TI 2022-044/VI
Tinbergen Institute Discussion Paper

Housing returns and intertemporal substitution in consumption: estimates for industrial economies

Lorenzo Pozzi[1,2]

1 Erasmus University Rotterdam

2 Tinbergen Institute

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Housing returns and intertemporal substitution in consumption: estimates for industrial economies

Lorenzo Pozzi*[1]

[1] *Erasmus University Rotterdam & Tinbergen Institute*

July 2022

Abstract

This paper uses housing returns to estimate the elasticity of intertemporal substitution (EIS) in consumption for fifteen advanced economies over the postwar period $1950 - 2015$. As housing is the main asset for the majority of households, returns on housing are better suited to estimate the EIS than the asset returns typically considered in the literature, i.e., equity and bill returns. An estimable regression equation for aggregate consumption growth and returns is obtained from the aggregation of the consumption Euler equations of heterogeneous agents. As the regression equation includes unobserved omitted variables, we use instrumental variables estimation. We exploit both the temporal and spatial dimensions of the panel by instrumenting the domestic return using its own lag and a cross-country average of foreign returns. Both instruments are strong and allow to test the overidentifying restriction. The restriction holds once we control for common international growth and financial factors in the regression equation. We report a baseline elasticity estimate of about 0.21. This is substantially larger than the elasticities estimated from equity and bill returns which, in line with the extant literature, are found not to be significantly different from zero.

JEL Classification: E21, C23

Keywords: consumption, intertemporal substitution, housing returns, panel data, instrumental variables

*Department of Economics, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. Email: pozzi@ese.eur.nl. Website: http://people.few.eur.nl/pozzi.

1 Introduction

The elasticity of intertemporal substitution (EIS) is one of the main parameters driving consumption behavior. It measures the willingness of consumers to substitute consumption across time periods in response to expected changes in the returns on their wealth. As an important behavioral parameter, the EIS is a key input into a wide range of economic models that feature intertemporal choice, such as macroeconomic DSGE models, asset-pricing models and environmental models. Unfortunately, there is no agreement in the economic and financial literature as to which values for this parameter are appropriate and researchers have used a wide range of values for the EIS when calibrating the aforementioned models. This is problematic as the model-based evaluation of economic events and policies often strongly depends on the values assumed for the EIS where even relatively small differences in calibration may lead to substantially different conclusions (see e.g., Havranek et al., 2015, for an illustration of this for monetary policy). The wide range of values used in calibration reflects the wide variety of EIS estimates that have been reported in the large empirical literature that estimates this important parameter. Since the seminal paper of Hall (1988) who argued that the EIS is close to zero, many studies have used a plethora of approaches to estimate the EIS. We refer to the relatively recent meta-analyses of Havranek et al. (2015) and Havranek (2015) for an overview of these studies, the methods they employ and the estimates they report. These analyses conclude that EIS estimates reported by macro-level studies that use aggregate returns and consumption are generally smaller than those based on micro-level studies that typically focus on particular asset holders. After controlling for selective reporting that biases the reported EIS estimates upward, Havranek (2015) reports a mean for the macro-based estimates equal to zero and a mean for the micro-based estimates that lies in the range of $0.3 - 0.4$.

An aspect of EIS estimation that has received surprisingly little scrutiny in all these studies, both micro- and macro-based, is the choice of the asset return. The asset return considered is typically either the real return on a Treasury bill (i.e., the short-run real interest rate) or the real return on equity. This is somewhat surprising as relatively few households directly own bonds and stocks and these financial assets are highly concentrated in the upper part of the income distribution. For the US, for example, only about 24% of households directly own bonds and stocks according to the 2019 Survey of Consumer Finances. This paper therefore contributes to the literature by looking at an asset class that has not typically been considered when estimating the EIS, namely housing. Housing is the main asset for the majority of households and a large fraction of households own a house. For the US, for example, about 65% of households own a home as a primary residence according to the 2019 Survey of Consumer Finances.[1]

[1] A recent Financial Times article entitled 'The oldest asset class of all still dominates modern wealth' (FT, November 14th 2021) notes '...for all the talk of digitalization, it seems that bricks and mortar are the new bricks and mortar'.

Hence, housing may be substantially more relevant for consumption than other assets. Interestingly, this seems to be confirmed by studies that investigate the impact of wealth on consumption and find that the marginal propensity to consume out of housing wealth is larger than that out of financial wealth (see e.g., Case et al., 2005; Carroll et al., 2011). As such, one would expect that the returns on housing matter for consumption as well, in particular, for its allocation across time. Intuitively, an expected increase in the return on housing leads to a higher consumption growth rate as consumption is shifted from the present to the future because consumers save more to increase the amount of housing they hold and benefit from the higher return. The opposite holds for an expected decrease in the housing return. Changing the amount of the (relatively illiquid) housing asset in portfolio need not be as drastic as buying or selling a house but can be achieved, for instance, by taking out or repaying mortgages or home equity loans. Hence, taking a closer look at the impact of housing returns on consumption growth and at the implied intertemporal substitution in consumption seems warranted. To the best of our knowledge, only two published studies have linked the housing market to the estimation of the EIS. Neither of these micro-based studies is concerned specifically with the impact of housing returns on consumption growth, however. Dacy and Hasanov (2011) include housing in a consumer portfolio (i.e., a synthetic mutual fund) consisting of many different assets and investigate the impact of that portfolio's return on intertemporal substitution while Best et al. (2020) focus on discrete jumps in UK mortgage interest rates as a source of variation to identify the EIS.[2]

The impact of housing returns on consumption growth and the implications for the EIS are investigated in this paper using a cross-country panel data analysis for fifteen industrial economies over the postwar period $1950-2015$. To this end, the paper uses Jordà et al. (2019)'s recently developed extensive dataset on aggregate housing returns for advanced economies which, over the postwar period, provides uninterrupted time series on these returns at the annual frequency for all the countries in the sample. The contribution of the paper is both theoretical and empirical. Theoretically, we consider a heterogeneous agent setting consisting of optimizing consumers who face preference shifters and incomplete financial markets. An estimable regression equation for aggregate consumption growth and returns is obtained from the aggregation of the consumption Euler equations of these agents. This model is substantially richer than the representative agent settings typically considered in macro-based studies (see e.g., Hall, 1988; Yogo, 2004) and its generality provides some important insights on estimating the EIS using housing returns and aggregate data. Specifically, our set-up shows that identifying the EIS from the returns on a particular asset such as housing depends on the extent to which this asset is held by consumers.

[2]The approach of Dacy and Hasanov (2011) is in the vein of Mulligan (2002) who measures intertemporal substitution in response to the return on a representative unit of total capital.

Moreover, it shows that identifying the EIS using aggregate data is feasible if complications related to aggregation such as asset ownership rates and aggregation biases are taken into account. Methodologically, we estimate the regression equation of aggregate consumption growth on housing returns for every country using instrumental variables and then average the estimated impact of housing returns across countries in a mean-group panel approach (see Pesaran and Smith, 1995). This heterogeneous panel approach that allows for a country-specific impact of returns on consumption growth is in line with cross-country heterogeneity in the EIS documented by Havranek et al. (2015) and is facilitated by the availability of long time series for housing returns. The use of instrumental variables estimation is motivated by reverse causality and measurement error considerations as well as by the presence of unobserved omitted variables in the regression equation which, according to our theoretical framework, are related to preference shifters, incomplete markets and aggregation. To instrument the domestic housing return, we propose two instruments that exploit both the temporal and spatial dimensions of the panel, namely the one-year lagged domestic housing return and the cross-sectional average of international housing returns. While lags of returns have been widely used in the literature to estimate the EIS from equity and bill returns, cross-sectional averages of returns have not been previously considered as instruments in this context. Their use can be motivated by noting that both theory - i.e., the international CAPM (see Harvey, 1991) - and empirical evidence suggest that international returns have explanatory power for domestic returns (see e.g., evidence by Hirata et al., 2013, for housing). We check for the strength (relevance) of our proposed instrument set using a weak instruments test and compare it to that of alternative sets. With two instruments to identify the one EIS parameter, we also check the validity (exogeneity) of our instrument set using an overidentifying restrictions test. As it is unlikely that the error term of a simple regression of aggregate consumption growth on housing returns is *iid*, the proposed instruments can be invalidated. This is plausible since, as noted above, the error term includes omitted variables that potentially are persistent and/or common across countries. With respect to the latter possibility, there is a large empirical literature that documents the presence of international business and financial cycles in domestic GDP growth and its components such as consumption growth (see e.g., Kose et al., 2003, 2012; Ha et al., 2020). We therefore also estimate a regression specification for consumption growth that includes persistent common international growth and financial factors as additional regressors. Finally, we obtain EIS estimates by combining the estimates for the impact of housing returns on consumption growth with data on homeownership rates. Our approach, in essence, reverses the typical micro-based approaches that estimate the EIS for specific asset holders (see e.g., Vissing-Jorgensen, 2002). Rather than considering only the consumers who hold the asset (in this case, housing), we initially consider all consumers by looking at aggregate consumption, but then apply a correction factor - i.e., the fraction of consumers who own a house - to obtain estimates of the EIS. Hence, in accordance with micro-based studies, we take into account asset ownership while maintaining the data advantages that come with a macro-based approach.[3]

The main findings of the paper are the following. First, our implemented statistical tests reveal that the instruments that we propose for the domestic housing return - i.e., the one-year lagged domestic housing return and the cross-sectional average of foreign housing returns - are strong (i.e., relevant) and valid (i.e., exogenous) when we consider the regression specification that controls for persistent common international growth and financial factors. Second, we find a positive and statistically significant average (mean-group) impact of aggregate housing returns on aggregate consumption growth for our panel of fifteen advanced economies over the period $1950-2015$. The implied average baseline EIS estimate equals 0.21. Our EIS estimates are generally larger than the small near-zero estimates typically obtained by macro studies that link aggregate consumption to the returns on less widely held assets like equity and bills. And they are generally smaller that the estimates in the 0.3-0.4 range typically obtained by micro studies that consider the consumption only of particular asset holders. Interestingly, our EIS estimates are larger than but generally not too different from the micro-based estimates for the UK reported recently by Best et al. (2020) who focus also on the housing market (albeit on mortgage rates rather than housing returns). Third, our estimations suggest that the EIS is higher during the globalization period $1985-2015$ compared to the pre-globalization period $1950-1985$. This result is indicative of time-variation in the EIS, a possibility which has barely been addressed by the literature. Finally, when applying our empirical approach to equity and bill returns, we find, in accordance with the extant macro-based literature, elasticities that are not significantly different from zero.

The paper proceeds as follows. Section 2 presents and discusses the theoretical framework that underlies our empirical approach to estimate the EIS from housing returns. Section 3 focusses on the estimation methodology, i.e., the empirical specifications are discussed as well as the IV and mean-group estimation approaches. Attention is further given to the data used in estimation. Section 4 shows the results obtained from estimating regressions for consumption growth and housing returns and presents the implied estimates for the EIS. Section 5 concludes.

[3]These advantages are the availability of data for many periods (which allows to link long time series for consumption to long time series for housing returns and to check for time-variation in the EIS) and for many countries (which allows to control for cross-country differences in the EIS). Micro-level EIS studies typically only consider the US or the UK as micro data for both consumption and returns are generally unavailable for other countries. Additionally, as noted by Havranek et al. (2015), the magnitude of the EIS estimates obtained with micro data depends on whether estimation relies on cross-sectional or time series variation in the rate of return.

2 Theoretical framework

This section presents the theoretical framework used to investigate the impact of the real rate of return on an asset on aggregate consumption growth and the implied elasticity of intertemporal substitution (EIS). First, we derive and discuss the Euler equation of a utility maximizing individual consumer who faces uncertainty about future income and asset returns and a potentially binding credit constraint. Then, we average the consumer-specific consumption growth rate across consumers to obtain an equation for aggregate consumption growth that depends on the aggregate real asset return and on a number of unobserved variables. In the ensuing discussion, we argue that the identification of the EIS is affected by the number of consumers who hold the asset under consideration and by the presence of unobserved variables in the equation.

2.1 Consumer Euler equation

The economy consists of consumers who face uncertain future labor income, uncertain returns on a portfolio of assets, and a potentially binding credit constraint. In every period each infinitely lived consumer j (with $j = 1, ..., J$) makes a consumption and portfolio decision by maximizing the expected value of lifetime utility, i.e., we have,

$$\max E_{jt} \sum_{l=0}^{\infty} \rho^t U(C_{j,t+l}, \delta_{j,t+l}) \tag{1}$$

where E_{jt} is the rational expectations operator conditional on consumer j's period t information set, where $0 < \rho \leq 1$ is the discount factor that reflects the rate of time preference, where $U(.)$ is an isoelastic contemporaneous utility function, where C_{jt} is consumer j's real consumption in period t and where δ_{jt} captures (unspecified) preference shifters that shift marginal utility over time. Maximization occurs subject to the budget constraint,

$$C_{jt} + A_{j,t+1} = \left[\sum_{k=1}^{K} \omega_{jt}^k R_{jt}^k \right] A_{jt} + Y_{jt}^L \tag{2}$$

and the credit constraint,

$$A_{j,t+1} \geq 0 \tag{3}$$

where A_{jt} denotes real asset wealth of consumer j at the beginning of period t, where Y_{jt}^L denotes consumer j's real labor income in period t, where R_{jt}^k is consumer j's period t gross real return rate on asset k (with $k = 1, ..., K$) and with ω_{jt}^k the share of wealth held by consumer j in asset k.

The maximization problem given by eqs.(1)-(3) implies a set of first-order conditions or Euler equations linking successive periods, i.e., for periods $t-1$ and t we have,

$$E_{j,t-1}\left(\rho R^k_{jt}\frac{U'(C_{jt},\delta_{jt})}{U'(C_{j,t-1},\delta_{j,t-1})}\right)+\lambda_{j,t-1}=1 \tag{4}$$

which holds for every asset k held by consumer j and where $\lambda_{j,t-1}\geq 0$ is the (normalized) Langrange multiplier associated with the credit constraint which is positive when the constraint is binding and zero when the constraint is not binding (see e.g., Zeldes, 1989). Eq.(4) can also be written as,

$$\left(\rho R^k_{jt}\frac{U'(C_{jt},\delta_{jt})}{U'(C_{j,t-1},\delta_{j,t-1})}\right)=1-\lambda_{j,t-1}+\eta^k_{jt} \tag{5}$$

where η^k_{jt} is an expectation error uncorrelated with period $t-1$ information, i.e., we have $E_{j,t-1}\eta^k_{jt}=0$. The contemporaneous utility function $U(.)$ is isoelastic and given by $U(C_{jt},\delta_{jt})=\frac{C_{jt}^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}\exp(\delta_{jt})$ where $\sigma>0$ denotes the elasticity of intertemporal substitution or EIS.[4] Hence, we assume that the EIS is homogeneous across consumers within the economy considered which is in line with recent micro-based evidence for the UK provided by Best et al. (2020). Using this, we can rewrite eq.(5) as,

$$\left(\rho R^k_{jt}\frac{C_{jt}^{-\frac{1}{\sigma}}\exp(\delta_{jt})}{C_{j,t-1}^{-\frac{1}{\sigma}}\exp(\delta_{j,t-1})}\right)=1-\lambda_{j,t-1}+\eta^k_{jt} \tag{6}$$

After taking logs of both sides of this expression and solving for the growth rate in consumption, Δc_{jt}, we obtain,

$$\Delta c_{jt}=\sigma\ln\rho+\sigma r^k_{jt}+\sigma\Delta\delta_{jt}+\sigma\nu^k_{jt} \tag{7}$$

where $c_{jt}=\ln(C_{jt})$, $r^k_{jt}=\ln(R^k_{jt})$ and $\nu^k_{jt}=-\ln(1-\lambda_{j,t-1}+\eta^k_{jt})$.

Apart from the term σr^k_{jt} which is our object of interest and which captures intertemporal substitution in consumption with respect to changes in the return on asset k, there are two additional time-varying terms in eq.(7). First, the term $\sigma\Delta\delta_{jt}$ captures the impact of preference shifters on consumption growth. Second, the unexpected part of ν^k_{jt} reflects new information arriving to the consumer while the expected part of ν^k_{jt}, i.e., the term $E_{j,t-1}\nu^k_{jt}=-E_{j,t-1}\ln(1-\lambda_{j,t-1}+\eta^k_{jt})$, reflects the incomplete (financial) markets component of consumption growth which is due to the presence of a precautionary saving motive and a liquidity constraint (see Parker and Preston, 2005).[5]

Finally, we write,

$$\Delta c_{jt}=\mu+\sigma r^k_{jt}+\epsilon^k_{jt} \tag{8}$$

[4] In our expected utility framework, the EIS σ equals the inverse of the coefficient of relative risk aversion which is given by $\frac{1}{\sigma}>0$.

[5] This component reduces period $t-1$ consumption and augments period t consumption thereby raising consumption growth from $t-1$ to t, i.e., we have $E_{j,t-1}\nu^k_{jt}>0$. To see this, we suppress sub- and superscripts and note that $\ln(E(1-\lambda+\eta))=\ln(1-\lambda)\leq 0$ (this follows from $E(\eta)=0$, $E(\lambda)=\lambda$ and $\lambda\geq 0$). For the concave log function, we have that $\ln(E(.))>E(\ln(.))$ so that $E(\ln(1-\lambda+\eta))<0$ and $-E(\ln(1-\lambda+\eta))>0$.

where $\mu = \sigma \ln \rho$ and $\epsilon_{jt}^k = \sigma \Delta \delta_{jt} + \sigma \nu_{jt}^k$, i.e., we collect the unobserved components $\sigma \Delta \delta_{jt}$ and $\sigma \nu_{jt}^k$ in the term ϵ_{jt}^k. If these unobserved components are persistent, the term ϵ_{jt}^k is characterized by autocorrelation which, a priori, is of unknown order.

2.2 Aggregation

When aggregating eq.(8), we take into account two complications. First, we note that the average of log consumption, $\frac{1}{J}\sum_{j=1}^{J} c_{jt}$, is not identical to the log of average consumption, $c_t = \ln(\frac{1}{J}\sum_{j=1}^{J} C_{jt})$. The aggregation in the model pertains to the former while the latter is in accordance with how the aggregate data that we use in estimation are constructed. The difference between both is denoted by Theil's entropy κ_t and should be considered when averaging the growth rates Δc_{jt} across J consumers (see Attanasio and Weber, 1993). Second, we note that not all J consumers hold asset k. As such, the aggregate real return r_t^k for which we have data differs from $\frac{1}{J}\sum_{j=1}^{J} r_{jt}^k$ as r_t^k is an average over only those consumers $J^k \subset J$ who hold asset k. Aggregation of eq.(8) then gives,

$$\Delta c_t = \mu + \psi^k r_t^k + \varepsilon_t^k \tag{9}$$

where $\Delta c_t = \frac{1}{J}\sum_{j=1}^{J} \Delta c_{jt} - \Delta \kappa_t$, $r_t^k = \frac{1}{J^k}\sum_{j=1}^{J^k} r_{jt}^k$ and $\varepsilon_t^k = \frac{1}{J}\sum_{j=1}^{J} \epsilon_{jt}^k - \Delta \kappa_t$. We note that $\psi^k = \frac{J^k}{J}\sigma \geq 0$ where the factor $0 \leq \frac{J^k}{J} \leq 1$ ensures the consistency of summing over J for Δc_t and summing over J^k for r_t^k.[6]

2.3 Discussion

The intuition behind eq.(9) is straightforward, i.e., higher (expected) returns on an asset coincide with higher consumption growth rates because consumption is shifted from the present to the future as consumers save more to increase their holding of the asset and benefit from the higher returns.[7] This equation - i.e., a regression of aggregate consumption growth Δc_t on the real asset return r_t^k - is generally in accordance with the specification introduced by Hall (1988) in a representative agent setting with complete markets and no preference shifters and constitutes the most widely used specification to estimate the EIS (see e.g., Havranek et al., 2015, for an overview). Our use of a more general heterogeneous agent framework to derive eq.(9) provides some important additional insights, however.

First, unlike in the representative agent case, the slope coefficient ψ^k is not equal to the EIS σ. Rather, it equals the EIS σ times a factor $\frac{J^k}{J}$ that reflects to what extent the asset is held by consumers. As such,

[6]Without loss of generality, we index the J^k consumers out of J who hold asset k from 1 to J^k. Since the consumers who do not hold the asset k implicitly obtain zero returns on k, the difference between r_t^k and $\frac{1}{J}\sum_{j=1}^{J} r_{jt}^k$ amounts to a mere factor of proportionality $\frac{J}{J^k}$, i.e., we have $r_t^k = \frac{J}{J^k}\left(\frac{1}{J}\sum_{j=1}^{J} r_{jt}^k\right)$.

[7]Specifically, a period $t-1$ expected increase in r_t^k reduces c_{t-1}, thereby increasing Δc_t.

the coefficient ψ^k constitutes a lower bound to the actual EIS and an estimate for the EIS σ is obtained from an estimate for ψ^k after multiplying it by a proxy for $\frac{J}{J^k}$. We note that if the asset is not widely held, then it is difficult to identify the EIS as $\psi^k \approx 0$ in this case. Hence, it makes sense to estimate the EIS based on an asset such as housing which is held by many consumers. Indeed, while it should, in principle, be possible to estimate the same EIS σ by estimating eq.(9) for any asset k (independently of all other assets), in practice this requires a sufficiently large ratio $\frac{J^k}{J}$ for this asset.

Second, the error term ε_t^k includes many unobserved components (preference shifters, incomplete market components, Theil's entropy measures) that can be correlated with the return r_t^k. Moreover, if these components are persistent, then ε_t^k is characterized by autocorrelation. Importantly, this autocorrelation is of unknown order a priori.[8]

Our set-up therefore illustrates that estimating the EIS from aggregate data (aggregate consumption and aggregate housing returns) is feasible on the condition that the complications related to aggregation are dealt with. Here, this means that the estimation strategy employed to identify the EIS must take into account the quantity $\frac{J^k}{J}$, i.e., the fraction of consumers that hold the asset. It must also take into account the characteristics of the error term ε_t which potentially includes, among other unobserved variables, aggregation biases (i.e., Theil's entropy). The estimation strategy is the topic of the next section.

3 Empirical specifications, methodology and data

This section presents and discusses the empirical specifications - i.e., the cross-country panel regression equations - used to estimate the EIS. Moreover, we elaborate on the methodology. We discuss the instrumental variables (IV) estimation method used at the country level where we particularly focus on identification, i.e., on the choice of instruments and on their strength and validity. We also provide details on the panel mean-group estimation approach. The section ends with a discussion of the data used in estimation, i.e., cross-country data on consumption, housing returns and other variables for fifteen industrial economies over the postwar period $1950 - 2015$.

[8]Note that the lack of a priori knowledge on the autocorrelation structure and order of the error term stands in contrast some of the earlier literature which, in the context of a representative agent setting with complete markets and no preference shifters, points to either an *iid* error term or to the potential presence of a simple $MA(1)$ component in the error term induced by time aggregation and/or measurement error (see e.g., Hall, 1988; Sommer, 2007).

3.1 Empirical specification

We consider the following regression of aggregate consumption growth on the real rate of return on housing, i.e., we estimate,

$$\Delta c_{it} = \mu_i + \psi_i r_{it} + \varepsilon_{it} \tag{10}$$

where Δc_{it} is the growth rate of per capita real aggregate consumption in period t and country i (with $i = 1, ..., N$), where r_{it} is the real rate of return on housing, where μ_i is a country fixed effect, where ψ_i denotes the country-specific impact of r_{it} on Δc_{it} and where ε_{it} is the error term. Note that, compared to the previous section, we omit the superscripts k from the parameter ψ_i, from the return variable r_{it}, and from the error term ε_{it}, as, with the exception of some additional results presented in Section 4, the paper focusses solely on one particular asset, i.e., on housing.

We note that, given the evidence in the literature of cross-country heterogeneity in the EIS (see Havranek et al., 2015) *and* given the relatively long time series at our disposal, we allow the parameter ψ_i to be country-specific, i.e., we consider a heterogeneous panel. Furthermore, the discussion in the previous section has made clear that the parameter ψ_i constitutes a lower bound to the EIS. A country-specific estimate for the EIS, i.e., an estimate for σ_i, is then obtained by dividing ψ_i by a proxy for the fraction of consumers who hold the housing asset. In Section 4.4 below, we use country-specific homeownership rates for this purpose (note that ψ_i and σ_i are equal if the homeownership rate in country i equals 100%).

When looking at the error term in eq.(10), we note that it need not be *iid*. First, as discussed in the previous section, ε_{it} can be autocorrelated with autocorrelation of unknown order. Second, as we discuss below in Sections 3.2 and 3.3, there can be cross-sectional dependence in ε_{it} due to the presence of unobserved common factors. Third, as emphasized for instance by Yogo (2004), ε_{it} can be heteroskedastic.

Most importantly, however, the error term ε_{it} most likely is correlated with the regressor r_{it}. There are three reasons for this. First, while the real housing return affects consumption growth, consumption growth most likely also affects the rate of return, i.e., there may be reverse causality. Moreover, as discussed in detail in the previous section, the error term may include time-varying unobserved omitted variables and these can be correlated with the regressor. Finally, there may be measurement error, not only in the dependent variable, but also in the regressor. Because of the potential contemporaneous correlation between r_{it} and ε_{it} - i.e., the endogeneity of r_{it} -, an instrumental variables approach to estimate eq.(10) is necessary. This is discussed in the next section.

3.2 Per country IV estimation

Given the potential endogeneity of r_{it}, one or more instrumental variables are necessary to identify the parameter ψ_i. More specifically, if $E(r_{it}\varepsilon_{it}) \neq 0$, we need one or more instruments collected in a vector z_{it} for which $E(z_{it}\varepsilon_{it}) = 0$ to identify ψ_i. Estimation of eq.(10) can then be conducted using instrumental variables (IV) estimation which provides consistent estimates of ψ_i if the instruments are relevant (i.e., not weak) and valid (i.e., exogenous). While, contrary to estimation with the generalized method of moments (GMM), IV is not robust to heteroskedasticity and autocorrelation, it does allow for heteroskedasticity- and autocorrelation-robust inference using standard errors by Newey and West (1987) at the country level. Importantly for our purposes, as noted by Andrews and Stock (2018), the existing and appropriate weak instrument pre-testing procedures that we implement are tailored to IV estimation. Hence, the estimations in the paper are mostly conducted using IV. As a robustness check, however, we also report estimates that are obtained from applying GMM. In what follows, we discuss the choice, strength and exogeneity of the instruments that we use in our IV approach.

3.2.1 Choice and strength of instruments

We propose two relevant instruments for the housing return r_{it}, i.e., the lagged domestic housing return $r_{i,t-1}$ and the cross-sectional average of international housing returns $\bar{r}_t$.

Lagged domestic return $r_{i,t-1}$

We exploit the temporal dimension of our panel and consider the lagged domestic return $r_{i,t-1}$ as a first instrument for r_{it}. This approach is in line with the extant literature on estimating the EIS that typically includes lagged returns in the instrument vector when estimating a regression equation like eq.(10)(see Havranek et al., 2015).[9] For housing returns, this approach seems to be even more appropriate as, based on preliminary AR regressions, we find that the persistence in housing returns is high and larger than the persistence that can be observed in the returns of other assets such as equity (these results are unreported but available upon request).

[9]Often, however, the earlier literature includes *only* deeper lags of returns (and other variables) as instruments based on a priori assumptions about the order of autocorrelation in the error term. This approach makes little sense based on our general theoretical framework because, as discussed Section 2.3, we do not a priori know the order of autocorrelation in ε_{it}. Moreover, the literature points to severe weak instrument problems when estimating the EIS using as instruments variables that are lagged twice or more (see e.g., Neely et al., 2001; Yogo, 2004; Montiel Olea and Pflueger, 2013). This is not surprising as using only deeper lags throws away a lot of information if predictability in the return stems mostly from its first lag. Some studies such as Yogo (2004) tackle this through weak-instrument robust inference but there are limitations to this inference as well, e.g., outcomes can be inconclusive with confidence sets for the parameter of interest that are very wide or, conversely, empty.

Cross-sectional average of foreign returns $\bar{r}_t$

Apart from the lagged domestic housing return $r_{i,t-1}$, we consider a second instrument for r_{it} by exploiting the cross-country availability of our returns data, namely the contemporaneous (weighted) average $\bar{r}_t$ of the returns r_{it} of the different countries in the sample. Most asset markets are to some degree integrated internationally. As such, we expect that this international return has a significant impact on the domestic return of the country considered. Indeed, there is a large literature on international financial market integration that provides evidence that domestic asset returns are driven by international asset returns.[10] With respect to housing, in particular, Hirata et al. (2013) document significant international synchronization of house prices in advanced economies which implies that housing returns co-move across these countries as well. The use of this type of instrument, while not previously considered in the estimation of the EIS, follows in the vein of recent studies that use cross-country (weighted) averages of variables to instrument their domestic counterparts.[11]

Testing for weak instruments

The first-stage regressions of our IV approach where r_{it} is regressed on the instruments $\bar{r}_t$ and $r_{i,t-1}$ (and a constant) allows to evaluate instrument strength. To this end, we consider both the standard F statistic and the effective F statistic of Montiel Olea and Pflueger (2013). As noted also by Andrews and Stock (2018), the latter is the appropriate statistic to use when considering a linear IV regression with potential heteroskedasticity and autocorrelation in the error term. The null hypothesis of weak instruments based on the effective F is the hypothesis that the bias of the IV estimator relative to that of a 'worst-case' benchmark (like OLS) exceeds the threshold $\tau\%$ where we set τ to respectively 30%, 20% and 10%. Under the null hypothesis of weak instruments, the effective F statistic follows a non-central χ-squared distribution.[12]

We note that, apart from looking at the F statistics which are informative about the instrument set as a whole, we also check whether the included instruments are individually significant in the first-stage

[10] We refer to Harvey (1991) for early work on the international CAPM and to Forbes and Chinn (2004) for an example of empirical work that links domestic asset returns to international factors using factor analysis.

[11] Examples are Jordà et al. (2015) for short-term nominal interest rates and Furceri and Loungani (2018) for capital account liberalization.

[12] More specifically, denote the effective F statistic by F^e and the effective degrees of freedom by K^e. F^e and K^e are calculated according to eqs.(3) and (4) in Montiel Olea and Pflueger (2013) using a heteroskedasticity- and autocorrelation-consistent covariance matrix for the OLS estimates of the empirical model in reduced form (after projecting out exogenous variables). We use the simplified test detailed in Section 2.2.2 of their paper which is more conservative. To this end, we set the variable x in their eq.(4) to the inverse of the threshold τ. P-values of the test for different threshold levels are obtained by noting that F^e is distributed as $\frac{\chi^2_{K^e}(K^e x)}{K^e}$ under the null with $\chi^2_{K^e}(K^e x)$ denoting a non-central χ^2 distribution with K^e degrees of freedom and non-centrality parameter $K^e x$.

regressions.[13]

3.2.2 Exogeneity of instruments

The question then remains whether the instruments $r_{i,t-1}$ and $\bar{r}_t$ are valid, i.e., whether they are uncorrelated with the error term ε_{it}. If ε_{it} were *iid* both in time (no autocorrelation) and in space (no cross-sectional dependence), then we could be sure that both instruments are exogenous. There are reasons to suspect that this is not the case, however. Again, we discuss each instrument in turn. We then discuss the overidentifying restriction test that can be implemented to formally test orthogonality between the error term and both relevant instruments.

Lagged domestic return $r_{i,t-1}$

With respect to the potential exogeneity of the lagged domestic return $r_{i,t-1}$, we note that the theoretical framework presented and discussed in Section 2 suggests that a plethora of unobserved omitted variables can show up in the error term ε_{it}. If these omitted variables are persistent, the error term is autocorrelated and this autocorrelation is of unknown order a priori. And if the instrument $r_{i,t-1}$ is correlated with the persistent omitted variables that are potentially present in the error term, it is not exogenous. While finding autocorrelation in the error term when estimating eq.(10) is informative, it is not sufficient to conclude that $r_{i,t-1}$ is invalid as an instrument. To determine the orthogonality between the instruments and the error term, an overidentified model is required. For that, a second relevant instrument, i.e., $\bar{r}_t$, is needed.

Cross-sectional average of foreign returns $\bar{r}_t$

With respect to the potential exogeneity of the international return $\bar{r}_t$, we note, first, that it is unlikely that domestic conditions have an impact on $\bar{r}_t$. To further minimize the potential impact of country i on $\bar{r}_t$ in estimation, we also consider versions of $\bar{r}_t$ that exclude the domestic return from the calculation of the cross-sectional average $\bar{r}_t$. Second, from a theoretical point of view, there is no reason to suspect that $\bar{r}_t$ has an impact on consumption growth Δc_{it} *in addition to* the influence that it exerts on consumption growth through the domestic return r_{it} as there is no separate role for international returns in standard

[13] In the extant earlier literature, for instance, often many lags of r_{it} are included as instruments but no check occurs to find out whether all included lags are individually significant. A problem with this approach is that even if the instrument set as a whole passes the weak instrument test, it may still contain irrelevant instruments. This complicates the implementation of the overidentifying restrictions test which evaluates the exogeneity of the instruments and which is discussed in Section 3.2.2. As an example, suppose $r_{i,t-1}$ and $r_{i,t-2}$ are used to instrument r_{it} when estimating the EIS. Based on the F tests, the instruments jointly pass the weak instrument test but only $r_{i,t-1}$ is significant in the first-stage regression. With two instruments and one parameter to estimate, the model appears overidentified. Since there is in fact only one relevant instrument, the model is really only just identified and the overidentifying restriction test should not be implemented.

consumption theory, i.e., in standard models domestic consumption and saving are driven by international returns only in so far as they are driven by domestic returns. Empirically, however, $\bar{r}_t$ is not necessarily exogenous if it is correlated with unobserved common factors that potentially drive aggregate consumption growth. Examples of common factors are international business or financial cycles or changes in trade or financial integration that occur simultaneously in most or all countries of the sample.[14] Indeed, there is a large literature that documents the presence of international business and financial cycles in domestic GDP growth and its components (see e.g., Kose et al., 2003, 2012; Ha et al., 2020). Hence, if unobserved common factors are indeed present in the error term ε_{it} - i.e., in aggregate consumption growth Δc_{it} after conditioning on the rate of return r_{it} - and if the instrument $\bar{r}_t$ is sufficiently correlated with these factors, then $\bar{r}_t$ is not exogenous. While the presence of cross-sectional dependence in the error term of eq.(10) is indicative of common international factors, it is not sufficient to conclude that $\bar{r}_t$ is invalid as an instrument. The joint validity of the instruments can be tested, however, using the overidentifying restrictions test.

Testing the overidentifying restriction

With two relevant - i.e., sufficiently strong - instruments to identify one parameter of interest, the model is overidentified and we can test the exogeneity of the instrument set - i.e., its orthogonality with the error term - using the Sargan-Hansen overidentifying restrictions test (see Sargan, 1958; Hansen, 1982). Under the null hypothesis that the instrument set is valid, this statistic is χ-squared distributed with one degree of freedom, i.e., the number of instruments minus the number of regressors. While the test does not allow to assess the individual validity of the instruments, it nonetheless provides valuable insight into the validity of the empirical model and the instrument set as a whole.

3.3 Specification with persistent common factors

From the discussion in the previous section, we note that persistence (autocorrelation) and commonality (cross-sectional dependence) in the error term ε_{it} of eq.(10) can invalidate the relevant instruments $r_{i,t-1}$ and $\bar{r}_t$ that we have at our disposal for the endogenous variable r_{it}, i.e., if ε_{it} is neither *iid* in time nor in space, we may have $E(\varepsilon_{it} r_{i,t-1}) \neq 0$ and $E(\varepsilon_{it}\bar{r}_t) \neq 0$. To deal with this, we take out (or soak up) a large part of the economically meaningful persistence and commonality from the error term that is likely to induce correlation with the instruments. To this end, we control for the presence of persistent

[14] How do these factors fit in the model of Section 2? We give one example. Just as the domestic return r_{it} can be driven by international return $\bar{r}_t$, so the unobserved domestic incomplete financial market component that is present in the error term ε_{it} of aggregate consumption growth (as discussed in the theory of Section 2) can be driven by the international financial cycle.

and common factors in the error term of eq.(10) by including as additional regressors proxies for an international (income) growth factor and for an international financial factor.[15] Hence, we consider the extended specification,

$$\Delta c_{it} = \mu_i + \psi_i r_{it} + \gamma_i^{inc} f_t^{inc} + \gamma_i^{fin} f_t^{fin} + \varepsilon_{it} \tag{11}$$

where f_t^{inc} denotes the common international (income) growth factor in period t and f_t^{fin} denotes the international financial factor in period t with respective country-specific factor loadings γ_i^{inc} and γ_i^{fin}. In line with our approach for the international return $\bar{r}_t$, we use the cross-sectional average of the per capita real GDP growth rates of the countries in the sample as a proxy for f_t^{inc} while we use the cross-sectional average of the per capita real credit growth rates of the countries in the sample as a proxy for f_t^{fin}. And as with $\bar{r}_t$, we calculate different versions of these cross-sectional averages (weighted or not and with the domestic GDP/credit growth rate included or excluded). We note that the international common factors are based on the growth rates of per capita real GDP and credit rather than on trend-cycle decompositions (see also e.g., Kose et al., 2012; Hirata et al., 2013). As such, as GDP an credit growth rates are potentially driven not only by transitory but also by permanent shocks, the common growth and financial factors do not merely reflect international business and financial cycles but may also capture more structural international changes such as trade and financial integration.

A major advantage of using these international factors as controls is that they can be considered exogenous, i.e., they most likely are not affected very much by the conditions in one particular country (especially when we consider the versions that exclude the country-specific GDP/credit growth rates). This implies that they can be instrumented by themselves. The empirical model is therefore virtually unchanged. It still consists of one endogenous regressor, namely r_{it}, so that the standard and effective F tests can be applied to test instrument strength.[16] And the number of overidentifying restrictions is also unchanged, i.e., the Sargan-Hansen test still has one degree of freedom as there are two additional regressors but also two additional instruments.

3.4 Mean-group panel approach

Eqs.(10) and (11) are characterized by country-specific slope coefficients so that these equations are estimated country-by-country using instrumental variables (IV) estimation. This is feasible given the

[15] In a heterogeneous panel setting, a typical approach to deal with unobserved common factors is to consider the data in the regression in deviations from cross-sectional means (see e.g., Bond et al., 2010). Since one of our instruments is the cross-sectional average $\bar{r}_t$ of r_{it}, taking the cross-sectional average out of the regressor r_{it}, however, would render $\bar{r}_t$ useless as an instrument for r_{it}.

[16] Instead of projecting out only the constant (i.e., a vector of ones), now three exogenous regressors are projected out of the data before calculating the tests (i.e., the constant and both common factors).

relatively long time series of the data that are available per country. Pesaran and Smith (1995) show that in a heterogeneous panel with a country-specific parameter vector Ψ_i and with a sufficiently large T and N, consistent estimates of the average effects $\Psi = N^{-1}\sum_{i=1}^{N}\Psi_i$ are obtained by averaging over the country-specific coefficient estimates, i.e., $\widehat{\Psi} = N^{-1}\sum_{i=1}^{N}\widehat{\Psi}_i$. The average over the N country-specific estimates is referred to as the mean-group (MG) estimator. It is consistent provided that the country-specific coefficients are consistently estimated. Following Pesaran et al. (1996), the covariance matrix Σ for the mean-group estimator is consistently estimated nonparametrically by,

$$\widehat{\Sigma} = \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\widehat{\Psi}_i - \widehat{\Psi}\right)\left(\widehat{\Psi}_i - \widehat{\Psi}\right)' \tag{12}$$

from which standard errors for the MG estimates are calculated.

For the conducted statistical tests such as the weak instruments test and the overidentifying restrictions test, we combine the per country p-values obtained for each test by calculating the harmonic mean p-value or HMP (see Wilson, 2019). Contrary to the more commonly applied combined probability test by Fisher (1925), the HMP avoids the somewhat restrictive assumption that the country-specific p-values are independent.[17] Since our mean-group estimates are equally weighted (or unweighted) averages, we also calculate the HMP's for every test using equal weights, i.e., the weights always equal $\frac{1}{N}$. As the HMP is anti-conservative when interpreted directly as a p-value, we instead compare it to a critical value. From Wilson (2019), with $N = 15$ and at the 5% level of significance, the critical value is about 0.04, i.e., the null hypothesis of the test is rejected at the 5% level of significance if the HMP is below 0.04.

3.5 Data

We use data over the postwar period $1950 - 2015$ for $N = 15$ industrial economies. These are Australia, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the UK and the US. The availability of data for nominal returns on housing, as reported by Jordà et al. (2019), determines the countries and periods included in our dataset.[18,19] To calculate real housing returns r_{it}, we deflate these nominal returns using the inflation rate calculated from the

[17] When applying Fisher's method, we generally draw identical conclusions to the ones reported in the paper, however.

[18] The data can be found at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GGDQGJ where the nominal housing returns have code 'housing-tr'. Details on the data sources and data construction are discussed in the text and online appendix of Jordà et al. (2019)'s paper.

[19] Before 1950, the available data for housing returns are characterized by multiple large gaps for many countries. This is problematic for estimation, especially since we use cross-sectional averages of returns (and other variables) in estimation. The time series of cross-sectional averages are characterized by artificial shifts when the number of cross-sections used in their calculation changes over time. From 1950 onward, only the housing return series for Belgium is characterized by a large gap over the period $1965 - 1976$ which is why Belgium is not included in the sample. For Germany and Japan, the housing returns data start later (1963 for Germany and 1960 for Japan) but including these countries in the sample at a later date does not introduce visible disruptions in the time series of the cross-sectional averages used in the estimations.

Consumer Price Index (CPI). With respect to the other variables, for log per capita real consumption c_{it}, we use the log of per capita real personal consumer expenditures. For log per capita output y_{it} which is used in the calculation of the common (income) growth factor f_t^{inc}, we use the log of per capita real GDP. For log per capita real credit $cred_{it}$ which is used in the calculation of the common financial factor f_t^{fin}, we use the log of per capita real total loans to the non-financial private sector where the real per capita total loans series is obtained by dividing the nominal total loans series by the CPI and by total population. The series for the CPI, for per capita real personal consumer expenditures, for per capita real GDP, for nominal total loans to the non-financial private sector and for population are taken from the Jordà-Schularick-Taylor macro-history database (see Jordà et al., 2016).[20]

As far as the cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} are concerned, we calculate them in different ways and check the robustness of our results to the different approaches used. Denote a cross-sectional average by $\bar{x}_t$ which equals either $\bar{r}_t$ or f_t^{inc} or f_t^{fin}. If all countries are included in the calculation of $\bar{x}_t$, we have $\bar{x}_t = \sum_{j=1,N} w_{jt}x_{jt}$ where x_{jt} equals either r_{jt} (when $\bar{x}_t = \bar{r}_t$) or Δy_{jt} (when $\bar{x}_t = f_t^{inc}$) or $\Delta cred_{jt}$ (when $\bar{x}_t = f_t^{fin}$) with r_{jt} the real housing return, y_{jt} the log of per capita output and $cred_{jt}$ the log of per capita real credit. If country i is excluded from the cross-sectional average $\bar{x}_t$ used in the regression for country i, we have $\bar{x}_t^{(-i)} = \sum_{j=1,N(j\neq i)} w_{jt}x_{jt}$ with x_{jt} as before. Note that in the latter case, the variable $\bar{x}_t$ differs for every country i. With respect to the weights, if equal weights are used in the calculation of $\bar{x}_t$, then w_{jt} equals one over the number of countries included in the summation $(\forall j, t)$. Conversely, if unequal weights are used in the calculation of $\bar{x}_t$, then we calculate the weight w_{jt} as country j's PPP-adjusted real GDP in period t divided by total PPP-adjusted real GDP in period t of all countries included in the summation.[21] Data for PPP-adjusted real GDP are also taken from the Jordà-Schularick-Taylor macro-history database.[22]

4 Results

This section presents the estimation results. First, in Section 4.1, we take a look at different instrument sets and decide that, in terms of strength, an instrument set consisting of the cross-sectional average of housing returns $\bar{r}_t$ and the lagged domestic housing return $r_{i,t-1}$ is the best choice. Second, in Section 4.2, we present and discuss the estimation results obtained from estimating the specification for

[20]The website is http://www.macrohistory.net/data. The data used have respectively codes 'cpi', 'rconpc', 'rgdppc', 'tloans' and 'pop'.

[21]An alternative population-based weighting scheme offers very similar results, i.e., when the weight w_{jt} is calculated as country j's population in period t divided by total population in period t of all countries included in the summation. These results are not reported but are available upon request.

[22]PPP-adjusted real GDP is calculated by multiplying PPP-adjusted per capita real GDP (code 'rgdpmad') by population (code 'pop').

consumption growth and housing returns without common factors. The obtained results suggest that this specification is not adequate. Third, in Section 4.3, we present and discuss the estimation results obtained from estimating the specification for consumption growth and housing returns that includes both a common international growth factor and a common international financial factor. Finally, from the specification with common factors, which is supported by the data, we obtain estimates for the elasticity of intertemporal substitution (EIS). These are presented in Section 4.4.

4.1 Choice of instruments

Table 1 presents the mean-group estimation results of first-stage regressions where the housing return r_{it} is regressed on different instrument sets. Based on our discussion in the previous section, we focus in particular on the cross-sectional average of housing returns given by $\bar{r}_t$ and on lags of the domestic return, i.e., on $r_{i,t-l}$ (with $l = 1, 2, 3$).[23] Hence, we focus not only on the first lag of r_{it} but we also take a look at the performance of deeper lags (lags two and three) given that the extant literature has often considered such lags when estimating the EIS using returns on equity or bills. The first three columns of the table look at instrument sets containing one variable (apart from the constant), i.e., either $\bar{r}_t$, $r_{i,t-1}$ or $r_{i,t-2}$. All instruments have a (highly) significant impact on r_{it} and, unsurprisingly, the effective F statistics strongly reject that these instruments are weak. We refer to Section 3.2.1 above for details on the weak instruments test.

Since we prefer an overidentified model to a just-identified one, the following five columns in the table report the first-stage regression results obtained with instrument sets that include two or more variables (apart from the constant). The instrument set consisting of $\bar{r}_t$ and $r_{i,t-1}$ (column four) is clearly preferred, both in terms of the magnitude of the reported F-statistics and in terms of the individual significance of the instruments. The instrument set consisting of $\bar{r}_t$ and $r_{i,t-2}$ (column five), while characterized by lower F statistics, also performs well, i.e., the effective F statistic rejects the null of weak instruments and both instruments are individually significant. This does not necessarily mean that an instrument set consisting of $\bar{r}_t$ and $r_{i,t-2}$ is appropriate for our empirical analysis, however. First, from columns six and seven, we observe that once the instrument set also includes the first lag of r_{it} (with or without $\bar{r}_t$), then the individual impact of $r_{i,t-2}$ for r_{it} is no longer significant (even though weakness of the instrument sets as a whole is still strongly rejected). Hence, the lag that matters for r_{it} is $r_{i,t-1}$ rather than $r_{i,t-2}$. Second, once we extend the empirical specification to include common factors as regressors and instruments as detailed in Section 3.3, we can no longer unambiguously reject the null hypothesis of weak instruments

[23] The results reported in the table are for $\bar{r}_t$ calculated with all countries included and with GDP-based weights. The conclusions drawn from Table 1 are unchanged when $\bar{r}_t$ is calculated using alternative approaches. We refer to Section 3.5 for details on the different approaches to calculate $\bar{r}_t$.

when working with $\bar{r}_t$ and $r_{i,t-2}$ (these results are not reported but available upon request).

We further note that the results reported in columns six and seven of the table are somewhat misleading. While, as noted, the weakness of these instrument sets is strongly rejected, they include an instrument, i.e., $r_{i,t-2}$, that is not significant. This implies that the instrument set of column six (consisting of $r_{i,t-1}$ and $r_{i,t-2}$) is in fact equivalent to the instrument set of column two (consisting only of $r_{i,t-1}$) implying a model that is just-identified rather than overidentified. And this implies that the instrument set of column seven (consisting of $\bar{r}_t$, $r_{i,t-1}$ and $r_{i,t-2}$) is in fact equivalent to the set of column four (consisting only of $\bar{r}_t$ and $r_{i,t-1}$) implying a model that is overidentified with one degree of freedom rather than overidentified with two degrees of freedom. Finally, the results reported in column eight of the table show that the effective F statistic cannot reject that an instrument set that includes only lags two and three of r_{it} is weak even though the coefficients on these variables are both individually significant. This is in line with the findings of Yogo (2004) and Montiel Olea and Pflueger (2013) and confirms that the practice of using instrument sets that consist of multiple deeper lags to estimate the EIS may be subject to severe weak instruments problems.

4.2 Estimation without common factors

We now report the results of estimating eq.(10) - i.e., our specification without common factors - using the variables $\bar{r}_t$ and $r_{i,t-1}$ as instruments. Table 2 first presents the mean-group results of the first-stage regressions of r_{it} on the instruments. Each column corresponds to a different calculation of the cross-sectional average $\bar{r}_t$, i.e., GDP-weighted or unweighted and with country i included or excluded from the cross-sectional average used as an instrument in the per country regression for i. We refer to Section 3.5 for details. We note that the results in column one correspond to the results presented earlier in column four of Table 1. From the table, we confirm the relevance of both instruments, irrespective of the calculation of $\bar{r}_t$. Using the effective F test detailed in Section 3.2.1 above, we strongly reject that the instrument set is weak. Moreover, the coefficients on both instruments considered individually are also highly significant.

Table 1: Mean-group results of regressing r_{it} on different instrument sets

	Dependent variable r_{it}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\bar{r}_t$	1.034			0.756	0.958		0.750	
	(0.139)			(0.151)	(0.141)		(0.143)	
$r_{i,t-1}$		0.423		0.341		0.487	0.402	
		(0.067)		(0.066)		(0.081)	(0.075)	
$r_{i,t-2}$			0.206		0.146	-0.071	-0.069	0.234
			(0.062)		(0.060)	(0.056)	(0.052)	(0.059)
$r_{i,t-3}$								-0.075
								(0.033)
F	15.122	31.007	9.430	22.822	12.827	18.341	17.099	4.192
effective F	15.124	28.973	9.947	17.759	10.447	14.965	15.091	3.727
30% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.045]
20% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.155]
10% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]	[0.032]	[0.000]	[0.000]	[0.658]

Notes: The table reports the mean-group results of estimating the first stage regression of r_{it} on instruments. Estimation is based on panel data for fifteen countries over the period $1950-2015$. Reported in the upper panel are the mean-group results obtained from country-by-country OLS estimation of the regression of the return r_{it} on a constant, on the cross-country average return $\bar{r}_t$ and/or on one ore more lags of the return $r_{i,t-l}$ (with $l=1,2,3$). Standard errors based on eq.(12) are in parentheses. Mean-group results for the constant are not reported. The cross-sectional average $\bar{r}_t$ used in all reported estimations is calculated as $\bar{r}_t=\sum_{i=1,N} w_{it} r_{it}$ with weight w_{it} given by country i's PPP-adjusted real GDP in period t divided by total PPP-adjusted real GDP in period t of all countries in the sample. Reported in the lower panel are the cross-country averages of the regular F statistic and the effective F statistic of Montiel Olea and Pflueger (2013). The null hypothesis of weak instruments based on the effective F is the hypothesis that the bias of the IV estimator relative to that of a 'worst-case' benchmark (like OLS) exceeds the threshold τ with $\tau=0.3,0.2,0.1$. More details on the calculation are provided in the text and in Montiel Olea and Pflueger (2013). Between square brackets are the harmonic means of the country-specific p-values (HMP) for the effective F. With $N=15$, the critical value of the harmonic mean p-value for a test at the 5% level of significance is about 0.04. We refer to Wilson (2019) for details.

Table 2: Mean-group results first-stage regression of r_{it} on $\bar{r}_t$ and $r_{i,t-1}$ (no common factors)

	(1)	(2)	(3)	(4)
$\bar{r}_t$	0.756	0.543	0.766	0.476
	(0.151)	(0.145)	(0.137)	(0.114)
$r_{i,t-1}$	0.341	0.378	0.325	0.366
	(0.066)	(0.072)	(0.068)	(0.068)
F	22.822	18.664	25.928	20.845
effective F	17.759	16.670	21.169	17.942
30% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]
20% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]
10% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]
$\bar{r}_t$	weighted	weighted	unweighted	unweighted
	i included	i excluded	i included	i excluded

Notes: The table reports the mean-group results of per country OLS estimation of the first stage regression of r_{it} on a constant, on the cross-sectional average $\bar{r}_t$ and on the first lag $r_{i,t-1}$. Estimation is based on panel data for fifteen countries over the period $1950-2015$. Standard errors based on eq.(12) are in parentheses. Mean-group results for the constant are not reported. Every column corresponds to a different calculation of the cross-sectional average $\bar{r}_t$. If all countries are included in the calculation of $\bar{r}_t$ ('i included'), we have $\bar{r}_t = \sum_{j=1,N} w_{jt} r_{jt}$. If country i is excluded from the cross-sectional average $\bar{r}_t$ used in the regression for country i ('i excluded'), we have $\bar{r}_t^{(-i)} = \sum_{j=1,N(j\neq i)} w_{jt} r_{jt}$ (note that in this case the variable $\bar{r}_t$ differs for every country i). For 'weighted', the weight w_{jt} is given by country j's PPP-adjusted real GDP in period t divided by total PPP-adjusted real GDP in period t of all countries included in the summation. For 'unweighted', w_{jt} equals one over the number of countries included in the summation ($\forall j, t$). For details concerning the F statistics reported in the table, we refer to the notes to Table 1.

Table 3 then reports the mean-group estimates and corresponding standard errors calculated from the IV estimates of ψ_i obtained when estimating eq.(10) per country (using $\bar{r}_t$ and $r_{i,t-1}$ as instruments). As before, each column in the table corresponds to a different calculation of the instrument $\bar{r}_t$. With respect to diagnostics, apart from the Sargan-Hansen overidentifying restrictions test which is discussed in Section 3.2.2 above, the table also reports the Cumby and Huizinga (1992) autocorrelation statistic which tests the null hypothesis of no autocorrelation in the error term against the alternative that the autocorrelations of the error term are nonzero at lags greater than zero. This autocorrelation test is particularly suitable as it is applicable when using estimators other than OLS such as IV and GMM, while it is valid also if the errors are conditionally heteroskedastic (see Cumby and Huizinga, 1992, for details). Additionally, we report the Durbin-Wu-Hausman endogeneity test (see Nakamura and Nakamura, 1981) which tests the null hypothesis that the regressor r_{it} is exogenous.[24] As with the F tests and the Sargan-Hansen

[24]This test adds the residuals of the first-stage regression for r_{it} to the second stage regression for Δc_{it}. The Durbin-

test, we calculate these statistics per country and report the cross-country averages. And as discussed in Section 3.4, the p-values reported in the table are calculated as the harmonic means of the per country p-values (HMP) of the tests.

Table 3: Mean-group results estimation of regression eq.(10)

	(1)	(2)	(3)	(4)
r_{it} (ψ)	0.258	0.315	0.285	0.397
	(0.042)	(0.102)	(0.055)	(0.160)
Sargan-Hansen	3.501	3.786	3.826	4.287
$[p - val.]$	[0.000]	[0.000]	[0.000]	[0.000]
Cumby-Huizinga	8.469	7.790	8.292	7.784
$[p - val.]$	[0.014]	[0.015]	[0.012]	[0.014]
Durbin-Wu-Hausman	4.180	3.342	4.150	2.495
$[p - val.]$	[0.000]	[0.000]	[0.000]	[0.027]
Instruments	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$
$\bar{r}_t$	weighted	weighted	unweighted	unweighted
	i included	i excluded	i included	i excluded

Notes: The table reports the mean-group results of per country IV estimation of eq.(10). The instrument set consists of a constant, $\bar{r}_t$ and $r_{i,t-1}$. Estimation is based on panel data for fifteen countries over the period $1950 - 2015$. Standard errors based on eq.(12) are in parentheses. Mean-group results for the constant are not reported. Every column corresponds to a different calculation of the cross-sectional average $\bar{r}_t$, i.e., weighted or unweighted and country i included or excluded. We refer to the notes to Table 2 for more details. The Sargan-Hansen test reported is the average of the country-specific Sargan-Hansen overidentifying restrictions statistics that test the null hypothesis of the joint validity of the instruments used (see Sargan, 1958; Hansen, 1982). The Cumby-Huizinga test shows the average of the individual countries' Cumby and Huizinga (1992) autocorrelation tests, testing the null hypothesis of no autocorrelation against the alternative that the autocorrelations of the error term are nonzero at lags greater than zero. The Durbin-Wu-Hausman statistic shows the average of the individual countries' Durbin-Wu-Hausman endogeneity tests (see Nakamura and Nakamura, 1981), testing the null hypothesis that the regressor r_{it} is exogenous. Between square brackets are the harmonic means of the country-specific p-values (HMP) of these tests. With $N = 15$, the 5% critical value of the harmonic mean p-value is about 0.04. We refer to Wilson (2019) for details.

The table reports estimates for ψ, i.e., the average impact of housing returns r_{it} on aggregate consumption growth Δc_{it}. These estimates - which can be considered lower bounds to the EIS σ - are, as expected, positive. They take on values between 0.25 and 0.40 and are highly significant. The Durbin-Wu-Hausman tests reject the exogeneity of the regressor r_{it} in all four cases which supports the conducted

Wu-Hausman endogeneity test is in fact a Wald test that checks the null hypothesis that the coefficient on these residuals is zero.

IV estimation approach. Nonetheless, the results of the Cumby-Huizinga autocorrelation test and those of the Sargan-Hansen overidentifying restrictions test cast doubt on the reliability of the IV estimates of ψ reported in the table. Indeed, the null hypothesis that there is no autocorrelation is always rejected. Given our use of the lagged return $r_{i,t-1}$ as an instrument this may be indicative of correlation between the instruments and the error term. The latter is confirmed by the results of the Sargan-Hansen test which strongly rejects the validity of the instruments in all cases. The empirical model consisting of eq.(10) and the instruments $\bar{r}_t$ and $r_{i,t-1}$ is therefore rejected by the data and we must consider another specification.

4.3 Estimation with common factors

4.3.1 Main results

As discussed in the previous section, we reject the overidentifying restriction when estimating eq.(10) using the relevant instruments $\bar{r}_t$ and $r_{i,t-1}$. To deal with this, our approach is to take out the persistence and commonality from the error term of eq.(10) that is likely to induce correlation with the instruments by controlling for the presence of persistent and common factors. Hence, we estimate eq.(11) which now includes, as additional regressors, proxies for an international (income) growth factor f_t^{inc} and for an international financial factor f_t^{fin} as detailed in Section 3.3. As the variables f_t^{inc} and f_t^{fin} can be considered exogenous, the first stage now regresses the endogenous variable r_{it} on the instruments $\bar{r}_t$, $r_{i,t-1}$, f_t^{inc} and f_t^{fin} (and a constant). The mean-group first-stage regression results are presented in Table 4. Each column in the table corresponds to a different calculation of the cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} (GDP-weighted or unweighted, country i included or excluded in the cross-sectional average used in the regression for country i). We refer to section 3.5 for details.

Table 4: Mean-group results first-stage regression of r_{it} on $\bar{r}_t$ and $r_{i,t-1}$ (with common factors)

	(1)	(2)	(3)	(4)
$\bar{r}_t$	0.558	0.277	0.858	0.262
	(0.189)	(0.161)	(0.175)	(0.146)
$r_{i,t-1}$	0.320	0.354	0.311	0.339
	(0.066)	(0.071)	(0.068)	(0.069)
f_t^{inc}	0.209	0.289	-0.053	0.292
	(0.120)	(0.118)	(0.238)	(0.222)
f_t^{fin}	0.158	0.218	-0.073	0.171
	(0.114)	(0.108)	(0.087)	(0.109)
F	15.752	13.853	15.783	13.157
effective F	14.341	13.847	15.716	13.508
30% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]
20% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]
10% thresh. $[p-val.]$	[0.000]	[0.000]	[0.000]	[0.000]
$\bar{r}_t$, f_t^{inc}, f_t^{fin}	weighted	weighted	unweighted	unweighted
	i included	i excluded	i included	i excluded

Notes: The table reports the mean-group results of per country OLS estimation of the first stage regression of r_{it} on a constant, on the cross-sectional average $\bar{r}_t$, on the first lag $r_{i,t-1}$ and on the factors f_t^{inc} and f_t^{fin}. Estimation is based on panel data for fifteen countries over the period 1950−2015. Standard errors based on eq.(12) are in parentheses. Mean-group results for the constant are not reported. Every column corresponds to a different calculation of the cross-sectional average $\bar{x}_t$ where $\bar{x}_t$ equals either $\bar{r}_t$ or f_t^{inc} or f_t^{fin}. If all countries are included in the calculation of $\bar{x}_t$ ('i included'), we have $\bar{x}_t = \sum_{j=1,N} w_{jt}x_{jt}$ where x_{jt} equals either r_{jt} (when $\bar{x}_t = \bar{r}_t$) or Δy_{jt} (when $\bar{x}_t = f_t^{inc}$) or $\Delta cred_{jt}$ (when $\bar{x}_t = f_t^{fin}$) with r_{jt} the real housing return, y_{jt} the log of per capita real GDP and $cred_{jt}$ the log of per capita real credit. If country i is excluded from the cross-sectional average $\bar{x}_t$ used in the regression for country i ('i excluded'), we have $\bar{x}_t^{(-i)} = \sum_{j=1,N(j\neq i)} w_{jt}x_{jt}$ with x_{jt} as before (note that in this case the variable $\bar{x}_t$ differs for every country i). For 'weighted', the weight w_{jt} is given by country j's PPP-adjusted real GDP in period t divided by total PPP-adjusted real GDP in period t of all countries included in the summation. For 'unweighted', w_{jt} equals one over the number of countries included in the summation ($\forall j, t$). For details concerning the F statistics reported in the table, we refer to the notes to Table 1.

From the results reported in the table, we note that we still strongly reject that the instrument sets considered are weak even though the F and effective F statistics are of considerably lower magnitude compared to those reported in Table 2. Both instruments $\bar{r}_t$ and $r_{i,t-1}$ are still individually significant as well, even though the impact and significance of the instrument $\bar{r}_t$ for r_{it} is now generally lower compared to the results of Table 2 (with the exception of the case reported in column three of the table). This is not surprising as the common factors f_t^{inc} and f_t^{fin} now capture part of the commonality in r_{it} which previously was only controlled for by $\bar{r}_t$. We also note that, with the exception of the results in column three which are insignificant, the impacts of f_t^{inc} and f_t^{fin} on r_{it} are positive, i.e., high international GDP or credit growth coincides with high domestic housing returns.

Table 5: Mean-group results estimation of regression eq.(11)

	(1)	(2)	(3)	(4)
r_{it} (ψ)	0.131	0.153	0.108	0.137
	(0.039)	(0.041)	(0.026)	(0.034)
f_t^{inc} (γ^{inc})	0.338	0.288	0.491	0.422
	(0.067)	(0.055)	(0.076)	(0.066)
f_t^{fin} (γ^{fin})	0.065	0.065	0.041	0.028
	(0.031)	(0.032)	(0.027)	(0.030)
Sargan-Hansen	1.067	1.011	1.798	1.335
$[p-val.]$	[0.192]	[0.221]	[0.072]	[0.138]
Cumby-Huizinga	7.373	7.126	6.040	6.528
$[p-val.]$	[0.112]	[0.118]	[0.220]	[0.167]
Durbin-Wu-Hausman	2.825	2.053	1.364	1.551
$[p-val.]$	[0.001]	[0.020]	[0.148]	[0.075]
Instruments	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$
	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}
$\bar{r}_t$, f_t^{inc}, f_t^{fin}	weighted	weighted	unweighted	unweighted
	i included	i excluded	i included	i excluded

Notes: The table reports the mean-group results of per country IV estimation of eq.(11). The instrument set consists of a constant, $\bar{r}_t$, $r_{i,t-1}$, f_t^{inc} and f_t^{fin}. Estimation is based on panel data for fifteen countries over the period $1950-2015$. Standard errors based on eq.(12) are in parentheses. Mean-group results for the constant are not reported. Every column corresponds to a different calculation of the cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin}, i.e., weighted or unweighted and country i included or excluded. We refer to the notes to Table 4 for more details. The Sargan-Hansen test is the average of the country-specific overidentifying restrictions statistics that test the null of the joint validity of the instruments used (see Sargan, 1958; Hansen, 1982). The Cumby-Huizinga test shows the average of the individual countries' Cumby and Huizinga (1992) autocorrelation tests, testing the null of no autocorrelation against the alternative that the autocorrelations of the error term are nonzero at lags greater than zero. The Durbin-Wu-Hausman statistic shows the average of the individual countries' Durbin-Wu-Hausman endogeneity tests (see Nakamura and Nakamura, 1981), testing the null hypothesis that the regressor r_{it} is exogenous. Between square brackets are the harmonic means of the country-specific p-values (HMP) of these tests. With $N = 15$, the 5% critical value of the harmonic mean p-value is about 0.04. We refer to Wilson (2019) for details.

Table 5 then presents the mean-group results of estimating eq.(11), again with each column corresponding to a different approach used to calculate the cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} (see

Section 3.5 for details). As noted above, the estimates for ψ, i.e., the average impact of housing returns r_{it} on aggregate consumption growth Δc_{it}, can be considered lower bounds to the EIS σ. We note that the country-specific IV estimates ψ_i on which the mean-group estimates of ψ are based are reported in Appendix A. Compared to the regression results for the specification without common factors eq.(10) reported in the previous section, the estimates of ψ are now of lower magnitude. They are still positive and highly significant, however, and take on values between 0.10 and 0.15. Additionally, the coefficients on the common international growth factor are positive and highly significant in all four reported cases in the table, i.e., high international output growth leads to high domestic consumption growth. The coefficients on the common international financial factor are also positive, suggesting high international credit growth leads to high domestic consumption growth, but they are significant only in columns one and two of the table. Furthermore, we note that the autocorrelation and overidentifying restrictions tests reported in the table are supportive of the regression specification with common factors. First, the reported Cumby-Huizinga tests cannot reject the null hypothesis that there is no autocorrelation in the residuals of eq.(11). This suggests that including persistent common factors in the regression equation captures the persistence that is present in the error term of eq.(10), i.e., the specification without common factors. More importantly, the reported Sargan-Hansen tests support the validity of the used instrument sets as, different from what we see when estimating the specification without common factors, the overidentifying restriction is not rejected. Again, this suggests that including persistent common factors in the regression equation has taken out persistence and commonality from the error term, thereby decreasing its correlation with the instruments.

To conclude, the specification that includes common factors is supported by the data and the estimates for ψ reported in Table 5 appear to be solid. In Section 4.4 below, we therefore calculate EIS estimates based on the country-specific IV estimates ψ_i that underlie the mean-group estimates of Table 5. In the next section, however, we first present some additional results obtained from estimating eq.(11).

4.3.2 Additional results

As additional results, we consider, in turn, estimation over different subperiods, estimation using GMM at the country level instead of IV, and estimation using the returns commonly used in the literature to estimate the EIS, i.e., returns on equity and Treasury bills.

Subperiods

First, we estimate eq.(11) over subperiods. Following, among others, Kose et al. (2012), we consider the pre-globalization period of $1950-1985$ and the globalization period of $1985-2015$. While the literature has been largely silent on the potential long-term structural evolution of the EIS parameter, our use of a relatively long sample period allows to check whether our estimates of ψ (and, therefore, of the EIS) differ across both subperiods. There are a priori reasons to suspect that the EIS has increased over time. Some of the reasons are equivalent to those that have been put forward to explain potentially different values for the EIS across countries (see Havranek et al., 2015, for an overview). First, as countries have become richer over time, the fraction of necessities in the consumption bundle of households has decreased. Necessities are harder to substitute across time. Second, financial liberalization and integration have increased asset market participation rates and asset market participants are more willing to substitute consumption across time. Finally, liberalization and integration have also reduced credit constraints and less constrained consumers have more possibilities to substitute consumption across time.

The mean-group results of estimating eq.(11) over both subperiods are presented in Table 6. We report only the results for cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} that are GDP-weighted and that include all countries, but the findings reported in the table generally also hold when we consider alternative approaches to calculate these variables. For both considered subperiods, the diagnostics reported in the table support the estimated specifications. In particular, we reject that the instruments are weak (irrelevant) and we do not reject that they are exogenous (valid). The model for the globalization period performs better as it is characterized by higher F statistics and a lower Sargan-Hansen statistic. Interestingly, our estimates for ψ support our a priori expectations regarding the evolution of the EIS over time. While the estimate for ψ is positive for both subperiods, it is substantially larger - i.e., twice as large - and significant only when estimated over the globalization period (1985-2015), i.e., a period characterized by high economic and financial liberalization and integration.

Table 6: Mean-group results estimation of regression eq.(11) over subperiods

	(1) Period $1950-1985$	(2) Period $1985-2015$
r_{it} (ψ)	0.076	0.155
	(0.059)	(0.036)
f_t^{inc} (γ^{inc})	0.284	0.372
	(0.072)	(0.073)
f_t^{fin} (γ^{fin})	0.067	-0.012
	(0.039)	(0.050)
F	6.659	12.917
effective F	8.264	14.341
30% thresh. [$p-val.$]	[0.000]	[0.000]
20% thresh. [$p-val.$]	[0.000]	[0.000]
10% thresh. [$p-val.$]	[0.000]	[0.000]
Sargan-Hansen	1.546	0.732
[$p-val.$]	[0.127]	[0.303]
Cumby-Huizinga	5.410	5.312
[$p-val.$]	[0.289]	[0.276]
Durbin-Wu-Hausman	3.171	2.070
[$p-val.$]	[0.002]	[0.011]
Instruments	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$
	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}
$\bar{r}_t$, f_t^{inc}, f_t^{fin}	weighted	weighted
	i included	i included

Notes: The table reports the mean-group results of per country IV estimation of eq.(11) over the subperiods $1950-1985$ and $1985-2015$. The instrument set consists of a constant, $\bar{r}_t$, $r_{i,t-1}$, f_t^{inc} and f_t^{fin}. Estimation is based on panel data for fifteen countries. Standard errors based on eq.(12) are in parentheses. Mean-group results for the constant are not reported. Results reported are for cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} that are GDP-weighted and that include country i. We refer to the notes to Table 4 for details. Reported are the cross-country averages of the regular F statistic and the effective F statistic of Montiel Olea and Pflueger (2013). The null hypothesis of weak instruments based on the effective F is the hypothesis that the bias of the IV estimator relative to that of a 'worst-case' benchmark (like OLS) exceeds the threshold τ with $\tau = 0.3, 0.2, 0.1$. The Sargan-Hansen test is the average of the country-specific overidentifying restrictions statistics that test the null of the joint validity of the instruments used (see Sargan, 1958; Hansen, 1982). The Cumby-Huizinga test shows the average of the individual countries' Cumby and Huizinga (1992) autocorrelation tests, testing the null of no autocorrelation against the alternative that the autocorrelations of the error term are nonzero at lags greater than zero. The Durbin-Wu-Hausman statistic shows the average of the individual countries' Durbin-Wu-Hausman endogeneity tests (see Nakamura and Nakamura, 1981), testing the null hypothesis that the regressor r_{it} is exogenous. Between square brackets are the harmonic means of the country-specific p-values (HMP) of these tests. With $N = 15$, the 5% critical value of the harmonic mean p-value is about 0.04. We refer to Wilson (2019) for details.

GMM

Next, we estimate eq.(11) using GMM at the country level as this method has often been applied in the literature to obtain EIS estimates (see e.g., Yogo, 2004). The mean-group estimation results of this exercise are presented in Appendix B. The results and conclusions based on GMM estimation generally coincide with and support those based on IV estimation.

Alternative returns

Finally, we estimate eq.(11) using the returns on the two assets that have generally been considered when estimating the EIS, namely equity and Treasury bills (see Havranek et al., 2015). The mean-group estimation results of this exercise are presented in Appendix C. The results presented and discussed in this appendix imply EIS estimates obtained from returns on equity or bills that are close to zero in magnitude (especially those obtained for equity) and not significantly from zero. These results are in accordance with other macro studies that estimate the EIS from equity or bill returns (see, in particular, Havranek, 2015, for a recent meta analysis).

4.4 EIS estimates

From the estimates of the impact of housing returns on aggregate consumption growth, we now calculate estimates for the EIS. The theory of Section 2 shows that the EIS parameter σ can generally be written as the impact ψ^k of asset k's return on aggregate consumption growth divided by the fraction of consumers who hold asset k, namely $\frac{J^k}{J}$ where J is the total number of consumers and J^k is the number of consumers who hold asset k. In addition to the estimates for ψ^k for the housing asset (discussed and presented in the previous sections), we therefore additionally need a proxy for the ratio $\frac{J^k}{J}$ for the housing asset. And given our per country approach to estimation, we prefer a country-specific proxy for this ratio. A variable that meets these requirements is the per country homeownership rate reported by, among others, the OECD. Table 7 presents the most recent homeownership rates for all countries in our sample.

Table 7: Homeownership rates for the countries in the sample

Country	Home ownership rate	Country	Home ownership rate
Australia	0.627	Norway	0.724
Denmark	0.527	Portugal	0.719
Finland	0.639	Spain	0.756
France	0.462	Sweden	0.580
Germany	0.438	Switzerland	0.374
Italy	0.713	UK	0.649
Japan	0.612	US	0.642
Netherlands	0.580		

Notes: Data for all countries but Japan are taken from the OECD Affordable Housing Database (reported figures are for 2019). For Japan, the number is from Statista.com (the reported figure is for 2018).

Per country EIS estimates $\widehat{\sigma}_i$ are then obtained as $\widehat{\sigma}_i = \frac{\widehat{\psi}_i}{\xi_i}$ with $\widehat{\psi}_i$ the per country IV estimate of the coefficient on the housing return r_{it} in eq.(11) and ξ_i the homeownership rate in country i. As, unfortunately, data for homeownership rates are not sufficiently available to calculate an average homeownership rate over the full sample period for every country, we instead use the most recent values for ξ_i, i.e., those reported in Table 7. As Andrews and Caldera Sánchez (2011), for instance, argue that homeownership rates have increased in recent decades in many OECD countries, using these relatively large recent values for ξ then implies estimates for σ that are relatively small, i.e., by using the most recent values for the homeownership rates, we may somewhat underestimate the true EIS. As in the literature EIS estimates are typically biased upward (see Havranek, 2015), we consider this less problematic than if the opposite were the case.

The mean-group estimates of the EIS parameter σ, calculated as the simple cross-country averages of $\frac{\widehat{\psi}_i}{\xi_i}$, are presented in Table 8.[25] We note that the columns in the table correspond to those of Table 5 with each column corresponding to a different set of regressors and instruments used when estimating the parameters ψ_i from eq.(11). From column one in the table, we report a baseline EIS estimate of 0.21. The EIS estimates reported in the other columns of the table are of similar magnitude. Moreover, all reported estimates are highly significant, i.e., at the 1% level. Hence, the reported EIS estimates - which, as noted above, may even be somewhat underestimated due to the use of relatively high values for the

[25] The standard error of the mean-group estimate $\widehat{\sigma} = \frac{1}{N}\sum_{i=1,N}\frac{\widehat{\psi}_i}{\xi_i}$ is given by $\widehat{se}(\widehat{\sigma}) = \sqrt{(\widehat{se}(\widehat{\psi}))^2\frac{1}{N}\sum_{i=1,N}\frac{1}{\xi_i^2}}$ where $\widehat{se}(\widehat{\psi})$ is the estimated standard error of the mean-group estimate $\widehat{\psi}$. This is obtained as follows. First, given $\sigma = \frac{1}{N}\sum_{i=1,N}\frac{\psi_i}{\xi_i}$ and given the assumption used in mean-group estimation that ψ_i are *iid* across countries, we write $V(\sigma) = \frac{1}{N^2}\sum_{i=1,N}\frac{1}{\xi_i^2}V(\psi_i)$. Second, with ψ_i being *iid* across countries, the variance of $\psi = \frac{1}{N}\sum_{i=1,N}\psi_i$ is given by $V(\psi) = \frac{V(\psi_i)}{N}$ from which we have $V(\psi_i) = NV(\psi)$. After substituting this into the expression for $V(\sigma)$, we have $V(\sigma) = V(\psi)\frac{1}{N}\sum_{i=1,N}\frac{1}{\xi_i^2}$ or $se(\sigma) = \sqrt{se(\psi)^2\frac{1}{N}\sum_{i=1,N}\frac{1}{\xi_i^2}}$.

homeownership rates - are significantly larger than the EIS estimates typically reported by other macro studies. As noted by Havranek (2015), the latter equal zero on average. And while our EIS estimates are smaller than the typical values for the EIS obtained by micro studies which are in the range $0.3 - 0.4$ (see also Havranek, 2015), they are larger than but generally not too different from the recent micro-based estimates for the UK reported by Best et al. (2020) who, interestingly, also focus on the housing market, in particular, on discrete jumps in UK mortgage interest rates as a source of variation to identify the EIS.[26]

Table 8: Mean-group EIS estimates

	Columns correspond to columns of Table 5			
	(1)	(2)	(3)	(4)
EIS (σ)	0.210	0.241	0.175	0.220
	(0.069)	(0.072)	(0.046)	(0.059)

Notes: The table reports mean-group estimates of the elasticity of intertemporal substitution (EIS) based on the estimates of ψ_i obtained when estimating eq.(11). Standard errors are in parentheses. The cases in the four columns correspond to the cases in the four columns of Table 5, i.e., estimation of eq.(11) using different approaches to calculate the cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin}. The mean-group estimate for the EIS σ is calculated as $\widehat{\sigma} = \frac{1}{N}\sum_{i=1,N}\frac{\widehat{\psi}_i}{\xi_i}$ with $\widehat{\psi}_i$ the per country IV estimate of the coefficient on the housing return r_{it} in eq.(11) and ξ_i the home ownership rate in country i as reported in Table 7. The standard error of the mean-group estimate $\widehat{\sigma}$ is calculated as $\widehat{se}(\widehat{\sigma}) = \sqrt{(\widehat{se}(\widehat{\psi}))^2 \frac{1}{N}\sum_{i=1,N}\frac{1}{\xi_i^2}}$ where $\widehat{se}(\widehat{\psi})$ is the estimated standard error of the mean-group estimate $\widehat{\psi}$ reported in Table 5.

To conclude, our empirical approach - i.e., our choice of asset (housing), our focus on asset holdings (homeownership rates) that stems from the heterogeneous agent setting that we consider, and the scrutiny that we give to the instruments (choice, relevance, validity) - delivers EIS estimates that are not typical for a macro-based approach but, rather, are in line with micro-based estimates.

5 Conclusions

The elasticity of intertemporal substitution (EIS) is one of the main parameters driving consumption behavior but, unfortunately, there is no agreement in the economic and financial literature as to which values are appropriate for this parameter. Motivated by the fact that asset choice has received relatively

[26]For the UK, we find IV estimates of ψ_i between 0.105 and 0.122 (see Table A-1 in Appendix A). Combined with a homeownership rate of 0.649 (see Table 7), we therefore obtain EIS estimates for the UK that lie between 0.162 and 0.188. The EIS estimates for the UK reported by Best et al. (2020) vary between 0.08 and 0.17 (depending on the structural assumptions made in their estimations).

little attention in an otherwise extensive empirical literature that tries to pinpoint the EIS, this paper proposes the use of housing returns to estimate this parameter. As housing is the main asset for the majority of households, it can be argued that returns on housing are better suited to estimate the EIS than the asset returns typically considered in the literature, i.e., equity and bill returns. We therefore use recently constructed data on housing returns for fifteen advanced economies over the postwar period $1950 - 2015$ to estimate the impact of these returns on consumption growth and to obtain EIS estimates. Theoretically, we base our estimations on a general heterogeneous agent model from which we derive an estimable equation that regresses aggregate consumption growth on aggregate housing returns. Our theoretical framework provides some important additional methodological insights compared to the representative agent settings that are typically considered in macro-level EIS studies. In particular, we emphasise the importance of taking into account the extent to which consumers hold the asset considered - i.e., housing - in the estimation of the EIS. Methodologically, because of omitted variables, reverse causality and measurement error considerations, we estimate our regression equation using an instrumental variables approach where the choice of instruments, their relevance and validity are carefully scrutinized. With respect to our instrument choice, we propose to instrument the domestic housing return using the one-year lagged domestic housing return and a cross-country average of international housing returns, thereby exploiting both the temporal and spatial dimensions of our dataset. We subsequently combine the per country regression estimates for the impact of housing returns on consumption growth with per country house ownership data to calculate per country and panel-wide mean-group EIS estimates.

Our findings show that both proposed instruments are strong. Moreover, they are valid (exogenous) once we include persistent common international growth and financial factors as additional regressors in our regression equation. On average, across countries, we find a highly statistically significant baseline elasticity estimate of about 0.21. Our EIS estimates are generally larger than the small near-zero estimates typically obtained by macro studies that link aggregate consumption to the aggregate returns on less widely held assets like equity and bills. And while our EIS estimates are generally smaller than the typical EIS values obtained by micro studies that consider the consumption only of particular asset holders, some micro-based studies report EIS estimates that are of similar magnitude or even smaller than ours. Hence, our empirical approach delivers EIS estimates that are not typical for a macro-based approach but, rather, are in line with micro-level estimates. Our findings reaffirm that intertemporal substitution in consumption is both economically and statistically significant.

References

Andrews, D. and Caldera Sánchez, A. (2011). The evolution of homeownership rates in selected OECD countries: demographic and public policy influences. *OECD Journal: Economic Studies*, 2011/1:207–243.

Andrews, I. and Stock, J. (2018). Weak instruments and what to do about them. *Lecture notes, NBER Summer Institute Methods Lectures.*

Attanasio, O. and Weber, G. (1993). Consumption growth, the interest rate and aggregation. *Review of Economic Studies*, 60:631–649.

Best, M., Cloyne, J., Ilzetzki, E., and Kleven, H. (2020). Estimating the elasticity of intertemporal substitution using mortgage notches. *Review of Economic Studies*, 87:656–690.

Bond, S., Leblebicioglu, A., and Schiantarelli, F. (2010). Capital accumulation and growth: a new look at the empirical evidence. *Journal of Applied Econometrics*, 25:1073–1099.

Carroll, C., Otsuka, M., and Slacalek, J. (2011). How large are housing and financial wealth effects? A new approach. *Journal of Money, Credit and Banking*, 43(1):55–79.

Case, K., Quigley, J., and Shiller, R. (2005). Comparing wealth effects: the stock market versus the housing market. *Advances in Macroeconomics*, 5(1):1–32.

Cumby, R. E. and Huizinga, J. (1992). Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica*, 60(1):185–195.

Dacy, D. and Hasanov, F. (2011). A finance approach to estimating consumption parameters. *Economic Inquiry*, 49(1):122–154.

Fisher, R. (1925). *Statistical methods for research workers.* Oliver and Boyd, Edinburgh.

Forbes, K. and Chinn, M. (2004). A decomposition of global linkages in financial markets over time. *Review of Economics and Statistics*, 86(3):705–722.

Furceri, D. and Loungani, P. (2018). The distributional effects of capital account liberalization. *Journal of Development Economics*, 130:127–144.

Ha, J., Kose, M., Otrok, C., and Prasad, E. (2020). Global macro-financial cycles and spillovers. *NBER Working Paper*, (26798).

Hall, R. (1988). Intertemporal substitution in consumption. *Journal of Political Economy*, 96:339–57.

Hansen, L. (1982). Large sample properties of generalized method of moment estimators. *Econometrica*, 50:1029–1054.

Harvey, C. (1991). The world price of covariance risk. *Journal of Finance*, XLVI:111–157.

Havranek, T. (2015). Measuring intertemporal substitution: the importance of method choices and selective reporting. *Journal of the European Economic Association*, 13(6):1180–1204.

Havranek, T., Horvath, R., Irsova, Z., and Rusnak, M. (2015). Cross-country heterogeneity in intertemporal substitution. *Journal of International Economics*, 96:100–118.

Hirata, H., Kose, M., Otrok, C., and Terrones, M. (2013). Global house price fluctuations: synchronization and determinants. *NBER International Seminar on Macroeconomics*, 9(1):119–166.

Jordà, O., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. (2019). The rate of return on everything, 1870-2015. *Quarterly Journal of Economics*, 134(3):1225–1298.

Jordà, O., Schularick, M., and Taylor, A. (2015). Betting the house. *Journal of International Economics*, 96:2–18.

Jordà, O., Schularick, M., and Taylor, A. (2016). Macrofinancial history and the new business cycle facts. In Eichenbaum, M. and Parker, J., editors, *NBER Macroeconomics Annual 2016*, volume 31. University of Chicago Press, Chicago.

Kose, M., Otrok, C., and Prasad, E. (2012). Global business cycles: convergence or decoupling? *International Economic Review*, 53(2):511–538.

Kose, M., Otrok, C., and Whiteman, C. (2003). International business cycles: world, region, and country-specific factors. *American Economic Review*, 93(4):1216–1239.

Montiel Olea, J. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business and Economic Statistics*, 31(3):358–369.

Mulligan, C. (2002). Capital, interest and aggregate intertemporal substitution. *NBER Working Paper*, (9373).

Nakamura, A. and Nakamura, M. (1981). On the relationships among several specification error tests presented by Durbin, Wu, and Hausman. *Econometrica*, 49(6):1583–1588.

Neely, C., Roy, A., and Whiteman, C. (2001). Risk aversion versus intertemporal substitution: a case study of identification failure in the intertemporal consumption capital asset pricing model. *Journal of Business and Economic Statistics*, 19(4):395–403.

Newey, W. and West, K. (1987). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Parker, J. and Preston, B. (2005). Precautionary savings and consumption fluctuations. *American Economic Review*, 95(4):1119–1143.

Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79–113.

Pesaran, M. H., Smith, R., and Im, K.-S. (1996). Dynamic linear models for heterogeneous panels. In Matyas, L. and Sevestre, P., editors, *The econometrics of panel data*, pages 145–195. Kluwer Academic Publishers, Dordrecht, second edition.

Sargan, J. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, 26:393–415.

Sommer, M. (2007). Habit formation and aggregate consumption dynamics. *The B.E. Journal of Macroeconomics - Advances*, 7:article 21.

Vissing-Jorgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4):825–853.

Wilson, D. (2019). The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences*, 116(4):1195–1200.

Yogo, M. (2004). Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics*, 86(3):797–810.

Zeldes, S. (1989). Consumption and liquidity constraints: an empirical investigation. *Journal of Political Economy*, 97(2):305–346.

Appendices

Appendix A Per country estimation results

This appendix presents per country results. Table A-1 presents the IV estimates for the main parameter of interest, ψ_i. These are obtained when estimating eq.(11). These estimates are used in the calculation of the mean-group estimates for ψ reported in Table 5. As with Table 5, each column corresponds to a different calculation of the cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} used in estimation. The ψ_i estimates, which can be considered lower bounds to the EIS, are generally - as expected - positive. Only for Norway and Switzerland do we find consistently - i.e., across all columns - negative values for ψ_i. In the extant literature which uses equity and/or bill returns to estimate the EIS, the finding of a negative impact of the rate of return on consumption growth is much more common, however.[1] Additionally, the reported estimates per country are rather similar across columns, i.e., the approach used to calculate the cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} has a limited impact on the estimates of ψ_i. An interesting exception, however, are the ψ_i estimates for the US reported in columns one and two of the table, i.e., estimates obtained using GDP-weighted cross-sectional averages for $\bar{r}_t$, f_t^{inc} and f_t^{fin}. Since the US contribute the most to total GDP of all countries in the sample, they receive a large weight in the calculation of the GDP-based averages. When estimating eq.(11) using US data and GDP-based weights, it therefore makes sense to remove the US from the calculation of these averages. As it turns out, this also provides more convincing results, i.e., the ψ_i estimate for the US is negative when the US is included in the calculation of the cross-sectional averages (column one) while it is positive when the US are excluded from the calculation of the cross-sectional averages (column two).

Appendix B GMM estimation

This appendix reports GMM-based estimation results. While the results reported in the paper are based on IV estimation of the regression specifications at the country level, the extant literature has often used a GMM approach to estimate the EIS (see e.g., Yogo, 2004). The set-up of Table A-2 below fully coincides with that of Table 5 in the text but, rather than reporting mean-group results that are based on per country IV estimation, it reports mean-group results that are based on per country GMM estimation. Upon comparing both tables, we conclude that the results based on GMM generally coincide with and support those based on IV.

[1] When using equity and bill returns in our empirical set-up instead of housing returns, as discussed in Appendix C, we also often obtain negative estimates of ψ_i.

Table A-1: Per country IV estimates of ψ_i in regression eq.(11)

	(1)	(2)	(3)	(4)
Australia	-0.012	-0.029	0.038	0.047
	(0.152)	(0.187)	(0.115)	(0.182)
Denmark	0.256	0.263	0.153	0.253
	(0.147)	(0.148)	(0.115)	(0.113)
Finland	0.180	0.174	0.196	0.180
	(0.131)	(0.121)	(0.160)	(0.147)
France	0.081	0.085	0.034	0.041
	(0.035)	(0.033)	(0.037)	(0.036)
Germany	0.058	0.060	0.061	0.081
	(0.098)	(0.102)	(0.093)	(0.099)
Italy	0.199	0.342	0.206	0.270
	(0.091)	(0.160)	(0.170)	(0.747)
Japan	0.219	0.262	0.202	0.233
	(0.039)	(0.040)	(0.050)	(0.048)
Netherlands	0.218	0.216	0.193	0.210
	(0.049)	(0.047)	(0.053)	(0.049)
Norway	-0.101	-0.086	-0.102	-0.137
	(0.270)	(0.262)	(0.145)	(0.284)
Portugal	0.514	0.517	0.290	0.383
	(0.147)	(0.150)	(0.140)	(0.191)
Spain	0.133	0.163	0.100	0.213
	(0.120)	(0.150)	(0.088)	(0.140)
Sweden	0.163	0.174	0.089	0.124
	(0.059)	(0.068)	(0.040)	(0.047)
Switzerland	-0.043	-0.048	-0.002	-0.014
	(0.083)	(0.085)	(0.060)	(0.070)
UK	0.122	0.113	0.114	0.105
	(0.041)	(0.045)	(0.038)	(0.043)
US	-0.019	0.089	0.050	0.069
	(0.067)	(0.093)	(0.072)	(0.074)
Instruments	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$
	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}
$\bar{r}_t$, f_t^{inc}, f_t^{fin}	weighted	weighted	unweighted	unweighted
	i included	i excluded	i included	i excluded

Notes: Reported are the per country IV estimates of ψ_i in eq.(11). Heteroskedasticity- and autocorrelation-robust Newey-West standard errors are in parentheses (see Newey and West, 1987). Estimation occurs over the period $1950 - 2015$. The reported estimates are used to calculate the mean-group estimates reported in Table 5. See that table for details.

Table A-2: Results estimation of regression eq.(11) using GMM

	(1)	(2)	(3)	(4)
r_{it} (ψ)	0.133	0.151	0.112	0.124
	(0.039)	(0.041)	(0.026)	(0.033)
f_t^{inc} (γ^{inc})	0.341	0.289	0.486	0.436
	(0.063)	(0.053)	(0.073)	(0.072)
f_t^{fin} (γ^{fin})	0.062	0.061	0.044	0.024
	(0.033)	(0.035)	(0.023)	(0.030)
Sargan-Hansen	0.799	0.745	1.394	1.117
$[p - val.]$	[0.298]	[0.333]	[0.114]	[0.174]
Cumby-Huizinga	7.249	7.007	6.209	6.689
$[p - val.]$	[0.119]	[0.132]	[0.193]	[0.161]
Durbin-Wu-Hausman	2.825	2.053	1.364	1.551
$[p - val.]$	[0.001]	[0.020]	[0.148]	[0.075]
Instruments	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$
	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}
$\bar{r}_t$, f_t^{inc}, f_t^{fin}	weighted	weighted	unweighted	unweighted
	i included	i excluded	i included	i excluded

Notes: The table reports the mean-group results of per country GMM estimation of eq.(11). The optimal weighting matrix used to calculate the GMM estimates is robust to heteroskedasticity and autocorrelation (see Newey and West, 1987). The instrument set consists of a constant, $\bar{r}_t$, $r_{i,t-1}$, f_t^{inc} and f_t^{fin}. Estimation is based on panel data for fifteen countries over the period $1950 - 2015$. Standard errors based on eq.(12) are in parentheses. Mean-group results for the constant are not reported. Every column corresponds to a different calculation of the cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin}, i.e., weighted or unweighted and country i included or excluded. We refer to the notes to Table 4 for more details. The Sargan-Hansen test is the average of the country-specific overidentifying restrictions statistics that test the null of the joint validity of the instruments used (see Sargan, 1958; Hansen, 1982). The Cumby-Huizinga test shows the average of the individual countries' Cumby and Huizinga (1992) autocorrelation tests, testing the null of no autocorrelation against the alternative that the autocorrelations of the error term are nonzero at lags greater than zero. The Durbin-Wu-Hausman statistic shows the average of the individual countries' Durbin-Wu-Hausman endogeneity tests (see Nakamura and Nakamura, 1981), testing the null hypothesis that the regressor r_{it} is exogenous. Between square brackets are the harmonic means of the country-specific p-values (HMP) of these tests. With $N = 15$, the 5% critical value of the harmonic mean p-value is about 0.04. We refer to Wilson (2019) for details.

Appendix C   Results for equity and bill returns

This appendix reports the results obtained from our empirical approach when using the asset returns that have been traditionally considered when estimating the EIS, i.e., real returns on equity and Treasury bills (see Havranek et al., 2015, for an overview). Data for nominal equity and bill returns are taken from Jordà et al. (2019).[2] To calculate real returns r_{it}, we deflate these nominal returns using the inflation rate calculated from the Consumer Price Index (CPI) which is taken from the Jordà-Schularick-Taylor macro-history database.

Table A-3 presents the mean-group results of estimating eq.(11) using real returns on equity (column one) and bills (column two) for r_{it}. As before, the instruments are $\bar{r}_t$ (now based on equity or bill returns), $r_{i,t-1}$, f_t^{inc} and f_t^{fin}. The cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} are GDP-weighted and include all countries of the sample. The results obtained using alternative approaches to calculate these averages provide similar results and conclusions (they are unreported, but available upon request). With respect to the diagnostics, we note that the effective F tests strongly reject the weakness of instruments for both returns. For equity returns, however, the lagged return $r_{i,t-1}$ has no individually significant impact on r_{it} in the first-stage regression (not reported), i.e., after controlling for the cross-sectional average $\bar{r}_t$, there is not much persistence left in equity returns. Furthermore, the overidentifying restrictions test does not reject the exogeneity of the instruments for bill returns but does reject the validity of the instruments for equity returns. As for equity returns only the cross-sectional average $\bar{r}_t$ is a relevant instrument, the model for equity returns is in fact just-identified rather than overidentified. Hence, it is not clear whether the result of the Sargan-Hansen test is very informative in this case. Irrespective of the diagnostics and their limitations, we note that the mean-group estimates obtained for ψ - i.e., the average impact of equity or bill returns on aggregate consumption growth - are close to zero and not significantly different from zero. This implies that EIS estimates based on these results - which, as with housing, can be obtained by taking into account the ownership rates of equity and bills - will also not be significantly different from zero. This is in line with other macro studies that estimate the EIS from equity or bill returns (see, in particular, Havranek, 2015, for a recent meta-analysis).

[2]The data can be found at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GGDQGJ where the nominal equity returns have code 'eq-tr' and nominal bill returns have code 'bill-rate'. Details on the data sources and data construction are discussed in the text and online appendix of Jordà et al. (2019)'s paper.

Table A-3: Results estimation of regression eq.(11) using equity and T-bill returns

	(1) equity	(2) T-bills
r_{it} (ψ)	0.005	0.028
	(0.004)	(0.044)
f_t^{inc} (γ^{inc})	0.392	0.389
	(0.066)	(0.063)
f_t^{fin} (γ^{fin})	0.160	0.165
	(0.022)	(0.028)
F	46.226	69.849
effective F	52.436	46.110
30% thresh. [$p-val.$]	[0.000]	[0.000]
20% thresh. [$p-val.$]	[0.000]	[0.000]
10% thresh. [$p-val.$]	[0.000]	[0.000]
Sargan-Hansen	3.303	1.311
[$p-val.$]	[0.027]	[0.154]
Cumby-Huizinga	8.863	8.664
[$p-val.$]	[0.018]	[0.027]
Durbin-Wu-Hausman	1.122	2.960
[$p-val.$]	[0.030]	[0.013]
Instruments	$\bar{r}_t$, $r_{i,t-1}$	$\bar{r}_t$, $r_{i,t-1}$
	f_t^{inc}, f_t^{fin}	f_t^{inc}, f_t^{fin}
$\bar{r}_t$, f_t^{inc}, f_t^{fin}	weighted	weighted
	i included	i included

Notes: The table reports the mean-group results of per country IV estimation of eq.(11) using real equity returns, respectively real T-bill returns for r_{it}. The instrument set consists of a constant, $\bar{r}_t$, $r_{i,t-1}$, f_t^{inc} and f_t^{fin}. Estimation is based on panel data for fifteen countries over the period 1950−2015. Standard errors based on eq.(12) are in parentheses. Mean-group results for the constant are not reported. Results reported are for cross-sectional averages $\bar{r}_t$, f_t^{inc} and f_t^{fin} that are GDP-weighted and that include country i (i.e., all countries are included). We refer to the notes to Table 4 for details. Reported are the cross-country averages of the regular F statistic and the effective F statistic of Montiel Olea and Pflueger (2013). The null hypothesis of weak instruments based on the effective F is the hypothesis that the bias of the IV estimator relative to that of a 'worst-case' benchmark (like OLS) exceeds the threshold τ with $\tau = 0.3, 0.2, 0.1$. The Sargan-Hansen test is the average of the country-specific overidentifying restrictions statistics that test the null of the joint validity of the instruments used (see Sargan, 1958; Hansen, 1982). The Cumby-Huizinga test shows the average of the individual countries' Cumby and Huizinga (1992) autocorrelation tests, testing the null of no autocorrelation against the alternative that the autocorrelations of the error term are nonzero at lags greater than zero. The Durbin-Wu-Hausman statistic shows the average of the individual countries' Durbin-Wu-Hausman endogeneity tests (see Nakamura and Nakamura, 1981), testing the null hypothesis that the regressor r_{it} is exogenous. Between square brackets are the harmonic means of the country-specific p-values (HMP) of these tests. With $N = 15$, the 5% critical value of the harmonic mean p-value is about 0.04. We refer to Wilson (2019) for details.