Conclusion

In this paper we have shown, using a simple dynamic nonlinear income tax model without commitment, that the threat of high-skill emigration can increase social welfare in the short-run and have no effect on social welfare over the long-run. While this result may be viewed as being primarily of theoretical interest, it also provides a new example of how restrictions on policy instruments that are clearly welfare-reducing when the government can commit, may in fact be beneficial when the government cannot commit.

6 Appendix

Proof of the Proposition

The Lagrangian corresponding to programme (2.1) − (2.3) is:

$$\mathcal{L}^2 = (1-\phi)\left\{u(c_L^2) - \frac{y_L^2}{w_L}\right\} + \phi\left\{u(c_H^2) - \frac{y_H^2}{w_H}\right\}$$

$$+ \lambda^2\left\{(1-\phi)\left[y_L^2 - c_L^2\right] + \phi\left[y_H^2 - c_H^2\right]\right\} + \alpha_H^2\left\{u(c_H^2) - \frac{y_H^2}{w_H} - V_H\right\} \tag{A.1}$$

where $\lambda^2 > 0$ and $\alpha_H^2 > 0$ are Lagrange multipliers. The first-order conditions are:

$$(1-\phi)u'(c_L^2) - \lambda^2(1-\phi) = 0 \tag{A.2}$$

$$-(1-\phi)\frac{1}{w_L} + \lambda^2(1-\phi) = 0 \tag{A.3}$$

$$\phi u'(c_H^2) - \lambda^2\phi + \alpha_H^2 u'(c_H^2) = 0 \tag{A.4}$$

$$-\phi\frac{1}{w_H}+\lambda^2\phi-\alpha_H^2\frac{1}{w_H}=0 \tag{A.5}$$

$$(1-\phi)\left[y_L^2-c_L^2\right]+\phi\left[y_H^2-c_H^2\right]=0 \tag{A.6}$$

$$u(c_H^2)-\frac{y_H^2}{w_H}-V_H=0 \tag{A.7}$$

By the Envelope Theorem:

$$\frac{\partial W^2(\cdot)}{\partial V_H}=\frac{\partial \mathcal{L}^2(\cdot)}{\partial V_H}=-\alpha_H^2=-\phi\left[\frac{w_H}{w_L}-1\right] \tag{A.8}$$

where the last equality follows from (A.3) and (A.5). This shows that $\partial W^2(\cdot)/\partial V_H<0$.

Under the assumption that $\widehat{U}_H^2>V_H$, the Lagrangian corresponding to programme $(2.4)-(2.6)$ is:

$$\mathcal{L}^1=(1-\phi)\left\{u(c_L^1)-\frac{y_L^1}{w_L}\right\}+\phi\left\{u(c_H^1)-\frac{y_H^1}{w_H}\right\}+\lambda^1\left\{(1-\phi)\left[y_L^1-c_L^1\right]+\phi\left[y_H^1-c_H^1\right]\right\}$$

$$+\theta_H^1\left\{u(c_H^1)-\frac{y_H^1}{w_H}+\delta V_H-u(c_L^1)+\frac{y_L^1}{w_H}-\delta u(c_L^2(\cdot))+\frac{\delta y_L^2(\cdot)}{w_H}\right\} \tag{A.9}$$

where $\lambda^1>0$ and $\theta_H^1>0$ are Lagrange multipliers, and use is made of the fact that $\widehat{U}_H^2=u(c_L^2(\cdot))-\frac{y_L^2(\cdot)}{w_H}$. The first-order conditions on y_L^1 and y_H^1 are, respectively:

$$-(1-\phi)\frac{1}{w_L}+\lambda^1(1-\phi)+\theta_H^1\frac{1}{w_H}=0 \tag{A.10}$$

$$-\phi\frac{1}{w_H}+\lambda^1\phi-\theta_H^1\frac{1}{w_H}=0 \tag{A.11}$$

By the Envelope Theorem:

$$\frac{\partial W^1(\cdot)}{\partial V_H}=\frac{\partial \mathcal{L}^1(\cdot)}{\partial V_H}=\delta\theta_H^1\left[1-u'(c_L^2)\frac{\partial c_L^2(\cdot)}{\partial V_H}+\frac{1}{w_H}\frac{\partial y_L^2(\cdot)}{\partial V_H}\right]=\delta\theta_H^1\left[1-\frac{1}{w_L}\frac{\partial c_L^2(\cdot)}{\partial V_H}+\frac{1}{w_H}\frac{\partial y_L^2(\cdot)}{\partial V_H}\right] \tag{A.12}$$

where the last equality makes use of (A.2) and (A.3).

From (A.2) and (A.3) we also obtain:

$$u'(c_L^2)-\frac{1}{w_L}=0 \qquad \Longrightarrow \qquad \frac{\partial c_L^2(\cdot)}{\partial V_H}=0 \tag{A.13}$$

Likewise, from (A.3) − (A.5) we obtain:

$$u'(c_H^2) - \frac{1}{w_H} = 0 \qquad \Longrightarrow \qquad \frac{\partial c_H^2(\cdot)}{\partial V_H} = 0 \tag{A.14}$$

From (A.7) and using (A.14) we obtain:

$$y_H^2 - w_H \left[u(c_H^2) - V_H\right] = 0 \qquad \Longrightarrow \qquad \frac{\partial y_H^2(\cdot)}{\partial V_H} = -w_H \tag{A.15}$$

From (A.6) and using (A.13) − (A.15) we obtain:

$$y_L^2 - c_L^2 - \frac{\phi c_H^2}{1-\phi} + \frac{\phi y_H^2}{1-\phi} = 0 \qquad \Longrightarrow \qquad \frac{\partial y_L^2(\cdot)}{\partial V_H} = \frac{\phi w_H}{1-\phi} \tag{A.16}$$

Equation (A.12) can now be simplified to:

$$\frac{\partial W^1(\cdot)}{\partial V_H} = \delta\phi \left[\frac{w_H}{w_L} - 1\right] \tag{A.17}$$

where use has been made of (A.10) and (A.11). This shows that $\partial W^1(\cdot)/\partial V_H > 0$.

Finally, equations (A.8) and (A.17) imply that $\partial W^1(\cdot)/\partial V_H + \delta \partial W^2(\cdot)/\partial V_H = 0$. ■

Derivation of Equation (3.1)

Assuming log utility, equations (A.2) − (A.7) can be solved to yield:

$$c_L^2 = w_L \qquad \text{and} \qquad y_L^2 = w_L - \frac{\phi w_H}{(1-\phi)} \left[\ln(w_H) - 1 - V_H\right] \tag{A.18}$$

Recall that $\widehat{U}_H^2 = u(c_L^2) - \frac{y_L^2}{w_H}$. Using (A.18) and after some algebraic manipulation, it can be shown that:

$$\widehat{U}_H^2 > V_H \qquad \Longleftrightarrow \qquad (1-\phi)\left[\ln(w_L) - \frac{w_L}{w_H}\right] + \phi\left[\ln(w_H) - 1\right] > V_H \tag{A.19}$$

which is equation (3.1). ■

References

[1] Apps, P. and R. Rees (2006), "Repeated Optimal Nonlinear Income Taxation", *mimeo.*

[2] Berliant, M. and J. Ledyard (2014), "Optimal Dynamic Nonlinear Income Taxes with No Commitment", *Journal of Public Economic Theory*, 16, 196-221.

[3] Brett, C. and J. Weymark (2008a), "Optimal Nonlinear Taxation of Income and Savings without Commitment", *Vanderbilt University Working Paper*, No. 08-W05.

[4] Brett, C. and J. Weymark (2008b), "The Impact of Changing Skill Levels on Optimal Nonlinear Income Taxes", *Journal of Public Economics,* 92, 1765-1771.

[5] Brett, C. and J. Weymark (2011), "How Optimal Nonlinear Income Taxes Change when the Distribution of the Population Changes", *Journal of Public Economics*, 95, 1239-1247.

[6] Chetty, R. (2006), "A New Method of Estimating Risk Aversion", *American Economic Review*, 96, 1821-1834.

[7] Fang, H. (2006), "Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices", *International Economic Review*, 47, 1151-1185.

[8] Goldin, C. and L. Katz (2007), "The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890-2005", *NBER Working Paper*, No. 12984.

[9] Guo, J-T. and A. Krause (2011), "Optimal Nonlinear Income Taxation with Habit Formation", *Journal of Public Economic Theory*, 13, 463-480.

[10] Guo, J-T. and A. Krause (2013), "Optimal Nonlinear Taxation of Income and Education Expenditures", *Oxford Economic Papers*, 65, 74-95.

[11] Guo, J-T. and A. Krause (2014), "Optimal Dynamic Nonlinear Income Taxation under Loose Commitment", *Macroeconomic Dynamics*, 18, 1403-1427.

[12] Guo, J-T. and A. Krause (2015a), "Dynamic Nonlinear Income Taxation with Quasi-Hyperbolic Discounting and No Commitment", *Journal of Economic Behavior and Organization*, 109, 101-119.

[13] Guo, J-T. and A. Krause (2015b), "Dynamic Income Taxation without Commitment: Comparing Alternative Tax Systems", *Economic Modelling*, 47, 319-326.

[14] Krause, A. (2009a), "Education and Taxation Policies in the Presence of Countervailing Incentives", *Economica*, 76, 387-399.

[15] Krause, A. (2009b), "Optimal Nonlinear Income Taxation with Learning-by-Doing", *Journal of Public Economics*, 93, 1098-1110.

[16] Laffont, J. and D. Martimort (2002), *The Theory of Incentives*: *The Principal-Agent Model*, Princeton University Press.

[17] Lehmann, E., L. Simula and A. Trannoy (2014), "Tax Me If You Can! Optimal Nonlinear Income Tax Between Competing Governments", *Quarterly Journal of Economics*, 129, 1995-2030.

[18] Leite-Monteiro, M. (1997), "Redistributive Policy with Labour Mobility Across Countries", *Journal of Public Economics*, 65, 229-244.

[19] Mirrlees, J. (1971), "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies*, 38, 175-208.

[20] OECD (2014), *Highlights from Education at a Glance*, OECD publications.

[21] Osmundsen, P. (1999), "Taxing Internationally Mobile Individuals—A Case of Countervailing Incentives", *International Tax and Public Finance*, 6, 149-164.

[22] Roberts, K. (1984), "The Theoretical Limits to Redistribution", *Review of Economic Studies*, 51, 177-195.

[23] Simula, L. (2010), "Optimal Nonlinear Income Tax and Nonlinear Pricing: Optimality Conditions and Comparative Static Properties", *Social Choice and Welfare*, 35, 199-220.

[24] Simula, L. and A. Trannoy (2010), "Optimal Income Tax under the Threat of Migration by Top-Income Earners", *Journal of Public Economics*, 94, 163-173.

[25] Simula, L. and A. Trannoy (2012), "Shall we keep the Highly Skilled at Home? The Optimal Income Tax Perspective", *Social Choice and Welfare*, 39, 751-782.

[26] Weymark, J. (1987), "Comparative Static Properties of Optimal Nonlinear Income Taxes", *Econometrica*, 55, 1165-1185.

TABLE

Numerical Example

$$u(c_i^t) - l_i^t = \ln(c_i^t) - l_i^t$$

Parameter values:			
φ	0.333	w_L	1.000
δ	0.456	w_H	1.600

	$V_H = -1.0$	$V_H = -0.9$
Low-skill utility: period 1	−1.082	−1.046
High-skill utility: period 1	−0.432	−0.478
Mimicker utility: *period 1*	*−0.710*	*−0.687*
Low-skill utility: period 2	−0.624	−0.704
High-skill utility: period 2	−1.0	−0.9
Mimicker utility $\hat{U}_H^2$: *period 2*	*−0.390*	*−0.444*
Social welfare: period 1	−0.866	−0.856
Social welfare: period 2	−0.749	−0.769
Social welfare: total	−1.208	−1.208