Pontificia Universidad Católica de Chile
Instituto de Economía
Magíster en Economía

TESIS DE GRADO
MAGÍSTER EN ECONOMÍA

Del Canto, Monge, Felipe Nicolás

Enero, 2020

Pontificia Universidad Católica de Chile
Instituto de Economía
Magíster en Economía

AGGREGATE PROBLEMS REQUIRE AGGREGATE SOLUTIONS? WHEN HETEROGENEITY IS EXPENDABLE

Felipe Nicolás Del Canto Monge

Comisión

Eugenio Bobenrieth

Felipe Zurita

Santiago, Enero de 2020

Aggregate problems require aggregate solutions? When heterogeneity is expendable

Felipe Del Canto M.*

January, 2020

> (...) all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind.
>
> *Empirical Model-Building and Response Surfaces (1987)*
> George Box and Norman Draper

Abstract

Aggregation is a tool used to reduce the complexity of economic models in order to draw more clear and succinct conclusions or simplify analyses. As any approximation, its use may be accompanied with errors researchers may not be willing to tolerate if they become aware of them. In this work I present how these errors appear using aggregation across goods and across consumers. To this end, I consider aggregation as a means to approximate probability distributions over parameters. Using this approach, I show ways to bound approximation errors by tailoring the parameters of the model. Further, I briefly discuss a methodology to study the goodness-of-fit of aggregate models in more general settings.

*Thesis written as a Master student at the Microeconomics thesis seminar, Department of Economics, Pontificia Universidad Católica de Chile. This work was funded by the CONICYT PFCHA/MAGISTER NACIONAL/2018 - 22180053. Any comment to the author's email address: fndelcanto@uc.cl

1. Introduction

Every model is by definition a simplified reality. On the bright side, abstracting from the complexity of the real world has allowed society to understand the sometimes subtle mechanisms that rule nature and human behavior. This does not mean that a model is useful for every purpose. Evidently, whilst some of them can illustrate certain dynamics of the real world very clearly, approximation in itself may carry errors that harm predictions. This last comment points directly to the question of which model is the most useful for a given problem. In particular, when the answer is *many* the modeler needs to make a choice based on the results she expects to highlight and the channels she wants to study. The dilemma is by no means alien to the field of economics. When describing an economy, the researcher faces several different assumptions that shape the complexity of the model. Although some of them may be made for feasibility reasons (e.g., because a highly detailed model cannot be solved or simulated or because data is not available to calibrate it), others serve a transparency purpose, that is, they intend to highlight the most important results without dwelling on the unnecessary details. Consequently, in the process of constructing a model, the investigator may choose to follow Occam's Razor principle: among the models that are consistent with the evidence, choose the one that makes the fewest assumptions. This criterion implies that the measure of a (correct) model is its complexity. However, as Milton Friedman said, "The ultimate goal of a positive science is the development of a 'theory' or, 'hypothesis' that yields valid and meaningful (...) predictions about phenomena not yet observed" and thus "Its performance is to be judged by the precision, scope, and conformity with experience of the predictions it yields".[1] Hence, in evaluating the validity of a theory, the robustness of its conclusions should be of critical importance.

Different strands in the economic literature have studied when predictions derived from a model are robust to different specifications. In Sutton (2007), the author discusses which mechanisms in the context of industrial organization still hold in conditions outside the classical models of, for example, Cournot and Bertrand. A similar motivation can be found in Kajii and Morris (1997), where they study how sensitive game theory conclusions are to the assumption of common knowledge of payoffs in a game. The interest in robustness in the context of mechanism design can be also found in ter Vehn and Morris (2011). An interesting approach is the one in Basu and Fernald (1997) where the authors try to estimate discrepancies due to "aggregation effects" when considering a model with a representative firm and one where heterogeneous effects

[1] Friedman (1953).

are considered. Similarly, in Hanushek et al. (1996) the authors try to reconcile contradictory results in the literature on estimating the value added by schools arguing an important role of aggregation in the magnitude of omitted variable bias, which can in principle invalidate previous estimations. Related to this literature there is a concern with models that make use of aggregated data. Different authors have studied what is called "aggregation bias" arguing that these models hide important mechanisms that could explain these differences.[2] All in all, different authors have tried to untangle the differences between predictions and realizations by appealing to the goodness-of-fit of the models used to produce such estimations. In the cases where the deviations are substantial, a revision to the model must be made.

As an example, consider the case of the representative agent model. The assumption that there is only one consumer in the economy is useful and has been key to understand important qualitative results, especially in macroeconomics. Nevertheless, employing this model to predict future realizations of certain key variables such as aggregate demand or marginal propensity to consume (MPC) may be inaccurate if heterogeneity effects are in place. In other words, there is a shadow price in the approximation (which the investigator can be willing to pay or not) if she wishes to use the model for another, more quantitative-driven purpose. This point is made clearly in Carroll (2000) in the context of the buffer-stock model: "Representative-agent models are typically calibrated to match an aggregate wealth-to-income ratio" but "the typical household's wealth is much smaller than the wealth of such a representative agent (...), this would lead one to expect that the behavior of the median household may not resemble the behavior of a representative agent with a wealth-to-income ratio similar to the aggregate ratio". The evidence quickly backs this view up: while the annual MPC predicted by the representative agent model is about 0.04, many empirical analyses estimate this parameter to lie between 0.2 and 0.5.[3]

The aforementioned model is a particular case of a common practice in economics: aggregation. The other canonical example of its use is aggregation across goods, where instead of describing the myriad of goods available in an economy they are grouped into one or more categories. Regarding these two implementations, precedent theoretical literature focused on one side of the problem: When is it possible to carry out this practice and describe precisely the same economy. In the case of the representative agent, the necessary and sufficient condition is that the indirect utility function of every consumer has the Gorman form.[4] When aggregation

[2] See, for example Lee et al. (1990); Ravallion (1998); Feenstra and Hanson (2000); Imbs et al. (2005).
[3] Carroll (2000).
[4] Gorman (1953).

is applied to goods, the answer has been more elusive but two results have arisen. First, the Hicks-Leontief (composite commodity) theorem allows aggregation if relative prices are constant within the group of goods that are to be bundled.[5] A somewhat weaker requirement is proposed by Lewbel (1996): bundling is feasible if all group relative prices are independent of price indexes and income. The second answer states that grouping some goods is allowed if preferences between them are "independent" of the remaining goods present in the economy.[6]

For the two kinds of aggregation mentioned before, the conditions for them to hold are highly restrictive and not typically met in econometric or theoretical applications. As mentioned before, the literature has assumed (disregarding these conditions) exact aggregation of both types in constructing models and making econometric estimations and this practice comes at a price. As stated in the preceding discussion, the validity of these models is directly attached to the size of such cost. If the question an investigator is seeking to answer allows for the use of aggregate models without incurring in a severe deviation from predictions, then not having exact aggregation is of minor importance. In other words, compliance of the conditions for aggregation is not a problem as long as the parameters of the simplified model are calibrated in such a way that the estimation error is below some tolerance predefined by the modeler. Hence, understanding and quantifying these differences is crucial in determining and measuring the goodness-of-fit of these models. Consequently, and in contrast with some of the articles mentioned earlier, in this work I intend to give a theoretical look at how these deviations appear using simple models, and understand how a researcher might limit them. To this end, the main insight I employ is that aggregation is trading off heterogeneity for simplicity and from here is where errors make their appearance. Explicitly, I model heterogeneity as a probability distribution over the relevant parameters of the economy, where relevance is understood depending on the research question being asked. In that context, aggregation is replacing the distribution of unknown relevant parameters by a Dirac distribution centered at a point chosen by the modeler. Thus, the approximation error is closely related to the nature of the original and the approximate distributions, specifically to some finite set of moments of them.

Reformulating the problem in this manner has consequences beyond the particular case of aggregate models. Diversity appears naturally in many applications and hence modeling it as a distribution allows the investigator to define the problem in clearer terms. After determining the relevant sources of heterogeneity in the specific context, it only remains to pick an appropriate

[5] Leontief (1936); Hicks (1946).
[6] See for example Gorman (1959).

measure of the error that, conditional on the precision desired, narrows the range of distributions available for replacement. The advantage of this line of thinking is that the optimal metric can be created following the Friedman guidelines mentioned above. This way, models are evaluated under the scope of the quality of their predictions and not on their complexity. Moreover, the methodology proposed here serves also as a decision rule, leaving arbitrary choices outside the modeling process.

The rest of the paper proceeds as follows. To motivate the subsequent discussion, in Section 2 I summarize the problems and existing results regarding conditions under which aggregation is permitted. Then, I present three settings to illustrate how approximation errors appear when using aggregate models to estimate future economic variables. First, in Section 3, I propose a representative agent model in the context of aggregate demand estimation. Second, in Section 4, I study a different setting with a representative agent model but where the objective is to estimate the MPC. The third model, in Section 5, mixes the two canonical types of aggregation by presenting an economy composed of several individuals consuming various goods, where good aggregation takes a different form for each of the agents. Finally, in Section 6 I discuss how to generalize the problem of heterogeneity approximation, detailing the precise methodology that must be followed in order to reformulate it.

2. Existing results in aggregation

The economic literature has recognized two forms of aggregation. First, the problem of consumer aggregation or the representative agent model seeks to describe the aggregate demand of a multi-person economy by focusing only on the aggregate determinants of demand (e.g., the aggregate income), as opposed to the distribution of such variables. The second class of aggregation focuses on describing demands for categories of goods without distinguishing the individual consumption of each element in the group. I refer to this last problem as the "Hicksian aggregation problem".

Both kinds of aggregation are widely used in economic models. The representative consumer is a salient feature of macroeconomic models, like the ones developed by Robert Lucas in 1978 to study asset prices and by Kydland and Prescott in 1982, that began the theory of real business cycles.[7] These authors also make use of aggregation across goods by employing a two-good model, although this feature has also been widely used to study the behavior of savings (in

[7] See Lucas (1978) and Kydland and Prescott (1982).

particular, precautionary savings during the life cycle) or consumption dynamics.[8] Empirical studies also make use of both forms of aggregation. Some of them assume all consumers are equal – which is an example of consumer aggregation – and others overcome problems in the data (e.g., availability or comprehensiveness) by stating that agents choose consumption on the category and not on individual goods.[9]

In what follows I formally describe the problems about aggregation the early literature tried to answer. These questions aimed at finding conditions that ensured aggregation did an exact fit of a heterogeneous model. In order to describe both problems I rely heavily on Varian (1992).

2.1. The representative agent problem

Consider an economy composed by n consumers indexed by $i = 1, \ldots, n$. Their demand functions of the k goods in the economy are summarized in the vector $\mathbf{x}_i(\mathbf{p}, y_i)$, where $\mathbf{p}$ is the price vector of the goods and y_i is the income of agent i. The aggregate demand vector is defined by

$$\mathbf{X}(\mathbf{p}, y_1, \ldots, y_n) = \sum_{i=1}^{n} \mathbf{x}_i(\mathbf{p}, y_i). \tag{1}$$

The question that automatically arises is: Can this aggregate demand function be regarded as the one generated by a single (or "representative") consumer? According to (1), the previous question is equivalent to looking for conditions under which $\mathbf{X}$ does not depend on the distribution but on the aggregate income

$$Y := \sum_{i=1}^{n} y_i.$$

The definitive answer to this problem came with Gorman (1953). According to his result, $\mathbf{X}$ is a function of Y if and only if for every $i \in \{1, \ldots, n\}$ the indirect utility function has the Gorman form

$$v_i(\mathbf{p}, y_i) = a_i(\mathbf{p}) + b(\mathbf{p})y_i,$$

where a_i, b are functions that must only depend on $\mathbf{p}$ and b has to be the same across consumers. This functional requirement is somewhat restrictive but at least two particular examples are worth mentioning: homothetic and quasilinear utility functions. For the first, the indirect

[8] See Carroll (1992), Gourinchas and Parker (2002), Gul and Pesendorfer (2004), Parker and Preston (2005) for some examples.

[9] See Kaplan and Violante (2014), Berger and Vavra (2015), Kan et al. (2017), Fagereng et al. (2017) for some examples.

utility functions is

$$v(\mathbf{p}, y) = b(\mathbf{p})y, \tag{2}$$

whereas for the second

$$v(\mathbf{p}, y) = a(\mathbf{p}) + y.$$

Both examples clearly have the Gorman form. However, some homothetic functions can differ in the function v between consumers and therefore aggregation is not possible. As an example, for $k = 2$, an economy of consumers with Cobb-Douglas utility functions

$$u(x_1, x_2) = x_1^{\alpha_i} x_2^{1-\alpha_i},$$

where at least two α_i are different does not allow for aggregation. In this case the function $v(\mathbf{p})$ in (2) is different among individuals and the Gorman form is not present.

Observe that the main objective of the representative agent problem is to find an alternative model that is coherent with the original. That is, this new, simpler model has to describe the exactly same economy as the disaggregate one. Backing this task is the idea that models should fit exactly in the context they intend to describe. However, since the original model is an approximation by itself, a more consistent approach is to consider the second fit as an additional layer of estimation and weigh its suitability in terms of the accuracy loss and the potential additional insight obtained.

2.2. Hicksian aggregation problem

For this problem consider the following setting. Assume the consumption vector of some agent is divided in two bundles $(\mathbf{x}, \mathbf{z})$. Accordingly, the price vector is separated into $(\mathbf{p}, \mathbf{q})$. Thus, if the utility function of the consumer is u, then the demand for the $\mathbf{x}$ goods is

$$\begin{aligned} \mathbf{x}(\mathbf{p}, \mathbf{q}, y) = \arg\max_{\mathbf{x},\mathbf{z}} \quad & u(\mathbf{x}, \mathbf{z}) \\ \text{s.t.} \quad & \mathbf{p}\mathbf{x} + \mathbf{q}\mathbf{z} = y. \end{aligned} \tag{3}$$

In numerous models, there is no interest in the consumption of each of the $\mathbf{x}$-goods but only in the demand for the bundle (e.g., focus on expenditure in savings against expenditure in different financial instruments). Hence, the Hicksian aggregation problem is finding conditions under which this approximation can be made without losing information. This implies finding a quantity index $X = g(\mathbf{x})$, a price index $P = f(\mathbf{p})$ and a new utility function $U(X, \mathbf{z})$ such

that the solution

$$X(P, \mathbf{q}, y) = \arg\max_{X, \mathbf{z}} \quad U(X, \mathbf{z}) \qquad \text{s.t.} \quad PX + \mathbf{qz} = y. \tag{4}$$

satisfies

$$X(f(\mathbf{p}), \mathbf{q}, y) = g(\mathbf{x}(\mathbf{p}, \mathbf{q}, y)).$$

In other words, the objective is to find an alternative model that is coherent with the original. This brings out again the discussion at the end of the preceding section: Aggregation can be regarded as a second layer of approximation and pondered in terms of its benefits and costs. In this particular setting, the nature of the approximation is different, but the essential objective of simplifying the world remains.

At least two situations exist under which these alternative models can be found: functional and Hicksian separability. In the first, assume that the preference relation represented by u has the following "independence" property

$$(\mathbf{x}, \mathbf{z}) \succ (\mathbf{x}', \mathbf{z}) \Longleftrightarrow (\mathbf{x}, \mathbf{z}') \succ (\mathbf{x}', \mathbf{z}') \qquad \forall\ \mathbf{x}, \mathbf{x}', \mathbf{z}, \mathbf{z}'. \tag{5}$$

This independence property implies that there exists a function v such that

$$u(\mathbf{x}, \mathbf{z}) = U(v(\mathbf{x}), \mathbf{z}),$$

where $U(v, \mathbf{z})$ is increasing in v. In this case, the consumer values consumption of $\mathbf{x}$ only through v. For example, consider $\mathbf{x}$ to be different types of food. Independence in this context can manifest if the agent only values the amount of calories she gets from $\mathbf{x}$. In that situation, all kinds of food are valued differently according to the amount of calories per unit each of them give but the consumer only cares about the total caloric intake. The preceding example suggests a particularity of this type of separability. Following the same nomenclature: The consumer chooses the amount of calories and the total amount to spend on food first and then she decides the individual consumption of each kind of food. Models that show this feature are often referred to as hierarchical consumption models.

By calling $m_{\mathbf{x}} := \mathbf{p} \cdot \mathbf{x}(\mathbf{p}, \mathbf{q}, y)$, it can be shown that the following equality holds

$$\mathbf{x}(\mathbf{p}, \mathbf{q}, y) = \arg\max_{\mathbf{x}} \quad v(\mathbf{x}) \qquad \text{s.t.} \quad \mathbf{px} = m_{\mathbf{x}}.$$

Therefore, if $e(\mathbf{p}, v)$ is the expenditure function of the previous problem, then

$$\begin{aligned} v(\mathbf{x}(\mathbf{p}, \mathbf{q}, y)) = \arg\max_{v, \mathbf{z}} \quad & U(v, \mathbf{z}) \\ \text{s.t.} \quad & e(\mathbf{p}, v) + \mathbf{q}\mathbf{z} = y. \end{aligned}$$

However, note that the latter problem does not have the same structure as (4). This only happens if v has a particular structure such that

$$e(\mathbf{p}, v) = e(\mathbf{p})v.$$

This property holds if, for example, v is homothetic and thus the Cobb-Douglas utility function is an example in which this kind of aggregation is admissible. This discussion shows that under functional separability this consumption model is hierarchical.

On the other hand, Hicksian separability is present in the following situation. Assume that $\mathbf{p} = t\mathbf{p}_0$, where t is a scalar and $\mathbf{p}_0$ is a fixed price vector. Letting $P := t$ and $X := \mathbf{p}_0\mathbf{x}$, define the following indirect utility function

$$\begin{aligned} V(P, \mathbf{q}, y) = \arg\max_{\mathbf{x}, \mathbf{z}} \quad & u(\mathbf{x}, \mathbf{z}) \\ \text{s.t.} \quad & P\mathbf{p}_0\mathbf{x} + \mathbf{q}\mathbf{z} = y. \end{aligned}$$

The solution U of the dual problem

$$\begin{aligned} U(X, \mathbf{z}) = \arg\min_{P, \mathbf{q}} \quad & V(P, \mathbf{q}, y) \\ \text{s.t.} \quad & PX + \mathbf{q}\mathbf{z} = y, \end{aligned}$$

by definition satisfies

$$\begin{aligned} V(P, \mathbf{q}, y) = \arg\max_{X, \mathbf{z}} \quad & U(X, \mathbf{z}) \\ \text{s.t.} \quad & PX + \mathbf{q}\mathbf{z} = y. \end{aligned}$$

showing that aggregation is possible with $g(\mathbf{x}) = \mathbf{p}_0\mathbf{x}$ and $f(\mathbf{p}) = t$. This result, presented by both Leontief (1936) and Hicks (1946), has been called the Hicks-Leontief theorem. As with functional separability, note that the theorem presents sufficient but not necessary conditions to find the functions g, f and U. In his work, Lewbel (1996) relaxes the assumptions of the Hicks-Leontief theorem by not asking for constant relative prices through time but instead for them to be independent of both the price index and income. In terms of data analysis,

the generalized composite commodity theorem of Lewbel does not ask for a correlation of one between the price of the **x**-goods and their index.

One of the most common applications of good aggregation under Hicksian separability assumptions are two-good models. Here, the interest is in the demand for one good while bundling all the other goods in only one category. These kind of models are frequently used in the macroeconomic literature where the study of aggregate consumption dynamics is greatly simplified by assuming agents just consume and save (and hence there are only two goods available). Observe that for this to happen and in accordance with the conditions in Lewbel (1996) it must be true that relative prices of all goods in the economy are independent of income and the price index. This requirement is stringent at its best. It is conceivable that high-income consumers suffer changes in some relative prices that low-income ones do not. This last situation crucially depends on the nature of the goods being bundled. In the extreme case where the two-good model is employed, the definition of the category *consumption* is too broad to ensure the conditions for the generalized theorem hold.

3. Aggregation across consumers

3.1. Description of the economy

Consider a two period economy ($t = 0, 1$) composed by agents that consume two goods $\mathbf{x}_t = (x_t, z_t)$ valued at prices $\mathbf{p}_t = (p_t, q_t) \in \mathbb{R}^2_{++}$. For simplicity, all subscripts t are omitted unless needed. Each individual in this setting is identified with a pair (y, α), where y is her income and α determines the form of her Cobb-Douglas utility function

$$u_\alpha(\mathbf{x}) := u_{(y,\alpha)}(\mathbf{x}) = x^\alpha z^{1-\alpha}.$$

Both y and α follow a joint distribution, F, with support $S \subset (0, \infty) \times (0, 1)$. Observe that in principle y and α can be correlated. The marginal distributions are F_y and F_α, respectively.

For every fixed vector $\mathbf{p}$, the agents choose the consumption bundle $\mathbf{x}(\mathbf{p}, y, \alpha)$ by maximizing u over $\mathbf{x}$ subject to their respective budget restriction. Specifically, the agent identified with the pair (y, α) solves

$$\begin{aligned} \max_{\mathbf{x}} \quad & u_\alpha(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{p} \cdot \mathbf{x} = y. \end{aligned}$$

Hence, for good 1

$$x(p, y, \alpha) = \alpha \frac{y}{p},$$

and thus its aggregate demand is

$$X(\mathbf{p}, F_y, F_\alpha) = \frac{1}{p} \int_S y\alpha \, dF(\alpha, y) = \frac{1}{p} \mathbb{E}[y\alpha] = \frac{1}{p} \Big(\mathbb{E}[y]\mathbb{E}[\alpha] + \mathrm{Cov}(y, \alpha) \Big).$$

Similar expressions are obtained for the personal and aggregate demands of good z.

3.2. The researcher's problem

Suppose now an investigator wishes to describe the aggregate demand of good 1 at time 1. She has access to a full description of the income distribution of the agents in the economy in $t = 0$ (that is, she knows F_{y_0}) but preferences at any time are not available to her. She also knows X_0, the aggregate demand of good 1 at time 0.[10] In this setting, if X_1 is the aggregate demand for cellphones then our researcher wishes to estimate consumption of mobiles at $t = 1$ by using data available at $t = 0$ on income and aggregate consumption of these devices.

Since individual preferences are unknown to her, she assumes all individuals are equal, that is, there exists $\overline{\alpha}$ such that every agent in this economy has utility function $u_{\overline{\alpha}}$. Given that aggregation is possible when the parameters of the Cobb-Douglas utility function are the same for all individuals (see Section 2.1), this assumption is equivalent to saying only one person exists in the economy. Put differently, what the researcher is doing is approximate F_α by $\delta_{\overline{\alpha}}$ where δ_x is the Dirac distribution centered at x. Analogously, she is approximating F with a joint distribution G over the same support S and with marginals F_y and $\delta_{\overline{\alpha}}$. This alternative interpretation goes in line with the discussion at the end of Section 2.1, where the use of aggregate models is an additional layer of approximation within the construction of a model.

Under the previous assumption, the representative agent has utility function

$$U_{\overline{\alpha}}(X, Z) = X^{\overline{\alpha}} Z^{1-\overline{\alpha}},$$

and since $E[y_t]$ is the aggregate income of this economy at time t, the best estimation for X_t is

$$\widehat{X}_t(p_t, F_{y_t}, \overline{\alpha}) = \overline{\alpha} \, \frac{E[y_t]}{p_t}.$$

[10] Observe that by assuming complete knowledge on the distribution of income and on the aggregate demand I am abstracting the discussion from the appearance of estimation errors due to sampling.

3.3. The estimation error

As mentioned in the introduction, the aggregation assumption imposes a shadow price on the estimation. The error in this case arises from the differences between the individual demands for x_1 and the ones estimated by the simplified model. Indeed, observe that

$$\begin{aligned} X(p,F) - \widehat{X}(p,F_y,\overline{\alpha}) &= \int_S \frac{y\alpha}{p} - \frac{y\overline{\alpha}}{p}\, dF(y,\alpha), \\ &= \int_S x(p,y,\alpha) - x(p,y,\overline{\alpha})\, dF(y,\alpha), \end{aligned}$$

and thus the accuracy in the estimation, although observable in macroeconomic variables, has its roots in microeconomic differences. Moreover, conditional on having the correct distribution of income, this discrepancy depends only on the choice of $\overline{\alpha}$.

Note that the monetized error (with sign) in the estimation at time t, as a function of $\overline{\alpha}$ is

$$D_t(\overline{\alpha}) := p_t\Big\{X_t(p_t,F_{y_t},F_{\alpha_t}) - \widehat{X}_t(p_t,F_{y_t},\overline{\alpha})\Big\} = \mathbb{E}[y_t]\left\{\mathbb{E}[\alpha_t] + \frac{\mathrm{Cov}(y_t,\alpha_t)}{\mathbb{E}[y_t]} - \overline{\alpha}\right\}, \tag{6}$$

and observe that this difference does not depend on prices. Consequently, the error can be reduced to zero if

$$\overline{\alpha} = \underbrace{\mathbb{E}[\alpha_t] + \frac{\mathrm{Cov}(y_t,\alpha_t)}{\mathbb{E}[y_t]}}_{:=M_t}.^{11} \tag{7}$$

Observe that this equation says that taking $\overline{\alpha}$ as the expectation of α_t is not enough. Instead, some heterogeneity correction (in the form of the covariance of income and preferences) is necessary to obtain an exact fit at time t. Recalling the interpretation of the single consumer assumption as a way to approximate F_t, then (7) says that the only features of the distribution that are important to have an exact fit at time t are its first and second moments. This finding suggests that what matters when estimating aggregate variables (with a certain degree of precision) using representative consumer models are a finite set of statistics of the underlying distribution of the population.

Since the researcher has access to X_0, then $\overline{\alpha}$ can be chosen optimally to reduce the estimation error at $t = 0$. Indeed, the optimal choice of $\overline{\alpha}$ given this information is M_0 and

[11] Note that $M_t = \mathbb{E}[\alpha_t] + \dfrac{\mathbb{E}[y_t\alpha_t] - \mathbb{E}[y_t]\mathbb{E}[\alpha_t]}{\mathbb{E}[y_t]} = \dfrac{\mathbb{E}[y_t\alpha_t]}{\mathbb{E}[y_t]}$. This is positive because y_t and α_t are. M_t is also smaller than 1 because $\alpha_t < 1$ and thus $\dfrac{\mathbb{E}[y_t\alpha_t]}{\mathbb{E}[y_t]} < \dfrac{\mathbb{E}[y_t]}{\mathbb{E}[y_t]} = 1$. Hence, $M_t \in (0,1)$.

consequently, by (6), the estimation error in $t = 1$ is

$$E[y_t] \Big\{ M_1 - M_0 \Big\}. \tag{8}$$

If preferences do not change, then letting $\alpha := \alpha_0 = \alpha_1$, it follows that

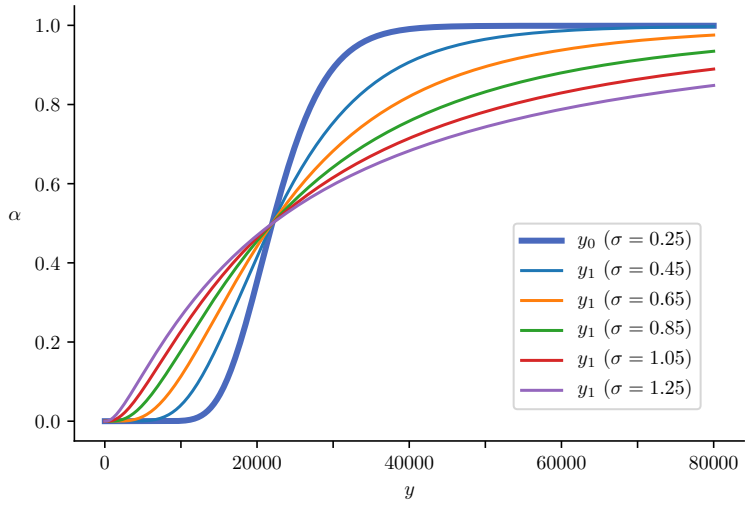
$$M_1 - M_0 = \frac{\text{Cov}(y_1, \alpha)}{\mathbb{E}[y_1]} - \frac{\text{Cov}(y_0, \alpha)}{\mathbb{E}[y_0]}. \tag{9}$$

Moreover, if $y_1 = Ry_0$ for some $R > 0$, then $\overline{\alpha} = M_0$ eliminates the estimation error at $t = 1$.

3.4. An example

To illustrate the appearance and magnitude of the estimation error in a model as the one described above, consider the following setting. Preferences are the same in both periods and $\alpha \sim U(0, 1)$ (the marginal distribution of α is uniform). At $t = 0$, the income follows a log-Normal distribution with parameters (μ, σ_0) with $\mu = 10$ and $\sigma_0 = 0.25$. The joint distribution F_0 is such that $S_0 = \{(y, \alpha) : P(y_0 \leqslant y) = \alpha\}$. This condition implies that $\text{Cov}(y_0, \alpha) \geqslant 0$. To construct this joint distribution I take 10^7 draws from each marginal distribution, then sort each list in ascending order to finally pair each of the 10^7 points of each distribution in the order given in the previous step. The support of the distribution constructed this way can be seen in the thick line in Figure 1.

Figure 1: Support of F_0 and F_1



The thick line represents the joint distribution at $t = 0$. The thin lines are five different examples of joint distributions at $t = 1$, varying in the value of σ_1.

Under these parameters and according to (7), the optimal choice of $\overline{\alpha}$ at $t = 0$ is

$$\overline{\alpha} = \mathbb{E}[\alpha_0] + \frac{\text{Cov}(y_0, \alpha_0)}{\mathbb{E}[y_0]} = 0.57, \tag{10}$$

which is 14% bigger than the mean of the distribution of α.[12] Now, suppose that between $t = 0$ and $t = 1$ there is a redistribution of income between each of the two halves of the distribution determined by the median. In particular, I assume there is a "transfer" from the poorest to the richest half. A modification like this can occur for example, after an unequal growth in income or after a financial crisis. To generate this change I take $y_1 \sim \log\text{-Normal}(\mu, \sigma_1)$, with $\sigma_1 \geqslant \sigma_0$. To see the effects that different values of σ_1 have in the estimation error I consider $\sigma_1 \in [0.25, 1.25]$. Note that both y_0 and y_1 have the same median but the tails of both distributions are heavier, as shown in Figure 2.[13] I assume that the joint distribution at $t = 1$, F_1, still satisfies $S_1 = \{(y, \alpha) : P(y_0 \leqslant y) = \alpha\}$ and thus I construct it using the same algorithm. This assumptions means that the redistribution did not change the rank of each individual. In order to illustrate the differences between S_0 and the different supports S_1, in Figure 1 I show both sets, using the same values for σ_1 as in Figure 2. From Figure 1 it is clear that given that more mass is moved to the tails of the income distribution, preferences grow more quickly at the beginning of the interval $(0, \infty)$ and higher values of α are restricted to high values of y.

It follows from Figure 1 that the covariance between α and y_1 changes depending on the value of σ_1. Hence, according to (8) and (9), the estimation error changes with σ_1. In Figure 3 I show the relationship between the estimation error, $D(\overline{\alpha})$ and σ_1 for $0.25 \leqslant \sigma_1 \leqslant 1.25$ in monetary terms. The maximum difference is close to 10 000 which representes nearly half of the median income.[14] The figure also shows that the error is monotone in σ_1 and weakly convex, as can be seen from the proximity of the curve to the straight line that joins the extremes. Finally, observe that the error is positive for all the considered values of σ_1 and the reason comes directly from the particular construction of this example. Since the covariance between y_0 and α is always positive and grows (or stalls) at $t = 1$, then $M_1 \geqslant M_0$ for all values of σ_1 (see (7)) and so, according to (8), $D_1(\overline{\alpha}) \geqslant 0$ for every σ_1.

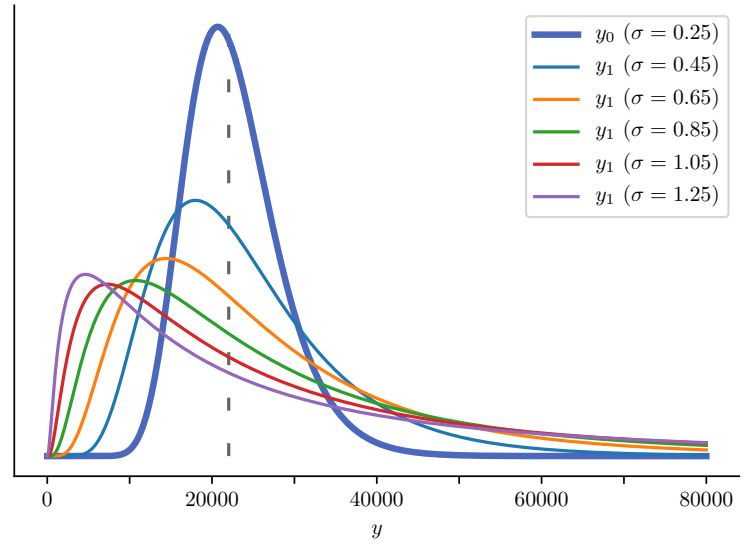
As noted in (10), the value of $\overline{\alpha}$ is slightly over the mean value of the preference marginal distribution. From the previous exercise at least two questions arise: How different is the optimal $\overline{\alpha}$ in (10) from those needed to reduce the error to zero at $t = 1$? Is it possible, in the

[12] Since $\alpha \sim U(0, 1)$, $E[\alpha] = 0.5$.

[13] The median of a log-Normal distribution with parameters μ and σ is e^{μ}. For $\mu = 10$, we have that $e^{\mu} \approx 22\,026$.

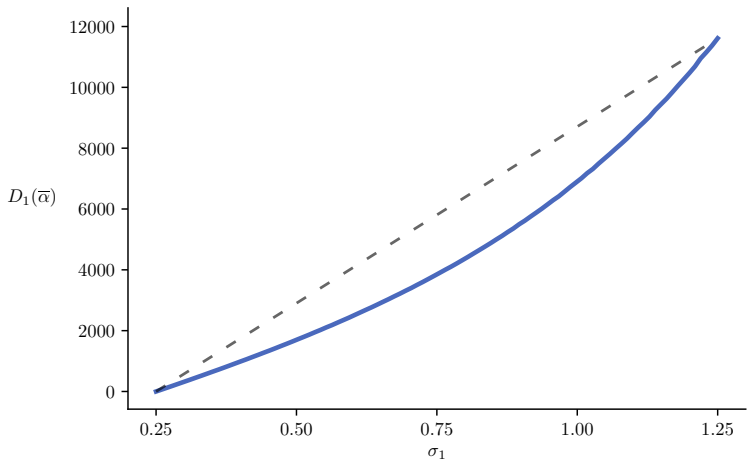
[14] 10 000 is approximately 0.45 times the median income.

Figure 2: Densities of y_0 and y_1 for different values of σ_1



All distributions have $\mu = 10$ and σ according to the legend. The thick line represents the income distribution at $t = 0$. The thin lines are five different examples of distributions at $t = 1$. The dashed gray line corresponds to the common median at, approximately, 22 026.

Figure 3: Estimation error $D_1(\overline{\alpha})$ as a function of σ_1



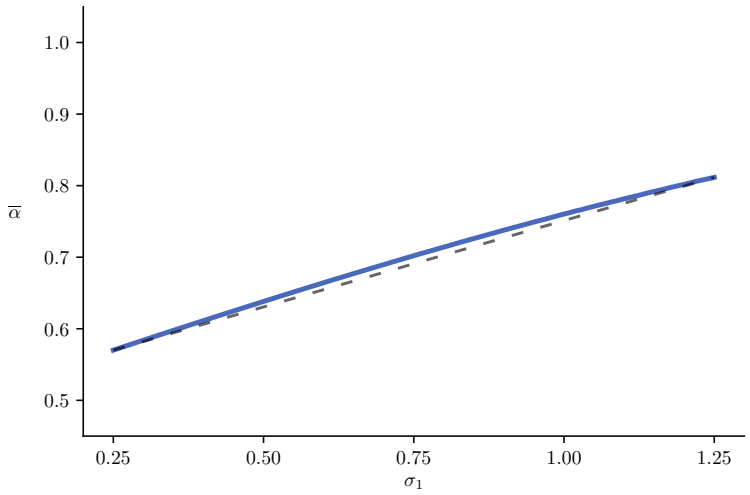
The blue line shows $D_1(\overline{\alpha})$ as a function of σ_1. The dashed gray line joins the two extremes of the curve.

context of this setting, to obtain optimal $\overline{\alpha}$ that completely cover the $(0, 1)$ interval?

Both questions are closely related, since each one requires studying the response of $\overline{\alpha}$ to changes in the value of σ_1. However, the answers point to different aspects of this setting. The solution to the first question may give additional insights about how sensitive is the model to the choice of parameters. For example, if the relationship between the optimal $\overline{\alpha}$ at $t = 1$ and σ_1 is strongly convex, then a representative agent model is very reactive to even the smallest changes in the underlying heterogeneity of the economy. On the other hand, the answer to the second question illustrates how unrestrained the error can be in the parameter selection even in a simple model like the one presented here. Observe, however, that in this setting, given that $\text{Cov}(y_1, \alpha) \geqslant 0$, the value of $\overline{\alpha}$ is always greater or equal than $\mathbb{E}[\alpha] = 0.5$ (see (7)).

To address the first question I compute M_1 (that according to (7) is equal to the optimal $\overline{\alpha}$ at $t = 1$) for the different values of $\sigma_1 \in [0.25, 1.25]$ and plot them in Figure 4. The figure shows that the optimal value of $\overline{\alpha}$ grows fairly linear with σ_1 in this range, which is illustrated by the closeness of both the blue and gray lines. Note that for $\sigma_1 = 1.25$ the optimal $\overline{\alpha}$ is approximately equal to 0.811 which is 42.3% bigger than the chosen value of 0.57.

Figure 4: Optimal $\overline{\alpha}$ at $t = 1$ as a function of σ_1

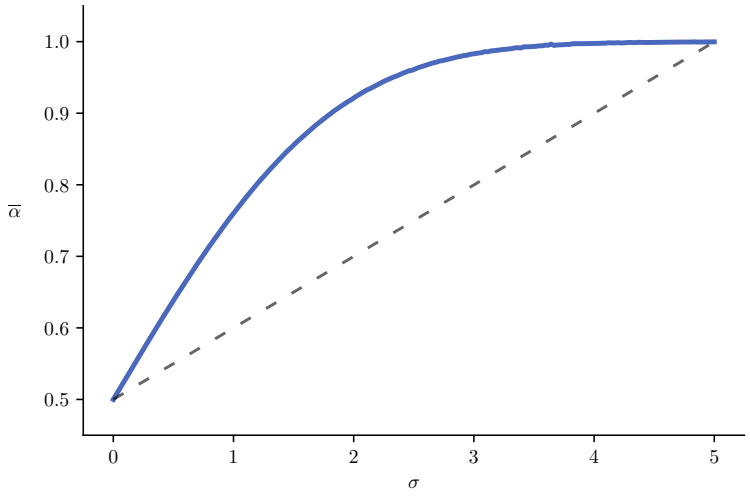


The blue line shows the optimal $\overline{\alpha}$ at $t = 1$ as a function of σ_1. The dashed gray line joins the two extremes of the curve. Note that for $\sigma_1 = 0$, the optimal value of $\overline{\alpha}$ is the same as the one computed at $t = 0$, approximately at 0.57 (see (10)).

Now, to find an answer to the second question, I conduct the following exercise. I take the same distributions as before, that is, $\alpha \sim U(0, 1)$, $y \sim \log\text{-Normal}(\mu, \sigma)$ with $\mu = 10$ and allow σ to move between 0 and 5. As before, the joint distribution of (y, α) satisfies the same condition over its support. Next, I compute the optimal $\overline{\alpha}$ using (7) for these parameters and plot them in Figure 5. As mentioned above, $\overline{\alpha} \geqslant 0.5$ for every σ in the considered range. Observe that the curve is highly concave, meaning that $\overline{\alpha}$ can be highly sensitive to the change of σ. In fact, the

growth in the optimal $\overline{\alpha}$ is very steep at the beginning, reaching 0.9 approximately at $\sigma = 1.83$.[15] Additionally, growth after $\sigma = 2.91$ is greatly reduced: The derivative at this point is 0.01 and since $\overline{\alpha} = 0.98$, then the semi-elasticiy is also close to 0.01.[16] The previous calculation implies that from 0.98 onwards the growth in $\overline{\alpha}$ is lower than 1%. It is also interesting to highlight the opposite curvature of the functions in Figures 3 and 5. According to (7) and (8), the error is a multiple of the difference between the optimal $\overline{\alpha}$ and as a result what drives the convexity of the curve in Figure 3 is the growth in the mean of y_1. As a final remark, note that, although this setting does not allow for $\overline{\alpha} < 0.5$, reversing the covariance between α and y produces a similar picture as that of Figure 5, but the curve is concave decreasing instead of increasing and with an asymptote at $\overline{\alpha} = 0$.

Figure 5: Optimal $\overline{\alpha}$ at $t = 1$ as a function of σ



The blue line shows the optimal $\overline{\alpha}$ as a function of σ, computed according to (7), with $\alpha \sim U(0,1)$ and $y \sim$ log-Normal(μ, σ), with $\mu = 10$. The dashed gray line joins the two extremes of the curve.

[15] For $\sigma = 1.809$, $\overline{\alpha} = 0.899$ and for $\sigma = 1.834$, $\overline{\alpha} = 0.902$.

[16] The semi-elasticity of $\overline{\alpha}$ in this context is $\dfrac{1}{\overline{\alpha}} \cdot \dfrac{\partial \overline{\alpha}}{\partial \sigma}$.

4. Consumer aggregation in dynamic models

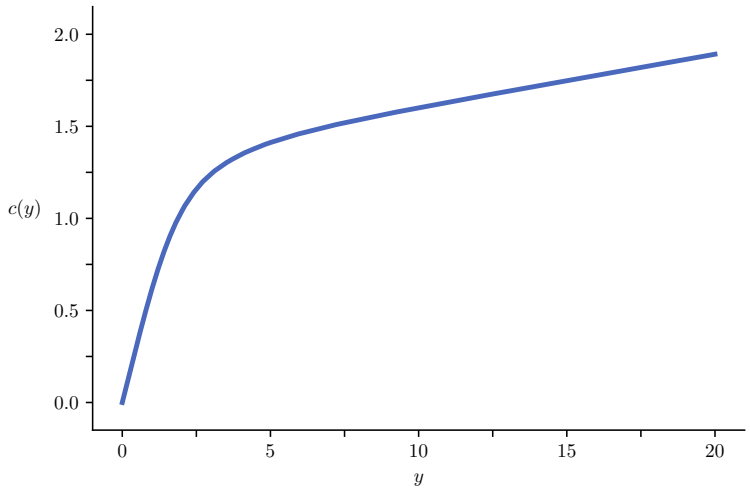
4.1. The model

Consider the following model where a consumer, identified with parameters (ρ, β) solves the following problem

$$\begin{aligned} \max_{\{c_t\}_t} \quad & \sum_{t=1}^{\infty} \beta^t \frac{c_t^{1-\rho}}{1-\rho}, \\ \text{s.t.} \quad & k_{t+1} = (1-\delta)(y_t - c_{t+1}), \\ & y_t = k_t + \theta_t k_t^{\alpha}, \end{aligned} \tag{11}$$

where C_t is consumption at time t, K_{t+1} is the capital at the start of period $t+1$ and Y_t represents the total resources available (income) at time t. Let $c(y)$ represent the infinite-horizon consumption as a function of current resources, where both variables are normalized by permanent labor income, that is, $y = \dfrac{Y}{wL}$ and similarly for c.[17] A salient feature of these models is that c is a highly concave function of y, but for low and high values of income the function is nearly linear, as in Figure 6. This behavior may harm predictions of the (aggregate) marginal propensity of consumption (MPC) using a representative agent model if the aggregate income is high but wealth is very unequally distributed. As expected, the difference between empirically estimated MPCs and the one predicted by the representative agent model is substantial, while the former are found between 0.2 and 0.5, the latter is only 0.04 under standard parameters.[18]

Figure 6: Consumption function



The blue line shows the (normalized) consumption as a function of the cash-on-hand y. The model that produces this type of function includes permanent and idiosyncratic shocks, as in Krusell and Smith (1998).

[17] If $c_t(y)$ is the optimal normalized consumption as a function of normalized current resources, then $c(y) := \lim_{n\to\infty} c_{T-n}(y)$, where T is the last period of the finite horizon counterpart of problem (11).

[18] See Carroll (2000).

One way to make better predictions is to modify the model to produce a skewed distribution of income, more similar to the data. Of course, if that same data hints how to choose the parameters of the representative agent model, then its use can produce more reasonable estimates without incorporating more complexity. Hereinafter I use the same argument as the one in Section 3 to understand how the basic representative agent model in (11) can be calibrated to obtain a better prediction for the MPC.

4.2. The problem of estimating the MPC

Suppose the economy is composed by agents identified by their preference parameters (β, ρ) and their income y, distributed according to F with marginals $F_{\beta,\rho}$ and F_y, respectively. According to the model in (11), for every pair (β, ρ) there is an optimal consumption $c_{\beta,\rho}$ and thus the (aggregate) MPC is

$$\operatorname{MPC}(F) = \int_S c'_{\beta,\rho}(y)\, dF,$$

where $S := \operatorname{supp} F$ is the support of the distribution F. Note that using the law of total expectation we can write

$$\operatorname{MPC}(F) = \mathbb{E}_y\Big[\mathbb{E}_{(\beta,\rho)\,|\,y}[c'_{\beta,\rho}(y)] \;\Big|\; y\Big]. \tag{12}$$

Note that if patience (β) and risk aversion (ρ) are known from income, then (12) gives a reasonable method to compute the MPC.

Suppose now that an investigator wishes to compute the MPC for this economy but instead of using a heterogeneous model (because data on preferences may not be available), she chooses a pair of parameters $(\overline{\beta}, \overline{\rho})$ to propose a representative consumer model. Solving (11) for these values gives a consumption function $c_{\overline{\beta},\overline{\rho}}$. Suppose this researcher has access to F_y, that is, she has data on income. As discussed in Section 3, the previous modeling decision is equivalent to approximating $F_{\beta,\rho}$ by a Dirac distribution $\delta_{\overline{\beta},\overline{\rho}}$ or, in more broader terms, approximate F by another distribution G with marginals $\delta_{\overline{\beta},\overline{\rho}}$ and F_y and with the same support. Hence, the MPC in this context is

$$\operatorname{MPC}(G) = \int_S c'_{\overline{\beta},\overline{\rho}}(y)\, dG = \mathbb{E}_y\Big[c'_{\overline{\beta},\overline{\rho}}(y)\Big]. \tag{13}$$

In consequence, using (12) and (13), the estimation error with sign is

$$D(\overline{\beta}, \overline{\rho}) := \mathbb{E}_y\Big[\mathbb{E}_{(\beta,\rho)\,|\,y}[c'_{\beta,\rho}(y) - c'_{\overline{\beta},\overline{\rho}}(y)] \;\Big|\; y\Big].$$

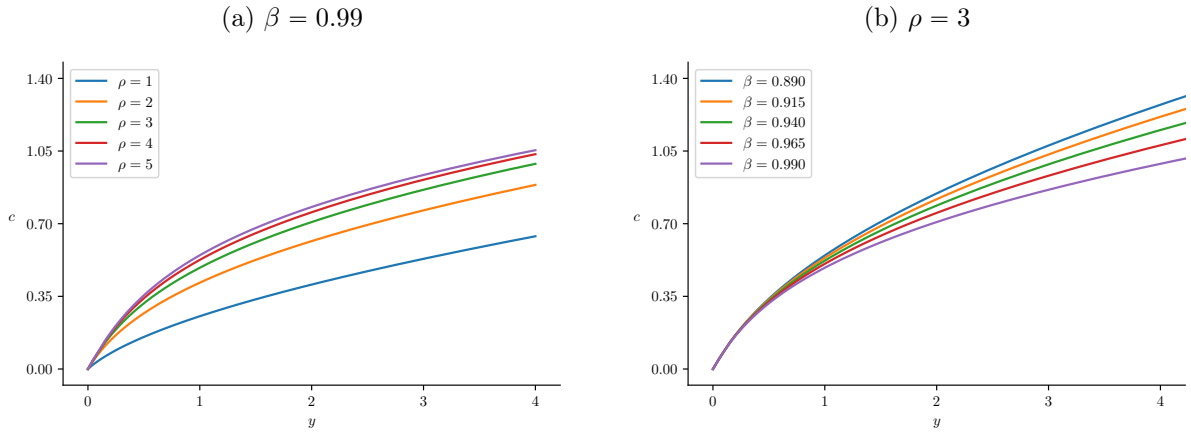
This means that, as shown in Section 3, the estimation error arises from differences in the "estimated" MPC for each agent. Of course, the aggregate model is not in principle capturing directly the individual MPC. Nevertheless when thought of it as approximating the heterogeneity in the economy, then it is possible to interpret that is indeed the case.

Sadly, for a model like (11) there is no analytic expression for $c_{\beta,\rho}$ and thus neither one for $c'_{\beta,\rho}$. Still, conditional on having a Taylor approximation to a given desired degree of precision, it is clear that these functions depend on a finite amount of moments of the joint distribution of β, ρ and y. Hence, $D(\overline{\beta}, \overline{\rho})$ also depends on a finite number of moments of F, allowing a tailoring in the approximation of the heterogeneity with the objective of reducing the estimation error.

4.3. Numerical example

In this section I show that an optimally chosen representative agent model can provide close estimates of the aggregate MPC without modeling the heterogeneity of the population. As mentioned above, every pair (β, ρ) in the economy is associated with a consumption function $c_{\beta,\rho}$ from which the MPC at the respective income value y is obtained. To assess how different the policy functions are depending on these parameters, observe Figure 7.

Figure 7: Consumption functions at different values of ρ and β



Each line is a different consumption function. In Panel (a), $\beta = 0.99$ for all values of ρ. Similarly, in Panel (b) for all different consumption functions, $\rho = 3$.

In Panel (a), where β is fixed at 0.99 and ρ varies from 1 to 5, the concavity of the function is greater as ρ grows. It is also interesting to note that the curves are almost parallel for high-income agents, meaning that for different values of ρ, the MPC is very similar if y is large enough. On the other hand, in Panel (b), ρ is fixed at 3 and β varies from 0.89 to 0.99. All functions are fairly similar but, again, the concavity increases as β gets closer to 0.99. Contrary

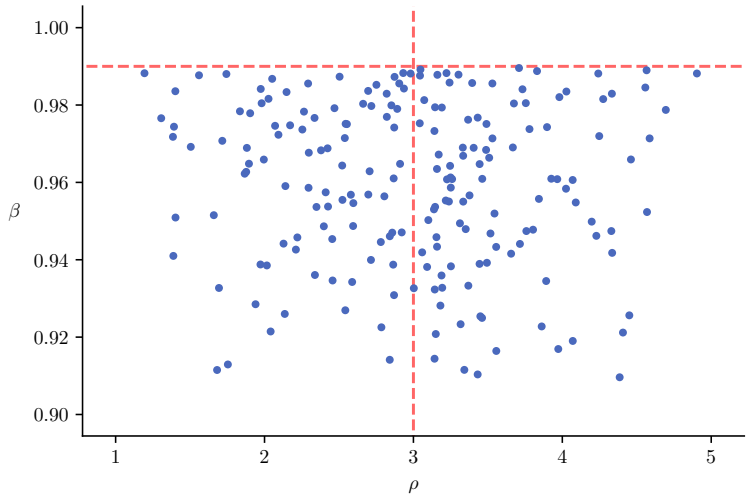
to what is shown in Panel (a), the curves have different slopes for large values of y, meaning that the MPC differs significantly for each β. As a final remark, note that the smooth transitions showed in this figure strongly suggest that the choice of the parameters of a representative agent model can balance the underlying heterogeneity of the agents' policy functions when estimating the aggregate MPC.

In order to make the last point explicit, consider the following setting. There are 200 agents in the economy with β, ρ and y independent random variables. The distribution F is obtained by first taking 200 independent draws from the following distributions:

- β: triangular distribution with lower limit 0.9, and with upper limit and mean equal to 0.99.
- ρ: triangular distribution with lower limit 1, upper limit 5 and mean 3.
- y: uniform distribution over the interval $[0, 10]$.

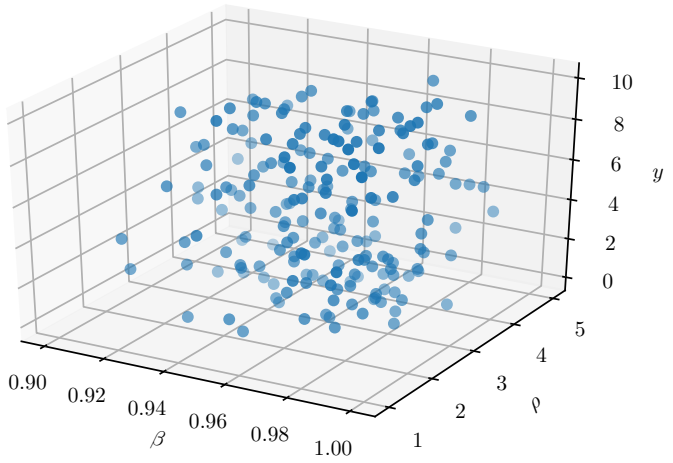
Then, a triple (β, ρ, y) is associated to each agent. The distribution of (β, ρ) in the economy is presented in Figure 8. As can be seen in the figure, the majority of agents are close to the intersection of the dashed lines, that correspond to the mean (and mode) of the underlying distributions. The joint distribution of the triples (β, ρ, y) is presented in Figure 9 to further illustrate the independency of the variables.

Figure 8: Distribution of β and ρ in the economy



The figure shows the marginal distribution $F_{\beta,\rho}$. Each dot represents an agent in the economy. The dashed lines correspond to $\rho = 3$ and $\beta = 0.99$, the means of the triangular distributions.

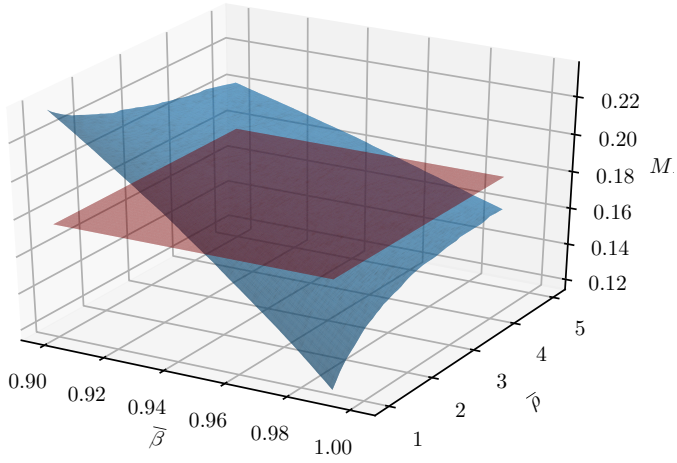
Figure 9: Joint distribution of β, ρ and y in the economy



The figure shows the distribution $F_{\beta,\rho,y}$. Each dot represents an agent in the economy.

According to the previous section, the joint distribution generated this way determines an aggregated MPC and the choice of $(\overline{\beta}, \overline{\rho})$ is aimed at keeping $D(\overline{\beta}, \overline{\rho})$ close to zero. Assuming the only known feature of the distribution of (β, ρ) is its boundaries – in this example, $\beta \in [0.9, 0.99]$ and $\rho \in [1, 5]$ – then I construct a grid with 10 000 points by taking 100 evenly spaced points from each interval and making all possible combinations. Next, for each pair $(\overline{\beta}, \overline{\rho})$ I compute the MPC associated to a representative agent model with these parameters. The results of this procedure are depicted in Figure 10.

Figure 10: Computed MPC for each pair $(\overline{\beta}, \overline{\rho})$

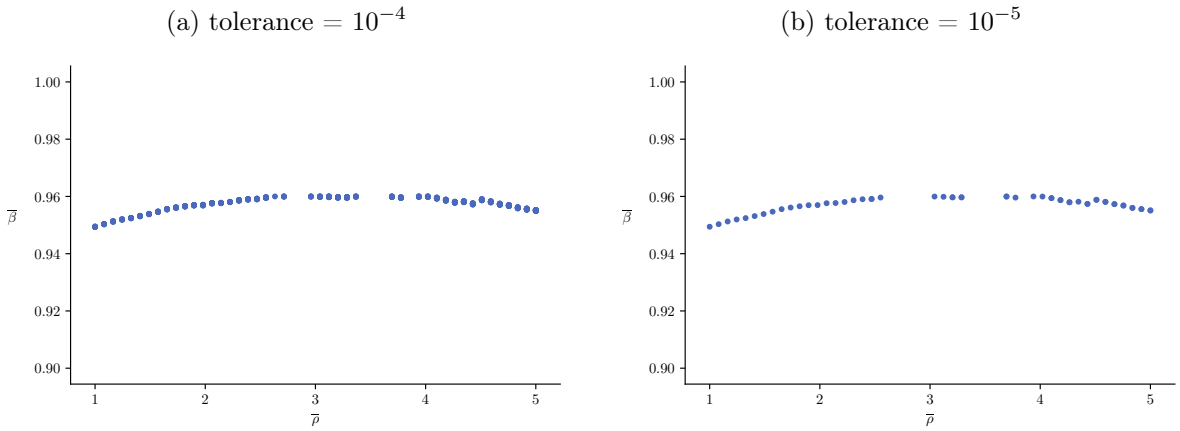


The figure shows the computed MPC for each pair $(\overline{\beta}, \overline{\rho})$ in a 10 000 point grid constructed in $[0.9, 1] \times [1, 5]$. The red plane is set at the real value of the MPC for this economy, approximately at 0.176.

From the figure it is clear that the blue and red planes intersect in more than one point. Moreover, the intersection is a curve. The figure also shows that the estimated MPC is more responsive to β than to ρ, which is clear from the higher slope that the blue surface has on the $\overline{\beta}$ axis relative to the $\overline{\rho}$ one. This last observation is in line with Carroll (2000), where the choice of two different β values is crucial to obtain a closer estimate of the MPC.

Finally, to give an idea of how the mentioned curve looks, in Figure 11 I graph the "zero" difference curves, that is, the points $(\overline{\beta}, \overline{\rho})$ for which $|D(\overline{\beta}, \overline{\rho})|$ is less than or equal than a certain tolerance. In Panel (a) the tolerance is set to 10^{-4} while in Panel (b) it is set to 10^{-5}. As seen in this figure, the curve is fairly smooth and again shows that the difference is more sensitive to the choice of $\overline{\beta}$ than that of $\overline{\rho}$. This can be seen in the fact that when the tolerance is 10^{-4}, for every value of $\overline{\rho}$ there is a $\overline{\beta}$ such that the difference is below the tolerance level but the converse is not true. This observation brings back the discussion over Figure 7. The reason why choosing $\overline{\beta}$ is more critical than choosing $\overline{\rho}$ can be derived following two steps. First, the MPC for different values of ρ is fairly similar for $y > 2$, as evidenced by the similarity of the slopes in Panel (a) of Figure 7. Second, the 80% of the uniform distribution over $[0, 10]$ is contained in the interval $[2, 10]$. Hence, the majority of the agents lie in a portion of the income distribution where the difference in ρ is not significant in the calculation of the MPC. On the contrary, since the MPC for different values of β may differ greatly, then the range of optimal $\overline{\beta}$ available can be greatly reduced if tolerance is small enough. In the same line, note that decreasing the tolerance by one order of magnitude strongly reduces the set of optimal parameters available.

Figure 11: "Zero" difference curves



The figure shows pairs $(\overline{\beta}, \overline{\rho})$ for which $|D(\overline{\beta}, \overline{\rho})|$ is less or equal than a certain tolerance.

5. Consumer aggregation and category goods

5.1. Description

Consider a two period economy composed by agents that consume $n+1$ goods: an essential good z (e.g., water) and n different goods grouped in the vector $\mathbf{x}$. Good z is valued at price $q \in \mathbb{R}_{++}$ and the $\mathbf{x}$ goods are valued at prices $\mathbf{p} \in \mathbb{R}^n_{++}$. In what follows and for simplicity of notation, the subscripts t for all variables are omitted unless needed. Each individual in this setting is identified with a pair (y, α), where $y \in (0, \infty)$ is her income and $\alpha \in (0, 1)$ determines the form of her pseudo-Cobb-Douglas (PCD) utility function

$$u_\alpha(\mathbf{x}, z) := u_{(y,\alpha)}(\mathbf{x}, z) = \sum_{j=1}^{n} x_j^\alpha z^{1-\alpha} = \left(\sum_{j=1}^{n} x_j^\alpha \right) z^{1-\alpha}.^{19}$$

Consumers choose how much of the $\mathbf{x}$ goods to consume in each period according to

$$\begin{aligned} \mathbf{x}(\mathbf{p}, q, y, \alpha) := \arg\max_{\mathbf{x}, z} \quad & u_\alpha(\mathbf{x}, z), \\ \text{s.t.} \quad & \mathbf{p}\mathbf{x} + qz = y. \end{aligned} \tag{14}$$

The pairs (y, α) follow a distribution F (with marginal distributions F_y and F_α) over their support $S \subset (0, \infty) \times (0, 1)$. In this setting, it is possible to aggregate consumption of the $\mathbf{x}$ goods into a single good X (see Section 2.2) given by

$$g_\alpha(\mathbf{x}) = \left(\sum_{j=1}^{n} x_j^\alpha \right)^{1/\alpha}. \tag{15}$$

In that case, we have

$$U(g_\alpha(\mathbf{x}), z) := g_\alpha(\mathbf{x})^\alpha z^{1-\alpha}, \tag{16}$$

which is the usual Cobb-Douglas utility function. Hence, by defining $\varepsilon := (1-\alpha)^{-1}$ and

$$P_\alpha(\mathbf{p}) := \left(\sum_{j=1}^{n} p_j^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}, \tag{17}$$

we have that the category demand

$$\begin{aligned} X(\mathbf{p}, q, y, \alpha) := \arg\max_{X, z} \quad & U(X, z), \\ \text{s.t.} \quad & P_\alpha(\mathbf{p})X + qz = y, \end{aligned} \tag{18}$$

[19] The choice of this particular utility function is based on the work of Helpman et al. (2008), where they use a similar function to study international trade. The price index in (17) is also obtained from this work.

is equal to

$$X(\mathbf{p}, q, y, \alpha) = y\, T_t(\alpha, \mathbf{p}) := y \frac{\alpha}{P_\alpha(\mathbf{p})}, \tag{19}$$

and satisfies

$$X(\mathbf{p}, q, y, \alpha) = g_\alpha\big(\mathbf{x}(\mathbf{p}, q, y, \alpha)\big). \tag{20}$$

This means that the disaggregate model is coherent with the aggregate one for each individual consumer. In other words, aggregation is an exact fit for each agent. Indeed, observe that the price and quantity indices are specific for each agent as they depend on α. This means that any attempt to describe this economy using category demands instead of the disaggregate consumption needs knowledge about the heterogeneity of the population in order to avoid approximation errors. Hence, although an exact representation for each consumer is available, doing the same for the economy as a whole requires knowledge about the different preferences, something that is possibly not available to the modeler.

5.2. The prediction problem

Consider now the following situation. An investigator at time $t = 0$ has data available on disaggregate consumption of the $\mathbf{x}$ and z goods, income (i.e., she knows F_{y_0}) and prices. Her objective is to estimate the aggregate category demand at $t = 1$ (e.g., the country demand for meat next year), $\mathbf{X}_1$. For simplicity she assumes that all agents have the same preferences, which means that all differences in consumption are accounted by differences in income. The last assumption also implies that she must choose a single parameter $\overline{\alpha}$ in order to obtain the category demand and the price index of each agent. Again, in the distribution-approximation sense, observe that the mentioned assumption has the same effect I discuss above: The aggregate model imposes an additional level of approximation by replacing F_α with $\delta_{\overline{\alpha}}$. However, note that in this case bundling goods into a category also represents an approximation but of a different nature. Both, F and the function $\mathbf{x}(\mathbf{p}, q, y, \alpha)$ induce a distribution G over the vector $\mathbf{x}$. Then, g_α induces a distribution over X that is in some sense an approximation of G, due to (20). Because of that, while the approximation is made directly over F, its implications reach G through g_α.

Under the previous assumption, the best estimation for the category demand of each agent at time t is

$$\widehat{X}_t(\mathbf{p}_t, y_t, \overline{\alpha}) = y_t\, T_t(\overline{\alpha}, \mathbf{p}_t) = \frac{\overline{\alpha}}{P_{\overline{\alpha}}(\mathbf{p}_t)}.$$

Thus, the best estimation for the aggregate category demand is

$$\widehat{\mathbf{X}}_t(\mathbf{p}_t, F_{y_t}, \overline{\alpha}) = \int_0^\infty y_t \, T_t(\overline{\alpha}, \mathbf{p}_t) \, dF_{y_t} = \mathbb{E}[y_t] \, T_t(\overline{\alpha}, \mathbf{p}_t). \tag{21}$$

5.3. A complex estimation error

To understand how the estimation errors appear in this setting I proceed in two steps. First, since category demands depend on α_t, then for agent (y_t, α_t) the true value differs from the estimated one by

$$\mathcal{D}_t(y_t, \alpha_t, \overline{\alpha}) := X(\mathbf{p}_t, y_t, \alpha_t) - \widehat{X}_t(\mathbf{p}_t, y_t, \overline{\alpha}) = y_t \Big(T_t(\alpha_t, \mathbf{p}_t) - T_t(\overline{\alpha}, \mathbf{p}_t) \Big).^{20} \tag{22}$$

It is not obvious that this difference is monotone in $\overline{\alpha}$ for a given pair (y_t, α_t). In fact, it is not. To see this I take $n = 2$, $\mathbf{p}_t = (1, 10)$ and present differences $\mathcal{D}_t$ relative to income (i.e., as a percentage) as a function of $\overline{\alpha}$ for different values of α_t in Figure 12.[21] For illustrative purposes I restrict $\overline{\alpha}$ to the interval $[0.15, 0.85]$.

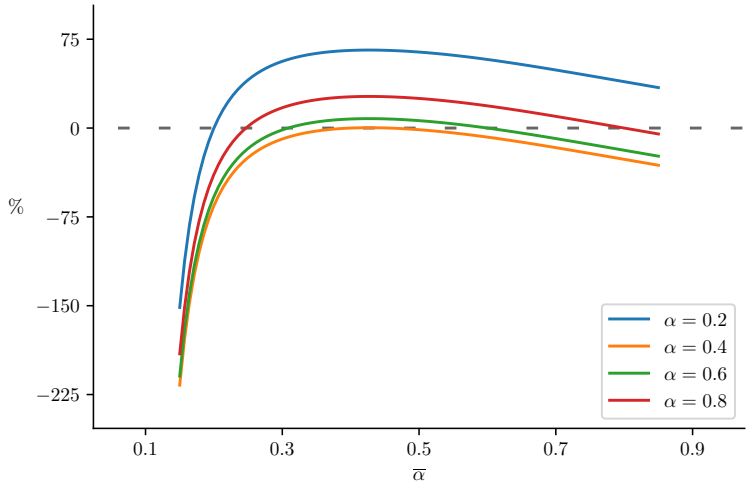
Each curve exhibits a strong non-monotonic behavior and is highly concave, increasing its value quite abruptly when $\overline{\alpha}$ is low and decreasing slowly when $\overline{\alpha}$ is high. Related to the former phenomenon and as showed in the figure, almost all curves cross the horizontal line twice: once when $\overline{\alpha} = \alpha$ in the increasing part of the curve and a second time in a greater value when the relative error is decreasing. The previous characteristic is important when trying to reduce the error to zero because it gives, at first glance, a broader range of optimal $\overline{\alpha}$ to choose. However, it may be the case that the error tolerance is small enough to restrict the attention to higher values of $\overline{\alpha}$ where the relative change in $\mathcal{D}_t$ is smaller. The last comment also highlights a curious feature of this setting: An investigator whose objective is reducing the error to zero might be interested in selecting a high value of $\overline{\alpha}$, away from the real values of preferences, in order to protect herself from small changes that could be more harmful when $\overline{\alpha}$ is small. A final insight can be extracted from the transition between the curves. Note that the blue line ($\alpha = 0.2$) moves downward when α grows to 0.4 but all the remaining curves move upward. This behavior relates closely to the one of T_t, as shown in Figure 13, where this variable displays the inverse behavior, decreasing rapidly for small values of α and then growing at a slower rate.

The second step of the analysis consists in studying the aggregate error. To this end, note that

[20] Note that in this setting X does not depend on q and consequently, to simplify notation, this variable is suppressed.

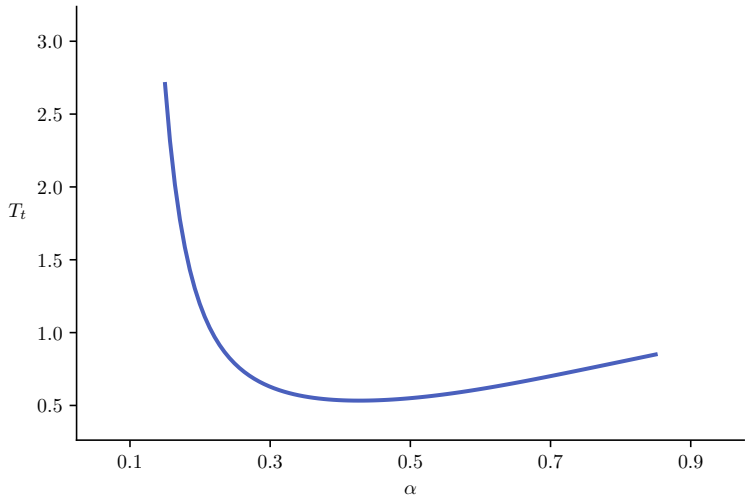
[21] Choosing this price vector has only clarifying purposes. All of the features mentioned in the rest of the paragraph are maintained if $\mathbf{p}_t$ is changed.

Figure 12: D_t relative to income as a function of $\overline{\alpha}$



The figure shows $\mathcal{D}_t$ relative to income (i.e., $\mathcal{D}_t/y_t \times 100$) as a function of $\overline{\alpha}$ for the five different α_t showed in the legend. The price vector used is $\mathbf{p}_t = (1, 10)$. The dashed gray line corresponds to $\mathcal{D}_t = 0$.

Figure 13: T_t as a function of α



The figure shows T_t as a function of α. The price vector used is $\mathbf{p}_t = (1, 10)$.

the estimation error (with sign) of $\mathbf{X}_{1t}$ is just the integral of $\mathcal{D}_t$ over S. To simplify notation let $T_t = T_t(\alpha, \mathbf{p}_t)$ and $\overline{T}_t = T_t(\overline{\alpha}, \mathbf{p}_t)$. In the simplest case where preferences do not change over time we have

$$\begin{aligned} D_t\left(\overline{T}_t\right) &:= \int_S \mathcal{D}_t(y_t, \alpha, \overline{\alpha}) \\ &= \mathbb{E}\big[y_t\, T_t\big] - \mathbb{E}[y_t]\,\overline{T}_t \\ &= \mathrm{Cov}\left(y_t, T_t\right) + \mathbb{E}[y_t]\mathbb{E}\left[T_t\right] - \mathbb{E}[y_t]\,\overline{T}_t \\ &= \mathbb{E}[y_t]\left\{\frac{\mathrm{Cov}\left(y_t, T_t\right)}{\mathbb{E}[y_t]} + \mathbb{E}\left[T_t\right] - \overline{T}_t\right\}. \end{aligned} \tag{23}$$

which is in principle very similar to (6) in Section 3.3. However, note now that D_t depends on prices and hence on their dynamics. In terms of distributions observe that in this case, the mixed setting implies that the distribution that matters is the one of (y_t, T_t) which includes an additional source of heterogeneity: $\mathbf{p}_t$. From (23) we know that the sufficient statistics needed to obtain an exact fit at time t are, just as in Section 3.3, the first and second moments of that distribution. The difference in this example is that the new source of variation, prices, may affect heavily the precision with which these statistics are estimated.

It is clear from (23) that an investigator interested in making the estimation error the smallest possible should choose $\overline{\alpha}$ such that

$$\overline{T}_t = \tilde{M}_t := \underbrace{\mathbb{E}\left[T_t\right] + \frac{\mathrm{Cov}\left(y_t, T_t\right)}{\mathbb{E}[y_t]}}_{:=\tilde{M}_t}. \tag{24}$$

Observe that changes in prices may affect the right hand side of (24). If the investigator only has access to $t = 0$ variables, knowing the price dynamics may not be enough to reduce the estimation error at $t = 1$ to zero. Indeed, note that the value of each term in $\tilde{M}_0$ is unknown at $t = 0$, thus any correction using knowledge about price changes is not feasible. It is interesting to note, however, that if relative prices do not change between $t = 0$ and $t = 1$, that is, if $\mathbf{p}_1 = \lambda \mathbf{p}_0$, then using that $P_\alpha(\mathbf{p})$ is homogeneous of degree 1 we have $T_1 = \lambda^{-1} T_0$,

$$\lambda^{-1}\tilde{M}_0 = \mathbb{E}[T_1] + \frac{\mathrm{Cov}\left(y_0, T_1\right)}{\mathbb{E}[y_0]},$$

and

$$\overline{T}_1 = \lambda^{-1}\overline{T}_0. \tag{25}$$

Hence, in this scenario, setting $\overline{T}_0 = \tilde{M}_0$ as in (24) (assuming that reducing the estimation error to zero at $t = 0$ is achievable) and then making the correction in (25) further reduces the

estimation error if

$$\begin{aligned}\left|\tilde{M}_1 - \lambda^{-1}\tilde{M}_0\right| &\leqslant \left|\tilde{M}_1 - \tilde{M}_0\right| \\ \left|\frac{\operatorname{Cov}(y_1, T_1)}{\mathbb{E}[y_1]} - \frac{\operatorname{Cov}(y_0, T_1)}{\mathbb{E}[y_0]}\right| &\leqslant \left|\mathbb{E}[T_1] - \mathbb{E}[T_0] + \frac{\operatorname{Cov}(y_1, T_1)}{\mathbb{E}[y_1]} - \frac{\operatorname{Cov}(y_0, T_0)}{\mathbb{E}[y_0]}\right| \\ \left|\frac{\operatorname{Cov}(y_1, T_1)}{\mathbb{E}[y_1]} - \frac{\operatorname{Cov}(y_0, T_1)}{\mathbb{E}[y_0]}\right| &\leqslant \left|(1-\lambda)\mathbb{E}[T_1] + \frac{\operatorname{Cov}(y_1, T_1)}{\mathbb{E}[y_1]} - \frac{\operatorname{Cov}(y_0, T_0)}{\mathbb{E}[y_0]}\right|.\end{aligned}$$

Moreover, if $y_1 = Ry_0$ for some $R > 0$, then implementing this correction eliminates the error, just as what happened in Section 3.3. Finally, note that if prices do not change, then $T_0 = T_1$ and (23) has the same form as (6). This phenomenon is in line with the previous discussion: By suppressing the variation or heterogeneity in prices over time we return to the results of the model in Section 3.

5.4. Numerical example

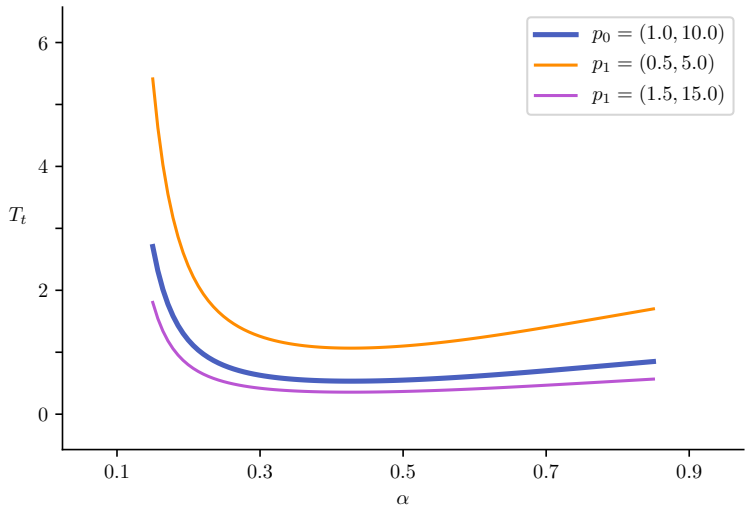
In this section I construct an example in order to understand this particular setting and study how the optimal aggregation parameters and the approximation accuracy change between one period and the next. For this purpose, I take $n = 2$ and $\mathbf{p}_0 = (1, 10)$, I assume preferences do not change (i.e., $\alpha_0 = \alpha_1 = \alpha$) and consider the same constructions for the joint distributions F_0 and F_1 as presented in Section 3.4, except for the fact that, for clarifying purposes, α is restricted to the interval $[0.15, 0.85]$.[22] In this case I choose $\mu = \ln 5.5$, $\sigma_0 = 0.25$ and $\sigma_1 \in [0.25, 1.25]$. This value of μ ensures that the median consumer can buy half of each of the $\mathbf{x}$ goods. In order to continue the discussion at the end of the preceding section I consider $\mathbf{p}_1 = \lambda\mathbf{p}_0$ and two values for λ, $\lambda_1 = 0.5$, $\lambda_2 = 1.5$. This way, with λ_1 the $\mathbf{x}$ goods reduce their cost (relative to z) and viceversa with λ_2. The difference between T_0 and both T_1 is presented in Figure 14. As can be seen in the image, lower prices rise the value of T_1 in contrast to T_0, while greater prices have the opposite effect, as expected from the functional form of T_t. Since prices appear in the denominator, the increase of T_t after a price fall is greater than its drop.[23]

In the foregoing section I mentioned that for this setting the important distribution is that of (y_t, T_t). Consequently, the next step is to investigate its support in a similar fashion as in Section 3.4. In Figure 15 I present the support of the three joint distributions of interest: the one of (y_t, α_t) in Figure 15a, (y_t, T_t) with $\lambda_1 = 0.5$ in Figure 15b and the one of (y_t, T_t) using

[22] This is, y_0 and y_1 follow *log*-Normal distributions with parameters (μ, σ_0) and (μ, σ_1), respectively, $\alpha \sim U(0.15, 0.85)$ and the joint distribution is such that individuals with higher α have higher income.

[23] Remember that in this scheme $T_1 = \lambda^{-1}T_0$ and hence, for a given value of T_0, the function $T_1(\lambda)$ behaves like x^{-1}.

Figure 14: T_0 and T_1 as a function of α



The figure shows T_0 (thick blue line) and T_1 (yellow and purple lines) as a function of α. T_0 is computed using $\mathbf{p}_0 = (1, 10)$ and each T_1 is computed with $\mathbf{p}_1 = \lambda_i \mathbf{p}_0$, where $\lambda_1 = 0.5$ and $\lambda_2 = 1.5$.

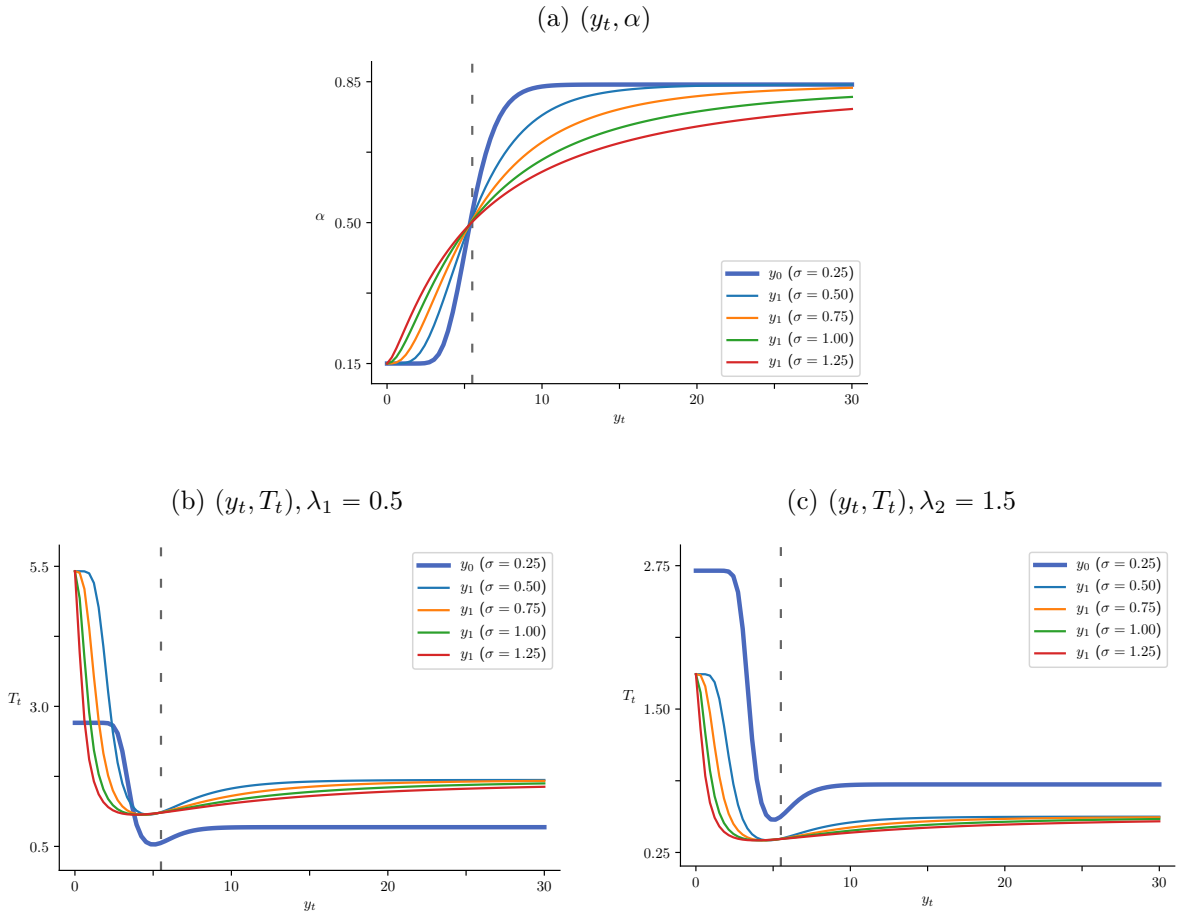
$\lambda_2 = 1.5$ in Figure 15c. It is possible to see that while y_t and α have positive covariance, for y_t and T_t it is negative and in both cases the relation is not monotonic. Additionally, in Figure 15b we see that this covariance is increased in absolute value between $t = 0$ and $t = 1$, as evidenced by the steeper line that joins the extremes of the $t = 1$ curves, which contrasts with the wider line of the $t = 0$ variables. The opposite effect is observed in Figure 15c. It is also worth noting that to the right side of the median, the relation (y_t, T_t) is almost constant in both Panels (b) and (c), while the highest variability happens to income values below the median. Hence, by looking back at (22), it is clear that choosing $\overline{\alpha}$ closer to the one shared by the high-income fraction of the population may reduce the estimation error further than focusing on the complete set of agents. Remembering that using the representative agent means changing the support to a straight horizontal line in each of the panels in Figure 15, then the previous comment means that in this particular setting, a two-agent model can be a much better approximation, while still maintaining the simplicity. In graphical terms, a two-agent model appears as a two-valued function in each of the plots in Figure 15.

For this example, the choice of optimal parameters at $t = 0$ is summarized in Table 1. The value of $\overline{\alpha}$ is chosen optimally to minimize the (absolute value of the) error at $t = 0$ and then $\overline{T}_0$ is computed using the definition. Finally, $\overline{T}_1$ is obtained following the aforementioned correction $\overline{T}_1 = \lambda^{-1}\overline{T}_0$. The first two rows of the column are equal precisely because the optimal parameters are obtained at $t = 0$, when no price change has yet occurred. The last row is in line with the discussion above. The optimal $\overline{T}_1$ in both cases is closer to the more constant part given by the rich half of the population. The last statement finds its explanation directly

from (22): Since income is higher for this half, is more important to diminish the difference for these agents than for the other half in order to reduce the error to zero.

The next step is to determine how big the estimation error is at $t = 1$. Figure 16 presents the difference $D_1(\overline{T}_1)$ as a function of σ_1 (see (23)), the parameter of redistribution. As we can see from the figure, the approximation error using this corrected parameter is near zero and in both cases the aggregate demand is overestimated. Although meaningless because of the order of magnitude, it is interesting to note that when prices drop, the discrepancy is bigger than when they increase. In other words, when prices fall from $t = 0$ to $t = 1$, the overestimation of the aggregate demand is more pronounced. This result should come with no surprise after analyzing Figure 15. In Panels (b) and (c) we see that the drop in T_t after a price increase is smaller than its rise after a price fall. Hence, the change in X_1 in the first case is smaller than in the second (see (19)).

Figure 15: Support of joint distributions



The figure shows the support of the joint distributions of (y_t, α), (y_t, T_t) with $\lambda_1 = 0.5$ and (y_t, T_t) with $\lambda_2 = 1.5$ for five different values of σ_1. The dashed gray lines correspond to $y_t = 5.5$, the median of the distribution. The price vectors used are $\mathbf{p}_0 = (1, 10)$ and $\mathbf{p}_1 = \lambda_i \mathbf{p}_0$, with $i = 1, 2$.

Table 1: Optimal parameters at $t = 0$

	λ	
	0.5	1.5
$\overline{\alpha}$	0.75	0.75
$\overline{T}_0$	0.75	0.75
$\overline{T}_1$	1.5	0.5

The table shows the optimal parameters $\overline{\alpha}$, $\overline{T}_0$ and $\overline{T}_1 = \lambda^{-1}\overline{T}_0$ for the two values of λ. The price vectors used are $\mathbf{p}_0 = (1, 10)$ and $\mathbf{p}_1 = \lambda_i \mathbf{p}_0$, with $i = 1, 2$.

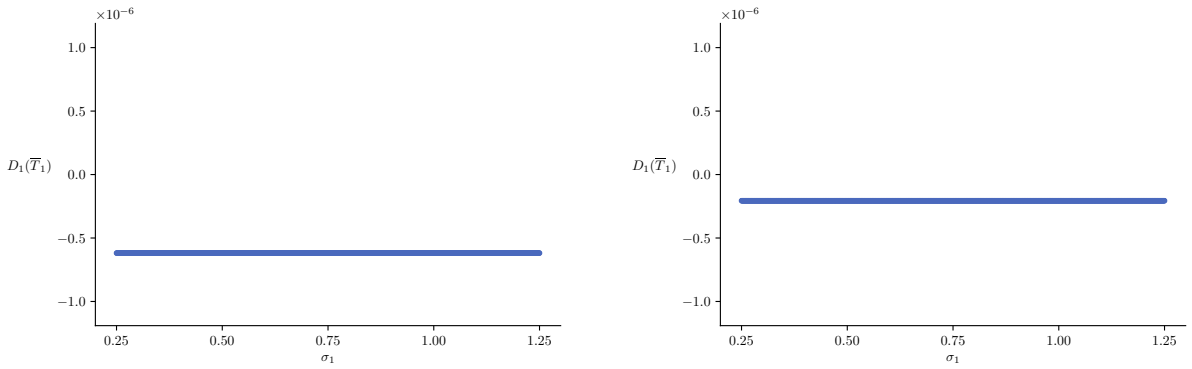
From the discussion at the end of Section 5.3, the correction in (25) is useful only if

$$\left|\tilde{M}_1 - \lambda^{-1}\tilde{M}_0\right| \leqslant \left|\tilde{M}_1 - \tilde{M}_0\right|.$$

Using this numerical example it is possible to determine if the change from $\overline{T}_0$ to $\overline{T}_1$ is capable of lowering the (absolute value of the) error by directly computing $D_1(\overline{T}_0)$. In Figure 17 I show how this difference changes with σ_1. Comparing this figure with Figure 16 we see that making the correction lowers the error by several orders of magnitude: While changing the parameter obtained at $t = 0$ using (25) gives an error of magnitude 10^{-7}, keeping $\overline{T}_0$ returns an estimate that differs with the real value by 10^{-1}, six orders of magnitude greater than in the previous case. Nevertheless, it must be borne in mind that in order to use $\overline{T}_0$ it is not necessary to know the value of λ, which may not be available at $t = 0$. If that is the case, the difference found may be small enough to choose to maintain the uncorrected parameter.

Figure 16: Estimation error $D_1(\overline{T}_1)$ as a function of σ_1

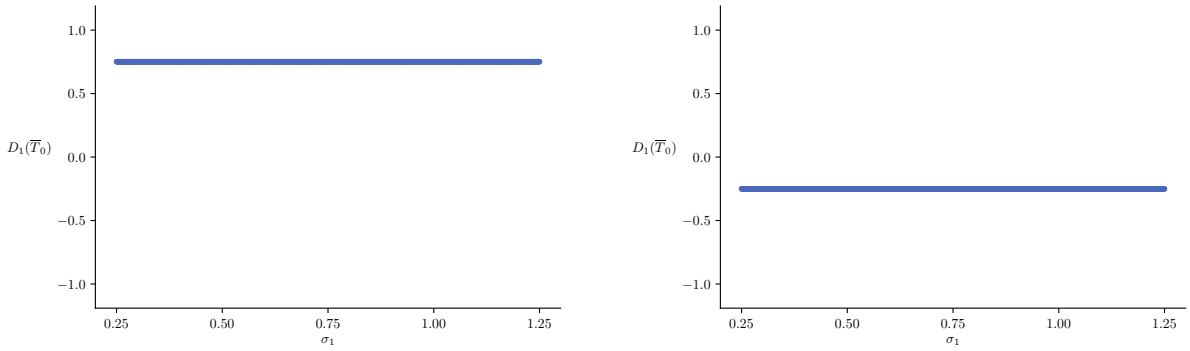
(a) $\lambda = 0.5$ (b) $\lambda = 1.5$



The blue line shows $D_1(\overline{T}_1)$ as a function of σ_1, computed using (23). $\overline{T}_1$ is computed using the correction in (25). The price vectors used are $\mathbf{p}_0 = (1, 10)$ and $\mathbf{p}_1 = \lambda_i \mathbf{p}_0$, with $i = 1, 2$.

Figure 17: Estimation error $D_1(\overline{T}_0)$ as a function of σ_1

(a) $\lambda = 0.5$ (b) $\lambda = 1.5$

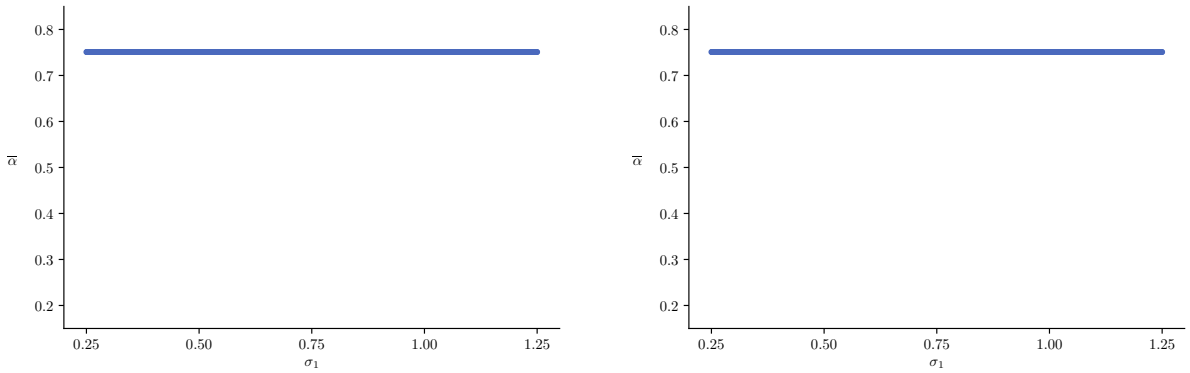


The blue line shows $D_1(\overline{T}_1)$ as a function of σ_1, computed using (23). The price vectors used are $\mathbf{p}_0 = (1, 10)$ and $\mathbf{p}_1 = \lambda_i \mathbf{p}_0$, with $i = 1, 2$.

To close this section I show how the optimal $\overline{\alpha}$ should evolve depending on the value of σ_1. The results shown in Figures 16 and 17 suggest that $\overline{\alpha}$ should be constant over time, since the error commited in both cases is small, especially in the first case. In Figure 18 I show that this is exactly the case: Both when prices drop and when they increase, the optimal $\overline{\alpha}$ is the same regardless the value of σ_1. The explanation for this phenomenon is, again, obtained directly from (24). In this example, $\text{Cov}(y_t, T_t)$ is extremely small, with values around 10^{-15}. This contrasts with the same statistic for (y_t, α_t) obtained in Section 3, where $\text{Cov}(y_t, \alpha_t)$ is large and increased monotonically with σ_1. The previous explanation implies that $\overline{T}_t$ is very close to $\mathbb{E}[T_t]$ for every σ_1 and for that reason the optimal parameter does not change noticeably.

Figure 18: Optimal $\overline{\alpha}$ at $t = 1$ as a function of σ_1

(a) $\lambda = 0.5$ (b) $\lambda = 1.5$



The blue line shows the optimal $\overline{\alpha}$ at $t = 1$ as a function of σ_1. Note that for $\sigma_1 = 0$, the optimal value of $\overline{\alpha}$ is the same as the one computed at $t = 0$: 0.75 (see Table 1). The price vectors used are $\mathbf{p}_0 = (1, 10)$ and $\mathbf{p}_1 = \lambda_i \mathbf{p}_0$, with $i = 1, 2$.

6. Concluding remarks

The preceding exercises and problems depict three situations where the inherent diversity of a model is replaced with a completely homogeneous alternative. In all three cases, aggregation is used as a tool to overcome the complexity imposed by differences between agents. This work diverges from previous attempts to study aggregate models not on the objective but on the means to attain it. Both approaches share a quest to obtain requisites that ensure coherence between homogeneous and heterogeneous models. However, while the existing literature seek to impose these requirements on the disaggregate model, I claim instead that a better approach is to introduce them on the aggregate one. The advantage of this alternative is that it suppresses all restrictions about what kind of model can be subject to aggregation. As a consequence, all past models that ignored the conditions to have exact aggregation can be now evaluated in terms of the quality of the fit, instead of the lack of the right context to apply it. In particular, this way of thinking the problem has implications on a variety of models where intrinsic differences are part of their essence.

In broader terms, the main idea I present in this paper is that aggregation can be regarded as a tool or solution used in the problem of heterogeneity approximation. Following the insights and lines of argumentation of the foregoing sections, I can reformulate this modeling problem as follows. Diversity is, in formal terms, the presence of a non-trivial distribution F over a given set of parameters in the economy. Thus, the question of finding a suitable substitute to it consists of two parts: 1) determine which aspects of F are relevant to the specific objective of the model, and 2) establish an appropriate distance between distributions that allows the researcher to select the best $\hat{F}$ to replace F. This way, the original problem, which may be ambiguous in its conception, is replaced by one that is both transparent and susceptible to be tackled directly using mathematical tools.

The first part of the question is equivalent to determining the most significant sources of heterogeneity and how to model them. For that reason, the result of this search is inevitably tied to the particular problem the researcher is trying to confront. However, this does not necessarily imply that obtaining the answer is difficult. On the contrary, this phase of the modeling problem was already present and, moreover, is allegedly easier since heterogeneous features within a model are generally the fundamental part of the question. Nevertheless, it is clear that when finding the answer, some decision about the extent of the disaggregate nature of the model has to be made. Indeed, note that the dual of the question asks which aspects of F are not relevant and, consequently, not included in the approximation problem. An additional

advantage is that reviewing the most important aspects of the problem at hand may give some hints regarding the next step: choosing the distance between distributions. By narrowing the sources of heterogeneity it is conceivable that an appropriate measure of error emerges naturally.

The second stage of the question, adopting a suitable distance, is probably the hardest of the two. In optimization terms, the problem requires to carefully define the loss function to use. In the examples developed throughout this text the metrics used are simple differences between the real and estimated values of the variable of interest. This later implied that the error can be written in terms of a certain number moments of both the original and approximate distributions. Arguably, other types of loss functions can result in better alternative models if their choice is capable of correcting deficiencies that may arise when using the ones I introduced throughout the text. However, in the case of a researcher interested in a given variable (or set of variables) X, the simple-difference choice may be backed by the fact that for a certain tolerance in the goodness-of-fit, a Taylor approximation of X around a previously know point (e.g., the known value at the moment of the modeling process) can be made. This polynomial provides the relevant moments of the variables that are needed to have a suitable fit. Therefore, a possibility is using another distribution $\hat{F}$ that shares those moments with F. Another option is using a simpler $\hat{F}$ that is calibrated using this same approximation, like the examples in the paper.

The above discussion makes it clear that taking into account the full diversity of the model and using a completely aggregate one are two extremes of a wide spectrum of possibilities. Depending on the particular problem to confront, the most desirable approximation likely lies between the two. For the models I present in this work, the aggregate extreme is always a feasible solution, but it is not necessarily optimal under every criteria. Moreover, this analysis only takes care of problems where the interest is predicting an aggregate variable. This is the crucial feature that implies that the estimation differences depend on some unknown moments of the original distribution. Specific examples remain to be studied where the interest is of a different kind (e.g., estimation of treatment effects or robustness of the conclusions of a model). As mentioned before, it is plausible that the main challenge in any other modeling problem is the choice of the appropriate measure of error or, in other words, the distance between the chosen distributions. All in all, for researchers dealing with problems in which diversity of any kind makes an appearance, this framework provides them the necessary tools to make informed modeling decisions. As a result, they can base their assumptions exclusively on the question at hand and the implications of its answer, instead of on stringent conditions that most of the time are not met.

Acknowledgements

The paper I developed in the preceding pages is the product of a process longer than the four months it took to write. Within each sentence my words bear a hidden meaning, a testament of the experiences I have been through. They speak about education, at home, at school, at college. They represent dedication, responsibility, tenacity. But mostly they are a reflection of support. This work would have not be the same without the encouragement and company of many people I met from the earliest to this latest part of my life. To everyone of you, I dedicate this work.

To my parents. I start with you because that is how it all began. A great part of the success of this (long) journey has been without doubt a result of countless efforts from both. Thanks for your unconditional support, for the values and virtues you taught and showed me since my birth. Thank you for proving me that hard-work, patience and discipline are key to achieve any objective. Thank you for giving me the privilege of a good school and a top-notch education in general. Above all, I am grateful for your unquestioning love, your company and the luck to have you as parents. I wish this paper makes you proud.

To my grandmother, Miss Glory. It is impossible to imagine this moment without you. You have been an extraordinary company in every stage of the way and for that I am grateful. As a kid I was kind of your young side-kick but I think we are more like partners now. I sincerely hope you feel proud of me and that you think of this achievement as part of you too. With all my love, this thesis is for you.

To my friends (in alphabetical order), Ceci, Coni, David, Iván, Nacho, Nico, Pancho, Pato, Raúl, Rubén. Maybe you are not aware but your presence in my life was key to reach this moment. From early in school to these last days of college your company and encouragement have shaped me into the person I am today. Thank you for that. It may sound trivial (pun intended) but having moments of leisure with you playing video games, having a burger together at a birthday or just meeting to chat about our lives will be without doubt one of the most valuable memories I will cherish forever. Again and always, thank you.

To my guides, Felipe and Eugenio. I can not imagine how tiresome and difficult it must be to help a student to reach high quality academic work in such a short amount of time. You showed dedication up to the last minute in spite of the unexpected conditions at the end of the semester. Not only that, you proved patient when we were lost and out of focus. Thank you for all your work in this four months.

Last but not least, to Nine. *Thank you* is not the word I'm looking for, there is so much more inside me now. It is evident that without you this thesis would have not been the document I feel proud of. Your comments were critical to reach this point. But your support does not even begin there. You enlightened the way when I thought I was writing nonsense. You encouraged me when I ran out of ideas. You were there just to hear me ranting when I needed. I love you for that, for your kindness, your intelligence and your dedication towards us. I feel the luckiest person in the world for finding the most wonderful teammate. I sincerely hope to be the partner you need me to be and to repay at least a drop of all the love you give me every day. G r a t i t u d e.

References

S. Basu and J. G. Fernald. Returns to scale in u.s. production: Estimates and implications. *Journal of Political Economy*, 105(2):249–283, 1997. ISSN 00223808, 1537534X. 2

D. Berger and J. Vavra. Consumption dynamics during recessions. *Econometrica*, 83(1):101–154, 2015. doi: 10.3982/ECTA11254. URL https://doi.org/10.3982/ECTA11254. 6

C. D. Carroll. The buffer-stock theory of saving: Some macroeconomic evidence. *Brookings Papers on Economic Activity*, 23(2):61–156, 1992. URL https://ideas.repec.org/a/bin/bpeajo/v23y1992i1992-2p61-156.html. 6

C. D. Carroll. Requiem for the representative consumer? aggregate implications of microeconomic consumption behavior. *American Economic Review*, 90(2):110–115, May 2000. doi: 10.1257/aer.90.2.110. 3, 18, 23

A. Fagereng, L. Guiso, and L. Pistaferri. Firm-related risk and precautionary saving response. *American Economic Review*, 107(5):393–97, May 2017. doi: 10.1257/aer.p20171093. URL http://www.aeaweb.org/articles?id=10.1257/aer.p20171093. 6

R. C. Feenstra and G. H. Hanson. Aggregation bias in the factor content of trade: Evidence from u.s. manufacturing. *American Economic Review*, 90(2):155–160, May 2000. doi: 10.1257/aer.90.2.155. 3

M. Friedman. The methodology of positive economics. In *Essays in Positive Economics*, A Phoenix book. Business economics, pages 3–43. University of Chicago Press, 1953. ISBN 9780226264035. 2

W. M. Gorman. Community preference fields. *Econometrica*, 21(1):63–80, 1953. doi: 10.2307/1906943. 3, 6

W. M. Gorman. Separable utility and aggregation. *Econometrica*, 27(3):469–481, 1959. doi: 10.2307/1909472. 4

P. Gourinchas and J. A. Parker. Consumption over the life cycle. *Econometrica*, 70(1):47–89, 2002. doi: 10.1111/1468-0262.00269. URL https://doi.org/10.1111/1468-0262.00269. 6

F. Gul and W. Pesendorfer. Self-control and the theory of consumption. *Econometrica*, 72(1):119–158, 2004. doi: 10.1111/j.1468-0262.2004.00480.x. URL https://doi.org/10.1111/j.1468-0262.2004.00480.x. 6

E. A. Hanushek, S. G. Rivkin, and L. L. Taylor. Aggregation and the estimated effects of school resources. *The Review of Economics and Statistics*, 78(4):611–627, 1996. ISSN 00346535, 15309142. 3

E. Helpman, M. Melitz, and Y. Rubinstein. Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2):441–487, 2008. URL https://EconPapers.repec.org/RePEc:oup:qjecon:v:123:y:2008:i:2:p:441-487. 24

J. R. Hicks. *Value and Capital: An Inquiry Into Some Fundamental Principles of Economic Theory*. Clarendon paperbacks. Clarendon Press, 1946. ISBN 9780198282693. 4, 9

J. Imbs, H. Mumtaz, M. O. Ravn, and H. Rey. PPP Strikes Back: Aggregation And the Real Exchange Rate*. *The Quarterly Journal of Economics*, 120(1):1–43, 02 2005. ISSN 0033-5533. doi: 10.1162/0033553053327524. 3

A. Kajii and S. Morris. The robustness of equilibria to incomplete information. *Econometrica*, 65(6):1283–1309, 1997. doi: 10.2307/2171737. 2

K. Kan, S.-K. Peng, and P. Wang. Understanding consumption behavior: Evidence from consumers' reaction to shopping vouchers. *American Economic Journal: Economic Policy*, 9(1):137–53, February 2017. doi: 10.1257/pol.20130426. URL http://www.aeaweb.org/articles?id=10.1257/pol.20130426. 6

G. Kaplan and G. L. Violante. A model of the consumption response to fiscal stimulus payments. *Econometrica*, 82(4):1199–1239, 2014. doi: 10.3982/ECTA10528. URL https://doi.org/10.3982/ECTA10528. 6

P. Krusell and A. A. Smith, Jr. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, 1998. doi: 10.1086/250034. URL www.jstor.org/stable/10.1086/250034. 18

F. E. Kydland and E. C. Prescott. Time to build and aggregate fluctuations. *Econometrica*, 50(6):1345–1370, 1982. doi: 10.2307/1913386. URL www.jstor.org/stable/1913386. 5

K. C. Lee, M. H. Pesaran, and R. G. Pierse. Testing for aggregation bias in linear models. *The Economic Journal*, 100(400):137–150, 1990. doi: 10.2307/2234191. 3

W. Leontief. Composite commodities and the problem of index numbers. *Econometrica*, 4(1): 39–59, 1936. doi: 10.2307/1907120. 4, 9

A. Lewbel. Aggregation without separability: A generalized composite commodity theorem. *The American Economic Review*, 86(3):524–543, 1996. doi: 10.2307/2118210. 4, 9, 10

R. E. Lucas. Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445, 1978. doi: 10.2307/1913837. URL www.jstor.org/stable/1913837. 5

J. A. Parker and B. Preston. Precautionary saving and consumption fluctuations. *American Economic Review*, 95(4):1119–1143, September 2005. doi: 10.1257/0002828054825556. URL http://www.aeaweb.org/articles?id=10.1257/0002828054825556. 6

M. Ravallion. Does aggregation hide the harmful effects of inequality on growth? *Economics Letters*, 61(1):73 – 77, 1998. ISSN 0165-1765. doi: https://doi.org/10.1016/S0165-1765(98)00139-6. 3

J. Sutton. Market structure: Theory and evidence. In M. Armstrong and R. Porter, editors, *Handbook of Industrial Organization*, volume 3, chapter 35, pages 2301–2368. Elsevier, 1 edition, 2007. 2

M. M. ter Vehn and S. Morris. The robustness of robust implementation. *Journal of Economic Theory*, 146(5):2093 – 2104, 2011. ISSN 0022-0531. doi: https://doi.org/10.1016/j.jet.2011.03.011. 2

H. Varian. *Microeconomic Analysis*. Norton International edition. Norton, 1992. ISBN 9780393960266. 6