



# A new investigation into whether households are excessively sensitive to predictable changes in income

Charles Grant<sup>1</sup>

Received: 2 October 2023 / Accepted: 26 June 2024 / Published online: 15 July 2024  
© The Author(s) 2024

## Abstract

Several tests of the permanent income hypothesis have examined whether predictable changes in income affect current consumption growth. The predictable income change is constructed using past variables as instruments in a two-stage regression. However, finding instruments that are truly exogenous but nevertheless predict income growth is very difficult. Moreover, identification requires a large number of time periods. This paper, using the CEX, shows that previous studies have given insufficient attention to properly testing the instruments. Using a longer number of time periods, and a valid set of instruments, the paper finds that households are indeed excessively sensitive to predictable changes in income.

**Keywords** Consumption · Excess sensitivity · Testing Instruments

**JEL Classification** D12 · D15

## 1 Introduction

This paper will investigate the consumption response of US households when their income changes. In common with the already extensive literature on this problem, it will concentrate on the response of households to predictable changes in income where changes in income are predicted using past information as instruments. In simple versions of the permanent income hypothesis (PIH), households should not react to predictable changes in income: this is known as the excess sensitivity test. The results depend crucially on the choice of instruments used to predict the change in income. In this paper we argue the previous literature has failed to adequately test the choice of instruments, and this failure has seriously affected the conclusions that this earlier literature has reached.

The author(s) states he/she has no relevant or material financial interests that relate to the research described in this paper.

✉ Charles Grant  
charles.grant@brunel.ac.uk

<sup>1</sup> Department of Economics and Finance, Brunel University, Uxbridge, London UB8 3PH, UK

While the first papers to test for excess sensitivity, such as Flavin (1981), used national account data, since at least the early 1990s Attanasio and Weber (1995) have shown the importance of using household data to study the response of households to income changes. Altonji and Siow (1987), using the PSID, showed that predictable changes in income did not predict current consumption growth, and hence, they did not reject the PIH. Similarly, Attanasio and Weber (1995) used the consumer expenditure survey (CEX), arguing that when household data are used, and household-specific characteristics are included in the regression, the excess sensitivity test does not reject the PIH. In contrast, Zeldes (1989) partitioned households in the PSID into a low-asset group (who, he argued, are likely credit-constrained) and a high-asset group and showed that lagged income growth predicted current consumption growth for the low-asset households. Using a more recent sample from the PSID, Fisher et al. (2020) confirmed Zeldes' earlier finding that low-asset households are sensitive to predictable changes in income.

Despite using household data, these papers have all used a fairly short panel of observations. Few years of data were available to the earlier studies, but even the most recent, Fisher et al. (2020), only used eight waves of the PSID. However, in the Euler equation the conditional expectation of the forecast errors (e.g. the expectation of the errors across households in each time period) must be zero. If there are aggregate shocks to the economy, then the average of the errors across individual households will not converge to zero as the number of households becomes large: Jappelli and Pistaferri (2010) explains identification really requires the number of time periods to be large. Even Attanasio and Weber (1995), who take the issue more seriously than other studies, only had quarterly data from the CEX for 1980–1990. This paper will considerably extend the number of time periods used in Attanasio and Weber (1995), and it will show that the short panel in previous studies is a genuine problem that has substantively affected the earlier results.

The previous literature has also failed to devote enough attention to testing the instruments which are used to predict income growth. Altonji and Siow (1987), for example, only reports an  $R^2$ -statistic for the first stage. Attanasio and Weber (1995) test the over-identifying instruments (although the test is likely to have very weak power in their short panel); they also report an  $R^2$ -statistic for the first stage in some of their regressions, but without formally testing for weak instruments. Fisher et al. (2020), although a much more recent paper, does not present any formal test of the instruments. This paper will present much more comprehensive tests of the instruments used to predict consumption, and it will show that the large number of instruments used in some previous studies are likely to have seriously affected the results they report.

The rest of the paper is organized as follows. Section 2 describes the methodology and the current literature which uses the excess sensitivity test to investigate whether households satisfy the Permanent Income Hypothesis. The data used in this study are the CEX, and Sect. 3 will describe this dataset and explain why it is preferred to using the PSID. The paper will report results in Sect. 4. The conclusions are drawn in Sect. 5.

## 2 Methodology

The papers which undertake the excess sensitivity test consider the standard problem of a consumer  $i$  who must choose consumption  $c$  in the current period  $t$  such that the consumer maximizes expected utility subject to an inter-temporal budget constraint. If the consumer can borrow and lend at the same interest rate  $r_t$ , then in the consumer's optimal solution, the expectation of marginal utility is held constant. The solution, as is well known (see the summary in Jappelli and Pistaferri, 2010), can be written formally:

$$u'(c_{it-1}) = (1 + \delta)^{-1} E_{t-1}[(1 + r_t)u'(c_{it})] \tag{1}$$

where  $u(\cdot)$  is the utility function and  $\delta$  is the discount rate and  $c_{it}$  is consumption for household  $i$  at time  $t$ . The Permanent Income Hypothesis argues that this relationship holds exactly. To test the PIH, the model can be linearized in the following way (where  $\Delta$  is the first difference).

$$\Delta \ln c_{it} = \sigma \ln(1 + r_t) + \beta \Delta X_{it} + \varepsilon_{it} \tag{2}$$

As Attanasio and Weber (1995) explain, this formulation can be derived by log-linearizing the first-order condition from an inter-temporally separable optimization problem with isoelastic preferences. (Some papers, such as Flavin, 1981, estimate a model in where the change in consumption is in levels rather than log-levels: such a relationship can be motivated by assuming quadratic preferences.) In this formulation  $X_{it}$  represent 'taste-shifters' (such as family composition) which shift the marginal utility of consumption. The innovation in consumption  $\varepsilon_{it}$  is the change in consumption between time  $t - 1$  and time  $t$  that is not predictable given tastes  $X_{it}$  or the interest rate or any other information which is available at time  $t - 1$ . This innovation in consumption  $\varepsilon_{it}$  is treated as an error term in regression models which are based on estimating Eq. 2.

An implication of the Euler equation framework is that variables known at time  $t - 1$  should not affect the current change in consumption  $\Delta \ln c_{it}$ . Many papers have tested the PIH by adding predictable changes in income to Eq. 2 so that the regression model that they report is based on the following equation.

$$\Delta \ln c_{it} = \alpha \Delta \ln y_{it}^p + \sigma \ln(1 + r_t) + \beta \Delta X_{it} + \varepsilon_{it} \tag{3}$$

In this framework  $\Delta \ln y_{it}^p$  is the predictable change in income where  $y_{it}$  is the level of income for household  $i$  in time  $t$  and the superscript  $p$  is used to emphasize that it is the predictable change in income which is included in this regression (where it is predictable based in information at  $t - 1$  or earlier). The coefficient  $\alpha$  can be interpreted as the marginal propensity to consume from changes in predictable income. According to the basic version of the Permanent Income Hypothesis, households should only react to unexpected changes in their income. Hence, a test for the PIH is a test for whether  $\alpha$  equals zero.

One approach is to argue that there was a change in income which was fully predicted by the household. For example, Parker (1999) used changes in social security pay-

ments, while Souleles (1999) and Johnson, Parker, and Johnson et al. (2006) looked at federal tax rebates. All three studies found that they could reject that  $\alpha$  equals zero in their regressions, at least for low-asset households. In contrast, Altonji and Siow (1987) and Attanasio and Weber (1995) used past values of income (or income growth) and other variables as instruments to predict the current change in income. Both papers strongly argued that when using appropriate instruments, and including a set of household characteristics  $X_{it}$  to reflect changes in tastes, then the coefficient  $\alpha$  was not identically different from zero. Hence, they argued that households followed the PIH when making their consumption decisions. The fact that different approaches to identifying predictable income changes have led to different conclusions about whether the PIH holds has been noted by Commault (2022).

Jappelli and Pistaferri (2010) argue that when instruments are used to predict income growth, the choice of instruments is crucial. The validity of the instruments needs to be thoroughly tested. In this paper we argue the previous literature has failed to adequately test the validity of the instruments used to predict income, and that this failure has seriously affected the results. For example, Altonji and Siow (1987) only report the  $R$ -squared values for the first-stage regression of the validity of the instruments (the rank test), but no Sargan test. Nevertheless, the implied value of the  $F$ -statistic for the first-stage regression reported in their Table 3, given the  $R$ -squared, is 12.60. This is not only statistically significant, but it is also above the rule-of-thumb value of 10, below which the instruments are thought to be weak. Nothing can be said about whether their instrument set passes the Sargan test. Attanasio and Weber (1995) report a Sargan test. However, this test is known to have weak power to reject the over-identifying instruments when the sample size is small and the number of instruments is large, which is the exactly the situation that arises in their paper. Their small sample size and large number of instruments are also likely to result in a low value for the rank test; they report an  $R$ -squared for their instruments for some of their regressions. Their first-stage  $R$ -squared for income growth is 0.24, which suggests they have a weak instrument problem. (The  $R$ -squared for the first-stage regression predicting the interest rate is much higher.) More surprisingly, Fisher et al. (2020), although a more recent paper, reports neither the rank test nor the Sargan test. The omission of these test statistics makes it difficult to judge the perspicacity of their results.

A further criticism of estimates of the marginal propensity to consume is discussed in Jappelli and Pistaferri (2010). They note the PIH implies that the conditional expectation of the forecast errors must be zero in Eq. 3. This implies that  $E_{t-1}(\varepsilon_{it})$  is zero. The empirical analog of this expectation is that the average holds over a large number of time periods rather than over a large number of households; there is no guarantee that the cross-sectional average of forecast errors will converge to zero as the number of households becomes large. This won't happen, for example, if the economy experiences aggregate, time-specific shocks. Empirical papers have typically used rather short time periods: for example, the sample used by Altonji and Siow (1987) includes 14 years of data, while Zeldes (1989) uses only six time periods. This problem is sometimes handled by including time dummies in the Euler equation, but as Jappelli and Pistaferri (2010) explain, time dummies do not solve the problem if the aggregate shock is distributed unevenly in the population.

This paper will address some of the concerns about the choice of instruments, and the short number of time periods included in previous studies of the Euler equation. Early studies using household data were limited in the number of time periods they could include, but by now data for both the CEX and the PSID surveys have been continuously collected for many years; for the CEX, for instance, around forty years of data are publicly available. This allows this paper to use a much longer time-series of observations than previous studies. The fact that more time periods can be included in the analysis also means the criticism that the Sargan test in Attanasio and Weber (1995) has weak power can be addressed. This paper will make a much more thorough analysis of the choice of instruments, showing that in the earlier literature, poor instruments are likely to have been a serious problem and to have substantively affected their conclusions. This paper will show when a smaller set of instruments are used in the first-stage regression to construct the predictable change in consumption, it is possible to obtain good estimates of the marginal propensity to consume. To make the analysis as transparent as possible, the paper will follow a similar approach to Attanasio and Weber (1995); this will ensure that discussion is centred on the choice of instruments and the longer time period, rather than other particularities of the regressions.

### 3 Data

The importance of using household-level data to investigate the PIH is widely understood. Estimates of marginal propensity to consume that use US household survey data have either used the consumer expenditure survey (CEX) or they have used the PSID: this paper will use the CEX. The CEX is a survey of US households conducted by the Bureau of Labor Statistics, representative of the US population. It was originally collected with the intention of enabling the US government to construct a measure of household inflation. The survey has been conducted on a continuous basis since 1980 and is designed to provide very detailed information on all aspects of households' expenditure. Around 5000 (or 7500 after 1998) randomly chosen households are interviewed each quarter: they report, in detail, their income, consumption and other particulars of their household circumstances (such as family composition, and the age and education of family members). Each household provides information on the household's spending over four successive quarters, with one-quarter of households dropping out of the survey and replaced each quarter.

An advantage of using the CEX is that it contains very detailed information on almost all categories of spending undertaken by the household and remains the most comprehensive dataset on household spending in the USA. Nevertheless, it seems to have been under-utilized in recent years in studies of household spending. Many US household studies have used the PSID, a survey of households which has operated continuously since 1968, in which the same households are continuously reinterviewed, originally, on an annual basis. Before 1999 the PSID measured food expenditure (from which some researchers have attempted impute broader measure based on household characteristics, despite the obvious problem in using these household characteristics to impute both a broader measure of consumption and at the same time using them in the estimation of the marginal propensity to consume). Since 1999 the PSID uses a

broader measure of consumption than just food spending and switched to surveying the households biannually at this time. If this fuller measure of household spending in the PSID had been utilized, then it necessarily entails a relatively short number of time periods. However, this measure captures only 72% of the expenditure reported in the CEX. Andreski et al. (2014) argue “*the CEX remains the most comprehensive household expenditure survey*” for US households. Given the emphasis this study attaches to having a long time-series of observations for a broad measure of consumption, this paper will use the CEX in the analysis.

The sample of households chosen in this study using the CEX excludes student households, households with multiple non-related adults, and households in which the principle earner is neither the head of household nor their spouse. It will also exclude the youngest and the oldest households from the analysis. The survey includes details on family size, the number of children in the household, the marital status of the household head, their gender, age and education, and information about whether the wife works, and all these variables will form part of the analysis. To aid comparison with the results reported in earlier research, the paper will look at non-durable consumption only. The construction of the measure of non-durable consumption closely follows that in Attanasio and Weber (1995): it includes food, alcohol, tobacco, fuel and utilities, transport services, personal care items, entertainment services, and housing services and maintenance. This measure excludes housing costs, consumer durables, cars, and health spending. The nominal values of household non-durable consumption are deflated by the consumer price index, published by the BLS (where the index equals 100 in 1983). The income measure for each household is the after tax income which includes transfers. The analysis will also include the real interest rate in the regressions. The measure used is the three-month treasury bill, adjusted for the rate of inflation.

A problem with using the CEX to estimate a dynamic model of consumption based on Eq. 3 is that each individual household is only observed for four quarters. However, we can define a synthetic cohort based on characteristics of the household that do not change over time. Since the year-of-birth of the household head is reported in the survey, it is possible to construct cohorts based on the year-of-birth of the head. For each year-of-birth cohort and each quarter, we can define the average of each of the variables included in the regression (where we take the average of log-consumption and log-income). This means that in Eq. 3,  $i$  represents the cohort rather than the individual household. Table 1 reports the choice of year-of-birth cohort used in the analysis. The average number of observations used to construct each cohort-quarter cell is 354. Cohorts are included in the analysis when their average age is between 29 and 65; this means, for instance, those household heads born between 1916 and 1920 are only included when observed between 1981 and 1983, while those household heads born between 1986 and 1990 are included when observed between 2017 and 2019.

**Table 1** Cohort based on year-of-birth

Cohort	Year-of-birth	Years included in sample
1	1916–1920	1981–1983
2	1921–1925	1981–1988
3	1926–1930	1981–1993
4	1931–1935	1981–1998
5	1936–1940	1981–2003
6	1941–1945	1981–2008
7	1946–1950	1981–2013
8	1951–1955	1982–2018
9	1956–1960	1987–2019
10	1961–1965	1992–2019
11	1966–1970	1997–2019
12	1971–1975	2002–2019
13	1976–1980	2007–2019
14	1981–1985	2012–2019
15	1986–1990	2017–2019

## 4 Results

The empirical analysis will start by running regressions which are similar to those in Attanasio and Weber (1995) who tested for excess sensitivity. The intention is not to run identical regressions (which would not be possible because the data sample they used is no longer available), but to show that results which are similar to those reported there can be obtained using our data: by similar results we mean that we can reach the same conclusions about the permanent income hypothesis as were reported by Attanasio and Weber (1995).

The first results are reported in Table 2; these results are for the Euler equation including similar control variables to those used in Attanasio and Weber (1995). The table reports results for 1981–1990 (the time-frame used in Attanasio and Weber, 1995), and for the full time period which extends to 2019. In either case, the change in income is instrumented by age and age-squared; the second and third lag of the number of earners; and the second, third, and fourth lag of income growth, inflation, and consumption growth. The real interest rate is also instrumented, with the second, third, and fourth lag included in the instrument set. These are similar (but not identical) to the instruments used in the regressions reported by Attanasio and Weber (1995); hence, the results should be broadly comparable to those reported in that paper without being an exact match.

The first two regressions reported on the left-hand side of Table 2 do not include the household controls in the regression. Whether using data from only 1981–1990, or using the full dataset, the table shows that the change in income is significant at the 1% level (even if the longer time period has a smaller coefficient). The real interest rate is not statistically different from zero in these regressions (and will never be significant in any of the regressions). The middle-two columns add the change in log-family size

**Table 2** Euler regression with large instrument set

	81–90	All	81–90	All	81–90	All
$\Delta y_{it}$	0.338** (0.123)	0.234** (0.082)	0.239* (0.121)	0.106 (0.082)	0.184 (0.116)	0.069 (0.081)
interest rate	-0.471 (0.328)	-0.031 (0.062)	-0.359 (0.308)	-0.018 (0.059)	-0.312 (0.301)	-0.030 (0.059)
$\Delta \ln(\text{family size})$			0.436** (0.096)	0.495** (0.051)	0.459** (0.167)	0.357** (0.092)
$\Delta$ wife-work					0.249* (0.099)	0.137** (0.044)
$\Delta$ couple					-0.051 (0.185)	0.162 (0.111)
$\Delta$ children					-0.009 (0.151)	0.035 (0.078)
<i>Wald</i>	7.57	8.20	3.89	1.64	2.52	0.73
( <i>p</i> -val)	(0.022)	(0.016)	(0.142)	(0.441)	(0.283)	(0.694)
<i>CLR</i>	6.21	1.83	3.43	0.12	2.07	0.00
( <i>p</i> -val)	(0.016)	(0.176)	(0.076)	(0.735)	(0.170)	(0.965)
Rank test	2.45	4.44	2.27	3.98	2.49	4.32
Under-I test	36.71	68.01	34.45	61.40	37.73	66.59
( <i>p</i> -val)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sargan Test	13.21	44.00	9.04	32.14	8.52	31.46
( <i>p</i> -val)	(0.510)	(0.000)	(0.828)	(0.003)	(0.860)	(0.004)
$R^2$	0.45	0.42	0.52	0.48	0.54	0.48
<i>N</i>	282	1,142	282	1,142	282	1,142

*Instrument Set* age, age-sq,  $r_{t-2}$ ,  $r_{t-3}$ ,  $r_{t-4}$ ,  $\Delta c_{it-2}$ ,  $\Delta c_{it-3}$ ,  $\Delta c_{it-4}$ ,  $\Delta y_{it-2}$ ,  $\Delta y_{it-3}$ ,  $\Delta y_{it-4}$ ,  $\Delta p_{it-2}$ ,  $\Delta p_{it-3}$ ,  $\Delta p_{it-4}$ ,  $earner_{t-2}$ ,  $earner_{t-3}$ .

Standard errors in parenthesis. The instruments include age, age-squared, and the lags of inflation, consumption, income, the number of earners, and the real interest rate  $r$ . Income and consumption are measured in logs. All regressions included quarter dummies; \* denotes significant at the 5% significance level while \*\* denotes significant at the 1% significance level. The Wald test and the CLR test are testing the variable  $\Delta y_{it}$  (*p*-values given in brackets beneath). Although both the CLR test and the AR test allow for weak instruments, the CLR test is preferred. The rank test is the Cragg and Donald minimum eigenvalue, for which Stock and Yogo (2005) have published critical values; in this case the critical value is 10.96. The under-I test is the under-identification test

to the regressions; this variable is highly significant in these two regressions. When the regression includes family size, the change in income is significant at the 5% level when using the 1981–1990 sample, but is not significant in the full sample. The last-two columns of Table 2 add the change in whether the wife works, the change in couple and the change in the number of children to the regressions. When these additional explanatory variables are added to the regression, the change in income is no longer significant whichever time-frame is chosen. Attanasio and Weber (1995) argued that consumption no longer responded to predictable changes in income when the regressions used household-level data and household-level taste shifters are included

in the regression. The results reported in Table 2 also clearly demonstrate this same conclusion.

The results in the table can be criticized on a number of grounds. A relatively minor criticism is that the additional household variables that are included in the regression must, necessarily, be interpreted as ‘taste-shifters’ and not as ‘income-shifters’ (nor a combination of both). If these variables are interpreted as “income-shifters”, then the fact that income growth is no longer significant when the household variables are included (that is, the results in the last four columns of Table 2) cannot be interpreted as confirmation of the permanent income hypothesis: the effect of changes in permanent income will be acting through these household variables and some of these household variables are significant in explaining changes in consumption. Fundamentally, this assumption is not testable, and it is for each reader to decide how to interpret the coefficients on the household variables. Attanasio and Weber (1995) have argued very strongly that they should be interpreted as taste-shifters rather than income-shifters. Nevertheless, there has been considerable debate about the interpretation of the household variables in these types of regressions (see Deaton, 1992 for a discussion); but the argument in favour of interpreting the household variables as taste-shifters has been widely accepted (see Jappelli and Pistaferri 2010, for instance). Moreover, the broad finding that predictable changes in income are not significant when household variables are included still holds when only changes in family size are included in the regression (see column 4 of Table 2), a variable which is rather less unlikely to be an income-shifter than, for example, the wife-work variable.

#### 4.1 Testing for weak instruments

A more serious criticism of the regressions in Table 2 is over the instrument set, and particularly that there is a weak instrument problem. The problems associated with weak instruments only became widely known following Staiger and Stock (1997), and it has since become an extremely active area of research. When weak instruments occur, the standard regression output is likely to under-predict the standard error of the estimates of the variables and thus inference based on the standard regression output is likely to be unsafe. Weak instruments occur when the instruments predict the instrumented variable, but with a low  $R$ -squared value. It is possible to test for whether the instruments are weak. With a single instrumented variable, the  $F$ -statistic for the first-stage regression of the instruments against the instrumented variable can be reported. However, it is not robust to heteroskedastic errors. In the case where there is one variable being instrumented and only one instrumental variable (e.g. the just identified case), then the correction suggested by Kleibergen and Paap (2006) can be used. But in the over-identified case (where there are many instruments), then the full correction suggested by Montiel Olea and Pflueger (2013) should be used. The instruments are weak if the value of the  $F$ -statistic (corrected for non-homoskedasticity) is below the critical value: critical values for a variety of scenarios have been reported by Stock and Yogo (2005), but it is also very common (and simpler) to use the rule-of-thumb critical value of 10 which is known to work reasonably well. When there are several variables being instrumented, then the  $F$ -statistic can be replaced by the Cragg–Donald

statistic (a generalization of the  $F$ -statistic): more accurately, Sanderson and Windmeijer (2016) showed that the Cragg and Donald (1993) minimum eigenvalue rank test statistic should be used. However, no counter-part to the adjustments suggested by Montiel Olea and Pflueger (2013) has yet been suggested in the literature on weak instruments when there is more than one instrumented variable. Nevertheless, Stock and Yogo (2005) have published critical values for this scenario (or the rule-of-thumb value of 10 can be used).

Table 2 reports rank tests for each of the regressions. For instance, in the first column (the regression using only data from 1981–1990 and without the household controls), the Cragg and Donald minimum eigenvalue takes the value of 2.45; a value well below the rule-of-thumb value of 10 (and also below the Stock and Yogo critical value of 10.96). Thus, it is clear that there is a weak instrument problem. In the second column the rank test reports a value of 4.44, again below the rule-of-thumb value of 10. The remaining columns also consistently show low values of the rank test. Hence, we can conclude that the rank test is failed in every column of the table, and that there is a weak instrument problem in each of the regressions reported in Table 2. The tables also report the under-identification test (this is the Anderson LM-test). In each case the instrument set passes the test. Note that when the interest rate is regressed against the instruments, the  $F$ -statistic is 233.14, while when income growth is regressed against the instruments, the  $F$ -statistic is 4.81. This suggests that income growth only is weakly identified.

When a weak instrument problem occurs the  $t$ -test (or the Wald test) for the variable being instrumented no longer has the standard distribution; hence, inference based on this value is unsafe. Nevertheless, results for the Wald test have been reported for each regression in the table, showing the test is no longer significant in columns 3–6 when household controls are included in the regression. In the over-identified case, Andrews et al. (2006) argue that the conditional likelihood ratio test (CLR test) proposed in Moreira (2003) should be used in place of the Wald test as the latter is biased and will typically over-reject the null hypothesis when instruments are weak. (They also discuss the Anderson–Rubin AR test, which could also be used.) Finlay, Finlay et al. (2013) have written a STATA routine called `weakiv` which implements these tests. Results for the CLR test of the coefficient on income growth are reported in Table 2. In the small sample without the household variables, the CLR test shows that income growth is significant at 1% (the coefficient is 6.21 with a  $p$ -value of 0.016). In the second column, where the full data sample is used, the CLR test finds that the coefficient on income growth is not significant at 5%. The third and the fourth column of Table 2 adds the change in family size to the regressions; including this variable means the change in income is no longer significant at the 5% level. In the last column, when the additional household variables are added, the CLR test shows that income growth is not significant.

## 4.2 Testing the over-identifying instruments

A third criticism of the results reported in Table 2 is over the choice of instruments. In the small sample, when data only from 1981–1990 are used, there are a fairly

small number of observations, and there are a large number of instruments. The over-identifying instruments can be tested using a Sargan test. However, the Sargan test is known to have very weak power to reject the null hypothesis in these types of situations (see Bowsher 2002). When the small sample is used (the first, third, and fifth column of results in Table 2), the Sargan test is always passed. However, when the full sample up to 2019 is used (the second, fourth, and sixth column of results), the over-identifying instruments are always rejected. A clear reason why the two datasets give contrasting results is that there are a small number of observations and a large number of instruments in the small sample, which is precisely the situation in which the Sargan test has weak power. Adding observations improves the power of the test and leads to the rejection of the over-identifying instruments; and this is what has happened in the large sample.

Rejecting the over-identifying instruments affects the performance of the CLR test since it widens the confidence bounds for a 95% confidence interval (or for a confidence region of any level of significance which the researcher might choose). In the second column of Table 2, which uses the large data sample and does not include the household variables in the regression, the 95% confidence region around the point estimate includes the 'entire grid' (e.g. all the possible values of the point estimate which could be included in the confidence region). That is, there is no coefficient value when estimating the effect of income growth on consumption growth which could reject the null hypothesis that the coefficient is zero and thus that households do not respond to predictable changes income: the model is not capable of rejecting the permanent income hypothesis. Adding the change in family size to the regression, the regression in column 4 of table 2, leads to exactly the same outcome, since again the 95% confidence interval for the CLR test includes the 'entire grid'. The last column, which includes all the household variables in the regression, also has extremely wide 95% confidence bounds for the CLR test. Rejecting the over-identifying instruments also affects the performance of the AR test; this is an alternative to the CLR test, also discussed by Andrews et al. (2006). Results for the AR test are not reported, but when the over-identifying instruments are rejected, the 95% confidence region is the empty set (and thus, using the AR test, every point estimate will be statistically significant regardless of the value it takes). Overall, the outcome of the Sargan test suggests that the results reported in table 2 are unreliable and that no firm conclusion can be drawn about the permanent income hypothesis based on the regression results reported in this table.

### 4.3 Using a reduced set of instruments

The evidence suggests that the regressions reported in Table 2 have too large an instrument set to be reliable, and the regressions should use fewer instruments. The key question is which of the instruments should be used if some are to be dropped. Obviously, the permanent income hypothesis does not offer specific guidance about the exact choice of instruments that should be included in the regression (other than the fact that it should include lagged variables). So, the researcher must make a decision about what choice should be made. In Table 3 the choice of instruments is restricted

**Table 3** Euler regression with reduced instrument set

	81–90	All	81–90	All	81–90	All
$\Delta y_{it}$	0.364** (0.137)	0.358** (0.092)	0.284* (0.135)	0.248** (0.092)	0.226 (0.130)	0.207* (0.091)
interest rate	-0.108 (0.397)	0.001 (0.066)	-0.089 (0.372)	0.006 (0.061)	-0.064 (0.364)	-0.008 (0.060)
$\Delta \ln(\text{family size})$			0.417** (0.100)	0.441** (0.054)	0.479** (0.169)	0.406** (0.094)
$\Delta$ wife-work					0.248* (0.099)	0.136** (0.044)
$\Delta$ couple					-0.081 (0.190)	0.035 (0.117)
$\Delta$ children					-0.043 (0.156)	-0.003 (0.079)
<i>Wald</i>	7.01	15.21	4.42	7.35	3.04	5.17
( <i>p</i> -val)	(0.030)	(0.000)	(0.109)	(0.025)	(0.218)	(0.075)
<i>CLR</i>	6.53	12.65	4.42	6.70	2.87	4.62
( <i>p</i> -val)	(0.011)	(0.000)	(0.036)	(0.010)	(0.097)	(0.029)
Rank test	3.97	7.39	3.75	6.55	4.01	6.81
Under-I test	20.70	56.84	28.33	50.71	30.37	52.70
( <i>p</i> -val)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sargan Test	5.54	9.15	2.63	3.06	3.14	3.35
( <i>p</i> -val)	(0.475)	(0.165)	(0.853)	(0.801)	(0.790)	(0.762)
$R^2$	0.44	0.38	0.51	0.47	0.53	0.48
<i>N</i>	282	1,142	282	1,142	282	1,142

*Instruments* age, age-sq,  $r_{t-2}$ ,  $r_{t-3}$ ,  $r_{t-4}$ ,  $\Delta y_{it-2}$ ,  $\Delta y_{it-3}$ ,  $\Delta y_{it-4}$ .

Standard errors in parenthesis. The instruments include age, age-squared, and lags of income and the real interest rate  $r$ . Income and consumption are measured in logs. All regressions included quarter dummies; \* denotes significant at the 5% significance level, while \*\* denotes significant at the 1% significance level. The Wald test and the CLR test are testing the variable  $\Delta y_{it}$  (*p*-values given in brackets beneath). Although both the CLR test and the AR test allow for weak instruments, the CLR test is preferred. The rank test is the Cragg and Donald minimum eigenvalue, for which Stock and Yogo (2005) have published critical values; in this case the critical value is 10.22. The under-I test is the under-identification test

to age, age-squared, and the second, third, and fourth lags of both the interest rate and income growth. These lags of income growth and the quadratic in age all significantly predict current income growth in the first-stage regression, which is the reason why this choice was made.

Results for the effect of predictable changes in income on consumption growth using the reduced instrument set are reported in Table 3. Results are reported for the small sample (using data from 1981–1990) and for the full sample. Using the small sample, the results mirror those reported in Table 2. When no household controls are included in the regression, reported in the first column, the *t*-test for income growth is significant at the 1% level. When family size is added to the regression, reported

in the third column, income growth remains significant at the 5% level. But when the full set of household characteristics is included in the regression, the fifth column in the table, income growth is no longer significant. For each of these regressions, the Sargan test is passed. Additionally, the under-identification test is passed and the over-identifying instruments are not rejected. The rank test (the Cragg and Donald minimum eigenvalue) takes a value of 3.97 in the first column, 3.75 in the third column, and 4.01 in the fifth column. In each case this value is well below the rule-of-thumb value of 10 (and also the Stock and Yogo critical value of 10.22). Hence, there is evidence that there is a weak instrument problem. To address this issue, the table reports the CLR test for income growth (which accounts for weak instruments). This test reports that income growth is significant at the 5% level when no household controls are used (the first column), significant at the 5% level when the change in family size is included in the regression (the third column) and significant at the 10% level when all the household variables are included in the regression (the fifth column).

Results for the full sample which includes all the years are also reported in Table 3. For each of the three regressions, the coefficient for income growth is similar to that reported for the small sample (but the larger sample means a smaller standard error). The  $t$ -test for income growth is significant at the 1% in the second column which excludes the household controls; is significant at the 1% level in the fourth column which includes the change in family size; and is significant at the 5% level in the last column, when all the household controls are included in the regression. Unlike in the earlier case, when the reduced instrument is used, the Sargan test is always satisfied and we cannot reject the over-identifying instruments. The under-identification test is always passed. The rank test (the Cragg and Donald minimum eigenvalue) reports a value of 7.39 in the second column, a value of 6.55 in the fourth column, and a value of 6.80 in the last column. In each case the value is below the rule-of-thumb value of 10; hence, each of these regressions has a weak instrument problem. Note that when the interest rate is regressed against the instruments, the F-statistic is 376.86, while when income growth is regressed against the instruments, the F-statistic is 7.50. This again suggests that income growth only is weakly identified. Due to the weak instrument issue, the table reports the CLR test for whether income growth is significant. In the second column, which does not include the household characteristics, the change in income is significant at the 1% level. The fourth column adds the change in family size to the regression, and the CLR test shows that income growth is significant at the 5% level. The last column includes all the household controls, and again income growth is significant at the 5% level.

In Table 3 the choice of instruments has been restricted to age, age-squared, and the second, third, and fourth lags of both the interest rate and income growth. These lags of income growth are significant in the first-stage regression predicting current income growth, as is age-squared (but not age). Nevertheless, other choices of instruments could be made: some brief comments on some other possibilities are appropriate (for brevity, these results are not presented). The reduced instrument set has excluded the lags of inflation and the lags of the number of earners since these variables did not predict income growth in the first stage. Including lagged consumption growth in the instrument set resulted in the Sargan test rejecting the over-identifying instruments regardless of how many lags were included.

Including the second, third, and fourth lags of income growth and the interest rate matches the number of lags used in Attanasio and Weber (1995). Including only the second lag, or only the second and third lag, produced much lower rank values, and hence affected the CLR values. But they did not otherwise substantively affect the results reported.

## 5 Conclusion

Estimates of the marginal propensity to consume of US households have often implemented the excess sensitivity test to examine whether households behave according to the permanent income hypothesis (PIH). Papers such as Altonji and Siow (1987), Zeldes (1989), and Attanasio and Weber (1995) have used econometric methods to identify a predictable component of the change in income and then investigated whether this predictable change in income also predicts changes in consumption: if it does then the permanent income hypothesis can be rejected. In these econometric methods, the predictable income change is constructed by using past variables as instruments in a two-stage regression. However, this literature has been criticized on the grounds that it is difficult to find good instruments that are truly exogenous and also predict current income (see Souleles, 1999 and Jappelli and Pistaferri, 2010). Furthermore, in the presence of aggregate shocks, identification of the marginal propensity to consume out of predictable changes in income requires a large number of time periods; but previous studies that have used household data in their estimates have mostly used rather short time intervals (due to longer series not being available to them). Few papers which have used instruments to test for excess sensitivity have given enough attention to motivating their choice of instruments through the extensive use of formal tests. While the lack of attention to thoroughly testing the choice of instruments is understandable in the earlier literature, it is surprising that a more recent paper, Fisher et al. (2020), also omits formal tests of the instruments.

This paper addresses the criticism made of the earlier literature by using a much larger number of time periods than previous studies and testing the instruments more thoroughly. It has used the CEX, for which there are now nearly forty years of data. (The survey has been operating on a continuous basis since 1980.) To make the results easily comparable to those reported in earlier studies, it has reported results from regressions which are similar (but not identical) to those presented in Attanasio and Weber (1995). Using data for 1981–1990 only, this paper is able to reproduce their key finding: (i) predictable changes in income predict current consumption growth when household variables are excluded from the regressions; (ii) predictable changes in income no longer predict consumption growth when these household variables (interpreted as taste-shifters) are added.

The results reported in Attanasio and Weber (1995) only use data for 1981–1990. (This paper also presents results for this time period.) These estimates also used a large number of instruments to predict income growth. However, the Sargan test is known to perform poorly when there are relatively few observations and a large number of instruments, as it has very weak power to reject the over-identifying instruments in this situation. This paper shows that when using data only for 1981–1990 and a

large number of instruments, the Sargan test is not able to reject the over-identifying instruments. These regressions are similar (but not identical) to those reported by Attanasio and Weber (1995), suggesting their results are likely to be affected by the same issue. When a longer time-frame is used, 1981–2019 (and thus many more observations are included in the regressions), the Sargan test clearly and obviously rejects the over-identifying instruments. Moreover, the instruments are found to be weak, since the instrument set includes a very large number of variables many of which poorly predict income growth. Similar problems may well to be present in other papers which test for excess sensitivity.

The clear conclusion of this paper is that much of the previous literature has used too many instruments, many of which poorly predict current income growth, and has not used enough time periods in their regressions. To address these criticisms, the paper reports results for a longer time period, and a substantively smaller set of instruments which excludes variables which do not predict income in the first-stage regression. This smaller instrument set passes the Sargan test; and although the rank test shows the instruments are weak, results can be reported using the CLR test. In these new regressions, predictable income growth is significant when the household variables are omitted. But unlike the earlier literature, predictable income growth remains significant when household variables are added to the regression. This paper therefore argues households do not follow the PIH and that the earlier papers only passed the excess sensitivity test because of their choice of instruments and the relatively short time-frame for which they had data.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Altonji JG, Siow A (1987) Testing the response of consumption to income changes with (noisy) panel data. *Quart J Econ* 102:293–328
- Andreski P, Li G, Samancioglu MZ, Schoeni R (2014) Estimates of annual consumption expenditures and its major components in the PSID in comparison to the CE. *Am Econ Rev* 104(5):132–135
- Andrews DWK, Moreira MJ, Stock JH (2006) Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74(3):715–752
- Attanasio O, Weber G (1995) Is consumption growth consistent with intertemporal optimization? Evidence from the consumer expenditure survey. *J Polit Econ* 106:1121–1157
- Bowsher CG (2002) On testing overidentifying restrictions in dynamic panel data models. *Econ Lett* 77:211–220
- Commault J (2022) Does consumption respond to transitory shocks? Reconciling natural experiments and semistructural methods. *Am Econ J Macroecon* 14(2):96–122
- Deaton A (1992) *Understanding Consumption*. Oxford University Press, Oxford

- Finlay K, Magnusson L, Schaffer M (2013) “weakiv: Weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models”. <http://ideas.repec.org/c/boc/bocode/s457684.html>
- Fisher JD, Johnson DS, Smeeding TM, Thompson JP (2020) Estimating the marginal propensity to consume using the distributions of income, consumption, and wealth. *J Macroecon* 65:1–14
- Flavin M (1981) The adjustment of consumption to changing expectations about future income. *J Polit Econ* 89:974–1009
- Jappelli T, Pistaferri L (2010) The consumption response to income changes. *Ann Rev Econ* 2:479–506
- Johnson DS, Parker J, Souleles NS (2006) Household expenditure and the income tax rebates of 2001. *Am Econ Rev* 96:1589–1610
- Kleibergen F, Paap R (2006) Generalized reduced rank tests using the singular value decomposition. *J Econ* 133:97–126
- Montiel Olea JL, Pflueger C (2013) A robust test for weak instruments. *J Bus Econ Stat* 31:358–369
- Moreira MJ (2003) A conditional likelihood ratio test for structural models. *Econometrica* 71:1027–1048
- Parker J (1999) The reaction of household consumption to predictable changes in social security taxes. *Am. Econ. Rev.* 89:959–973
- Sanderson E, Windmeijer F (2016) A weak instrument F-test in linear IV models with multiple endogenous variables. *J. Econ.* 190:212–221
- Souleles NS (1999) The response of household consumption to income tax refunds. *Am. Econ. Rev.* 89:947–958
- Staiger D, Stock JH (1997) Instrumental variables regression with weak instruments. *Econometrica* 65:557–586
- Stock J, Yogo M (2005) Testing for weak instruments in linear IV regression. In: Andrews DW (ed) *Identification and Inference for Econometric Models*. Cambridge University Press, New York, pp 80–108
- Zeldes SP (1989) Consumption and liquidity constraints: an empirical investigation. *J Polit Econ* 97:305–346

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.