

EPRU Working Paper Series

**Economic Policy Research Unit
Institute of Economics
University of Copenhagen
Studiestræde 6
DK-1455 Copenhagen K
DENMARK
Tel: (+45) 3532 4411
Fax: (+45) 3532 4444
Web: <http://www.econ.ku.dk/epru/>**

A Revised Efficiency Principle
for the Taxation of Couples

Henrik Jacobsen Kleven
Claus Thustrup Kreiner

2004-09

ISSN 0908-7745

**The activities of EPRU are financed by a grant from
The National Research Foundation**

A REVISED EFFICIENCY PRINCIPLE FOR THE TAXATION OF COUPLES*

Henrik Jacobsen Kleven
University of Copenhagen and EPRU

Claus Thustrup Kreiner
University of Copenhagen, EPRU, CESifo

August 2004

Abstract

We reconsider the result that efficient taxation involves a lower marginal tax on secondary earners than on primary earners. Introducing labor force participation responses into the analysis, we show that a second-earner tax allowance is better than selective marginal tax rates.

JEL Classification: H21, J22, D10

Keywords: Optimal Taxation, Married Couples, Labor Supply Behavior

*We wish to thank Peter Birch Sørensen for helpful comments. The activities of EPRU (Economic Policy Research Unit) are supported by a grant from The Danish National Research Foundation. *Correspondence:* claus.thustrup.kreiner@econ.ku.dk or henrik.kleven@econ.ku.dk.

1 Introduction

Whether the taxation of married couples should be based on the individual or the family is a classic debating point in public finance. Since the work of Boskin and Sheshinski (1983), an efficiency argument for individual-based taxation has been widely accepted. The argument is based on the empirical observation that the labor supply of secondary earners (typically wives) is more elastic than the labor supply of primary earners (typically husbands). In this case, a traditional Ramsey-type argument calls for a lower marginal tax rate on the secondary earner to minimize distortions of labor supply. This efficiency principle is met by a progressive individual-based income tax, because the secondary earner has a lower income than the primary earner. By contrast, under a fully joint income tax based on family income, the marginal tax rates on the two spouses are identical so that the Boskin-Sheshinski efficiency principle is violated.

The literature has focused exclusively on labor supply responses along the intensive margin, i.e., hours of work for those who are working. Labor supply responses along the extensive margin – the margin of entry and exit – are not included in the analysis. This is at odds with the modern empirical labor market literature which emphasizes the importance of accounting for extensive responses. An emerging consensus is that most of the observed variation in labor supply reflects changes in labor force participation. While estimates of hours-of-work elasticities tend to be close to zero for both males and females, the participation elasticity seems to be very high for married women (Heckman, 1993; Blundell and MaCurdy, 1999).

These empirical insights call for a reconsideration of the efficient tax treatment of married couples. The reason is that the distortion of labor force participation is related to a different tax rate than the one distorting hours of work. While working hours depend on the marginal tax rate, entry-exit decisions are influenced by the total tax burden on labor income and therefore

the average tax rate. Intuitively, since the observed elasticity differential between spouses is mainly due to participation effects, the Ramsey argument calls for a lower average tax rate on secondary earners. By contrast, there is no reason to differentiate marginal tax rates to increase labor supply and efficiency. In practise, the efficient tax system could be implemented by a joint income tax featuring tax allowances for two-earner couples. Interestingly, the United States had a system like that in the beginning of the 1980's.

The following sections establish our result formally and provide a numerical example showing the quantitative importance of our result. Of course, since we focus exclusively on economic efficiency, it would be premature to make firm policy conclusions based on our analysis. Yet, since the efficiency principle for individual taxation has received considerable attention over the years, it is worthwhile investigating its sensitivity to the assumption on which it is based. For an analysis of family taxation with distributional concerns, we refer to Apps and Rees (1999) and Kleven, Kreiner and Saez (2004).

2 The Analysis

2.1 Family Labor Supply Behavior

In this section, we set up a model accounting for both intensive and extensive labor supply responses. In the modelling of extensive responses, we allow for discrete entry into the labor market, consistent with empirical evidence showing that workers rarely choose very low hours of work. This type of behavior is typically explained by non-convexities in preferences or budget sets created by work costs (e.g. Cogan, 1981). These work costs may be monetary costs (say, child care), time losses (e.g. commuting time) or emotional costs associated with working. Below we adopt a stylized framework incorporating fixed work costs which may capture some of these factors.

Like Boskin and Sheshinski (1983), we consider the family to be the decision making unit and assume that behavior is determined by the maximization of a household utility function. Each family consists of two individuals, a primary earner (p) and a secondary earner (s), who differ in their preferences for work and in their market productivities. Both earners face fixed work costs which are denoted by q_p and q_s , respectively. To obtain smooth changes in labor force participation, we allow work costs to vary across families. The distribution of work costs is captured by a joint density function $f(q_p, q_s)$. The welfare of each family is represented by a quasi-linear utility function,¹

$$u(c, h_p, h_s) = c - d_p(h_p) - d_s(h_s), \quad (1)$$

where c is family consumption, h_p and h_s are working hours, while d_p and d_s denote disutility of work for the two spouses. Each disutility term is given by

$$d_i(h_i) = \begin{cases} q_i + v_i(h_i) & \text{for } h_i > 0 \\ 0 & \text{for } h_i = 0 \end{cases} \quad \text{for } i = p, s, \quad (2)$$

where $v'_i > 0$ and $v''_i < 0$. This expression decomposes the cost of labor market participation into the fixed cost, q_i , as well as a variable cost depending on the number of working hours. The disutility of non-participation is normalized to zero. The budget constraint for the household equals

$$c \leq w_p h_p - T_p(w_p h_p) + w_s h_s - T_s(w_s h_s), \quad (3)$$

where w_p and w_s are wage rates, while $T_p(\cdot)$ and $T_s(\cdot)$ are separate and spouse-specific tax functions. The assumption of separate tax functions is conventional in the literature, and it

¹The quasi-linear specification excludes income effects as often done in problems of optimal income taxation (e.g. Diamond, 1998; Saez, 2002). In our context, with family taxation and endogenous labor market entry, the problem becomes quite complicated to solve with a general preference specification.

greatly simplifies the analysis.² Moreover, we focus on tax schedules of the following form

$$T_i(w_i h_i) = \begin{cases} t_i w_i h_i + r_i & \text{for } h_i > 0 \\ 0 & \text{for } h_i = 0 \end{cases} \quad \text{for } i = p, s, \quad (4)$$

where t_i is the marginal tax rate on earnings, and r_i is a tax which is conditional on positive earnings but unrelated to the size of earnings. In other words, r_i is a tax on labor market entry or, if negative, an employment credit. Boskin and Sheshinski (1983) did not include this instrument, but it is important in our context due to endogenous labor force participation.

The household maximizes utility subject to the budget constraint. Given participation, i.e. $h_i > 0$, the optimal labor supply is characterized by the standard marginal condition

$$(1 - t_i) w_i = v'_i(h_i^*) \quad \text{for } i = p, s. \quad (5)$$

This expression shows that, if an individual chooses to work, the number of hours worked is independent of the fixed work costs. But for the individual to enter the labor market in the first place, the utility from participation must be greater than or equal to the utility from non-participation. This gives the following participation constraint

$$q_i \leq w_i h_i^* - t_i w_i h_i^* - r_i - v_i(h_i^*) \equiv q_i^* \quad \text{for } i = p, s. \quad (6)$$

Individuals with a fixed cost below the threshold-value q_i^* decide to work h_i^* hours while those with a fixed cost above the threshold choose to stay outside the labor force. Notice that the participation constraint depends on the total tax burden and therefore the average tax rate, whereas hours of work conditional on working (eq. 5) is related to the marginal tax rate.

The fraction of secondary earners who decide to participate in the labor market is given by $F_s(q_s^*) \equiv \int_0^{q_s^*} \int_0^\infty f(q_s, q_p) dq_p dq_s$, and similarly for primary earners. The aggregate labor

²For an analysis of a fully general non-separable and non-linear income tax for couples, we refer to Kleven *et al.* (2004).

supply of primary and secondary earners, respectively, becomes

$$L_i = h_i^* F_i(q_i^*) \quad \text{for } i = p, s, \quad (7)$$

where we have normalized the population of families to one. Hence, aggregate labor supply is a product of hours of work for those who are working and the labor force participation rate. The responsiveness along the two margins of labor supply may be captured by an hours-of-work elasticity and a participation elasticity. The hours-of-work elasticity ε_i is defined with respect to the marginal net-of-tax wage, $(1 - t_i) w_i$, and the participation elasticity η_i is defined with respect to the net-of-tax income gain from entry, $w_i h_i^* - t_i w_i h_i^* - r_i$. We obtain

$$\varepsilon_i \equiv \frac{v_i'(h_i^*)}{v_i''(h_i^*) h_i^*}, \quad \eta_i \equiv F_i'(q_i^*) \frac{w_i h_i^* - t_i w_i h_i^* - r_i}{F_i(q_i^*)} \quad \text{for } i = p, s, \quad (8)$$

where we have used the first-order conditions (5) and (6).

2.2 Optimal Tax Rules

To study the properties of an efficient tax treatment of married couples, we derive aggregate utilitarian welfare by integrating utility over all households. This gives

$$\begin{aligned} U &= \int_0^{q_p^*} \int_0^\infty [w_p h_p^* - t_p w_p h_p^* - r_p - q_p - v_p(h_p^*)] f(q_p, q_s) dq_s dq_p \\ &\quad + \int_0^{q_s^*} \int_0^\infty [w_s h_s^* - t_s w_s h_s^* - r_s - q_s - v_s(h_s^*)] f(q_p, q_s) dq_p dq_s. \end{aligned} \quad (9)$$

The government budget constraint equals

$$F_p(q_p^*) (t_p w_p h_p^* + r_p) + F_s(q_s^*) (t_s w_s h_s^* + r_s) \geq R, \quad (10)$$

where R is an exogenous revenue requirement. To derive the efficient tax system, we maximize aggregate utilitarian welfare (9) with respect to the tax parameters t_p , t_s , r_p , and r_s subject to the government budget constraint (10) and the labor supply decision rules (5)-(6). However,

before proceeding to this problem, let us briefly look at the more restrictive problem of proportional separate tax schedules ($r_p = r_s = 0$), also considered by Boskin and Sheshinski (1983). Assuming that the problem is well-behaved, we obtain the optimal tax rule (see Appendix A)

$$\frac{t_p / (1 - t_p)}{t_s / (1 - t_s)} = \frac{\varepsilon_s + \eta_s}{\varepsilon_p + \eta_p}. \quad (11)$$

This policy rule is reminiscent of the Boskin-Sheshinski result that we should differentiate the taxation of spouses according to their labor supply elasticities. There is a subtle difference, however, in that our version of the optimality condition discerns hours-of-work responses from participation responses. The formula thus emphasizes that taxation should be differentiated according to *total* elasticities including both margins of labor supply response:

Proposition 1 *With proportional and separate tax schedules, the spouse with the lower total labor supply elasticity, $\varepsilon_i + \eta_i$, should face the higher tax rate.*

The analysis underlying Proposition 1 stipulates that marginal and average tax rates are identical for each spouse. This is an important limitation in our context, because marginal and average taxes operate through different margins of labor supply response. To separate the average from the marginal tax rate, we incorporate the instrument r_i into the optimal tax analysis. However, with no further restrictions, the solution would become trivial. It is easy to see that r_i is a more efficient instrument than t_i , such that the optimal policy would involve zero marginal tax rates.³ This result is not interesting, since it derives solely from the fact that we do not incorporate distributional concerns. To avoid it, we introduce an additional restriction

$$r_p F_p(q_p^*) + r_s F_s(q_s^*) = 0. \quad (12)$$

³This is because r_i distorts only the participation decision, whereas the marginal tax rate t_i distorts participation and hours of work at the same time.

This constraint implies that the r_i 's cannot affect the aggregate tax revenue collected; they affect only the distribution of tax payments on primary and secondary earners. For example, a policy with $r_s < 0$ (and hence $r_p > 0$) corresponds to a second-earner tax allowance financed by a tax on the primary earner. In this way, the policy maker can shift average tax burdens across spouses without changing their marginal tax rates, which is clearly a realistic policy option. The extra degree of freedom would not be important if one were to consider only intensive margin responses, but becomes important once participation effects are accounted for.

The solution to the revised optimization problem is found by maximizing (9) with respect to t_p , t_s , r_p , and r_s subject to eqs (5), (6), (10), and (12). Letting $a_i \equiv T_i(\cdot) / (w_i h_i^*)$ denote the average tax rate, the optimal tax system is characterized by (see Appendix B)

$$\frac{t_p / (1 - t_p)}{t_s / (1 - t_s)} = \frac{\varepsilon_s}{\varepsilon_p}, \quad (13)$$

$$\frac{\frac{1-t_p}{1-a_p} \left[\frac{a_p}{t_p} (1 - \Omega) + \Omega \right]}{\frac{1-t_s}{1-a_s} \left[\frac{a_s}{t_s} (1 - \Omega) + \Omega \right]} = \frac{\eta_s / \eta_p}{\varepsilon_s / \varepsilon_p}, \quad (14)$$

where $\Omega \equiv \frac{t_i}{1-t_i} \varepsilon_i < 1$. From these two formulae, we obtain

Proposition 2 (i) *For the spouse with the higher (lower) hours-of-work elasticity, the marginal tax rate should be lower (higher). (ii) For the spouse with the higher (lower) participation elasticity relative to hours-of-work elasticity, the average tax rate should be lower (higher) than the marginal tax rate.*

Comparing Propositions 1 and 2, we see that the results change remarkably once a more general tax system is allowed for. First of all, information about total labor supply elasticities is no longer sufficient for policy design; we need to distinguish the two margins of labor supply response. Relative marginal tax rates are determined by hours-of-work elasticities alone. Al-

though the marginal tax rates t_p and t_s do affect the extensive margin, the marginal deadweight losses on this margin may be equalized across spouses through a suitable shift in tax allowances (r_p and r_s). Hence, the policy maker should focus entirely on the intensive margin when setting marginal tax rates.

The differentiation of average tax rates, on the other hand, is determined by the ratio of participation elasticities (η_s/η_p) over the ratio of hours-of-work elasticities ($\varepsilon_s/\varepsilon_p$). Only in the implausible case where relative participation elasticities equal relative hours-of-work elasticities is it optimal to differentiate average and marginal tax rates to the same degree. As mentioned in the Introduction, estimated participation elasticities are much higher for secondary earners than for primary earners, while hours-of-work elasticities tend to be very low for both husbands and wives. Proposition 2 then shows that marginal tax rates should be identical, whereas the average/total tax burden should be differentiated in favor of the secondary earner.

2.3 A Numerical Example

To get a sense of the quantitative importance of our results, it is useful to consider a numerical example. In line with the empirical literature, we assume that the hours-of-work elasticity is 0.1 for both spouses, and that the participation elasticity for the secondary earner is five times higher than for the primary earner. Moreover, we assume that the government must collect a tax revenue equal to 40 per cent of aggregate earnings, that secondary earnings (conditional on entry) equal 2/3 of primary earnings, and that the secondary earners' participation rate is 2/3 of the primary earners's rate. In this example, it is straightforward to see that the efficient tax system features a 40% marginal tax rate for both spouses. Moreover, by solving numerically eqs (12) and (14), we find that the optimal average tax rate of the secondary earner is around 15%, while the average tax rate of the primary earner is circa 50%. To conclude, the efficiency

principle derived in this paper may call for a large differentiation of average tax rates across spouses, implemented by a second-earner tax allowance. Marginal tax rates, on the other hand, need not be differentiated for efficiency purposes.

A Derivation of eq. (11)

The government's problem is to maximize (9) subject to (10) under the assumption that $r_p = r_s = 0$. The Lagrangian becomes

$$\begin{aligned} L = & \int_0^{q_p^*} \int_0^\infty [w_p h_p^* - t_p w_p h_p^* - q_p - v_p(h_p^*)] f(q_p, q_s) dq_s dq_p \\ & + \int_0^{q_s^*} \int_0^\infty [w_s h_s^* - t_s w_s h_s^* - q_s - v_s(h_s^*)] f(q_p, q_s) dq_p dq_s \\ & + \lambda [F_p(q_p^*) t_p w_p h_p^* + F_s(q_s^*) t_s w_s h_s^* - R], \end{aligned}$$

where λ is a Lagrangian multiplier and where h_i^* and q_i^* are determined by eqs (5) and (6). The first-order condition with respect to t_p equals

$$\frac{\partial L}{\partial t_p} = -w_p h_p^* F_p(q_p^*) + \lambda \left[F_p(q_p^*) w_p h_p^* + F_p(q_p^*) t_p w_p \cdot \frac{\partial h_p^*}{\partial t_p} + F_p'(q_p^*) t_p w_p h_p^* \cdot \frac{\partial q_p^*}{\partial t_p} \right] = 0,$$

where we have used the envelope theorem and the definition $F_p(q_p^*) \equiv \int_0^{q_p^*} \int_0^\infty f(q_p, q_s) dq_p dq_s$.

From eqs (5) and (6), we obtain $\partial h_p^* / \partial t_p = -w_p / v''(h_p^*)$ and $\partial q_p^* / \partial t_p = -w_p h_p^*$. Inserting these derivatives and the elasticities (8), the first-order condition may be rewritten to

$$\frac{t_p}{1 - t_p} (\varepsilon_p + \eta_p) = \frac{\lambda - 1}{\lambda}.$$

Since the two earners are symmetric in the model, the first-order condition for t_s may be obtained simply by changing the subscript from p to s in the above equation. Combining the two first-order conditions to eliminate λ , we obtain eq. (11).

B Derivation of eqs (13) and (14)

The government's problem is to maximize (9) subject to (10) and (12). The Lagrangian becomes

$$\begin{aligned}
L &= \int_0^{q_p^*} \int_0^\infty [w_p h_p^* - t_p w_p h_p^* - r_p - q_p - v_p(h_p^*)] f(q_p, q_s) dq_s dq_p \\
&\quad + \int_0^{q_s^*} \int_0^\infty [w_s h_s^* - t_s w_s h_s^* - r_s - q_s - v_s(h_s^*)] f(q_p, q_s) dq_p dq_s \\
&\quad + \lambda_1 [F_p(q_p^*) (t_p w_p h_p^* + r_p) + F_s(q_s^*) (t_s w_s h_s^* + r_s) - R] \\
&\quad + \lambda_2 [r_p F_p(q_p^*) + r_s F_s(q_s^*)],
\end{aligned}$$

where λ_1 and λ_2 are Lagrangian multipliers and where h_i^* and q_i^* are determined by eqs (5) and (6). The first-order conditions for t_p and r_p are given by (there are of course symmetric conditions for t_s and r_s):

$$\begin{aligned}
\frac{\partial L}{\partial t_p} &= -w_p h_p^* F_p(q_p^*) + \lambda_1 \left[F_p(q_p^*) w_p h_p^* + F_p(q_p^*) t_p w_p \cdot \frac{\partial h_p^*}{\partial t_p} + F_p'(q_p^*) (t_p w_p h_p^* + r_p) \cdot \frac{\partial q_p^*}{\partial t_p} \right] \\
&\quad + \lambda_2 r_p F_p'(q_p^*) \cdot \frac{\partial q_p^*}{\partial t_p} = 0, \\
\frac{\partial L}{\partial r_p} &= -F_p(q_p^*) + \lambda_1 \left[F_p(q_p^*) + F_p'(q_p^*) (t_p w_p h_p^* + r_p) \cdot \frac{\partial q_p^*}{\partial r_p} \right] \\
&\quad + \lambda_2 \left[F_p(q_p^*) + r_p F_p'(q_p^*) \cdot \frac{\partial q_p^*}{\partial r_p} \right] = 0
\end{aligned}$$

where we have used the envelope theorem and the definition $F_p(q_p^*) \equiv \int_0^{q_p^*} \int_0^\infty f(q_p, q_s) dq_p dq_s$.

From eqs (5) and (6), we obtain $\partial h_p^* / \partial t_p = -w_p / v''(h_p^*)$, $\partial q_p^* / \partial t_p = -w_p h_p^*$, and $\partial q_p^* / \partial r_p = -1$.

Inserting these derivatives along with (8), the above conditions may be rewritten to

$$-1 + \lambda_1 \left(1 - \frac{t_p}{1 - t_p} \varepsilon_p - \frac{t_p w_p h_p^* + r_p}{w_p h_p^* - t_p w_p h_p^* - r_p} \eta_p \right) - \lambda_2 \frac{r_p}{w_p h_p^* - t_p w_p h_p^* - r_p} \eta_p = 0, \quad (15)$$

$$1 + \lambda_1 \left[-1 + \frac{t_p w_p h_p^* + r_p}{w_p h_p^* - t_p w_p h_p^* - r_p} \eta_p \right] + \lambda_2 \left[-1 + \frac{r_p}{w_p h_p^* - t_p w_p h_p^* - r_p} \eta_p \right] = 0. \quad (16)$$

By adding these two equalities and simplifying, we obtain

$$\frac{t_p}{1 - t_p} \varepsilon_p = -\frac{\lambda_2}{\lambda_1} \equiv \Omega. \quad (17)$$

This relationship may be combined with a symmetric relationship for the other spouse so as to obtain eq. (13).

In order to derive eq. (14), we combine (16) with the similar equation for the other spouse.

This gives

$$\frac{\frac{t_p w_p h_p^* + r_p}{w_p h_p^* - t_p w_p h_p^* - r_p} - \frac{r_p}{w_p h_p^* - t_p w_p h_p^* - r_p} \Omega}{\frac{t_s w_s h_s^* + r_s}{w_s h_s^* - t_s w_s h_s^* - r_s} - \frac{r_s}{w_s h_s^* - t_s w_s h_s^* - r_s} \Omega} = \frac{\eta_s}{\eta_p}.$$

By inserting the definition of the average tax rate, a_p , and noting that $a_p - t_p = \frac{r_p}{w_p h_p^*}$, the above relationship may be rewritten to

$$\frac{\frac{a_p}{1-a_p} + \frac{t_p - a_p}{1-a_p} \Omega}{\frac{a_s}{1-a_s} + \frac{t_s - a_s}{1-a_s} \Omega} = \frac{\eta_s}{\eta_p}$$

Finally, eq. (14) is obtained by dividing the above expression with (13).

In order to see that $\Omega < 1$, notice that eq. (16) and the symmetric equation for the other spouse imply

$$\begin{aligned} 1 - \lambda_1 - \lambda_2 &= -\lambda_1 \frac{t_p w_p h_p^*}{w_p h_p^* - t_p w_p h_p^* - r_p} \eta_p - (\lambda_1 + \lambda_2) \frac{r_p}{w_p h_p^* - t_p w_p h_p^* - r_p} \eta_p, \\ 1 - \lambda_1 - \lambda_2 &= -\lambda_1 \frac{t_s w_s h_s^*}{w_s h_s^* - t_s w_s h_s^* - r_s} \eta_s - (\lambda_1 + \lambda_2) \frac{r_s}{w_s h_s^* - t_s w_s h_s^* - r_s} \eta_s. \end{aligned}$$

Since the Lagrangian multiplier λ_1 reflects the utility loss of collecting additional revenue, this parameter is positive. This implies that the first term on each right-hand side is negative. The second terms on the right-hand sides have opposite signs because r_p and r_s have opposite signs.

Then it must be the case that $1 - \lambda_1 - \lambda_2 < 0$, which implies

$$\Omega = -\frac{\lambda_2}{\lambda_1} < \frac{\lambda_1 - 1}{\lambda_1} < 1,$$

where the last inequality follows from $\lambda_1 > 0$.

References

- Apps, P.F. and R. Rees** (1999). "On the Taxation of Trade within and between Households." *Journal of Public Economics* 73, 241-263.
- Blundell, R. and T. MaCurdy** (1999). "Labor Supply: A Review of Alternative Approaches," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* vol. 3A. Elsevier Science B.V.: Amsterdam.
- Boskin, M.J. and E. Sheshinski** (1983). "Optimal Tax treatment of the Family: Married Couples." *Journal of Public Economics* 20, 281-297.
- Cogan, J.F.** (1981). "Fixed Costs and Labor Supply." *Econometrica* 49, 945-963.
- Diamond, P.** (1998), "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, **88**, 83-95.
- Heckman, J. J.** (1993). "What Has Been Learned About Labor Supply in the Past Twenty Years?" *American Economic Review Papers and Proceedings* 83(2), 116-121.
- Kleven, H.J., C.T. Kreiner, and E. Saez** (2004). "The Optimal Taxation of Couples." Working Paper, UC Berkeley.
- Saez, E.** (2002). "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses." *Quarterly Journal of Economics* 117(3), 1039-1073.