

Supplemental Appendix

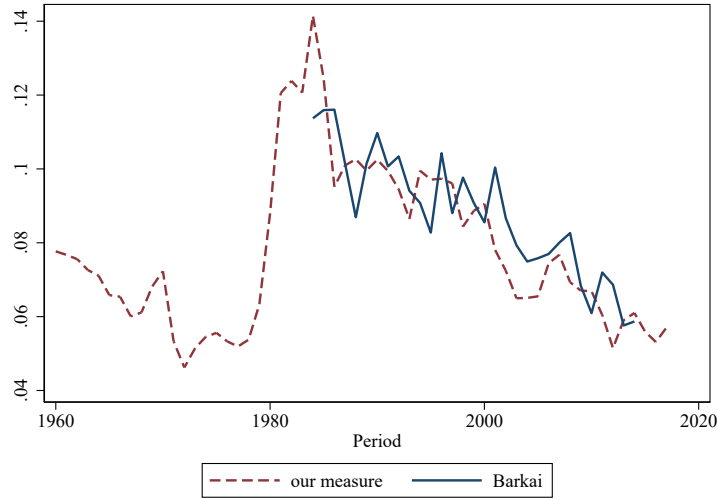
Consumption Inequality in the Digital Age

Kai Arvai and Katja Mann

A Data Appendix

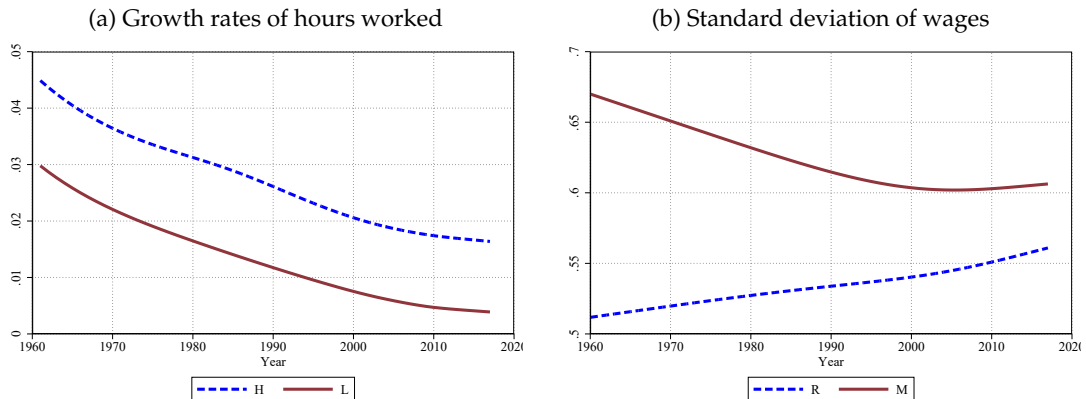
A.1 Additional Tables and Figures

Figure A1: Comparison of costs of capital to ?



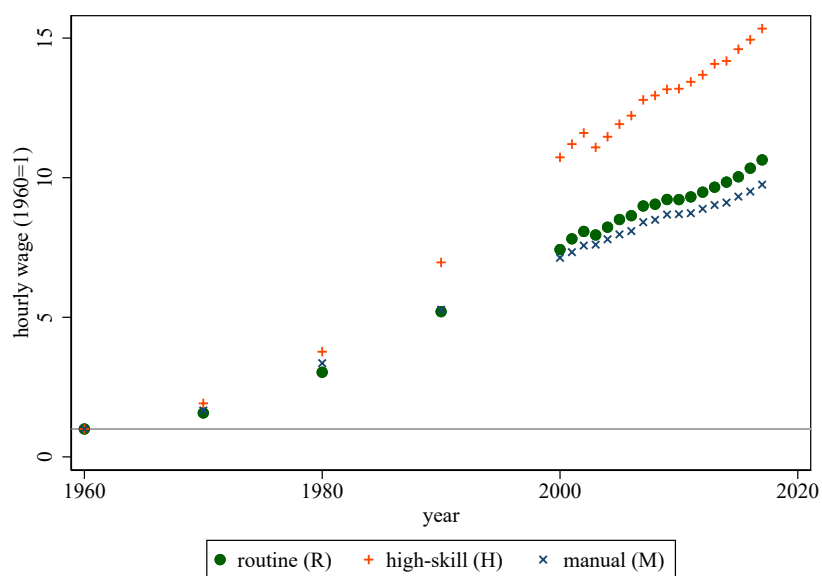
Note: The graph shows the costs of capital as calculated by $\tilde{r}_t$ in eq. (??) with the weighted average costs of capital of ?, Figure 2A. *Source:* ?, BEA and own calculations.

Figure A2: Labor supply and standard deviation of wages by worker type



Note: The left panel shows annual growth rates of high-skill and low-skill workers. High-skill workers are defined as workers in non-routine cognitive occupations, low-skill workers as all others. The right panel shows the annual standard deviations of wages for routine and non-routine manual workers as obtained from a truncated maximum likelihood estimation. The occupational classification follows ?. Numbers show a trend from applying an HP filter. Between Census years 1960, 1970, 1990 and 2000, numbers are interpolated. *Source:* ACS, Census and ?.

Figure A3: Hourly wages, ACS



Note: The graph shows the development in hourly wages by occupation group between 1960 and 2017 based on data from the Census and ACS. Hourly wages are normalized to 1 in 1960

Table A1: ICT assets: equipment

asset code	asset name	ICT asset
EP1A	Mainframes	✓
EP1B	PCs	✓
EP1C	DASDs	✓
EP1D	Printers	✓
EP1E	Terminals	✓
EP1F	Tape drives	✓
EP1G	Storage devices	✓
EP1H	System integrators	✓
EP20	Communications	
EP34	Nonelectro medical instruments	
EP35	Electro medical instruments	
EP36	Nonmedical instruments	
EP31	Photocopy and related equipment	
EP12	Office and accounting equipment	
EI11	Nuclear fuel	
EI12	Other fabricated metals	
EI21	Steam engines	
EI22	Internal combustion engines	
EI30	Metalworking machinery	
EI40	Special industrial machinery	
EI50	General industrial equipment	
EI60	Electric transmission and distribution	
ET11	Light trucks (including utility vehicles)	
ET12	Other trucks, buses and truck trailers	
ET20	Autos	
ET30	Aircraft	
ET40	Ships and boats	
ET50	Railroad equipment	
EO11	Household furniture	
EO12	Other furniture	
EO30	Other agricultural machinery	
EO21	Farm tractors	
EO40	Other construction machinery	
EO22	Construction tractors	
EO50	Mining and oilfield machinery	
EO60	Service industry machinery	
EO71	Household appliances	
EO72	Other electrical	
EO80	Other	

Note: The data are retrieved from BEA's detailed fixed assets accounts.

Table A2: ICT and non-ICT assets: intellectual property products

asset code	asset name	ICT asset
ENS1	Prepackaged software	✓
ENS2	Custom software	✓
ENS3	Own account software	✓
RD11	Pharmaceutical and medicine manufacturing	
RD12	Chemical manufacturing, ex. pharma and med	
RD23	Semiconductor and other component manufacturing	✓
RD21	Computers and peripheral equipment manufacturing	✓
RD22	Communications equipment manufacturing	
RD24	Navigational and other instruments manufacturing	
RD25	Other computer and electronic manufacturing, n.e.c. ¹	
RD31	Motor vehicles and parts manufacturing	
RD32	Aerospace products and parts manufacturing	
RD0M	Other manufacturing	
RD70	Scientific research and development services	
RD40	Software publishers ¹	
RD50	Financial and real estate services	
RD60	Computer systems design and related services ¹	
RD80	All other nonmanufacturing, n.e.c.	
RD91	Private universities and colleges	
RD92	Other nonprofit institutions	
AE10	Theatrical movies	
AE20	Long-lived television programs	
AE30	Books	
AE40	Music	
AE50	Other entertainment originals	

Note: The data are retrieved from BEA's detailed fixed assets accounts. As this paper excludes structures, the list contains only equipment capital and intellectual property products. ¹: Excluded due to short time series dimension.

Table A3: Income and wealth by education group, SCF 1998-2013

	Labor income	Capital income	Total income	Net wealth
college graduates	56,443	50	59,300	142,044
non-college graduates	27,873	0	29,707	27,634

Note: All numbers are in 2019 US Dollars and refer to the median household in each education group. Total income is the sum of labor income, capital income and transfer income as defined by ?. *Source:* SCF.

A.2 Details on Data Sources and Dataset Construction

NIPA

The goods and assets price series as well as aggregate consumption measures are constructed using data from the BEA's National Income and Product Accounts (NIPA). For asset prices, we use NIPA Tables 5.5.3, 5.5.4 and 5.5.5, which contain quantities, prices and expenditures for equipment, and Tables 5.6.3, 5.6.4 and 5.6.5 for intellectual property products. We merge the two categories and create aggregate Törnqvist price indices for ICT and non-ICT capital, following the classification used for the Detailed Data for Fixed Assets and described in the main text. For the price series, we use NIPA Tables on personal consumption expenditures by type of product. We use tables 2.4.3U, 2.4.4U, 2.5.4U to construct series for consumed quantities, prices and personal consumption expenditures. To match these with the ICT share measures at the level of IO final goods and services, we use a crosswalk from NIPA PCE lines to UCC codes provided by the BLS. Then we apply our matching of UCC codes to IO final commodities. We construct price indices for the composite ICT- and non-ICT good as described in the main text.

Input-Output Accounts

We use the BEA's Input Output Accounts and focus on the detailed Input-Output tables after re-definitions. We use producer-value tables, which means that the distribution margin (i.e. the cost of wholesaling, retailing and transportation) is modeled as a flow from the distribution industries to the final consumer rather than as a flow from the producing industry to the final consumer. This is the standard approach and is appropriate here as we have no reason to believe that the digitalization content of distribution differs between the different goods and services. We match our 61 BEA industries to the approximately 400 IO product categories.

The Input-Output accounts are presented in a set of different tables, among them *use*, *make* and *direct requirements* tables. We start from the commodity-by-industry direct requirements table, which shows the amount of each commodity (i.e., intermediate or final good or service) that is required by an industry to produce one dollar of the industry's output. The problem with this table is that

the same commodity can be produced by different industries, e.g. ice-cream can be produced by the dairy product manufacturing industry and the ice-cream manufacturing industry. We follow ? in creating a commodity-by-commodity direct requirements matrix that takes the market shares of each industry in producing certain commodities into account. With this matrix, we can calculate the digitalization share of final goods and services as a weighted mean of the digitalization shares of its intermediate inputs and of value added. Note that all sectors related to housing are matched to the real estate sector.

We pursue an iterative approach. We initially assign to each commodity the digitalization share of the industry that is the ultimate producer of the commodity. Then, we consider all commodities that use this specific commodity as intermediate input. We update the output digitalization share by calculating a weighted mean of the digitalization shares of all inputs and the digitalization share of value added of the final producer. We again assign this share to the inputs used to produce other commodities, and update the output digitalization share again, etc. We continue this procedure until the commodity digitalization shares have converged to fixed values.

For the import share used as a control variable in Table ??, we use IO use tables after redefinitions at producer value. The import share is calculated as the (negative of the) value of imports of goods and services divided by the value of total output.

CEX

The Survey of Consumer Expenditures is the most detailed expenditure survey in the United States, carried out at the household level. Next to providing detailed expenditure data, it also contains a host of demographic and income information on the households. We use the interview survey data of the public-use microdata, which covers the full consumption basket starting from 1990. Sample households are interviewed at a quarterly frequency for four to five consecutive months. We treat each interview round as independent observation. Households report monthly expenditures at the universal classification code (UCC) level for the three months prior to the interview month. Additionally, the data contain quarterly expenses at the level of summary product categories. We consider only households where the head is between 16 and 64 years old and in the labor force. This is because we want to focus on households where labor income is the main source of income. In defining the head, we divert from the CEX convention by making the head the man in mixed couples.

We measure income as total before-tax annual household income. Income is self-reported prior to 2004 (FINCBTAX) and imputed from 2004 onward (FINCBTXM). We only consider households with strictly positive income. We drop the top and bottom 5% of the income distribution in order to mitigate the effect of outliers and top-coding, following ?. The main income component is labor income. Next to labor income, the income measure includes farm and non-farm business income, social security income, interest on savings accounts or bonds, income from dividends, royalties, estates and trusts and rental income. For renters, we define housing expenditures by rent paid. For home owners, we use self-reported rental equivalence rather than out-of-pocket expenditures for repairs and maintenance and mortgage payments. We also drop financial expenditures such as cash contributions, loans and interest payments, alimony payments and cash gifts, as well as

trade-in allowances. We calculate annual expenditures at the level of calendar years. As survey weights refer to the interview year, we need to adjust survey weights for expenditures that relate to the previous year. This concerns households interviewed in January through March. E.g. for a household interviewed in February, the reported expenditures refer to November and December of the previous year, as well as January of the interview year. In the adjustment we follow the instructions published in the Public Use Microdata Users' Documentation.

Before 1990, the CEX only contains spending on summary product categories. For 1984-1989, we follow ? to impute UCC-level expenditures. They use microdata for 1990 and assume that the expenditure shares for any given income percentile within each of the summary product categories are the same in 1984-1989 as in 1990. Our final sample has between 9,000 to 18,000 observations per year.

The CEX is known for under-reporting of expenditure, in particular by richer households (?; ?). This problem has been increasing over time. In consequence, inequality measures like the total consumption expenditure of rich relative to poor households are biased downwards. For the purpose of this analysis, we consider the spending of households of a specific income bin on different products relative to their total expenditure. As long as rich households underreport all expenditures to an equal extent, our results should not be biased.

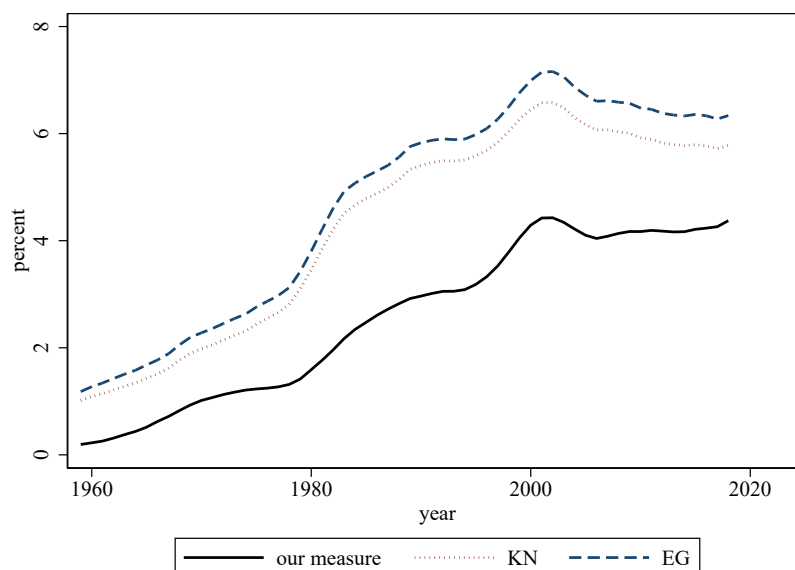
We observe only total expenditure on each item, and cannot separate between prices paid and quantities purchased. This is problematic, because households may face different prices for the same good, or choose differently-priced products within the narrow categories of the CEX. This issue has been recognized by the literature on consumption inequality (see ? for a discussion) and means that we are potentially overestimating consumption inequality.

Census and ACS

Most of our empirical measures of employment and wages are constructed using data from the Decennial Census (for the years 1960, 1970, 1980 and 1990) and the American Community Survey (ACS, available annually from 2000). For the years in between, we interpolate the data. In the sample composition as well as the definition of routine and non-routine workers, we follow ?. We keep only individuals between ages 16-64 that are employed and non-institutionalized. Our key measures are the number of hours worked per year and the hourly wage. The number of hours worked is the number of hours usually worked during a week times the number of weeks worked within a year. As income is top-coded, we replace income in the top bracket with 1.5 times the actual value and then divide by the number of hours worked to obtain the hourly wage. We report all Dollar values in constant 1999 US Dollars. We create a consistent occupational classification using the crosswalks provided on David Dorn's website. We classify the following occupations as routine: sales and administrative support, precision production and craft workers and machine operators. We classify as non-routine cognitive workers managers, professionals and technicians. Non-routine manual occupations are those working in construction and mining, services, mechanics, repairers and transportation. We drop individuals working in the agricultural sector or the military. Out of the total of 318 occupations, 210 are routine or non-routine manual.

A.3 Comparison of ICT share to related measures

Figure A4: ICT share comparison



Note: The figure compares the ICT share in value-added of this paper with ? and ?. For the sake of comparability, all three measures are calculated using quality-adjusted asset prices. Note that the numbers for our ICT share differ slightly from the numbers shown in Figure ?? (a) since we calculate a weighted average here, whereas Figure ?? is an unweighted average across industries. *Source:* BEA and own calculations.

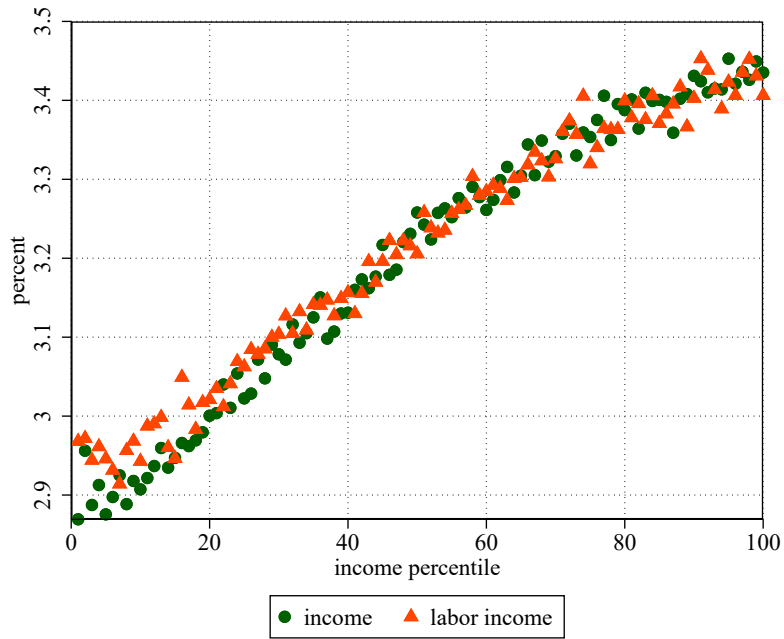
Both ? and ? calculate the ICT share in value-added, but using a slightly different definition of ICT assets. Their set of ICT assets is broader than ours, also including medical and nonmedical instruments, photocopy equipment, office and accounting equipment and communications equipment. In contrast, ? do not include intellectual property products that relate to computer and electronic manufacturing and software publishers. Computers and software, which are at the core of our measure, have seen stronger price declines than other assets such as medical and nonmedical instruments, photocopy equipment or office and accounting equipment. None of the other two papers takes quality changes in ICT capital into account. For the sake of comparability, we have computed their measures analogously to ours, i.e. adjusting for quality improvements. Figure A4 shows ICT shares for the aggregate US economy. Due to the broader ICT definition in ? and ?, their ICT shares are larger. However, they exhibit similar time trends.

A.4 ICT share in the Consumption Basket: Extensions and Robustness

Sorting Households by Labor Income

As an alternative to grouping households by total income, we group households into percentiles of labor income, dropping households with non-positive labor income. As Figure A5 shows, it makes almost no difference whether we consider percentiles of total income or labor income, the only difference being the lowest percentiles.

Figure A5: ICT share along the labor income distribution

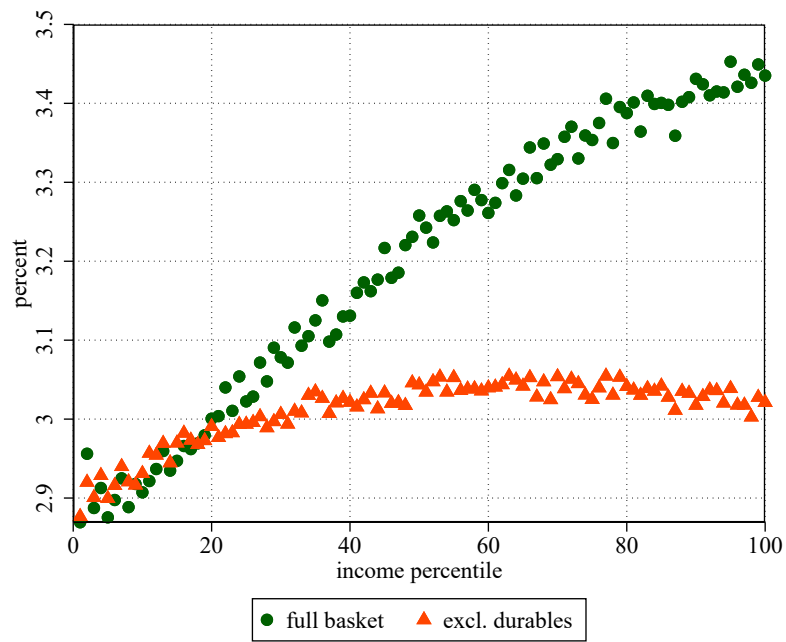


Note: The graph shows the ICT share of the consumption basket comparing percentile of total household income with percentiles of labor income. *Source:* BEA, CEX and own calculations.

Non-durable Goods

Consumption in most macroeconomic models can best be understood as consumption of non-durable goods, therefore we exclude in this extension spending on durable goods such as housing, vehicles or education. Figure A6 is the equivalent to Figure ?? but excluding durable goods. This excludes around 35% of expenditures. Among those dropped are expenditures for housing, household equipment, vehicle expenses, health and education. We follow ? in the non-durable/durable classification. The ICT share still increases along the income distribution, but the pattern is significantly weakened. Durable goods are an important driver of differences in ICT share across households.

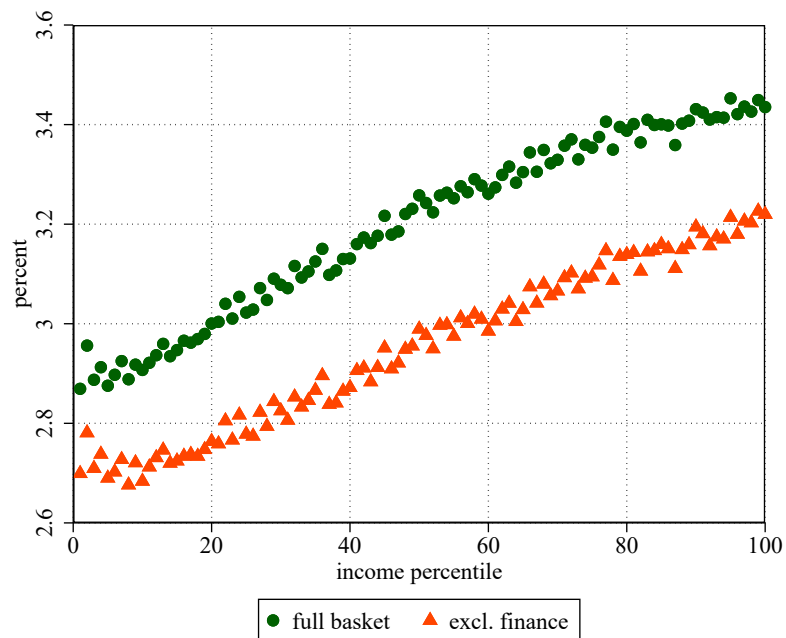
Figure A6: ICT share along the income distribution, only non-durable goods



Note: The graph shows the ICT share of the consumption basket by percentile based only on non-durable goods. *Source:* BEA, CEX and own calculations.

Excluding Finance and Insurance

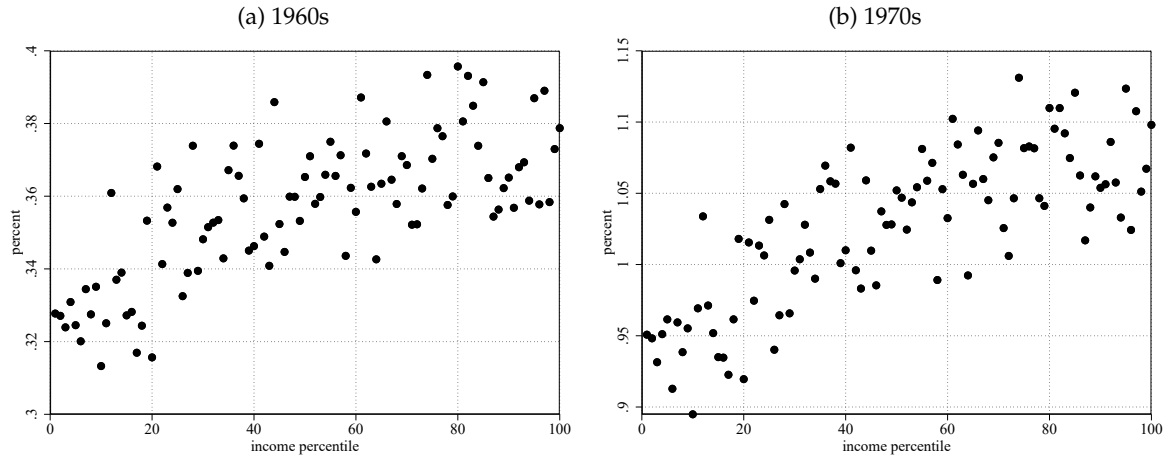
Figure A7: ICT share along the income distribution, excluding finance and insurance



Note: The graph shows the ICT share of the consumption basket by percentile for the full sample period including and excluding finance and insurance. *Source:* BEA, CEX and own calculations.

Historical Consumption Data

Figure A8: ICT share along the income distribution, historical data



Note: The graph shows the ICT share of the consumption basket by income percentile for the 1960s and 1970s. *Source:* BEA, CEX, ? and own calculations.

We use historical data constructed by ? for the 1960s and 1970s. These data are based on several imputations: first, between the years with available data (1960, 1972, 1984). Second, for detailed consumption categories (UCCs), which were not reported in the early waves. ? use expenditure on summary categories and assume that expenditure shares within these categories were the same as in 1990. As a result, the data are much noisier. The early data do not contain some of the expenditure categories with the highest ICT shares such as computers and TVs, and in consequence the overall level of the ICT share is much lower. However, the positive income gradient is also observable in the 1960s and 1970s.

Demographic Factors

In this subsection, we study what role the demographic factors age and education play for our results. Household income tends to increase with age, which is why we might be concerned that rather than capturing an increasing ICT share of consumption along the income distribution, we capture higher spending shares on digital goods among the age distribution. Similarly, income also tends to be higher for households with college education, who at the same time might consume more digitally. Whether a household is located in an urban or rural area might also place a role for digital consumption patterns, while at the same time it correlates with income. To refute these concerns, we first run regressions of the ICT share on income, controlling for the age and the education of the household head and for whether the households lives in a urban or rural area. Then, we show Figure ?? separately for young households (below age 32), middle-aged households (aged 32-47) and old households (above age 47), households with and without college education, and urban and rural households.

Table A4 shows regressions of ICT share on income and demographic controls at the level of income percentiles. The demographic controls are measured as averages across all households within the income percentile, where age and education are for the household head. Thus, age is

Table A4: Regression of ICT share on income and age

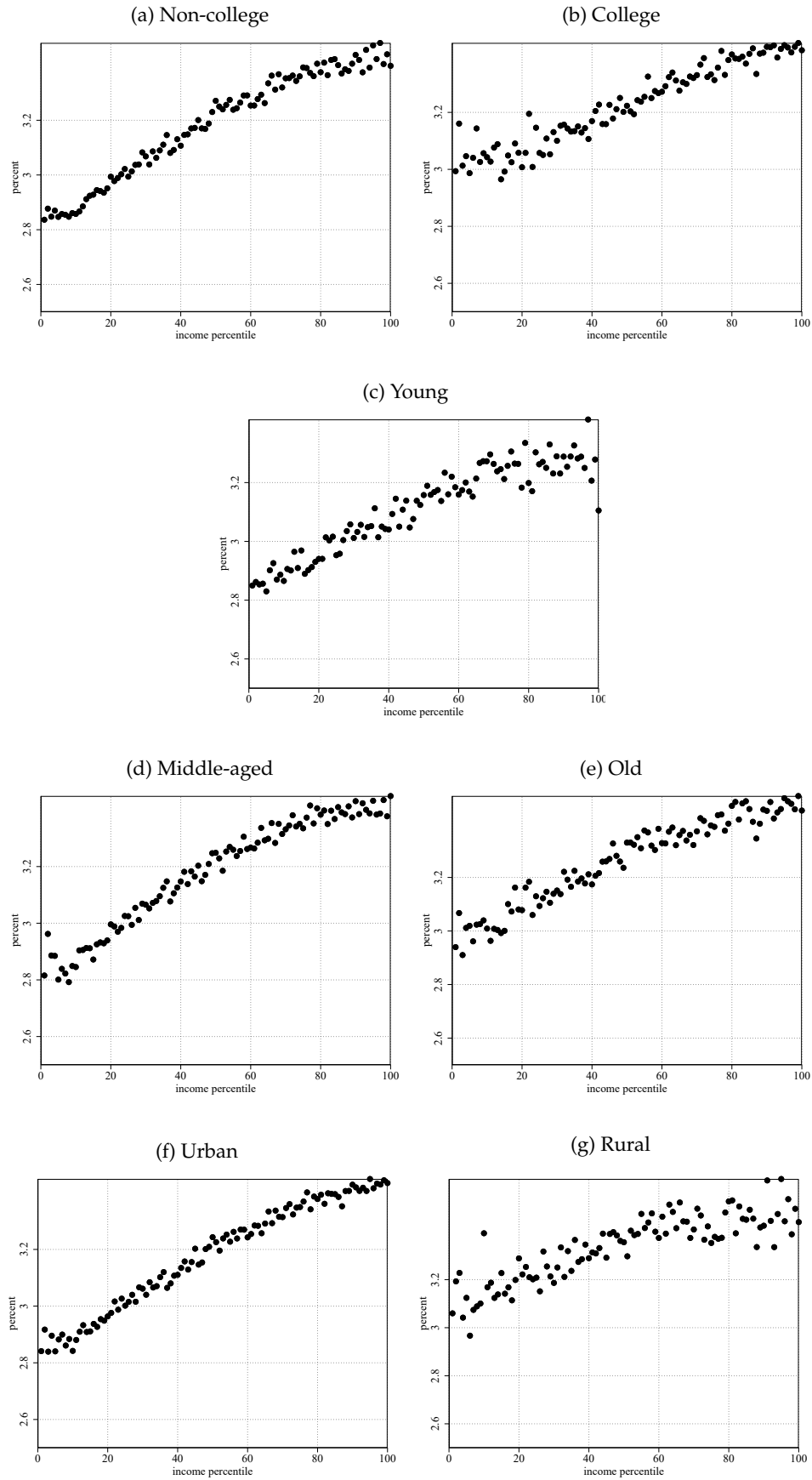
	(1)	(2)	(3)	(4)	(5)
income	0.483*** (0.007)	0.323*** (0.010)	0.223*** (0.013)	0.492*** (0.007)	0.174*** (0.013)
age		2.617*** (0.134)			1.747*** (0.137)
educ			48.092*** (2.089)		42.639*** (2.235)
urban				-17.928** (7.094)	-61.737*** (6.588)
Constant	181.311*** (1.329)	84.531*** (5.108)	87.143*** (4.272)	197.328*** (6.476)	88.354*** (7.209)
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	3,400	3,400	3,400	3,400	3,400
R-squared	0.958	0.962	0.963	0.958	0.966

Note: Income is in thousands of constant 2015 USD. ICT share is in %. Education and age refer to the household head. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

the average age of the household head, educ the average level of education (where at the individual level, 1 means no high-school degree, 2 means high-school but no college degree, and 3 college degree), and urban the share of households living in urban areas. The coefficient on income is highly significant and remains significant once demographic control variables are added. As expected, the magnitude becomes smaller when adding age and education. In contrast, controlling for the urban share makes the income gradient slightly steeper. Including all demographic controls in column (5), the coefficient on income is still positive and significant, but only about one third of the size of column (1).

In Figure A9, the ICT share increases with income across all demographic subgroups. The non-college group shows a larger income gradient than the college group, so the difference between the bottom and top percentile is more pronounced for non-college households, suggesting that they are a more heterogeneous group than college-educated households. The median for both groups is very similar (2.99 vs. 2.94). The ICT share is slightly higher for older households than for middle-aged and younger ones. The median is 2.88 for the young compared to 3.04 for the old. The rural sample is much smaller than the urban sample and therefore exhibits more dispersion. The median in the rural sample is – perhaps counterintuitively – higher than in the urban sample (2.96 vs. 3.96).

Figure A9: ICT share along the income distribution - subsamples



Note: The graph replicates Figure ?? for households (a) without college degree, (b) with college degree, (c) aged below 32, (d) aged between 32-47, (e) aged above 47, (f) living in urban areas, (g) living in rural areas. For the ease of exposition, we only show the average across the whole time period 1984-2017. *Source:* BEA, CEX and own calculations.

A.5 Relative Price Changes: Extensions and Robustness

Additional Price Regressions

Table A5 shows several robustness checks for Table ?? : In columns (1) and (2), we include fixed effects at the level of IO goods and services instead of major industries. In columns (3), (4) and (5) we use as explanatory variable the beginning-of-decade ICT share instead of end-of-decade ICT share. In columns (6), (7) and (8), we include a quadratic ICT share. The baseline results of Table ?? are robust to each of the specifications. Notably, the ICT share seems to have a non-linear effect on price changes, and is less negative at higher ICT shares. The net effect would turn positive from an ICT share of 42.0, which is however outside of the bounds of our sample. The highest ICT share in our sample is 33.0, which is found for the software publishing industry (511200).

Table A5: Price changes and ICT share

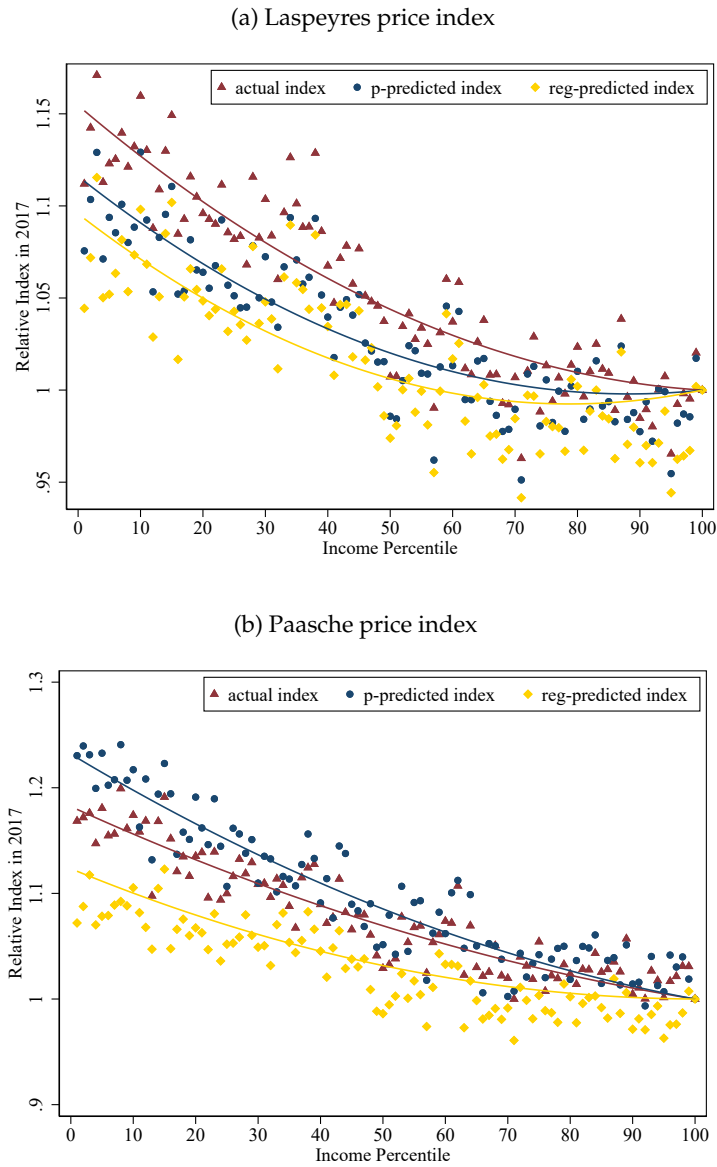
VARIABLES	(1) 60-17	(2) 60-17	(3) 67-17	(4) 67-17	(5) 67-17	(6) 60-17	(7) 60-17	(8) 60-17
ICT share	-0.0743*** (0.00953)	-0.0804** (0.0331)				-0.151 (0.140)	-0.184** (0.0726)	-0.294*** (0.111)
L.10 ICT share			-0.205*** (0.0643)	-0.163*** (0.0392)	-0.182*** (0.0458)			
ICT share ²						-0.000562 (0.00424)	0.00170 (0.00220)	0.00495 (0.00320)
Constant	0.0394 (0.0401)	0.0226 (0.185)	0.328 (0.327)	0.212 (0.205)	0.377 (0.237)	0.298 (0.338)	0.270 (0.212)	0.207** (0.0991)
Observations	3,912	3,912	3,275	3,275	3,275	3,912	3,912	3,912
R-squared	0.531	0.533	0.091	0.334	0.338	0.102	0.344	0.351
industry FE	no	no	no	yes	yes	no	yes	yes
IO FE	yes	yes	no	no	no	no	no	no
decade FE	no	yes	no	no	yes	no	no	yes

Note: The table shows the regression of annualized decadal price changes on the ICT share at the end or beginning of each decade. The unit of observation are product categories. All variables are measured in %. Standard errors clustered at the level of final goods and services in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Alternative Price Indices

Figure A10 replicates the result of Figure ?? but using two different price indices: the Laspeyres index and the Paasche index. It is a well-known fact that the Laspeyres index, due to its reliance on period t-expenditure shares, tends to overstate inflation, the Paasche index, relying on period t+1-expenditure shares, understates inflation. ? showed that inflation inequality across the income distribution looks similar for all three indices. Figure A10 confirms this pattern for the longer 1984-2017 period while also showing that the effect of digitalization on prices is similar for Törnqvist, Laspeyres and Paasche.

Figure A10: Inflation and ICT, alternative price indices

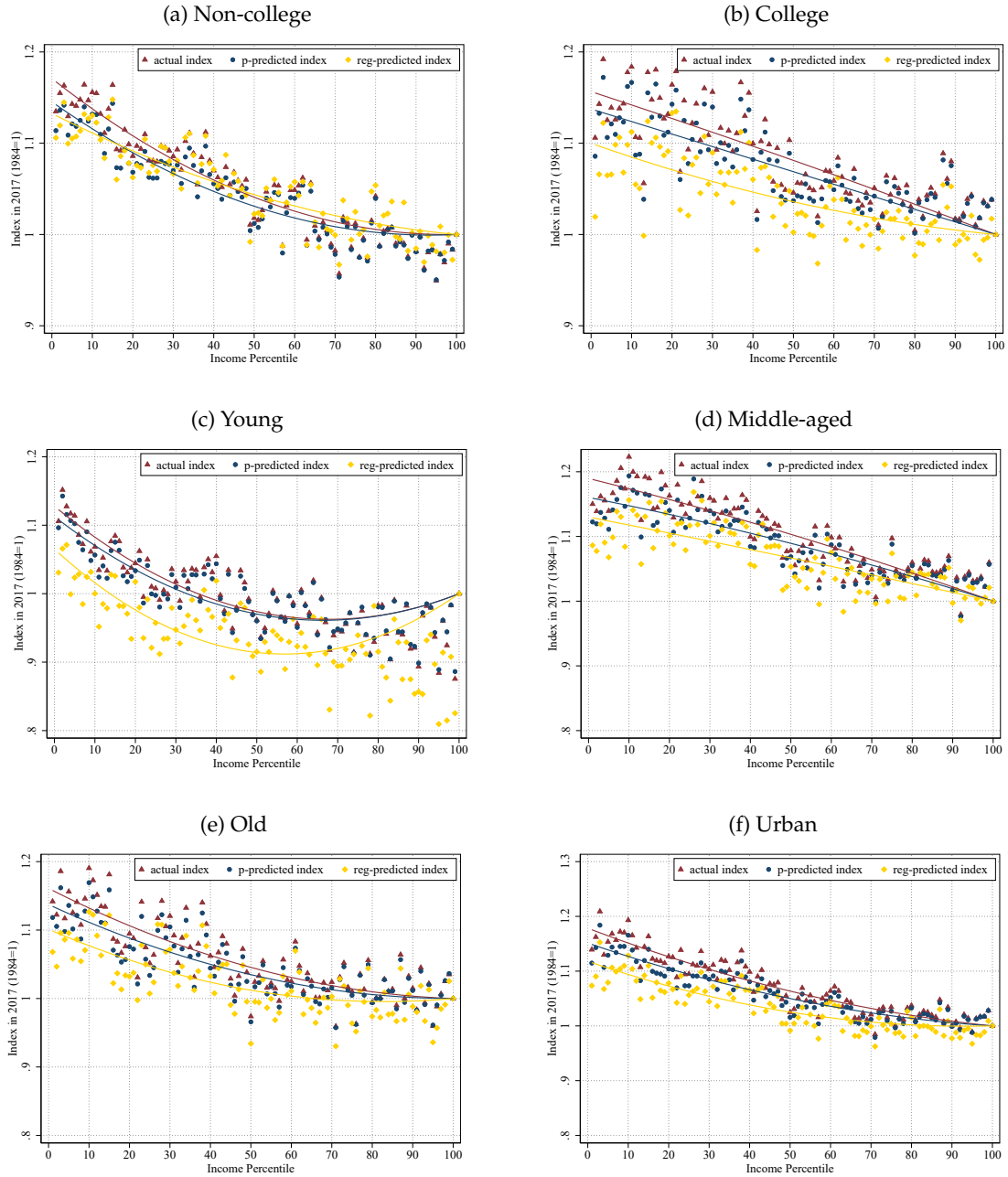


Note: The top graph plots Laspeyres price indices, the bottom graph Paasche price indices. Both price indices are shown for 2017, indexed to 1 in 1984 and relative to the price in the highest income percentile. The red triangles indicate the actual data, whereas the blue circles and the yellow diamonds indicate ICT-adjusted indices following the same computation as in Figure ???. The blue ICT-adjusted Paasche index indicates that ICT-intensive products would benefit the poor, but this is due to stronger substitution among the poor, which gets a disproportionate weight. *Source:* BEA, CEX and own calculations.

Demographic Factors

Figure A11 replicate Figure ?? for the five demographic subgroups discussed in Appendix Section A.4: Young households (below age 32), middle-aged households (aged 32-47) and old households (above age 47), households with and without college education, and households living in urban areas. Due to the low number of observations, the group of rural households is omitted.

Figure A11: ICT share along the income distribution - subsamples



Note: The graph replicates Figure ?? for households (a) without college degree, (b) with college degree, (c) aged below 32, (d) aged between 32-47, (e) aged above 47, (f) living in urban areas. The rural group is omitted due to lack of observations. *Source:* BEA, CEX and own calculations.

A.6 Alternative Digitalization Measures

Input- vs. Output-based Digitalization Measures

In the main text, we define digitalization based on the inputs used in the production process. A good is classified as digital according to the extent to which it has been produced with digital assets like computer hardware or software. An alternative would be an output-based measure, sorting final consumption goods into digital or non-digital goods. Different attempts have been made to classify final consumption goods. We rely on the BEA's Digital Economy Satellite Account, which represents the BEA's approach to calculating the contribution of digital goods- and service-producing industries to U.S. GDP ("digital economy"). The BEA defines the digital economy as comprising three types of goods and services: digital infrastructure, which primarily refers to ICT, e-commerce and priced digital services, i.e. services related to computing and communication that require a fee (?). The BEA provides data for the digital economy's value added at the industry level. We use the share of digital value-added in total value-added as output-based ICT share measure. Since the industry classification is the same as the one we use to construct our output-based measure, we can easily compare it to our input-based measure.

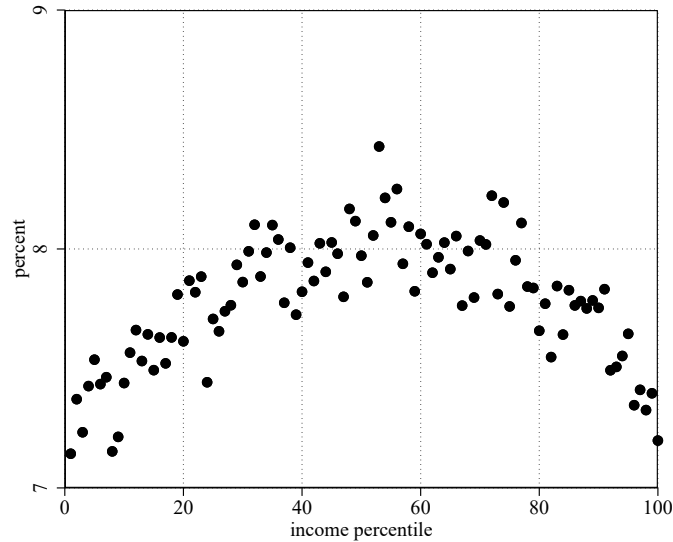
The BEA Digital Economy Satellite Account does not (yet) attempt to estimate the value of free digital goods and services. It is sometimes approximated by the advertising revenue generated by the companies offering it, however, this will insufficiently reflect the consumer surplus generated by these products (?). ? and ? present an alternative approach to estimate the value of free digital goods. They carry out choice experiments to estimate consumers' reservation prices, asking them about the monetary compensation at which they are willing to give up access to a digital good such as Google Maps, Facebook or Wikipedia. These studies show that the consumer surplus generated from free goods can be substantial. However, the research in this area is still in its infancy and the existing literature does not study differences between high- and low-income households systematically. Therefore, it is not possible to assess the bias created by ignoring the consumer surplus of free digital goods.

The correlation between our input-based and the BEA's output-based measures is 0.78. While both measures rely highly on the ICT sector (Naics 334* (excl. 3345), 5112, 5191, 5415 and 518), the following categories have a high ICT share according the input-based measure, but not the output-based measure: navigational, measuring, electromedical and control instruments (3345), financial services and insurance carriers (52*), newspaper and book publishers (5111). These tend to be consumed proportionately more by high-income households. Categories with a low input ICT-intensity but high output ICT-intensity are in particular radio and television broadcasting (5151) and telecommunications (517). The expenditure share of these categories is higher for low-income households. Due to these differences, the ICT share along the income distribution looks different when using the output-based definition. Figure A12 shows the digital share of the consumption basket along the income distribution for the output-based measure. According to this measure, the ICT share is hump-shaped along the income distribution, with middle-income households having the highest ICT share, whereas top and bottom are very similar to each other.

This underlines that it matters how digital goods are defined. Choosing one measure or the other

might lead to different implications about who benefits and who loses from digitalization. We have opted for an input-based measure in this paper since we are interested in comparing income and price effects of changes in production processes, but we acknowledge that there could also be benefits in studying output-based measures, and that the findings would be complementary to ours.

Figure A12: ICT share along the income distribution, output-based classification



Note: The graph shows the ICT share of the consumption basket by percentile for 2005-2017 period for the output-based series constructed based on the BEA's Digital Economy Satellite Account. *Source:* BEA, CEX and own calculations.

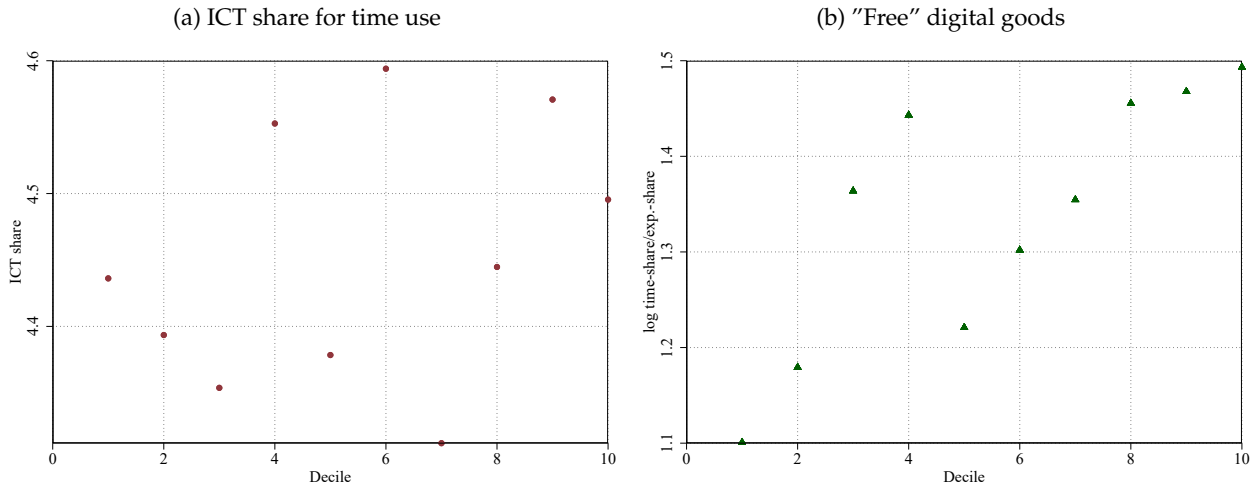
Digitalization and Time Use

We supplement our main analysis with data from the American Time Use Survey (ATUS). ATUS documents how Americans allocate their time across different activities. It has been running since 2003 and is based on a subsample drawn from the Current Population Survey (CPS) of between 10,000 to 20,000 respondents. Within a 24-hour period, the respondents document their primary activities within 15 minute windows, choosing among more than 400 different activity categories. We exclude from our analysis the time when individuals are working and sleeping as well as the use of government services.

To develop a concordance that matches ATUS activity codes to IO categories, we follow ?. However, while ? match broad ATUS categories to broad CEX expenditure categories, our concordance is more disaggregate.

When matching all activities to their corresponding expenditures, there is a near zero correlation between expenditure and time-shares for all income groups. This is illustrated by Table A6 which focuses on the richest and poorest 10 percent. This implies there is no overall pattern suggesting time use and expenditures are substitutes or complements. This is also in line with the fact that the average ICT share is similar for both time use and consumption expenditures, as illustrated by comparing Figure ?? with A13 (a). Among categories that are substitutes (high time use and

Figure A13: Time use across income deciles



Note: Graph (a) shows the ICT share of the overall time use by income decile for different sub-periods between 2003 and 2017. Graph (b) shows the log ratio of time-expenditure-shares for "free" digital products and services, or more precisely digital activities with zero marginal costs. Free digital products and services are defined as ATUS activities computer use, e-mail services, playing games and listening to music. Average between 2003-2017. *Source:* BEA, ATUS, CEX and own calculations.

low expenditures or vice versa) are activities related to housework. Those are slightly less ICT intensive and more important for the poor.

Table A6: Correlation between time share and expenditure share

Time Period	Top 10%	Bot 10%
2003-2010	0.04	0.06
2011-2017	0.01	0.05

Sources: BEA, CEX, ATUS and own calculations

Next, we compute the average ICT share of time use of an individual by summing up the time-share weighted activities multiplied with their ICT shares. We track the time share of activities since 2003 and compute the average ICT time use intensity across income deciles¹. Analogous to the ICT share of expenditures in Figure ??, Figure A13 (a) shows the ICT share of time use along the income distribution. The ICT share in time use, while noisy, moves similar to the expenditure share between 4.3 and 4.6% and features a slight upward trend along the income deciles.

ATUS partly captures free digital products and services in the categories computer use, e-mail services, playing games and listening to music. The notion of "free digital good" relates to near zero marginal costs, once these goods are accessible. While the CEX allows measuring the entry costs for these goods, ATUS shows how much time households dedicate for these activities. We compute the log ratio between time shares and expenditure shares for each income decile, to assess whether a group dedicates relatively more time to free digital goods. The ratio is upward sloping in Figure A13 (b) which implies that rich households spend relatively more time on 'free' activities relative to their expenditures than poor households. This could suggest that entry costs (such as buying a computer or having internet access) are relevant, which is problematic for poor

¹The household income data of ATUS are reported in few income bins, which is why we consider deciles here.

households.

The results based on time use are in line with those based on expenditure data. This implies that even though we disregard some parts of consumption when focusing on expenditure, we still get a fairly accurate picture of the differences in ICT shares of the consumption baskets of rich and poor households. Considering the evidence on free digital goods and home production, the digital inequality based on expenditure data is likely to be understated. In Section ??, we assume that all consumption shows up in form of purchases of goods or services at a market price, which therefore constitutes a conservative estimate of the impact digitalization has on inequality.

B Model Appendix

B.1 Sectoral Classification

In the main part of the paper, we sort IO industries into ICT- and non-ICT sector based on their factor inputs, i.e. whether the industries *use* ICT capital. Below we explore two alternative classifications: (1) classification based on whether industries *produce* ICT capital, (2) a categorization following the structural change literature into agriculture, manufacturing and services (?) or alternatively into agriculture/manufacturing, high-skill services and low-skill services (?).

Classification Based on Capital Flows

Table 11 in NIPA's 1997 Capital Flow Table shows the input-output commodity composition of private equipment and software expenditure categories. We focus specifically on the assets which we classify as ICT assets in Table A1. Computers and peripheral equipment are produced by BEA industries 3340 (computer and electronic product manufacturing; about 94% of the total value of production) and 5415 (computer systems design and related services, the remaining 6%). Software is produced by 5110 (Publishing industries, except internet, including software; 33% of the total value of production) and by 5415 (66%), the rest falls on non comparable imports. About 80% of the value of non-ICT assets is produced by non-ICT industries. Industries 3340, 5415 and 5110 are a subset of our five ICT-using industries of the main part.

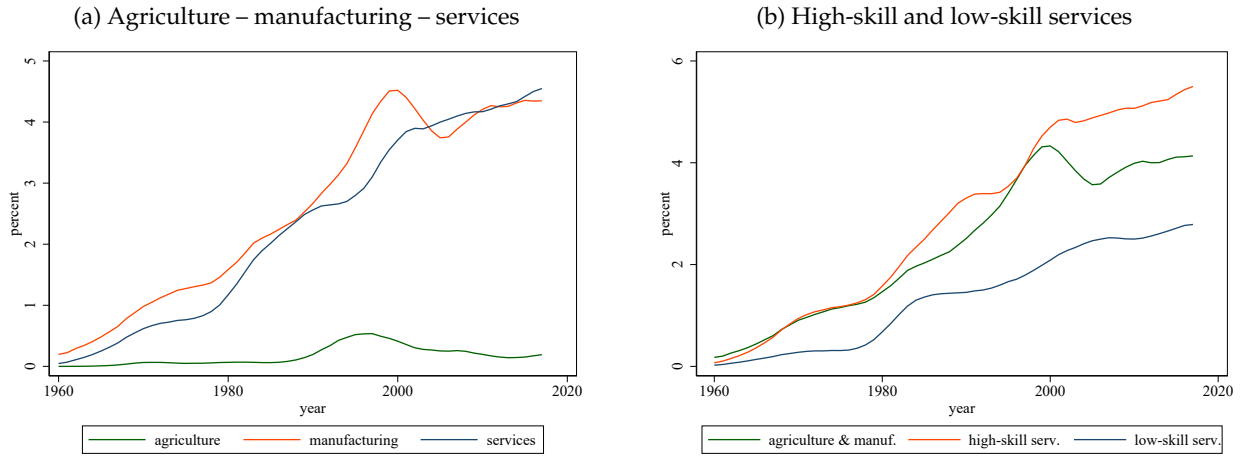
Classification Following the Structural Change Literature

Agriculture – manufacturing – services: Manufacturing and services have an almost identical ICT share (Figure B1 a). Whereas among the industries with the highest ICT share are mostly service industries (especially 5110, 5230, 5415, 5140), the industry on top of the list lies within manufacturing (3340, computer and electronic products manufacturing, with an average ICT share of 17.0%). This distinction is therefore not very informative about digitalization.

Agriculture and manufacturing – high-skill services – low-skill services: Like the definition in the main part, the high-skill vs. low-skill services distinction aims at the factor inputs, but in contrast to our definition, it focuses exclusively on labor inputs. ? classify services as high- and

low-skill based on the share of college graduates among the workers employed in the sector. They argue that growth in services is driven by the movement of consumption into more skill-intensive output. This is actually very similar to our framework: high-skill workers are complements to ICT capital, whereas routine workers (a subset of low-skill workers) is substitutable. In fact, all of the high-ICT service industries 5110, 5230, 5415, 5140 are high-skill services. Not surprisingly, the ICT share of high-skill services is higher than that of the other two categories in Figure B1 b). Thus, the rise in ICT-intensive production goes along with the rise in skill-intensive services. A major difference, however, is the role played by manufacturing, or specifically, the role played by the computer and electronic products manufacturing industry. Another difference is that high-skill services according to ? include health and education, which are low-ICT.

Figure B1: ICT shares of alternative industry composites



Note: The graph shows the ICT share of composite sectors. The left graph classifies sectors into agriculture, manufacturing (including also mining and construction) and services. The right graph classifies sectors into agriculture/manufacturing, high-skill services (comprising BEA industries 50*, 51*, 52*, 53*, 54*, 55*, 61*, 62* and 72* as well as 4810) and low-skill services (comprising all other services). *Source:* BEA and own calculations.

B.2 Calibration Algorithm

We calibrate the model in the following way:

1. Global search: We solve the model for 20,000 random draws from a holton set of the nine parameter values $\eta, \epsilon, \phi_1, \phi_2, \gamma_1, \gamma_2, \nu, \theta, \rho, g_{a1}, g_{a2}$ within permissible bounds. We define the loss function as the weighted sum of squared errors in the set of data moments (scaled by the corresponding moment). This means that we minimize the percentage error of moments. We choose a weighting matrix that over-weights by a factor of ten the following, more relevant, moments: the relative price, the skill premium, the wage premium of routine relative to manual workers, and expenditure share differences between the top 1 and bottom 1 percent.
2. Local search: We solve the model for 10,000 random draws in the vicinity of the 10 best solutions, obtain a new set of 10 best solutions and repeat this procedure three times. From the resulting final set of best solutions, we select the one where the loss function is smallest.

B.3 Link between Parameters and Moments

Several moments link directly to one or more parameters. The difference in the expenditure shares of the ICT-intensive good between two income groups is used as a moment to inform ρ . From eq.(??),

$$\frac{P_1 C_{1M}}{e_M} / \frac{P_1 C_{1H}}{e_H} = \left(\frac{e_H}{e_M} \right)^\rho \Delta_M$$

where we use as empirical approximation the expenditure shares of the top and bottom 10% of the income distribution as derived in Section ??.

θ is pinned down by the increase in the expenditure share of ICT-intensive goods over time. At any t, s ,

$$\frac{P_{1t} C_{1t}}{e_t} / \frac{P_{1s} C_{1s}}{e_s} = \frac{\left(\frac{P_{2t}}{e_t} \right)^\rho \left(\frac{P_{1t}}{P_{2t}} \right)^\theta \Delta_t}{\left(\frac{P_{2s}}{e_s} \right)^\rho \left(\frac{P_{1s}}{P_{2s}} \right)^\theta \Delta_s}$$

The relative quantities consumed of good 1 and good 2 are informative for all three parameters of the utility function:

$$\frac{C_1}{C_2} = \frac{\nu \frac{e_o}{P_1} \left(\frac{P_2}{e_o} \right)^\rho \left(\frac{P_1}{P_2} \right)^\theta}{\frac{e_o}{P_2} \left(1 - \nu \left(\frac{P_2}{e_o} \right)^\rho \left(\frac{P_1}{P_2} \right)^\theta \right)}$$

where $o = H, R, M$.

To pin down the parameters of the production function, we choose moments that are informative about the relative use of the inputs R, M, H, ICT and K (which can move freely across sectors).

The skill premium of H over R is given by

$$\frac{W_H}{W_R} = \frac{\gamma_i}{1 - \gamma_i} \frac{R_i^{\frac{1}{\eta}}}{H_i^{\frac{1}{\epsilon}}} AW^{\frac{\epsilon-1}{\epsilon} - \frac{\eta-1}{\eta}}$$

Using log difference, γ is not time varying and the premium is therefore strongly influenced by the two elasticities.

The wage premium of R over M additionally incorporates information about the two elasticities:

$$\frac{W_{Rt}}{W_{Mt}} = \frac{AW_{it}^{\frac{\epsilon-1}{\epsilon}} (1 - \gamma_i) (1 - \psi_i) \phi_i M_{it}}{\psi_i R_{it}^{\frac{1}{\eta}} AW_{it}^{\frac{\eta-1}{\eta}} SW_{it}^{\frac{\epsilon-1}{\epsilon}}}$$

Considering log differences, the two parameters that remain to match this moment are ϵ and η . The time series for W_{Rt}/W_{Mt} is used to match the employment share of routine and manual workers over time.

To match input factors and their returns, we choose the following additional moments: the ICT

income share in the aggregate economy, the return on ICT capital and the labor share, which is informative about the use of capital vs. labor in the production process. ? show that the decline in the labor share is fully explained by the rise in intellectual property products (whose rents are attributed to capital income by the BEA). A major share of intellectual property products is software. ?, ?, ? and ? also argue that automation is an important driver of the labor share. ? choose a similar calibration strategy to ours where they also require that automation explains the declining labor share and the increasing skill premium fully. Finally we match the evolution of the relative price from Figure ???. This is decisive for the overall strength of the price effect and also helps to inform other parameters in the production function. GDP per capita growth rates and sectoral value added shares then pin down the contribution of TFP growth.

B.4 Validation

Sector-specific Non-targeted Moments

Table B1 shows how the model matches additional non-targeted moments mostly relating to the sectoral composition. Compared to the ICT increase in the overall economy, the increase in the ICT intensive sector is smaller, both in the model and in the data. Furthermore, the increase in the share of hours worked in the ICT intensive sector of high-skill and low-skill workers is very well matched as well. The model slightly over predicts the increase in hours shares for both skill groups, which is probably due to the fact that workers can switch sector at no cost and at any time in the model, whereas in reality there could be substantial frictions connected with switching industry, such as acquiring new skills in the form of additional education or on-the-job training, building a new professional network, etc.

Table B1: Validation: Model and data increases by factors in sector 2

Moment	Data	Model
ICT intensity	3.9125	3.4559
Share hours H	1.064	0.89321
Share hours L	0.21221	0.21668

Note: All moments are percentage changes between 1960 and 2017. Sources for the data: ACS and BEA.

Data-Model Comparison along the Income Distribution

In Figure B2, we plot the model predictions against the data for income, prices and consumption expenditures. To highlight differences across the income distribution, we normalize the mean across low-skill percentiles to 1 for both data and model, thus interpreting each value relative to the average change. Since high-skill workers are homogeneous in the model, the values will appear flat on the right-hand side of the graphs, and there is a discrete jump between percentile 77 and 78. In reality, the change should be more gradual, and the values for percentiles 78-100 should be interpreted as averages across the top percentiles. Accordingly, we show data averages for percentiles 78-100.

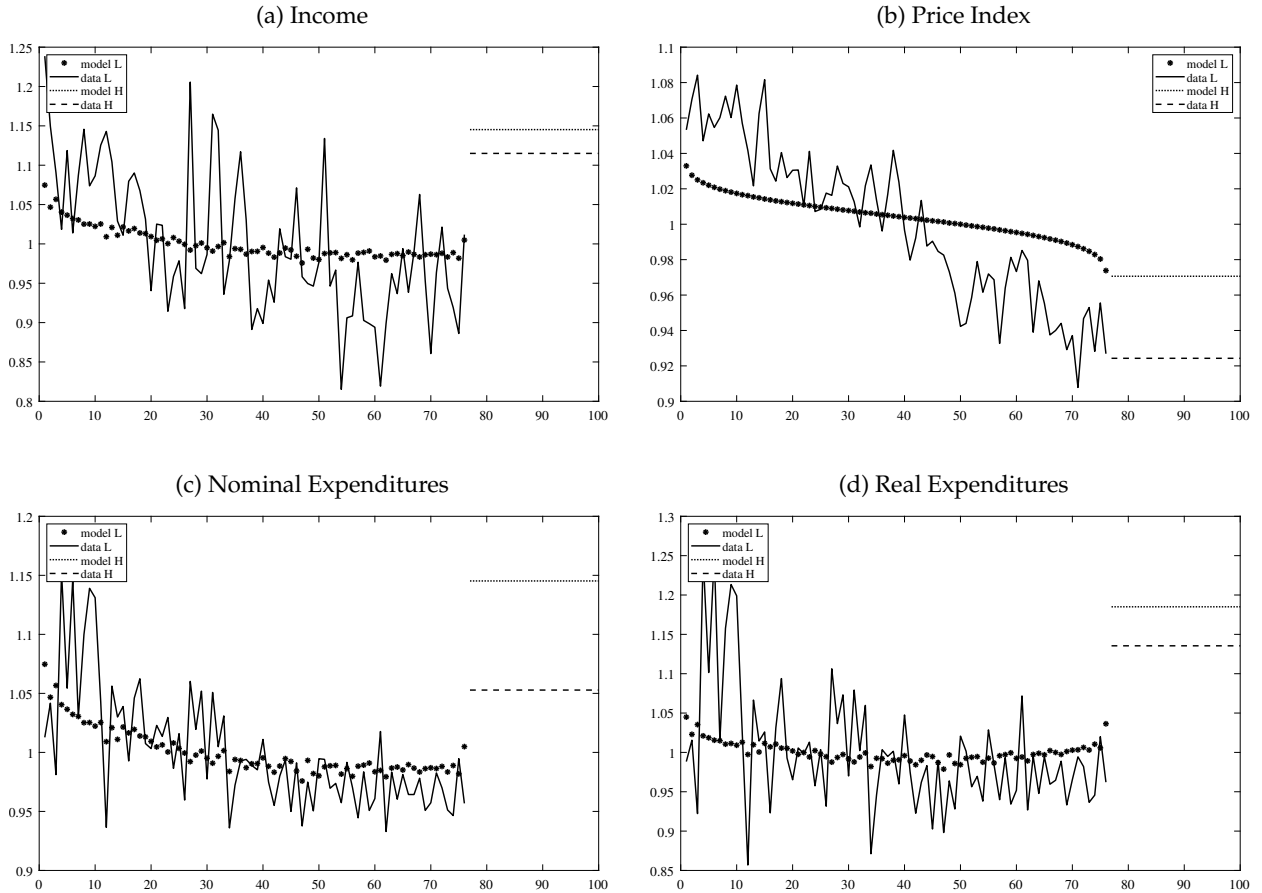
Figure B2 (a) plots the percentage change in real hourly wages between 1960 and 2017 in the ACS and Census against the model. The model captures both the negative slope of the low-skill households, with the flattening out of the curve towards the upper end of the distribution and the wage increase for high-skill households. Replicating the U-shaped polarization pattern, the bottom and top of the distribution experience an income growth 8% and 15% higher than the average, respectively, while large parts of the middle of the distribution experience relative income declines. Figure B2 (b) shows changes in the Törnqvist price index between 1960 and 2017 in the model and in the data. Be reminded that the model does not target the empirical differences in price indices directly, but matches the evolution of relative prices, the difference in expenditure patterns and the rise in income inequality. With these targets matched, the model generates a downward sloping pattern of the price index as in the data, although slightly less steep, indicating that the purchasing power of richer households has increased relative to poorer households.

Lastly, we focus on consumption inequality. In Figure B2 (c), we show expenditures that are normalized in the same way as panels (a) and (b). We use data from ?. In panel (d), we deflate in addition each percentile of consumption expenditures with the percentile-specific price index of panel (b). This means we measure changes in consumption expenditures by the price that people actually need to pay for their bundles. This approach approximates the real costs of living of the households, which is the more relevant concept for consumers.

We match well the data for the low-skill households. For high-skill households, the model captures that their expenditures are above the other percentiles, but the difference is slightly too large.

While expenditures in panel (c) differ by about 10 percent between the top and bottom of low-skill workers, this difference shrinks to almost zero when adjusting for percentile-specific price changes. In contrast, the difference between high-skill workers and the first percentile increases from about 10 percent to about 20 percent. While actual expenditures move closely with the evolution of income for lower income groups, the price effect partly overturns the pure income effect and leads to a different shape of the increase in real expenditures. This is especially true for the bottom part of the distribution that suffers a relative decline in purchasing power. This shows that it is important to consider real units when discussing consumption inequality. To the workers, it is not nominal expenditures that matter, but what they can actually buy with the money they spend.

Figure B2: Data and model along the income distribution



Note: The figure compares key model variables (dots/dotted lines) with data moments (solid line/dashed line). Percentiles 1–77 show the distribution of low-skill workers, percentiles 78–100 the homogeneous high-skill workers. The data and model series are normalized by the low-skill mean. Panel (a) shows log changes in income, (b) the Törnqvist price index, (c) log expenditure increase and (d) log expenditure increase deflated with the percentile-specific price index. We use our CEX data from Section ??, where we calculated price changes between 1984 and 2017. Because the time frame is longer now, we multiply each value with a factor 57/33. Sources: ACS, Census, NIPA, CEX, ?.

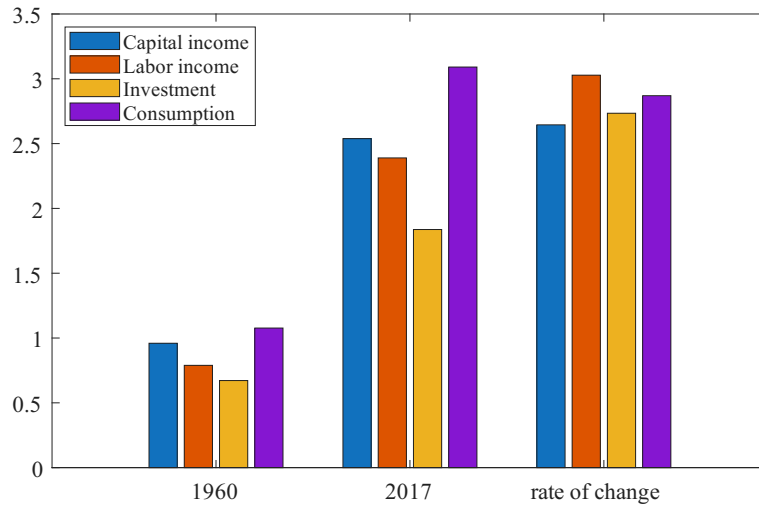
B.5 The Role of Capital Income and Investment

According to eq.(??), high-skill households receive two types of income, labor and capital income, which they need to allocate across income and investment. Is this choice consequential? Figure B3 shows the different types of incomes and expenditures for 1960 and for 2017 and their percentage change. In both years, capital and labor income each make for about half of total income of high-skilled households. They both increase to about 2.5 times the initial value. So the increasing skill premium and the rising capital share move in tandem. ? and ? also stress the importance of the increase in the capital share, and our findings are in line with theirs. Given that changes in both the labor share and the skill premium were targeted in the model calibration, this is not surprising.

High-skill households allocate about 60% of their income to consumption and the rest to investment. Importantly, both consumption and investment grow by about the same rate between 1960 and 2017. As the model is solved with a focus on changes, the fact that all four variables grow by about the same rate implies that it does not make a great difference for the income effect whether

capital and investment are included or excluded.

Figure B3: Income and expenditure of high-skill households



Note: The graph shows model predictions for per-capita capital income, labor income, investment and consumption of high-skill workers. Numbers are per-capita.

B.6 Counterfactual Exercise with Muted Labor Supply Growth

Table B2 compares model outcomes and data moments for a counterfactual where we mute labor supply growth. This means that both high-skill and low-skill labor stay at their 1960 value during the whole simulation period. This scenario results in a more pronounced income polarization. On the one hand, the skill premium now grows much faster. On the other hand, the premium of routine over manual wages declines much more. This result complements the exercise in Section ??, in which the labor demand increase was muted. In absence of an increase in (primarily high-skill) labor supply, the effects of digitalization on incomes would have been more extreme. Also noteworthy is that the labor share declines much less in this scenario. The declining labor share is thus not only driven by technological advances, but also by the fact that labor supply has grown.

Muting the increase in labor supply also results in a lower GDP per capita than in the baseline. This shows that complementarities between technology and labor are of crucial importance for GDP growth.

Lastly, note that the muted growth in labor supply has almost no effect on the relative price and on consumption patterns. Labor supply changes do not explain any of the price effect .

Table B2: Counterfactual $g_L = g_H = 0$, moments vs. data

Moment	Data	Model
Skill premium, 1960-2017	39.9	241.5
Relative price, 1960-2017	244.2	227.8
ICT intensity	643.1	526
Relative consumption, 1960-2017	-85.4	-88.7
Labor share, 1960-2017	-11.7	-4.2
Rate of return to ICT capital, 1960-2017	-99.7	-98.1
Median increase in exp. share good 2, 1960-2017	70.7	64.3
Percent diff., exp. share good 2 Top 1 Bot 1, 2017	23.3	26.7
Routine-manual wage premium, 1960-2017	-31.7	-52.9
GDP per capita, 1960-2017	281.6	159.9
Value-added share of sector 2, 1960-2017	71.4	57.9

Note: All moments are expressed in percentage changes, except for the difference in expenditure shares.