

LABOR SUPPLY: A REVIEW OF ALTERNATIVE APPROACHES

RICHARD BLUNDELL*

University College London and Institute for Fiscal Studies

THOMAS MACURDY*

Department of Economics and The Hoover Institution, Stanford University

Contents

Abstracts	1560
JEL codes	1560
1 Introduction	1560
2 How have tax and welfare policies changed?	1563
2.1 US tax and welfare programs	1564
2.2 UK tax and welfare programs	1569
3 Recent empirical trends	1572
3.1 Data sources	1574
3.2 Participation	1577
3.3 Hours of work	1580
3.4 Real wages	1584
4 A framework for understanding labor supply	1586
4.1 The static labor supply model	1587
4.2 Multiperiod models of labor supply under certainty	1591
4.3 Multiperiod models of labor supply under uncertainty	1596
4.4 Basic empirical specifications	1598
4.5 Which elasticities for policy evaluation?	1603
5 Policy reforms and the natural experiment approach	1607
5.1 The natural-experiment approach and the difference-in-differences estimator	1608
5.2 Does the difference-in differences estimator measure behavioral responses?	1613
5.3 A review of some empirical applications	1615
6 Estimation with non-participation and non-linear budget constraints	1617
6.1 Basic economic model with taxes	1618

* We would like to thank John Pencavel for providing the US data, Christian Dustmann for the German data, Howard Reed for the British data, and Lennart Flood for the Swedish data. Jed DeVaro and Mika Kuusmanen provided able research assistance and many helpful comments. We also thank Soren Blomquist, James Heckman, Ian Walker and Valerie Lechene for comments on sections of earlier drafts. Blundell thanks the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at IFS for financial support. MaCurdy gratefully acknowledges research support from NIH grant HD32055-02. Opinions expressed in this chapter are those of the authors and do not represent the official position or policy of any agency funding this research.

6.2	Instrumental-variable estimation	1622
6.3	Maximum likelihood: convex differential constraints with full participation	1626
6.4	Maximum likelihood: convex piecewise-linear constraints with full participation	1629
6.5	Maximum likelihood: accounting for fixed costs of participation and missing wages	1635
6.6	Welfare participation and non-convex budget constraints	1638
6.7	An approach for computational simplification and discrete hours choices	1643
6.8	Survey of empirical findings for non-linear budget constraints models	1644
7	Family labor supply	1657
7.1	The basic economic model of family labor supply	1657
7.2	The collective model of family labor supply	1661
7.3	Some empirical findings for the family labor supply model	1665
8	Structural dynamic models	1672
8.1	The standard intertemporal labor supply model with participation	1672
8.2	Learning by doing and human capital	1676
8.3	Habit persistence	1680
8.4	Review of empirical results for structural dynamic models	1680
9	Closing comments	1684
	Appendix A. Specifications of within-period preferences	1686
	References	1689

Abstract

This chapter surveys existing approaches to modeling labor supply and identifies important gaps in the literature that could be addressed in future research. The discussion begins with a look at recent policy reforms and labor market facts that motivate the study of labor supply. The analysis then presents a unifying framework that allows alternative empirical formulations of the labor supply model to be compared and their resulting elasticities to be interpreted. This is followed by critical reviews of alternative approaches to labor-supply modeling. The first review assesses the difference-in-differences approach and its relationship to natural experiments. The second analyzes estimation with non-linear budget constraints and welfare-program participation. The third appraises developments of family labor-supply models including both the standard unitary and collective labor-supply formulations. The fourth briefly explores dynamic extensions of the labor supply model, characterizing how participation decisions, learning-by-doing, human capital accumulation and habit formation affect the analysis of the lifecycle model. At the end of each of the four broad reviews, we summarize a selection of the recent empirical findings. The concluding section asks whether the developments reviewed in this chapter place us in a better position to answer the policy-reform questions and to interpret the trends in participation and hours with which we began this review. © 1999 Elsevier Science B.V. All rights reserved.

JEL codes: J21; J22; J24; C21; C24

1. Introduction

Consistent with its tradition, research on labor supply during the past decade has been at the forefront of developments in empirical microeconomics. At the same time, an important component of this research has rebuffed sophisticated estimation approaches in favor

of simple methods for evaluating behavioral responses underlying hours-of-work decisions. The attention devoted to the study of labor supply arises from intense interests in assessing the consequences of a wide array of public policies, ranging from tax and welfare programs to the alteration of institutional features of labor markets. A further motivation concerns the curiosity of economists in explaining the factors underlying the dramatic changes in employment patterns that have occurred in recent years, trends that show no evidence of stabilizing in the near future.

After presenting a brief overview of the phenomena stimulating recent analyses of labor supply, this chapter pursues its main purpose of reviewing the empirical developments and findings produced by this research. It focuses on work done since the surveys of Pencavel (1986) and Killingsworth and Heckman (1986), which ably summarized the labor supply literature in the previous *Handbook of Labor Economics*. We draw widely on existing research in the labor supply literature. Our discussion of methodological developments presents simplified examples to highlight essential ideas, not attempting to attribute each development to specific authors and, thus, omitting most references in this discussion. We do not claim originality in this survey, and our discussion of applications refers to many of the studies that have made the major contributions to this research area since the earlier *Handbook* surveys. It is inevitable that we have omitted references and we apologize for such omissions.

The influence of governmental programs on people's employment and hours of work is often a critical consideration in the design of policies. Indeed, the primary objective of many recent reforms in both tax and welfare programs in North America, the UK, Scandinavia and other parts of Europe has been to encourage participants to increase their work effort. Few decades match the most recent in terms of how much change has occurred in tax and welfare policies. Understanding labor supply behavior is vital in formulating proposals that build in work incentives while providing income support.

This chapter begins with a cursory description of how tax and welfare policies have changed in recent years, considering how these changes enter the picture of labor supply and its empirical analyses. For this discussion, we focus on reforms in the US and the UK. This is not simply because of our own local knowledge but also because these two countries have been at the forefront of introducing welfare and tax reforms designed to encourage work effort – in particular, the move toward in-work benefits. These “Welfare to Work” proposals form a particularly attractive background against which to motivate labor supply analysis as they are generally reforms directly aimed at addressing the decline in participation among certain types of workers. The analysis of participation in work is key to the evaluation of welfare-to-work reforms and this is the margin over which labor supply responses may be most responsive. However, to properly evaluate the impact of a welfare-to-work policy reform, such as the Earned Income Tax Credit in the US, requires a careful examination of the balance between the labor supply decisions of those individuals already working who may now face a higher benefit (or credit) reduction rate and the labor supply decisions of those individuals who may be induced to enter by such a reform. We

provide a detailed analysis of how recent policy reforms in the US and UK have changed the shape of the budget constraint facing many workers.

Any analysis of labor supply requires an understanding of the background changes in wages, participation and hours of work. In Section 3, we provide an analysis of labor supply facts highlighting the important changes in labor market participation and in the dispersion of wages. As a comparison with the US and the UK, we document the changes in these aspects of labor supply for two additional countries: Germany and Sweden. It is these changes in participation and working hours that labor supply models attempt to explain. The success of labor supply models will be judged largely in terms of their ability to explain and enhance our understanding of the changes in participation and hours.

Having motivated our analysis of labor supply with important policy questions and labor supply facts, our aim in the remaining sections is to present a comprehensive evaluation of alternative approaches to modeling labor supply. This seeks to achieve three broad objectives: to make different studies comparable by providing a unifying framework by which the results of each can be interpreted; to provide a description of the mechanics of implementing each approach and the data and assumptions required; and to identify gaps in our knowledge which can motivate future research. We have attempted to review the state of empirical knowledge on labor supply responses, and we end each section with a discussion of relevant empirical results.

The unifying framework we develop in Section 4 is designed to compare across alternative basic labor supply specifications. It should be noted at the outset that individual labor supply responses may be reflected in the choice of hours across firms rather than within any establishment. Complexities that arise from non-linear taxation, fixed costs, welfare programs, dynamics, etc. are taken up in detail in the following sections. A simple multi-period framework is used to compare across alternative static formulations, two-stage budgeting models, the Frisch model and fully-specified lifecycle models. The aim is not to dictate a single approach to estimation, but rather to evaluate precisely what can be learned from different datasets and different approaches to estimation. The wage coefficient in each approach is related to alternative elasticity measures and we ask which measure is appropriate for the evaluation of policy reforms. Even the simplest tax reform typically involves an unanticipated shift in the profile of wages. None of the standard elasticity measures fully reflects responses to such a shift and Section 4 precisely documents what is required to answer such policy questions.

Sections 5–8 consider alternative aspects and approaches to labor supply that have been adopted in the literature. We begin with a review of the application of difference in differences and natural experiments in labor supply estimation. Our aim here is to emphasize the structural assumptions underlying this approach and to relate the estimated parameters to those needed for policy analysis. A number of influential studies that have used this approach, and related approaches, are then reviewed.

Procedures by which a researcher can fully account for non-linear taxation, fixed costs, welfare participation and missing wages in estimation and simulation motivate the discus-

sion in Section 6. Again, the emphasis here is to lay out the precise assumptions and restrictions placed on behavior by alternative models. The practical issue of how to account for multiple program participation and the interactions between the tax and benefit system are highlighted. The empirical literature in this area is vast. This aspect of labor supply continues to attract considerable research interest, reflecting the recurring importance placed on the labor supply responses to tax and benefit reforms.

Placing the labor supply problem in a context where there is potentially more than one supplier of labor in the household is covered in Section 7, which reflects two important developments in this area. The first is to acknowledge the complex set of incentives faced by multiple workers once the full tax and welfare system is accounted for. The second is the introduction of alternative models of labor supply decision-making when multiple workers are located in the same family. These alternative models that seek to account for collective choices that are solutions to bargaining within the family are still in their infancy as far as empirical application are concerned. However, we are able to compare them to the standard “unitary” model and review the empirical literature that has developed to date.

Our review of alternative formulations of the labor supply model is completed in Section 8 with a discussion of dynamic models. Here we highlight generalizations of the basic multiperiod model described in Section 4 that allow for human capital and non-participation. The first-order conditions for the standard multiperiod model can be severely distorted in the presence of human capital choices. Human capital choices, or purely exogenous learning by doing, can break the separability of the intertemporal decision rule that allows simple Frisch and two-stage budgeting formulations. This is also shared by models that allow for habits. We describe the appropriate adaptation of the multiperiod model to cover these extensions and review the results from the empirical literature. We also consider the complications that arise in these models once non-participation and fixed costs are allowed for. We evaluate the trade-off between realism and computational tractability and set up the standard discrete dynamic programming formulation for this problem.

In Section 9, we conclude this chapter with a brief assessment of what has been achieved by recent research on labor supply and ask whether we are now in a better position to answer the policy-reform questions raised in Section 2 and better able to understand the labor supply facts described in Section 3. We document a large number of significant contributions across a wide range of labor supply issues but we also identify significant gaps in our knowledge which will continue to place research on labor supply at the forefront of research in labor economics for some time to come.

2. How have tax and welfare policies changed?

In few decades have we seen the marked changes in tax and welfare policies that have occurred since the early 1980s. In the US, the number of tax brackets sharply diminished with the passage of the federal tax reform in 1986. In the UK, the number and level of

higher brackets were reduced following the 1979 move away from direct taxation and towards indirect taxation. Sweden and other European countries subsequently followed this direction in reforming their income tax systems during the late 1980s and early 1990s. In both the US and the UK, in-work benefits increasingly became the main platform for encouraging low-income families to increase their work effort and incomes. In the US, the earned income tax credit (EITC) was greatly enhanced in 1993, while in the UK the Family Credit (FC) system, based on a minimum number of weekly hours worked, reduced the limit in 1992 from 24 to 16 h per week and significantly increased the number of recipients.

In 1996, the US adopted sweeping reforms in its welfare systems, all designed to induce recipients to support themselves through work. In the UK, the Family Credit system was extended to incorporate a 30-hour benefit supplement. In the 1998 budget, Family Credit was made more generous and was renamed Working Families Tax Credit (WFTC) to signify that payments would be paid through the tax system. The motivation of much research on labor supply is to predict the consequences of such reforms for hours of work and earnings. Researchers often devote considerable attention to modeling the institutional features of tax and transfer policies. This section briefly summarizes the changes that have occurred during the last decade in tax and welfare policies. We focus on policy changes for the US and the UK. The following sections explain how labor supply analyses have exploited these changes to assess their impacts on work behavior.

2.1. US tax and welfare programs

Perhaps the easiest way to convey the complexities introduced by the US tax and welfare system is to describe the number of programs in which individuals participate when they work. Workers must pay federal income taxes which account for an array of deductions, social security tax, state income tax and a variety of health and insurance taxes. If a worker's family has sufficiently low income, it may be eligible for benefits from a patchwork of different programs. These public assistance programs provide support in the form of cash income as well as in-kind support for necessities such as food, housing, medical care or home heating. The six major programs that offer the core of resource support for poor families in the US are: Aid to Families with Dependent Children (AFDC), Food Stamp Program (FSP), Supplemental Security Income (SSI), Housing Assistance, Medicaid, and the Earned Income Tax Credit (EITC). AFDC, SSI, and EITC pay cash assistance to low-income families. FSP provides food vouchers denominated in dollars to low-income households. Housing assistance programs come in two varieties: rent subsidies for occupancy of private dwellings, and low-income public housing which is built, managed and maintained by government agencies. Finally, Medicaid is an in-kind benefits program providing medical assistance to poor persons.

Describing how all of these programs have changed individually during the past decade would occupy many papers, yet this exercise would still fall short of characterizing how these policy alterations influence labor supply, as the most profound and

disconcerting effects occur when families simultaneously participate in multiple programs. Each program has its own benefit reduction rate which determines how much benefits decline as earnings increase. These rates act as tax rates on earnings, in that they dictate how much families get to keep out of any incremental earnings they receive while collecting benefits. Because benefit reduction rates are independent across programs, the combined benefit reduction rate that results when a family participates in several programs rises to staggeringly high levels that no policymaker ever intended. This, in turn, produces significant disincentives for families to work. The relevant factor in assessing the impact of these policies on labor supply is the combined effect of these programs through time.

2.1.1. How do programs in the US combine to tax earnings and provide income support?

Fig. 1 shows how net governmental transfers change as a family's earnings rise, given participation in various combinations of public assistance programs. The figure depicts three scenarios: the lower curve indicates transfers when the family receives benefits from just EITC; the middle curve gives the total benefits received when the family collects food stamps in conjunction with EITC; and the upper curve measures the total transfers when the family participates in the AFDC program as well. The curves are for a single-parent family with two children living in California – only the AFDC benefit schedule depends on California residency. Other than the social security tax (about 7.5%), families at the low income level pay no federal or state income taxes. As earnings increase (i.e., moving left to right in the figure), net transfers initially rise due to the increase in EITC, regardless of the

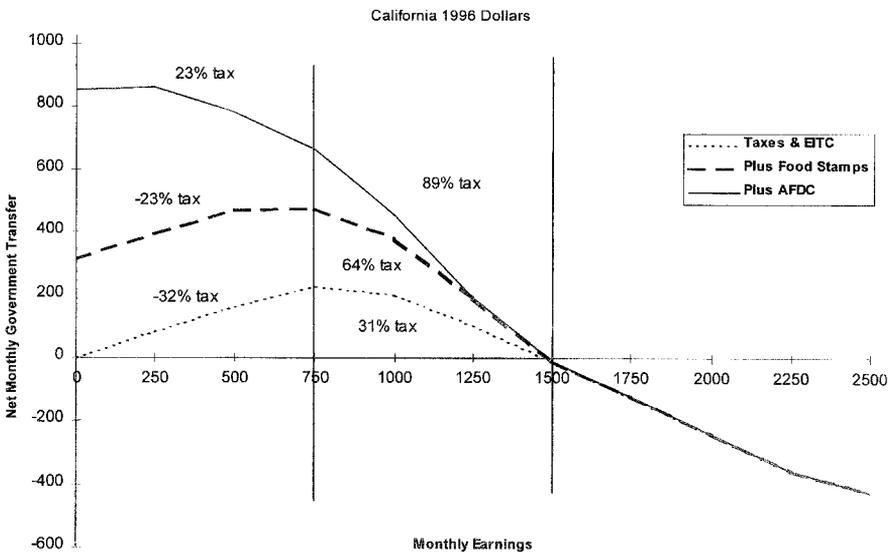


Fig. 1. Net transfers/taxes for California in 1996.

combination of programs in which the family participates. However, eventually these transfers decline with higher levels of earnings. The reversal is fastest when the family collects AFDC, food stamps and EITC simultaneously, and slowest when collecting only EITC.

For a family participating in all three programs, the uncoordinated nature of the programs leads to some unintended and undesirable features. As the family's earnings rise within the first \$750/month earned (=30 h per week at \$5.75) in 1996, the EITC provides a tax credit increasing the value of work by 40%. If this were the only program, the family would face an implicit tax rate of -32% (a negative tax), paying only social security taxes. However, since both food stamps and AFDC benefits decline more rapidly with earnings than EITC rises, a family who also participates in these programs ends up losing about 23 cents out of every \$1 earned up to \$750/month. This translates into an effective positive tax rate of 23% on earnings. Earning \$750/month, this family still receives benefits from all three programs. Increasing family earnings from \$750 to \$1500/month would put it in an income range with effective tax rates of about 89%, meaning that it would retain only 11 cents out of every dollar earned.

Ironically, this high tax rate is the result of changes during the past five years that were designed to increase work incentives. Recent federal legislation increased the generosity of the EITC, and at about the same time California lowered the benefit reduction rates through the passage of "30 and a third" reforms in AFDC. Comparing the benefit structure and tax rates in 1996 to those in 1992 reveals that these federal and California state changes decreased the effective taxes for families in the lowest earnings range. The marginal tax rate for the first \$750 of earnings fell from 71% to 23%. However, these changes simultaneously raised the marginal tax rate for the second \$750 of earnings from 59% to 89%.

Knowing that AFDC participants do not work extensively under the current system says little about their motivation or prospects for working, because the existing benefit structure creates strong disincentives to working. It is quite rational for AFDC recipients to work little or not at all. The current rules tax income highly as earnings increase. These work disincentives become more severe the more a recipient works and the closer he or she gets to self-sufficiency.

Under the system today, an AFDC recipient would need to work 40 h per week at \$6.90/h to make enough to leave AFDC (= \$1104/month). She would need to earn \$7.88/h to lose food stamps as well (= \$1261/month). Yet in moving from \$750/month to \$1500/month, her net income would rise by only \$82 due to a combination of benefit reductions in both AFDC and food stamps and a reduction in the EITC as earnings enter a "phase out range." Unfortunately, the resulting 89% tax rate falls precisely on the earnings range that makes the difference between welfare receipt and self-sufficiency.

2.1.2. How do programs differ across states?

Fig. 2 illustrates how differing AFDC programs across states affect the benefit amounts

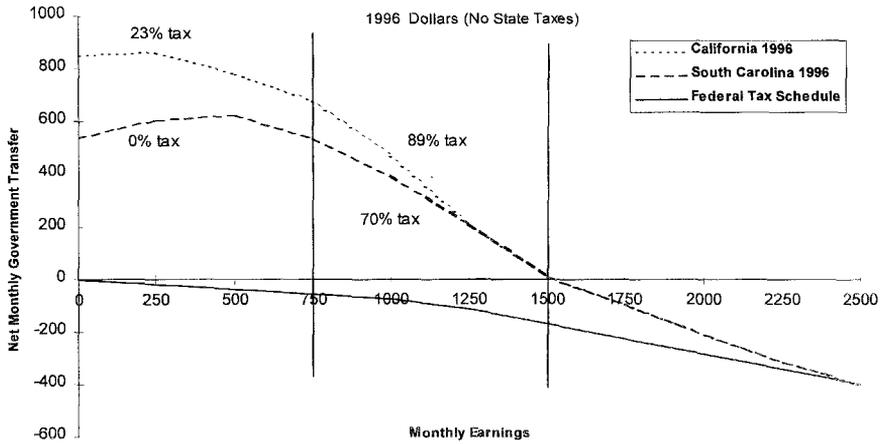


Fig. 2. Comparison of net transfers/taxes for California and South Carolina in 1996.

received by a family participating in all three programs. The top curve is for our California family and the middle curve is for an identical family living in South Carolina. We select South Carolina as the comparison state for California because, in the early 1990s, it occupied an opposite position in the distribution of state AFDC benefit levels: whereas California had the fifth most generous state AFDC program, South Carolina had the fifth least generous. The lowest curve corresponds to the taxes a family would pay if it participated in no low-income transfer programs.

Since South Carolina paid lower AFDC benefits than those in California, the net transfers received by the South Carolina family are everywhere below those of the California family until monthly earnings reach between \$1250 and \$1500 when both AFDC programs cease to pay benefits. The higher generosity of California's program has a serious downside: California's implicit tax rates on earnings are much higher. The benefit reduction rates are similar, but more is lost for every dollar earned in California because the reduction rate applies to a larger benefit amount.

Reduction rates are still quite large for South Carolina residents. Even though the tax rate faced by a South Carolina family increasing its earnings from \$750 to \$1500/month is almost 20 percentage points below the rate faced by a California family with the same earnings increase, this lower rate is still 70%. Such tax rates are staggeringly high and are very likely to discourage work.

2.1.3. How have programs changed in the US?

Fig. 3 shows how net governmental transfers have changed in California during 1985-1996. Similar changes have occurred in other states. These changes reflect a combination of factors, the most prominent being decreased benefit reduction rates for welfare programs and increased generosity in the EITC.

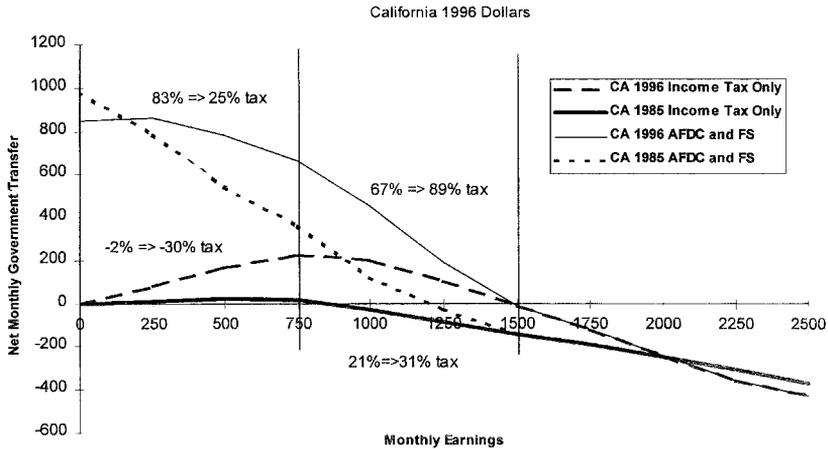


Fig. 3. Net transfers/taxes for California in 1996 versus 1985.

The two lines that begin lowest on the graph represent the net benefit receipt from EITC, combined with taxes, alone. The difference between 1985 and 1996 is striking. In 1985, the peak EITC benefit was a mere \$20, while in 1996 this figure had increased more than tenfold to \$224. As a result, the average tax rate on the first \$750 for a family receiving just EITC fell from -2% in 1985 to -30% in 1996. This reduction in tax rates leads many to argue that EITC has a strong pro-work effect. However, after \$750 EITC benefits decline, yielding a tax rate of 31% in 1996 versus 21% in 1985. Hence, the increased generosity of EITC has led to increased marginal tax rates for families seeking to increase their income from welfare-dependent levels to more self-sufficient levels, as noted above.

When combined with changes in AFDC and Food Stamps, this shift in incentives is even more pronounced. The top two lines in Fig. 3 depict total AFDC, Food Stamp and EITC benefit levels in 1985 and 1996. From 1985 to 1996, the monthly AFDC and Food Stamp benefit for a family with no earnings was reduced from \$980 to \$852. Associated with this reduction was a flattening of the benefit versus earnings graph, as shown in the figure. As a result, the combined tax rate fell from 83% to 25% on the first \$750 in earnings during this period. However, a flatter benefit reduction schedule for the first \$750/month simply required a steeper schedule for the next \$750. As a result, the average tax rate for the second \$750 in income/month – as noted, that income required to move off welfare – rose from 67% to a staggering 89% from 1985 to 1996.

In summary, changes from 1985 to 1996, which were advanced as measures to increase work incentives created markedly higher tax rates on income between \$750 and \$1500/month. This range of income is very important for families seeking to move off welfare. While it may be possible to enact further reforms that push this region of high benefit reductions and marginal tax rates to higher incomes, this can only be done at the cost of substantially reduced benefit levels or substantially increased program costs. This is the

fundamental policy dilemma facing those seeking to change tax and transfer policies. Before undertaking this effort, it is important to understand the exact nature of the labor supply changes induced by changing tax and transfer policy.

2.2. UK tax and welfare programs

There are effectively four important components of the British direct tax and welfare system as it affects labor supply. The first is the individual tax allowance on earned income, below which no direct taxes are paid. Couples in the UK are taxed independently and the tax allowance is also individually based. In 1996, it was £3650 per year (almost \$6000) and was sufficiently large to exempt from direct taxes many part-time low wage workers, especially married women. Approximately 36% of working women married to employed men had earnings below this limit. The majority of workers with earned income above this limit pay direct taxes at a flat basic rate, which has fallen from around 33% to 24% in the 15 years to 1996.¹

The second component is the National Insurance system which acts like a tax on earnings between a lower and an upper limit. This is also individually based, adds between 2 and 9 percentage points to the basic tax rate and is paid in full once earnings rise above the lower limit. Therefore, unlike the basic tax rate, the NI premium is payable on all earnings. Moreover, as NI payments stop at approximately the level of the higher tax rate, the overall tax rate through the direct tax system rarely exceeds 40%. Third is the “in-work” benefit Family Credit described in Section 2.2.1 (reformed and renamed Working Families Tax Credit in the 1998 Budget). The last of the four components is the multitude of largely means-tested income assistance programs that cover unemployment insurance and housing benefits; child support is a flat-rate non-means tested benefit examined in more detail below. Although the welfare system is designed to acknowledge interdependencies in benefit reduction rates so that no effective tax rate exceeds 100%, combining the tax system with the welfare system implies some severe disincentives for work, especially for low-wage families. This motivated the introduction of an “in-work” tax credit.

It is also worth noting that, over this period, the rate of Value Added Tax, paid on all goods except food and children’s clothing has risen from 9% in 1979 to 17.5%.²

2.2.1. An hours-based “in-work” benefit

An important component of the British tax and welfare system is the “in-work” benefit program called Family Credit (FC). Introduced in 1988 as an extension to Family Income Supplement, it has many features in common with the EITC program in the US. However, eligibility is based on a minimum weekly working hours requirement. The new Working Families Tax Credit, which replaces FC in October 1999, has exactly the same minimum weekly hours requirement. In this respect, the British in-work benefit system has simi-

¹ In 1993, a lower band of 20% was introduced on a relatively small initial part of taxable income. Less than 15% of workers pay tax at the higher 40% rate.

² Domestic energy was exempt from VAT but now attracts a reduced rate of 8%.

rities to the Canadian Self-Sufficiency Program (see Card and Robins, 1996). However, it should be pointed out that the SSP, which is only currently operating on an experimental basis, is time-limited and only available to parents with long durations of welfare receipt and unemployment. The FC system was designed to encourage part-time work and to support the income of part-time working parents. It has subsequently been extended with a small supplement for full-time work.

The basic FC scheme is generous but has a high withdrawal or benefit reduction rate. Family Credit becomes payable to individuals in families with children if their working hours exceed 16 per week and their overall income falls below some specified level, which varies with the number and age of children. The credit is then progressively withdrawn at a reduction rate of 70% as income rises (to be lowered to 55% in the WFTC reform). This rate is considerably higher than that for the EITC in the US.

Since the introduction of FC in 1988, the structure of the in-work benefit system has witnessed three major reforms: a reduction in the hours rule from 24 to 16 in 1992, the introduction of childcare disregards to help recipients with child-care costs in 1994, and the introduction of an additional credit at 30 h. During this period, the number of recipients doubled to well over 500,000. The Working Families Tax Credit reform only marginally changes the structure with a more generous level of payment and a lower benefit reduction rate of 55%. Consequently, more individuals in work who would not have received FC will now receive WFTC.

For most low-income individuals, working less than 16 h per week, the income support and housing benefit system renders the budget constraint virtually flat, so that FC can act as an important jump in the in-work income for low-wage working parents. The high benefit reduction rate, however, implies a reasonably flat constraint above 16 h, providing a potentially strong incentive for those working more to reduce their hours. Consequently, questions similar to those of the EITC arise as to the effectiveness of the system.

Since 1980 there has been no earnings-related unemployment insurance in the UK. Benefits for the unemployed, called job-seekers' allowance (JSA), are flat-rate at a level similar to the level of basic Income Support. This is worth about 20% of median full-time male net weekly earnings and is withdrawn at a rate of 100% against earnings provided weekly hours of work are fewer than 16. At higher hours, no income support is available. However, a child benefit of approximately £10 per week per child is payable to all families regardless of income. Consequently, for childless workers with low housing costs, income out of work is relatively low. For families with children, in particular for lone parents, this is not the case.

Fig. 4 shows the implied net government transfers for a single parent earning £4/h with two pre-school children in the UK. The FC at 16 h produces a large jump in net income. The additional supplement at 30 h is also evident. Fig. 5 displays the same budget constraint but in terms of weekly hours of work. This highlights the minimum hours requirement in the British in-work credit system. We assume a rent level of £50 per week. Housing Benefit (HB) is paid to all individuals with a sufficiently low income, and covers all rent whether the individual is in private or public rental housing. Once

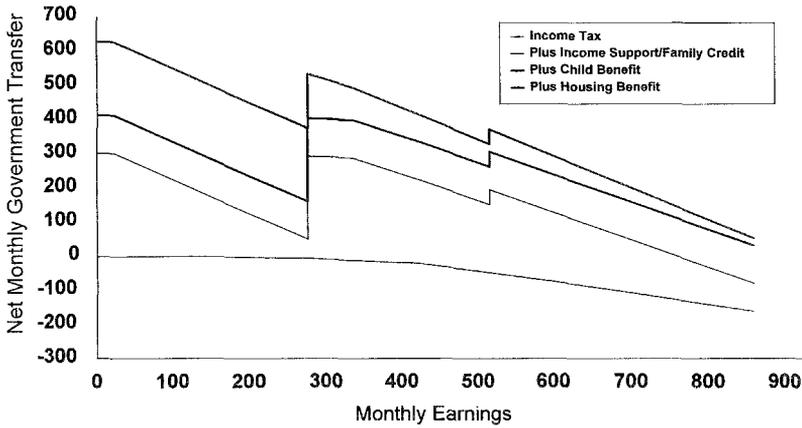


Fig. 4. UK net transfers 1996–1997: lone parent.

income reaches a ceiling, the benefit is withdrawn at a rate of 65%. This is further enhanced since the 65% withdrawal is made after income tax, NI and local taxes have been paid. Income support and other benefits, such as one-parent benefits, can be seen to fall in line with the increase in earned income up to 16 h per week. After that point, FC enters. National Insurance payments also become important and the budget constraint is further flattened by the high benefit reduction rate for FC. The total disposable income line in Fig. 5 shows the combined impact of the UK tax and benefit system.

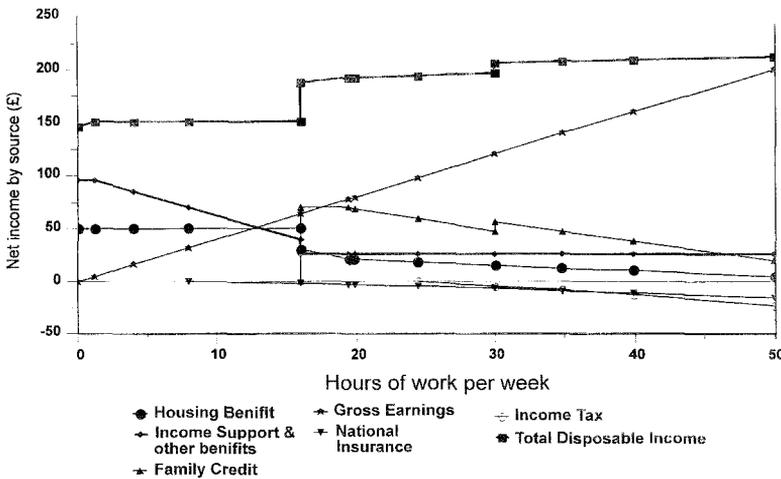


Fig. 5. UK budget constraint, 1996–1997.

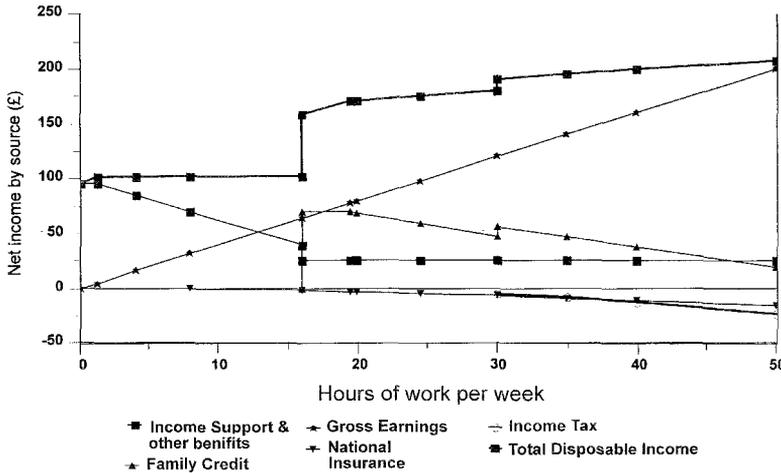


Fig. 6. The impact of removing housing benefit.

2.2.2. Interactions among British programs: family credit, housing benefit and income support

As we noted above for the US, for the purposes of analyzing labor supply it is important to recognize the interactions between benefits and in-work credits. These are critically important for low-wage families and raise similar issues to those discussed in the context of the EITC program. In Fig. 6, we show the impact of excluding Housing Benefit (HB). Since FC is treated as income in determining eligibility for HB, the impact of FC is considerably reduced by the HB program. A comparison of Figs. 5 and 6 shows there is now a much larger increase in net income at 16 h.

Since 1986, Family Credit, Housing Benefit and Income Support have all become important components of the British welfare system. This is revealed in a comparison with Fig. 7 which shows the net government transfers for the same lone parent facing the 1986–1987 welfare and tax system. Lower housing benefits (a mean of £35 rather than £50 per week in 1996 prices) in 1986 reflect the lower level of social rents in public sector housing which, paradoxically, reduced the incentive problem facing low-wage workers. It is probably the rise in public housing rents together with the decline in the relative real wages of low-skilled workers that most significantly changed the balance between work and unemployment for low-wage families in the UK.

3. Recent empirical trends

In conjunction with the large changes in tax and benefit policies detailed in Section 2, the 1980s and early 1990s have seen dramatic changes in participation, hours of work and

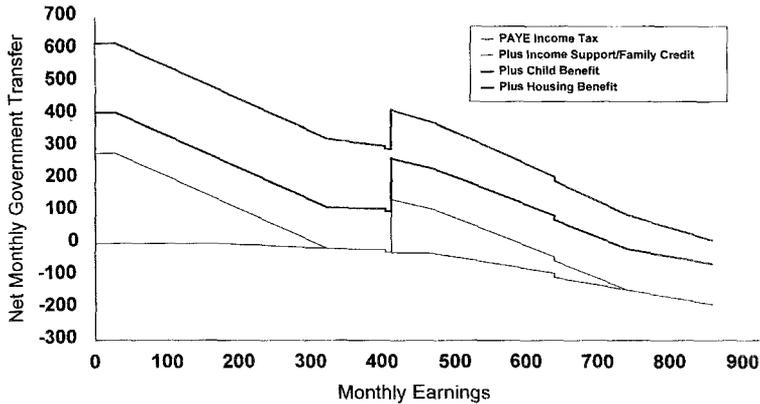


Fig. 7. UK net transfers 1986–1987: lone parent.

hourly wages. In this section, we provide a brief documentary of these changes, drawing on evidence from the US, the UK, Germany and Sweden.

It is the changes in participation and working hours that labor supply models attempt to explain. The success of these models must therefore be judged according to their ability to explain and enhance our understanding of the changes in participation and hours. Moreover, movements in the structure of real wages, in addition to reforms of the tax and benefit system, provide the variation needed to explain these changes. To the extent that trend differences in real wages, government transfers and marginal tax rates across groups can be argued to be exogenous to changes in preferences for labor supply, they provide the most convincing data, outside social experiments, for recovering reliable estimates of labor supply responses. This explains the central role we place on these empirical regularities in this survey.

The changes in participation, hours and real wages have varied widely across economic and demographic groups. For example, higher-educated workers in the UK and US have seen strong growth in real wages, while less-educated workers have experienced stagnant or falling real wages. In contrast, the real wages in all education groups in Germany appear to have risen steadily during this period. There have also been strong differences in labor market attachment across age groups. An increase in the overall participation of women has been matched by a drop off in the participation of males, particularly pronounced among older men in Europe.

In this section, we first discuss changes in participation. These are analyzed by education level for men and women separately. The contrast by education group is striking, as are the differences in the trend changes between men and women. Next, we move to an analysis of hours of work to highlight the changes in the average weekly and annual hours worked by different education and gender groups since the end of the 1970s. Finally, we consider changes in gross hourly wage rates. The detailed changes in these are documen-

ted elsewhere in this Handbook (see the chapter by Katz). However, our aim is to focus on contrasts by education and gender and to evaluate differences in these between the US, the UK, Germany and Sweden (Figs. 8–14).

3.1. Data sources

We draw from a variety of country-specific data sources. Our samples contain men and

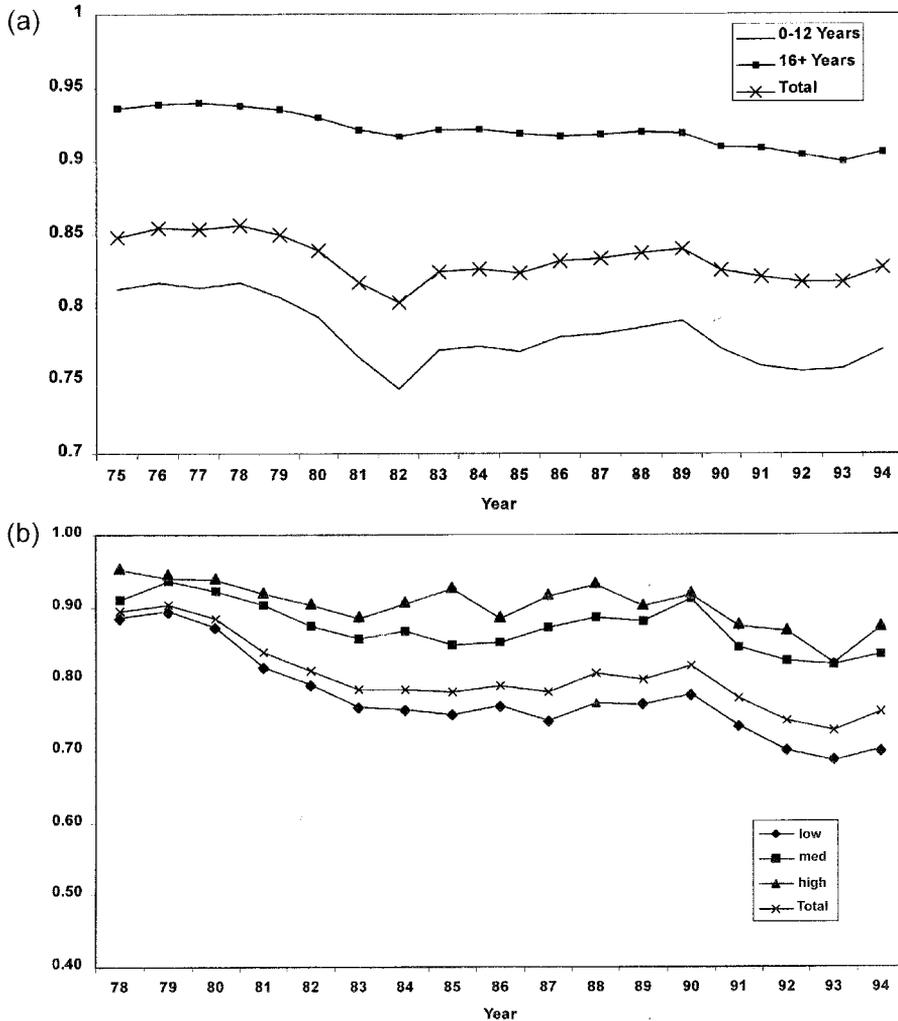


Fig. 8. Men's employment to population ratio by education: (a) US; (b) UK; (c) Germany; (d) Sweden.

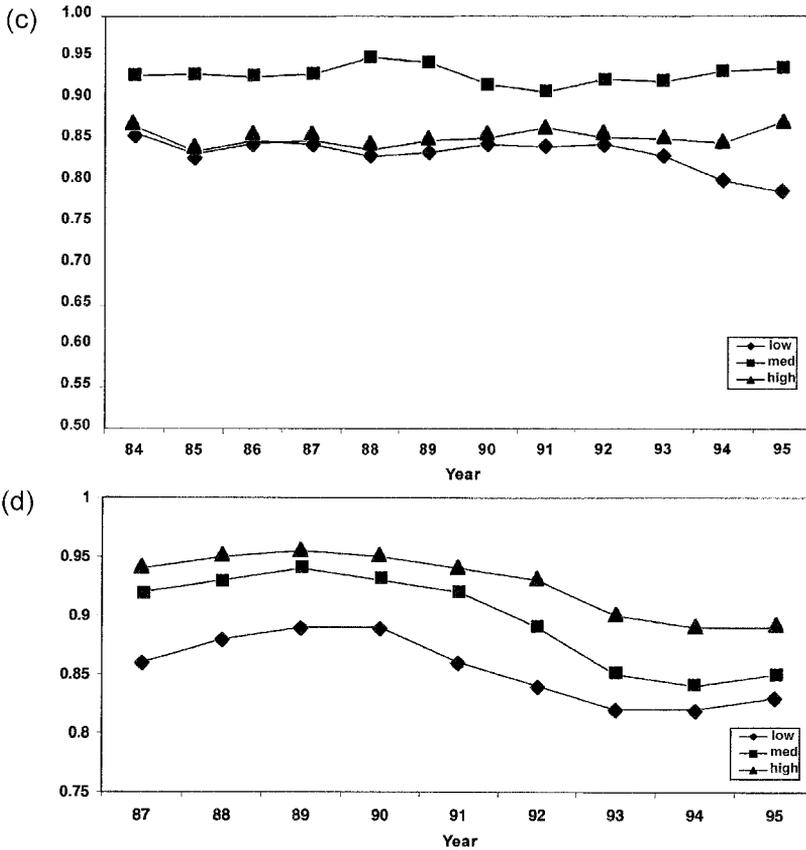


Fig. 8 (continued).

women above minimum school leaving age and below the standard retirement age. Participation is defined as the proportion in employment out of all individuals of working age in a specific group. For the US, the primary source of data is the Current Population Survey, a monthly survey of approximately 60,000 households. A group of CPS interviewees stays in the sample for 4 months, is out of the sample for the next 8 months, and then returns to the sample for the following 4 months. We consider data from 1975 to 1994, for men aged 26–64 and for women in multiple-year birth cohorts ranging from 1920–1926 to 1950–1964. For the UK it is the Family Expenditure Survey (FES), a repeated cross section. Each FES survey consists of around 7000 households. All individuals aged between 18 and 59 years of age are used except those in full-time education, self-employment or the armed services. The “low” education group includes those that left formal schooling at the minimum school leaving age (currently age 16). The “med” education group includes those in schooling until age 18. The “high” group includes those with college education.

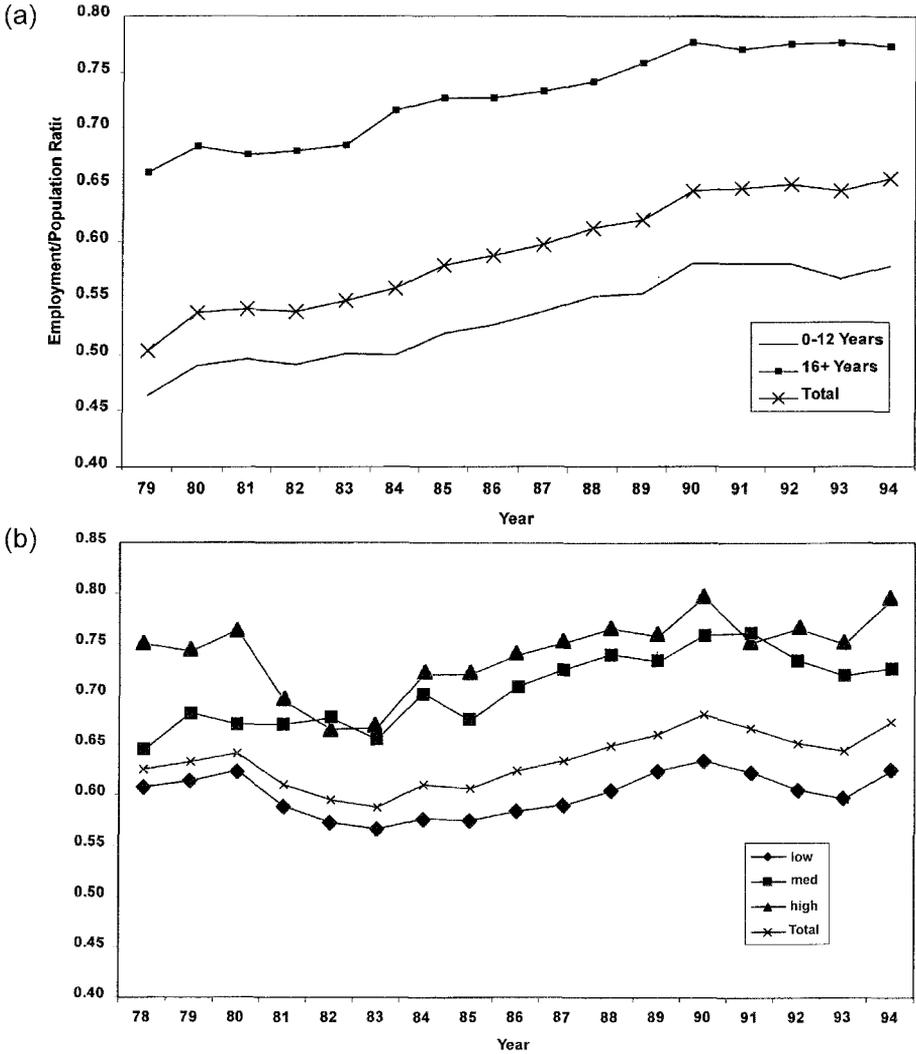


Fig. 9. Women's employment to population ratio by education: (a) US; (b) UK; (c) Germany; (d) Sweden.

For Germany, a similar selection criterion is used and we draw on individuals from the first 12 waves (1984–1995) of the German Socio-Economic Panel (GSOEP). The design of the GSOEP is similar to that of the US Panel Study of Income Dynamics (PSID), see Wagner et al. (1993). All figures reported below refer to individuals located in the geographic area of the former West Germany. The precise details of the data construc-

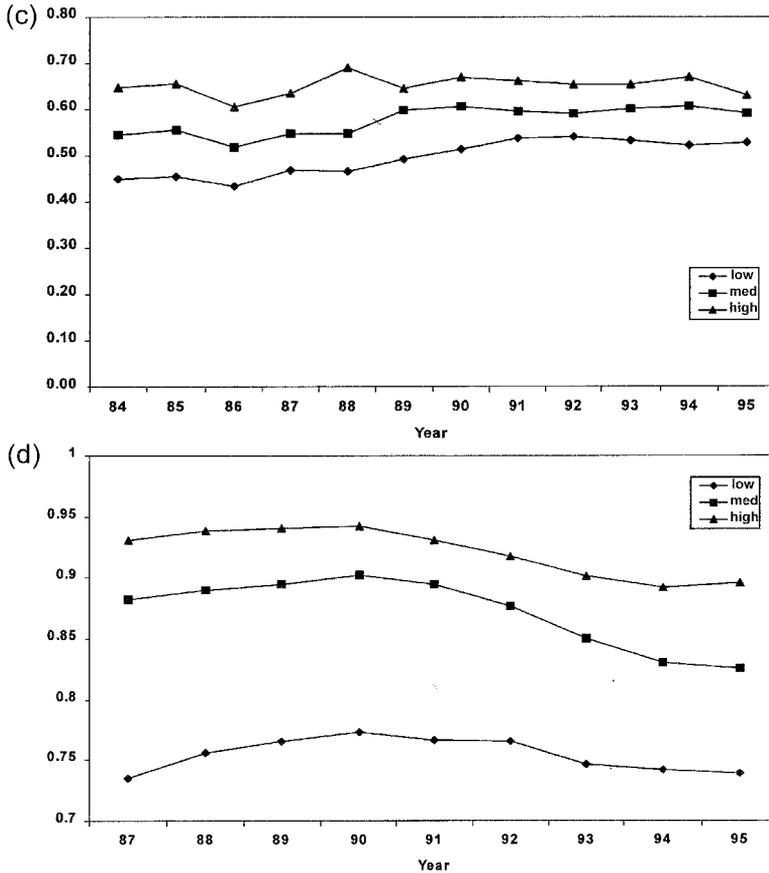


Fig. 9 (continued).

tion follow the work of Dustmann and Van Soest (1997). For Sweden, three different data sources have been used. An income survey (HINK) and the Swedish Labor Force Survey (AKU), both from Statistics Sweden. This is supplemented with data from the Swedish survey, Market and Non-Market Activities (HUS) (see Flood et al., 1997 for details).

3.2. Participation

Participation in work has seen some important changes since the late 1970s. Fig. 8a provides the evidence for US men by education level (here measured by years of schooling). The cyclical nature of participation for the lower education group and the much lower participation rates stand out clearly in the data. If anything, there is a slight downward

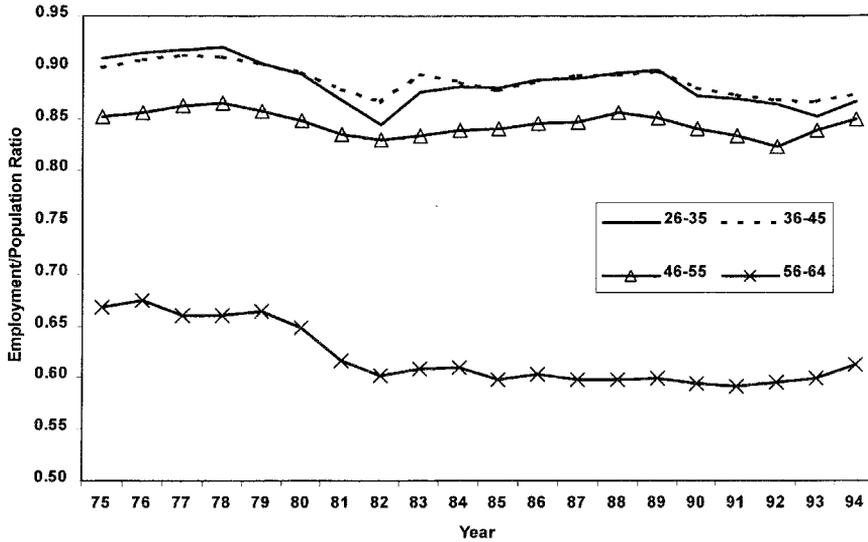


Fig. 10. Men's employment to population ratio by age: US.

trend in participation using this employment-to-population ratio definition. Notice how this differs for women, where Fig. 9a shows that both education groups saw a strong increase in participation until the early 1990s.

This picture for male and female employment in the US has many features in common with the experience in the UK, although, as Fig. 8b shows, male participation has fallen dramatically in the UK since the end of the 1970s. Notice that, even at the top of the boom in 1990, participation did not return to its 1979 levels. This is in contrast to the pattern for women, where Fig. 9b reveals that the participation rate approached 70% in the 1990 boom. In Germany, participation throughout the late 1980s was much more stable than it was in either the US or the UK. Fig. 8c shows that the fall in participation among lower-educated men in Germany only set in after 1992. Remember that these data refer to the West German region both before and after reunification. For German women (Fig. 9c), participation has been slowly rising for all groups until 1992. Finally, in Sweden, we only have a consistent split by education available on an annual basis after 1987. However, until that point, participation rates rose steadily for women and stayed fairly flat for men. The onset of the 1991 recession in Sweden is clear from Figs. 8d and 9d.

The decline in participation for men, which has been experienced to some degree in all countries, is particularly reflected in the working behavior of older age groups. For example, Fig. 10 shows a strong fall in the US employment-to-population ratio for men in the 56–64 year old age group. This declining attachment to the labor market by older men is mirrored in the UK and Germany (see Blundell and Johnson, 1998; Borsh-Supan and Schnabel, 1998). Interestingly, for the UK and Germany, it is the younger birth cohorts

as they age that are seeing larger declines. For older women, this picture is attenuated by the steady rise in participation across time and across birth cohorts. In the US and the UK, there has been an increase in participation for younger birth cohorts of women and, consequently, at the same age, younger cohorts of women have higher participation rates.

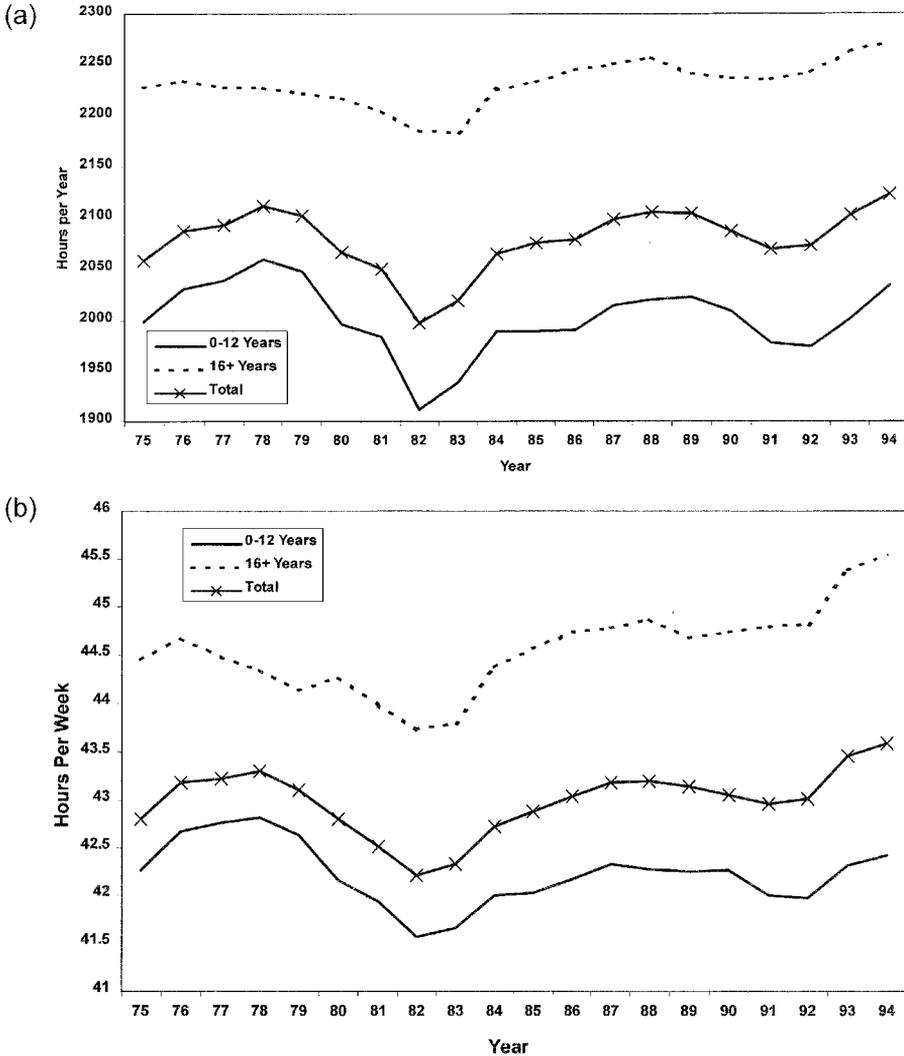


Fig. 11. (a) Men's annual hours worked by education level: US. (b) Men's weekly hours worked by education level: US. (c) Men's weekly hours worked by education level: UK. (d) Men's weekly hours worked by education level: Germany.

3.3. Hours of work

Annual hours of work for men in the US display a strong cyclical pattern and, especially during the last decade, an increasing trend. A similar story is true for weekly hours. These two measures of working hours in the US are presented in Fig. 11a,b. For the UK, the nature of our survey data means that we can only present weekly hours (that include

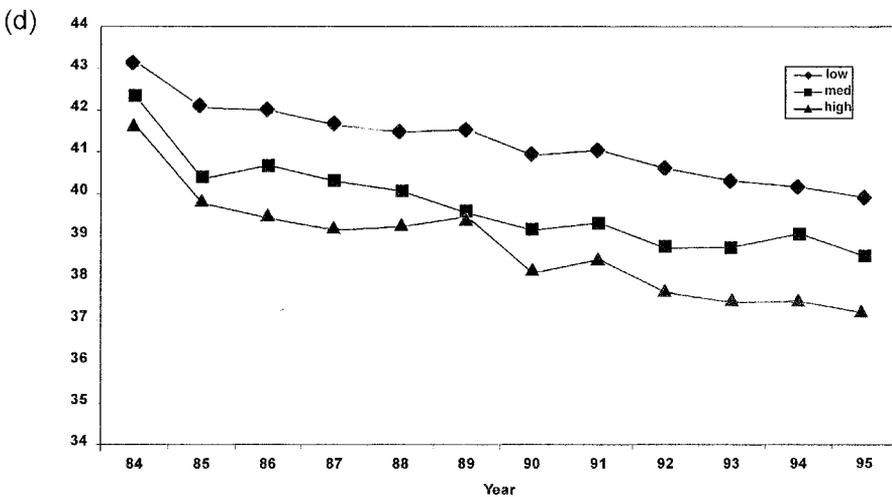
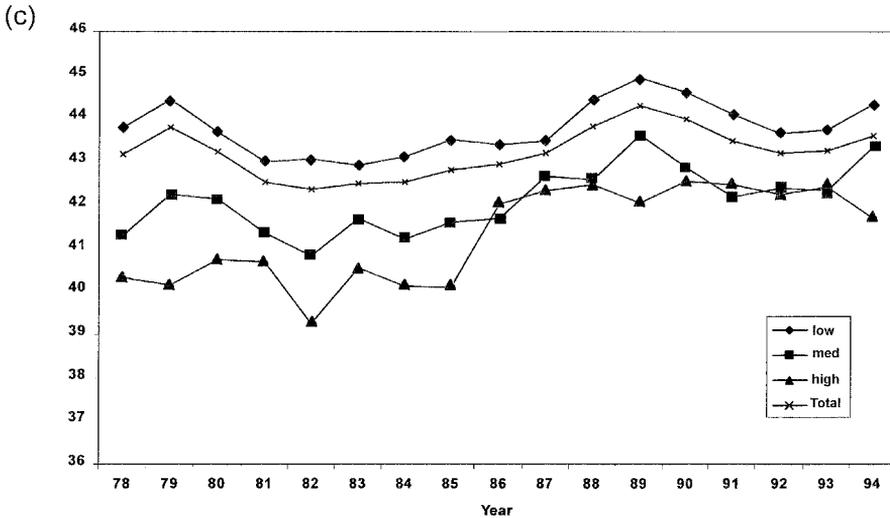


Fig. 11 (continued).

normal overtime hours). Fig. 11c shows that this measure of hours worked reveals a similar strong cycle and trend increase although, in contrast to the US, it is the higher-educated group in the UK that has tended to work fewer weekly hours on average. What is notable in both of these countries is that the trend increase in weekly hours is more accentuated for the higher-education group. Interestingly, as we shall see below, this is precisely the group that has seen a trend rise in real wages.

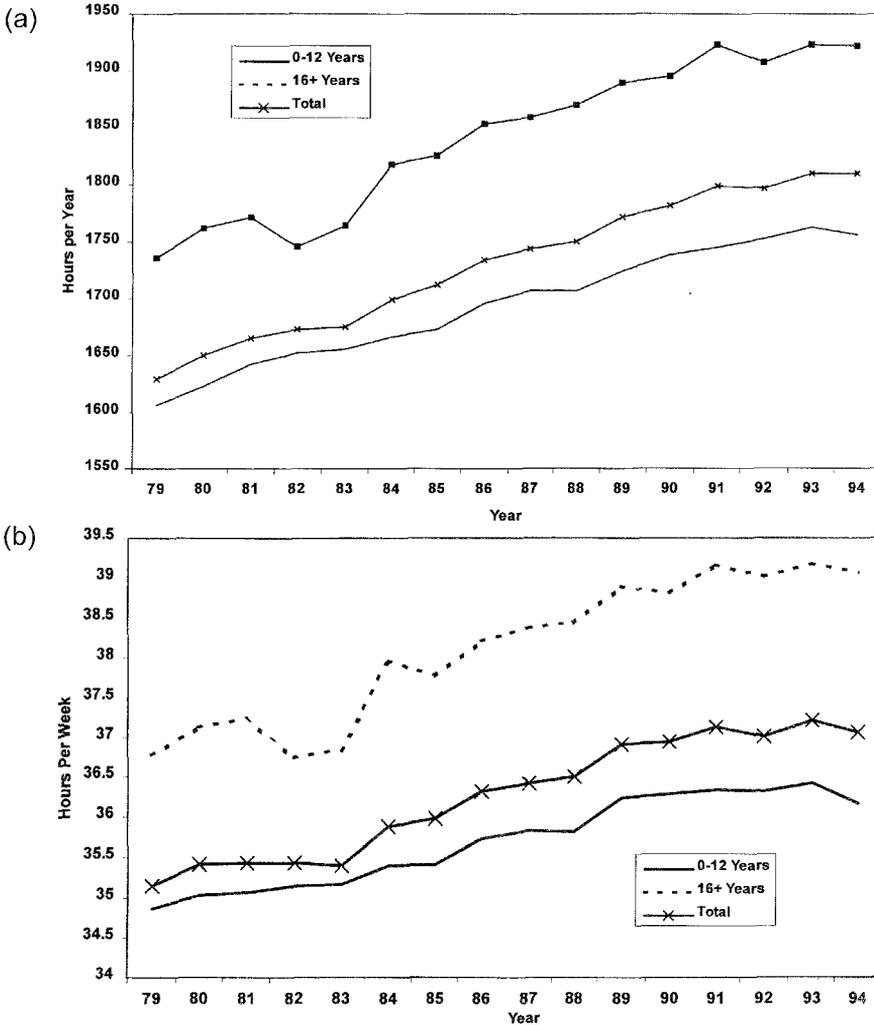


Fig. 12. (a) Women's annual hours worked by education level: US. (b) Women's weekly hours worked by education level: US. (c) Women's weekly hours worked by education level: UK. (d) Women's weekly hours worked by education level: Germany.

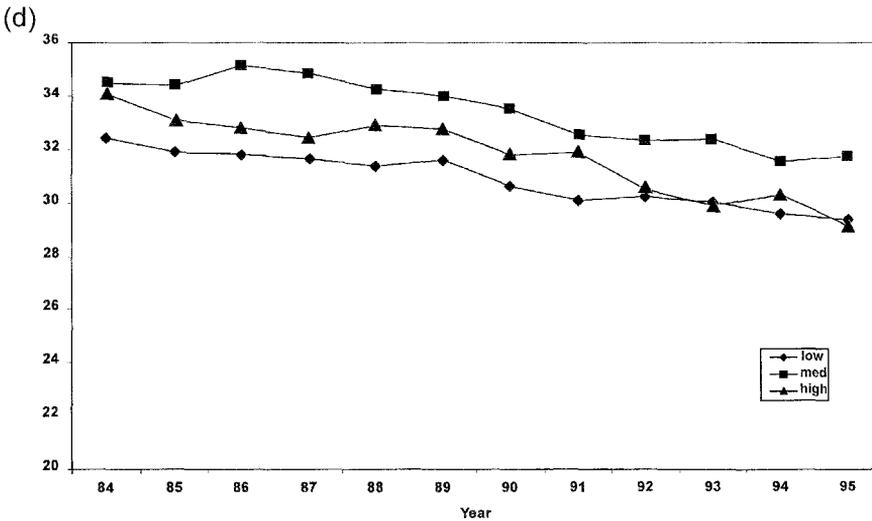
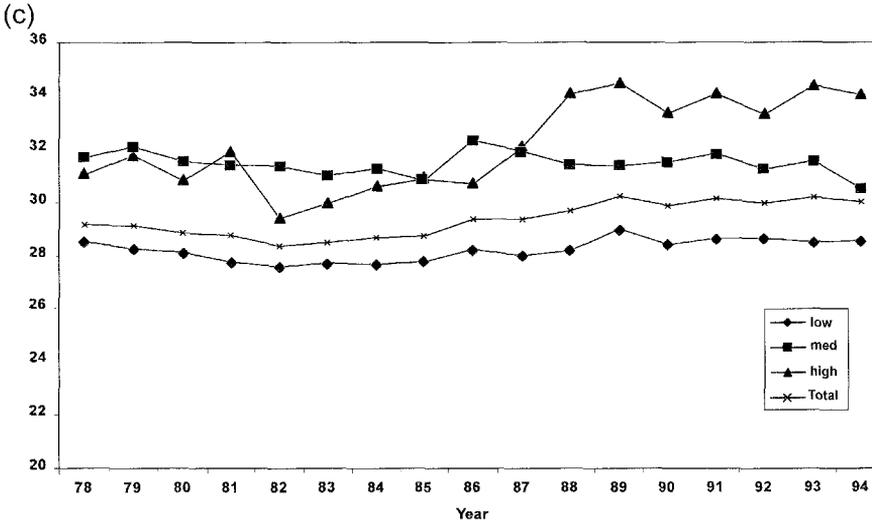


Fig. 12 (continued).

Fig. 12a-c shows that a similar story is true for the weekly working hours of women in both the US and the UK, although it is the higher-educated group that works longer weekly hours in the UK. Fig. 12b shows that, if anything, this gap has grown during the recent past. Annual hours have shown a strong trend increase in the US, as seen in Fig. 12a. In the UK this is probably less pronounced, at least for weekly hours of work. None the less, the

UK has seen a steady rise in women's weekly hours since the early 1980s when the cyclical downturn in 1980 and 1981 had a depressing effect on female and male hours of work alike. Although not reported here, working hours in Sweden for employed males have been quite stable despite a major tax reform in 1991. After 1993 there is a small increase for highly educated workers. For females there has been an upward trend in hours. This is especially pronounced for the highly educated. Working hours in Germany have seen a slow and smooth decline, as evidenced in Figs. 11d and 12d.

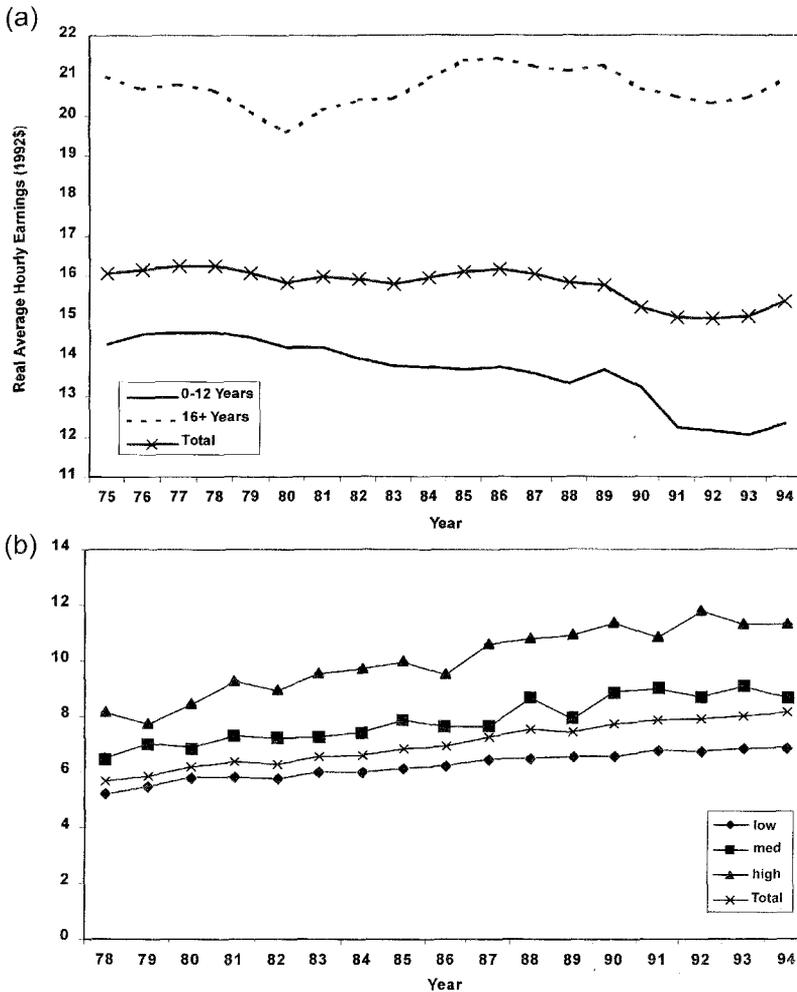


Fig. 13. (a) Men's real average hourly earnings by education level: US. (b) Men's real average hourly earnings by education level: UK. (c) Men's real average hourly earnings by education level: Germany.

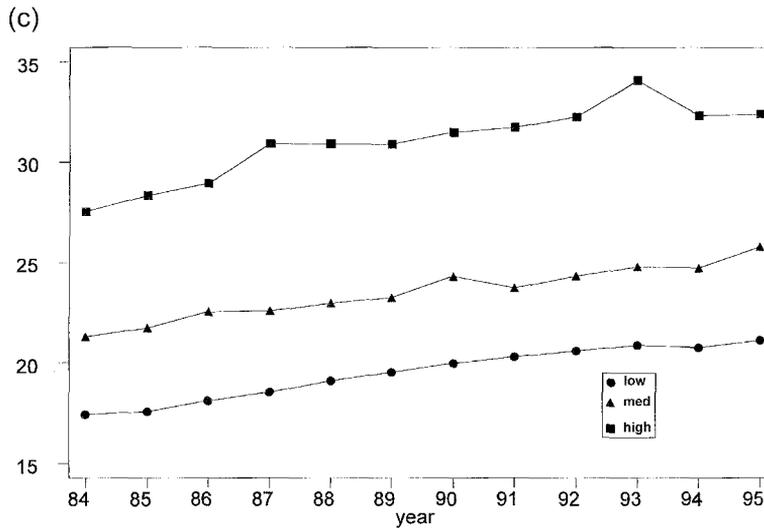


Fig. 13. (continued)

3.4. Real wages

The contrasts among the US, the UK, and continental Europe are probably most stark when it comes to a comparison of the level and growth of real wages. This is especially the case when split by education level. However, there are serious pitfalls in the interpretation of raw wage trends. First, there is a considerable change in composition across time both in terms of the total group of employees and in terms of the different education groups, and these composition changes are very different across countries. Second, there is the dubious comparability of definitions of education levels across countries.

The first issue is really at the heart of labor supply analysis itself, since it relates to the changing composition of those in work over time. For example, if lower real wages at the bottom of a cycle mean fewer lower-ability workers supplying labor at that time, then this systematically biases upwards the real wage at the bottom of the cycle. Similarly, if the increasing levels of non-participation by older men reflect a higher proportion of lower-ability workers leaving employment due to a relatively generous social security and benefit system, then this results in an upward bias in measured real wages and in measured returns to experience for low educated workers. This biases upward the trend increase in real wages for lower-educated workers and biases downward the apparent return to education. Any comparison of the growth of real wages and returns to education between the US and European economies must therefore acknowledge the impact of differential changes in composition on real wages.

With these points in mind, turning first to the US, Fig. 13a for men tells a dramatic story. For the lower-education group, real wages have fallen almost relentlessly since the late 1970s. Consequently, the education differential has widened significantly. As Fig. 14a shows, this is less clear-cut for women but, given the rise in participation for the lower-education group of women, a comparison over time may be less interpretable. None the

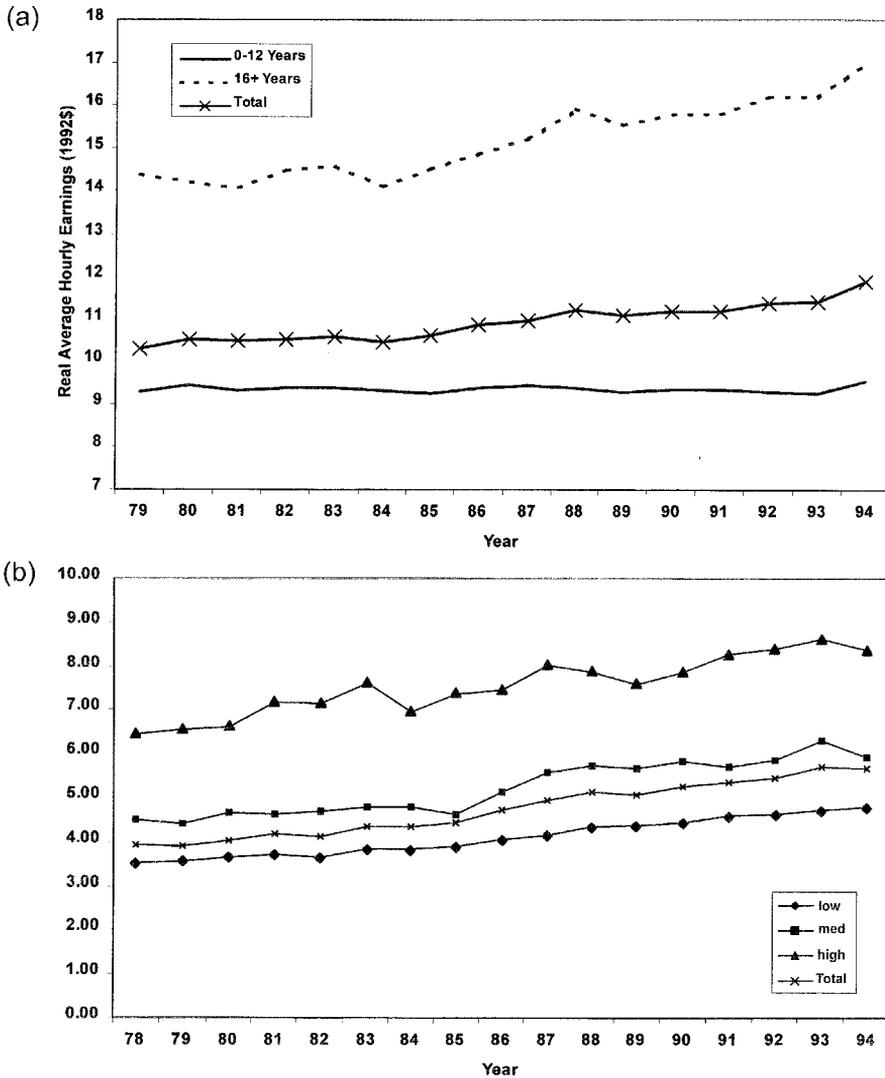


Fig. 14. (a) Women's real average hourly earnings by education level: US. (b) Women's real average hourly earnings by education level: UK. (c) Women's real average hourly earnings by education level: Germany.

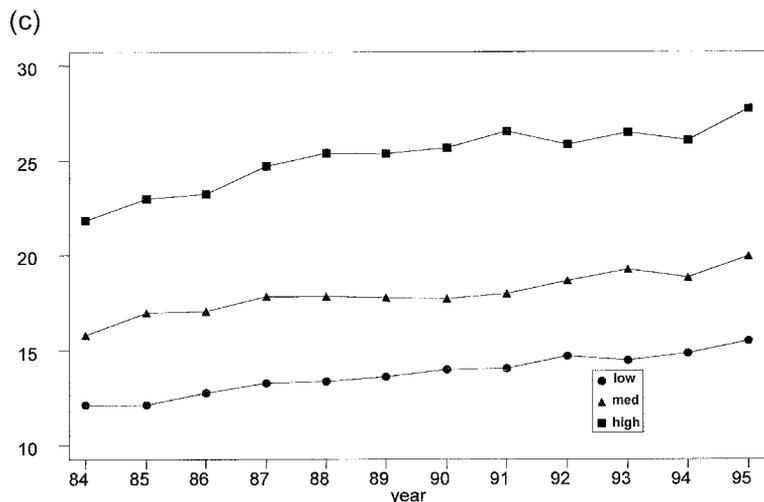


Fig. 14. (continued)

less, the increase in the differential is clear and the rise in the real earnings of the higher-educated women is quite spectacular, with a consequent fall in the raw gender differentials. For the UK, Fig. 13b shows an increase in the educational differential for men but, in contrast to the US, no fall in real wages for the lower education group. Although this lower-education group refers only to individuals who left school at 16 or earlier, it still makes up nearly 70% of the UK sample. For this group, composition changes are likely to be quite severe since, as we have already seen, there was a dramatic fall in participation during this period. By contrast, in Sweden there has been an increase in wages for low-educated workers. The wage difference has decreased over time. In Germany, we also see no decline in real wages for the lower-education group. Indeed, at least for the decade 1986–1995, if anything Fig. 13c points to a slight fall in the raw educational differential for men. Given the stable employment rates during most of this period, it is difficult to attribute this rise to a composition effect.

4. A framework for understanding labor supply

Evaluating and interpreting labor supply estimates requires economic models to provide a context for comparison. Estimates often diverge simply because studies focus on evaluating behavioral responses corresponding to different wage and income effects. Sometimes empirical analyses are not precise about what model underlies their estimates. Is a static or

a lifecycle model used? What do substitution effects hold constant? Does the analysis recognize taxes and joint decision-making by family members? Does the model assume perfect certainty or can it allow for uncertainty? Does it assume a representative agent or is individual heterogeneity allowed? When researchers discover divergence in their labor supply estimates, they frequently cite sampling or data differences to explain discrepancies. Equally important, and often more informative for economists seeking to reconcile them, are the differences in economic frameworks used in the studies.

Many empirical studies of labor supply leave the reader to deduce the underlying model from the set of outcome and control variables incorporated in the analysis. Apart from hourly wages and other income, are controls for lifetime wages included? Is a measure of property income included and, if so, how is it measured? Do researchers account for expected changes in income sources? What demographic characteristics are included as controls? Differences in the included conditioning variables implicitly determine the economic framework as well as the response parameters estimated within that framework. Hence, a clear understanding of the implication of these decisions is necessary for any comparison of divergent estimates.

This section presents a unifying framework in which different basic labor supply models can be compared. By considering existing empirical work in one consistent framework, we can determine whether individual studies estimate meaningful parameters and, if so, which parameters are comparable across papers. Empirical studies will have to contend with practical issues concerning non-linear taxation, measurement error and the discreteness in choices. This section abstracts from these complexities so as to focus on the differences in interpretation across models. These complexities are then taken up in the remaining sections of the paper where specific empirical studies are also reviewed.

The development of a unifying structure for interpreting labor supply studies should not suggest that there is one correct way to estimate labor supply equations. Quite to the contrary, we recognize that many of the differences across existing empirical models reflect differences in data availability; our approach seeks to provide a synthesis in which results from each data source can be compared. Data vary in the forms of income that are included, the definition of hours and wage variables, and whether observations are longitudinal or cross-sectional, but meaningful and comparable results can be derived from each if the implications of the estimated function are carefully considered. An understanding of labor supply is greatly enhanced by any available source of exogenous wage and income variation, and no dataset providing this information should be discarded. At the same time, results from varying studies must be comparable, and the framework presented here seeks to facilitate these comparisons.

4.1. The static labor supply model

To set the scene, we begin by outlining the standard static, within-period labor supply model. This is an application of basic consumer theory. Assume each individual has a quasi-concave utility function

$$U(C_t, L_t, X_t) \tag{4.1}$$

in which C_t , L_t , and X_t are within-period consumption, leisure hours and individual attributes, in period t .³ Utility is assumed to be maximized subject to the budget constraint

$$C_t + W_t L_t = Y_t + W_t T, \tag{4.2}$$

where W_t is the hourly wage rate, Y_t is non-labor income, T is the total time available and a single consumption good is taken as the numeraire. The right-hand side of (4.2), then, includes the full value of one's endowment of time as well as all other sources of income. This is often defined as "full income" from which the consumer purchases consumption goods and leisure. We denote this income concept as M_t , so that

$$M_t \equiv Y_t + W_t T. \tag{4.3}$$

In static models, non-labor income, Y_t , is typically the sum of two components: asset income and other unearned income. We return to the measurement of non-labor income in our analysis of multiperiod models below.

First-order conditions take the familiar form

$$U_C(C_t, L_t, X_t) = \lambda_t, \quad U_L(C_t, L_t, X_t) \geq \lambda_t W_t, \tag{4.4}$$

where λ_t is the marginal utility of income. If the inequality in (4.4) holds strictly then the individual is not working and $L_t = T$. The wage, W_{Rt} , such that $U_L(Y_t, T, X_t) = \lambda_t W_{Rt}$, is the reservation wage below which the individual will not work.

Many have the mistaken impression that labor supply analyses rely on the assumption that individuals can freely choose their hours of work at a fixed wage with a single employer. The behavioral models considered here can readily be thought of as characterizing situations wherein persons choose their hours of work by selecting across employers offering different wage packages. In such instances, the labor supply function approximates the "average" relationship describing consumers' preferences for work hours and hourly earnings. Moreover, one can also allow for "wages" to vary as a function of hour of work with relatively straightforward modifications of the subsequent analyses.

4.1.1. Alternative representations of labor supply

An equivalent expression for the labor supply conditions (4.4) can be given in terms of the marginal rates of substitution (*MRS*). Eliminating λ_t from the first-order conditions (4.4) yields the equation

$$U_L/U_C \equiv MRS_L(C_t, L_t, X_t) \geq W_t. \tag{4.5}$$

This equation contains all information necessary to relate the level of leisure to the level of consumption.

³ X_t includes all consumer attributes in this specification – observed and unobserved. Since many individual attributes will not be fully observed by the econometrician, it is important to consider the treatment of unobserved heterogeneity in analyzing empirical specifications.

Solving the first-order conditions yields the Marshallian demand functions:

$$C_t = C(W_t, M_t, X_t), \quad L_t = L(W_t, M_t, X_t) \leq T. \quad (4.6)$$

Equivalently, using $H_t = T - L_t$ and the definition of M_t in terms of Y_t , we have the hours of work rule,

$$H_t = H(W_t, Y_t, X_t). \quad (4.7)$$

Many empirical studies of labor supply seek to estimate forms of (4.7). They vary widely in the measurement of the wage W_t , the income variable Y_t and the demographic controls incorporated in the specification. Depending how these issues are resolved, our full lifecycle framework, developed below, shows that “static” estimates can represent several types of substitution and income effects, ranging from those predicting responses to intertemporal movements in wages to those predicting responses to shifts in entire wage profiles.

Studies generally focus on the wage elasticity of the Marshallian supply function in (4.7), and on the associated utility-constant Hicksian wage elasticity. The Marshallian (uncompensated) wage elasticity is defined as

$$K_u = \partial \ln(H_t) / \partial \ln(W_t). \quad (4.8)$$

Denoting the Hicksian (compensated) wage elasticity by K_c , the Marshallian and Hicksian wage elasticities are linked by the Slutsky equation

$$K_u = K_c + \frac{W_t H_t}{Y_t} \frac{\partial \ln(H_t)}{\partial \ln(Y_t)}, \quad (4.9)$$

where the share $W_t H_t / Y_t$ is the size of earnings relative to non-labor income. The standard sign, homogeneity, and symmetry restrictions from consumer demand theory apply to the Hicksian supply function and have been used to check on the theoretical predictions of the model. Assuming that leisure is a normal good, this expression implies that the Hicksian compensated elasticity is larger than the Marshallian elasticity – the well known result that income and substitution effects work in opposite directions in Marshallian demand.

4.1.2. Family labor supply

Placing labor supply in a family or household context adds a number of important dimensions. Many tax and benefit policies designed to influence labor supply behavior can only be properly understood within a family labor supply framework. Moreover, the changes in the structure of wages facing men and women presented in Section 3, as well as changes in fertility, have important consequences in understanding the changing balance between men and women in family labor supply.

The standard “unitary” family labor supply model treats the family as a single decision-making unit. The attractions of this formulation are that standard welfare results from consumer theory are available and that the family labor supply model can be placed easily within the intertemporal framework. However, although with sufficient separability in

household members' utility the unitary approach can allow for decentralization of within-household allocations, these allocations continue to satisfy the Slutsky symmetry restrictions from consumer theory and also the "income-pooling" restrictions in which the marginal value of non-labor income is equalized across decision-making units within the family. Slutsky symmetry and income pooling are often considered to be unreasonable restrictions and a popular alternative framework that relaxes these latter two restrictions is the collective family labor supply model. This alternative representation of joint labor supply decisions also implies testable restrictions and is sensitive to the introduction of household production. A full discussion of the collective model and its relationship to the standard unitary framework is presented in Section 7. That section also provides a detailed evaluation of the empirical studies of family labor supply and considers the introduction of non-linear taxation and welfare programs. Here we simply outline the basic family labor supply model.

Suppose a family or household consists of two working-age individuals. Children and any other dependents are included in the vector of household attributes, X_t . Families are assumed to maximize joint utility over consumption, C_t , and the leisure of each family member, L_{1t} and L_{2t} . For such a household, utility may be written

$$U_t(C_t, L_{1t}, L_{2t}, X_t). \quad (4.10)$$

The budget constraint now takes the form

$$C_t + W_{1t}L_{1t} + W_{2t}L_{2t} = Y_t + W_{1t}T + W_{2t}T, \quad (4.11)$$

with full income now given by

$$M_t = Y_t + W_{1t}T + W_{2t}T. \quad (4.12)$$

The unearned income term, Y_t , combines all sources of non-labor income.

For the present discussion we consider consumption measured as a single aggregate, C_t .⁴ The first-order condition for consumption (4.4) continues to hold, but now it governs family consumption. For leisure choices, (4.4) is extended to give

$$U_C(C_t, L_{1t}, L_{2t}, X_t) = \lambda_t,$$

$$U_{L1}(C_t, L_{1t}, L_{2t}, X_t) \geq \lambda_t W_{1t}, \quad (4.13)$$

$$U_{L2}(C_t, L_{1t}, L_{2t}, X_t) \geq \lambda_t W_{2t}.$$

Reservation wages can be computed for each family member exactly as above. Demand functions now take the form

⁴ Typically, consumption, even for privately consumed goods, is recorded at the household level. Consequently, measuring individual consumption is difficult. However, for goods such as clothing, separate measurement is often recorded and can be helpful in identifying individual preferences within a household. This is described further in Section 7.

$$C_t = C(W_{1t}, W_{2t}, M_t, X_t),$$

$$L_{1t} = L_1(W_{1t}, W_{2t}, M_t, X_t) \leq T, \quad (4.14)$$

$$L_{2t} = L_2(W_{1t}, W_{2t}, M_t, X_t) \leq T.$$

This model provides a useful framework for thinking about household labor supply decisions. Clearly, if utility is weakly separable in the individual leisures then, provided the appropriate definition of income is used, individual labor supplies can be modeled in the usual way. However, separability is a strong restriction and one that would typically fail in a model that allowed for household production. In Section 7 we describe the family labor supply model more fully, particularly its relation to collective models of labor supply, household production and the analysis of discrete choices. We also present an overview of recent empirical results on family labor supply.

4.2. Multiperiod models of labor supply under certainty

Although its study is often placed in a static framework, labor supply is clearly part of a lifetime decision-making process. Individuals attend school early in life, accumulate wealth while in the labor force, and make retirement decisions late in life; each of these activities can only be understood in a lifecycle framework. We know that savings from labor earnings are often required to sustain individuals, or their dependents, during periods when they are out of the labor market. In addition, variations in health status, family composition and real wages provide incentives for individuals to vary the timing of their labor market earnings for income-smoothing and insurance purposes. In this section, we present the basic components of a single-agent lifecycle labor supply model assuming perfect certainty.

A full lifecycle model, starting at time t , is characterized by a utility function of the form

$$U_t = U(C_t, L_t, X_t, C_{t+1}, L_{t+1}, X_{t+1}, \dots, C_\tau, L_\tau, X_\tau). \quad (4.15)$$

The intertemporal budget constraint can be represented by the time path of assets, A , as

$$A_{t+1} = (1 + r_{t+1})(A_t + B_t + W_t H_t - C_t), \quad (4.16)$$

where A_{t+1} is the real value of assets at the beginning of period $t + 1$, r_{t+1} is the real rate of return earned on assets between t and $t + 1$, and B_t represents unearned-non-asset income. Individuals maximize (4.15) subject to the series of constraints given by (4.16) for all t through some fixed horizon τ ; τ is assumed to be known for simplicity.

This full model is empirically intractable, so virtually all studies assume some form of separability in time. That is, they assume that the utility function can be written

$$U_t = U(U^1(C_t, L_t, X_t), U^{t+1}(C_{t+1}, L_{t+1}, X_{t+1}), \dots, U^\tau(C_\tau, L_\tau, X_\tau)). \quad (4.17)$$

In this case, the marginal rate of substitution between leisure and consumption in period s can be written as

$$MRS_{L_s} \equiv (\partial U / \partial L_s) / (\partial U / \partial C_s) = U_L^s / U_C^s. \quad (4.18)$$

Combining this with the intertemporal budget constraint, we see that a necessary condition for maximization is

$$MRS_{L_s} \geq W_s. \quad (4.19)$$

So, with separability across time, the within-period marginal rate of substitution condition continues to characterize the relative amounts of leisure and consumption. All that remains, then, is to find a summary statistic that captures the impact of other periods on this decision, thus allowing one to pin down the levels of leisure and consumption. The two common methods for this are two-stage budgeting and marginal-utility-of-wealth-constant labor supply.

4.2.1. Two-stage budgeting

The idea behind two-stage budgeting is simple.⁵ Since the within-period marginal rate of substitution conditions continue to characterize behavior, we only need an allocation of full income, M_t , to each period to allow each maximization problem to be solved exactly as it was in the static problem. Hence, the decision rule can be decomposed into two stages: first, determine an allocation of wealth across periods; second, within each period, solve the standard static maximization problem. The solution to this problem can be found by reversing the two stages: first, maximize each period's utility, given some M_t ; this yields an indirect utility function, $V_t(M_t, W_t)$, for each period. Then, insert the V_t 's into U_t and choose the M_t 's to maximize this function given current wealth and future wages (expected wages under uncertainty). This solution can be represented by the demand Eqs. (4.6) and (4.7) together with an equation for M_t :

$$M_t = M(A_{t-1}^*, r_t, W_t, Y_t, X_t, Z_t), \quad (4.20)$$

where we have defined A_{t-1}^* to be the end of period $t - 1$ assets⁶ and Z_t represents future values of W , Y , r and X .

To compare this specification to the static specification introduced in Section 4.1, notice that the two-stage budgeting model automatically corrects the full income measure for the change in assets appropriate in the multiperiod model. From the definition of M_t , we may write

$$M_t = C_t + W_t L_t = r_t A_{t-1}^* + \Delta A_t^* + B_t + W_t T_t,$$

where now $r_t A_{t-1}^*$ is the real interest income available for expenditure on consumption at

⁵ Gorman (1959, 1968) is widely credited with developing the full implications of two-stage budgeting. MaCurdy (1983) and Blundell and Walker (1986), among others, have applied this concept in empirical analyses of labor supply.

⁶ In the discussion of multiperiod models, we use two definitions of assets: A_{t+1} in Eq. (4.16) is the beginning of period $t + 1$ assets and is therefore equal to $(1 + r_{t+1})A_t^*$. We do this to make our definition of the within-period budget for the two-stage budgeting problem consistent with the intertemporal constraint on assets (4.16).

the beginning of period t and ΔA_t^* is the adjustment in the level of real assets by the end of period t . In contrast, the full income variable in the static model simply includes real interest income and other non-asset income and is given by

$$Y_t + W_t T = r_t A_{t-1}^* + B_t + W_t T,$$

omitting the term ΔA_{t-1}^* which captures the intertemporal adjustment in assets.

The hours-of-work rule in the two-stage budgeting framework, mirroring (4.7), has the form

$$\max U(C_t, H_t, X_t)$$

subject to the budget constraint

$$C_t + W_t H_t = Y_t^C,$$

where we define the consumption-based other income variable

$$Y_t^C \equiv r_t A_{t-1}^* + \Delta A_t^* + B_t.$$

The first stage allocation (4.20) becomes

$$Y_t^C = Y^C(A_{t-1}^*, r_t, W_t, B_t, X_t, Z_t). \quad (4.20')$$

Note the appeal of the two-stage budgeting formulation. If consumption and leisure (work) hours for the period are observed, then M_t is observable via the within-period budget constraint

$$M_t = C_t + W_t L_t. \quad (4.21)$$

The appropriate adjustment of full income M_t or other income Y_t can be made either with information on assets across periods or with information on consumption. Hence, given some specification for the expectational variables Z_t , one can estimate (4.6) and (4.20) just as in the static framework. Marshallian elasticities can be derived by conditioning on Y_t^C in place of Y_t and can be converted to compensated elasticities via the Slutsky equation, yielding estimates of all other response parameters of interest. Of course, even if the static model were true, or some variant in which there were borrowing restrictions, the within-period allocations that condition on the consumption based measure of full income remain valid.⁷

In evaluating studies using this framework, one must keep two important considerations in mind. First, the appropriate measure of income is the value of consumption plus the value of leisure – that is the full income allocated to the period. Many researchers perform their analyses in a static framework, arguing that they are estimating the second stage of the two-stage budgeting process, but define income as current wages plus unearned

⁷ Note that this specification places no restrictions on the path of wages or interest rates, so that employment or capital market constraints can be accounted for, with a wage of 0 indicating no acceptable employment opportunities and an interest rate of ∞ indicating completely constrained capital markets.

income. As we have shown, in a lifecycle setting, these current income figures are irrelevant to current period work and consumption decisions except in so far as they impact the determination of M_t . Second, often elasticities or other response parameters estimated in this basic framework take M_t as fixed and exogenous, just as it is in the static model. Not only does this require the far-fetched notion that consumption is exogenous (if full income is valued with a consumption measure), but it misses any of the response to shocks that occurs through the first stage – that is, through a reallocation of the M_t s. In general, it is only by estimating both stages of the intertemporal allocation model that such responses can be fully accounted for. We take up this issue further in our discussion of multiperiod models under uncertainty.

4.2.2. Frisch labor supply equations and the Euler condition

Marginal-utility-of-wealth-constant labor supply functions, known as Frisch functions, provide an alternative and extremely useful method for analyzing lifecycle maximization problems. In this framework, the marginal-utility-of-wealth parameter, λ_t , serves as the sufficient statistic which captures all information from other periods that is needed to solve the current-period maximization problem. Our discussion critically relies on intertemporal strong separability in preferences, and, for simplicity, the analysis assumes a non-stochastic interest rate.

A useful representation of the problem is given by the functional equation formulation of dynamic programming. Consumers choose consumption and leisure according to the value function

$$V(A_t, t) = \max[U(C_t, L_t, X_t) + \kappa V(A_{t+1}, t + 1)] \quad (4.22)$$

subject to the asset accumulation rule (4.16). κ represents the consumer's discount factor. Standard dynamic programming techniques yield the following first-order conditions:

$$U_C(C_t, L_t, X_t) = \lambda_t,$$

$$U_L(C_t, L_t, X_t) \geq \lambda_t W_t, \quad (4.23)$$

$$\lambda_t = \kappa(1 + r_{t+1})\lambda_{t+1},$$

where λ_t is the marginal utility of wealth, $\partial V/\partial A_t$. These are the same first-order conditions as in the static problem, with the addition of the Euler equation for λ . This equation is central to the solution method since it determines the rule for the allocation of wealth across periods. In this formulation, the consumer chooses savings so that the marginal utility of wealth in period t equals the discounted value of the marginal utility of wealth in period $t + 1$, where the rate of discount is $\kappa(1 + r_{t+1})$.

These first-order conditions imply consumption demand and hours-of-work supply functions of the form

$$C_t = C(\lambda_t, W_t, X_t), \quad H_t = H(\lambda_t, W_t, X_t) \geq 0. \quad (4.24)$$

These are commonly referred to as Frisch demand functions.⁸ Their functional form depends only on the form of the utility function and whether a corner solution is chosen for hours of work at age t . These functions decompose consumption and labor supply decisions into components observed in the current period, X and W , and λ , which summarizes the relevant information from all other periods. Variables such as future wealth, wages, or personal characteristics affect consumption and labor supply only by changing the value of λ_t . Thus, λ_t serves the role of sufficient statistic, just as M_t was a sufficient statistic in the two-stage budgeting model.⁹

These Frisch labor supply functions are a third type of labor supply function along with the Marshallian and Hicksian functions previously discussed. Whereas Marshallian functions hold income constant and Hicksian functions hold utility constant, Frisch functions hold the marginal utility of wealth constant. One can calculate wage elasticities of the form $\partial H_t / \partial W_t$ for Frisch functions just as for Marshallian and Hicksian functions. We saw above that the Hicksian elasticity is larger than the Marshallian when leisure is a normal good; MaCurdy (1981) and Browning et al. (1985) show that the Frisch elasticity is the largest of the three.

The Euler equation implies a time path for λ of the form

$$\ln \lambda_t = b_t + \ln \lambda_{t-1} \quad (4.25)$$

where $b_t = -\ln(\kappa(1 + r_t))$. Repeated substitution yields

$$\ln \lambda_t = \sum_{j=1}^t b_j + \ln \lambda_0. \quad (4.26)$$

Hence, the λ term in (4.24) can be captured as an individual fixed effect, λ_0 , plus a function of age which is common across consumers.¹⁰ This ability to model differences in λ as individual effects is very important in the empirical specifications discussed below.

Estimation of (4.24) only allows computation of the Frisch elasticity. This measures the effect of a change in wages holding λ constant. As shown above, in this world of perfect certainty, the path of λ through time is determined solely by the known path of interest rates and the discount factor. Hence, for a given individual, changes in wages have no impact on λ and thus the Frisch elasticity is the correct elasticity for assessing the impact of wage changes through time on labor supply. However, researchers are often interested in comparing the impact of wage variation across consumers on labor supply. In this case, we do not simply examine evolutionary wage changes through time, but rather variation in

⁸ We have presented the hours-of-work supply function here, rather than the equivalent leisure demand function. With only two uses for a consumer's time, the two are obviously related by the identity $L_t = T - H_t$.

⁹ Note that if labor and capital income is taxed jointly by a non-linear tax, these conditions may need adapting (see, e.g., Blomquist, 1985).

¹⁰ The b_j terms are functions of κ and r which are assumed constant across consumers. Note that if we assume the rate of time preference, ρ (where $\kappa = 1/(1 + \rho)$), equals the rate of interest, b_j is 0 for all j and λ is constant over time.

the entire wage profile. This variation certainly impacts the value of λ_0 and, thus, the Frisch wage elasticity is inappropriate for measuring the effect of such wage variation. To estimate the full impact of wages requires a specification of the impact of the wage profile on λ_0 . We consider this further at the end of this section where we evaluate the appropriate elasticity measures for alternative policy questions.

4.2.3. Multiperiod models of family labor supply

The family labor supply model becomes more complicated with the addition of multiple periods or uncertainty, as family composition may change over time. As long as the unitary model is maintained, however, its analysis is straightforward. The marginal conditions for the λ -constant, Marshallian and marginal rate of substitution labor supply equations described above follow naturally from the first-order conditions given by (4.13). Notice that there is still only a single marginal utility of wealth, λ_w , and, therefore, there remains only one Euler condition as in the third equation of (4.23). Consequently, allocations to each individual in this time separable model satisfy equality of marginal utility of wealth. However, to avoid strong separability assumptions between each family member's leisure, careful choice of specification for Frisch labor supplies is required. Further extensions of the multi-period to the family labor supply case are presented in Section 7.

4.3. Multiperiod models of labor supply under uncertainty

The concepts developed in the certainty case essentially carry over to a lifecycle model that recognizes that individuals make labor supply choices in an environment in which they are uncertain about their futures. This requires replacing the deterministic dynamic programming characterization of behavior that we considered in the previous section with a formulation in which agents optimize expected lifetime utility.¹¹

4.3.1. Two-stage budgeting under uncertainty

Accounting for uncertainty in two-stage budgeting is inconsequential. Eqs. (4.6) and (4.20) continue to summarize choices. Actually solving for the optimum period-specific expenditure allocation, M_t , is now more difficult since wealth cannot be allocated once at the beginning of life and instead must be reallocated each period as information is revealed. One can solve this problem, however, via standard dynamic programming formulations. Instead of including realized values, the variables Z_t in (4.20) (or (4.20')) now include attributes of the distribution of future wages and income, and future determinants of preferences. In Section 4.5, we consider approaches to estimation of the full lifecycle model which combine the two-stage budgeting formulation with the intertemporal first-order conditions on consumption. This turns out to be a useful way of characterizing the preference restrictions underlying various empirical specifications.

¹¹ Much of this framework comes from MaCurdy (1985). We refer the reader to this reference for details of the development of these specifications.

4.3.2. Frisch labor supply under uncertainty

With the introduction of uncertainty over future wages, the dynamic programming representation of the consumer’s problem changes to

$$V(A_t, t) = \max\{U(C_t, L_t, X_t) + \kappa E_t[V(A_{t+1}, t + 1)]\} \tag{4.27}$$

subject to the asset accumulation rule (4.16).¹² The first-order conditions now include (4.23) and (4.24) together with a modified Euler equation,

$$\lambda_t = \kappa E_t[\lambda_{t+1}(1 + r_{t+1})]. \tag{4.28}$$

The only change from the certainty case is that λ_{t+1} is now a random variable which is not realized until the start of period $t + 1$. The savings allocation rule, given by (4.28), determines the path followed by λ through time. Given that the consumer cannot perfectly control the level of his wealth, his environment changes as he acquires information and λ_t is stochastic. Condition (4.28) describes how the consumer allocates his resources to account for unanticipated shocks. He sets his savings policy so that the expectation of next period’s marginal utility of wealth is revised by the full amount of the unanticipated elements; in other words, the consumer revises the means of all future values of λ to account for all forecasting errors when they are realized. The standard Euler equation for consumption is derived by replacing λ_t and λ_{t+1} in (4.28) by $U_c(C_t, L_t, X_t)$ and $U_c(C_{t+1}, L_{t+1}, X_{t+1})$, respectively.

A useful characterization of the stochastic process for λ implied by (4.28) takes the form

$$\ln \lambda_t = b_t^* + \ln \lambda_{t-1} + \varepsilon_t^*,$$

where the coefficient b_t^* depends on the discount factor, κ , the interest rate, r_t , and the moments of the forecast error, ε_t^* .¹³ Repeated substitution yields

$$\ln \lambda_t = \sum_{j=1}^t b_j^* + \ln \lambda_0 + \sum_{j=1}^t \varepsilon_j^* = b^* t + \ln \lambda_0 + \sum_{j=1}^t \varepsilon_j^*, \tag{4.29}$$

where the last expression assumes $b_t^* = b^*$ for expositional simplicity.

Combining Eq. (4.29) with the consumption and labor supply conditions in (4.23) suggests a simple view of lifecycle behavior under uncertainty. At the start of the lifetime, the consumer sets the initial value of λ_0 to include all available information. As he ages, he

¹² For simplicity of presentation, we continue to assume a non-stochastic interest rate, although the extension to random interest rates is trivial.

¹³ To develop this expression, write $\ln \lambda_t = E_{t-1}\{\ln \lambda_t\} + \varepsilon_t^*$. This relation implies $E_{t-1}\{\lambda_t\} = \exp[E_{t-1}\{\ln \lambda_t\}]E_{t-1}\{\exp[\varepsilon_t^*]\}$, which in turn yields $\lambda_t = [E_{t-1}\{\exp[\varepsilon_t^*]\}]^{-1}E_{t-1}\{\lambda_t\}\exp[\varepsilon_t^*]$. Inserting the condition for $E_{t-1}\{\lambda_t\}$ given by Eq. (4.28) into this latter relation yields $\lambda_t = [E_{t-1}\{\exp[\varepsilon_t^*]\}]^{-1} \times (\kappa(1 + r_t))^{-1} \lambda_{t-1} \exp[\varepsilon_t^*]$. Taking natural logs gives $\ln \lambda_t = b_t^* + \ln \lambda_{t-1} + \varepsilon_t^*$ where $b_t^* \equiv -\ln(\kappa(1 + r_t)) - \ln(E_{t-1}\{\exp[\varepsilon_t^*]\})$.

responds to new information by updating λ according to (4.29). At each age, the consumer only needs the updated λ , along with current wages and characteristics, to determine his optimal consumption and labor supply.

A substantial complication to the labor supply models described in this section arises if we relax the assumption of time-separable utility. For example, consider allowing an individual's wage to be a function of human capital, which this person chooses to acquire by training. In this case, the wage is endogenous, as it is determined by an individual's training decision. The primary method of dealing with this complication is to move to a fully structural model of lifetime decision-making in which parameter values are chosen to match closely the observed work, training and consumption decisions. We defer a discussion of this and other related dynamic generalizations of the labor supply model that relax the time separability assumption to Section 8, which considers dynamic structural models.

4.4. Basic empirical specifications

A prototype empirical specification that encompasses many economic models of labor supply takes the form

$$\ln H_t = \alpha \ln W_t + \beta Q_t + e_t, \quad (4.30)$$

where α and β are parameters, Q_t is a vector of "controls" and e_t is a stochastic term unobservable to the economist. In what follows, we consider alternative specifications for βQ_t .

Studies also often use alternative transformations of H as the dependent variable. For example, a popular alternative is the semi-log specification

$$H_t = a \ln W_t + b Q_t + v_t, \quad (4.30')$$

which is particularly attractive for dealing with non-participation. One also finds various formulations for wages as right-hand side variables (such as after-tax wages or non-linear functions of wage rates). In each case, additivity between the log wage variable, unearned income variables and the other controls will imply restrictions on preferences. The preference restrictions underlying these and other popular labor supply specifications are reviewed in Appendix A.

The value of α in (4.30) determines the substitution effect associated with the response of labor supply to changes in wages. As discussed above, the interpretation of this substitution effect varies according to precisely which controls one includes in the vector Q , and which of these controls are treated as exogenous.

4.4.1. Static specifications

The conventional static specification involves estimating Eq. (4.30) with controls set according to

$$\beta Q_t = \rho X_t + \theta Y_t, \quad (4.31)$$

where X_t is a vector of observable “taste shifter” controls and Y_t is a measure of non-labor income. Non-labor income is typically measured as the sum of interest income $r_t A_{t-1}$ and exogenous income B_t . This static specification is only appropriate if the static model of Section 4.1 is correct. This could be the case if consumers behave completely myopically or if capital markets are completely constrained so that it is impossible to transfer capital across periods. If the static model is correct, the wage coefficient in specification (4.30) measures

$$\alpha = \alpha_y = \text{uncompensated substitution elasticity given income } Y. \quad (4.32)$$

The parameter α_y corresponds to the Marshallian wage elasticity in the static model. Its estimation requires instrumental-variable techniques to account for the endogeneity of the wage, arising from unobservable characteristics affecting both W_t and H_t or from measurement error. Nevertheless, if consumers adjust their behavior to account for factors in future periods, the coefficient on log wage lacks economic meaning, no matter what econometric methods are applied. That is, if the labor supply decision has any lifecycle elements, static regressions confuse shifts of wage profiles with movements along wage profiles and, thus, yield parameters that lack economic interpretation.

4.4.2. Two-stage budgeting specifications

To estimate a labor supply equation within a two-stage budgeting framework, set

$$\beta Q_t = \rho X_t + \theta Y_t^C, \quad (4.33)$$

where Y_t^C is the consumption-based income measure defined in (4.20'). (Alternatively, one can condition on the full income measure M_t , defined by (4.20)). In applying these controls, one should note that Y_t^C (or M_t) is defined by leisure and consumption choices and, thus, is endogenous. Appropriate instrumental-variable techniques must, therefore, be applied to obtain consistent estimators. The wage elasticity coefficient, α , can then be interpreted as

$$\alpha = \alpha_c = \text{uncompensated substitution elasticity given total consumption } C \quad (4.34)$$

This wage effect determines the impact of wages on hours worked, holding the first-stage income allocation constant. Hence, it captures the impact of anticipated wage movements through time, but does not capture the impact of shifts of the entire wage profile, as these shifts would also impact hours through their effect on the allocation of Y_t^C . In general, one needs a model of Y_t^C that includes the impact of all current and future wages to assess the impact of wage profile shifts. We take up this issue further in the discussion of relevant elasticities for policy evaluation in Section 4.5.

4.4.3. Frisch specifications

To create a Frisch labor supply function in the form of (4.30), suppose the contemporaneous utility function for period t takes the form

$$U_t = G(C_t, X_t) - \Psi_t(H_t)^\sigma, \quad (4.35)$$

where G is a monotonically increasing function of C_t , $\sigma > 1$ is a time-invariant parameter common across consumers and Ψ_t is a function of consumer characteristics. We take Ψ_t to be $\exp(-X_t \rho^* - \nu_t^*)$ where ν_t^* reflects the contribution of unmeasured characteristics and ρ^* is a vector of preference parameters.

Assuming an interior optimum, the implied Frisch hours-of-work function takes the form of (4.30) with

$$\beta Q_t = F_t + \rho X_t, \quad (4.36)$$

where $F_t = \alpha(\ln \lambda_t - \ln \sigma)$, $\alpha = 1/(\sigma - 1)$, $\rho = \alpha \rho^*$, and $e_t = \alpha \nu_t^*$. Modifying Eq. (4.29) by assuming that the b_t^* terms are constant across consumers and time, and substituting this into (4.36) yields

$$\beta Q_t = F_0 + bt + \rho X_t, \quad (4.37)$$

where $b = \alpha b^*$, and e_t now includes sums of forecast error terms. So, the necessary controls are the exogenous variables X_t , age and an individual effect F_0 . Taking first differences of this form of Eq. (4.30) yields

$$\Delta \ln H_t = b + \rho \Delta X_t + \alpha \Delta \ln W_t + \Delta e_t. \quad (4.38)$$

Given the availability of instruments for the change in wage, one can fit this equation on panel data to yield an estimate of α . In these specifications α corresponds to the Frisch wage elasticity discussed above, which the literature commonly designates

$$\alpha = \alpha_t = \text{intertemporal substitution elasticity}. \quad (4.39)$$

This elasticity holds marginal utility of wealth constant, and it describes how changes in wages induced by movements along an individual's wage profile influence hours of work. Individuals fully anticipate these wage movements and this is why F_0 remains fixed. For this reason, they are often referred to as evolutionary wage changes.¹⁴

If we wish to measure the impact of wage variation across consumers, or unanticipated shifts of an individual's wage profile, we must complete the model and provide an empirical specification of the evolution of wages and other incomes as well as accounting for the impact of these shifts on F_0 . Hence, we need an empirical specification for λ and, thus, for F . This is provided by the lifecycle specifications that we now consider.

4.4.4. Lifecycle specifications

For this empirical specification, we assume that one can approximate $\ln \lambda_0$ by the equation

$$\ln \lambda_0 = D_0 \varphi_0^* + \sum_{j=0}^{\tau} \gamma_{0j}^* E_0 \{\ln W_j\} + \theta_0^* A_0 + a_0^*, \quad (4.40)$$

¹⁴ Although the particular form for utility, (4.35), conveniently implies a log-linear Frisch labor supply equation, it also places strong restrictions on the form of within period and intertemporal preferences. In this specification, labor supply and consumption are explicitly additive in utility both within period and across periods.

where D_0 is a vector of demographic characteristics either observed at 0 or anticipated in future periods, and a^*_0 is an error term. This implies a form for F_0 :

$$F_0 = D_0\varphi_0 + \sum_{j=0}^{\tau} \gamma_{0j}E_0\{\ln W_j\} + \theta_0A_0 + a_0, \tag{4.41}$$

where the parameters and error term equal their superscript “*” counterparts multiplied by α , and with the intercept defined to include the term $-\alpha \ln \sigma$. This empirical specification imposes strong simplifying restrictions – it assumes that the consumer knows he will work τ periods and it incorporates any effect of interest rates or time preference into the intercept and other parameters.

Relations (4.41) and (4.37) yield a formulation for (4.30) with

$$\beta Q_t = D_0\varphi_0 + \sum_{j=0, j \neq t}^{\tau} \gamma_{0j}E_0\{\ln W_j\} + \theta_0A_0 + bt + X_t\rho, \tag{4.42}$$

$\alpha = \alpha_t + \gamma_{0t}$, where the disturbance in (4.30) is $e_t = a_0 + \nu_t - \gamma_{0t}(\ln W_t - E_0\{\ln W_t\})$. So, Q_t now includes all start-of-life controls used to form λ_0 and all controls needed for the period- t utility function: age, initial wealth and the expected wage profile as of age 0. Estimation of this equation yields an estimate of $\alpha_t + \gamma_{0t}$, the wage elasticity of hours corresponding to a shift in the period t wage rate, as well as estimates of the γ_{0s} s determining the impact of a shift in the entire wage profile. As we argue below, this formulation also provides us with precisely the parameters we need for the analysis of tax reform.

Implementing (4.42) requires the econometrician to have consistent predictions of the consumer’s expected future wages. Assume that the lifetime wage path anticipated in period 0 is

$$E_0\{\ln W_t\} = \pi_0 + \pi_1t + \pi_2t^2 + u_t, \tag{4.43}$$

where the π ’s are deterministic functions of time invariant characteristics of the consumer and u_t is an error term assumed to be uncorrelated with all demographic variables in M_0 as well as with those used to predict wages or wealth (below).

A researcher also requires a specification for initial wealth since most datasets do not include this variable. If we assume that property income, Y_t , follows a path similar to wages (with the similar properties for errors and parameters)

$$E_0\{Y_t\} = \zeta_0 + \zeta_1t + \zeta_2t^2 + \eta_t, \tag{4.44}$$

then using the fact that $Y_0 = (A_0/1 + r_0)r_0$, we see that initial wealth can be predicted by $\zeta_0(1 + r_0)/r_0$.

Combining these forms for wages and wealth with (4.41), we arrive at an expression for the individual effect:

$$F_0 = D_0\varphi_0 + \pi_0\bar{\gamma}_0 + \pi_1\bar{\gamma}_1 + \pi_2\bar{\gamma}_2 + \zeta_0\bar{\theta} + \mu, \tag{4.45}$$

where

$$\bar{\gamma}_k = \sum_{j=0}^{\tau} j^k \gamma_{0j}, \quad \text{for } k = 0, 1, 2, \quad \bar{\theta} = \theta_0 r_0 / (1 + r_0)$$

and μ is a disturbance depending on the errors a_0 , u_i 's and η_i 's. This equation relates a consumer's individual effect to the parameters of his wage and income profile. Relations (4.37) and (4.45) yield a formulation for (4.30) with

$$\beta Q_t = D_0 \varphi_0 + \pi_0 \bar{\gamma}_0 + \pi_1 \bar{\gamma}_1 + \pi_2 \bar{\gamma}_2 + \zeta_0 \bar{\theta} + bt + X_t \rho, \quad (4.46)$$

$$\alpha = \alpha_i + \gamma_{0i},$$

where the disturbance e_t in (4.30) now incorporates the error component μ . Hence, in this formulation, D_0 and age remain as controls, but initial wealth and the expected wage profile are replaced by the parameters describing wage and property income profiles through time. Simultaneous estimation of (4.43), (4.44) and (4.46) yields estimates of all parameters needed to compute the response of hours of work to both evolutionary and parametric wage changes. In this formulation, only wages and property income are endogenous.

4.4.5. Interpreting cross-sectional specifications in a lifecycle framework

Many labor supply studies attempt to estimate "wage elasticities" using cross-sectional variation in wages. As we have seen above, the term wage elasticity is ambiguous – it is crucial that the researcher distinguish between evolutionary and parametric wage shifts. Since most do not, the reader is left trying to compare elasticity estimates that may not be comparable. Add to this the difficulty of identifying any lifecycle effects in a cross-sectional setting and, even if there were no data measurement differences, it would not be surprising to see many different elasticity estimates.

Utilizing the above framework, we can evaluate what cross-sectional specifications actually allow meaningful lifecycle parameter estimates to be recovered and which specific parameters are being estimated given the included control variables. To develop a simple expression for (4.46) which can be compared to those of existing cross-sectional studies, assume that $D_0 \phi_0$, π_0 , π_1 , π_2 , and ζ_0 are linear functions of the variables contained in a vector, K . Then we have

$$\ln H_t = Kq + bt + X_t \rho + (\alpha_i + \gamma_{0i}) \ln W_t + e_t, \quad (4.47)$$

where q is a vector of coefficients. Alternatively, we could assume that D_0 contains only an intercept and that the coefficients on age and age-squared for the lifetime wage and income paths (i.e., π_0 , π_1 , π_2 , and ζ_0) are constant across consumers. Then one can write (4.46) as

$$\ln H_t = d_1 + d_4 t + d_5 t^2 + \bar{\theta} Y_t + X_t \rho + (\alpha_i + \bar{\gamma}_0) \ln W_t + e_t, \quad (4.48)$$

where

$$d_1 = g_0 + \pi_1 \bar{\gamma}_1 + \pi_2 \bar{\gamma}_2,$$

$$d_4 = b - \pi_1 \bar{\gamma}_0 - \gamma_2 \bar{\zeta}_1,$$

$$d_5 = -\pi_2 \bar{\gamma}_0 - \theta \alpha_2 \bar{\zeta}.$$

So, there are two equations which one can estimate using instrumental variable techniques on cross-sectional data to yield meaningful lifecycle parameter estimates. If a researcher regresses log hours of work on age; all age-invariant characteristics determining lifetime wages, preferences, and initial permanent income; and log wage, then the coefficient on the current wage rate is α , the Frisch elasticity. Intuitively, this approach controls for differences in the value of F_0 across consumers and leaves higher-order age variables as instruments to identify wage variation. Hence, only evolutionary wage variation along the age-wage path is included.

If, alternatively, a researcher regresses log hours worked on property income, age, age squared, and log wage, the coefficient on wage is the response of labor supply to a parametric wage shift – including both the intertemporal substitution effect, α , and the reallocation of wealth across periods captured by a change in F . Intuitively, this approach controls for age effects and leaves individual characteristics as instruments for wage. Changes in these characteristics capture full profile shifts, rather than movements along the age-wage path. The static equations presented in (4.30) fit neither of these patterns, however, as they include property income together with personal characteristics rather than age and age squared. Hence, as noted above, given the existence of lifecycle effects they confuse the effect of movements along the wage profile with shifts in the profile and, thus, yield parameters without an economic interpretation.

4.5. Which elasticities for policy evaluation?

This section has highlighted four “core” wage elasticities which correspond to four key specifications for control variables that can be found in the empirical literature on labor supply. Two are within-period elasticities: α_V relating to the purely static formulation (4.31) and α_C relating to the two-stage budgeting specification (4.33). Two are lifecycle elasticities: α_I the intertemporal elasticity of substitution relating to the Frisch specification (4.38) and measuring responses to evolutionary movements along the lifecycle wage profile, and $\alpha_I + \gamma_0$ relating to the full lifecycle specification (4.42) and measuring responses to parametric shifts in the lifecycle profile itself. As most tax and benefit reforms are probably best described as once-and-for-all unanticipated shifts in net-of-tax real wages today and in the future, the most appropriate elasticity for describing responses to this kind of shift is $\alpha_I + \gamma_0$. Here we examine the relations among each of the elasticities and consider their relevance for policy evaluation.

4.5.1. Relationships among the lifecycle elasticities

The Frisch specification treats the individual marginal-utility-of-wealth as a fixed effect and allows the researcher to estimate only the intertemporal substitution elasticity, α_I . Given that appropriate methods are employed to account for the fixed effect (generally first

differencing in panel data), the relevant independent variables, apart from the wage, are simply within-period characteristics and age.¹⁵ The Frisch elasticity, by ignoring this (unexpected) shift in wealth from a once-and-for-all change in real wages, is larger than the policy-relevant elasticity $\alpha_l + \gamma_0$ and overestimates the impact of a reform.

Direct estimation of the simple parameterization of the full lifecycle model, required to recover $\alpha_l + \gamma_0$, relies on specifications for both within-period utility and the individual marginal-utility-of-wealth effect. As a result, controls are needed for all of the following: “start of life” characteristics which impact the initial setting of F_0 , current-period characteristics which affect the within-period utility function, age, expected wages as of time 0, and initial wealth. Expected wages are unobservable and initial wealth is generally not included in data sets, so these should be replaced with the parameters governing the time path of wages and property income, which must be jointly estimated with the labor supply equation. Estimation of this full framework allows computation of both the intertemporal substitution elasticity and the elasticity of labor supply in reaction to a full, parametric wage profile shift. However, it is also the most demanding in terms of data.

It is worth noting that the elasticity derived from the static specification, α_γ , can be placed in an intertemporal setting but is economically meaningful only under a strong assumption of either complete myopia or perfectly constrained capital markets. Otherwise, this elasticity confuses movements along wage profiles with shifts of these profiles and, thus, yields response parameters which are a mixture of these. Such hybrid estimates lack an economic interpretation and are not generally useful in policy evaluation.

However, we have also described several formulations which appear essentially static, but which vary greatly based on included controls. Under simplifying assumptions, formulation (4.47) allows the researcher to compute the intertemporal substitution elasticity using cross-sectional data alone. Age and age-invariant consumer characteristics are the required controls. In contrast, formulation (4.48) allows one to estimate the response to a parametric wage shift. Required controls here are property income in period t , age, and age squared.

4.5.2. Relationships among within-period and lifecycle elasticities

In general, a tax policy reform will lead to a change in the optimal level of consumption and full income. The within-period elasticity, α_C , based on the two-stage budgeting framework, does not account appropriately for intertemporal adjustments in consumption. So how should we interpret elasticity α_C from the two-stage budgeting formulation? Under the strong assumption of either complete myopia or perfectly constrained capital markets, this elasticity is identical to α_γ . But in the lifecycle model with capital markets, the precise relationship between the policy-relevant elasticity, $\alpha_l + \gamma_0$, and α_C is ambiguous. However, since α_C is bounded above by the Slutsky compensated elasticity and α_l is

¹⁵ In the model with uncertainty the fixed effect is replaced by a random walk (see Section 4.3.2), but the first difference solution to estimation is retained with appropriate adjustment for the endogeneity of differenced wages.

bounded below by the Slutsky elasticity, α_C is no greater than the Frisch elasticity. It may well be much smaller and, unlike α_t , can be negative.

Indeed, in certain cases, α_C precisely reflects the labor supply adjustment induced by the shift in wealth, capturing exactly the impact of the parametric shift in the wage profile that corresponds to a policy reform involving an unexpected and permanent change in real wages. To see this, consider the case where within-period preferences are of Stone–Geary form

$$U_t = \theta \ln(\gamma_H - H_t) + (1 - \theta) \ln(C_t - \gamma_C), \tag{4.49}$$

where θ , γ_H and γ_C are preference parameters. Suppose also that intertemporal preferences are explicitly additive over U_t .¹⁶ The labor supply specification from the two-stage budgeting approach has the form

$$H_t = \gamma_H - (\theta/W_t)\{Y_t^C - \gamma_C + \gamma_H W_t\} \tag{4.50}$$

and the within-period elasticity is

$$\alpha_C = \frac{W}{H} \cdot \frac{\partial H}{\partial W} \Big|_{\gamma_C} = \frac{\gamma_H}{H} (1 - \theta) - 1. \tag{4.51}$$

To compare this elasticity with $\alpha_t + \gamma_{0t}$, we can compute the following expression for λ_t^{-1} :

$$\lambda_t^{-1} = A_{t-1} + \sum_j (\kappa(1 + r))^{-j} (\gamma_H W_j - \gamma_C). \tag{4.52}$$

Now consider a permanent change in the wage, W . Assume (i) $\kappa(1 + r) = 1$ and (ii) future real wages remain at this new level. The corresponding elasticity is

$$\frac{W}{H} \frac{\partial H}{\partial W} = \frac{\gamma_H}{H} (1 - \theta) - 1, \tag{4.53}$$

which, in this case, is identical to the within-period uncompensated elasticity from the two-stage budgeting formulation (4.51). In this case, it turns out that the consumption-based measure of other income, Y_t^C , is constant for a permanent uniform shift in real wages and, consequently, α_C matches the policy-relevant elasticity. Consumption levels adjust but are exactly offset by the change in $W_t H_t$ in the definition of $Y_t^C = C_t - W_t H_t$. This example shows that, in certain cases, the adjustment for the wealth effect needed to account for the unexpected and permanent change in future wages arising from a policy change is completely captured in the two-stage budgeting formulation. It also highlights the degree to which the intertemporal substitution elasticity overestimates the policy relevant effect.

For completeness, consider now the Frisch elasticity for this Stone–Geary specification. The Frisch labor supply has the form

¹⁶ See Ashenfelter and Ham (1979) and Bover (1989) for further discussion of this specification.

$$H_t = \gamma_H - (\theta/W_t)\lambda_t^{-1}, \quad (4.54)$$

with elasticity given by

$$\alpha_l = \frac{W}{H} \frac{\partial H}{\partial W} \Big|_{\lambda} = \frac{\gamma_H}{H} - 1. \quad (4.55)$$

This intertemporal substitution elasticity must be non-negative since $\gamma_H \geq H_t$ and, since θ lies between zero and one, this elasticity is larger than α_C from the two-stage budgeting formulation.

In general, the equivalence between α_C and $\alpha_l + \gamma_0$ found in this Stone–Geary example without uncertainty does not hold. The Stone–Geary preference specification and explicit additivity over time places strong restrictions on preferences. In Appendix A we describe the properties of this and other popular preference models for within-period labor supply.

One general way to exploit the simplicity of the second stage of the two-stage budgeting formulation under uncertainty is to use the linkages between within-period and intertemporal preference restrictions.¹⁷ This combines the within period stage, which conditions on consumption, with an Euler equation for the marginal utility of wealth under uncertainty. All preference parameters needed to describe both stages of the intertemporal allocation model under uncertainty are identified by combining the second stage of the two-stage budgeting framework with the Euler equation for consumption. Within-period allocations between consumption and leisure are completely described by the labor supply equations that condition on the consumption-based measure of full income or the marginal rate of substitution condition between consumption and hours. The Euler condition on the marginal utility of wealth then recovers the remaining parameters describing intertemporal allocations.

This approach of combining the two-stage budgeting formulation with the Euler equation for the marginal utility of consumption has many potential advantages over the Frisch and full lifecycle approaches. Frisch labor supply models specify hours of work directly in terms of wages and the marginal utility of wealth. The strong restrictions on preferences in the standard log linear specification can be seen directly from the implied form of utility in (4.35). Utility is explicitly additive over time, goods and leisure. In general, for the Frisch labor supply model to be log linear in the wage and log marginal utility, the intertemporal utility must be explicitly additive over time, consumption and hours. However, the two-stage budgeting approach requires accurate measurement of consumption as well as labor supply and real wages. Moreover, there are many potential pitfalls. Additive heterogeneity

¹⁷ Consider writing the period-specific utility function $U(C_t, L_t, X_t)$ in the intertemporal program (4.27) as $U(C_t, L_t, X_t) = G(u(C_t, L_t, X_t), X_t)$, where G is some positive monotonic transformation of a quasi-concave, differentiable, within-period utility, u . This expression is convenient since the marginal within-period allocation conditions (4.4) become $G_u u_c(C_t, L_t, X_t) = \lambda_t$ and $G_u u_L(C_t, L_t, X_t) \geq \lambda_t W_t$. The marginal rate of substitution $u_L/u_C \equiv MRS_L(C_t, L_t, X_t) \geq W_t$ does not depend on G . Consequently, within-period allocations place no restrictions on G and, therefore, provide no information on the identification of G . In contrast, the Euler condition (4.28) involves the derivatives of G and u . Given u , the form of G places restrictions on intertemporal preferences.

at the within-period level does not fit easily into a non-linear Euler equation. Similar issues arise with measurement error, endogeneity and non-participation.¹⁸ As with static labor supply models, simple specifications may be preferred in empirical applications where heterogeneity and measurement error are considered to be overriding issues.

4.5.3. Summary and some qualifications

This section has demonstrated the importance of understanding which elasticity is being recovered in the empirical analysis of labor supply and has shown that this depends crucially on the conditioning variables included in estimation. We have identified four “core” elasticities that are commonly estimated and which differ substantially in their interpretation. For this purpose we have abstracted, in this section, from important issues such as non-linear taxation, discreteness in choices and flexibility in the specification of preferences, so as to highlight the differences in interpretation of coefficients across alternative specifications. We have argued that, in general, a full lifecycle parameterization of the model is needed to evaluate policy reforms. However, we have shown how key policy-relevant elasticities can be recovered from the analysis of available data sources.

The analysis presented here and elsewhere in this chapter is conducted in a partial equilibrium framework and, therefore, considers only one side of the market. To analyze the impact of a policy reform, a general equilibrium analysis will sometimes be required, though discussion of this is outside the scope of this chapter. The model specifications examined in this section have been stylized and often relate to simple linear formulations, which place strong restrictions on preferences.¹⁹ Furthermore, in focusing on one side of the market, these specifications may not directly capture short-term constraints on the adjustment of labor supply. Nevertheless, they do include error terms to reflect this and should be viewed as representing “average” behavior. Ham (1986a,b) provides evidence of the importance of short-run constraints. Extreme liquidity constraints may also limit the usefulness of the intertemporal model. Finally, it may be that these simple intertemporal models are inappropriate for certain types of workers. For example, in a (unionized) bargaining model, hours-wages contracts might implicitly allow for smoothing consumption via clauses that provide for a steady stream of income in exchange for additional effort from the workers in good times. See Card (1994) for a critical review of the intertemporal labor supply model. In Section 8 we consider many extensions of the basic intertemporal model, though we focus only on those extensions that allow for human capital, habits and discrete participation choices.

5. Policy reforms and the natural experiment approach

“Natural experiments” have gained considerable popularity recently, and the simplicity of

¹⁸ Section 8 considers the introduction of participation in this formulation.

¹⁹ In Appendix A, we summarize the preference restrictions underlying popular parameterizations of labor supply.

this estimation method will undoubtedly make its popularity enduring among empirical economists for some time to come. This method often goes by the name of the difference-in-difference estimator. This section interprets the essence of this approach, and it relates those applications that estimate how tax and welfare policies influence labor supply to the empirical models surveyed elsewhere in this chapter. Although the discussion focuses on labor supply analyses, the evaluation presented here applies to any implementation of the natural-experiment approach.

The natural-experiment approach is not new, nor is it a method that is “non-structural”. The statistical apparatus underlying this approach has been extensively applied in the labor-economics literature since the inception of empirical work in the field. The basic idea is to compare (at least) two groups, one of which experienced a specific policy change, and another with similar characteristics whose behavior was unaffected by this policy change. The second group is assumed to mimic a control environment in experimental terminology. Such comparisons provide the foundation for most empirical work in labor economics. The problem comes in creating a control environment, which is done either by including exogenous variables in an analysis designed to adjust for relevant differences among sample observations, or by selecting observations in a manner that permits a matched-pair type of analysis.

Contrary to many researchers’ perceptions, the natural experiment approach relies on restrictive structural assumptions analogous to those of most other methods. In fact, this approach is entirely equivalent to the fixed-effects model popularized in the 1970s. By writing the model in this way, we are able to compare it with the alternative structural models outlined in the previous section and to state the conditions under which a structural interpretation can be placed on estimates from studies that use this approach.

5.1. The natural-experiment approach and the difference-in-differences estimator

Suppose one is interested in estimating the influence of a policy instrument on an outcome for a group, say outcome y_{it} measuring hours of work or participation. The group consists of individuals $i = 1, \dots, N$, with these individuals observed over a sample horizon $t = 1, \dots, T$. (Individuals here may refer to data on groups such as the average in a state or in a specific demographic category.) Suppose further that the policy instrument changes in particular period t for only a segment of the group. Let δ_{it} be a zero-one indicator that equals unity if the policy change was operative for individual i in period t . Members of the group who experience the policy change react according to a parameter γ . A framework for estimating γ expressed in terms of a conventional fixed-effect model takes the form

$$y_{it} = \gamma\delta_{it} + \eta_i + m_t + \varepsilon_{it}, \quad (5.1)$$

where η_i is a time-invariant effect unique to individual i , m_t is a time effect common to all individuals in period t , and ε_{it} is an individual time-varying error distributed independently across individuals and independently of all η_i and m_t .

Estimation of coefficients in “error-components” models, of which Eq. (5.1) is a special

case, occupies an extensive econometrics literature. Balestra and Nerlove (1966) and Nerlove (1971) discuss a variety of estimation procedures under various assumptions regarding the distributions of η_i and m_i . When η_i and m_i are random components, meaning their distributions are independent of observed right-hand side variables, then conventional generalized least squares produces an estimator that is consistent and asymptotically efficient.²⁰ When the distributions of η_i and m_i depend on right-hand side variables, the literature implements a differencing procedure to calculate consistent estimators, where the form of differencing depends on the particular nature of the simultaneity problems induced by η_i and m_i . Analysts commonly refer to these as “within” estimators because they rely only on variation within groups in calculations. The fixed-effect estimator, which treats η_i and m_i as parameters, is a special case of such an estimator.

5.1.1. *Difference-in-differences estimators*

Suppose both η_i and m_i are believed to be dependent on δ_{it} in some unknown manner, and one wants to compute a consistent estimate of γ in (5.1). A popular version of a within estimator involves first differencing (5.1) over time to obtain

$$\Delta_t y_{it} = \gamma \Delta_t \delta_{it} + \mu_t + \Delta_t \varepsilon_{it}, \tag{5.2}$$

where $\Delta_t y_{it} \equiv y_{it} - y_{i(t-1)}$ and $\mu_t \equiv \Delta_t m_t$. The operator Δ_t differences an individual’s observation across periods, and μ_t is merely defined to be a parameter representing the difference in common time effects.

Suppose, for simplicity, that the sample consists of only two periods: period $t - 1$ which is before the implementation of the policy instrument and period t which is after. Let group e represent the “experimentals”, the individuals who experienced the change in the policy instrument – and let group c denote the “controls” – the individuals who encountered no policy change. Then least squares applied to (5.2) yields the estimators

$$\hat{\gamma} = \Delta_t \bar{y}_t^e - \Delta_t \bar{y}_t^c, \quad \hat{\mu} = \Delta_t \bar{y}_t^c, \tag{5.3}$$

where

$$\Delta_t \bar{y}_t^k = \bar{y}_t^k - \bar{y}_{t-1}^k, \quad k = e, c,$$

$$\bar{y}_t^k = \frac{\sum_{i \in k} y_{it}}{N_k}, \quad k = e, c,$$

where \bar{y}_t^k is the average outcome for group k .²¹

The estimator $\hat{\gamma}$ in (5.3) is identical to what is now known in the literature as the

²⁰ This estimator accounts for the autocorrelation implied by the disturbance $\eta_i + m_i + \varepsilon_{it}$ for an individual, and for correlation across individuals implied by the disturbances m_i .

²¹ The notation $\sum_{i \in k}$ designates that summation is over all individuals included in group k , and N_k is the total number of individuals in group k .

difference-in-difference estimator. The fixed-effect and difference-in-difference estimators do not merely share the same asymptotic distribution; they are computationally identical.

The literature considers many generalizations of fixed-effects models, which in turn imply generalizations of the natural-experiment approach. A common extension incorporates covariates in (5.1) to obtain

$$y_{it} = \gamma\delta_{it} + Z_{it}\theta + \eta_i + m_t + \varepsilon_{it}, \quad (5.4)$$

where Z_{it} includes observed exogenous and/or endogenous variables.²² A further generalization of this model allows for treatment effects to vary randomly across individuals. Under the stringent structural assumptions on time effects and composition highlighted below, the difference-in-differences estimator can be shown to recover the average treatment effect for the treated (i.e., the parameter $E(\gamma \mid \delta_{it} = 1)$). Unfortunately, this parameter is subject to conventional sample selection biases and in general cannot be used to simulate policy responses.

5.1.2. Structural assumptions maintained by the difference-in-difference estimator

Applications of the natural experiment approach typically suggest that it is a “non-structural” estimation procedure, but its equivalence to error-components models clearly indicates that all of the restrictions required for consistent estimation of these models must also hold for the difference-in-difference estimator to measure a behaviorally meaningful parameter. The literature has never interpreted the fixed-effect model as non-structural. The requirement of two sets of structural restrictions are likely to challenge the credibility of many natural-experiment applications concerned with estimating behavioral responses in labor supply.

Assumption 1. Time effects in (5.1) (or (5.4)) must be common across experimentals and controls.

More flexible specifications of (5.1) include the following:

$$y_{it} = \gamma\delta_{it} + \eta_i + m_{ct} + m_{et} + \varepsilon_{it} \quad (5.5)$$

and

$$y_{it} = \gamma\delta_{it} + \lambda_t\eta_i + m_t + \varepsilon_{it}. \quad (5.6)$$

Many factors can lead to these generalizations, including failure to include relevant time-varying variables in Z_{it} that differ across experimentals and controls. Specification (5.5) recognizes that experimentals and controls might experience dissimilar trends and/or cyclical effects. Such an event is likely, for example, when the demographic composition of experimentals and controls differs; empirical analysis usually shows that the trends and

²² Hausman and Taylor (1981) and Amemiya and MaCurdy (1986), for example, develop asymptotically efficient estimators for model (5.4).

cycles differ for married and single people, for men and women, and for high- and low-skilled workers. Specification (5.6) allows individual effects to influence outcomes differentially over time. This phenomenon often happens in analyses of work or wage outcomes over the life cycle. An analysis of the differential time trends, before and after the policy intervention, for each group provides useful information in assessing the reliability of this assumption.

Assumption 2. The composition of both experimentals and controls must remain stable before and after the policy change.

The averages in (5.3) presume that the same individuals make up each group in both period t and period $t - 1$. If this is not the case, then differencing does not eliminate averages of the individual effects η_i . Instead, the terms

$$\Delta_t \eta^e = \bar{\eta}^{e_t} - \bar{\eta}^{e_{t-1}}, \quad \Delta_t \eta^c = \bar{\eta}^{c_t} - \bar{\eta}^{c_{t-1}},$$

with

$$\bar{\eta}^{k_j} = \frac{\sum_{i \in k_j} \eta_{ij}}{N_{k_j}}, \quad k_j = e_t, e_{t-1}, c_t, c_{t-1}$$

contaminate the estimate of γ given by (5.3). Even when the groups e_t and e_{t-1} consist of different individuals, it can still happen that $\Delta_t \eta^e$ vanishes asymptotically keeping $\hat{\gamma}$ consistent. These circumstances typically involve random selection mechanisms. However, selection into groups made up of workers, as is the case in analyses of labor supply, is invariably not random since it depends intricately on the nature of the policy change. For example, a tax change can be expected to alter who works and who does not in a systematic manner. As a consequence, sample selection terms prevent $\Delta_t \eta^e$ from vanishing. Exactly the same problem arises for a shifting composition of the control groups c_t and c_{t-1} , which keeps $\Delta_t \eta^c$ from disappearing.

5.1.3. Grouping estimators

Applications occasionally have grouped data available for their analyses, or they may have a discrete grouping variable (instrument) G_{it} that allocates individuals into $g = 1, \dots, J$ groups of size N_{gt} in each period $t = 1, \dots, T$. A modest modification of fixed-effect model (5.1) (or (5.4)) provides a framework for estimating relevant coefficients in many of these cases. Suppose also that the discrete grouping variable satisfies the assumption

$$y_{it} = \gamma \delta_{it} + \theta_g + \eta_i + m_t + \varepsilon_{it}, \quad (5.7)$$

where θ_g is a time-invariant effect unique to group g and η_i is now an error reflecting the deviation of a particular observation's individual effect around its respective group mean. Defining the group averages

$$\bar{y}_{gt} = \frac{\sum_{i \in g} y_{it}}{N_g}, \quad \bar{\delta}_{gt} = \frac{\sum_{i \in g} \delta_{it}}{N_g}, \quad \bar{\varepsilon}_{gt} = \frac{\sum_{i \in g} \varepsilon_{it}}{N_g},$$

and averaging Eq. (5.7) over groups yields

$$\bar{y}_{gt} = \gamma \bar{\delta}_{gt} + \theta_g + m_t + \varepsilon_{gt}. \quad (5.8)$$

This is just another version of a fixed-effect model, as long as one maintains the structural assumptions for the error components θ_g , m_t and $\bar{\varepsilon}_{gt}$ for grouped data analogous to those outlined in Section 5.1.2 for the components η_i , m_t , and ε_{it} using individual data.

Estimation of model (5.8) – or its variant with the grouped covariates \bar{X}_{gt} also included – involves no complications beyond those already discussed.²³ Under the structural assumptions presumed for the conventional fixed-effect model, differencing eliminates the source of endogeneity for $\bar{\delta}_{gt}$. The quantity $\bar{\delta}_{gt}$ represents the proportion in group g receiving the treatment. The asymptotically efficient estimators developed for model (5.4) apply here as well, with instrumental variables now specified for groups. When there are two groups and when the grouping instrument coincides exactly with the policy reform dummy variable δ_{it} , then this estimator is identical to the difference-in-differences estimator. In any particular application, the objective is to find a suitable grouping instrument such that the resulting grouped error components satisfy the structural conditions of the fixed-effect specification.

5.1.4. Repeated cross-section or panel data?

Since the difference-in-differences estimator and the instrumental variable estimator defined by Eq. (5.3) are expressed in terms of sample means, they can be computed equally well using either repeated cross-section or panel data. Panel data only become useful when the instrumental variable method uses an historic individual variable as an instrument. For example, if past employment status or past tax status is the instrument, then this estimator would typically not be available using cross-section data.

In both the panel data and the repeated cross-section case, the structural conditions are still needed to pursue the difference-in-difference estimator. Provided there is no systematic attrition across groups, panel data allow the groups to be determined in a time-invariant way and, therefore, the difference-in-difference approach completely eliminates the individual fixed effects η_i . Thus, no restrictions need be placed on the distribution of the individual effects. Repeated cross section data, on the other hand, must satisfy the assumption that the unobservable individual effects are drawn from the same population distribution across periods before and after the reform. Otherwise, the difference-in-differences estimator and the instrumental-variable estimator suffer from composition bias. Panel data applications still require the strong restrictions on the distribution of the individual “transitory” time-varying effects and must retain the common-trend assumption.

²³ See also Angrist (1991) and Moffitt (1993).

5.2. Does the difference-in-differences estimator measure behavioral responses?

Most advocates of the natural-experiment approach would answer this question as NO, and they would be right if behavioral responses refers to substitution and income effects familiar in labor supply analyses. Indeed, researchers applying a difference-in-difference procedure often emphasize that they have no intention of estimating such effects.

What, then, is the interpretation of γ in Eq. (5.1) (or Eq. (5.4))? Clearly, under ideal circumstances, γ measures the total response of a policy change, or, more precisely, how a shift in a policy regime influences the average outcome for a worker in the experimental group. But one can seldom translate this response into interpretable behavioral effects because most shifts in policy regimes involve simultaneous changes in marginal wages and net income, and rarely are these changes the same for all individuals making up a group.

To illustrate the issues, reconsider the prototype empirical specification given by Eq. (4.30), which we repeat here for convenience:

$$\ln H_{it} = \alpha \ln W_{it} + \beta Q_{it} + e_{it}. \quad (5.9)$$

Suppose that a policy shift results in changes in $\ln W$ and Q equal to $\Delta_t \ln W$ and $\Delta_t Q$, respectively. A translation of this model into the simple fixed-effect framework,

$$y_{it} = \gamma \delta_{it} + \eta_i + m_t + \varepsilon_{it},$$

is possible by specifying

$$y_{it} = \ln H_{it}, \quad (5.10a)$$

$$\gamma = \overline{\alpha \Delta_t \ln W_{it} + \beta \Delta_t Q_{it}}, \quad (5.10b)$$

$$\eta_i + m_t = \alpha \ln W_{it} + \beta Q_{it}, \quad (5.10c)$$

$$\varepsilon_{it} = e_{it} + \delta_{it} [\alpha \Delta_t \ln W_{it} + \beta \Delta_t Q_{it} - \gamma]. \quad (5.10d)$$

The coefficient γ is the average of $\alpha \Delta_t \ln W_{it} + \beta \Delta_t Q_{it}$ among the experimentals, and the error ε_{it} includes the difference between $\alpha \Delta_t \ln W_{it} + \beta \Delta_t Q_{it}$ and its mean as one of its components. This is one interpretation of the heterogeneous treatment-effects model discussed in Section 5.1.1. Formulation (5.9) assumes that only experimentals experience the change in policy, with $\delta_{it} = 1$ signaling the periods and individuals affected by the change. Specifications (5.9) restrict the permissible variation in W and Q across individuals and time; a form of variation satisfying this property occurs when both $\ln W_{it}$ and Q_{it} can be represented as the sum of an individual and time effect. (Of course, consideration of fixed-effect formulation (5.4) permits some relaxation of these variability restrictions.) The natural-experiment framework requires ε_{it} to be independent of δ_{it} , meaning neither the structural error, e_{it} , nor changes in W and Q provide any information indicating whether an individual is in the experimental group or not.

In this idealized model, the difference-in-difference estimator for γ measures a weighted substitution-income effect given by Eq. (5.10b). We know from our discussion in Section 4 that the interpretation of this combined effect depends on the other control variables included in Q . If one imagines a situation in which Q properly includes a measure of static income or within-period expenditure (such as Eqs. (4.31) or (4.33)), then the substitution effect α corresponds to an uncompensated substitution elasticity. If, on the other hand, Q incorporates age-invariant characteristics controlling for lifetime wages, preferences and initial permanent income (see Section 4.6), then α conforms to the intertemporal substitution elasticity. For still another interpretation, if Q now includes controls for age, initial wealth and the expected wage profile (such as Eq. (4.46)), then α measures the wage elasticity of hours corresponding to a shift in the entire wage profile.

Without, then, carefully specifying the labor-supply model underlying the fixed-effect formulation, it is difficult to know exactly what combination of parameters is being estimated by the natural-experiment approach. Including variables in Q needed for an interpretation of γ invariably implies that one must rely on the generalized fixed-effect specification given by Eq. (5.4), meaning that covariates Z_{it} must be accounted for when calculating the difference-in-difference estimator. In addition to entering specifications directly, the presence of Z_{it} typically alters the formulation of $\Delta_t Q$ which further complicates the interpretation of γ .

Another critical qualification revealed by this attempt to interpret the difference-in-difference estimator involves the requirement that only the experimental group experiences the policy change. If controls also undergo a change at the same time, albeit a different change, then the appropriate specification for Eq. (5.1) becomes

$$y_{it} = \gamma_e \delta_{it} + \gamma_c (1 - \delta_{it}) + \eta_i + m_t + \varepsilon_{it}, \quad (5.11)$$

where γ_e and γ_c represent the behavioral response of the experimentals and controls, respectively.

Such a circumstance would arise, for example, in the case of the 1986 US tax reform. A particular change in the tax code may have directly impacted only a segment of taxpayers (experimentals), but many changes were made to the tax code simultaneously and literally all taxpayers were affected. This would also be the case if there were general equilibrium effects of the policy intervention that affected all wages (or prices) in the economy.

With the term $\gamma_c(1 - \delta_{it})$ present in Eq. (5.11), the difference-in-difference estimator $\hat{\gamma}$ loses its interpretation as a response to any policy change. Fixed-effect estimation of Eq. (5.11) directly can in principle, recover behavioral responses γ_e and γ_c with interpretations analogous to Eq. (5.10), but most formulations imply correlation between δ_{it} and ε_{it} , rendering least squares inconsistent. Such correlation arises when the size of the policy change is systematically different across experimentals and controls, and this occurs almost by definition since it is the nature of the policy change that distinguishes experi-

mentals and controls. With endogeneity induced by this correlation, instrumental-variable procedures must be implemented to estimate (5.11).

5.3. A review of some empirical applications

The empirical applications reviewed here all consider the impact of tax reforms on labor supply. The usual strategy adopted for estimation in these studies is to include the policy dummy with some controls for the wage, other income and demographic variables. As our analysis in Section 5.1 has shown, the interpretation of the estimates from these studies depends on which control variables were included and whether, for the groups chosen, the required assumptions on the unobservable error terms are plausible.

A difference-in-differences estimator was used by Eissa (1995a) to evaluate the effects of the US 1986 Tax Reform Act (TRA) on married women's labor supply. She uses the repeated cross sections of the March Current Population Surveys (CPSs) and compares data from the 1984–1986 surveys just preceding the reform and the 1990–1992 surveys sometime after. Her study compares the behavior of wives married to high-earning husbands (those who were at or above the 99th percentile of the CPS income distribution) to that of wives of lower earning husbands (between the 75th and 80th percentile of the income distribution). The two groups were affected differentially by the 1986 tax reform.

Estimates are provided for both participation and hours. In particular, a reduced form probit equation for participation and an hours equation which included an inverse Mills ratio control for selection were estimated. Demographic variables were entered in the model and some specifications allowed for interactions of the response coefficient with education level. These adjustments were found to significantly reduce the elasticity estimates. The reported wage elasticities for hours were between 0.6 and 1 while, for participation, elasticity estimates were surprisingly smaller – between 0.1 and 0.6. The choice of grouping is controversial since it might be thought that, even given the observed included controls, husband's income is not exogenous for the change in his spouse's labor supply. Moreover, given the increasing dispersion of incomes and wages among all groups during that period, the common time effects (common trends) assumption among the unobservable components across the two groups may not be satisfied.

Eissa's approach was also followed in a recent panel data study of the 1987 Danish tax reform by Graversen (1996). He considered the participation and hours worked of women, split according to marital status. In both cases, for controls he used a group for which predicted tax rate changes were small, using pre-reform hours and wages on the post-reform tax parameters. No exclusion restrictions appear to have been used to identify the selection term. For the difference-in-differences estimates with no controls for observable individual characteristics, he found perversely signed effects, but including numbers and age of children, for example, resulted in small but positive responses. This sensitivity of the difference-in-differences parameter estimates to the inclusion of observable time-varying characteristics is indicative of the importance of the conditions placed on the distribution of unobservables within each group over time.

Eissa and Liebman (1995) focused on the effects of TRA and EITC on single women with children. Again, they used data from the March CPSs for the US. Their identification strategy was to compare the change in labor supply for women with children to the change in labor supply for women with no children. They found that the participation of single women with children increased by 1.9–2.8 percentage points relative to single women with no children (from a base of 73%). Eissa and Lieberman also found a rather more surprising result that the EITC expansion in the Tax Reform Act had no perceptible effect on the hours of work of single women with children who were already in employment. The use of women with no children as a control group is open to criticism on a number of grounds. First, the conditions on the time and composition effects among the unobservables is unlikely to be satisfied in the repeated cross-sections of the CPS, even given the included regressors. Second, women with no children are probably working closer to their upper bound, as far as participation is concerned, and would not, therefore, be expected to increase participation. This is really a failure of the common trends assumption since such women may not, therefore, be able to absorb an upward common trend to labor supply on the participation margin.

Blundell et al. (1998b) consider the use of the sequence of tax reforms in the UK over the 1980s and early 1990s to study the hours responses of married women from a long time series of repeated cross-sections. A semi-log linear labor supply equation (see Eq. (4.54)) was specified with additive controls for other income, children, education and birth cohort. In contrast to the other studies discussed in this section, the hours equation included the log of the post-tax hourly wage rate and other income as well as a number of demographic controls. Other income was defined by the difference between consumption and the product of hours worked and the post-tax marginal hourly wage. This definition of other income is consistent both with intertemporal two-stage budgeting in the absence of liquidity constraints and with the presence of liquidity constraints as described in Section 4 above. The estimated labor supply model allowed the demographic variables to interact with the log wage and other income variables.

Two alternative estimators were considered. The first was a difference-in-differences estimator that grouped the sample by taxpayers and non-taxpayers. This was argued to be invalid because, under very general conditions, the composition of the two groups could be expected to change in a non-random way in response to the tax reforms. The second approach grouped by education and age cohort. This exploited the systematically changing distribution of wages by education and cohort group in the UK described in Section 3.4. The idea was that the differential growth in wages across birth cohorts by education group reflects changes in the demand for labor, possibly due to skill-biased technical change, and could be excluded from the labor supply equation. The log marginal hourly wage, which was included directly in the labor supply specification regression, together with the other income variable and participation in work, were treated as endogenous. The estimator can, thus, be interpreted as a (grouping) instrumental variable estimator in which the changes in the demand for the different skills of each education and cohort group are assumed to be exogenous and validly excluded from labor supply given the inclusion of the wage and

income variables. The education and cohort interactions with time, which were the excluded instruments, were found to be jointly significant in the wage and other income reduced forms.

The reported uncompensated labor supply elasticities, although small, were all positive and highest for women with children of pre-school age. The income elasticities were all negative, except for those women with no children, for whom they were essentially zero. As a result, the compensated wage effects, which matter for welfare, were all positive and the model was found to be consistent with standard theory everywhere in the data. In comparison, the estimates that use taxpayer status as a grouping instrument showed a significant *negative* wage elasticity. This negative estimate was argued to reflect the systematic change in the composition of the taxpaying group. During the period considered, there were many new entrants into this taxpaying group who had systematically lower hours. This non-random change in composition invalidates the second assumption of Section 5.1.2 on the composition of groups across time.

A number of additional experiments were reported that varied the control variables and instruments. In one experiment the time effects and the cohort/education effects were excluded. Only age and age squared were entered along with the demographic variables, the log marginal wage and the other income variable. This makes the specification similar to that in traditional cross section studies, such as those reviewed by Mroz (1987), except that the data contain a large number of time periods. Both the other income and wage elasticities became much larger. The resulting estimators are similar to those reported in the Arellano and Meghir (1992) study on the UK where education was used as the identifying instrument. As in the Eissa (1995b) study, controlling for education in the labor supply equation has the property of reducing the wage elasticity.

Eissa (1996) considers the case of labor supply responses to the sequence of tax reforms during the 1980s in the US. As in her previous studies, she uses March CPS data but this time over a longer period – 1976 to 1993. Moreover, the grouping was by education level. She finds only weak evidence of an increase in male labor supply in response to the Tax Reform Act. This poses an interesting issue relating to the larger effects on taxable income that have been found in studies that use tax-return data directly such as Feldstien (1995).²⁴

6. Estimation with non-participation and non-linear budget constraints

To analyze how tax and welfare policies influence hours of work, there has been a steady expansion in the use of sophisticated statistical models characterizing distributions of discrete-continuous variables that jointly describe work and program participation. Considered at the forefront of research in this area, these models offer a natural mechan-

²⁴ This could be reconciled if it can be shown that certain groups of individuals respond to tax reforms on other margins.

ism for capturing the institutional features of both tax and welfare programs. This section describes how to estimate the effect of these programs on labor supply using these models.

These models build on standard approaches for dealing with censored and missing data. The basic principles underlying these approaches are well documented in the econometrics literature. In what follows, we provide explicit details on applying these methods to incorporate fixed costs, missing wages and discrete program participation in models of labor supply behavior with taxes and welfare.

6.1. Basic economic model with taxes

Consider a model of static labor supply where individuals determine hours of work and consumption by maximizing a utility function $U(C,h)$ subject to the budget constraint

$$C = Wh + Y - \tau(I), \quad (6.1)$$

where C is the consumption, W is the gross wage/h, h is the hours of market work, Y is the non-labor income, τ is the taxes determined by the function $\tau(\cdot)$, I is the taxable income per year, $I = Wh + Y - D$, and D is the deductions per year.

Due to different marginal tax rates in the various income brackets combined with the existence of non-labor income, the budget set is inherently non-linear in most instances. The literature applies two approaches for modeling the non-linearities induced by taxes: piecewise-linear functions that recount the brackets making up tax schedules; and smooth differentiable relations that summarize the tax rates implied by bracketed schedules. This section outlines each of these approaches, along with the procedures implemented to estimate labor-supply parameters associated with each approach.

In the absence of taxes, maximization of the utility function subject to the budget constraint defines the labor supply function

$$h = f(W, Y, v), \quad (6.2)$$

where v is an error reflecting the contribution of factors relevant to economic agents and unobserved by the econometrician.²⁵ With W and Y reinterpreted as “after-tax” measures, the construction of which is presented below, f continues to describe hours-of-work behavior even when complex non-linearities affect budget constraints, as is the case with taxes. The objective of most labor-supply analyses is to estimate the parameters of the function f .

6.1.1. Structure of taxes

The institutional features of income and program taxes occupy a great deal of attention in

²⁵ It is straightforward to replace Eq. (6.2) by $f(W, Y, X, v)$ where X is a vector incorporating measured variables affecting agents' choices. We suppress X for notational convenience. f , of course, depends on a parameter vector which we also suppress.

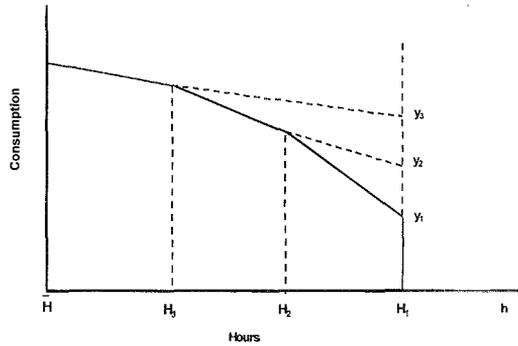


Fig. 15. Budget constraint with income taxes.

the labor supply literature. As described in Section 2, the complexities introduced by the US and the UK tax system, for example, contort the budget constraint faced by a typical worker. Modeling this constraint is often thought to be essential in labor supply analysis for capturing the opportunities available to individuals. For example, the overall tax schedule in the US consists of five components:

$$\tau(Y, E) = \text{FEDTX} + \text{STATX} + \text{EITC} + \text{SSTAX} + \text{WELFARE}, \tag{6.3}$$

where $\tau(Y, E)$ is the overall tax schedule, E is earned income, FEDTX is the federal income tax schedule, STATX is the state income tax schedule, EITC is the earned income tax credit schedule, SSTAX is the social security tax schedule, and WELFARE is net transfers from public assistance programs.

Each of these schedules has its own method of computing “taxable” income, but all in some way base calculations on a distinction between Y and E . We ignore these considerations here. Both federal and state income tax schedules compute taxes based on income brackets, which induces piecewise linear budget constraints. The other programs are applicable over only part of the income range which also creates brackets.

6.1.2. Piecewise linear constraints

Fig. 15 shows a hypothetical budget constraint for an individual in the US faced with federal income taxes alone, state income taxes alone, or both.²⁶ In this diagram, h denotes hours of work, and “Consumption” denotes total after-tax income or the consumption of market goods. The segments of the budget constraint correspond to the different marginal tax rates that an individual faces. In particular, he faces a tax rate of t_A between H_0 hours and H_1 hours (segment 1) and tax rates of t_B and t_C , respectively, in the intervals (H_1, H_2) and (H_2, \bar{H}) (segments 2 and 3). Thus, the net wages associated with each segment are:

²⁶ Note that $h = H_0 = 0$ corresponds to 0 h of work. As we move from right to left in these figures, hours of work increase.

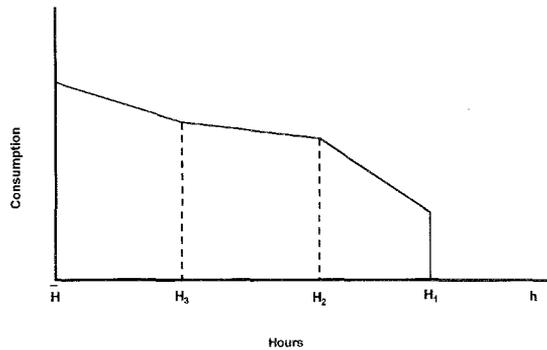


Fig. 16. Budget constraint with EITC.

$w_1 = (1 - t_A)W$ for segment 1, $w_2 = (1 - t_B)W$ for segment 2, and $w_3 = (1 - t_C)W$ for segment 3. Virtual income for each segment (i.e., income associated with a linear extrapolation of the budget constraint) is calculated as: $y_1 = Y - \pi(Y, 0)$; $y_2 = y_1 + (w_1 - w_2)H_1$; and $y_3 = y_2 + (w_2 - w_3)H_2$. Changes in tax brackets create the kink points.

Fig. 16 shows a budget constraint affected only by the EITC schedule,²⁷ and Fig. 17 shows a budget constraint that reflects the effects of the social security tax alone.²⁸ As seen in Fig. 18, welfare benefit programs create a budget set that resembles the one for Social Security. All of these taxes induce non-convexities in opportunity sets.

Summing these various tax components creates an overall tax-transfer schedule with two noteworthy features. First, the schedule faced by a typical individual includes a large number of different rates. Translated into the hours-consumption space, this implies a large number of kink points in the budget constraint. Second, for most individuals the tax schedule contains non-convex portions, which arise from four potential sources. The first arises from a fall in the EITC tax rate at the break even point. In Fig. 16, that point occurs at H_2 where the tax rate falls from a positive value to zero. The second source occurs when the social security tax hits its maximum (at H_1 in Fig. 17), where the corresponding tax rate goes from a positive value to zero. A third source is the non-convexity introduced by the structure of the standard deduction. Finally, if a worker's family participates in any welfare program, then significant non-convexities arise as benefits are withdrawn when earnings increase.

²⁷ The EITC is a negative income tax scheme which can induce, in the simplest case, two kinks in a person's constraint: one where the proportional credit reaches its maximum (H_1 in Fig. 16), and one at the break even point where the credit is fully taxed away (H_2 in the figure). The tax rates associated with the first two segments are t_A , which is negative, and t_B , which is positive. Thereafter, the EITC imposes no further tax.

²⁸ The social security tax is a proportional tax on earnings up to a specified earnings level, after which the amount of tax paid is the same regardless of earnings. As a result, Fig. 17 shows a constraint with a single interior kink (given by H_1 in the figure) corresponding to the maximum proportionally taxed earnings level. The tax rate on the segment leading up to that kink is t_A , switching to zero on the second segment.

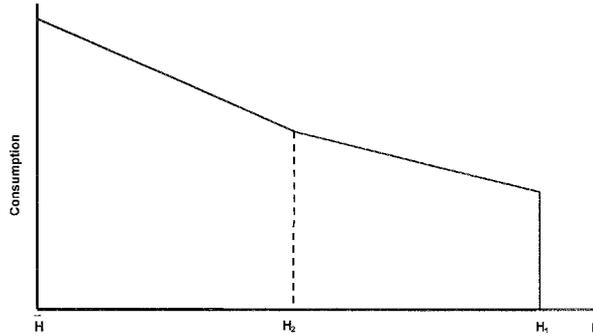


Fig. 17. Budget constraint with Social Security Tax.

6.1.3. Constructing a differentiable constraint

Approximating the tax schedule by a differentiable function leads to a simple approach for developing an empirical model of labor supply that recognizes the influence of taxes. A convenient approach for constructing this function is to approximate the marginal tax rate schedule – a step function – by a differentiable function. This approximation must itself be easily integrable to obtain a simple closed form for the tax function.

An elementary candidate for constructing a differentiable approximation that can be made as close as one desires to the piecewise-linear tax schedule has been applied in MaCurdy et al. (1990). To understand the nature of the approximation, return to Fig. 15. One can represent the underlying schedule as follows:

$$\begin{aligned}
 \tau'(I(h)) &= t_A && \text{from } I(H_0) \text{ to } I(H_1) \\
 &= t_B && \text{from } I(H_1) \text{ to } I(H_2) \\
 &= t_C && \text{above } I(H_2),
 \end{aligned}
 \tag{6.4}$$

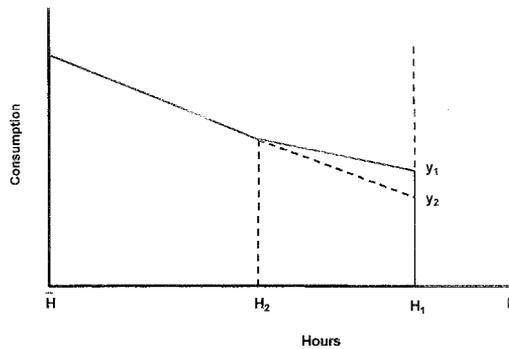


Fig. 18. Budget constraint with welfare.

where $\tau'(I(h))$ is the marginal tax rate, $I(h)$ is taxable income at h hours of work, and t_i is the marginal tax rate, $i = A, B, C$. For expositional simplicity, suppose that $t_A = 0$. Consider the following approximation of this schedule which uses three flat lines at the heights $t_A (= 0)$, t_B and t_C and weight functions parameterized to switch the three lines on and off at appropriate points:

$$\hat{\tau}'(I(h)) = t_B[\Phi_1(I(h)) - \Phi_2(I(h))] + t_C[\Phi_2(I(h))], \tag{6.5}$$

where the weight functions are given by $\Phi_i(I(h)) =$ the cumulative distribution function with mean μ_i and variance σ_i^2 , $i = 1, 2$. The middle segment of the tax schedule has height t_B and runs from taxable income $I(H_1)$ to $I(H_2)$. To capture this feature, parameterize $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ with means $\mu_1 = I(H_1)$ and $\mu_2 = I(H_2)$, respectively, with both variances set small. The first distribution function, $\Phi_1(\cdot)$ takes a value close to zero for taxable income levels below $I(H_1)$ and then switches quickly to take a value of one for higher values. Similarly, $\Phi_2(\cdot)$ takes a value of zero until near $I(H_2)$ and one thereafter. The difference between the two equals zero until $I(H_1)$, one from $I(H_1)$ to $I(H_2)$ and zero thereafter. Thus, the difference takes a value of one just over the range where t_B is relevant. Notice that we can control when that value of one begins and ends by adjusting the values μ_1 and μ_2 . Also, we can control how quickly this branch of the estimated schedule turns on and off by adjusting the variances of the cumulative distribution functions, trading off a more gradual, smoother transition against more precision. In general, adjusting the mean and variance parameters allows one to fit each segment of a schedule virtually exactly, switch quickly between segments, and still maintain differentiability at the switch points.

A generalization of this approximation takes the form

$$\hat{\tau}'(I(h)) = \sum_{i=1}^k [\Phi_i(I(h)) - \Phi_{i+1}(I(h))]b_i(I(h)), \tag{6.6}$$

where the functions $b_i(I(h))$ are polynomials in income. With the Φ_i denoting normal c.d.f.s, function (6.6) yields closed form solutions when it is either integrated or differentiated.²⁹ The resulting approximation can be made to look arbitrarily close to the budget constraint drawn in Fig. 15, except that the kink points are rounded.

6.2. Instrumental-variable estimation

Conventional non-linear instrumental-variable procedures offer a robust method for estimating particular forms of the labor-supply function f in Eq. (6.2), forms that permit the specification of structural equations that are linear in all sources of disturbances. As discussed in Section 4, the development of such specifications is a substantial challenge

²⁹ Total taxes are given by: $\tau(I) = \int \tau'(I)dI$. The following relations enable one to calculate an explicit form for $\tau(X)$: $\int \Phi dI = I\Phi + \varphi$, $\int I\Phi dI = (1/2)I^2\Phi - (1/2)I\varphi + (1/2)I\varphi$, $\int I^2\Phi dI = (1/3)I^3\Phi - (2/3)\varphi + (1/3)I^2\varphi$, $\int I^3\Phi dI = (1/4)I^4\Phi - (3/4)I\varphi + (3/4)I\varphi + (1/4)I^3\varphi$. In this expression, Φ refers to any Φ_i 's, and φ designates the density function associated with Φ_i .

for it proves difficult to discover a preference map that produces additivity in structural disturbances – errors reflecting unobserved differences among people (heterogeneity) – while at the same time permitting measurement errors in hours and wages to enter linearly.

6.2.1. A useful characterization of labor supply with taxes

The introduction of a non-linear tax schedule into a model of labor supply poses few analytical difficulties when the schedule generates a strictly convex constraint set with a twice-differentiable boundary. Utility maximization in this case implies a simple characterization of the hours-of-work choice.

With τ denoting the smooth function that approximates the tax schedule, specify the marginal wage rate and “virtual” income as

$$\omega = \omega(h) = (1 - \tau')W,$$

$$y = y(h) = Y + E - \tau - \omega h = Y + \tau'Wh - \tau = C - \omega h, \quad (6.7)$$

where $E = Wh$ is gross earnings, and τ and τ' (the derivative of the tax function with respect to income) are evaluated at income level $I = I(h) = Y + Wh - D$ which directly depends on the value of h . In Eq. (6.7) we write the marginal wage $\omega = \omega(h)$ and virtual income $y = y(h)$ as functions to emphasize their dependence on hours h .

Utility maximization implies a solution for hours of work that obeys the implicit equation

$$h = f(\omega(h), y(h), v), \quad (6.8)$$

where we write the marginal wage $\omega = \omega(h)$ and virtual income $y = y(h)$ as functions to emphasize their dependence on hours. Figs. 15 and 18 illustrate this representation of the solution for optimal hours of work. This characterization follows from work on taxes and labor supply (e.g., Hall, 1973) that represents a consumer as facing a linear budget constraint in the presence of non-linear tax programs. This linear constraint is constructed in a way to make it tangent to the actual non-linear opportunity set at the optimal solution for hours of work. The implied slope of this linearized constraint is $\omega(h)$ and the corresponding value of virtual income is $y(h)$. Eq. (6.8) constitutes a structural relationship that determines hours of work. By applying the Implicit Function Theorem to specification (6.8), we can solve this implicit equation for h in terms of W , Y , and other variables and parameters entering the functions τ and f . This operation produces the labor supply function applicable in the non-linear tax case.

6.2.2. A structural equation of labor supply with taxes

Relation (6.8) directly provides the basis for formulating a structural equation that can be estimated by standard instrumental-variable procedures. Consider, for example, the semi-log specification³⁰:

³⁰ See Eq. (A.5) in Appendix A.

$$h = f(\omega, y, v) = \mu + Z\gamma + \alpha \ln \omega + \beta y + v, \quad (6.9)$$

where μ , γ , α , and β are parameters, Z is a vector of observed determinants of labor supply (e.g., age, family size, etc.), and v is a structural disturbance capturing unobserved factors influencing hours-of-work decisions. The marginal after-tax wage ω enters this specification in a natural log, so α represents a hybrid of an uncompensated substitution effect and elasticity. The coefficient β corresponds to an income effect.

Conventional instrumental-variable procedures offer a robust method for estimating the coefficients of the semi-logarithmic specification of the labor supply function given by Eq. (6.9). In the absence of measurement error, inspection of Eq. (6.9) reveals that the error term enters linearly into the specification. Consequently, variables that are orthogonal to the structural disturbance v can serve as instruments for estimating the parameters determining substitution and income effects. The implementation of such procedures imposes no parametric restrictions and it allows one to consider a wide variety of exogeneity assumptions.

In many data sets there are serious suspicions that hours of work and wages are reported with error. Suppose h^* denotes measured hours of work and that the function $h^*(h, \varepsilon)$ relates h^* to actual hours, h , and to an error component, ε . An interesting specification for characterizing the form of reporting error is given by the multiplicative structure:

$$h^* = h^*(h, \varepsilon) = h e^\varepsilon, \quad \text{with } W^* = E/h^*, \quad (6.10)$$

where ε is distributed independently of h and v , and the distribution of the measurement error component ε satisfies the moment condition $E(e^\varepsilon) = 1$, implying that h^* and h have the same expected value. Whereas $W = E/h$ defines the true hourly wage rate, W^* designates the data available on wages. This formulation presumes not only measurement error in hours, but also the existence of reporting error in hourly wage rates due to its construction. Assuming accurate observations on E , calculating wages by dividing total labor earnings by reported hours induces a reciprocal relation in the measurement error linking data on hours and wages.

Incorporating the multiplicative measurement error model given by Eq. (6.10) into the semi-logarithmic specification of labor supply presented in Eq. (6.9) yields the empirical relations

$$h^* = \tilde{\mu} + Z\gamma + \alpha \ln \omega^* + \beta y + u, \quad (6.11)$$

where

$$\ln \omega^* = \ln(E/h^*) + \ln(1 - \tau'),$$

$$\tilde{\mu} = \mu - \alpha \sigma_\varepsilon^2 / 2, \quad (6.12)$$

$$u = v + \alpha(\varepsilon + \sigma_\varepsilon^2 / 2) + (h^* - h) = v + \alpha(\varepsilon + \sigma_\varepsilon^2 / 2) + h(e^\varepsilon - 1).$$

Relation (6.7) continues to define the variable y . This virtual income quantity and the

marginal tax rate τ' are not contaminated by measurement error because they are functions of Y , E and τ' which are known without errors. The variable $\ln \omega^*$ represents the natural logarithm of the after-tax wage rate evaluated at observed hours, which differs from the actual marginal wage due to the presence of reporting error in hours. The disturbance u possesses a zero mean since $E(\varepsilon) = -\sigma_\varepsilon^2/2$, $E(e^\varepsilon) = 1$, and the error ε is distributed independently of all endogenous components determining h , including the heterogeneity disturbance v .

Interpreting relation (6.11) as a structural equation describing labor supply, instrumental-variable methods continue to offer a flexible scheme for consistent estimation of substitution and income parameters. Due to the heteroscedasticity of the disturbance u in Eq. (6.11), the estimation procedure must compute robust standard errors to produce valid test statistics. For consistent estimation of the parameters of Eq. (6.11), one needs to be able to identify a set of variables X that are orthogonal to the structural disturbance v , independent of measurement error ε , and are capable of predicting the endogenous variables ω^* and y . Selecting alternative formulations of X offers the opportunity to entertain a variety of exogeneity assumptions, even with measurement error present, thereby indicating the direction of potential biases in estimated work disincentive effects arising from these assumptions. The maximum-likelihood approaches discussed below typically maintain that all sources of income are exogenous determinants of work hours, including W , and Y (i.e., the gross wage rate, non-taxable non-labor income, and non-labor taxable income). Judicious inclusion and exclusion of these income sources in X provides a basis for judging whether endogeneity of wages and/or incomes is a problem. Of course, the ability to test these exogeneity assumptions critically relies on there existing a sufficient number of elements in X that satisfy exclusion restrictions in Eq. (6.11).

6.2.3. Lifecycle considerations

As outlined in Section 4, substituting an alternative measure for the variable y in Eq. (6.9) creates a labor-supply specification that is consistent with decision-making in a lifecycle context, and this specification can in turn be modified to account for the existence of income taxes.³¹ In a static analysis with taxes, one specifies virtual income as

$$y = Q - \omega h, \quad \text{with } Q = Y + E - \tau, \quad (6.13)$$

where the income components making up the quantity Q represent current income in the period. In such an analysis $Q = C$ by assumption. However, in an intertemporal setting it need not be the case that $y = Y + E - \tau - \omega h$ since one can have $C - \omega h \neq Y + E - \tau - \omega h$ due to saving or borrowing. As shown in Section 4, in a multiperiod framework, with or without an uncertain future, the construction of the virtual income variable y (or, more precisely, the quantity Q) must account for net savings in the period.

Given the availability of data for each family's total consumption, a formulation for Q that obviously accomplishes this task is to set

³¹ For details beyond the discussion presented in this subsection, see MaCurdy (1983).

$$y = Q - \omega h, \quad \text{with } Q = C \quad (6.14)$$

as the measure of virtual income. Given this construction for y , and lifetime utility maximization with strongly separable preferences over time, the function $h = f(\omega, y, v)$ characterizes the optimal lifecycle choice of labor supply in the period under consideration.

Standard two-stage least squares procedures continue to provide a computationally simple method for consistently estimating the parameters of the function f , assuming, of course, that the empirical specification of f is linear in disturbances—such as specification (6.9). One can apply linear or non-linear instrumental-variable procedures to estimate coefficients depending on whether the specification of f is linear or non-linear in parameters, with robust standard errors computed when appropriate.

6.3. Maximum likelihood: convex differential constraints with full participation

Maximum-likelihood estimation of labor-supply models with a tax schedule described by a twice-differentiable boundary implying a convex budget set poses few difficulties. Provided the gross wage variable and the other income variable are assumed free from measurement error and independent of unobserved heterogeneity, such an estimation approach need not heavily rely on exclusion restrictions to identify parameters. In contrast to the case when implementing instrumental-variable procedures, even though marginal wages and virtual incomes are endogenous, non-linearities introduced through distributional assumptions provide a valuable source of identification. Because exclusion restrictions are often difficult to justify, many researchers turn to maximum likelihood to avoid making ad hoc exclusion properties. Of course, the independence assumptions on the distribution of unobserved heterogeneity in these maximum likelihood approaches are strong and are precisely what is being relaxed in the fixed effects models that underlie the difference-in-differences and related approaches outlined in Section 5.

6.3.1. Specification of likelihood functions with multiplicative measurement error

Considering maximum-likelihood estimation of the model analyzed in Section 6.2, suppose the heterogeneity-error-component v in the labor-supply function (6.9) and the disturbance ε in the measurement-error equation (6.10) for hours of work is independent of the gross wage and other income and possesses the joint distribution: $(v, \varepsilon) \approx g_{v\varepsilon}$, where $g_{v\varepsilon}$ designates a density function. Using relations (6.9) and (6.10) to perform a standard change in variables from the errors v and ε to the variables h and h^* produces the likelihood function needed to compute maximum-likelihood estimates. The transformation from (v, ε) to (h, h^*) is monotonic for a wide range of functional forms for f as long as the underlying preferences satisfy quasiconcavity and budget sets are convex.

Without measurement error, the likelihood function for hours of work, h , takes the form

$$l = \frac{dv}{dh} g_v(h - \mu - Z\gamma - \alpha \ln W - \alpha \ln(1 - \tau') - \beta y), \quad (6.15)$$

where g_v is the marginal density for v , and the Jacobian term is

$$\frac{dv}{dh} = 1 + \left(\left(\frac{\alpha}{W(1 - \tau')} - \beta h \right) W^2 \frac{\partial \tau'}{\partial I} \right), \tag{6.16}$$

which is required to be non-negative. In Eqs. (6.15) and (6.16), the derivative τ' is evaluated at $I = Wh + Y - \pi(Wh + Y)$.

With multiplicative measurement error, the likelihood function for observed hours h^* becomes

$$l = \int_0^{\text{maxwage}} \int_0^{\text{maxhours}} \frac{dv}{dh} g_{\nu\epsilon}(\ln h^* - \ln h, h - \mu - Z\gamma - \alpha \ln \omega - \beta y) \psi(W) dh dW, \tag{6.17}$$

where integration occurs over the hourly wage, which is unobserved, using its density $\psi(W)$.^{32,33} The non-negativity of the Jacobian term clearly places restrictions on the behavioral parameters and we discuss these restrictions further below.

6.3.2. A popular linear empirical specification

One of the most numerous commonly applied empirical specification for labor supply implemented in maximum likelihood analyses – particularly those using the piecewise-linear approach discussed below – takes the linear form:

$$h = f(\omega, y, v) = \mu + \alpha\omega + \beta y + Z\gamma + v \equiv \hat{h} + v, \tag{6.18}$$

where the unobserved error component v represents heterogeneity in preferences with $v \approx g_{\ν}$, where g_{ν} denotes the marginal density of ν . In conjunction with this specification, analyses also presume measurement error in hours of work possessing the classical

³² This likelihood function fundamentally differs from the one proposed in Eq. (D.5) in Appendix D of MaCurdy et al. (1990). The particular form of the labor-supply model considered in MaCurdy et al. (1990) is

$$(a) \ h^* = \tilde{\mu} + Z\gamma + \alpha\omega^* + \beta y + u$$

with $h^* = he^{\epsilon}$. The analog of (6.17) for this linear specification is

$$(b) \ l = \int_0^{\text{maxwage}} \int_0^{\text{maxhours}} \frac{dv}{dh} (\ln h^* - \ln h, h - \mu - Z\gamma - \alpha\omega - \beta y) \psi(W) dh dW$$

with $(dv/dh) = 1 + (\alpha - \beta h)W^2(\partial\tau'/\partial I)$. Likelihood function (b) is the valid specification for estimating model (a), whereas likelihood function (D.5) presented in MaCurdy et al. is not – unbeknownst, unfortunately, to the authors of MaCurdy et al. Specification (D.5) of MaCurdy et al. implicitly conditions on the true wage rate W , even though earnings, E , rather than W appears in (D.5). W is an unobserved variable in the analysis and, therefore, must be integrated out of (D.5) to obtain a valid formulation. Specifications (b) and (6.17) incorporate this integration. MaCurdy recognized this oversight when reconciling some Monte Carlo findings done by Lennart Flood during his visit to Stanford in 1996; Lennart’s assistance in revealing this problem is gratefully acknowledged.

³³ If W is not independent of ν and ϵ , then (6.17) is replaced by

$$l = \int_0^{\text{maxwage}} \int_0^{\text{maxhours}} \frac{dv}{dh} g_{\nu\epsilon W}(\ln h^* - \ln h, h - \mu - Z\gamma - \alpha \ln \omega - \beta y, W) dh dW,$$

where $g_{\nu\epsilon W}$ is the joint density of ν , ϵ , and W .

linear functional form

$$h^* = h^*(h, \varepsilon) = h + \varepsilon, \quad (6.19)$$

where $\varepsilon \approx g_\varepsilon$, with ε and v independent. The measurement error component ε represents reporting error that contaminates observations on h for individuals who work.

The derivation of likelihood functions for this case is straightforward given the assumptions about preferences and budget constraints maintained to this point. Assuming no measurement error (i.e., $h^* = h$), a change in variables from the heterogeneity error v to actual hours h using relation (6.18) yields the likelihood function for h :

$$g^h(h) = \frac{dv}{dh} g_v(h - \mu_v - Z\gamma - \alpha w - \beta y), \quad (6.20)$$

where the Jacobian term is

$$\frac{dv}{dh} = 1 + (\alpha - \beta h)W^2 \frac{\partial \tau'}{\partial I}. \quad (6.21)$$

This Jacobian term is restricted to be non-negative over the admissible range. Maximizing (6.20) yields maximum-likelihood estimates for the parameters of the labor supply function f , which provide the information needed to infer the work disincentive effects of taxation.

If hours are indeed contaminated by additive measurement error, then the likelihood function for observed hours $h^* = h + \varepsilon$ is given by:

$$g_{h^*}(h^*) = \int_0^{\text{maxhours}} g_\varepsilon(h^* - h)g_h(h)dh. \quad (6.22)$$

This expression resembles relation (6.20) except that integration occurs over hours to account for the existence of reporting error, and h^* replaces actual hours h in the Jacobian term in (6.17).

6.3.3. Imposition of behavioral restrictions with differentiable constraints

The implementation of maximum likelihood procedures imposes interesting and important restrictions on behavioral parameters in the presence of non-linear budget constraints. Consider, for example, likelihood function (6.22). For this specification to be a properly-defined likelihood functions, the Jacobian (6.21) must be non-negative. Violation of this condition implies that the density function for h is negative, which obviously cannot occur. Relation (6.20) indicates that this non-negativity condition translates into the property

$$\frac{\partial h^s}{\partial w} - \frac{\partial h^s}{\partial y} h \geq -\left(\frac{\partial \tau}{\partial I} W^2\right)^{-1} \leq 0, \quad (6.23)$$

where $h^s (=f)$ refers to the labor supply function. The left-hand side of this inequality is the Slutsky term. This inequality result does not require compensated substitution effects to be

positive as quasi-concave preferences mandate, only that these effects cannot become too negative.

Maximum likelihood procedures yield nonsensical results unless Eq. (6.23) holds. Without measurement error, estimated parameter values cannot imply a violation of Eq. (6.23) at any of the data combinations $(h, w(h), y(h))$ actually observed in the sample. If a violation occurs, then the evaluation of Eq. (6.22) for the observation associated with this combination would result in a non-positive value which causes the overall log likelihood function to approach minus infinity—which clearly cannot represent a maximum. With measurement error, maximum likelihood estimation applied to Eq. (6.22) ensures that a weighted average of Eq. (6.22) holds, with weighting occurring over all combinations of hours, marginal wages, and virtual income lying in the feasible range of the budget constraint of any individual included in the sample. Since maximum likelihood procedures assume the validity of such restrictions when calculating estimates of the coefficients of h^s , the resulting estimated labor supply function can be expected to exhibit compensated substitution effects that obey inequality (6.23) over a very wide range of hours, wages, and incomes.³⁴ Section 6.4.3 revisits these restrictions, relating them to those invoked in cases when maximum likelihood is used with non-differentiable (piecewise-linear) tax functions.

6.4. Maximum likelihood: convex piecewise-linear constraints with full participation

The majority of empirical labor-supply studies incorporating taxes treat the tax schedule as a series of brackets implying a piecewise-linear budget set. With such a tax function, the familiar change-in-variables techniques implemented in conventional maximum likelihood do not apply due to the non-existence of the Jacobian over measurable segments of the sample space, which occurs since the functional relationships characterizing hours-of-work choices are not differentiable. Moreover, a piecewise-linear budget set creates endogenous variables (hours and after-tax wages) that are both discrete and continuous in character, complicating the use of instrumental-variable procedures, which require the inclusion of sample-selection terms in equations to produce disturbances with zero means.

6.4.1. Characterization of labor supply with piecewise-linear constraints

To illustrate the derivation of an estimable labor supply model using the piecewise-linear approach for the model described in Section 6.1.2., consider the simple case of a budget set with only three segments as presented in Fig. 15. The preceding discussion defines the variables y_j , ω_j , and H_j appearing in this figure. To locate the kinks and slopes of the budget constraint for an individual, a researcher must know the individual's level of non-labor

³⁴ It is, of course, computationally feasible to use (6.22) in estimation and not require g_h to be defined over the entire range of its support. Computationally one merely requires g_h to be non-negative over a sufficiently large region to ensure $(6.22) > 0$. Of course, not requiring $g_h \geq 0$ over its relevant range produces a nonsensical statistical model.

income, gross wage rate, hours of work, and the structure of the tax system. The hours of work at which kinks occur are given by $H_j = (I_j - Y + D)/W$, where Y and D , respectively, represent taxable non-labor income and deductions, and I_j is the maximum taxable income for segment j . The slope of each segment is given by the marginal wage rate for that segment: $\omega_j = W(1 - t_j)$, where j denotes the segment, t_j signifies the marginal tax rate for that segment, and W is the gross wage rate/h. Finally, the non-labor income at zero hours of work – the intercept of the budget line – is $y_1 = V + Y - \tau(Y - D)$, where $\tau(\cdot)$ is the tax function evaluated at the individual's taxable income at zero earnings. Given this intercept value, virtual incomes or the intercepts associated with successive budget segments are computed by repeated application of the formula: $y_j = y_{j-1} + (\omega_{j-1} - \omega_j)H_{j-1}$.

Given a convex budget constraint, an individual's optimization problem amounts to maximizing $U(C, h)$ subject to

$$\begin{aligned}
 C &= y_1 && \text{if } h = 0, \\
 &= \omega_1 h + y_1 && \text{if } H_0 < h \leq H_1, \\
 &= \omega_2 h + y_2 && \text{if } H_1 < h \leq H_2, \\
 &= \omega_3 h + y_3 && \text{if } H_2 < h \leq \bar{H}, \\
 &= \omega_3 \bar{H} + y_3 && \text{if } h = \bar{H},
 \end{aligned} \tag{6.24}$$

The solution of this maximization problem decomposes into two steps. First, determine the choice of h conditional on locating on a particular segment or a kink. This step yields the solution

$$\begin{aligned}
 h &= 0 && \text{if } h = 0 && \text{(lower limit),} \\
 &= f(\omega_1, y_1, v) && \text{if } 0 < h < H_1 && \text{(segment 1),} \\
 &= H_1 && \text{if } h = H_1 && \text{(kink 1),} \\
 &= f(\omega_2, y_2, v) && \text{if } H_1 < h < H_2 && \text{(segment 2),} \\
 &= H_2 && \text{if } h = H_2 && \text{(kink 2),} \\
 &= f(\omega_3, y_3, v) && \text{if } H_2 < h < \bar{H} && \text{(segment 3),} \\
 &= \bar{H} && \text{if } h = \bar{H} && \text{(kink 3 = upper limit).}
 \end{aligned} \tag{6.25}$$

Second, determine the segment or the kink on which the person locates. The following relations characterize this solution: choose

$$\begin{aligned}
 & \text{if } f(\omega_1, y_1, v) \leq 0 && 0, \\
 & \text{if } H_0 < f(\omega_1, y_1, v) < H_1 && \text{(Segment 1),} \\
 & \text{if } f(\omega_2, y_2, v) \leq H_1 < f(\omega_1, y_1, v), && \text{(Kink 1),} \\
 & \text{if } H_1 < f(\omega_2, y_2, v) < H_2 && \text{(Segment 2),} \\
 & \text{if } f(\omega_3, y_3, v) \leq H_2 < f(\omega_2, y_2, v) && \text{(Kink 2)} \\
 & \text{if } H_2 < f(\omega_3, y_3, v) < \bar{H} && \text{(Segment 3),} \\
 & \text{if } f(\omega_3, y_3, v) \geq \bar{H} && \text{(Kink 3).} \tag{6.26}
 \end{aligned}$$

Combined, these two steps imply the values of h and C that represent the utility-maximizing solutions for labor supply and consumption.

6.4.2. *Specification of the likelihood function with measurement error: all participants*

The linear specification of f given by Eq. (6.18) implies the following stochastic specification for labor supply:

$$\begin{aligned}
 \hat{h}_1 + v + \varepsilon & \quad \text{if } 0 < \hat{h}_1 + v \leq H_1 && \text{(segment 1),} \\
 H_1 + \varepsilon & \quad \text{if } \hat{h}_2 + v < H_1 < \hat{h}_1 + v && \text{(kink 1),} \\
 h^* = \hat{h}_2 + v + \varepsilon & \quad \text{if } H_1 < \hat{h}_2 + v \leq H_2 && \text{(segment 2),} \\
 H_2 + \varepsilon & \quad \text{if } \hat{h}_3 + v < H_2 < \hat{h}_2 + v && \text{(kink 2),} \\
 \hat{h}_3 + v + \varepsilon & \quad \text{if } H_2 < \hat{h}_3 + v \leq \bar{H} && \text{(segment 3),} \\
 \bar{H} + \varepsilon & \quad \text{if } \hat{h}_3 + v \geq \bar{H} && \text{(upper limit).} \tag{6.27}
 \end{aligned}$$

This represents a sophisticated variant of an econometric model that combines discrete and continuous choice elements.

All studies implementing the piecewise-linear approach assume the existence of measurement error in hours of work. With the linear measurement error model given by Eq. (6.19), observed hours $h^* = h + \varepsilon$. As long as the measurement error component ε is continuously distributed, so is h^* . In contrast to information on h , knowledge of h^* suffices neither to allocate individuals to the correct branches of the budget constraints nor to identify the marginal tax rate faced by individuals, other than at zero hours of work. The state of the world an individual occupies can no longer be directly observed, and one confronts a discrete data version of an errors-in-variables problem. The interpretation of

measurement error maintained in this analysis is that ε represents reporting error that contaminates the observation on h for persons who work.³⁵

The log-likelihood function for this model is given by $\sum_i \log g_{h^*}(h_i^*)$, where i indexes observations. Defining $v_j = H_{j-1,i} - \hat{h}_{ji}$ and $\bar{v}_j = H_{j,i} - \hat{h}_{ji}$, the components $g_{h^*}(h_i^*)$ are given by

$$\begin{aligned}
 g_{h^*}(h^*) &= \sum_{j=1}^3 \int_{\underline{v}_j}^{\bar{v}_j} g_2[h^* - \hat{h}_j, v] dv && \text{(segments 1, 2, 3),} \\
 &+ \sum_{j=1}^2 \int_{\bar{v}_j}^{\underline{v}_{j+1}} g_1[h^* - H_j, v] dv && \text{(kinks 1, 2),} \\
 &+ \int_{\bar{v}_3}^{\infty} g_1[h^* - \bar{H}, v] dv && \text{(upper limit),}
 \end{aligned} \tag{6.28}$$

where $g_1(\cdot, \cdot)$ and $g_2(\cdot, \cdot)$ are the bivariate density functions of (ε, v) and $(\varepsilon + v, v)$, respectively. Maximizing the log-likelihood function produces estimates of the coefficients of the labor supply function f . These estimates provide the information used to infer both substitution and income responses, which in turn provide the basis for calculating the work disincentive effects of income taxation.

6.4.3. Comparisons of the piecewise-linear approach with other estimation procedures

The piecewise-linear approach for estimating the work disincentive effects of taxes offers both advantages and disadvantages relative to other methods. Concerning the attractive features of this approach, piecewise-linear analyses recognize that institutional features of tax systems induce budget sets with linear segments and kinks. This is important if one believes that a smooth tax function does not provide a reasonably accurate description of the tax schedule. The piecewise-linear approach admits randomness in hours of work arising from both measurement error and variation in individual preferences and it explicitly accounts for endogeneity of the marginal tax rate in estimation, but so do the instrumental-variable and differentiable likelihood methods discussed above. As we will see below, the piecewise-linear approach more readily incorporates fixed costs of holding a job, regressive features of the tax code, and multiple program participation than other procedures due to the discrete-continuous character of hours-of-work choices induced in these environments. These features of the piecewise-linear method make it a vital approach in empirical analysis of labor supply.

³⁵ Note that expected hours of work, in this convex piece-wise linear case, is additive in each hours choice weighted by the probability of each segment or kink. Each term in this sum being at most a function of two marginal wages and two virtual incomes. Blomquist and Newey (1997) exploit this observation to develop a semi-parametric estimator for hours of work with piece-wise linear taxation, imposing the additivity through a series estimator.

On the other hand, the following shortcomings of the piecewise-linear procedure raise serious doubts about the reliability of its estimates of work disincentive effects. First, the piecewise-linear methodology assumes that both the econometrician and each individual in the sample have perfect knowledge of the *entire* budget constraint that is relevant for the worker in question. Errors are permitted neither in perceptions nor in measuring budget constraints. Taken literally, this means that: all income and wage variables used to compute each sample member's taxes are observed perfectly by the econometrician; individuals making labor supply choices know these variables exactly prior to deciding on hours of work; each individual and the econometrician know when the taxpayer will itemize deductions and the amount of these itemizations; and each taxpayer's understanding of the tax system is equivalent to that of the econometrician (e.g., the operation of such features as earned-income credits). Clearly, given virtual certainty that most of these assumptions are violated in empirical analyses of labor supply, the estimates produced by methods relying on these assumptions must be interpreted very cautiously. The differentiable-likelihood methods rely on the same assumptions. The instrumental-variable methods do not, so they are likely to be more robust.

Second, measurement error plays an artificial role in econometric models based on the piecewise-linear approach. Its presence is needed to avoid implausible predictions of the model. The statistical framework induced by the piecewise-linear approach implies that bunching in hours of work should occur at kink points if hours precisely measure h . However, for the vast majority of data sources currently used in the literature, only a trivial number of individuals, if indeed any at all, report hours of work at interior kink points. Unless one presumes that the data on hours do not directly represent h , such evidence provides the basis for immediately rejecting the distributional implications of the above specifications. Considering, for example, the labor-supply characterization proposed in Eq. (6.27), almost any test of the distributional assumptions implied by this specification would be readily rejected because observed hours would take the values H_0 , H_1 , H_2 , and \underline{H} with only a trivial or zero probability. Instead, observed hours essentially look as if they are distributed according to a continuous distribution. When a continuously-distributed measurement error ε is added to the model, observed hours h^* are continuously distributed. This provides an essential reason for introducing measurement error in the data, for without it the piecewise-linear structure provides a framework that is grossly inconsistent with the data. Of course, several sound reasons exist for admitting measurement error in a labor supply model, including the widespread suspicion that reporting error contaminates data on hours of work. However, measurement error in hours of work implies measurement error in wages, since they are typically computed as average hourly earnings. Current applications of the piecewise-linear analysis mistakenly ignore this by assuming perfectly measured budget constraints.³⁶ The unnatural role played by measurement error raises questions about the credibility of findings derived from the piecewise-

³⁶ It is possible to argue that this error does not result in measurement error in the hourly wage, if the measurement error is interpreted as an "optimization" error.

linear approach.³⁷ In contrast to the piecewise-linear approach, it is not essential to introduce measurement error in either the differentiable-likelihood or the instrumental-variable approach because hours in the distribution of h are continuous without measurement error.

Third, existing research implementing the piecewise-linear methodology relies on very strong exogeneity assumptions. Other than hours of work, all variables involved in the calculation of taxes are presumed to be exogenous determinants of labor supply behavior, both from a statistical and from an economic perspective. These variables include gross wages, the various components of non-labor income, and deductions. In light of the evidence supporting the view that wages and income are endogenous variables in labor supply analyses, particularly in the case of wages,³⁸ suspicions arise regarding the dependability of estimated substitution and income effects based on procedures that ignore such possibilities. Most of the exogeneity assumptions are also maintained in the differentiable-likelihood approach, but are easily relaxed when applying instrumental-variable procedures (given the availability of a sufficient number of other instrumental variables).

Fourth, some concerns about the reliability of estimates produced by the piecewise-linear approach ensue due to the static behavioral framework maintained in the formulation of empirical relations. Piecewise-linear studies invariably rely on the textbook one-period model of labor supply as a description of hours-of-work choices, and impose it to estimate parameters. Existing implementations of the differentiable-likelihood approach suffer from the same problem. Everyone acknowledges that individuals are not simply myopic optimizers; they transfer income across periods to achieve consumption plans that are infeasible without savings. A serious question arises concerning the relevance of such considerations in estimating substitution and income effects used to predict responses to tax policy.

6.4.4. Imposition of behavioral restrictions with convex constraints

The econometric model produced by this piece-wise linear formulation implicitly imposes parametric restrictions that constrain the signs of estimated substitution and income effects. As developed in MaCurdy et al. (1990), particular inequality restrictions must hold in the application of estimation procedures with piecewise-linear budget constraints for likelihood functions to be defined (i.e., to ensure that the components of these functions are non-negative). More specifically, in applications of such procedures the Slutsky condition must be locally satisfied at all interior kink points of budget sets that represent feasible options for any individual in the sample such that the compensated substitution effect must be positive. For the linear specification of the labor supply function considered in the preceding discussion, the specific inequality constraints imposed are

$$\alpha - \beta H_{ji} \geq 0, \quad \forall i, j, \quad (6.29)$$

where the quantities H_{ji} represent the hours-of-work values that correspond to interior kink

³⁷ See Blomquist (1996) for some Monte Carlo comparisons of ML and IV with measurement error.

³⁸ See Pencavel (1986) for a summary of this evidence.

points j on a sample member i 's budget set. Because many values of H_{ji} exist in most analyses of piecewise-linear constraints, fulfillment of relations (6.29) essentially requires global satisfaction of the Slutsky condition by the labor supply function. Such a requirement, in essence, globally dictates that the uncompensated substitution effect of a wage change on hours of work must be positive for the labor supply specification considered in the preceding discussion, and the income effect for hours of work must be negative. The imposition of these restrictions, especially for men, is highly suspect given the available evidence from other studies. These restrictions carry over to more general labor supply functions.

6.5. Maximum likelihood: accounting for fixed costs of participation and missing wages

6.5.1. Fixed costs

As mentioned above, some applications of the piecewise-linear approach incorporate fixed costs to working – costs such as transportation that must be paid for any amount of work but which may vary across individuals. This significantly complicates the analysis because the optimized level of work under the budget constraint while working may not represent the optimal choice overall; one must explicitly consider the option of not working and thus avoiding the fixed costs. For any level of fixed costs, a minimum number of hours worked is implied creating an attainable range in the observable hours of work distribution; individuals will not work unless the gain is large enough to overcome the fixed costs. In essence, these complications arise because the budget constraint is not convex, invalidating simple maximization procedures.

If an individual must pay fixed monetary costs, F , to work, then non-labor income, Y , in the above budget constraints is replaced by

$$\begin{aligned} Y - F, & \quad \text{if } h > 0, \\ Y, & \quad \text{if } h = 0. \end{aligned} \tag{6.30}$$

F is partially unobservable and, thus, modeled as a stochastic element, varying across individuals. Hence, we see that the budget constraint discontinuously jumps down by F when the individual chooses to work.

To solve for the optimum when faced with this budget constraint, two regimes must explicitly be considered: working and not working. Estimation proceeds by finding the maximum utility under each regime and then comparing these to determine which option is chosen. In either regime, the utility function $U(C, h, \nu)$ – where we explicitly note the unobserved component, ν – is maximized subject to Eq. (6.1) modified by Eq. (6.30).

In the no-work regime, the solution is simple. We know h is 0, so utility is given by $U(Y - \pi(Y - D), 0, \nu)$.

The solution in the work regime closely follows the solution presented in Section 6.3. Again utilizing the labor supply function, $f(\omega, y, \nu)$ yields the solution for h given in Eq. (6.25), where the virtual income y now subtracts fixed costs F . However, to compute

maximum utility in this regime requires associating a utility level with each possible hours choice. Utility along any segment, i , is given by the indirect utility function, $V(\omega_j, y_j, \nu)$. At kinks, the direct utility function must be used, so the utility at kink j is given by $U(\omega_j H_j + y_j, H_j, \nu)$. Hence, utilizing exactly the same solution procedure derived in Section 6.3, we can define maximized utility when working, V^* :

$$\begin{aligned}
 -\infty, & & f_1 \leq 0, \\
 V(\omega_1, y_1, \nu), & & 0 < f_1 < H_1, \\
 U(\omega_1 H_1 + y_1, H_1, \nu), & & f_2 < H_1 \leq f_1, \\
 V^*(\omega, y, \nu) = V(\omega_2, y_2, \nu), & & H_1 < f_2 < H_2, \\
 U(\omega_2 H_2 + y_2, H_2, \nu), & & f_3 < H_2 \leq f_2, \\
 V(\omega_3, y_3, \nu), & & H_2 < f_3 < H_3, \\
 U(\omega_3 \bar{H} + y_3, \bar{H}, \nu), & & f_3 \geq \bar{H},
 \end{aligned} \tag{6.31}$$

where

$$f_j = f(\omega_j, y_j, \nu) \equiv \frac{V_\omega(\omega_j, y_j, \nu)}{V_y(\omega_j, y_j, \nu)}, \tag{6.32}$$

with V_ω and V_y denoting the partial derivatives of V ; relation (6.32) is, of course, Roy's identity defining the labor supply function, f , evaluated at wage and income levels ω_j and y_j . The use of $-\infty$ for $h = 0$ simply indicates that $h = 0$ is not included in this regime and, thus, selecting it indicates that the no-work regime is preferred. Given functional forms for V and U , finding V^* is straightforward.

Given maximized utility under each regime, the final step in the solution is to compare the two regimes. An individual chooses to work at the hours specified by the solution in Eq. (6.25) if

$$V^*(\omega, y, \nu) \geq U(Y - \pi(Y - D), 0, \nu) \tag{6.33}$$

and chooses not to work otherwise. For any level of ν , treating Eq. (6.33) as an equality implies a critical level of fixed costs, $F^*(\nu)$ above which the individual will choose not to work; F enters this relation through the virtual income variable y . Because desired hours of work increase with ν , this critical value will generally be increasing in ν – greater propensity to work implies that higher fixed costs are required to prefer the no-work option. If restrictions are placed on the support of F , such as $F > \underline{F}$, there will be values of ν low enough to rule out the work regime, thus implying a hole at the low end of the h distribution.

6.5.2. Missing wages

As a final step before deriving the likelihood function, note that in the no-work regime, gross wage, W , is not observed and, thus, the budget constraint cannot be derived. Hence, W must be endogenized. This can be accomplished by the simple function

$$W = W(Z) + \eta \tag{6.34}$$

where Z includes all observable variables determining W and η is the unobservable component. In a richer model, the equation for W could be derived as an equilibrium condition.

To derive the likelihood function, first consider the likelihood contribution of an individual who does not work. We assume this no-work decision can be observed, so there is no measurement error. In the no-work case, one of two situations applies: (i) fixed costs are sufficiently high with $F > F^* \equiv F^*(\nu, \eta)$ for any given ν and η , or (ii) if this fixed-cost threshold falls below the lowest admissible value for F (i.e., $F^* \leq F$), then desired hours are sufficiently low with $\nu < \nu^* \equiv \nu^*(\eta)$ for any η .³⁹ The probability of this event is

$$l_0 = \int_{-\infty}^{\infty} \int_{-\infty}^{\nu^*} \int_{F^*}^{\infty} g_{\nu\eta F}(\nu, \eta, F) dF d\nu d\eta, \tag{6.35}$$

where $g_{\nu\eta F}$ is joint density of (ν, η, F) .

For the work regime, the likelihood contribution looks very much like that derived in Eq. (6.28), as we continue to assume the linear hours of work function and the form of measurement error assumed there. The only changes are the addition of terms for δ and F (accounting for the fact that $F < F^*(\nu)$) and the removal of the term for the lower limit which is no longer part of that regime and is now perfectly observable. Using g_1 and g_2 to denote the distribution of $(\varepsilon, \nu, \eta, F)$ and $(\varepsilon + \nu, \nu, \eta, F)$ yields

$$l_1 = \sum_{j=1}^3 \int_{\underline{\nu}_j}^{\bar{\nu}_j} \int_0^{F^*} g_2[h^* - f_j, \nu, W - W(Z), F] dF d\nu + \sum_{j=1}^2 \int_{\bar{\nu}_j}^{\underline{\nu}_{j+1}} \int_0^{F^*} g_1[h^* - H_j, \nu, W - W(Z), F] dF d\nu + \int_{\bar{\nu}_3}^{\infty} \int_0^{F^*} g_1[h^* - \bar{H}, \nu, W - W(Z), F] dF d\nu, \tag{6.36}$$

where

$$\underline{\nu}_j \text{ solves the equation } f(\omega_j, y_j, \underline{\nu}_j) = H_{j-1},$$

$$\bar{\nu}_j \text{ solves the equation } f(\omega_j, y_j, \bar{\nu}_j) = H_j. \tag{6.37}$$

All variables are defined as in Section 6.4. Define $P_k = 1$ if the individual works and 0 otherwise. Then the likelihood function for an individual is given by

³⁹ The critical value ν^* solves relation (6.33) treated as an equality with virtual income y evaluated at F .

$$l = (l_1)^{P_\varepsilon} (l_0)^{1-P_\varepsilon}. \quad (6.38)$$

Estimation proceeds by maximizing the sum of log likelihoods across individuals, as always.

This is quite complex in this case, requiring knowledge of both the direct utility U and the indirect utility V , and also requiring comparisons across regimes for all individuals and all parameter values.

6.6. Welfare participation and non-convex budget constraints

A common source of non-linearity in budget constraints involves participation in welfare programs. To illustrate this situation, consider the simplest case in which the only taxes faced by an individual result from benefit reduction on a single welfare program. Fig. 18 presents this scenario. As discussed in Section 2 of this survey, individuals face very high effective tax rates when they initially work due to large reductions in their benefits occurring when earnings increase. Once benefits reach 0, the tax rate drops to a lower level, creating a non-convex kink in the budget constraint. This non-convexity invalidates the simple procedures of Section 6.4 implemented to divide sample spaces into locations on budget sets.

6.6.1. Simplest welfare case with no stigma

In this simple case, an individual maximizes $U(C, h, v)$ subject to the budget constraint

$$C = Wh + Y + B(I(h)), \quad (6.39)$$

where benefits are given by the simplest benefit schedule:

$$B(I(h)) = \begin{cases} G - bWh, & \text{if } G - bWh > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (6.40)$$

G gives the guarantee amount which is reduced at the benefit reduction rate b as the earnings, Wh , increase. This implies a kink point at $H_1 = G/bW$ where benefits reach 0 and, thus, the marginal wage rises to W . So, the individual faces two segments: segment 1 has $h < H_1$ with net wage $\omega_1 = (1 - b)W$ and virtual income $y_1 = Y + G$; and segment 2 has $h > H_1$ with net wage $\omega_2 = W$ and virtual income $y_2 = Y$.⁴⁰

Because the budget constraint is non-convex, the solution cannot be characterized simply by finding a tangency with the budget constraint as it was in Section 6.3. Multiple tangencies are possible and these must be directly compared to determine the optimum. Hence, the regime shift approach of Section 6.5 is needed.

Consider first the regime in which positive benefits are received; that is, $h < H_1$. Maximization, given the effective wage and income, on this linear segment follows the

⁴⁰ We continue to use N to denote unearned non-taxable income for ease of notation. In addition, we ignore any upper bound on hours worked for simplicity.

method of Section 6.3. We can characterize the optimal choice according to the function $f(\omega_1, y_1, \nu)$. Denote the value of ν which implies $f(\omega_1, y_1, \nu) = 0$ as ν_0 . Then the optimal hours choice along that segment is given by

$$h = f(\omega_1, y_1, \nu), \quad \nu > \nu_0, \quad h = 0, \quad \nu \leq \nu_0. \tag{6.41}$$

The optimized value on this segment (including the zero work option), accounting for the fact that $h > H_1$ is not allowed, is given by

$$V_1^*(\omega_1, y_1, \nu) = \begin{cases} V(\omega_1, y_1, \nu), & 0 < f_1 \leq H_1 \\ U(y_1, 0, \nu), & f_1 \leq 0 \\ -\infty, & f_1 > H_1. \end{cases} \tag{6.42}$$

where Eq. (6.32) defines f_1 .

Next, consider the regime without benefits, that is with $h \geq H_1$. Again the optimal choice, given the wage and income, on this segment is given by the labor supply function $f(\omega_2, y_2, \nu)$. The optimized value, accounting for the fact that $h < H_1$ is not admissible, is given by⁴¹

$$V_2^*(\omega_2, y_2, \nu) = \begin{cases} V(\omega_2, y_2, \nu), & f_2 \geq H_1, \\ -\infty, & f_2 < H_1. \end{cases} \tag{6.43}$$

Hence, the individual selects regime 1, with welfare receipt, if $V_1^* > V_2^*$, and regime 2 otherwise. Since work propensity increases with ν , this can be characterized by a cutoff value, ν^* , defined by

$$V_1^*(\omega_1, y_1, \nu^*) = V_2^*(\omega_2, y_2, \nu^*). \tag{6.44}$$

For values of ν above ν^* , regime 2 is chosen; and for values below ν^* , regime 1 is realized.

We can define three sets, Ω_0 , Ω_1 , and Ω_2 , such that for $\nu \in \Omega_0$ the individual chooses not to work, for $\nu \in \Omega_1$ the individual locates on segment 1 with positive hours of work, and for $\nu \in \Omega_2$ the individual locates on segment 2. We must consider two cases to define these sets exactly. First, suppose $\nu^* > \nu_0$. Then we have

$$\Omega_0 = \{ \nu \mid \nu \leq \nu_0 \},$$

$$\Omega_1 = \{ \nu \mid \nu_0 < \nu \leq \nu^* \},$$

$$\Omega_2 = \{ \nu \mid \nu > \nu^* \}. \tag{6.45}$$

Alternatively, if $\nu^* \leq \nu_0$, then the switch to regime 2 occurs before positive hours are worked in regime 1, that is

⁴¹ In the following formulation, we implicitly assume that the event $f_2 \geq \bar{H}$ occurs with zero probability

$$\Omega_0 = \{ \nu \mid \nu \leq \nu^* \},$$

$$\Omega_1 = \emptyset,$$

$$\Omega_2 = \{ \nu \mid \nu > \nu^* \}. \tag{6.46}$$

Hence, for certain individuals and parameter values, no value of ν exists such that they will locate on segment 1 with positive hours of work.

To characterize the likelihood function we again need a functional form for the gross wage of the form $W = W(Z) + \eta$. We ignore measurement error here for simplicity, and because there is no problem with individuals failing to locate at the kink in this non-convex case. Define $P_B = 1$ if the individual receives benefits, and $P_E = 1$ if the individual works, both 0 otherwise. The likelihood function is given as follows, incorporating $g_{\nu\eta}(\eta, \nu)$ and the general inverse function $\nu = \nu(h)$:

$$P_B = 1, \quad P_E = 1, \quad l_{11} = \frac{\partial \nu}{\partial h} g_{\nu\eta}(\nu(h), W - W(Z))I(\nu \in \Omega_1),$$

$$P_B = 0, \quad P_E = 1, \quad l_{01} = \frac{\partial \nu}{\partial h} g_{\nu\eta}(\nu(h), W - W(Z))I(\nu \in \Omega_2),$$

$$P_B = 1, \quad P_E = 0, \quad l_{10} = \int_{\Omega_0}^{\infty} g_{\nu\eta}(\nu, \eta) d\nu d\eta, \tag{6.47}$$

where $I(\cdot)$ represents an indicator function equal to 1 if the condition in the parentheses is true. Because the value of ν implied by the hours choice may be inconsistent with the value implied by the regime choice, it is possible to have “holes” in the hours distribution around the kink point. For example, an individual on segment 1 must have $\nu \leq \nu^*$. If his hours choice is too close to the kink, this may imply a value of $\nu > \nu^*$ and thus an observation with zero likelihood.

The overall likelihood function is given by

$$l = (l_{11})^{(P_B)(P_E)} (l_{01})^{(1-P_B)(P_E)} (l_{10})^{(P_B)(1-P_E)}. \tag{6.48}$$

Estimation proceeds by maximizing this sum of the log likelihoods across individuals.

6.6.2. Welfare stigma

The above analysis assumes that all individuals eligible for welfare are on welfare. Individuals working less than h_0 but failing to receive welfare are operating below the implied budget constraint, a possibility not permitted in the analysis. Yet, many individuals are in exactly this situation. This is generally explained by assuming the existence of some utility loss or stigma associated with welfare.

To capture welfare stigma the utility function is modified to take the form

$$U = U(C, h, \nu) - P_B \xi, \tag{6.49}$$

where ξ is the level of welfare stigma which is greater than 0 and varies across individuals.⁴² With this modification we again consider the welfare and non-welfare regimes. Since the welfare stigma term does not affect the marginal decisions, given that the individual is on welfare, the discussion of hours of work presented above for regime 1 is still valid. The optimal utility is now given by

$$V^*(\omega_1, y_1, \nu) = \begin{cases} V_1(\omega_1, y_1, \nu) - \xi, & 0 < f_1 \leq H_1, \\ U(y_1, 0, \nu) - \xi, & f_1 \leq 0, \\ -\infty, & f_1 > H_1. \end{cases} \quad (6.50)$$

The analysis for regime 2 is altered in this case, because an individual can be observed on welfare for any value of h – that is, given welfare stigma, it is possible to observe an individual with $h < H_1$, but $P_B = 0$. So regime 2 is now defined solely by $P_B = 0$. Optimal hours of work, given ω_2 and y_2 , are given by $f(\omega_2, y_2, \nu)$. Defining the value of ν for which $f(\omega_2, y_2, \nu) = 0$ as ν^+ , hours of work under this regime are now given by

$$h = f(\omega_2, y_2, \nu), \quad \nu > \nu^+,$$

$$h = 0, \quad \nu \leq \nu^+. \quad (6.51)$$

Optimized utility is now

$$V_2^*(\omega_2, y_2, \nu) = \begin{cases} V(\omega_2, y_2, \nu), & f_2 > 0 \\ U(y_2, 0, \nu), & f_2 \leq 0. \end{cases} \quad (6.52)$$

Choice of regime still proceeds by comparing V_1^* and V_2^* , as done in Eq. (6.44). For any ν in the sets Ω_0 or Ω_1 defined by Eq. (6.45) or (6.46), there is now some critical level of $\xi^* \equiv \xi^*(\nu)$, which depends on ν , such that regime 2 is chosen when $\xi > \xi^*$; regime 1 is chosen otherwise.

Given this characterization, we can derive the likelihood function for each combination of P_B and P_E , using the joint densities $g_{\nu\xi\eta}(\nu, \xi, \eta)$ and $g_{\nu\eta}(\nu, \eta)$:

$$P_B = 1, \quad P_E = 1, \quad l_{11} = \frac{\partial \nu}{\partial h} \int_0^{\xi^*} g_{\nu\xi\eta}(\nu(h), \xi, W - W(z))I(\nu \in \Omega_1)d\xi,$$

$$P_B = 0, \quad P_E = 1, \quad l_{01} = \frac{\partial \nu}{\partial h} g_{\nu\eta}(\nu(h), \xi, W - W(z))I(\nu \in \Omega_1)$$

$$+ \frac{\partial \nu}{\partial h} \int_{\xi^*}^{\infty} g_{\nu\xi\eta}(\nu(h), \xi, W - W(z))I(\nu \in \Omega_1)d\xi,$$

⁴² This additive form is used for simplicity. More general forms can be used, but change none of the substantive points presented here.

$$\begin{aligned}
 P_B = 1, \quad P_E = 0, \quad l_{10} &= \int_{-\infty}^{\infty} \int_{\Omega_0} \int_0^{\xi^*} g_{\nu\xi\eta}(\nu, \xi, \eta) d\xi d\nu d\eta, \\
 P_B = 0, \quad P_E = 0, \quad l_{00} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\nu^*} \int_0^{\xi^*} g_{\nu\xi\eta}(\nu, \xi, \eta) d\xi d\nu d\eta.
 \end{aligned}
 \tag{6.53}$$

Estimation proceeds as in the non-stigma case by selecting the appropriate likelihood branch for each individual and then maximizing the sum of the log likelihoods.

As with the fixed cost case, the likelihood function is complex even in this extremely simplified welfare case. For each possible set of parameter values, the maximum must be computed for each regime and then compared to compute ξ^* . Adding the tax codes, with their implied kinks, increases computational complexity. As a result, the literature has adopted a simplifying methodology which we present in Section 6.8.

6.6.3. Multiple program participation

In principle, the extension to the case of multiple program participation is straightforward. For simplicity, we consider a case in which the individual can choose between participating in no welfare programs, participating in welfare program 1, participating only in program 2, or participating in both welfare programs 1 and 2. We extend the utility function as follows:

$$U = U(c, h, \nu) - P_1\xi - P_2\chi \tag{6.54}$$

where $P_1 = 1$ if the individual participates in program 1, and $P_2 = 1$ if the individual participates in program 2.⁴³ Benefits from program 1, $B_j(I(h))$, are given:

$$B_j(I(h)) = \begin{cases} G_j - b_jWh, & \text{if } G_j - b_jWh > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{6.55}$$

Benefits from both together are given as

$$B_1(I(h)) + B_2(I(h)) = \begin{cases} G_1 + G_2 - b_1Wh - b_2Wh = G - bWh, & \text{if } G - bWh > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6.56}$$

where $G = G_1 + G_2$ and $b = b_1 + b_2$. In general, the benefit functions for programs 1 and 2 will have different breakeven points, implying the values of hours defining kinks (H_1 in Fig. 18) will not be the same.

This formulation expands the model considered in Section 6.4.3. To adapt this earlier model, one must designate three distinct regimes in place of regime 1 specified above: regime 1a indicating an individual participates only in program 1, regime 1b signifying this person collects benefits only from welfare program 2, and regime 1c designating

⁴³ The use of two additive errors is a simplifying assumption which ensures that the stigma from both programs is higher than stigma from program 1 alone.

participation in both programs. Optimal hours and utility for participation in a regime are given by (6.41), (6.42), (6.50), (6.51), and (6.52), with net wages and virtual income in these formulations specified as $\omega_j = W(1 - b_j)$ and $y_j = Y + G_j$, with $j = 1a, 1b,$ or $1c$. In particular, relations analogous to (6.41) and (6.42) define the labor supply and utility functions for each of the new regimes for the “on-welfare” segments associated with relevant combination of welfare programs. Relations (6.51) and (6.52) still define the labor supply and utility functions for the non-welfare regime. The set of relations define thresholds for ν demarcating the regions of unobserved tastes determining when a person works (ν_0 in (6.41) and ν^+ in (6.51)). Maximization again requires selection of a regime. Relations analogous to (6.50) and (6.52) characterize utilities corresponding to the various regimes. Conditional on values ν , these relations in turn imply thresholds for the stigma errors ξ , χ , and $\xi + \chi$ that determine individuals’ welfare participation. The likelihood function for this model takes a form similar to Eq. (6.53), with more branches appearing in the function reflecting the additional regimes analyzed in this formulation.

Again, note the complexity of these, extremely simplified welfare cases, even these involve significantly financial burden. For each possible set of parameter values, one must compute the maximum for each regime, account for the benefit structure, and then compare these to compute the error ranges for the likelihood function. When the individual is unemployed, one must perform these calculations for all possible wage values and all values of ν consistent with the no-work decision. Adding the tax code, with its implied kinks, increases computational difficulties. Introducing additional sources of unobserved heterogeneity enlarges the number of dimensions over which one must calculate integrals, requiring sophisticated numerical procedures and considerable computer resources. As a result, the literature has adopted simplifying methodologies, a topic to which we now turn.

6.7. An approach for computational simplification and discrete hours choices

To make estimation problems manageable, a popular method is to presume that consumers face only a limited set of hours choices. For example, a worker may choose only full-time work, part-time work, or no work, with each of these options implying a prescribed number of hours. Formally, this is done by assuming that unobservable tastes components, ν , possess a discrete distribution, usually characterized as a multinomial distribution conditional on covariates. Combined with a 0/1 welfare decision, this finite set of hours choices yields a relatively small set of discrete states, say S states, over which the utility function must be maximized.

Given a specific form for the preference function, utility can be readily evaluated at each of the hours choices and the maximum can be determined. Given an assumed joint distribution for unobservable tastes components, ν , for the error component determining wages, η , and for welfare stigma, ξ , one can compute a probability that a family selects alternative “ j ”. This in turn defines a sample log likelihood of the form

$$l = \sum_{j=1}^S d_j \ln P(j | X, \Theta), \quad (6.57)$$

where d_j is an indicator for whether individual i chooses alternative j , X is a vector of observable characteristics, and $P(j | X, \Theta)$ is the probability of choosing alternative j with Θ the set of unknown parameters. Such formulations are substantially less complicated than the specifications considered above because one avoids the intricate process of calculating thresholds and dealing with combined continuous-discrete endogenous variables; only discrete choices are allowed for here.

This formulation requires each individual to be placed into a limited set of preassigned work states, even though observed hours worked take many more values, making hours look as if they were continuously distributed. To overcome this issue, analyses applying this approach necessarily introduce measurement error in hours of work to admit hours to deviate from the discrete values assumed for the choice set. Hence, conditional on ν , each alternative “ j ” contributes some positive probability $P(j | X, \Theta, \nu)$ which now depends on the value of the unobservable measurement error variables.

We illustrate this approach by considering the linear measurement error model given by Eq. (6.19) where the reporting error $\varepsilon \sim g_\varepsilon$, with ε and ν independent. Further, as typically assumed, we specify that hours are not subject to measurement error in no-work states. The likelihood function for hours now takes the form

$$l = \left(\sum_{j \in S_0} d_j \ln P(j | X, \Theta) \right)^{1-P_E} \left(\sum_{j \in S_1} d_j \ln (g_\varepsilon(h - h_j) P(j | X, \Theta)) \right)^{P_E}, \quad (6.58)$$

where P_E denotes a 0/1 variable with 1 indicating that the individual works, S_0 designates the set of all states associated with the individual not working, the set S_1 includes all states in which the individual works, and h_j denotes the admissible values of true hours. Earnings depend on the values of h_j and wages. In Eq. (6.58), observed hours are continuously distributed among workers.

6.8. Survey of empirical findings for non-linear budget constraints models

Having developed a theoretical framework for analyzing the effects of taxes on labor supply, we proceed in this section to briefly survey the body of empirical literature that seeks to estimate labor supply elasticities in the presence of welfare and taxes. The selection of studies considered here illustrate the empirical methodologies developed in the preceding subsections.⁴⁴ The survey begins with the empirical work involving maxi-

⁴⁴ See also Burtless and Hausman (1978), Heckman (1979c), Hausman (1980, 1985a,b), Cogan (1981), Nakamura and Nakamura (1981), Ashenfelter (1983), Moffitt (1983, 1986, 1992a,b), Fraker et al. (1985), Robins (1985), Blundell et al. (1986, 1988), Blau and Robins (1988), Fraker and Moffitt (1988), Zabalza (1988), Kell and Wright (1989), Moffitt and Wolfe (1992), Ribar (1992), Bingley et al. (1995), Blomquist (1996) and Bingley and Walker (1997).

imum likelihood estimation with convex budget sets. This case, which we discussed in Sections 6.3 and 6.4, has received the most attention in the received empirical literature. We then consider papers that report at least one model estimated by instrumental variables or involving non-convex budget sets, the cases we considered in Sections 6.2 and 6.5. The survey concludes with an application of the multiple welfare program participation model which we presented in Section 6.6. In what follows we restrict our attention to post-1980 analyses of the United States and Western Europe. Tables 1 and 2 summarize the results for men and women, respectively.

6.8.1. Maximum likelihood estimation with convex budget sets

Blomquist (1983) estimates labor supply functions for prime-age Swedish males, using a piecewise-linear analysis with a convex budget set to account for the highly progressive Swedish income tax. His approach follows closely that of Hausman (1981). The 1973 cross section of 688 males, aged 25–55, used in his study is derived from a survey conducted by the Swedish Institute for Social Research. Estimation is based on the following linear model:

$$h_i^* = \alpha\omega_i + \beta y_i + \gamma Z + \varepsilon_i, \quad (6.59)$$

where α , β , and γ are preference parameters, h_i^* is hours worked in 1973, Z is a vector of individual characteristics, ω_i is the net wage rate on the i th extended budget segment, y_i is the virtual income for this segment, and ε_i is a disturbance. Non-labor income is defined as the spouse's after-tax income plus the family's capital income after tax and family allowances, where after-tax capital income is computed as it would have been if the person had worked no hours.

Blomquist assumes that α and γ are constant across individuals, whereas each person's β is assumed to be a draw from $f(\beta)$, the normal density function with upper truncation at zero. Since the individual β_i are not identified, Blomquist estimates μ_β and σ_β^2 , the parameters of f , in addition to α and γ . Estimation by maximum likelihood yields an income elasticity of -0.03 , a compensated wage elasticity of 0.11 , and an uncompensated wage elasticity of 0.08 . The author also reports results from estimation with the restriction $\sigma_\beta^2 = 0$ imposed. Although the resulting estimates are similar to those of the unconstrained model, a likelihood ratio test rejects the restriction at conventional levels.

MaCurdy et al. (1990) analyze the labor supply of prime-age married males using the 1975 cross-section from the Michigan Panel Study of Income Dynamics (PSID). They show that maximum likelihood estimation of a consumer-choice problem with non-linear budget sets implicitly relies on the satisfaction of inequality constraints that translate into behaviorally meaningful restrictions. These constraints arise from the requirement to create a well-defined statistical model, and not as a consequence of economic theory (see Section 6.4.4). The authors then present empirical results suggesting that these implicit constraints play a major role in explaining the disparate results found in the literature on men's labor supply. The empirical work is based both on the piece-wise linear approach

Table 1
 Non-linear budget constraint models: summary of some empirical results for men

Study	Data source and sample selection	Variables: H, hours; W, wage; Y, income	Functional form of labor supply and budget set structure	Estimation method and stochastic specification ^a	Uncompensated wage elasticity	Income elasticity
Blomquist (1983)	Swedish Level of Living Survey 1974: sample size 688, all employed, married, aged 25-55	H, annual hours for 1973 (weeks worked × average hours per week) W, directly observed Y, spouse's net income + family allowances + net capital income	Linear labor supply, convex (piecewise linear)	ML ML random preferences (on income coefficient)	0.08 0.08	-0.03 -0.04
Blomquist and Hansson-Brusewitz (1990)	Swedish Level of Living Survey 1981: sample size 602, all employed, married, aged 25-55	H, annual hours W, directly observed Y, spouse's net income + family allowances + net capital income	Linear and quadratic labor supply Convex and non-convex (piecewise linear)	<i>Linear labor supply</i> ML-convex ML-non-convex ML-convex random preferences <i>Quadratic labor supply</i> ML-convex ML-convex, random preferences	0.08 0.08 0.13	0.002 -0.008 -0.01
Bourguignon and Magnac (1990)	French Labour Force Survey 1985: sample size 1992, all employed, married, aged 18-60	H, normal weekly hours W, hourly net wage (monthly earnings / hours) Y, family allowances	Linear labor supply Convex (piecewise linear)	Linear labor supply Convex (piecewise linear)	0.12 0.1	-0.008 -0.07
Blundell and Walker (1986)	British Family Expenditure Survey 1980: sample size 1378, all employed, married, aged 18-59	H, usual weekly hours W, weekly earnings/hours Y, consumption based two-stage budgeting ^b	Gorman polar form/translog Convex (piece-wise linear)	ML-convex, random preferences	0.024	-0.287
Flood and MaCurdy (1992)	Swedish Household Market and Non-market Survey (HUS)	H, annual hours W, hourly wage (annual earnings/)	Linear and semi-logarithmic Convex (piecewise	<i>Linear labor supply</i> ML-piecewise linear, random preferences	0.16	-0.1

1984: sample size 492, all employed, aged 25-65	annual hours, asset income, UI, housing allowances etc.	linear and differentiable	(on income coefficient) additive measurement error	
			ML-differentiable, random preferences (on income coefficient) measurement error:	
			Additive	0.14
			Multiplicative	0.04
			None	0.07
			<i>Semi-log labor supply</i>	
			ML-differentiable, random preferences, measurement error:	
			Multiplicative	0.21
			None	0.25
			IV across 7 different specifications	{ -0.25, 0.21 }
			ML, random preferences (on income coefficient)	{ 0.00, 0.03 }
				{ -0.09, -0.1, { -0.11, 0.04 } }
				{ -0.95, -1.03 }
Hausman (1981)	US Panel Study of Income Dynamics 1975: sample size 1085, all employed, married, aged 25-55	H, annual hours W, directly reported hourly wage rates Y, other income assuming 8% return to financial assets	Linear labor supply	
Kaiser et al. (1992)	German SocioEconomic Panel 1983: sample size 939 non-employed, married, non-retired	H, annual hours W, hourly wage (annual income/annual hours) Y, income from rents, capital income and transfer payments	Convex and non-convex (piecewise-linear)	
			ML-convex Linear labor supply convex (piecewise linear)	-0.004 -0.28

Table 1 (continued)

Study	Data source and sample selection	Variables: H, hours; W, wage; Y, income	Functional form of labor supply and budget set structure	Estimation method and stochastic specification ^a	Uncompensated wage elasticity	Income elasticity
MaCurdy et al. (1990)	US Panel Study of Income Dynamics 1975: sample size 1017, all employed, married, aged 25-55	H, annual hours W, average hourly wage (earnings/annual hours) Y, rent, interest, dividends, etc.	Linear labor supply Convex and non-convex (piecewise-linear and differentiable budget constraints)	ML-convex, random preferences ^c (on income coefficient)	0	-0.01
Triest (1990)	US Panel Study of Income Dynamics 1983: sample size 978, all employed, married, aged 25-55	H, yearly hours in all jobs held in 1983 W, average hourly earnings (earnings/annual hours) Y, rents, dividends, interest income, trust funds, etc.	Linear labor supply Convex (piecewise linear)	ML-convex, random preferences ^c (on income coefficient)	0.05	0 ^d
van Soest et al. (1990)	Dutch Strategic Labor Market Research Survey 1985: sample size 801 employed, 49 non-employed	H, average weekly hours W, net hourly wage (earnings/hours) Y, other incomes	Linear Labor Supply Convex (piecewise linear)	ML-convex	0.12	-0.01

^a This column indicates whether preferences are treated as random in specifications, in addition to "optimization" or measurement errors that are always incorporated in specifications. Unless stated otherwise, random preferences without further indications means that only intercept coefficients are stochastic. For further information, see Eq. (6.59) and related discussion.

^b See Eq. (4.33) and related discussion.

^c For other results see article.

^d Estimated coefficient constrained at zero.

Table 2
Non-linear budget constraint models: summary of some empirical results for married women

Study	Data source and sample selection	Variables: H, hours; W, wage ^a ; Y, income	Functional form of labor supply and budget set structure	Estimation method and stochastic specification ^b	Uncompensated wage elasticity	Income elasticity
Arellano and Meghir (1992)	British Family Expenditure Survey (FES) 1983 and British Labor Force Survey (LFS) 1983: sample size 11,535 employed, 13,200 non-employed, aged 20–59	H, weekly hours W, hourly earnings ^c Y, consumption based other income measure	Semi-log labor supply Convex (piecewise linear)	Instrumental variables/selection	(0.29, 0.71)	{ -0.13, -0.40 }
Arrufat and Zabalza (1986)	British General Household Survey 1974: sample size 2002 employed, 1493 non-employed, aged < 60	H, weekly hours W, gross hourly earnings, SS Y, net weekly unearned family income + husband's earnings	CES utility based labor supply Convex (piecewise linear)	ML-convex, random preference (log normal on CES leisure coefficient) ^d	2.03	-0.2
Blomquist and Hansson-Brusewitz (1990)	Swedish Level of Living Survey 1981: sample size 795 full sample, 640 employed, aged 25–55	H, annual hours W, directly observed, SS ^e Y, spouse's net income + family allowances + net capital income	Linear and quadratic labor supply Convex and non-convex (piecewise linear)	Linear labor supply ML-non-convex ML-non-convex, random preferences (on income coefficient)	0.79 0.77	-0.24 -0.06
Blundell et al. (1988)	British Family Expenditure Survey 1980: sample size 1378 employed, aged 18–59	H, usual weekly hours W, hourly earnings ^f Y, consumption based two-stage budgeting measure	Generalized linear expenditure system Convex (piecewise linear)	Quadratic Labor Supply ML-convex Truncated ML, random preferences	0.58 0.09	-0.05 -0.26

Table 2 (continued)

Study	Data source and sample selection	Variables: H, hours; W, wage ^a ; Y, income	Functional form of labor supply and budget set structure	Estimation method and stochastic specification ^b	Uncompensated wage elasticity	Income elasticity
Bourgiugnon and Magnac (1990)	French Labor Force Survey 1985: sample size 1175 employed, 817 non-employed, aged 18-60	H, normal weekly hours W, hourly net wage, SS, (earnings /hours) Y, spouse's net income + family allowances	Linear labor supply Convex (piecewise linear)	ML-convex, random preferences ML with fixed costs, random preferences	1 0.05	-0.3 -0.2
Colombino and Del Boca (1990)	Turin Survey of Couples 1979: sample size 338 employed, 494 non-employed	H, yearly hours (weeks worked × average weekly hours) W, hourly wage, SS, (annual earnings/annual hours) Y, total net non-labor	Linear labor supply Convex (piecewise linear)	ML-convex	{1.18, 0.66}	0.52
Hausman (1981)	US Panel Study of Income Dynamics 1975: sample size 575 participants, 510 non-participants	H, annual hours of work W, directly reported hourly wage rates, SS ^a Y, transfer and asset income with 8% return to financial assets	Linear labor supply Convex (piecewise-linear) and non-convex (fixed costs)	ML-convex, random preferences ML-fixed costs random preferences (on income coefficient)	0.995 0.906	-0.121 -0.132
Kaiser et al. (1992)	German SocioEconomic Panel 1983: sample size 1076 employed, 2284 non-employed, non-retired	H, yearly hours W, hourly wage, SS (annual earnings/annual hours) Y, income from rents, capital income and transfer payments	Linear labor supply Convex (piecewise linear)	ML-convex	1.04	-0.18

Kuismanen (1997)	Finnish Labor Force Survey 1989 sample size: 1541 employed 485 non-employed aged 25–60	H, yearly hours W, hourly wage, SS, (annual earnings/linear) Y, Income from rents, dividends, capital income, etc.	Semi-log labor supply Convex (piecewise linear)	Survey data ML-convex fixed preferences Tax register data ML-convex, random preferences	-0.01 0.01	0.27 0.11
Triest (1990)	US Panel Study of Income Dynamics 1983: sample size 715 employed, 263 non-employed, aged 25–55	H, yearly hours in all jobs W, average hourly earnings, SS, ^h (earnings/hours) Y, rents, dividends, interests, etc.	Linear labor supply Convex and piecewise linear	Full sample ML-convex, random preferences (on income coefficient) Workers only ML-convex, random preferences (on income coefficient)	0.97	-0.33 -0.17
van Soest et al. (1990)	Dutch Strategic Labor Market Research Survey 1985: sample size 331 participants, 470 non-participants	H, average weekly working hours W, net hourly wage, SS (earnings/hours) Y, other incomes	Linear labor supply Convex (piecewise linear)	M-convex	0.79	-0.23

^a SS signifies that wages are predicted via linear selectivity adjusted regression, in this particular study selection bias was argued not to be important. In each of the studies for married women reported here, we leave readers to refer to the original source for details of identification strategy and included regressors.

^b This column indicates whether preferences are treated as random in specifications, in addition to “optimization” or measurement errors that are always incorporated in specifications. Unless stated otherwise, random preferences without further indications means that only intercept coefficients are stochastic. For further information, see Eq. (6.59) and related discussion.

^c Hourly earnings and consumption based other income are introduced to the LFS via an instrumental regression on the FES.

^d See Eq. (6.62).

^e In the results reported here, unobserved wages are predicted using selection adjusted wage equations except for the fixed preference linear labor supply case in which unobservable wages are integrated out of likelihood.

^f Only truncated sample of workers used in estimation.

^g Predicted wage used for those out of employment, selection bias is argued not to be important for this sample.

^h The results we report for the “workers only” use observed wages; the author reports similar results using imputed wages.

and the differentiable constraint case, and the authors only consider convexified budget sets.

MaCurdy, Green, and Paarsch consider three specifications of the labor supply function, along with both an additive and a multiplicative structure for the measurement error term. The first is a linear labor supply function with substitution and income effects constant across individuals. The second assumes the substitution coefficient to vary across individuals, and the third allows the income effects to vary. The third approach is also taken by Hausman (1981) and Blomquist (1983). In the piece-wise linear equations, the authors use only the additive structure for measurement error rather than the multiplicative, since the latter implies the unattractive feature that earnings are observed without error. The evidence from all models suggests a strong influence of the implicit inequality restrictions invoked by the maximum likelihood procedure. This offers an explanation for the divergent results of previous research relying on various empirical methodologies.

Arrufat and Zabalza (1986) use British cross-sectional data on married women from the 1974 General Household Survey to estimate a model of female labor supply that reflects the joint decision on labor force participation and hours, the non-linear budget constraint created by income taxation, the effect of heterogeneous preferences, and the existence of optimization errors. These optimization errors cause agents' actual position on the budget constraint to differ from their preferred position. The structural model is based on the following CES family utility index defined over net family income, x , and wife's leisure, l :

$$u = [x^{-\rho} + \alpha l^{-\rho}]^{-1/\rho} \quad (6.60)$$

with error structure

$$\alpha = \exp[\beta Z - \xi], \quad x/l = (x/l)^* \exp(\varepsilon), \quad (6.61)$$

where $(x/l)^*$ is the utility-maximizing income-leisure ratio, Z is a vector of personal characteristics, β is a vector of parameters, and (ξ, ε) is distributed bivariate normal $(0, 0, \sigma_\xi^2, \sigma_\varepsilon^2, 0)$. The budget constraint looks like Fig. 15 in (x, l) space. The maximum likelihood estimator yields an estimated elasticity of substitution of 1.21. The elasticities with respect to own wages, husband's wages, and unearned family income are 2.03, -1.27 , and -0.20 . This own wage elasticity of approximately two is larger than those estimated in previous studies using British data.

Blundell et al. (1988) estimate a generalized version of the Stone–Geary labor supply model using a sample of almost 1400 married women from the British Family Expenditure Survey for 1980. A truncated likelihood approach was used that considered hours of work conditional on participation. The preference specification was chosen according to standard likelihood diagnostics.⁴⁵ Although uncompensated wage elasticities were small, the compensated elasticities were found to be quite large and positive across a wide range of demographic groups. This model was then used to simulate a number of reforms to the British tax system in Blundell et al. (1988).

⁴⁵ See Blundell and Meghir (1986).

Friedberg (1995) analyzes data from the United States March Current Population Survey. She uses a convex budget set with a piece-wise linear constraint for studying progressive taxes and the social security earnings test, and assumes a linear functional form for the labor supply equation. The Heckman sample selection technique is used to predict non-participant wages in the labor supply equation. Maximum likelihood estimates of the model yield a compensated wage elasticity of 1.12, an uncompensated wage elasticity of 0.36, and an income elasticity of -0.76 .

Van Soest et al. (1990) analyze a cross-section of Dutch households from a 1985 labor mobility survey by the Organization of Strategic Labor Market Research. They consider a piece-wise linear framework with a convex budget set and normally distributed random preferences and optimization errors. As a second specification, they estimate a simple reduced form model of the demand side of the labor market, in which employers offer wage-hours packages and individuals choose among a limited number of these offers. The authors impose the distributional assumptions stated following Eq. (6.61), and estimate the models using maximum likelihood. In their second specification, the error term ν is replaced by a job offer mechanism, which treats the number of hours worked as a discrete rather than a continuous random variable. Their results imply wage-rate elasticities of 0.65 and 0.79 for women and 0.12 and 0.10 for men. These and the estimated income elasticities are in harmony with previous work using Dutch data.

6.8.2. Non-convex budget sets: maximum likelihood and instrumental variable estimation

Hausman (1981) estimates the effect of taxation and transfers on the labor supply of a subsample of prime-age husbands, wives, and female family heads who have children under the age of eighteen from the 1975 PSID, treating the husband as the primary earner and the wife as the secondary earner. For husbands and wives he considers two cases: the non-convex piece-wise linear case representing a tax and transfer schedule based on actual law, and a convexified tax schedule where the effects of FICA, the earned income credit, and the standard deduction are approximated by a consistently progressive convex budget set. For female household heads he considers only the non-convex case because of the large initial non-convexity introduced by AFDC.

Hausman assumes a linear functional form for the labor supply equation. Although the wage coefficient in the hours equation is assumed to be constant across individuals, the coefficient of virtual income is assumed to vary. Blomquist (1983) also uses this approach, as discussed at the beginning of this survey. Hausman assumes that the coefficient of virtual income is the mean of the truncated normal distribution. Since it is assumed that this coefficient is non-positive, the relevant part of the distribution is to the left of zero. Hausman considers the possibility of selection bias, since market wages are unobserved for non-workers, but finds that it is not a problem in his sample. Estimation is by maximum likelihood.

For husbands, he finds that the uncompensated wage coefficient is essentially zero which accords with previous empirical findings. However, his finding of a significant income effect is at odds with prior work. Since the wage and income variables from the

convex and non-convex budget sets are similar, Hausman concludes that for estimation purposes it is probably reasonable to smooth the non-convexities created by the earned income credit, social security taxes, and the standard deduction. For wives, he finds substantial uncompensated wage and income elasticities. In addition to the convex and non-convex cases, a specification that explicitly accounts for the fixed costs of working is included for wives. The resulting wage elasticities are midway between those of husbands and those of wives.

Triest (1990) considers the sensitivity of Hausman's results to changes in the model specification. To this end, he estimates several variants of Hausman's model using a 1983 subsample of the PSID. Both the labor supply equation and the measurement error equation are linear, with the distributional assumptions stated following Eq. (6.61). A specification representing preference heterogeneity as a random income coefficient, rather than an additive disturbance, is also estimated following Hausman. Triest considers maximum likelihood estimation under the assumptions of preference heterogeneity only, measurement error only, and both heterogeneity and measurement error, in addition to instrumental variables estimation assuming only heterogeneity. In the heterogeneity-only model for women, GMM was used to estimate an IV version of the Tobit model. Triest follows Hausman by treating the convex hull of the budget set as the effective budget set in estimation.

The results, which are consistent across model specifications, suggest that the labor supply of prime-aged married men is relatively invariant to the net wage and virtual income. The finding of no virtual income effect, however, starkly contrasts with Hausman's result. Furthermore, the estimated net wage elasticities are positive and of larger magnitude than the one reported by Hausman. The results for women are more sensitive to the specification of the labor supply function. Net wage elasticities resulting from a censored estimator are similar to those of Hausman. But when a truncated estimator is used (conditioning on positive hours), estimated wage elasticities are much smaller. The same is also true (in absolute value) of the virtual income elasticities.

Blomquist and Hansson-Brusewitz (1990) estimate a potpourri of labor supply functions for married men and women in Sweden. They consider both linear and quadratic supply functions, with and without random preferences. For males, the linear fixed-preference specification is estimated first with the non-convex budget set and then using the convex hull as an approximation. The random-preference linear specification also uses this convex approximation. The fixed-preference model is a special case of the random-preference model when the constraint $\sigma_{\xi}^2 = 0$ is imposed. A likelihood-ratio test rejects this constraint at the 1% significance level. The fourth model for males includes a quadratic term in wages. A likelihood-ratio test of the null that this coefficient is zero is rejected at the 1% significance level. All estimation results for males imply a substantial compensated wage rate elasticity and a smaller income elasticity.

For females, the authors correct for sample selection bias using Heckman's two-stage technique. They offer four specifications for female labor supply: (i) linear supply function, fixed preferences, Heckman method; (ii) quadratic supply function, fixed preferences,

Heckman method; (iii) linear supply function, random preferences, Heckman method; (iv) linear supply function, fixed preferences, full-information maximum likelihood. As was the case for men, an asymptotic likelihood-ratio test rejects the null hypothesis of fixed-preferences. The wide differences in compensated wage elasticities between women and men, which are reported in Tables 1 and 2, are somewhat misleading since the wage rate elasticities for both groups are evaluated at different points on the labor supply functions. Using a quadratic supply function and evaluating the female wage rate elasticity at the mean male sample values yields an estimate of 0.10, comparable to the 0.12 estimate for males.

Bourgiugnon and Magnac (1990) estimate labor supply functions separately for a sample of French married men and women, using a piece-wise linear constraint and a convexified budget set. They assume that family labor-supply decisions are sequential, with the men first choosing their labor supply under the assumption of no other labor income in the family. Then the other family members choose their own labor supply, taking the household head's labor supply as given. Under the assumption that (ε, ξ) is distributed bivariate normal $(0, 0, \sigma_\varepsilon^2, \sigma_\xi^2, 0)$, where ε represents preference heterogeneity and ξ is a measurement error term, the authors estimate the model using maximum likelihood. The authors also consider the joint labor supply model, assuming that the original kinked budget constraint is approximated by some differentiable function as in Section 6.1.3. They use an instrumental variables estimator to estimate this model.

Flood and MaCurdy (1992) apply the full spectrum of methods for convex budget sets to a 1983 cross section of prime-age, married, Swedish men from the Swedish Household Market and Non-market Activities Survey (HUS), in hopes to reconcile the discrepant results of previous work on the disincentive effects of Swedish income taxes. They consider the piece-wise linear and differentiable constraint approaches, estimation using both instrumental variables and maximum likelihood, various functional forms for both labor supply and the structure of measurement error in hours worked, and extensions to incorporate family labor supply and lifecycle considerations. The authors also explore the viability of the standard exogeneity assumptions that underlie the maximum likelihood estimation approach.

Flood and MaCurdy report maximum likelihood results for the following specifications: piecewise-linear and the differentiable method with additive errors, linear labor-supply with and without multiplicative error, and logarithmic labor supply with and without multiplicative error.⁴⁶ These specifications yield uncompensated and compensated wage elasticities of around 0.15 and 0.20, slightly higher than those reported by Blomquist (1983).⁴⁷ The authors note the minor consequences both of accounting for measurement error and of using the piece-wise linear as opposed to the differentiable approach. This is

⁴⁶ See Section 6.2.2 for a discussion of the multiplicative measurement error structure and the logarithmic labor supply function. Also, recall from Section 6.4.3 that specifications relying on differentiable budget constraints need not assume any measurement error to render the empirical model data-consistent.

⁴⁷ Blomquist and Newey (1997) find slightly lower wage elasticities and slightly higher income elasticities using their non-parametric formulation of the piece-wise linear labor supply model.

consistent with the findings of Hausman (1981). The instrumental variable estimation results are summarized in Table 2. The key insight from these results is that the data reject the exogeneity assumptions maintained by the maximum-likelihood procedures. These assumptions dramatically influence the estimates of the substitution and income effects; conventional endogeneity tests reject the exogeneity of gross wages and all components of non-labor income. Finally, the results of Flood and MaCurdy suggest that altering the form of the structural labor-supply function produces only small changes, and neither lifecycle adjustments in the computation of virtual income nor attempts to explore the interaction of husband's and wife's labor choices substantively change the results.

Blundell et al. (1998a) present instrumental variable estimates of a labor supply model for the hours of work of married women in the UK that accounts for the endogeneity of gross wages and other income as well as accounting for selection and non-linear taxation. This model and its results are fully documented in our discussion of difference-in-differences specifications in Section 5.

6.8.3. Multiple welfare program participation

We close this section with a look at the labor supply effects of multiple welfare programs, as addressed in the working paper by Keane and Moffitt (1995).⁴⁸ They use a single-actor labor supply model to consider the joint decision of whether to work, whether to participate in AFDC, and whether to participate in the Food Stamps program. This necessitates estimation of the labor supply equation jointly with two welfare participation equations to account for the correlation between unobservables. The authors limit agents to full-time, part-time, and no work. Together with the 0/1 decision for two welfare programs, this implies twelve alternatives over which the utility function must be maximized.

Keane and Moffitt estimate the model using a sample from the 1984 SIPP of 968 female heads of households with children. Explanatory variables used include education, age, number of children, region, SMSA, and state characteristics. Using their estimates, they compute the uncompensated wage elasticity, at variable means, as 1.94. This is at the high end of prior estimates, which seems reasonable since this is a study of female-heads rather than married women. They estimate an income elasticity of -0.21 , a small (in absolute value) estimate which they attribute to measurement error in unearned income. The estimate of λ , the parameter indicating the extent to which welfare stigma is additive, is 0.05.

In addition, the authors simulate policy changes in the AFDC and Food Stamp programs. First, they consider the impact on predicted choices of reducing the AFDC benefit reduction rate from 100% to 50%. This has limited effect on labor supply but increases both AFDC and Food Stamp participation. Second, they find that a reduction of both AFDC and Food Stamp benefit reduction rates to 10% would increase average labor supply by two hours, but would also increase AFDC participation by one third and Food

⁴⁸ See also the discussion of family labor supply and program participation models in Section 7.

Stamp participation by one fourth. This would lead to an 80% increase in net costs even accounting for the increase in tax revenue. Third, they find that increasing gross wages by one dollar would increase average labor supply by about 3.5 h and reduce AFDC and Food Stamp participation, but that a government financed minimum wage of five dollars could accomplish the same changes at lower cost. Finally, they simulate the impact of the 1981 increase in the AFDC tax rate from 67% to 100%, by comparing predictions for the 1984 sample using both the 1980 and the 1984 welfare rules. They find decreased AFDC participation, with many AFDC recipients working part-time in 1980 either leaving AFDC to work full-time or quitting their jobs. As a result, they find an increase of 14.6% in the percentage of AFDC recipients who do not work. All of these results closely match those that were actually observed.

7. Family labor supply

This section considers two important developments to the family labor supply model. The first concerns the extension to cover non-participation and non-convex budget constraints. The second refers to the development of a collective framework for the study of family labor supply. Both are likely to be critical to our understanding of the impact of tax and welfare reforms discussed in Section 2 and our interpretation of the changing patterns of female and male labor supply documented in Section 3.

We develop the analysis of non-participation and non-convex budget constraints in a family labor supply context in two steps. The first simply accounts for non-participation via a corner solution in the labor supply of one of the individuals. The second incorporates a more general specification for welfare programs and fixed costs.

The discussion of the collective labor supply model that follows draws heavily from the recent literature on the specification and identification of these models. We also consider the robustness to alternative model specifications and to the introduction of home production. We round up this section with a review of the results from recent empirical applications of the family labor supply model.

7.1. The basic economic model of family labor supply

The standard approach to family labor supply modeling, discussed in Section 4.1.2, extends the consumption-leisure choice problem to include two leisure decisions. As will be clear from our discussion of collective family labor supply models in Section 7.2, this simple extension of the standard model is controversial. However, it is attractive because it extends naturally to cover multiperiod labor supply decisions⁴⁹ and, perhaps more interestingly, it can be used to place the discussion of non-linear budget constraints, fixed costs and participation problems introduced in Section 6 in a family labor supply setting.

⁴⁹ In Section 8, we consider in detail the issues that arise in a multiperiod labor supply model with participation.

7.1.1. Family labor supply with participation

The standard family labor supply model concerns the labor supply behavior of a household comprised of two working-age individuals. Children and other dependants are included in the vector of observable household characteristics, X . We assume that families maximize joint utility over consumption, C , and the leisure time of both workers $U(C, L_1, L_2, X)$ where L_1 and L_2 are the hours of leisure for two family members. For expositional reasons, we also consider non-participation for the second individual. The first-order conditions for this problem (see Eqs. (4.15) and (4.16)) can be written

$$U_{L_1} = \lambda W_1 \quad \text{and} \quad U_{L_2} \geq \lambda W_2, \tag{7.1}$$

where strict equality holds in the latter marginal condition when individual 2 works. Substituting out for the marginal utility of consumption $U_c = \lambda$ results in

$$U_{L_1} - U_c W_1 = 0 \quad \text{and} \quad U_{L_2} - U_c W_2 \geq 0, \tag{7.2}$$

in which each marginal utility is a function of L_1 and L_2 since from the budget constraint we can write consumption as $C = Y + W_1(T - L_1) + W_2(T - L_2)$.

The optimal labor supply choices in this framework satisfy the standard consumer demand restrictions of symmetry, negative semidefiniteness of the Slutsky substitution matrix, and homogeneity of degree zero in wages, prices and full income. Homogeneity is satisfied by specifying the labor supply model in terms of real wages and real incomes. Symmetry requires equality between the Slutsky cross-substitution terms

$$\frac{\partial L_i}{\partial W_j} + L^j \frac{\partial L_i}{\partial M} = \frac{\partial L_j}{\partial W_i} + L^i \frac{\partial L_j}{\partial M} \quad \text{for } i \neq j. \tag{7.3}$$

The negativity restriction generalizes the Slutsky condition on the sign of compensated labor supply by requiring the matrix of the own- and cross-Slutsky substitution terms to be negative semidefinite. To complete the specification, we may add taste heterogeneity terms to the marginal utility conditions to produce

$$U_{L_1} - U_c W_1 - \varepsilon_1 = 0 \tag{7.4}$$

and

$$U_{L_2} - U_c W_2 - \varepsilon_2 \geq 0, \tag{7.5}$$

with joint density $g(\varepsilon_1, \varepsilon_2)$. These terms are introduced directly into marginal utility rather than into the labor supply equations themselves (in contrast to Section 6) to preserve the taste heterogeneity interpretation of the error terms in a model with multiple labor supply decisions.

These first order conditions describe two regimes of behavior:

- (i) both spouses participate: $H_1 \equiv T - L_1 > 0, H_2 \equiv T - L_2 > 0,$
- (ii) individual 2 does not participate: $H_1 \equiv T - L_1 > 0, H_2 \equiv T - L_2 = 0,$

where H_1 and H_2 are the hours of work choices of each of the two adults in the family.

The sample likelihood for this model has two contributions and is similar to the sample likelihood for the single worker corner solution model described in Section 6. Ignoring taxation and measurement error, and additionally assuming wages are known and exogenous, the likelihood contribution for families observed in the first regime where both spouses work is given by

$$l_{H_1>0, H_2>0} = |J|g(U_{L_1} - U_C W_1, U_{L_2} - U_C W_2), \tag{7.6}$$

where the term $|J|$ is the Jacobian term that corresponds to Eq. (6.15) in the single worker case. This term is the determinant of the own and cross derivative matrix of ε_1 and ε_2 in terms of hours of work. It recognizes that ε_1 and ε_2 are non-linear functions of H_1 and H_2 .

For the non-participation regime we note that $\varepsilon_2 > U_{L_2} - U_C W_2$ defines a reservation wage condition, so that the choice of L_1 involves solving the marginal conditions with $L_2 = T$ which we write as $\tilde{U}_{L_1} - \tilde{U}_C W_1 - \varepsilon_1 = 0$. Consequently, the likelihood contribution for observations on families in the regime where the second worker does not participate is given by

$$l_{H_1>0, H_2=0} = |K| \int_{U_{L_2} - U_C W_2}^{\infty} g(\tilde{U}_{L_1} - \tilde{U}_C W_1, \varepsilon_2) d\varepsilon_2, \tag{7.7}$$

where again the term $|K|$ is the corresponding Jacobian term. It is interesting to note that the Slutsky symmetry and negativity conditions are sufficient to guarantee that both of the matrices J and K in the Jacobian terms are positive definite.⁵⁰

Missing wages, and also the endogeneity of gross wages, is best addressed by rewriting the marginal conditions (7.4) and (7.5) so that they are log linear in wages, i.e.

$$\ln\left(\frac{U_{L_1}}{U_C}\right) - \ln W_1 - \tilde{\varepsilon}_1 = 0 \tag{7.8}$$

and

$$\ln\left(\frac{U_{L_2}}{U_C}\right) - \ln W_2 - \tilde{\varepsilon}_2 \geq 0, \tag{7.9}$$

in which case wage equations of the form $\ln W_j = Z_j' \gamma_j + \xi_j$ can be easily incorporated. Education variables, typically excluded from preferences but included in the Z variables in each wage equation, can then be used to identify the model – under the strong assumption that education is uncorrelated with unobserved heterogeneity in labor supply.

Finally, in order to estimate the wage equation on the sample of observed wages for which $H_2 > 0$, one needs to account for the selection bias induced by correlation in the unobservables and ξ_2 . The parameters of the wage equation are identified through the exclusion of the exogenous income variable Y which does enter the determination of participation.

⁵⁰ See Ransom (1987) and Van Soest et al. (1990).

7.1.2. Family labor supply with taxes and program participation

The extension of these models to allow for convex piecewise linear budget constraints is a straightforward adaptation of the discussion presented in Section 6. Non-convexities in the budget constraint and welfare program participation pose further difficulties because direct comparisons of utilities are required as documented in Section 6.

Consider the problem of jointly modeling the work and welfare participation decisions of a two-worker family. Suppose we assume that families maximize a standard utility function of the form

$$U = U(L_1, L_2, C, \varepsilon) - \eta P_B, \quad (7.10)$$

where, in keeping with the notation in Section 6.6.3, P_B is a 0-1 program participation indicator. Unobservable preference heterogeneity is entered directly in utility through the vector ε which correspond to the ε_1 and ε_2 terms in Eqs. (7.4) and (7.5). As in Eq. (6.49) the ηP_B term is included so as to capture the costs of being on welfare, including “welfare stigma”. The budget constraint that determines consumption is given by

$$C = W_1 H_1 + W_2 H_2 + Y - T(Y, W_1 H_1, W_2 H_2) + B P_B, \quad (7.11)$$

where Y is unearned income, $T(\cdot)$ is a tax function, and B is program benefits.

Due to the computational difficulties encountered when considering the hours and participation of two persons with non-linear budget constraints, the approach outlined in Section 6.7 offers a tractable method for estimating the family labor supply model. In particular, given an assumed joint distribution for unobservable tastes components, errors determining wages, and welfare stigma, one can compute a probability that each family member selects among alternative employment and program participation states. This in turn defines a sample log likelihood of the form Eq. (6.57). As described in Section 6.7, this formulation requires that each individual be placed into a limited set of preassigned work states, even though observed hours worked take many more values. To overcome this issue, analyses applying this approach invariably introduce measurement error in hours of work to admit hours to deviate from the discrete values assumed for the choice set, as described in likelihood (6.58).

7.1.3. Drawbacks of the standard family labor supply

The “unitary” model described in Section 7.1 implies three broad groups of testable restrictions. The first set of restrictions covers the standard consumer demand restrictions of symmetry, negative semidefiniteness of the Slutsky substitution matrix, and homogeneity of degree zero in wages, prices and full income, see Eq. (7.3) and the related discussion above. The second set of restrictions refer to income pooling. This is the condition which implies that, as far as the household’s utility-maximizing choice of family labor supplies are concerned, one can combine all sources of non-labor income into a single unearned income measure, Y . If, for example, each of the two individuals has *private* unearned income Y_1 or Y_2 respectively, then pooling implies

$$L_i^{Y_1} = L_i^{Y_2}, \quad \text{for } i = 1 \text{ and } 2, \text{ where } L_i^{Y_1} \equiv \frac{\partial L_i}{\partial Y_1}. \quad (7.12)$$

This is a controversial assumption in the welfare reform debate since it implies that the source of non-labor income is irrelevant in within-family labor supply decisions.

Finally, there are the non-participation or “corner solution” conditions which state that if one individual is at a corner solution, it is the reservation wage of that individual rather than the market wage that affects the labor supply decision of the partner. As in the case of the income pooling assumption, this is far from innocuous, implying as it does that the “outside option” value of paid work for a non-participant does not influence the allocation of consumption and leisure within the household.

7.2. The collective model of family labor supply

Recent research has focused on relaxing the assumptions of symmetry and income pooling, seeking instead solutions from efficient bargaining theory. The advantages of the unitary model of family labor supply are well known. As we have seen they allow the direct utilization of consumer theory, recovering preferences from observed behavior in an unambiguous way and providing a framework for interpretation of empirical results. One can then use standard welfare economics to evaluate tax and welfare reform. An argument often raised by critics of the standard model is that it treats all individuals in the family as a single decision making unit rather than as if they were a collection of individuals. Moreover, researchers often conclude that allocations within the family derived from the unitary model cannot be recovered in a meaningful way. This conclusion is too strong. The standard decentralization theorems from consumer theory⁵¹ apply equally well to individual members’ utilities in a “unitary” household.

Suppose there are no public goods and that individual utilities are weakly separable over their private consumption and leisure. Let C_1 , C_2 , L_1 , and L_2 refer to the private consumption and leisure choices of individuals 1 and 2. Defining the private consumption of the second individual in the same way, we may write the within-period family utility as

$$F(U_1(C_1, L_1, X), U_2(C_2, L_2, X)), \quad (7.13)$$

where $U_1(C_1, L_1, X)$ is the sub-utility for the husband and $U_2(C_2, L_2, X)$ is the sub-utility for the wife. Where family utility has this weakly separable form, decentralization follows two-stage budgeting: total household (full) income is allocated among all household members, and then individuals act as if they are making their labor supply and consumption decisions conditional on this initial stage outlay.

Even if consumption goods are privately consumed, they are typically only measured at the household level – so that the individual consumptions are unobserved or “latent” to the economist. However, a single observed (privately-consumed) good – labor supply in this case – per sub-utility is often sufficient to identify decentralized preferences. This condi-

⁵¹ See Gorman (1958), for example.

tion on a single exclusive good per sub-utility corresponds to the identification condition in generalizations of weak separability that allow overlapping goods across groups.

So what advantages does the collective approach offer? It effectively relaxes the income allocation rule among individuals so that this allocation may depend on relative wages and other variables in a way that reflects the bargaining position of individuals within the family, rather than reflecting the marginal conditions underlying the joint optimizing framework of the traditional unitary approach. Even when individuals within the family are altruistic and allocations are Pareto Efficient, the allocation rule can deviate from the optimal rule in the traditional model.

7.2.1. A summary of the collective labor supply model

In this work,⁵² each family member either maximizes an “egoistic” utility, $U_1(C, L_1, X)$, or a “caring” utility function, $F_j(U_1(C_1, L_1, X), U_2(C_2, L_2, X))$, for $j = 1$ and 2 . Notice that this mirrors the separability assumption in Eq. (7.13). That is, the only way L_2 enters the (sub-) utility of individual 1 is through the (sub-)utility of individual 2; there is no direct impact on the utility of the partner.

Applications of this model assume that the decision process generates Pareto-efficient outcomes, all goods are privately consumed and there is no household production. The implications of relaxing these latter two assumptions are important and we consider them below.

The collective framework states the family labor supply problem as follows:

$$\max[\theta U_1 + (1 - \theta)U_2], \quad \text{s.t.} \quad C_1 + C_2 + W_1L_1 + W_2L_2 = M, \quad (7.14)$$

where θ is the utility weight for person 1, given by some non-negative function $\theta = f(W_1, W_2, M)$. This is equivalent to a sharing rule, or decentralized solution, in which individual 1 gets income $M - \varphi(W_1, W_2, X, M)$ and then allocates according to the rule

$$\max U_1, \quad \text{s.t.} \quad C_1 + W_1L_1 = M - \varphi(W_1, W_2, X, M), \quad (7.15)$$

where $\varphi(W_1, W_2, X, M)$ is defined as the sharing rule.

Given Pareto efficiency and the standard neoclassical assumptions on individual utilities, the conditions identifying preferences and the sharing rule (up to a linear translation) simply require one observable and assignable private good – here assumed to be the individual’s leisure. The intuition behind identification is simple: under the exclusive good assumption, the spouse’s wage can only have an effect through the sharing rule. Variation of income and the wage then permit consistent estimation of the marginal rate of substitution in the sharing rule. A researcher can do this for both spouses and, since the sharing rule must sum to one, recover the partial derivatives of the sharing rule.

Although the standard symmetry, income pooling, and participation conditions are not implications of this model, one can derive alternative testable restrictions. If separate

⁵² The most lucid statement of this argument occurs in the papers on household labor supply by Chiappori (1988, 1992).

income sources are unobservable to the econometrician and both individuals work in the labor market (i.e., there are no corner solutions for leisure), the only restrictions implied are those corresponding to the Slutsky conditions. These are expressed in terms of the derivatives of the labor supply equations with respect to the wage and income variables. Assuming the income derivatives are non-zero the collective model implies the differential equations:

$$\alpha_m \frac{L_1^{W_2}}{L_1^M} + \alpha \frac{\partial}{\partial M} \frac{L_1^{W_2}}{L_1^M} - \alpha_{w_2} = 0, \quad \beta_M \frac{L_2^{W_1}}{L_2^M} + \beta \frac{\partial}{\partial M} \frac{L_2^{W_1}}{L_2^M} - \beta_{w_1} = 0, \quad (7.16)$$

in which α is given by

$$\alpha = - \left[\frac{\frac{\partial}{\partial M} \left(\frac{L_1^{W_2}}{L_1^M} \right) \frac{L_2^{W_1}}{L_2^M} - \frac{\partial}{\partial W_2} \left(\frac{L_2^{W_1}}{L_2^M} \right)}{\frac{\partial}{\partial M} \left(\frac{L_2^{W_1}}{L_2^M} \right) \frac{L_1^{W_2}}{L_1^M} - \frac{\partial}{\partial W_1} \left(\frac{L_1^{W_2}}{L_1^M} \right)} \right]^{-1},$$

$\beta = 1 - \alpha$ and where superscripts denote partial derivatives. The terms $\alpha_M, \beta_M, \alpha_{w_j}, \beta_{w_j}$ are the corresponding income and wage derivatives. Eqs. (7.16) are analogous to the Slutsky symmetry conditions, while the Slutsky inequalities are matched by

$$\frac{L_1^{W_1}}{L_1^M} + \left(T - L_1 - \frac{\beta}{\alpha} \frac{L_1^{W_2}}{L_1^M} \right) \leq 0, \quad \frac{L_2^{W_2}}{L_2^M} + \left(T - L_2 - \frac{\beta}{\alpha} \frac{L_2^{W_1}}{L_2^M} \right) \leq 0. \quad (7.17)$$

These restrictions are sufficient for recovering preferences and the sharing rule (up to an additive constant). Indeed, the derivatives of the sharing rule, $\varphi(W_1, W_2, X, M)$, have the form

$$\frac{\partial \varphi}{\partial M} = \alpha, \quad \frac{\partial \varphi}{\partial W_2} = \frac{\partial L_1^{W_2}}{\partial L_1^M} \alpha, \quad \frac{\partial \varphi}{\partial W_1} = \frac{\partial L_2^{W_1}}{\partial L_2^M} (\alpha - 1). \quad (7.18)$$

Consequently, having estimated unrestricted family labor supply functions in terms of wages for each individual and full income, the researcher can recover individual preferences and the sharing rule.

7.2.2. Household production

The introduction of household production is problematic for estimation of the collective model since, as we have seen, this model exploits the exclusion restriction on the other individual's wage to identify the sharing rule under egoistic or caring preferences. Unless we assume that the household production good is marketable, identification up to an additive constant is lost.

It is reasonable to assume that for many families, non-market time is spent in the active production of home produced goods.⁵³ These may include activities for which a perfect substitute is directly available in the market, housework or home decoration for example;

⁵³ See also Apps and Rees (1997, 1998).

but they may also include activities for which a perfect substitute is not readily available, childcare, for example. What is of particular interest is when there is no direct substitute available and both spouses non-market time enter the production of the home produced good. For the standard non-separable (unitary) model of household labor supply, this has little direct impact on the labor supply function – it simply acts as if it were leisure time. However, that is not the case in the separable model.

Suppose there is a home-produced good, G , that requires inputs of time by both household members. Denoting these time inputs by t_1 and t_2 , one can write the production technology as

$$G = g(t_1, t_2), \quad (7.19)$$

where we assume that g is a concave function. Time not spent in the labor market can be used for two purposes, pure leisure or home production. If t_1 and t_2 are recorded by individuals in a time-use diary survey,⁵⁴ then the characteristics of $g(t_1, t_2)$ can be recovered. However, since leisure enters household utility in a general way in the family utility function, and since $g(t_1, t_2)$ is concave, family utility remains a concave function of non-market time and consumption. Consequently, the labor supply equations describing hours of work and labor market participation are observationally equivalent to those for the model without household production.

An interesting special case occurs when family utility is separable in the non-market time of each individual. In this case, family utility with household production can be written

$$F(U_1(C, L_1, G_1, X), U_2(C, L_2, G_2, X)), \quad (7.20)$$

where G_1 and G_2 are the private consumptions of the home-produced good so that $G_1 + G_2 = G$. (Alternatively, if the home-produced good is a public good such as childcare, then G itself enters each sub-utility.)

If the consumption of household production is not observed then the presence of G in each sub-utility does upset the separability assumption. To see this, suppose household production technology exhibits constant returns to scale. Then the implicit price, or unit cost, of household production is simply a function of the two wage rates:

$$P^* = r(W_1, W_2). \quad (7.21)$$

In the model without G , the weak separability condition is sufficient for each labor supply to be written in terms of the own wage and the allocation of full income. Introducing G implies that P^* , and therefore W_1 and W_2 , enter each labor supply. Consequently, the household production function is sufficient to break the separability condition and therefore the exclusion restriction on the other household member's wage in the labor supply equation. In this case, individual utilities are not recoverable. The only case in which this does not occur is when the household production good, G , is marketable and when the

⁵⁴ See Kapteyn and Kooreman (1993), for example.

solution is interior rather than at a corner. In this case, one sets P^* equal to the observable market price for the home produced good which will not depend on individual wages. This issue becomes more problematic for the collective model described below in which the exclusion restriction on the other individual's wage is required for identification. Of course, if the household production technology exhibits constant returns then P^* in (7.21) depends only on the two wages and the income terms in the sharing rule provide testable restrictions.⁵⁵

7.3. Some empirical findings for the family labor supply model

7.3.1. The unitary model

Recent studies build on the original work of Ashenfelter and Heckman (1974), Rosen (1978), Wales and Woodland (1976) and Smith (1977). These include Attanasio and MaCurdy (1997), Blundell and Walker (1982, 1986), Browning et al. (1985), Hausman and Ruud (1986), Hoynes (1996), Kooreman and Kapteyn (1986), Ransom (1987) and Van Soest (1995). Perhaps the most important issues raised in these studies are those concerning the degree of substitution between the labor supplied by different family members and the mixture of continuous hours and discrete participation choices.

Researchers have taken two modeling approaches. The first is to work with the bivariate censored model and allow continuous choices over hours of work. Ransom (1987), for example, takes this approach. The second approach is to simplify the hours choices to a set of discrete alternatives but to allow for fixed costs and program participation. Hoynes (1996) is an example of this. In addition to accounting for the discrete or censored nature of the data in a bivariate framework, researchers who have implemented empirical models of family labor supply have also been concerned with choosing the appropriate conditioning variables. Attanasio and MaCurdy (1997), for example, adopt a marginal rate of substitution framework for their analysis while Blundell and Walker (1986) use a consumption-based measure of non-labor income in a Marshallian model of family labor supply. Browning et al. (1985) work with a Frisch representation of family labor supplies and commodity demands (see Section 4 for a detailed discussion of these alternative choices of conditioning variables and the interpretation of the resulting elasticities). Rather than covering all studies in this discussion we have decided to single out a small number of studies that provide a useful guide to empirical models in the literature.

7.3.2. Continuous hours models with censoring

Ransom (1987) provides an analysis of family hours-of-work decisions using a sample of 1210 intact families drawn from the 1976 PSID. He restricts the sample to families with no self-employed members and in which the husband is working. Consequently, the only censoring occurs for female hours of work. This study makes a particularly convenient starting point for describing structural estimation in family labor supply models. However,

⁵⁵ Chiappori (1997) shows identification of the sharing rule up to some function of W_1 and W_2 in this case.

although Ransom accounts for censoring he does not account for the endogeneity of hourly wages or virtual income. Attanasio and MaCurdy (1997) relax these exogeneity restrictions and we discuss their study further below.

To interpret Ransom's results, consider the marginal utility conditions (7.4) and (7.5) for quadratic utility:

$$\begin{aligned}
 -UL_1 + W_1 U_C &= \alpha_1 + \alpha_3 W_1 - \beta_{11} H_1 - \beta_{33} W_1 (W_1 H_1 + W_2 H_2 + Y) - \beta_{12} H_2 \\
 &\quad + \beta_{13} (2W_1 H_1 + W_2 H_2 + Y) + \beta_{23} W_1 H_2
 \end{aligned} \tag{7.22}$$

and

$$\begin{aligned}
 -UL_2 + W_1 U_C &= \alpha_2 + \alpha_3 W_2 - \beta_{22} H_2 - \beta_{33} W_2 (W_1 H_1 + W_2 H_2 + Y) - \beta_{12} H_2 \\
 &\quad + \beta_{23} (2W_2 H_2 + W_1 H_1 + Y) + \beta_{13} W_2 H_1.
 \end{aligned} \tag{7.23}$$

Ransom then allows α_1 and α_2 to each be a linear function of observable characteristics and unobservable mean zero normal random variates ε_1 and ε_2 .

In contrast, Hausman and Ruud (1986) work directly with quadratic labor supply curves in their estimation of family labor supply for a sample of 1991 families in the 1976 PSID based on a similar selection to that of Ransom. This eliminates the need for jacobian terms but makes the introduction of random preference errors more difficult. The Hausman and Ruud indirect utility has the form

$$V(W_1, W_2, Y) = \exp(\beta_1 W_1 + \beta_2 W_2) Y^*, \tag{7.24}$$

where Y^* is given by the quadratic

$$Y^* = Y + \theta + \delta_1 W_1 + \delta_2 W_2 + 0.5(\gamma_1 W_1^2 + \gamma_2 W_2^2 + \alpha W_1 W_2). \tag{7.25}$$

From Roy's identity, hours of work are given by

$$H_1 = \delta_1 + \beta_1 Y^* + \gamma_1 W_1 + \frac{\alpha}{2} W_2, \tag{7.26}$$

$$H_2 = \delta_2 + \beta_2 Y^* + \gamma_2 W_2 + \frac{\alpha}{2} W_1, \tag{7.27}$$

where Y^* is defined in Eq. (7.25). Hausman and Ruud append additive normal errors to Eqs. (7.26) and (7.27) and estimation follows from the maximizing of a bivariate censored likelihood whose contributions are of the form (7.6) and (7.7). They take careful account of the non-linear surface of the budget constraint induced by the piecewise linear nature of the tax system. However, as was noted above, one cannot easily interpret this additive error structure as random preference variation.

The estimates Ransom presents are plausible. He finds a small compensated elasticity of 0.04 for men, and a larger one of 0.73 for women. He also finds smaller income elasticities for men, -0.03 versus -0.21 for women. (The uncompensated wage elasticity for men is

slightly negative.) These average elasticity estimates are close to those reported in Hausman and Ruud. However, the Ransom model does not perform particularly well in replicating the within-sample distribution of female hours of work. While actual mean hours of work for women are 1376 per year, the model predicts only 634. Moreover, although fewer than 50% of women in the sample are recorded as working, the model predicts that the majority of women work. Although there are many ways in which the model might be misspecified, the most likely culprits are the “Tobit” assumption on participation that rules out fixed costs of work (see Cogan, 1980, 1981) and the use of predicted wages (although corrected for selectivity) in a non-linear labor supply model. Interestingly, the Hausman and Ruud specification allows for a fixed cost parameter for female labor supply which is found to be significant.

Attanasio and MaCurdy (1997) address endogeneity issues and generalize the form of censoring in their study of families in the repeated cross-sections of the US Consumers Expenditure Survey (CEX). They choose log-linear forms for the marginal rate of substitution functions Eqs. (7.8) and (7.9). The authors selected a sample of 20,297 households from the CEX for the period 1981–1992. This CEX dataset has the dual advantages of providing consumption data directly and allowing variation over time. Attanasio and MaCurdy adopt a semi-parametric approach to correct for non-participation, and this relaxes the normality and Tobit assumptions. Their results imply a slightly negative male hours elasticity, whereas for women the corresponding elasticity is much larger in absolute value and implies a strongly upward sloping labor supply curve.

Kooreman and Kapteyn (1986) follow a similar approach but do not fully allow for random preference variation, in their study of the joint labor supply decisions of 315 households from the 1982 CBS survey for the Netherlands. As in Blundell and Walker (1986) they work directly with a Marshallian demand system and specify a second-order flexible form, derived from an Almost Ideal indirect utility function. In contrast, the Blundell and Walker study uses a Gorman polar form which retains the linearity in full income. This linearity obviates the need to specify a value for total available hours T . Kooreman and Kapteyn perform estimation using a censored likelihood in which they predict wages from a selectivity-adjusted log wage equation. They use education-level dummies as excluded instruments. Although Kooreman and Kapteyn do not provide elasticity estimates, they do report small and negative own wage responses for men and larger and positive own wage responses for women. Cross elasticities show a strong response of female hours to male wages. These results closely match those of Blundell and Walker who used a truncated likelihood estimator on a similar sample of 1378 British families from the 1980 Family Expenditure Survey, in which both husbands and wives worked. Blundell and Walker report a full set of elasticities by demographic type, and find small, positive labor supply elasticities for males and larger, positive ones for females. They also report small but significant positive cross elasticities.

7.3.3. Discrete hours choices and program participation

Van Soest (1995) introduces discrete hours choices in a study of the labor supply and

participation decisions of a sample of 2859 families from the 1987 Social Economic Panel for the Netherlands. He models a non-convex budget constraint for each family to explicitly account for the Dutch tax and benefit system. Six fixed hours intervals are defined, for husbands and for wives, resulting in a total of thirty-six possible discrete states.⁵⁶ A translog direct utility function for leisure hours and full income determines the utilities associated with each choice. Marginal utilities are, therefore, linear in log leisure hours and full income and are rendered stochastic by a choice-specific extreme-value distributed error term. Consequently choices follow a multinomial logit rule. Van Soest further extends this by introducing a jointly normal random parameter variation. He completes the specification by adding choice-specific constants and estimates using a simulated maximum likelihood estimator.

The estimation results reveal an important role for the random preference terms and the choice specific constants. Moreover, the reported elasticities are quite sensitive to changes in their specification. The most general specification suggests a small positive hours of work elasticity of around 0.1 for men and a larger elasticity of around 0.5 for females, with small, negative cross elasticities. These are reasonably consistent with the results in the Blundell and Walker study for the UK. However, Van Soest does report smaller income elasticities which often have signs at odds with theory. This may be attributable to his unearned income measure which, unlike that used in the Blundell and Walker study, is not consumption based.

Hoynes (1996) addresses the problem of jointly modeling the discrete work and welfare participation decisions of a two-worker family, in the context of the unitary labor supply model for a sample of 1010 observations on two-parent families from the 1984 Survey of Income and Program Participation (SIPP). She considers the labor supply impacts of the AFDC-UP (Aid for Families with Dependent Children – Unemployed Parent) program in the US in 1988. This program was available in 26 states in 1988 and provided AFDC benefits to two-parent families with children if the “principal earner” in the family worked less than one hundred hours/month. Hoynes models families as maximizing a standard Stone–Geary utility function (see Eq. (4.59)) of the form

$$U = \beta_1 \log(\gamma_1 - H_1) + \beta_2 \log(\gamma_2 - H_2) + \beta_C \log(C - \gamma_C) - \eta P_B, \quad (7.28)$$

where the notation is as in Eq. (7.10) and P_B is a 0-1 program participation indicator. Program benefits B in (7.11) are determined by

$$B = G - N - t_A(W_1 H_1 + W_2 H_2), \quad \text{if } B > 0 \text{ and } H_p < 100, \quad (7.29)$$

with $B = 0$ if the conditions are not met, where G is a government minimum guarantee, t_A is a government set benefit-reduction rate on earned income, and H_p is hours worked by the principal earner, either the husband or the wife.⁵⁷ Hoynes assumes that families who

⁵⁶ See also Ilmakunnas and Pudney (1990), Dickens and Lundberg (1993) and Aaberge et al. (1995) for important variations on this type of model that allows finite discrete choice sets.

⁵⁷ The principal earner is determined by program guidelines.

choose to receive AFDC-UP ($P_b = 1$) also receive food stamps. To avoid the issue of multiple program choices, (see the discussion of the Keane and Moffitt study in Section 6) and to make the problem manageable, she also assumes that only three work decisions are possible for each spouse: full-time work (40 h per week), part-time work (20 h per week), or no work. Combined with a 0/1 welfare decision, this yields 18 states over which the utility function must be maximized.

Hoynes introduces unobserved heterogeneity into the problem via the β and η parameters. She models the β 's as

$$\beta_i = \exp(X' \alpha_i + \varepsilon_i) / \Sigma[\exp(X' \alpha_i + \varepsilon_i)], \quad i = 1, 2, c, \quad (7.30)$$

where $\alpha_c = \varepsilon_c \equiv 0$.⁵⁸ η is given by

$$\eta = Z' \alpha_s + \mu + e, \quad (7.31)$$

where $e \sim N(0, \sigma_s^2)$. To further ease the estimation problem she assumes that ε and μ have a discrete support over M points (where $M = 6$ in the analysis) such that

$$\Pr(\varepsilon_1 = \varepsilon_{1k}, \varepsilon_2 = \varepsilon_{2k}, \mu = \mu_k) = \pi_k. \quad (7.32)$$

To understand the computational issues involved, consider the implications of this setup when $e = 0$ for all families. Then, for each of the M points of support, the values of ε and μ can be plugged directly into the utility function and the optimal choice over the 18 alternatives computed. Summing across the M states using the probability, π_k , of each yields the probability of each work/welfare alternative for each family. Replacing a non-zero e term complicates this only slightly – e behaves like the continuous error in a standard discrete choice model. For high enough e in each state a non-welfare option will be chosen, whereas a low e implies choice of a welfare option. Thus, within each state (ε, μ pair), two possibilities exist, with their probabilities determined by the level of ν required to make “welfare stigma” too high for welfare participation. Hoynes uses predicted wages from separate wage regression for all observations. Hence, the variables she uses to predict wages are included in the variables X, Z .

Hoynes assumes a measurement error term for each spouse, ν_1 and ν_2 , such that actual hours worked, h_1 and h_2 , relate to predicted hours worked, H_1 and H_2 , according to the functions:

$$h_1 = \exp(\nu_1)H_1 \quad \text{and} \quad h_2 = \exp(\nu_2)H_2, \quad (7.33)$$

where $\nu_j \sim N(-\sigma_j^2/2, \sigma_j^2)$.

Given the parameter estimates from this estimation procedure, Hoynes carries out several interesting simulations. First, she considers the impact of increasing the AFDC-UP guarantee amount, G , by 20%. This leads to an 18% increase in predicted participation in the program. In the population as a whole it leads to a slight reduction in employment, as

⁵⁸ This form satisfies the standard Stone–Geary restrictions that $\beta_i \geq 0$ and $\sum \beta_i = 1$.

should be expected since this is a pure income effect. However, among welfare recipients it leads to an increase in average hours worked, a composition effect caused by the addition of many working families whose incomes qualify for the program only after the policy shift. If unaccounted for, such a composition effect could lead researchers to reach incorrect conclusions concerning the impact of guarantees on labor supply. Second, Hoynes considers the impact of lowering the benefit reduction rate, t_A , by 20%. This leads to a 6% increase in participation and virtually no change in employment, as a tax rate change has both income and substitution effects. Third, she considers the impact of eliminating the $h_p < 100$ rule for eligibility. While this greatly increases eligibility, from 10.9% to 15% of the sample, it has almost no effect on program participation, since those who become eligible are already working and, thus, would receive small benefits which may be overwhelmed by other welfare costs (η). Finally, she finds that eliminating the program altogether would increase average hours worked of current recipients by 33 h for women and 46 h for men. However, this would not compensate for the loss of welfare income, as average family income for this group would still fall by approximately \$83/month.

7.3.4. Bargaining and collective models

There are relatively few empirical studies of family labor supply outside the unitary model. The original motivation for these developments came from the original studies by McElroy (1981) and Manser and Brown (1980). A number of more recent studies have used micro data to evaluate the pooling hypothesis or to recover collective preferences using exclusive goods, but these studies typically look at private consumption rather than labor supply. For example, Browning et al. (1996) use Canadian household expenditure data to examine the pooling hypothesis and to recover the derivatives of the sharing rule. Clothing in this analysis is the exclusive good providing identification.

Recent empirical studies concerning family labor supply include Lundberg (1988), Apps and Rees (1997), Kapteyn and Kooreman (1990) and Fortin and Lacroix (1997). Each of these aims to “test” the unitary model and to recover some parameters of collective preferences. Lundberg attempts to see which types of households, distinguished by demographic composition, come close to satisfying the hypotheses implied by the unitary model. The other three studies take this a step further by directly specifying and estimating labor supply equations from a collective specification. Apps and Rees (1996) specify a model to account for household production. Kooreman and Kapteyn (1990) use data on preferred hours of work to separately identify individual from collective preferences and, consequently, to identify the utility weight. Fortin and Lacroix (1997) follow closely the Chiappori framework and allow the utility weight to be a function of individual wages and unearned incomes. We briefly consider the results from each of these studies.

Kooreman and Kapteyn (1990) specify a Stone–Geary model of individual private utilities and they estimate the utility weight, which they assume to be independent of wages and income, as a constant parameter. Using data from the same 1982 Dutch survey

exploited in their 1986 study described above, they find an estimated utility weight within the unit interval, but rather imprecisely determined.⁵⁹

The focus of the Apps and Rees study is on household production and they analyze a sample of 1384 families from the Australian Bureau of Statistics 1985/86 Income Distribution Survey Sample file. All families are selected so that the male works and there is at least one child aged under 15 years. They specify a constant returns technology for household production so that the unit cost function has the form (7.21). This is then parameterized as a unit Translog function. Individual sub-utilities are given an Almost Ideal form. Since the sample does not contain information on individual consumptions of home produced or market goods, they identify the model by setting the individual income shares to the individual full incomes $W_{it} + M_{it}$. This would appear to be a rather restrictive assumption. Finally, only interior solutions are considered. They find an important role for exchange within the family with the female specializing in home production activity.

Fortin and Lacroix (1997) consider a sample of 4496 couples drawn from the 1986 Canadian Census. They follow the Chiappori framework closely and allow the utility weight to be a function of individual wages and unearned incomes. They specify the resulting sharing rule as a linear function of wages and individual unearned incomes, while they allow indirect utilities to be quadratic in own wages and individual unearned income allocations. For comparison, they specify a unitary model with a quadratic indirect utility in the two wages and total unearned income. Both specifications result in non-linear labor supply equations. For estimation, they use the sample of two working couples with instrumental variable procedures applied for the wage and income variables. Instruments were age and education polynomials, immigration dummies and regional dummies.

Fortin and Lacroix provide results for two age subgroups. For the majority of groups they reject the pooling hypothesis. The collective model restrictions are only rejected for the case in which preschool children are present, while symmetry is rejected across all groups. These results are interesting and, if confirmed across specifications accounting for endogenous participation in work and unobserved heterogeneity, they would challenge the standard family labor supply model. The results also suggest extensions to the collective model for families with young children where "home production" and public goods are likely to be of central importance.

One potentially important drawback of these models is their inability to allow for both preference heterogeneity and non-participation. This is common in modern specifications of the unitary family labor supply model as we have seen in the earlier discussion of this section. To properly assess the collective framework as an alternative empirical model, these developments are essential. This is the motivation for the Blundell et al. (1998a) study which considers the full non-parametric identification of the collective model with participation and hours choices. A general identification result is presented which is then extended to cover the introduction of unobserved heterogeneity. For the heterogeneous

⁵⁹ It should be noted that the estimated parameters and their identification rest heavily on their interpretation of data from the preferred hours question in the Dutch survey.

case a parametric form for preferences and the sharing rule is assumed. This result allows the empirical implementation of the collective model of family labor supply to be placed on an equal footing with the traditional model.

8. Structural dynamic models

This section explores extensions of the standard multiperiod model, introduced in Section 4, to allow for important dynamic features of labor supply behavior. The first considers the problem of participation, which plays a fundamental role in understanding all aspects of lifecycle behavior. Empirical models incorporating participation are obviously important for the analysis of female labor supply and retirement decisions. However, even in the simple case of continuous hours decisions examined in Section 4, we could not specify relations useful for policy simulations without assuming when a person works during the lifetime, for specifications depend on past, current, and future wages. If a person plans not to work in a period, then the wage for that period does not enter as a determinant of hours-of-work choices in other periods. To characterize the factors governing when individuals work significantly complicates empirical multiperiod models of labor supply, and the use of these models in simulations of policy scenarios. However, development of these more-elaborate models is essential to learn what is needed to account for many policy features. Given the scarcity of research on this topic, intertemporal models with non-participation or corners and saving offers many research opportunities.

The second extension considers two lifecycle models in which individuals can affect their wage growth through current investment activities: learning-by-doing models in which current work experience enters directly into the determination of future wages, and conventional human capital models in which workers endogenously choose schooling and training separately from work experience to enhance their future wages. Both of these developments imply that future events enter the optimal decision rule for hours of work and participation decisions in a more complex way.

Finally, the third extension relaxes the intertemporal separability assumption on preferences underlying the standard labor supply framework, implying that past levels of hours and consumption directly impact the marginal utility of work. Non-separabilities occur through primarily two routes: a habit persistence model, or a dynamic extension of the home production model in which inputs of time are used to produce future consumption.

8.1. The standard intertemporal labor supply model with participation

This section begins with an overview of an intertemporal labor supply model with participation which will serve as a framework for discussing the additional dynamic refinements in later subsections. Although decisions over continuous hours choices and consumption retain the simple marginal rate of substitution and Euler condition formulation described in Section 4, the participation no longer fits this simple framework. To

highlight the complexities introduced by participation, in this basic multiperiod model we presume that individuals can only choose between working and not working in a period.

8.1.1. Economic formulation

The optimization problem for participation with borrowing and saving is the solution to

$$\max_{P_t} V_t(P_t, A_t, W_t, Z_t), \quad (8.1)$$

where P_t is a zero-one dummy variable equaling unity if the individual participates in period t , V_t is the period- t value function, A_t represents beginning-period assets, W_t denotes period- t earnings from participation, and Z_t designates all non-wage variables relevant for lifecycle decision making that are not controlled by the decision maker. The elements of Z_t may be stochastic, with some uncertain in the future to the consumer. Decisions over time are linked through the asset accumulation constraint

$$A_{t+1} = (1 + r_t)(A_t - C_t + W_t P_t + Y_t), \quad (8.2)$$

where r_t is the return on assets, and Y_t is a component of Z_t representing income not attributable to earning or returns on assets. Eq. (8.2) assumes perfect capital markets.

The formulation for the value function follows from first principles in dynamic economics. Let $U(P_t, C_t, Z_t)$ be the utility function for period t , which need not depend on all or any elements of Z_t ; we include Z_t as an argument, rather than some subset of this vector, to save notation. We can write the value function as

$$V_t(P_t, A_t, W_t, Z_t) \equiv V_t^P = P_t V_t^1 + (1 - P_t) V_t^0 = V_t^P(A_t, W_t, Z_t), \quad (8.3)$$

where

$$V_t^1 = \max_{C_t} \left[U(1, C_t, Z_t) + \kappa E_t \left(\max_{P_{t+1}} V_{t+1}^P ((1 + r)(A_t - C_t + W_t + Y_t), W_{t+1}, Z_{t+1}) \right) \right], \quad (8.4)$$

$$V_t^0 = \max_{C_t} \left[U(0, C_t, Z_t) + \kappa E_t \left(\max_{P_{t+1}} V_{t+1}^P ((1 + r)(A_t - C_t + Y_t), W_{t+1}, Z_{t+1}) \right) \right],$$

with the operators E_t designating the consumer's expectation about the variables W_{t+1} and Z_{t+1} conditional on information I_t at time t , which includes W_t and Z_t . The term κ is a discount rate. The first-order condition of (8.4) with respect to C_t yields the Euler condition (4.28), which continues to relate the marginal utilities of consumption in adjacent periods even in this model with participation.

Alternative useful expressions for V_t^1 and V_t^0 are

$$V_t^1 = \max_{C_t} [U(1, C_t, Z_t) + \kappa E_t (\text{Prob}(P_{t+1} = 1 | I_t) V_{t+1}^1 + \text{Prob}(P_{t+1} = 0 | I_t) V_{t+1}^0)], \quad (8.5)$$

$$V_t^0 = \max_{C_t} [U(0, C_t, Z_t) + \kappa E_t (\text{Prob}(P_{t+1} = 1 | I_t) V_{t+1}^1 + \text{Prob}(P_{t+1} = 0 | I_t) V_{t+1}^0)],$$

where, for instance, $\text{Prob}(P_{t+1} = 1 \mid I_t)$ designates the consumer's probability of making the decision $P_{t+1} = 1$ conditional on information I_t . The value function in the last period, τ , is

$$V_\tau^p = P_\tau V_\tau^1 + (1 - P_\tau)V_\tau^0 = V_\tau^p(A_\tau, W_\tau, Z_\tau), \tag{8.6}$$

where

$$V_\tau^1 = \max_{C_\tau} U(1, C_\tau, Z_\tau) \quad \text{s.t.} \quad C_\tau = A_\tau + Y_\tau, \tag{8.7}$$

$$V_\tau^0 = \max_{C_\tau} U(0, C_\tau, Z_\tau) \quad \text{s.t.} \quad C_\tau = A_\tau + Y_\tau,$$

Solving recursively using backward induction yields formulations for each period's value functions and optimal choices.

8.1.2. Empirical formulation

An empirical model characterizes how the values of P_1, P_2, \dots, P_τ vary across a population, relating these participation decisions to economic factors relevant in the past, now, or in the future. Creating the likelihood function for the P_i 's requires specifying the densities describing the joint distributions of the W_i 's and Z_i 's, and identifying the partitions of W_i 's and Z_i 's associated with making particular decisions.

Consider, first, decisions in the final period. Define the sets:

$$\begin{aligned} \Theta_{\tau 1} &= \{(W_\tau, Z_\tau) : V_\tau^1 > V_\tau^0\}, \\ \Theta_{\tau 0} &= \{(W_\tau, Z_\tau) : V_\tau^1 \leq V_\tau^0\}. \end{aligned} \tag{8.8}$$

For combinations of W_τ and Z_τ falling in the set $\Theta_{\tau 1}$, the individual chooses $P_\tau = 1$; and when $(W_\tau, Z_\tau) \in \Theta_{\tau 0}$ this person does not work in period T . The sets $\Theta_{\tau 1}$ and $\Theta_{\tau 0}$ are functions of all decisions and variables observed in previous periods.

Now considering period $\tau - 1$, define the sets:

$$\begin{aligned} \Theta_{(\tau-1) 1}^1 &= \{(W_{\tau-1}, Z_{\tau-1}) : V_{\tau-1}^1 > V_{\tau-1}^0\}, \\ \Theta_{(\tau-1) 1}^0 &= \{(W_{\tau-1}, Z_{\tau-1}) : V_{\tau-1}^1 \leq V_{\tau-1}^0\}, \end{aligned} \tag{8.9}$$

The individual works when $(W_{\tau-1}, Z_{\tau-1}) \in \Theta_{(\tau-1) 1}^1$, and does not work otherwise. Once again, the sets $\Theta_{(\tau-1) 1}^1$ and $\Theta_{(\tau-1) 0}^0$ depend on decisions and variables observed in periods $\tau - 2, \tau - 3, \dots, 1$.

Letting $g(\cdot)$ denote the joint density function of the W_i 's and Z_i 's, the probability of the event $(P_1, P_2, \dots, P_\tau)$ is

$$l_{P_1, P_2, \dots, P_\tau} = \int_{\theta_1, P_1} \dots \int_{\theta_{P_i}} \dots \int_{\theta_\tau, P_\tau} g(W_1, Z_1, \dots, W_\tau, Z_\tau) dW_\tau dZ_\tau \dots dW_1 dZ_1. \tag{8.10}$$

The density function $g(\cdot)$ can readily be made conditional on those observed Z_t that are exogenous or fixed and known. The joint density $g(\cdot)$ need not be the distribution that individuals use to account for the uncertainty they perceive about the future; $g(\cdot)$ describes the stochastic properties of the variables unobserved by the econometrician.

The sets Θ_{jp} are usually quite complicated to calculate. A popular simplifying assumption is to presume that individuals cannot save. In this case, $A_t = 0$ and $C_t = W_t + Y_t$ in Eqs. (8.5) and (8.7). With these assumptions we see that

$$V_t^p = U(P_t, P_t W_t + Y_t, Z_t) + \kappa E_t[\max V_{t+1}^p], \tag{8.11}$$

where the second term on the right-hand side of this expression does not depend at all on P_t . This formulation greatly simplifies computation of both the value functions and the sets Θ_{jp} . To simplify computation further, researchers also often assume that the variables W_t and Z_t are serially (and sometimes contemporaneously) independent.

8.1.3. Multiple values of hours

We can extend the above model beyond the simple decision to participate by admitting a limited set of hours choices. The approach shares many of the attributes of the computational-simplification procedure described in Section 6.7, with the complication that we must infer the value function appropriate for evaluating options. To illustrate this approach in a lifecycle context, suppose a worker may choose among full-time work, part-time work, and no work in each period, with each option implying a prescribed number of hours. This finite set of hours choices yields a relatively small set of discrete states, say J states in each period, over which the lifetime utility function must be maximized. Let P_{jt} designate a zero-one dummy variable equaling unity if an individual selects option j hours in period t , and let W_{jt} denote the earnings received from this option.

The value function now becomes

$$V_t = \sum_{j=1}^J P_{jt} V_t^j = \sum_{j=1}^J P_{jt} V_t^j(A_t, W_{jt}, Z_t), \tag{8.12}$$

where

$$V_t^j = \max_{C_t} \left[U(P_{jt}, C_t, Z_t) + \kappa E_t \left(\max_{P_{j(t+1)}} \sum_{j=1}^J P_{j(t+1)} V_{t+1}^j((1+r)(A_t - C_t + W_{jt} + Y_t), W_{j(t+1)}, Z_{t+1}) \right) \right].$$

One can express V_t^j in a way similar to Eq. (8.5) which assists in computing value functions in many instances. The value function in the last period, τ , is

$$V^\tau = \max_{C_\tau} U(P_{j\tau}, C_\tau, Z_\tau) \quad \text{s.t.} \quad C_\tau = A_\tau + W_{j\tau} + Y_\tau. \tag{8.13}$$

A backward recursive solution once again permits computation of each period's value functions and optimal choices.

Developing the likelihood function for the P_{jt} 's requires partitioning the sample space of W_{jt} 's and Z_t 's corresponding to the particular decisions. Within period t , the decisions P_{jt} are mutually-exclusive and exhaustive. For notational convenience, suppose W_t now denotes a vector including all of the W_{jt} 's as elements. Define the sets

$$\Theta_{ij} = \{(W_t, Z_t) : V_t^j > V_t^k \text{ for } k = 1, \dots, J, k \neq j\}. \quad (8.14)$$

When $(W_t, Z_t) \in \Theta_{ij}$, the individual chooses $P_{jt} = 1$. The set Θ_{ij} are functions of all decisions and variables observed in periods $t - 1, t - 2, \dots, 1$.

The likelihood function for this more general case is given by Eq. (8.10), with the sets Θ_{ij} now replacing the sets Θ_{iP_t} . With this modification, $l_{P_1 P_2 \dots P_T}$ represents the probability of observing the event $(P_{j1}, P_{j2}, \dots, P_{jT})$.

Allowing for continuous choices in a lifecycle model involves insurmountable computational burden when participation is an issue, unless one relies on very strong behavioral and stochastic assumptions. In effect, this amounts to expanding the set J to a large number of values. Even in the simple case considering only participation, the above discussion shows that the entire lifecycle problem must be solved to characterize decisions in any period. The two-stage budgeting and Euler-condition approaches utilized in Section 4 are of little use in simplifying the estimation problem. Other behavioral features of lifecycle models diminish the usefulness of these approaches as well by invalidating the separability properties needed by them, even when participation is not a source of violation.

8.2. Learning by doing and human capital

Saving and the accumulation of assets is just one way that past labor supply choices can affect today's decisions. In learning-by-doing models, past work experience has a direct effect on the determination of market wages. A similar mechanism operates in human capital models. Past labor market decisions have an impact not just through the level of accumulated assets but also through the wage. These considerations significantly change the nature of the optimal labor supply decisions. For example, learning by doing introduces a trade-off between the increase in utility that can be achieved by reducing current work effort and the increase in future productivity that can be achieved from learning on the job. This implies that the current wage is no longer the appropriate measure of the return to working. An additional "dynamic rent" term must be included to account for increased future wages resulting from the accumulation of experience capital while working. Hence, the methods of Section 6, which are designed to deal with non-linearities in current wages arising from tax and transfer policies, are not directly useful here.

These dynamic generalizations of the standard model also imply that individuals who would have otherwise chosen to leave work may now choose to stay in employment. This property is also exhibited in search models that allow state dependence through asymmetry in layoff and job arrival rates. In this situation individuals may choose to remain in

employment so as to enhance the probability of being in employment when future returns to employment are high. For example, mothers of young children may choose to stay in employment simply to exploit the higher probability of subsequently being in work when children reach school age.

In the following we separate our discussion of these models into models with participation and those with continuous hours choices.

8.2.1. Learning by doing

8.2.1.1. Learning by doing with participation The learning-by-doing model posits that wages grow with experience. Individuals in these models do not decide whether or not to engage in human capital investment, the simple state of being in employment generates returns in its own right. The wage, W_t , is now determined as a function of experience capital, K_t . Experience capital in turn depends positively on past participation through a dynamic equation of the form

$$K_{t+1} = G(K_t, P_t). \quad (8.15)$$

Wages depend positively on K_t according to the function

$$W_t = W_t(K_t, \eta_t), \quad (8.16)$$

where η_t represents the unobservable component of wages as in Eq. (6.34). This implies that work not only brings immediate returns, but also increases future wages by adding to experience. For simplicity, we assume that the only uncertainty in the model enters through the wage error η_t .

The value functions in the period- t participation decision, corresponding to Eq. (8.4) have the form $V_{t+1}^p(A_{t+1}, W_{t+1}, K_{t+1}^p, \eta_{t+1}, X_{t+1})$, where X_t represents the elements of Z_{t+1} that remain after removing the K_{t+1}^p and η_{t+1} variables; now X_t incorporates all non-wage variables relevant for lifecycle decision making that are not controlled by the decision maker. This is done to explicitly acknowledge the dependence of K_{t+1} on P_t , and also to separate out the source of uncertainty η_{t+1} .

The solution to the individual's participation problem follows closely that outlined in Section 8.1. In period t , individuals choose participation to maximize utility as described by Eq. (8.4) but acknowledging the impact of P_t on K_{t+1} in V_{t+1}^p . Since the only uncertainty enters through η , the participation decision defines a "reservation value" for the wage error η_t^* , which in turn defines the sets (8.8). This reservation value depends on the value of K_t and thus, to solve the problem a solution must be found for each of the τ possible values of accumulated work experience. The definition of η_t^* for all periods and all possible value of K_t captures all of the economics of the problem. In each period, the individual realizes a wage shock and makes a work decision to maximize utility given accumulated experience. Accumulated experience impacts the decision both by increasing wages and by impacting the disutility of work. The impact of current work decisions on future utility is accounted for by the EV_{t+1} terms – working today changes the value of η^*

tomorrow and, thus, impacts the probability of future work and expected future utility. Given the value of η^* for all periods and all possible values of K_t , estimation is straightforward.

Following Eq. (8.12) this model can be extended to allow for additional discrete states, for example, part-time and full-time participation. Particular functional forms for $G(\cdot)$ may be also chosen to allow for interactions between K_t , participation and hours of work. We return to a discussion of specific parameterizations in the review of empirical studies in Section 8.4.

8.2.1.2. Learning by doing with continuous hours choices Often in a learning-by-doing model, the level of hours of work, rather than participation alone, determines wage growth. To introduce learning by doing in a model with continuous hours choices, we abstract from the participation decision and replace Eq. (8.15) with

$$K_{t+1} = G(H_t, K_t), \quad (8.17)$$

where H_t is hours of work in period t , and G is an increasing function of H_t . Choices over hours and consumption are made by maximizing equation

$$V(A_t, W_t, K_t, \eta_t, X_t) = \max_{C_t, L_t} [U(C_t, L_t, X_t) + \kappa E_t(V(A_{t+1}, W_{t+1}, K_{t+1}, \eta_{t+1} X_{t+1}))]. \quad (8.18)$$

Notice that the value function in period t is made a function of the beginning of period t experience capital K_t as well as the financial capital A_t and the maximization takes place subject to the accumulation equations for experience capital and asset capital. The Euler equation for consumption continues to hold. However, the first order conditions for the allocation of time generalize to account for the role of experience capital. Assuming an interior solution for this continuous hours problem we have

$$U_L(C_t, L_t, X_t) = \lambda_t W_t + \kappa E_t\{\Gamma_{t+1}(\partial G/\partial H_t)\}, \quad (8.19)$$

$$\Gamma_t = \lambda_t(\partial W/\partial K_t)H_t + \kappa E_t\{\Gamma_{t+1}(\partial G/\partial K_t)\}, \quad (8.20)$$

where $\lambda_t = \partial V_t/\partial A_t = \partial U_t/\partial C_t$ and $\Gamma_t = \partial V_t/\partial K_t$.

The basic change from the standard hours of work model discussed in Section 4, is that the value of work is no longer simply the wage, but now includes the return to experience. This return depends on all future work decisions through the term Γ_{t+1} which measures the return to human capital. As such, standard hours of work equations of the sort we have been considering are inappropriate, in that they relate work to current wage which is no longer relevant on its own. All future wages and implied work decisions must also be included in determining the value of work.

8.2.2. Human capital

8.2.2.1. Human capital models with participation Consider an individual who, in each period, can now choose between participation in work, P_t , and participation in human

capital investment, P_t^* . The wage, W_t , is determined as a function of human capital, K_t , according to the function, $W_t = W_t(K_t, \eta_t)$ where K_t depends on past investment decisions. Suppose human capital accumulates according to the dynamic equation:

$$K_{t+1} = G(K_t, P_t^*). \quad (8.21)$$

The problem is, in principle, more complicated now because the individual must choose among three activities. However, this problem can be solved applying the multiple values of hours formulation outlined in Section 8.1.3. This is done simply by reinterpreting the discrete hours choices as options over the four states characterized by the four possible values combinations of P_t and P_t^* . There is the added complication that the wage is state dependent, but this is readily handled within Eqs. (8.12) and (8.13) given the definitions of the value functions V_t^j . Typically, the applications make the additional assumption of no savings to take attain computational simplifications. We discuss particular specifications in our review of the empirical applications at the end of this section.

8.2.2.2. Human capital models with continuous hours choices In the continuous hours-of-work problem individuals choose how much time in each period to spend in three activities: leisure L_t , hours of work H_t , and human capital investment S_t . Their choice problem is to choose L_t and C_t ,

$$V(A_t, W_t, K_t, \eta_t, X_t) = \max_{C_t, L_t} [U(C_t, L_t, X_t) + \kappa E_t V(A_{t+1}, W_{t+1}, K_{t+1}, \eta_{t+1}, X_{t+1})] \quad (8.22)$$

subject to the human capital equations $K_{t+1} = G(K_t, S_t)$, the asset accumulation conditions and $L_t + H_t + S_t = T$. This results in two additional conditions:

$$U_L(C_t, L_t, X_t) = \kappa E_t \{ \Gamma_{t+1} (\partial G / \partial S_t) \}, \quad (8.23)$$

$$\Gamma_t = \lambda_t H_t (\partial W_t / \partial K_t) + \kappa E_t \{ \Gamma_{t+1} (\partial G / \partial K_t) \}. \quad (8.24)$$

Given these expressions, the marginal utility of leisure still equals λ times the wage rate, the marginal rate of substitution between consumption and leisure still equals W , and the Euler equation for consumption continues to hold.

Since time must be allocated among 3 activities, this problem becomes more complicated. Eq. (8.23) indicates that the return to training must also equal the return to leisure which equals the return to work. The return to schooling depends on the marginal value of a unit of human capital, Γ , and Eq. (8.24) gives a Euler equation for its time path. So, levels of both leisure and training must be selected to equate their marginal values with λ times the wage – these two choices together imply the number of hours worked. However, if hours of work can be measured separately from hours of training then the labor supply equation can be estimated directly since the standard marginal conditions for the choice of working hours remain valid.

8.3. Habit persistence

Habit persistence nullifies the intertemporal separability property for preferences through the dependence of current utility on past labor supply and consumption choices. In the framework introduced in Section 8.1, we can think of these past choices entering Z_t . For example, period- t utility may be written as $U_t(P_t, C_t, P_{t-1}, C_{t-1}, X_t)$. In this formulation we have divided Z_t into one set of elements controlled by the individual's previous behavior, namely P_{t-1} and C_{t-1} , and a second set designated as X_t that are not influenced by the decision maker. Typically this is set up as a household production model in which past non-market (leisure) time and past consumption influence today's utility. Consequently, one may wish to add further lags of participation and consumption. Our review of empirical applications considers such specifications.

The problem for the consumer is analogous to that described in Eq. (8.4), but now recognizing that Z_{t+1} is a function of current consumption and current participation. The backward recursion follows the same form as Eqs. (8.6) and (8.7). With the wage innovation η_t being the only source of uncertainty, the estimation is the same as in the learning-by-doing model.

This analogy with the learning-by-doing model also holds in the continuous hours choice framework without participation. In this case $U_t(C_t, L_t)$ is replaced by $U_t(C_t, L_t, C_{t-1}, L_{t-1}, X_t)$. Further lags may be included without changing the basic intuition underlying this model. The optimization problem in that case becomes

$$V(A_t, C_{t-1}, L_{t-1}, X_t) = \max_{C_t, L_t} [U(C_t, L_t, C_{t-1}, L_{t-1}, X_t) + \kappa E_t V(A_{t+1}, C_t, L_t, X_{t+1})]. \quad (8.25)$$

The first-order conditions for an interior solution for leisure becomes

$$U_L(C_t, L_t, C_{t-1}, L_{t-1}, X_t) + \kappa E_t \{U_L(C_{t+1}, L_{t+1}, C_t, L_t, X_{t+1})\} \\ = \kappa E_t \{U_C(C_{t+1}, L_{t+1}, C_t, L_t, X_{t+1})(1+r)W_t\}. \quad (8.26)$$

A similar relation exists for consumption. As in the learning-by-doing model the value of work is no longer simply the wage. Now it includes the dynamic rent in terms of the impact on future marginal utility.

8.4. Review of empirical results

8.4.1. The basic intertemporal labor supply model

There are many applications of the basic intertemporal labor supply model. These are generally extensions of the Heckman and MaCurdy (1980) and MaCurdy (1981) studies.⁶⁰ For example, Browning et al. (1985) work directly with Frisch labor supply equations (see Section 4.4.3) and use a Pseudo cohort approach on the time series of repeated cross sections on consumption and family labor supply available in the British Family Expen-

⁶⁰ See also Altonji (1982, 1986).

diture Survey. They do not allow for non-participation. Blundell et al. (1993) incorporate corner solutions in their study of intertemporal hours of work decisions among married women in the UK. They work directly with the marginal conditions (4.6) and (4.20). The within-period consumption-leisure choices are modeled using an Almost Ideal form for preferences. The Euler equation is then used to identify a Box-Cox monotonic transformation of within-period utilities (as also adopted in MaCurdy (1983)). Their results point to intertemporal (Frisch) labor supply elasticities for married women in the 0.5–1 range depending on demographic characteristics – women with younger children having the bigger elasticities. As expected, estimated Marshallian elasticities are quite a bit smaller, in the 0.2–0.5 range. The intertemporal elasticity of substitution for consumption is approximately 0.6 which suggests a moderate degree of risk aversion.

8.4.2. Learning-by-doing models

Shaw (1989) estimates a learning-by-doing model in her study of the continuous hours choices of a similar sample of men from the PSID. She selects 526 men in the 18–64 year age range during the period 1967–1980. As in the Hotz et al. (1988) study, a Translog direct utility is chosen but this is specified in terms of *current* non-market time and consumption. There are no habit terms. However, in contrast to that earlier study, the stock of experience enters the wage equation. A quadratic specification is used for the capital accumulation function (8.17) to reflect the possibly concave nature of the lifecycle earnings profile. This is then used to define an estimable dynamic wage equation by assuming $W_t = \rho_t K_t$ where ρ_t is the rental rate of experience capital. This rental rate is assumed constant across individuals in any particular year. Shaw again finds strong evidence of non-separability – this time entering through the wage experience relationship rather than through the utility function. She finds a large positive effect which implies that a temporary 25% increase in hours of work increases wages by 12.8% starting from the initial mean values. The Shaw study is restricted to men and does not consider the problem of non-participation.

This is tackled in the Eckstein and Wolpin (1989) study which estimates a discrete model using a sample of 318 women from the NLS of mature women survey. They specify within-period utility to have the form (8.27). To simplify the problem, they assume that there is no saving or borrowing, so the within-period budget constraint reduces to

$$C_t = W_t P_t + Y_t. \quad (8.28)$$

Wages are assumed to be log linear in schooling, experience capital and the unobservable η_t , with $\eta_t \sim \text{iid}N(0, \sigma_\eta^2)$. Under these assumptions, the sample likelihood is given by⁶¹

⁶¹ Obviously, the reservation wage cannot be bigger than the smallest wage observed for each individual of a particular type in the sample. Eckstein and Wolpin (1989) allow for measurement error in wages to avoid this restriction.

$$\prod_{i=1}^N \prod_{t=1}^{T_i} [\Phi(\eta_i^*/\sigma_\eta)]^{1-P_t} \left[\frac{1}{\sigma_h} \varphi(\eta_i/\sigma_\eta) \right]^{P_t}, \quad (8.29)$$

where η^* is derived from the structural utility maximization framework outlined in previous sections. Hence, the within-period problem is a standard Tobit formulation. Because the errors are serially uncorrelated, these within-period Tobit likelihood functions are simply multiplied together to yield the overall likelihood function.

The sample of women used in the Eckstein and Wolpin (1989) study were aged between 39 and 44 in 1967 and have at least four consecutive years of data on labor force participation beginning in 1966. The basic findings of the model are best summarized by the simulations the authors conduct, manipulating the value of each variable in the model and observing predicted work effort. First, they find that at any age, the probability of work increases with experience. Hence the positive impact of experience on wages overcomes the fact that the disutility of work increases with experience. Second, for any experience level, work effort decreases with age – as age advances there are less future gains available from increasing experience and, thus, the value of work declines. This explanation for declining work with age is missed by any static model. Third, work effort decreases with husband's earnings and increases with schooling. Finally, increasing the slope of the wage/experience profile substantially increases work effort over the lifetime. Again, this effect would be missed by any static labor supply model.

At this point, it is important to reiterate the extreme simplifying assumptions that have been made to make the problem manageable. First, a 0/1 work decision has been assumed. Second, individuals cannot save or borrow. These two assumptions together reduce the choice problem to a simple work, no-work decision, and limit the dynamic elements of the problem to the accumulation of human capital. Third, no unobserved heterogeneity is admitted in the utility function. The only error term in the model is the wage error, η , which is assumed serially uncorrelated and normally distributed. As we have seen, this reduces the dynamic problem to a series of standard Tobit problems and eliminates any concerns about initial conditions.

Altug and Miller (1990) combine certain aspects of both of these approaches in their study of labor supply and consumption. They use the Euler equation for consumption and the continuous hours information to recover some of the preference parameters. Utility is assumed explicitly additive in consumption and leisure but current-period utility is allowed to depend on past labor supply choices. Wages also have a multiplicative form in aggregate shocks, individual heterogeneity and a term capturing the effect of past labor supply choices. A log-differenced wage equation can, therefore, be estimated across individuals without adjustments for selection. To identify their model they are obliged to make certain additional assumptions on unobserved heterogeneity. First, they assume that, conditional on participation, there is no unobserved heterogeneity in hours of work. Second, they assume Pareto efficient allocations across all individuals in the economy. This latter assumption implies that the marginal utility of consumption is simply the

product of an individual and a time effect. A fixed cost of work parameter is introduced and recovered directly from the value function comparison. The forward looking terms in this comparison are simplified using the idea of Hotz and Miller (1993) which assumes sufficient stationarity to replace future value comparisons with current observed transition rates.

Estimation takes place using a sample of 2169 women from the PSID for 1973–1985. Consumption is restricted to food consumption. They find an important effect of past labor supply on wages. They also report important non-separabilities over time in utility. Current and past labor supplies are found to be substitutes.

8.4.3. Some extensions

This dynamic model has been extended in a number of papers to include endogenous fertility and marital decisions. For example, drawing on the earlier work of Heckman and Willis (1975) and Moffitt (1983), Hotz et al. (1988) develop a semi-reduced form representation of fertility and labor supply decision rules. Francesconi (1995) places this model in the Eckstein and Wolpin framework which is extended to allow endogenous fertility. Van Der Klaauw (1996) also presents an extension of this framework to allow for endogenous marital decisions, although he maintains the exogeneity of fertility.

Separability in the decision rule can also be relaxed through the introduction of asymmetric job layoff and arrival rates. This is the model presented in Blundell et al. (1997, 1998c) who developed earlier work on discouraged workers by Blundell et al. (1987) to allow for active search, layoffs and saving in a model of labor market transitions. Estimation is shown to be possible without recourse to the full dynamic programming solution using the information in the consumption Euler equation, labor market transition rates and the consumption policy function. However, strong restrictions are placed on the distribution of unobservable preference heterogeneity and on the distribution of wages.

8.4.4. Habit persistence models

The habit persistence model as discussed in Section 8.3 was investigated extensively in Hotz et al. (1988) although, as in the Shaw study, they do not consider non-participation. Their study further assumes that within-period utility over C_t and K_t in Eq. (8.26) is described by a Translog direct utility and they do not allow for learning by doing. Habits enter utility in the form

$$K_t = L_t + \alpha \Psi_t, \quad (8.30)$$

where Ψ_t is the habit stock of leisure

$$\Psi_t = (1 - \theta) \Psi_{t-1} + L_{t-1}. \quad (8.31)$$

The parameter α represents the substitution between current “leisure” and past leisure capital in the production of K_t . Notice that when the depreciation parameter, θ , in the definition of Ψ_t is unity then it is only last period’s leisure (or labor supply) that matters for

today's marginal utility of income.⁶² Allowing θ to be less than unity generalizes the first-order conditions slightly since now all future utilities depend on L_t through the stock term Ψ_t .

This specification results in two stochastic dynamic estimating equations which are estimated by generalized method of moments. Their application is to the hours and consumption choices of working men from the PSID panel for the period 1967–1978 (specifically 482 white household heads aged between 23 and 52). These two groups are subsequently split into a younger and older group. Although there is some evidence of misspecification in the consumption Euler equation, there is reasonably strong evidence of non-separable preferences and the parameters α and $1-\theta$ turn out to be precisely estimated at around 0.6 and 0.65, respectively, for the group of younger males who were aged 23–36 in 1967. For the sample of older men the α parameter is somewhat higher and the $1-\theta$ parameter slightly lower.

9. Closing comments

The aim of this chapter has been to critically review existing approaches to modeling labor supply and to identify important gaps in the literature that could be addressed in future research. We began with a look at the kind of policy reform proposals that labor supply models are now required to address and the set of labor market facts that labor supply models are designed to interpret. In the sections that followed, we developed a unifying framework and provided a brief assessment of each modeling approach, reviewing relevant empirical studies at the end of each section. In this concluding section, we ask: Have the recent advances in labor supply research, reviewed in this chapter, placed us in a better position to answer the policy reform questions raised in Section 2 and enabled us to provide a more reliable interpretation of the trends in participation and hours described in Section 3?

It is certainly true that this chapter has documented some significant advances in labor supply research since the original Handbook chapters on labor supply were written in the first half of the 1980s. Even relative to the important appraisal of the area by Heckman (1993), the marked changes in tax and welfare policies highlighted in Section 2 have forced labor supply research to increasingly acknowledge the importance of the extensive margin and discreteness in observed behavior. Likewise, the renewed focus on human capital in the policy debate has created the need for new generalizations in intertemporal models. We have also noted the innovations in our understanding of interactions between individuals within households concerning their labor supply decisions, brought about by the collective approach to family labor supply.

However, we have also identified some significant gaps in our knowledge which make it difficult to assert confidently that we are in a position to examine reliably many of the

⁶² See, for example, the Johnson and Pencavel (1984) specification.

important current policy reform proposals or to assess accurately the main determinants of participation and hours-of-work changes. This, in turn, explains why labor supply remains an active and productive area for research. What are these gaps in our current knowledge? Perhaps the overriding difficulty remains with modeling participation. This is key in any analysis of welfare reform. Even in the simplest dynamic model without fixed costs, we have seen that the reservation wage depends on the whole future of wages and other unobservables. Estimation of behavioral parameters and simulation of policy reforms is, therefore, considerably complicated.

Some studies have attempted to restrict the margins for intertemporal decisions so as to focus on the discrete participation decision. Although these studies have provided important insights into modeling techniques and enhanced our understanding of behavior, it is difficult to believe that they provide sufficiently good approximations to actual behavior to give robust policy guidance. For example, we have argued that saving decisions should be modeled alongside labor supply decisions. Studies that model saving and labor supply allowing for discrete behavior are few and far between and no robust view of how these interactions work is currently available.

Some analysts have been content to measure the overall impact of past policy reforms on either participation or hours of work using a difference-in-differences or natural experiment approach. However, even where the stringent assumptions required for consistent estimation of interpretable parameters are satisfied, the estimated parameters do not provide sufficient information for extrapolation or simulation. We have argued that simulation of tax and welfare proposals cannot be completed without a structural model. Here a gap in the literature is revealed. Structural models allowing for discrete choices over labor-force and welfare participation that acknowledge dynamic decisionmaking are still not available in the empirical literature. A central part of this survey has been to assemble the building blocks necessary for such an analysis.

In a similar spirit, developments of the family labor supply model that allow for collective behavior must also be placed in an intertemporal context. Much evidence suggests that the strong pooling assumptions underlying the traditional family labor supply model are untenable, which is worrying for any analysis of the impact of welfare reform on family labor supply. However, only very recently have the simplest collective labor supply models been extended to allow for discrete choice and unobserved heterogeneity, both necessary ingredients of any empirical study. Moreover, allowing for the possibility of household production in these models requires more detailed data on time use.

Structural models that allow for the interactions between family members and the non-convexities in the incentive structure facing individual workers typically place strong requirements on the individual's and the economist's knowledge of budget constraints and the distribution of unobservables. We have seen that it is often the desire for flexibility along these dimensions that motivates empirical studies that adopt the difference-in-differences approach. Even simple structural models often do not account for correlation between unobservable individual effects in labor supply and the wage and income vari-

ables. Additionally, they do not allow for mismeasurement of the budget constraint or the wage and income variable themselves. Since structural models are required for many purposes for which labor supply analysis is undertaken, precisely how much these measurement issues matter for different datasets and different modeling approaches should remain an active area for research.

We have devoted much attention to the specification of labor supply models that account for non-convexities in the budget constraint, induced by high welfare withdrawal rates and fixed costs of work. This is no coincidence; the evaluation of the labor supply responses to welfare policy reforms remains the most significant recent contribution of standard labor supply models. We have already pointed out the need for further research that places this analysis in a dynamic setting. We have also noted the importance of research designed to assess the robustness of alternative approximations to the shape of the budget constraint and the packaging of hours choices into discrete bundles.

There remain a number of big issues that we have not touched on in this chapter but that are important for labor supply analysis. Many of these issues are discussed elsewhere in this Handbook. Among the most important is the modeling of the retirement decision. In a general sense, this is implicitly covered in our discussion of participation, but to properly understand the retirement decision requires careful treatment of the specific institutional structure of retirement programs and the way in which they interact with disability schemes and rules for earning after retirement (see the forthcoming volume by Gruber and Wise (1998) for a useful selection of country specific studies of the retirement behavior and the structure of social security systems). Another issue relates to the process of job search and job matching.

We should also acknowledge the potential importance of general equilibrium effects from tax and transfer programs. These make it even more difficult to think of groups of individuals wholly unaffected by reforms, as is required in the difference-in-differences approach, and imply different welfare calculations from those from models that assume gross wages and prices are unaffected by transfer and tax reforms.

Finally, we reiterate the main theme of our review: to formalize the assumptions that are required for interpretation of elasticities recovered from alternative modeling approaches and data sources. We hope that this has satisfied the twin goals of making clear precisely what is being estimated in any specific study and making it possible to compare estimates across studies.

Appendix A. Specifications of within-period preferences

This appendix briefly reviews some popular within-period (or contemporaneous or static) labor supply specifications.⁶³ Specification (4.30), used to illustrate our discussion in Section 4, corresponds to a within-period labor supply model of the form⁶⁴

$$\ln H = \alpha \ln W + \theta Y + \rho. \quad (\text{A.1})$$

Here we suppress the t subscript and allow the single quantity, ρ , to represent observed and unobserved heterogeneity. Specification (A.1) is one of a number of popular alternative three-parameter specifications that allow a single parameter for each of the wage, income effects and heterogeneity terms. Such models place strong restrictions on preferences and modern research on consumer behavior strives to relax these restrictions using more flexible representations.

One important restriction on preferences in within-period labor supply models is on the sign of the wage response. In theory there is no requirement for the wage response to be the same sign over all hours choices and, although it is required to be positive at the participation margin where the income effect is zero, it can become negative as hours increase. The precise shape of the hours – wage relationship is also likely to vary with income and demographic composition. A second restriction is on the income response, which determines the extent to which leisure is a normal good and whether it is a luxury or necessity. Most evidence from consumer behavior suggests that this varies widely across different goods and different types of consumers. Models that are linear in income (quasi-homothetic preferences) as in (4.30'), or that imply constant elasticities as in Eq. (4.30) are typically rejected.⁶⁵

Restrictions on within-period preferences are usefully summarized by the specification of the indirect or direct utility function. The additivity between wage and income, implicit in (A.1), and the constancy of the wage elasticity for all hours choices, are reflected in the following additive exponential form of the indirect utility function:

$$v(W, Y) = \frac{W^{\alpha+1}}{\alpha + 1} - \frac{e^{-\theta Y}}{\theta e^{-\rho}}. \quad (\text{A.2})$$

Many alternative three-parameter specifications of this kind are popular in empirical applications. The relationship among these specifications and the preference restrictions they imply are helpful in comparing studies. Here we list a number of them and provide a brief commentary.

Linear labor supply:

$$H = \alpha W + \theta Y + \rho \quad (\text{A.3})$$

$$v(W, Y) = \exp(\theta W) \left(Y + \frac{\alpha}{\theta} W - \frac{\alpha}{\theta^2} + \frac{\rho}{\theta} \right). \quad (\text{A.4})$$

⁶³ See Stern (1986) for a comprehensive review of these and more non-linear parametric static labor supply specifications and their implied indirect and direct utility functions.

⁶⁴ To provide a within-period interpretation of these preferences in a two-stage budgeting context, we would replace Y by the consumption-based measure, Y^C .

⁶⁵ The assumption of quasi-homothetic preferences provides a very poor approximation in empirical work on consumer behavior. More data-coherent specifications require terms not only in M_t , but also in $M_t \ln(M_t)$ and even higher-order interactions.

Although popular, the linear model imposes the same sign on the wage response throughout and implies quasi-homothetic preferences.

Semi-log labor supply:

$$H = \alpha \ln W + \theta Y + \rho \quad (\text{A.5})$$

$$v(W, Y; X) = \frac{\exp(\theta W)}{\theta} (\theta Y + \rho + \alpha \log W) - \frac{\alpha}{\theta} \int_{\theta W} \frac{\exp(\theta W)}{\theta W} d(\theta W). \quad (\text{A.6})$$

The semi-log model allows some non-linear curvature in wage effects so that the wage elasticity declines with hours but its sign is positive throughout and it is still linear in income. This formulation is attractive where non-participation is an issue and where there may be measurement error or endogeneity in wages and income. The log linearity in wage allows proportional taxes to enter linearly and is also a popular specification for reduced forms for gross hourly wages.

Semi-log labor supply (generalization 1):

$$H = \alpha \ln W + \theta Y^* + \rho \quad (\text{A.7})$$

with $Y^* = WH + Y - \alpha W(1 - \exp(-H/\alpha))$. There is no easy form for the indirect utility with this generalization of the semi-log model but it is interesting for a number of reasons. First, it can be rewritten as a specification for the log marginal rate of substitution function which is linear in H and Y^* . Therefore, it produces a particularly simple form for the reservation wage. Second, it permits negative wage responses as hours increase. As H tends to zero, it approaches the standard semi-log model (A.6).

Semi-log labor supply (generalization 2):

$$H = \alpha \ln W + \theta Y/W + \rho \quad (\text{A.8})$$

$$v(W, Y) = \frac{W^{\alpha+1}}{\alpha+1} \left(\frac{Y}{W} (1+\theta)^2 + \alpha \ln W + \rho - \frac{\alpha}{(1+\theta)} \right). \quad (\text{A.9})$$

This generalization has the attraction of allowing a change in sign for the wage elasticity as Y is reduced, which would typically correspond to an increase in labor supply. It also facilitates the introduction of higher-order terms in $\ln W$. However, this specification retains the assumption of linearity in Y and introduces an awkward non-linearity in W .

Stone–Geary (LES) labor supply:

The direct utility is probably the most familiar characterization:

$$u(H, C) = [\theta \ln(\gamma_H - H) + (1 - \theta) \ln(C - \gamma_C)] \quad (\text{A.10})$$

with labor supply

$$WH = (1 - \theta) \gamma_H W - \theta Y + \theta \gamma_C. \quad (\text{A.11})$$

The Stone–Geary specification, although popular in early work on household behavior, has been used less frequently in recent years. It can allow negative wage responses but it

corresponds to a direct utility that is explicitly additive in hours and consumption. That is, the log marginal rate of substitution is additive in consumption and hours. It is also quasi-homothetic. Notice, however, that it is equivalent to the second generalization of the semilog model (A.8) with $\ln W$ replaced by $1/W$.

CES labor supply:

This is a useful generalization of the LES labor supply and corresponds to choosing a direct utility of the form

$$u(H, C) = [\theta(\gamma_H - H)^{-\mu} + (1 - \theta)(C - \gamma_C)^{-\mu}]^{-1/\mu}. \quad (\text{A.12})$$

It also implies an additive log marginal rate of substitution function and, therefore, explicit additivity between consumption and labor supply. However, it generalizes the substitution patterns between consumption and hours, and allows negative wage responses.

References

- Aaberge, R., J. Dagsvik and S. Strom (1995), "Labour supply responses and welfare effects of tax reforms", *Scandinavian Journal of Economics* 97: 635–659.
- Abbott, M. and O. Ashenfelter (1976), "Labor supply, commodity demand and the allocation of time", *Review of Economic Studies* 42: 389–411.
- Ackum-Agell, S. and C. Meghir (1995), "Male labour supply in sweden: are incentive effects important?" Tax reform evaluation report no.12 (National Institute of Economic Research).
- Agell, J., P. Englund and J. Sodersten (1996), "Tax reform of the century – the Swedish experiment", *National Tax Journal*, in press.
- Altonji, J.G. (1982), "The intertemporal substitution model of labour market fluctuations: an empirical analysis", *Review of Economic Studies* 49: 783–824.
- Altonji, J.G. (1986), "Intertemporal substitution in labor supply: evidence from micro data", *Journal of Political Economy* 94 (part II): S176–S215.
- Altug, S. and R. Miller (1990), "Household choices in equilibrium", *Econometrica* 58: 543–570.
- Amemiya, T. and T. MaCurdy (1986), "Instrumental variable estimation of an error components model", *Econometrica* 54: 869–881.
- Angrist, J. (1991), "Grouped data estimation and testing in simple labor supply models", *Journal of Econometrics* 47: 243–265.
- Apps, Patricia, F. and Ray Rees (1988), "Taxation and the household", *Journal of Public Economics* 35: 355–369.
- Apps, Patricia F. and Ray Rees (1993), "Labor supply, household production and intra-family welfare distribution", Discussion paper (Australian National University).
- Apps, Patricia F. and Ray Rees (1997), "Collective labor supply and household production", *Journal of Political Economy* 105: 178–190.
- Arellano, M. and C. Meghir (1992), "Female labour supply and on-the-job search: an empirical model estimated using complementary data sets", *Review of Economic Studies* 59: 537–559.
- Arrufat, J.L. and A. Zabalza (1986), "Female labour supply with taxation, random preferences, and optimization errors", *Econometrica* 54: 47–63.
- Ashenfelter, O. (1983), "Determining participation in income-tested social programs", *Journal of the American Statistical Society* 78: 517–525.
- Ashenfelter, O. and J. Ham (1979), "Education, unemployment and earnings", *Journal of Political Economy* 87: S99–S166.

- Ashenfelter, O. and J.J. Heckman (1974), "The estimation of income and substitution effects in a model of family labor supply", *Econometrica* 42: 73–85.
- Attanasio, O.P. and T. MaCurdy (1997), "Interactions in family labor supply and their implications for the impact of EITC", Mimeo. (Stanford University, Stanford, CA).
- Balestra, P. and M. Nerlove (1966), "Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas", *Econometrica* 34: 585–612.
- Bingley, P. and I. Walker (1997), "The labor supply, unemployment and participation of lone mothers in in-work transfer programmes", *Economic Journal* 107: 1375–1390.
- Bingley, P., G. Lanot, E. Symons and I. Walker (1995), "Child support reform and the labor supply of lone mothers in the UK", *Journal of Human Resources* 30: 256–279.
- Blau, D. and P. Robins (1988), "Child-care costs and family labor supply", *Review of Economics and Statistics* 70: 374–381.
- Blomquist, N.S. (1983), "The effect of income taxation on the labour supply of married men in Sweden", *Journal of Public Economics* 22: 169–197.
- Blomquist, N.S. (1985), "Labour supply in a two-period model: the effect of a non-linear progressive income tax", *Review of Economic Studies* 52: 515–529.
- Blomquist, S. (1996), "Estimation methods for male labour supply functions: how to take account of nonlinear taxes", *Journal of Econometrics* 70: 383–405.
- Blomquist, S. and U. Hansson-Brusewitz (1990), "The effect of taxes on male and female labour supply in Sweden", *Journal of Human Resources* 25: 317–357.
- Blomquist, S. and W. Newey (1997), "Nonparametric estimation of labor supply functions generated by piece wise linear budget constraints", Working paper no. 1997:24 (Department of Economics, Uppsala University, Sweden).
- Blundell, R.W. (1986), "Econometric approaches to the specification of life-cycle labour supply and commodity demand behaviour", *Econometric Reviews* 6(1): 147–151.
- Blundell, R.W. and P. Johnson (1998), "Pensions and labor market participation in the UK", *American Economic Review (Papers and Proceedings)* 88: 168–172.
- Blundell, R.W. and C. Meghir (1986), "Selection criteria for a microeconomic model of labour supply", *Journal of Applied Econometrics* 1: 55–81.
- Blundell, R.W. and C. Meghir (1987), "Bivariate alternatives to the tobit model", *Journal of Econometrics* 34: 179–200.
- Blundell, R.W. and I. Walker (1982), "Modelling the joint determination of household labour supplies and commodity demands", *Economic Journal* 92: 58–74.
- Blundell, R.W. and I. Walker (1986), "A life cycle consistent empirical model of labour supply using cross section data", *Review of Economic Studies* 53: 539–558.
- Blundell, R.W., C. Meghir, E. Symons and I. Walker (1986), "A labour supply model for the simulation of tax and benefit reforms", in: R.W. Blundell and I. Walker, eds., *Unemployment, search and labour supply* (Cambridge University Press, Cambridge, UK).
- Blundell, R.W., J. Ham and C. Meghir (1987), "Unemployment and female labour supply", *Economic Journal* 97: 44–64.
- Blundell, R.W., C. Meghir, E. Symons and I. Walker (1988), "Labour supply specification and the evaluation of tax reforms", *Journal of Public Economics* 36: 23–52.
- Blundell, R.W., V. Fry and C. Meghir (1990), "Preference restrictions in microeconomic models of life-cycle behaviour under uncertainty", in: J.P. Florens, ed., *Microeconometrics: surveys and applications* (Blackwell, Oxford, UK).
- Blundell, R.W., A. Duncan and C. Meghir (1992), "Taxation and empirical labour supply models: lone parents in the UK", *Economic Journal* 102: 265–278.
- Blundell, R., C. Meghir and P. Neves (1993), "Labour supply: an intertemporal substitution", *Journal of Econometrics* 59: 137–160.

- Blundell, R.W., M. Browning and C. Meghir (1994), "Consumer demand and the life-cycle allocation of household expenditure", *Review of Economic Studies* 161: 57–80.
- Blundell, R.W., T. Magnac and C. Meghir (1997), "Savings and labor market transitions", *Journal of Business and Economic Statistics* LXXIX: 527–539.
- Blundell, R.W., P.-A. Chiappori, T. Magnac and C. Meghir (1998a), "Collective labor supply and participation", Working paper W98/20 (Institute for Fiscal Studies).
- Blundell, R.W., A. Duncan and C. Meghir (1998b), "Estimating labour supply responses using tax policy reforms", *Econometrica* 66: 827–861.
- Blundell, R.W., J. Ham and C. Meghir (1998c), "Unemployment, discouraged workers and female labour supply", *Research in Economics* 52: 103–131.
- Boskin, Michael J. and Eytan Sheshinski (1983), "Optimal tax treatment of the family", *Journal of Public Economics* 20: 281–297.
- Bourgiugnon, F. and T. Magnac (1990), "Labour supply and taxation in France", *Journal of Human Resources* 25: 358–389.
- Bourgiugnon, F., M. Browning, P.-A. Chiappori and V. Lechene (1994), "Incomes and outcomes: a structural model and some evidence from French data", *Journal of Political Economy* 102: 1067–1096.
- Borsh-Supan, A. and R. Schnabel (1998), "Social security and declining labor force participation in Germany", *American Economic Review* 88: 173–178.
- Bover, O. (1989), "Estimating intertemporal labour supply elasticities using structural models", *Economic Journal* 99: 1026–1039.
- Browning, M. and C. Meghir (1991), "Testing for separability between goods and leisure using conditional demand systems", *Econometrica* 59: 925–952.
- Browning, M., A. Deaton and M. Irish (1985), "A profitable approach to labor supply and commodity demand over the life cycle", *Econometrica* 53: 503–543.
- Burtless, G. and J. Hausman (1978), "The effect of taxes on labour supply", *Journal of Political Economy* 86: 1103–1130.
- Card, D. (1994), "Intertemporal labour supply: an assessment", in: C. Sims, ed., *Advances in econometrics* (Cambridge University Press, Cambridge, UK).
- Card, D. and P.K. Robins (1996), "Do financial incentives encourage welfare participants to work? Initial 18-month findings from the self-sufficiency project", (Social Research Corporation, Vancouver, BC, Canada).
- Chiappori, Pierre-Andre (1988), "Rational household labor supply", *Econometrica* 56: 63–90.
- Chiappori, Pierre-Andre (1992), "Collective labor supply", *Journal of Public Economics* 437–467.
- Chiappori, Pierre-Andre (1997), "Introducing household production in collective models of labor supply", *Journal of Political Economy* 105: 191–209.
- Cogan, J. (1980), "Labor supply with costs of labor market entry", in: James Smith, ed., *Female labor supply: theory and estimation* (Princeton University Press, Princeton, NJ) pp. 327–364.
- Cogan, J.F. (1981), "Fixed costs and labor supply", *Econometrica* 49: 945–964.
- Colombino, U. and D. Del Boca (1990), "The effect of taxes and labour supply in Italy", *Journal of Human Resources* 25: 390–414.
- Dustmann, C. and A. Van Soest (1997), "Wage structures in the private and public sectors in West Germany" *Fiscal Studies*, in press.
- Eckstein, Z. and K. Wolpin (1989), "Dynamic labour force participation of married women and endogenous work experience", *Review of Economic Studies* 56: 375–390.
- Eissa, N. (1995a), "Taxation and labor supply of married women: the Tax Reform Act of 1986 as a natural experiment", Working paper no. 5023 (NBER, Cambridge, MA).
- Eissa, N. (1995b), "Labor supply and the Economic Recovery Tax Act of 1981", in: M. Feldstien and J. Poterba, eds., *Empirical foundations of household taxation*, conference volume (NBER, Cambridge, MA).
- Eissa, N. (1996), "Tax reforms and labor supply", in: James Poterba, ed., *Tax policy and the economy*, Vol. 10 (NBER, Cambridge, MA).

- Eissa, N. and J. Liebman (1995), "Labor supply responses to the earned income tax credit", Working paper no. 5158 (National Bureau of Economic Research, Cambridge, MA).
- Feldstien, M. (1995), "The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act", *Journal of Political Economy* 103: 551–572.
- Flood, L.R. and T. MaCurdy (1992), "Work disincentive effects of taxes: an empirical analysis of Swedish men", *Carnegie-Rochester Conference Series on Public Policy* 37: 239–278.
- Flood, L.R., A. Klevmarcken et al. (1997), in: P. Olovsson, ed., *Household market and nonmarket activities (HUS)*, Vols. 3–6 (Department of Economics, Uppsala University, Uppsala, Sweden).
- Fortin, Bernard and Guy Lacroix (1997), "A test of neoclassical and collective models of household labor supply", *Economic Journal* 107: 933–955.
- Fraker, T. and R. Moffitt (1988), "The effect of food stamps on labor supply: a bivariate selection model", *Journal of Public Economics* 35: 25–56.
- Fraker, T., R. Moffitt and D. Wolf (1985), "Effective tax rates and guarantees in the AFDC program, 1967–1982", *Journal of Human Resources* 20: 251–263.
- Francesconi, M. (1998), "A joint dynamic model of fertility and work of married women", Working paper of the ESRC Research Centre on Micro-Social Change, 98-2 (University of Essex, Colchester).
- Friedberg, L. (1995), "The labor supply effects of the social security earnings test", Mimeo. (MIT, Boston, MA).
- Gorman, W.M. (1959), "Separable utility and aggregation", *Econometrica* 21: 63–80.
- Gorman, W.M. (1968), "The structure of utility functions", *Revue of Economic Studies* 32: 369–390.
- Graversen, E.K. (1996), "Measuring labour supply responses to tax changes by use of exogenous tax reforms", Working paper no. 96-17 (Centre for Labour Market and Social Research, University of Aarhus, Aarhus, The Netherlands).
- Gruber, J. and J. Wise (1998), "Social security and retirement: an international comparison", *American Economic Review* 88: 158–163.
- Hall, R. (1973), "Wages, income and hours of work in the US labor force", in: G. Cain and H. Watts, eds., *Income maintenance and labor supply* (Chicago University Press, Chicago, IL).
- Ham, J. (1986a), "Testing whether unemployment represents life-cycle labor supply behaviour", *Review of Economic Studies* LIII: 559–578.
- Ham, J. (1986b), "On the interpretation of unemployment in empirical labour supply analysis", in: R.W. Blundell and I. Walker, eds., *Unemployment, search and labour supply* (Cambridge University Press, Cambridge, UK) pp. 121–142.
- Hanoch, G. (1980), "A multivariate model of labor supply: methodology and estimation", in: James Smith, ed., *Female labor supply: theory and estimation* (Princeton University Press, Princeton, NJ).
- Hausman, J. (1980), "The effect of wages, taxes and fixed costs on women's labor force participation", *Journal of Public Economics* 14: 161–194.
- Hausman, J. (1981), "Labor supply. how taxes affect economic behavior", in: H. Aaron and J. Pechman, eds., *Tax and the economy* (Brookings Institution, Washington, DC).
- Hausman, J. (1985a), "The econometrics of nonlinear budget sets", *Econometrica* 53: 1255–1282.
- Hausman, J. (1985b), "Taxes and labor supply", in: A. Auerbach and M. Feldstein, eds., *Handbook of public economics*, Vol. 1 (North Holland, Amsterdam).
- Hausman, J. and P. Ruud (1986), "Family labor supply with taxes", *American Economic Review (Papers and Proceedings)* 74: 242–248.
- Hausman, J. and W. Taylor (1981), "Panel data and unobservable effects", *Econometrica* 49: 1377–1398.
- Heckman, J.J. (1974a), "Shadow prices, market wages and labor supply", *Econometrica* 42: 679–694.
- Heckman, J.J. (1974b), "Life-cycle consumption and labor supply: an explanation of the relationship between income and consumption over the life cycle", *American Economic Review* 64: 188–194.
- Heckman, J.J. (1974c), "Effects of child-care programs on women's work effort", *Journal of Political Economy* 82(2): S136–S163.
- Heckman, J.J. (1976), "Life-cycle model of earnings, learning and consumption", *Journal of Political Economy* 84: S11–S44.

- Heckman, J.J. (1978), "Dummy endogenous variables in a simultaneous equation system", *Econometrica* 46: 931–959.
- Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica* 47: 153–162.
- Heckman, J.J. (1993), "What has been learned about labor supply in the past twenty years", *American Economic Review (Papers and Proceedings)* 83: 116–121.
- Heckman, J.J. and T.E. MaCurdy (1980), "A life-cycle model of female labour supply", *Review of Economic Studies* 47: 47–74.
- Heckman, J.J. and B. Singer (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data", *Econometrica* 52: 271–320.
- Heckman, J.J. and R.J. Willis (1977), "A beta-logistic model for the analysis of sequential labor force participation by married women", *Journal of Political Economy* 85: 27–58.
- Hotz, V.J. and R.A. Miller (1993), "Conditional choice probabilities and the estimation of dynamic models", *Review of Economic Studies* 60: 497–530.
- Hotz, V.J., F.E. Kydland and G.L. Sedlacek (1988), "Intertemporal substitution and labor supply", *Econometrica* 56: 335–360.
- Hoynes, H.W. (1996), "Welfare transfers in two-parent families: labor supply and welfare participation under AFDC-UP", *Econometrica* 64(2): 295–332.
- IImakunnas, S. and S. Pudney (1990), "A model of female labour supply in the presence of hours restriction", *Journal of Public Economics* 41: 183–210.
- Johnson, T.R. and J.H. Pencavel (1984), "Dynamic hours of work functions for husbands, wives and single females", *Econometrica* 52: 363–389.
- Kaiser, H., U. van Essen and P.B. Spahn (1992), "Income taxation and the supply of labour in West Germany", *Jahrbucher fur Nationalokonomie und Statistik* 209/1-2: 87–105.
- Kapteyn, A. and I. Woittiez (1998), "Social interactions and habit formation in a model of female labour supply", *Journal of Public Economics* 70: 185–205.
- Keane, M.P. and R. Moffitt (1995), "A structural model of multiple welfare program participation and labor supply", *International Economic Review* 39(3): 553–589.
- Kell, M. and J. Wright (1989), "Benefits and the labour supply of women married to unemployed men", *Economic Journal Conference Supplement*: 1195–1265.
- Killingsworth, M. and J. Heckman (1986), "Female labor supply: a survey", in: O. Ashenfelter and R. Layard, eds., *Handbook of labor economics*, Vol.1 (North-Holland, Amsterdam) pp. 103–204.
- Kooreman, P. and A. Kapteyn (1986), "Estimation of rationed and unrationed household labour supply functions using flexible functional forms", *Economic Journal* 96: 308–322.
- Kooreman, P. and A. Kapteyn (1990), "On the empirical implementation of some game theoretic models of household labour supply", *Journal of Human Resources* 25: 584–598.
- Lundberg, S. (1988), "Labor supply of husbands and wives", *Review of Economics and Statistics* 70: 224–235.
- MaCurdy, T.E. (1981), "An empirical model of labour supply in a life-cycle setting", *Journal of Political Economy* 89: 1059–1085.
- MaCurdy, T.E. (1983), "A simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty", *International Economic Review* 24: 265–289.
- MaCurdy, T.E. (1985), "Interpreting empirical models of labour supply in an intertemporal framework with uncertainty", in: J.J. Heckman and B. Singer, eds., *Longitudinal analysis of labour market data* (Cambridge University Press, Cambridge, UK).
- MaCurdy, T.E. (1992), "Work disincentive effects of taxes: a re-examination of some evidence", *American Economic Review* 82: 243–249.
- MaCurdy, T., D. Green and H. Paarsch (1990), "Assessing empirical approaches for analyzing taxes and labour supply", *Journal of Human Resources* 25: 415–490.
- Manser, M. and M. Brown (1980), "Marriage and household decision making", *International Economic Review* 21: 31–44.

- McElroy, M.B. (1981), "Empirical results from estimates of joint labor supply functions of husbands and wives", in: R.G. Ehrenberg, ed., *Research in labor economics*, Vol. 4 (JAI Press, Greenwich, CT) pp. 53–64.
- McElroy, M. (1990), "The empirical content of nash-bargained household behavior", *Journal of Human Resources* 25: 559–583.
- Moffitt, R. (1983), "An economic model of welfare stigma", *American Economic Review* 73: 1023–1035.
- Moffitt, R. (1986), "The econometrics of piecewise-linear budget constraints: survey and exposition of the maximum likelihood method", *Journal of Business and Economic Statistics* 4: 317–327.
- Moffitt, R. (1992a), "Estimating the value of an in-kind transfer: the case of food stamps", *Econometrica* 57: 385–409.
- Moffitt, R. (1992b), "Incentive effects of the US welfare system: a review", *Journal of Economic Literature* 15: 1–161.
- Moffitt, R. (1993), "Identification and estimation of dynamic models with a time series of repeated cross sections", *Journal of Econometrics* 59: 99–124.
- Moffitt, R. (1993), "The effect of work and training programs on entry and exit from the welfare caseload", Discussion paper no. 1025-93 (Institute for Research on Poverty, University of Wisconsin, Madison, WI).
- Moffitt, R. and B. Wolfe (1992), "The effect of the medicaid program on welfare participation and labor supply", *Review of Economics and Statistics* 74: 615–626.
- Mroz, T.A. (1987), "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions", *Econometrica* 55: 765–800.
- Nakamura, A. and M. Nakamura (1981), "A comparison of the labour force behaviour of married women in the United States and Canada, with special attention to the impact of income taxes", *Econometrica* 49: 451–489.
- Nerlove, M. (1971), "A note on error components models", *Econometrica* 39: 359–382.
- Newey, W.K., J.L. Powell and J.R. Walker (1990), "Semiparametric estimation of selection models: some empirical results", *American Economic Review* 80(2): 324–328.
- Pagan, A. (1986), "Two stage and related estimators and their applications", *Review of Economic Studies* 53: 517–538.
- Pencavel, J. (1986), "Labor supply of men: a survey", in: O. Ashenfelter and R. Layard, eds., *Handbook of labor economics*, Vol. 1 (North-Holland, Amsterdam) pp. 3–102.
- Ransom, M.R. (1987), "An empirical model of discrete and continuous choice in family labor supply", *The Review of Economics and Statistics* 59: 465–472.
- Ribar, D. (1992), "Child care and the labor supply of married women", *Journal of Human Resources* 27: 134–165.
- Robins, P. (1985), "A comparison of the labor supply findings from the four negative income tax experiments", *Journal of Human Resources* 20: 567–582.
- Rosen, H.S. (1978), "The measurement of excess burden with explicit utility functions", *Journal of Political Economy* 86: s121–s135.
- Shaw, K. (1989), "Life-cycle labour supply with human capital accumulation", *International Economic Review* 30(2): 431–457.
- Smith, J.P. (1977), *Female labor supply: theory and estimation* (Princeton University Press, Princeton, NJ).
- Smith, R.J. and R.W. Blundell (1986), "An exogeneity test for the simultaneous equation tobit model", *Econometrica* 54: 679–685.
- Stern, N. (1986), "On the specification of labor supply functions", in: R.W. Blundell and I. Walker, eds., *Unemployment, search and labour supply* (Cambridge University Press, Cambridge, UK) pp. 121–142.
- Triest, R. (1990), "The effect of income taxation on labor supply in the US", *Journal of Human Resources* 25: 491–516.
- Van Soest, A., I. Woittiez and A. Kapteyn (1990), "Labour supply, income taxes and hours restrictions in the Netherlands", *Journal of Human Resources* 25: 517–558.
- Van Soest, A. (1995), "Structural models of family labor supply: a discrete choice approach", *Journal of Human Resources*, 30: 63–88.

- Wagner, Gert G., Richard V. Burkhusser and Friederike Behringer (1993), "The English language public use file or the German socio-economic panel", *Journal of Human Resources* 28(2 Spring): 429–433.
- Wales, T.J. and A. Wodland (1976), "Estimation of household utility functions and labor supply response", *International Economic Review* 17: 69–80.
- Zabalza, A. (1983), "The CES utility function, nonlinear budget constraints and labor supply", *Economic Journal* 93: 312–330.