Affordability and the demand for separate accommodation

J R Miron

University of Toronto, 455 Spadina Avenue, Toronto, Ontario M5S 2G8, Canada

Received 20 December 1993; in revised form 10 March 1995

Abstract. The number of private households in Canada surged in the 1970s; at the same time, rental housing was becoming more affordable. Then, in the first half of the 1980s, affordability declined and household formation slowed. Were these changes in household formation attributable to changing affordability? In this paper, a model is estimated by means of logistic regression, in which the living arrangement of a family (or a nonfamily adult) is a choice made on the basis of the income and other characteristics of the family (or adult), and the price of maintaining one's own dwelling. In the model current income is partitioned into expected and transitory components, indicators of income expectations are included, and the endogeneity of living arrangement and income is considered. Income variables are measured in terms of the units of standard rental housing that the family or individual can purchase. The model is estimated empirically by using pooled data drawn from the 1981 Census and 1986 Census public-use samples. The estimation is comprehensive in coverage; separate model estimates are presented for up to 44 distinct pools of consumers defined by age, sex, family status, and income. Estimates derived from the model show that affordability had an effect on living arrangement that, although often statistically significant, was numerically small and that declining affordability does not predict well changes in living arrangement in metropolitan Canada for 1981 and 1986.

The number of private households in Canada surged in the 1970s. Formation averaged 225 000 net new households annually from 1971 to 1981 (see table 1). At the time, this surge was widely seen to be demographic in origin. Members of the postwar baby-boom cohort had just begun to leave the parental home and enter the housing market on their own. Also contributing to net formation was vigorous immigration from abroad and improved longevity. In addition, the 1970s marked the beginning of the marriage bust; conventional first marriage and remarriage rates

Table 1. Households and potential households in Canada, 1971–86 (source: quinquennial Census of Canada).

	1971	1976	1981	1986
Potential number of households (in thousands):				
Persons aged 20 years or older not living with spouse				
never married	2 272	2 550	3 034	3 557
separated or spouse absent	567	637	727	789
widowed	941	1 042	1 157	1 249
divorced	174	302	500	690
Husband–wife families	4 605	5 169	5 611	5 881
Total	8 559	9 699	11 029	12 166
Actual number of households (in thousands)	6 035	7 166	8 282	8 992

Note: since 1981, the census has included common-law couples in counts of husband–wife families; the number of persons who are separated or with spouse absent is estimated as the number of married persons less twice the number of husband–wife families and includes a small number of persons separated or with spouse absent who are aged 15–19 years.

slumped, and the incidence of divorce and separation rose. The number of households not headed by a husband–wife couple increased quickly as a consequence.

Private-household formation slowed in the first half of the 1980s: averaging only 142 000 net new households from 1981 to 1986. At the same time, however, the demographic trends that had been thought to underlie rapid formation of households in the 1970s still had not changed substantially by 1986. One measure of demographic change is the number of private households formed if every husband–wife couple (regardless of age) and every other person (that is, every other person not living with a spouse) aged 20 years or older lived in separate accommodation. As shown in table 1, the number of ‘potential households’ increased by approximately 227 000 units annually from 1981 to 1986: down only modestly from the 247 000 averaged from 1971 to 1981. The experience was not unique to Canada: Blank and Rosen (1990, table 2) found that the proportions of families and individuals in large US cities who head a dwelling also fell overall during the period 1977–87.

Why, then, did private-household formation slump in Canada between 1981 and 1986? A body of empirical research has emerged in recent decades wherein living arrangements are viewed as choices from among alternatives; choices that are determined in part by the relative price or ‘affordability’ of each alternative (see Hickman, 1974; Miron, 1988; 1989; 1993; Smith, 1984). In such research, affordability is commonly measured as the ratio of the annual price (rent) paid for accommodation to the income of the consumer. Available empirical evidence suggests that the affordability of housing in Canada improved during the 1970s, then worsened in the first half of the 1980s (see table 2). It could be argued that these changes in affordability caused household formation to slump between 1981 and 1986.

Table 2. The affordability of housing in Canada, 1971–86 (source: Statistics Canada, various years).

	1971	1976	1981	1986
Housing component of the consumer price index ^a	100	148	226	301
Personal disposable income (Can\$ billion)	61	128	238	338
Potential households (in thousands) ^b	8 559	9 699	11 029	12 166
Personal disposable income per potential household ^a	100	185	302	389
Affordability index ^{a, c}	100	80	75	77

^a Indexed to equal 100 in 1971.

^b As estimated in table 1.

^c Ratio of the housing component of the consumer price index to the personal disposable income per potential household.

However, was the worsening of affordability substantial enough to account for the drop in household formation? The ‘traditional’ family—a husband–wife couple and their children—that forms the most common kind of household is widely thought to make a living-arrangement choice that is largely independent of affordability: high-income and low-income traditional families alike have a high propensity to form a separate household. However, as noted above, the number of households increased quickly during the 1970s because of the formation of ‘nontraditional’ households (lone-parent families and unattached individuals) and young and elderly couples who typically had low incomes compared with traditional families. The higher the incidence of nontraditional households, the more plausible the argument that household formation has become more sensitive to affordability.

In table 3 I provide evidence that the drop in household formation from 1981 to 1986 was concentrated among nontraditional and younger households. Consider, for example, the 25–34 years age-group. In the 1981 Census, 1 758 000 consumers (husband–wife families, lone-parent families, and nonfamily adults) in the 25–34 years age-group (age of husband in husband–wife families) were reported in selected metropolitan areas. Of these, 1 476 000, or 84%, maintained their own dwelling. In 1986, the corresponding figure was only 73%. The decline was most marked among young single women and separated men.

Table 3. The estimated number (in thousands) of census families and nonfamily persons in the sample pools, showing in parentheses the proportion maintaining own dwelling by age-group of husband, lone parent, or nonfamily person, and by income status (above or below the low-income cutoff), in selected Canadian metropolitan areas, 1981 and 1986 (data are calculated from the sample pools described in table 4).

Pool	25–34 years age-group		35–60 years age-group	
	1981	1986	1981	1986
Husband–wife family				
above cutoff	1 077 (0.98)	519 (0.97)	1 285 (0.99)	1 480 (0.99)
below cutoff	44 (0.97)	60 (0.98)	99 (0.99)	123 (0.98)
Female lone-parent family				
above cutoff	31 (0.85)	33 (0.83)	100 (0.94)	127 (0.93)
below cutoff	44 (0.95)	61 (0.95)	65 (0.95)	82 (0.97)
Male lone-parent family				
above cutoff	5 (0.75)	8 (0.78)	36 (0.95)	41 (0.93)
below cutoff	1 (0.87)	2 (0.87)	4 (0.87)	9 (0.92)
Nonfamily person:				
divorced woman				
above cutoff	17 (0.71)	15 (0.73)	31 (0.87)	45 (0.87)
below cutoff	6 (0.56)	7 (0.50)	17 (0.81)	25 (0.76)
separated woman				
above cutoff	14 (0.78)	11 (0.76)	15 (0.84)	17 (0.90)
below cutoff	6 (0.57)	6 (0.55)	12 (0.81)	12 (0.72)
single woman				
above cutoff	139 (0.64)	162 (0.58)	85 (0.72)	89 (0.76)
below cutoff	49 (0.39)	74 (0.37)	29 (0.50)	32 (0.55)
widowed woman				
above cutoff	0 (–)	1 (–)	25 (0.91)	20 (0.91)
below cutoff	0 (–)	0 (–)	20 (0.76)	19 (0.78)
divorced man				
above cutoff	20 (0.70)	17 (0.66)	40 (0.80)	54 (0.79)
below cutoff	4 (0.52)	7 (0.52)	12 (0.64)	21 (0.68)
separated man				
above cutoff	24 (0.72)	21 (0.62)	35 (0.81)	38 (0.77)
below cutoff	5 (0.59)	7 (0.47)	10 (0.71)	12 (0.68)
single man				
above cutoff	202 (0.52)	234 (0.51)	89 (0.68)	95 (0.71)
below cutoff	70 (0.37)	121 (0.33)	33 (0.49)	44 (0.55)
widowed man				
above cutoff	0 (–)	0 (–)	7 (0.86)	6 (0.93)
below cutoff	0 (–)	0 (–)	3 (0.62)	4 (0.71)
Total in pool	1 758	1 363	2 052	2 394
Total maintaining own household	1 476 (0.84)	993 (0.73)	1 894 (0.92)	2 219 (0.93)

– Sample size too small for reliable estimation.

My purpose in this paper is to estimate the effect of changes in affordability on the formation of private households in Canada in the early 1980s. Because the effect of affordability on living arrangement can vary with the circumstances of demography and income, a comprehensive set of model estimates are presented that cover all family heads, spouses, and nonfamily individuals aged 25–60 years. A model is estimated by using logistic regression wherein the living arrangement of each sampled family (or nonfamily adult) is seen to be a choice between two alternatives, made on the basis of the income and other characteristics of the family (or adult) and the price of maintaining one's own dwelling.⁽¹⁾ The exogenous variables include several demographic traits. However, in a subtle but important distinction, the aim is not so much to explain or replicate variations in living arrangements as to assess the importance of affordability as an explanation. To do this, the current-income variable is partitioned into expected and transitory components and is supplemented with indicators of income expectations: six exogenous variables in all. The comprehensiveness of the income measures used here helps to clarify the role that affordability plays in choice of living arrangement, as well as to address the endogeneity of income and living arrangement. In the base runs described below, income variables are measured in terms of the units of standard rental housing that families or individuals can purchase; hence each variable measures affordability. The model is estimated empirically by using pooled data drawn from the 1981 Census and 1986 Census public-use samples; the appropriateness of pooling is considered below.

The model

Suppose that consumer i (here referred to in the singular but taken to be a husband–wife family, lone-parent family, or nonfamily adult) has an income of Y_i^m if maintaining a dwelling, and an income of Y_i^n if not, and that social and demographic characteristics are represented by a vector X_i . Presumably, the incomes Y_i^m and Y_i^n differ to the extent that work-force participation (and its complement in 'leisure' or household production) and living arrangements are chosen jointly. Suppose that i makes a choice between two alternatives, maintaining a dwelling of his or her own (indexed here by superscript m) or lodging with others (indexed by superscript n). Here, 'lodging' includes any living arrangement in which the consumer is not maintaining his or her own dwelling: for example, a person paying rent to live in the dwelling of another consumer, or a person living rent-free with a relative or friend. Let the relative price for the service provided by a standard dwelling of one's own be P^m , and the relative price for a standard dwelling in which the consumer lodges (that is, lives in a dwelling maintained by someone else) be P^n . Suppose further that the indirect utility for each alternative— U_i^m and U_i^n , respectively—can be written as:

$$\left. \begin{aligned} U_i^m &= f^m(Y_i^m, P^m, X_i), \\ U_i^n &= f^n(Y_i^n, P^n, X_i). \end{aligned} \right\} \quad (1)$$

Because P^n is unobserved in the sample, simply assume that it is a fixed fraction of P^m . To simplify the model further, set P^m as the numeraire and measure income as the number of standard own dwellings from which services can be purchased.

⁽¹⁾ In contrast, the dependent variable in Blank and Rosen (1990) is the proportion of consumers in a city who maintain a dwelling of their own. Their observations are thus for cities at particular dates rather than for individual consumers.

Offsetting factors determine the utility functions, f^m and f^n . On the positive side, lodging can enable shared domestic production, and the resulting reduction in shelter expense reduces the need for paid work. In economic terms, lodging may enable a different division of household labour (both inside and outside the home). Economies of scale in the use of pooled furnishings (for example, a sofa) may also help make lodging an efficient living arrangement. On the negative side is the loss of privacy, control, primacy, and personal security that may arise in lodging, coupled with congestion, transactions costs, and other negative externalities in the use of shared facilities.

The economic literature on living arrangements dates back to the seminal work of Beresford and Rivlin (1966), and includes Ermisch (1981), McElroy (1985), Börsch-Supan (1986), Blank and Rosen (1990), Avery et al (1992), and Haurin et al (1993; 1994). In much of this literature, income, y , is exogenous (that is, $y_i^m = y_i^n$) and the maintaining of a dwelling is seen as a superior good. Implicit or explicit in this literature is the view that an adequate income gives the consumer the freedom to choose a preferred living arrangement, but that a low income compels the consumer to lodge. Elsewhere, the endogeneity of living arrangement and income has been modeled explicitly.

In this paper, a consumer's income is divided into four components: investment income (Y^z),⁽²⁾ expected noninvestment income (Y^e),⁽³⁾ positive transitory noninvestment income (Y^a), and negative transitory noninvestment income (Y^b). Y^e is defined to be the average noninvestment income of persons in similar circumstances. In turn, Y_i^a , Y_i^b , and the consumer's total income, Y_i , are defined as follows:

$$Y_i^a = \begin{cases} (Y_i - Y_i^z) - Y_i^e & \text{if this is positive,} \\ 0, & \text{otherwise;} \end{cases} \quad (2)$$

$$Y_i^b = \begin{cases} Y_i^e - (Y_i - Y_i^z) & \text{if this is positive,} \\ 0, & \text{otherwise;} \end{cases} \quad (3)$$

and

$$Y_i = Y_i^z + Y_i^e + Y_i^a - Y_i^b. \quad (4)$$

It is well known that permanent income, which is here proxied by expected income, is important in housing demand overall. What is not yet clear is whether permanent income affects household formation to the same extent that it affects tenure choice and possibly the quantity of housing services demanded. In this study I estimate empirically the impacts on household formation of both expected income and transitory income. I also estimate separately the impacts of positive versus negative transitory income. If consumers treat all dollars of income equivalently, the coefficients of these four components should be similar in magnitude (although the coefficient of Y_i^b , as defined, will be opposite in sign).

⁽²⁾ In the data sets employed in this study, investment income includes dividends, interest, and other investment income: for instance, interest received from deposits in financial institutions, savings certificates, and bonds and debentures; stock dividends; net rents from real estate; mortgage and loan interest received; regular income from an estate or trust fund; and interest from insurance policies. It does not include net receipts from an unincorporated business or professional practice, or royalties.

⁽³⁾ Expected noninvestment income is largely, but not exclusively, expected earnings.

By decomposing income into expected and transitory components one can address the consumer's sense of the 'permanence' of current income compared with the incomes of peers. Y_i^e indicates the typical income of persons of the same age, geographic location, level of schooling, and work status; hence, deviations from it may represent impermanence in the sense of chance. However, Y_i^e does not measure how consumers expect their incomes to change in the near future either because of career-path effects or because of a change in work status (from working part-time to working full-time, for instance). In this study, two additional measures of income expectation are used. Let Y_i^5 be the difference between the expected income of a peer aged 5 years older than the consumer and the consumer's own expected income. Y_i^5 is a career-path measure indicating how the consumers expect their own incomes to change over the next 5 years. Let Y_i^f be the difference in expected income between a fully employed peer and consumer i ; hence $Y_i^f = 0$ if consumer i is currently fully employed. Y_i^f is a measure of how much consumers might expect their own incomes to change if they were to work full time.⁽⁴⁾

McElroy (1985) and Haurin et al (1993) argue that current income, which consists primarily of earnings, is endogenous in a model of living arrangements.⁽⁵⁾ Earnings are the product of a wage rate and quantity of labour supplied, and these authors argue that quantity of labour supplied may well be a decision made jointly with household formation. Of course, the jointness of labour-force and living arrangements is an empirical question; McElroy models living arrangements and labour-force activity choices, and finds evidence of a significant simultaneity in a sample of out-of-school US men aged 19–23 years. Haurin et al take a simpler approach. They argue (1993, page 284) that "the correct concept for the household formation decision is the real wage that could be earned should the respondent take on the responsibility of independent living". However, they do not test this assertion. In this paper, $Y_i^e + Y_i^f$ corresponds to a full-employment wage. If Haurin et al are correct, the coefficients of Y_i^f and Y_i^e in the model presented here should be similar. Haurin et al (1993, page 289) themselves find that goodness of fit declines when predicted wage is used instead of current income. To be fair to them, however, their argument is that coefficient of current income is biased and not that a full-employment wage necessarily improves goodness of fit. A related argument is that consumers attach a monetary value to domestic production and to the labour so engaged. As a measure of that activity, Y_i^f may influence living arrangement, albeit not necessarily with the same coefficient as Y_i^e .

Maximum likelihood logistic regression is used in this estimation. For a group of consumers of which i is a member, let $LIV_i = 1$ if the consumer is observed to maintain a dwelling of his or her own, and 0 otherwise.⁽⁶⁾ Let p_i be the probability that consumers in the group with characteristics similar to those of i maintain a dwelling of their own; that is, $p_i = \text{prob}(LIV_i = 1 | \text{model})$. Given that the annual price, P^m , of maintaining a standard dwelling of one's own is the numeraire, let the following

⁽⁴⁾ Ignored here is sample selection bias; namely, that persons working full time earn a different hourly wage than peers working part time. In their study of US youth, Haurin et al (1993, page 286) use a Mill's-ratio procedure to correct for such a bias but find no significant difference.

⁽⁵⁾ Blau and Robins (1991) make a related point in modeling the child-care and labour-supply decision of young mothers.

⁽⁶⁾ In contrast, Haurin et al (1993) analyze the decision to leave a parental home and the decision to live alone.

model be assumed to underlie behaviour of each consumer, i , in the group:

$$\left. \begin{aligned} p_i &= \frac{\exp z_i}{1 + \exp z_i}, \\ \text{where} \\ z_i &= \alpha_0 + \beta X_i + \gamma_1 Y_i^z + \gamma_2 Y_i^c + \gamma_3 Y_i^a + \gamma_4 Y_i^b + \gamma_5 Y_i^s + \gamma_6 Y_i^f. \end{aligned} \right\} \quad (5)$$

and where α_0 , β , and $\gamma_1, \dots, \gamma_6$, are regression coefficients. In this choice model, demographic variables X_i enter by translation.⁽⁷⁾

Data sets, pools, and empirical approach

In this paper the parameters in equation (5) are assumed to vary among categories, or 'pools', of consumers. A separate set of parameter estimates is derived for each pool. The data sets used in this paper are pooled subsamples from the 1981 Census and 1986 Census public-use samples. Pooling of the two censuses is appropriate because in this paper the aim is to study whether the change in affordability between these two years—rather than a change in the structural form of the model—explains the change in incidence of maintainership. For this study I have employed separate pools for husband–wife families, for lone-parent families, and for non-family adults. In modern Canadian censuses, a (nuclear) family consists of a husband–wife couple (conventional marriage or common-law partners) living with any never-married children, or a lone parent living with one or more never-married children. Children can be of any age. Any other individual is a nonfamily person. This is identical to the definition used by Blank and Rosen (1990). The census 1% family file is used to derive the samples for husband–wife families and lone-parent families; the 2% individual file is used for nonfamily adults. The census family file is limited to families living in private dwellings (and hence excludes families in institutional accommodation and in other collectives); in drawing the subsample of nonfamily adults, a similar restriction was used. In Canadian censuses, individuals and families live either in private dwellings or in collectives (which include institutional residences, dormitories, barracks, work camps, residential hotels, and rooming houses). Whereas most cross-sectional studies of living arrangements in North America focus on persons living in private dwellings, Blank and Rosen (1990, footnote 10) appear to include persons and families in collective dwellings. Restrictions on the observations included in each pool are outlined in table 4.

In this paper, consumers are seen to make a choice of living arrangement on the basis of income, preferences, and the price of unsubsidized rental accommodation. The living arrangement of a consumer who receives a housing subsidy may well be different. Smith (1984) finds that, among the elderly, household formation in Canada is sensitive to the availability of subsidized (senior-citizen) housing. However, the census public-use samples do not distinguish between persons living in subsidized private accommodation (for instance, public housing, employee housing, or accommodation subsidized by a relative living elsewhere) and those living in unsubsidized private accommodation. As many of those who live in subsidized accommodation have low incomes, consumers are treated separately in this paper on the basis of income status (that is, above or below the low-income cutoff). The Statistics Canada 1978 low-income cutoffs, updated by changes in the consumer price index, are used for the husband–wife and lone-parent family pools. These

⁽⁷⁾ Goodman (1990) discusses the use of translation versus scaling in the introduction of demographic variables to models of housing demand.

cutoffs vary with size of family and population size of the area of residence and they test the total income of the economic family that includes the nuclear family. The nonfamily-adult pools do not include information about these low-income cutoffs; in this case, consumers were defined to be low-income if their average monthly income was not at least $2\frac{1}{2}$ times the average monthly rent for a standard-quality dwelling in their city. Blank and Rosen (1990) distinguish three income groups, the lowest of which corresponds to the below-cutoff pools used here. As I do not know which of the sampled consumers occupy subsidized housing, empirical findings for low-income consumers here should be interpreted with caution. Blank and Rosen find that the availability of public and rent-subsidized housing has a significantly positive effect on the propensity to maintain a dwelling among poor and near-poor consumers.

The model is initially estimated for 44 separate pools of consumers. In the case of husband-wife families, pools are defined by age-group of husband (25–34 years or 35–60 years) and income status. Among lone-parent families, pools are further

Table 4. Subsample restrictions in the construction of pools.

Husband-wife family pool	Lone-parent family pool	Nonfamily adult pool
<i>Census metropolitan areas included:</i>		
Calgary, Edmonton, Halifax, Hamilton, Kitchener, London, Montreal, Ottawa-Hull, Quebec, St Catharines, Toronto, Vancouver, Windsor (1986 only), and Winnipeg	Calgary, Edmonton, Halifax, Hamilton, Kitchener, London, Montreal, Ottawa-Hull, Quebec, St Catharines, Toronto, Vancouver, Windsor (1986 only), and Winnipeg	Calgary, Edmonton, Halifax (1981 only), Hamilton, Kitchener, London (1981 only), Montreal, Ottawa-Hull, Quebec (1981 only), Regina (1986 only), St Catharines, Saskatoon (1986 only), Toronto, Winnipeg, and Vancouver
<i>Family structure:</i>		
No restrictions	Includes only lone parents living with never-married child(ren) (aged under 25 years)	Not applicable
<i>Income:</i>		
Includes only families with nonzero total family income	Includes only families with nonzero total family income	Includes only persons with nonzero total individual income
<i>Age-group:</i>		
Husband and wife each aged 25–34 years or 35–60 years	Lone parent aged 25–34 years or 35–60 years	Person aged 25–34 years or 35–60 years
<i>Shelter expense:</i>		
Includes only families living in households that report a nonzero average monthly shelter expense (owner's major payments or gross rent) in 1980 or 1985	Includes only families living in households that report a nonzero average monthly shelter expense (owner's major payments or gross rent) in 1980 or 1985	Includes only persons living in households that report a nonzero average monthly shelter expense (owner's major payments or gross rent) in 1980 or 1985
<i>Low-income status:</i>		
Excludes families for which low-income cutoff status is not available	Excludes families for which low-income cutoff status is not available	Not applicable

defined by the gender of the parent. Among nonfamily adults, the pools are further differentiated by marital status (single, widowed, divorced, and separated). All categories are further divided into above-cutoff and below-cutoff pools. To summarize, the disaggregations are as follows:

Category	Number of pools	Breakdown
Husband-wife families	4	Income status (2 statuses), by age of husband (2 categories)
Lone-parent families	8	Age (2 categories) and sex of lone parent, by income status
Nonfamily adults	32	Marital status (4 statuses), by income status, by sex and age

Because in this paper the aim is to look only at the living arrangements of consumers within each pool, factors that cause the number of consumers to be found in a particular pool to vary over time (that is, the selection problem) are not considered. Affordability may, for example, affect the incidence of home-leaving among young adults, nuptiality, or marital breakdown. Haurin et al (1993) endogenize decisions about home-leaving and marriage in their study of living arrangements among young adults. Although the effects of affordability are unquestionably more pervasive, the purpose in this paper is to show the effect on incidence of maintainership within each pool.⁽⁸⁾

The living arrangements of almost all adults from the age of 25 years to 60 years are considered in this paper. Within this sample, a focus is put on those aged 25–34 years, the age-group wherein changes in living arrangement between 1981 and 1986 were most substantial. In a similar approach, Haurin et al (1993; 1994) restrict their analysis to young adults (aged 22–29 years). The living arrangements of adults under the age of 25 years are not considered in the present study. Although the age of 25 years is arbitrary, younger adults are excluded from this model, because our model of living arrangement is based upon measures of expected income that are conditional upon education. As younger adults have not yet completed their schooling, it is not possible to infer their income prospects. Of course, some individuals aged 25 years or older also have still not yet completed their schooling, but they are relatively few in number.

Variables

The dependent variable is LIV_i . In the 1981 and 1986 Canadian Censuses, respondents were asked to identify the person (or one of the persons) “who lives here and is responsible for paying the rent, or mortgage, or taxes, electricity, etc. for this dwelling”. That person is deemed to be the maintainer. Every dwelling has at most one maintainer; however, a few have none (that is, when no one living in the dwelling makes any such payment). The maintainer and the nuclear family (be it husband-wife or lone-parent family), if any, to which that maintainer belongs are said to maintain their own homes. Everyone else is a nonmaintainer.

Maintainership is an appropriate dependent variable as the focus in this paper is on household formation. Studies that are focused on another aspect of living

⁽⁸⁾ An extension to the present analysis that would consider both choice of pool and choice of maintainership might be patterned after the work of Goodman (1990) and others on simultaneous models of tenure choice and housing consumption.

arrangements may well use a different measure. For example, Haurin et al (1993) estimate the probability of home-leaving among young adults, as well as the conditional probability of living alone. Readers need to be mindful that the estimated effect of affordability on living arrangement may well vary with the dependent variable used.

The independent variables include a four-variable decomposition [Y^e (YSM), Y^a (YTA), Y^b (YTB), and Y^z (YINV)] of family income (individual income in the case of nonfamily adults) and two expected incomes [Y^f (YFT) and Y^s (YEXP)]. Social and demographic variables include age of the husband (AGEH) in the husband-wife pools, and age of the lone-parent or nonfamily adult (AGE) in other pools. AGE and AGEH are included to proxy changes in tastes and variations in wealth that occur across age-groups. In the husband-wife pools, AGED is the age of the husband minus the age of the wife; FSIZE, size of family, is used in the family pools; DCH6, a dummy variable for the presence of children under the age of six, is used in the family pools; IMM is a dummy variable for recent immigrants. LIV and the set of independent variables employed are defined in table 5.

To calculate expected noninvestment incomes, samples were taken from the 1981 and 1986 Census public-use samples (individual files only). Each sample included all persons aged 25 years or older living in private households in the following census metropolitan areas (CMAs): Calgary, Edmonton, Halifax (1981

Table 5. Variable definitions and application by sample: base runs.

Label	Definition
LIV	Equal to 1 if person or family maintains own dwelling, equal to 0 otherwise
AGE	Age at last birthday of person or lone parent (excludes husband-wife pools)
AGEH	Age at last birthday of husband (for husband-wife pools only)
AGED	The age of the husband less the age of the wife (for husband-wife pools only)
DCH6	Equal to 1 if a child under the age of 6 years is present, equal to 0 otherwise (excludes nonfamily adult pools)
FSIZE	Size of census family (excludes nonfamily adult pools)
IMM	Equal to 1 if person immigrated to Canada within past 8 years, equal to 0 otherwise; in husband-wife pools, IMM = 1 if either spouse immigrated within the past 8 years
YSM	Expected noninvestment income of the person, measured in units of P^m ; in the case of husband-wife families, YSM is the sum of expected noninvestment income for the two spouses
YTA	Transitory noninvestment income if positive, measured in units of P^m , equal to 0 otherwise; in the case of husband-wife families, YTA is the amount by which family noninvestment income exceeds the sum of the expected noninvestment incomes of the husband and wife
YTB	Negative transitory noninvestment income if negative, measured in units of P^m , equal to 0 otherwise; in the case of husband-wife families, YTB is the amount by which family noninvestment income is less than the sum of the expected noninvestment incomes of the husband and wife.
YINV	Investment income of individuals or family, measured in units of P^m
YEXP	Difference in expected income between the person and a peer aged 5 years older, measured in units of P^m ; in the case of husband-wife families, YEXP is the sum of the difference for each spouse
YFT	Difference in expected income, measured in units of P^m , for fully-employed person if not currently fully-employed, equal to 0 otherwise; in the case of husband-wife families, YFT is the sum of the difference for each spouse

Note: P^m , relative price for the service provided by a standard dwelling of one's own, that is, the annual rent, including utilities, that could be charged for a dwelling of standard quality.

only), Hamilton, Kitchener, London (1981 only), Montreal, Ottawa–Hull, Quebec (1981 only), Regina (1986 only), St Catharines, Saskatoon (1986 only), Toronto, Winnipeg, and Vancouver. Individuals were classified as either working full time in the preceding year or not: persons were deemed to work full time in the previous year (1980 or 1985) if they worked at least 49 weeks of mainly full-time work. Three levels of schooling were defined: low (persons with no more than grade 8 schooling or equivalent), moderate (those with more than grade 8 but no post-secondary schooling), and advanced (those with any postsecondary schooling). Each sample was disaggregated by sex, work status, and level of schooling into 12 distinct groups; average noninvestment income was calculated for each year of age; and expected noninvestment income was calculated as the centered running means⁽⁹⁾ of length 3 of these average individual incomes.⁽¹⁰⁾

This approach to estimating expected income is unusual, but not without merit. Expected income is here calculated by assuming that the individual does not change from part-time to full-time work or vice versa. However, in the living-arrangement model of equation (5) it is envisaged that a consumer working part time takes into account the potential increment in expected income from full-time employment. The advantage of this approach is that the estimation itself shows how the consumer incorporates expectations of change in work status.

The numeraire for all income measures is P^m , the annual rent, including utilities, charged for a dwelling of standard quality (hereinafter called 'shelter price'). Hence, all income variables can be thought to be indicators of the affordability of housing. Smith (1984) takes a similar approach, except that the affordability variable is specified in reciprocal form as the rent/income ratio rather than as the income/rent measure used here.

A hedonic price model was estimated for rents by using microdata from the 1974 Survey of Housing Units. A separate model was estimated for each CMA in that sample: a sampling of 63 020 households in 23 cities, conducted by Statistics Canada on behalf of Canada Mortgage and Housing Corporation. By using weighted least squares, monthly rent, including utilities, was regressed linearly against variables describing the total net floor area of the dwelling, average room size, state of repair, dwelling type, period of construction, toilet facilities, and neighbouring land uses. These estimated models were used to predict the average monthly rent (again including utilities) in each CMA for a hypothetical mix of dwellings corresponding to the mix of rental dwellings in CMAs across Canada. The 1974 shelter prices were then projected to 1980 and 1985 by using the rent index of each city's consumer price index (see table 6, see over).

⁽⁹⁾ Except for 25-year olds, where expected noninvestment income is simply the average noninvestment income among 25-year olds themselves.

⁽¹⁰⁾ An alternative approach is to estimate expected earnings by means of a regression model. However, the standard model includes a quadratic (or related) function of age or experience plus binary independent variables to measure variations attributable to schooling, region, gender, work status, and other factors. Although widely used, quadratic functions of age (or experience) are empirically problematical. In quadratic models symmetry of earnings above and below the age at which earnings peak is assumed, when, in fact, earnings tend to rise faster at younger ages than they fall at older ages. In the labour-market literature, there is no consensus on a replacement for the quadratic model. In this paper, age is entered into the model as a set of single-year dummy variables. When age dummies are combined with other explanatory variables that are each binary as well, a regression model becomes an analysis of subgroup means.

Table 6. The average monthly rent in selected Canadian dwellings, for the national mix of rental dwellings, 1980 and 1985 (source: Survey of Housing Units, 1974).

City	Average rent (Can\$ per month)		City	Average rent (Can\$ per month)	
	1980	1985		1980	1985
Calgary	287	375	Quebec	233	353
Edmonton	291	391	Regina ^c	262	384
Halifax	271	396	Saskatoon ^c	243	347
Hamilton ^{a, b}	248	351	St Catharines ^{a, b}	223	316
Kitchener ^{a, b}	237	336	Toronto	314	444
London ^b	242	342	Vancouver	311	437
Montreal	221	333	Windsor ^b	269	380
Ottawa-Hull	284	427	Winnipeg	255	373

Note: 1974 shelter prices were projected to 1980 and 1985 prices by using the rent index of each city's consumer price index.

^a For individual file, averaged to Can\$334 per month in 1985.

^b City consumer price index rent index not available. Annual change was estimated by using the corresponding change for Toronto.

^c For individual file, averaged to Can\$366 per month in 1985.

Estimation of the model

The model described by equation (5) is estimated by using the logistic regression procedure of SPSS (version 4.1). As the 1981 Census public-use family file is a stratified sample with weights provided, the SPSS procedure WEIGHT was used for the husband-wife and lone-parent family pools. The 1981 file is weighted because it is a combined household and family file. The weights are intended to inflate sample counts correctly to a 1% estimate of the population of families; hence, the mean weight in the 1981 family sample is 1.0. In the pooled samples, observations from the 1986 file (which is a simple 1% random sample) are each given a weight of 1.0. No weighting is used in estimation with the individual files because they are simple random samples. Further, because YTA tends to be almost always zero among consumers with incomes below the low-income cutoff, it is not included as an explanatory variable for such consumers.

Younger consumers above the low-income cutoff

In table 7 model estimates are given for husband-wife families where the husband is in the age-group 25-34 years. In table 8 the model estimates are given for male-led and female-led lone-parent families for lone parents, again in the age-group 25-34 years. In each case, the sample mean is shown for each variable in the pooled sample, as well as the logistic regression coefficients (showing statistical significance) and the estimated probability partial derivatives (calculated at the variable means). For example, expected noninvestment income (YSM), was, on average, 8.76 affordability units for young husband-wife families with incomes above the low-income cutoff. From the logistic regression, the effect of an increase in YSM of one affordability unit (varying from one city to the next, but typically about Can\$300 per month) at the mean income is to increase the probability of maintaining one's own dwelling by a modest 0.007. The equivalent results for nonfamily adults are presented in table 9.

Also shown in table 7 are the sample sizes for each pool. Some of these samples (notably, husband-wife families and nonfamily single persons) are much larger than others. With large samples, even a small numerical difference can be statistically

Table 7. Sample means, sample sizes, and model estimation for husband–wife families, with income above the cutoff level, where husband is aged 25–34 years, for Canadian metropolitan areas, 1981–86.

Variable ^a	Sample mean	Coefficient	Probability partial derivative ^b
AGED	0.94	0.015	0.000
AGEH	30.39	−0.024	−0.000
DCH6	0.55	−0.397*	−0.005
FSIZE	3.18	0.421*	0.005
IMM	0.05	−0.681*	−0.009
YEXP	1.23	0.041	0.001
YFT	2.45	0.316*	0.004
YINV	0.17	0.147	0.000
YSM	8.76	0.535*	0.007
YTA	1.61	0.074	0.001
YTB	0.91	−0.361*	−0.005
Intercept		−1.393	
Sample size	11810		
Likelihood ratio χ^2 statistic ^c	263*		
Pseudo R^2 statistic ^d	0.101		

* Test is significant at 95% level. For logistic coefficients the test statistic is that for the Wald test for which the null hypothesis is that the corresponding population coefficient is zero.

^a Definitions of the variables are provided in table 5.

^b Calculated at variable means.

^c The number of degrees of freedom is the number of slope coefficients, and the null hypothesis is that the slope coefficients are simultaneously all zero.

^d Calculated as per Maddala (1983, page 40).

significant. Therefore, in comparing statistical significance between pools, one needs to be mindful of the impact of sample size.

Affordability variables

Tables 7–9 provide interesting evidence of the effect of affordability on living arrangement. In every pool, the coefficient of expected noninvestment income (YSM) is positive; it is also statistically significant for all pools except for nonfamily men. Although statistically significant in several of the pools, the evidence suggests also that the effect of YSM is numerically small. Although the probability partial derivative is large for a few pools (notably male lone parents), it is quite small for many others. Among single nonfamily men, for example, a one-unit increase in YSM leads to only an 0.010 increase in LIV.

The coefficients of the transitory noninvestment income variables, YTA and YTB, reveal asymmetric effects on living arrangement. In tables 7–9, the coefficient of YTB is everywhere negative as expected, and almost always statistically significant. Further, that the coefficients of YTB are often similar in absolute value to those of YSM suggests that the marginal effect of income on living arrangement is the same for expected income as for downside transitory income; put differently, consumers with incomes at or below YSM do not differentiate among dollars of noninvestment income. In contrast, the coefficients of YTA are smaller in absolute value than are those of YSM and YTB. In other words, the likelihood of maintaining a home increases with family income above YSM but responds less to a dollar increase in income above YSM than to a dollar decrease below it.

Returning to the endogeneity argument, how does the coefficient of YFT compare with that of YSM? Among husband–wife families (table 7), the coefficient

of YFT is statistically significant, positive, and about one-half the coefficient for YSM. Among lone-parent families (table 8), YFT is not statistically significant in either regression. Among nonfamily adults (table 9), YFT is statistically significant only for single men, but there it has the wrong sign. These findings support the argument that the effect of living arrangement on work status is more important among husband-wife families (presumably because of variations in work-force participation among wives) than among other pools. At the same time, these results do not suggest that YSM can be replaced simply by expected full-time income (that is, YSM+YFT) even for husband-wife families. The argument of Haurin et al (1993) that the correct concept for the household-formation decision is the wage that the resident could earn is not borne out here. At the same time, their endogeneity argument is more appropriate in the context of their own sample which includes younger adults and those living with a parent.

The present finding runs counter to the evidence in McElroy (1985). In part, this may be because McElroy's sample consisted only of single men in the age-group 19–23 years: a group for whom labour-force participation may be more discretionary. In this paper an older age-group, 25–34 years, has been covered, wherein the incentive or need to work is presumably more pressing.

Table 8. Sample means, sample sizes, and model estimation for lone-parent families, with income above the cutoff level, where the parent is aged 25–34 years, for Canadian metropolitan areas, 1981–86.

Variable ^a	Female-led families			Male-led families		
	sample mean	coefficient	probability partial derivative ^b	sample mean	coefficient	probability partial derivative ^b
AGE	30.29	0.084	0.009	30.68	0.012	0.002
DCH6	0.46	–0.279	–0.029	0.50	–0.752	–0.107
FSIZE	2.39	0.043	0.004	2.42	–0.337	–0.048
IMM	0.05	–0.406	–0.042	0.05	1.666	0.237
YEXP	0.21	0.017	0.002	0.96	–0.606	–0.086
YFT	0.99	–0.055	–0.006	0.96	0.446	0.064
YINV	0.07	–0.468	0.048	0.04	–0.738	–0.105
YSM	3.65	0.545*	0.056	5.55	0.921*	0.131
YTA	1.29	0.458*	0.047	1.11	–0.042	–0.006
YTB	0.30	–0.671*	–0.069	0.71	–0.576*	–0.082
Intercept		–2.841			–2.139	
Sample size		646			133	
Likelihood ratio χ^2 statistic ^c		487*			27*	
Pseudo R^2 statistic ^d		0.320			0.116	

* Test is significant at 95% level. For logistic coefficients the test statistic is that for the Wald test for which the null hypothesis is that the corresponding population coefficient is zero.

^a Definitions of the variables are provided in table 5.

^b Calculated at variable means.

^c The number of degrees of freedom is the number of slope coefficients, and the null hypothesis is that the slope coefficients are simultaneously all zero.

^d Calculated as per Maddala (1983, page 40).

Social and demographic variables

In this paper the focus is on the effect of affordability on living arrangement. However, what about the other independent variables? Age has a positive and statistically significant effect in five of the six above-cutoff, young, nonfamily person pools (table 9). Age is not significant, however, for either of the two corresponding

Table 9. Sample means, sample sizes, and model estimation for nonfamily adults, with income above the cutoff level, aged 25 – 34 years, for Canadian metropolitan areas 1981 – 86.

Variable ^a	Nonfamily adult women ^e			Nonfamily adult men ^e		
	sample mean	coefficient	probability partial derivative ^b	sample mean	coefficient	probability partial derivative ^b
Divorced						
AGE	30.12	0.145 *	0.027	30.53	0.149 *	0.031
IMM	0.04	1.154	0.218	0.05	0.653	0.138
YEXP	0.25	0.656	0.124	0.95	1.045 *	0.222
YFT	0.70	0.220	0.041	0.86	-0.059	-0.013
YINV	0.11	0.051	0.010	0.07	0.267	0.057
YSM	(3.94)	0.490 *	0.093	5.56	0.093	0.020
YTA	1.35	0.120	0.023	1.18	0.026	0.005
YTB	0.31	-0.465 *	-0.088	0.73	-0.271 *	-0.057
Intercept		-5.586			-5.055	
Sample size		635		733		
Likelihood χ^2 statistic ^c		66 *		52 *		
Pseudo R^2 statistic ^d		0.048		0.030		
Separated						
AGE	29.20	0.029	0.005	29.84	0.142 *	0.030
IMM	0.03	-1.543 *	-0.264	0.06	0.679	0.145
YEXP	0.28	0.059	0.010	0.96	0.440	0.094
YFT	0.64	0.170	0.029	0.73	-0.163	-0.035
YINV	0.13	0.099	0.017	0.09	0.143	0.031
YSM	3.78	0.339 *	0.058	5.37	0.120	0.026
YTA	1.29	0.029	0.005	1.17	0.070	0.015
YTB	0.29	-0.210	-0.036	0.62	-0.304 *	-0.065
Intercept		-0.931			-4.337	
Sample size		489		902		
LIV		0.77		0.67		
Likelihood ratio χ^2 statistic ^c		21 *		68 *		
Pseudo R^2 statistic ^d		0.022		0.031		
Single						
AGE	28.53	0.102 *	0.024	28.36	0.162 *	0.040
IMM	0.05	-0.131	-0.031	0.06	0.373 *	0.093
YEXP	0.34	0.208	0.049	1.08	0.724 *	0.181
YFT	0.69	0.052	0.012	0.86	-0.067 *	-0.017
YINV	0.13	-0.322 *	-0.076	0.15	0.031	0.008
YSM	3.81	0.218 *	0.052	5.05	0.038	0.010
YTA	1.27	0.175 *	0.042	0.99	0.118 *	0.029
YTB	0.28	-0.336 *	-0.080	0.67	-0.217 *	-0.054
Intercept		-3.484			-5.443	
Sample size		6010		8707		
LIV		0.60		0.52		
Likelihood ratio χ^2 statistic ^c		356 *		647 *		
Pseudo R^2 statistic ^d		0.022		0.026		

* Test is significant at 95% level. For logistic coefficients the test statistic is that for the Wald test for which the null hypothesis is that the corresponding population coefficient is zero.

^a Definitions of the variables are provided in table 5.

^b Calculated at variable means.

^c The number of degrees of freedom in the number of slope coefficients, and the null hypothesis is that the slope coefficients are simultaneously all zero.

^d Calculated as per Maddala (1983, page 40).

^e Because of small sample sizes, results are not reported for widowed women ($n = 15$) and widowed men ($n = 6$).

lone-parent pools (table 8) and the age of the husband, AGEH, and AGED are not statistically significant either (table 7). Family size, FSIZE, and presence of children under the age of 6 years, DCH6, are each statistically significant in only one of the logistic regressions (table 7). The coefficient of IMM is variable: significantly negative for husband-wife families and for separated women; significantly positive for single men. In summary, for husband-wife families, only FSIZE, DCH6, and IMM among the social and demographic variables ever have statistically significant coefficients. The propensity to maintain one's own dwelling increases with size of family, and decreases with the presence of young children or where either spouse had recently immigrated to Canada. Among lone-parent families, none of the social and demographic variables are statistically significant. Among nonfamily adults, the coefficient of AGE is generally significant (and positive); the results for IMM vary.

Overall fit of the model

Maddala's pseudo R^2 statistic provides a measure of overall fit (again, see tables 7–9). The pseudo R^2 statistic is 10% for husband-wife families. It is better for the lone-parent samples, even with one fewer explanatory variables: 12% for male-led families and 32% for female-led lone-parent families. However, among nonfamily adults, the pseudo R^2 statistics are only 2%–6%; smaller than found for husband-wife families.

In spite of this, the goodness-of-fit values reported here are better than, or at least similar to, the findings from other studies that also use disaggregated cross-sectional data. Haurin et al (1993), although not reporting goodness of fit directly, appear to find pseudo R^2 statistics that may be in the range of only 0.5%–3%.⁽¹¹⁾ Börsch-Supan (1986, page 147) also fails to report a comparable goodness-of-fit measure but does comment that “estimation ... tends to be unsatisfactory because of the high noise-to-signal ratio”. Blank and Rosen (1990) report adjusted R^2 statistics that, although not directly comparable, are numerically larger than the pseudo R^2 statistics reported in this paper. However, their data are aggregated to the city level, and hence hide individual variations in living arrangement. Also, important contributors to the R^2 statistic are the city-specific dummy variables and a time trend. Blank and Rosen conclude, in fact, that changes in propensity to maintain a dwelling are not primarily the result of changes in housing markets and incomes. Instead, they assert the importance of broad social and cultural developments that they feel are being proxied by the time trend. That is similar to my finding that affordability, although often statistically significant, accounts for only a small proportion of the variation in propensity to maintain a dwelling. Taken together, these findings reinforce the conclusion that affordability plays only a modest role in explaining choice of living arrangement.

However, we need to be mindful here of multicollinearity. Casual examination of the correlation matrixes of the logistic coefficients suggests that multicollinearity is generally not a problem in these estimations; most of these correlations are under 0.25 in absolute value. However, there are some notable exceptions. The strongest correlations consistently are between the coefficients of YSM and YFT; YSM is higher, and YFT is simultaneously zero, when consumers work full time. Many pools display a strong positive correlation between the coefficients of AGE (or AGEH) and YSM [and between AGE (or AGEH) and YEXP] because YSM increases (and YEXP decreases) steadily in conjunction with age.⁽¹²⁾ There is also evidence of a correlation between YSM and YEXP that may be a result of this.

⁽¹¹⁾ Haurin et al (1993, page 289) report log-likelihoods of –1984.9 to –1995.7 on a sample size of 2573.

⁽¹²⁾ Haurin et al (1993; 1994) also show evidence of such multicollinearity in that the effect of age is negative in some estimated household-formation equations and is positive in others.

These results indicate the limitation of the estimation. First, although the various measures of income and income prospects and the social and demographic variables do help improve the predictability of living arrangement, much remains unexplained. Whatever causes the variation in living arrangement is explained only in small part by the variables employed in this study. Second, the effect of affordability on living arrangement may be obscured by the strong correlation between YSM and AGE (or AGEH) in these samples of young adults.

Other pools

Among individuals in the age-group 35–60 years, the correlation between YSM and AGE (or AGEH) is much smaller. Expected income increases only modestly past the age of 35 years and begins to decline for most groups of persons by their 50s. Can the effect of affordability be more clearly discerned in this group? In this paper, the model described by equation (5) has been estimated for each of 44 pools of consumers. Selected empirical results are summarized in table 10 (see over). Displayed there are the statistical significance of age and YSM, their estimated probability partial derivatives, sample size, and pseudo R^2 statistics.

Consider the pools in table 10(a). In general, the YSM and AGE or AGEH probability partial derivatives are smaller for the 35–60 years age-group, above-cutoff pools, than for the corresponding 25–34 years age-group pools. This suggests that living arrangement is even less sensitive to affordability than was indicated in tables 7–9.

Given the confounding effect of multicollinearity between YSM and AGE (or AGEH) in the 25–34 years age-group, do the results for the older age-group clarify consumer behaviour in the younger group? To evaluate this argument, the model, represented by equation (5), was reestimated pooling the two age-groups. In all, 22 logistic regressions were estimated; selected results from each regression are presented in table 11. The age-aggregated results yield probability partial derivatives for YSM reported in table 11 generally in between the estimates for 25–34-year olds and 35–60-year olds reported in table 10. However, for the two above-cutoff pools (single women and separated men) that, in table 3, show the largest drop in propensity to maintain a dwelling between 1981 and 1986, the YSM derivatives in table 11 are higher than those obtained for either age-group pool individually in table 10.

Otherwise, the base runs presented in table 11 are broadly similar to those reported for the age-disaggregated pools. The null hypothesis that the slope coefficients of the two age-groups are the same can be tested by using the likelihood ratio statistic as presented in table 11. The likelihood ratio test indicates whether separate estimation for the two age-groups (table 10) differs from the age-combined base runs (table 11). If this χ^2 statistic is large, we can reject the hypothesis that behaviour in the two age-groups is generated by the same model. By this test, the null hypothesis is rejected in only 11 of the 22 pools.

These findings suggest that the model estimates which are shown in table 11, wherein the multicollinearity between AGE (or AGEH) and YSM is less problematic, provide useful estimates of the responsiveness of living arrangements of young adults to the affordability of accommodation. To repeat, these estimates substantiate the principal conclusions reported above with respect to tables 7–10. In table 11(a) the coefficient of YSM is everywhere positive, statistically significant for all pools but one, but the elasticity is generally small (and lower among husband–wife families compared with other pools).

Table 10. Estimated probability partial derivatives calculated at variable means and statistical significance for AGE (or AGEH) and YSM, sample size, and goodness of fit, by pool, in Canadian metropolitan areas, 1981–86.

Pool	25–34 years age-group				35–60 years age-group			
	AGE or AGEH	YSM	sample size	pseudo R^2	AGE or AGEH	YSM	sample size	pseudo R^2
(a) Consumers above low-income cutoff								
Family								
husband–wife	–0.000	0.007*	11 810	0.101	0.000	0.003*	27 652	0.071
female lone-parent	0.009	0.056*	646	0.320	0.000	0.017*	2 268	0.085
male lone-parent	0.002	0.131*	133	0.116	0.003*	0.012	774	0.070
Nonfamily person								
divorced woman	0.027*	0.093*	635	0.048	0.005*	0.013	1 534	0.027
separated woman	0.005	0.058*	489	0.022	0.006*	0.032*	632	0.033
single woman	0.024*	0.052*	6 010	0.022	0.002	0.040*	3 485	0.012
widowed woman	–	–	15	–	0.003	0.026*	913	0.028
divorced man	0.031*	0.020	733	0.030	0.003	0.033*	1 887	0.021
separated man	0.030*	0.026	902	0.031	0.006*	0.022*	1 467	0.019
single man	0.040*	0.010	8 707	0.026	0.008*	0.032*	3 682	0.013
widowed man	–	–	6	–	–0.001	0.006	260	0.057
(b) Consumers below low-income cutoff								
Family								
husband–wife	0.000	0.009*	1 042	0.085	0.000	0.003*	2 221	0.031
female lone-parent	–0.001	0.001	1 050	0.037	0.000	0.000	1 466	0.013
male lone-parent	–	–	35	–	0.001	0.046	128	0.215
Nonfamily person								
divorced woman	0.026*	–0.008	255	0.027	0.007*	0.038	840	0.013
separated woman	–0.003	0.081	238	0.029	0.012*	0.096*	464	0.040
single woman	0.021*	0.118*	2 457	0.014	0.006*	0.094*	1 215	0.011
widowed woman	–	–	17	–	0.003	0.027	767	0.031
divorced man	–0.023	0.127	227	0.021	0.014*	0.069*	668	0.018
separated man	0.030	0.001	228	0.034	0.011*	0.036	445	0.011
single man	0.020*	0.074*	3 818	0.019	0.009*	0.037*	1 536	0.008
widowed man	–	–	7	–	0.002	0.035	123	0.031

Note: for consumers above the low-income cutoff and in the age-group 25–34 years data are selected logistic regression estimates for the 11 pools of tables 7–9; in the remainder of the table equivalent results for the other 33 pools are presented; not shown here are estimated probability partial derivatives for IMM, YEXP, YFT, YINV, YTB, YTA (above-cutoff pools only), DCH6 (family pools only), FSIZE (family pools only), and AGED (husband–wife pools only); variables are defined in table 5.

* Statistically significant at 95% level.

^a Calculated as per Maddala (1983, page 40).

Alternative model specifications

In this section, three criticisms of the base-run estimates are considered. In each case, it is argued here that there is at best little empirical support for the alternative over the base-run estimate, and that the base-run estimates of table 11 are superior in some respects.

Separate price and income effects (PM model)

One criticism might be that for the base-run estimates it was improperly presumed that the relevant price and income elasticities are equal in absolute value (though opposite in sign). To evaluate this, the model was reestimated to include annual shelter price (denoted P^m , and measured in 1980 Canadian dollars) and annual

Table 11. Estimated probability partial derivatives (calculated at variable means) and statistical significance for selected variables, the sample size, and goodness of fit, by age-aggregated pool, in Canadian metropolitan areas, base runs, 1981–86.

Pool	AGE or AGEH	YEXP	YFT	YSM	YTB	Sample size	χ^2 ^a
(a) Consumers above low-income cutoff							
Family							
husband–wife	0.000	0.000	0.002*	0.004*	–0.003*	38 420	31.5*
female lone-parent	0.001*	–0.005	0.004	0.027*	–0.034*	2 914	12.3
male lone-parent	0.004*	0.002	0.002	0.017*	–0.015*	907	10.3
Nonfamily person							
divorced woman	0.009*	0.074*	0.022*	0.037*	–0.044*	2 169	25.1*
separated woman	0.006*	0.010	0.021	0.040*	–0.013	1 121	4.4
single woman	0.007*	–0.006	0.023*	0.061*	–0.053*	9 495	78.0*
widowed woman	0.003*	0.003	0.009	0.025*	–0.015	928	15.6*
divorced man	0.007*	0.045	0.009	0.038*	–0.031*	2 620	19.2*
separated man	0.049*	0.069*	–0.005	0.032*	–0.036*	2 369	15.3
single man	0.016*	0.114*	0.015*	0.028*	–0.041*	12 389	98.9*
widowed man	0.003	–0.001	–0.024*	0.014	–0.016	266	11.2
(b) Consumers below low-income cutoff							
Family							
husband–wife	0.000	0.005	0.001	0.006*	–0.003*	3 263	11.3
female lone-parent	0.000	–0.014	0.002	–0.001	–0.002	2 517	6.2
male lone-parent	0.000	–0.005	0.006	0.063*	–0.030*	163	19.1*
Nonfamily person							
divorced woman	0.011*	–0.052	0.016	0.032	0.007	1 095	18.5*
separated woman	0.011*	0.044	0.067*	0.098*	–0.001	702	9.3
single woman	0.008*	0.002	0.046*	0.117*	–0.058*	3 672	17.7*
widowed woman	0.004	0.089	0.036	0.018	0.017	784	21.6*
divorced man	0.014*	0.112*	0.012	0.066*	–0.040	895	9.3
separated man	0.014	0.121	–0.003*	0.040	–0.002*	673	6.1
single man	0.013*	0.092*	0.024*	0.060*	–0.030*	5 354	38.6*
widowed man	–0.001	–0.178	0.051	0.064	–0.023	130	10.6

Note: not reported here are estimated probability partial derivatives for IMM, YINV, YTA (above-cutoff pools only), DCH6 (family pools only), FSIZE (family pools only), and AGED (husband–wife pools only); variables are defined in table 5.

* Statistically significant at 95% level.

^a The number of degrees of freedom is the number of slope coefficients. The likelihood ratio tests the null hypothesis that the slope coefficients of the two age-group submodels (25–34 years, and 35–60 years) are identical (see Griffiths et al, 1993, page 750).

income variables now measured in 1980 Canadian dollars (denoted $\tilde{Y}$). For 1986 data, the 1985 dollar amounts are deflated to 1980 dollars by using the overall consumer price index. We thus obtain:

$$p_i = \frac{\exp z_i'}{1 + \exp z_i'}, \quad (6)$$

where

$$z_i' = \alpha_0 + \beta X_i + \gamma_1 \tilde{Y}_i^z + \gamma_2 \tilde{Y}_i^c + \gamma_3 \tilde{Y}_i^a + \gamma_4 \tilde{Y}_i^b + \gamma_5 \tilde{Y}_i^5 + \gamma_6 \tilde{Y}_i^f + \delta_1 P_i^m.$$

The model described by equation (6), hereinafter called the PM model, gives results similar to those obtained in the base runs. If L^{PM} is the maximum likelihood in the PM model and L^* the maximum likelihood in the base run, model fit can be

calculated as

$$\Delta L = -2(\ln L^* - \ln L^{PM}), \quad (7)$$

where ΔL , shown in table 12, is an indicator of how much better the PM model performs compared with the base-run model. If the base-run model had been nested within the PM model, ΔL would be the corresponding likelihood ratio χ^2 statistic. The estimated coefficient for P^m itself, although not shown here, is negative for 10 of the above-cutoff pools (although statistically significant in only 4), and significantly positive in the other pool (nonfamily widowed men).

Table 12. Likelihood ratios (ΔL) for split by year and type of income, compared with base runs.

Pool	PM model ^a		YT model ^b	PL model ^c	
	above cutoff	below cutoff	above cutoff	above cutoff	below cutoff
Family					
husband-wife	14	1	128*	18	14
male lone-parent	2	0	7	7	-
female lone-parent	16	0	13*	10	12
Nonfamily person					
divorced woman	1	0	7	9	6
separated woman	1	-1	9*	11	9
single woman	103	22	128*	59*	17*
widowed woman	2	5	3	7	24*
divorced man	10	-2	13*	5	4
separated man	5	2	7	21*	7
single man	157	-3	57*	26*	31*
widowed man	9	0	2	20*	17*

Note: the χ^2 likelihood ratio test is described in Train (1986, pp 51–53).

- Sample size too small for estimation.

* Statistically significant at a 95% confidence level.

^a Because the base run is not nested within the PM model, ΔL is not a χ^2 statistic, and hence statistical significance cannot be assessed. However, ΔL does measure improvement in model fit in the PM model vis-à-vis the base run—a negative value ΔL means that the PM model performs worse than the base run;

^b In YT model versus the base-run model, the χ^2 statistic has 3 degrees of freedom. The test cannot be applied to below-cutoff pools because YTA is in this case excluded in the base-run estimation;

^c In PL model versus the base-run model, the χ^2 statistic has degrees of freedom equal to the number of independent variables in the model described by equation (5).

In many pools, the PM model does not perform substantially better than the simpler base run (ΔL is positive but small) and in other cases actually performs worse (ΔL is negative). The PM model performs best among nonfamily single men and women in the case studied here. Among singles above the low-income cutoff, the probability partial derivative for YSM is almost an order of magnitude larger than in the base-run case (0.248 versus 0.028, respectively, for men, 0.293 versus 0.061, respectively, for women); meanwhile the estimated derivative for P^m is much smaller (−0.016 for men, −0.007 for women).

Mixed results for the effect of housing costs have also been reported in other studies. Blank and Rosen (1990) regress the proportion maintaining a dwelling as a function of average consumer income in the given city in that income group, the city's new house-price index relative to its consumer price index, the city's rental price index relative to its consumer price index, and other variables. Blank and

Rosen (1990) report similar results for the cost of housing by using their three price variables. Sometimes, these variables have significantly negative coefficients as expected; however, in other cases, the coefficients are not significant or are even significantly positive. It can be argued that the dependent variables (living with parent, living alone) employed by Haurin et al (1993) should be more sensitive to affordability than is the dependent variable (maintaining a dwelling) used in this study. Nonetheless, after aggregating different marital statuses, Haurin et al do find their sole rent variable has an effect that is negative, though not always statistically significant. However, neither study measures the improvement in goodness of fit similar to the comparison here of PM model and base-run model.

The effect of disaggregation of income (YT model)

Is the disaggregation of consumer income into four components (that is, Y_i^c , Y_i^a , Y_i^b , and Y_i^z), as is done in the base runs, unnecessarily complex? Could these be replaced simply by Y_i , as is done, for example, by Börsch-Supan (1986, page 152). The base runs of table 11 were reestimated by means of the following model (hereinafter called the YT model):

$$p_i = \frac{\exp z_i''}{1 + \exp z_i''}, \quad \left. \begin{array}{l} \\ \text{where} \\ z_i'' = \alpha_0 + \beta X_i + \gamma_1 Y_i + \gamma_2 Y_i^5 + \gamma_3 Y_i^f. \end{array} \right\} \quad (8)$$

In the above-cutoff pools, the YT model is nested with the corresponding base run. This is the χ^2 likelihood ratio test proposed in Train (1986, pages 51–53) with 3 degrees of freedom. The test cannot be applied in the below-cutoff pools as YTA is excluded in the estimation of equation (5). Using a likelihood ratio test, one can evaluate the hypothesis that each of the components of income in the base run have the same marginal effect on living arrangement. Letting L^{YT} be the likelihood obtained in the YT model, the likelihood ratio test (ΔL) shown in the middle column of table 12 is calculated as follows:

$$\Delta L = -2(\ln L^{YT} - \ln L^*). \quad (9)$$

The hypothesis that the coefficients of the four components are identical is rejected in 6 of the 11 pools. Although not displayed here, the probability partial derivative for Y_i is generally only about one half the estimate for YSM.

The YT and base-run model estimates differ the most in the nonfamily singles pools. Here, the probability partial derivative for YSM is almost an order of magnitude larger than it is in the base-run case (0.248 versus 0.028, respectively, for men, 0.293 versus 0.035, respectively, for women); meanwhile, the estimated derivative for P^m is much smaller (−0.016 for men, −0.007 for women).

The effect of pooling 1981 and 1986 data (PL model)

A third criticism might be that it is inappropriate to pool 1981 and 1986 data, if the model coefficients have changed over time. To test this hypothesis, the base-run model of equation (5) was reestimated with only the 1981 sample data, then again with only the 1986 sample data. If L^{1981} and L^{1986} are the maximum likelihoods obtained by using the 1981 and the 1986 samples alone, respectively, then the following likelihood ratio test can be estimated [this χ^2 likelihood ratio test has degrees of freedom equal to the number of independent variables in the model represented by equation (5)]:

$$\Delta L = -2(\ln L^* - \ln L^{1981} - \ln L^{1986}). \quad (10)$$

The test results are presented in table 12. In most pools, the differences in model fit between the 1981 and 1986 samples were small. The null hypothesis that the slope coefficients were the same for the two years could be rejected in only 8 of the 22 pools.

Conclusions

Let us now return to the puzzle with which this paper began. From 1981 to 1986, the net formation of households in Canada slipped to 142 000 units annually, down from the 225 000 units averaged from 1971 to 1981. To be sure, there were accompanying changes in the demography of Canada's population, but a simple measure suggests that the formation of 'potential households' slowed only modestly. Why, then, the slump in household formation?

In this paper I have explored the possibility that a worsening in the affordability of separate accommodation contributed to the slump in household formation. Model estimates have been presented, based on cross-sectional evidence for 1981 and 1986, which show that income and the price of rental accommodation do play a role in determining whether consumers will maintain a dwelling of their own. The evidence is strongest for single men and single women. They had relatively large income elasticities and, for them, the incidence of maintaining their own dwelling dropped the most between 1981 and 1986.

At the same time, evidence suggests that, overall, the impact of reduced affordability may not have been large. The goodness of fit of the estimated models is not high, particularly for husband-wife families and nonfamily adults. Further, the estimated net change in household maintainers that could be attributed to worsening affordability does not bear a close relationship to changes in the number of households formed.

Why is the evidence on the effects of diminished affordability not stronger? Five arguments come to mind that suggest areas for future research.

Living arrangement and housing consumption: the impact of reduced affordability may be obscured in this study by covariation of living arrangement and the quality or quantity of housing services consumed. Consumers may adjust to reduced affordability by opting for smaller or lower-quality separate accommodation rather than by opting for lodging. Simultaneous modeling of living arrangement and housing consumption might better capture the effect of diminished affordability.

Maintaining versus sharing: in the data sets used in this study, each household contains at most one maintainer. Where one consumer rents a room within a dwelling, one might reasonably expect the landlord to be enumerated as the maintainer. However, where consumers share accommodation, as in a three-generation household, the arbitrariness of the maintainer designation becomes problematic.

Advantages of maintaining a dwelling: the advantages of maintaining one's own dwelling have been presented in this paper mainly as the advantages of not sharing accommodation. However, shared dwellings also have maintainers. Put differently, maintainers can opt to share their dwellings; lodgers cannot. This suggests the need to rethink how one characterizes the advantages and disadvantages of maintaining a dwelling.

The price of rental accommodation: rental accommodation varies considerably in location and quality. The assumption underlying this paper is that all accommodation in a given metropolitan area rents for the same price per standard unit of housing services. Whether this kind of equilibrium exists across submarkets within each metropolitan area is an open question. Such an equilibrium is, however, a critical assumption in this paper as the advantage of shared accommodation is

enhanced if rents happen to increase less than proportionately with size of dwelling (or more generally the flow of housing services). One also needs to consider explicitly the effect of subsidization—by governments, employers, relatives, and others—on living arrangements.

Renting versus owning: in a similar vein, in this study I have ignored the effect of differences in the prices of owner and rented accommodation between cities. Implicit in this paper is an assumption that local markets for housing are in an equilibrium whereby the rent for a rented dwelling bears a fixed relationship to the market price of owner-occupied housing. Such an assumption may be called into question given the existence of rent regulation in several Canadian provinces. Also, it can be argued that the real user cost of homeownership rose sharply in the early 1980s as real mortgage rates escalated. This may have slowed the transition to ownership among young adults and extended the period of their life that is typically spent renting with unrelated persons. Implicit in this argument is an assumption that young consumers widely 'double up' in their living arrangements in order to save for homeownership.

Acknowledgements. Financial support of the Social Sciences and Humanities Research Council of Canada, grants 410-89-0344 and 410-91-0228, is gratefully acknowledged, as is the research assistance of A Suchy, L Lam, and J Gibbs.

References

- Avery R, Goldscheider F, Speare A, 1992, "Feathered nest/gilded cage: parental income and leaving home in the transition to adulthood" *Demography* **29** 375–388
- Beresford J C, Rivlin A M, 1966, "Privacy, poverty, and old age" *Demography* **3** 247–258
- Blank R M, Rosen H S, 1990, "Recent trends in housing conditions among the urban poor" *Research in Urban Economics* **8** 15–34
- Blau D M, Robins P K, 1991, "Child care demand and labor supply of young mothers over time" *Demography* **28** 333–351
- Börsch-Supan A, 1986, "Household formation, housing prices, and public policy impacts" *Journal of Public Economics* **30** 145–164
- Ermisch J F, 1981, "An economic theory of household formation: theory and evidence from the general household survey" *Scottish Journal of Political Economy* **28**(1) 1–19
- Goodman A C, 1990, "Demographics of individual housing demand" *Regional Science and Urban Economics* **20** 83–102
- Griffiths W E, Hill R C, Judge G G, 1993 *Learning and Practicing Econometrics* (John Wiley, New York)
- Haurin D R, Hendershott P H, Kim D, 1993, "The impact of real rents and wages on household formation" *Review of Economics and Statistics* **75** 284–293
- Haurin D R, Hendershott P H, Kim D, 1994, "Housing decisions of American youth" *Journal of Urban Economics* **35** 28–45
- Hickman B G, 1974, "What became of the building cycle?", in *Nations and Households in Economic Growth* Eds P A David, M W Reder (Academic Press, New York) pp 291–314
- McElroy M B, 1985, "The joint determination of household membership and market work: the case of young men" *Journal of Labour Economics* **3** 293–316
- Maddala G S, 1983 *Limited Dependent and Qualitative Variables in Econometrics* (Cambridge University Press, Cambridge)
- Miron J R, 1988 *Housing in Postwar Canada* (McGill–Queen's University Press, Montreal)
- Miron J R, 1989, "Household formation, affordability, and housing policy" *Population Research and Policy Review* **8**(1) 55–77
- Miron J R, 1993, "Demographic and economic factors in housing demand", in *House, Home, and Community: Progress in Housing Canadians: 1945–1986* Eds J R Miron, L S Bourne, G Fallis, A Skaburskis, M Steele, P A Streich (McGill–Queen's University Press, Montreal) pp 22–40
- Smith L B, 1984, "Household headship rates, household formation, and housing demand in Canada" *Land Economics* **60**(2) 180–188

Statistics Canada, various years a *Consumer Prices and Price Indexes* (Statistics Canada, Ottawa)

Statistics Canada, various years b *National Income and Expenditure Accounts: Annual Estimates* (Statistics Canada, Ottawa)

Train K, 1986 *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand* (MIT Press, Cambridge, MA)