How to measure parenting styles?

By Christopher Rauh and Laëtitia Renée*

Draft: July 28, 2021

In this paper, we measure parenting styles through unsupervised machine learning in a panel following children from age 5 to 29 months. The algorithm classifies parents into two distinct behavioral types: "positive" and "negative". Parents of the positive type tend to respond to their children's expressions in a supportive manner and describe to children features of their environment, while parents of the negative type are less likely to engage with their children in an encouraging manner. In the language of developmental psychology, positive (negative) parents exhibit both high (low) warmth and control. Although types reveal some persistence, the share of parents with positive styles decreases with the age of the child. Overall, parenting styles are systematically related to socio-economic characteristics and positive parenting is more likely amongst educated mothers. Moreover, children of positive parents see their human capital improve relative to children of parents of the negative type.

Keywords: Parenting styles; human capital; latent Dirichlet allocation; inequality; machine learning

* Rauh: University of Cambridge, Trinity College Cambridge (email: cr542@cam.ac.uk). Renée: McGill (email: laetitia.renee@mail.mcgill.ca). Rauh would like to thank the FRQSC for financial support (grant number 2020-NP-267422).

1. Introduction

Early childhood investments have been shown to be crucial for children's human capital development (Cunha, Heckman and Schennach 2010, Del Boca, Flinn and Wiswall 2014, Attanasio, Meghir and Nix 2020), but the measurement of interactions between different dimensions of investment is challenging due to their complexity (Attanasio 2015). Parental time investments generally are captured through different activities parents engage in with their children, such as visits to museums or the frequency of a parent reading to their children. The number of activities considered is either vast or restricted arbitrarily. When many investments are considered, they generally are combined (log-)linearly in latent factor models or using principal component analysis. More recently, the debate about parenting styles has emerged in economics emphasizing that not only investments but also the style of investing matters (Doepke and Zilibotti 2017).[1] In developmental psychology, Baumrind (1967) already classified parenting styles and related them to behavioral traits among pre-school children. This approach was extended by McCoby and Martin (1983) to four styles along two dimensions: warmth and control.

Parenting is characterized by a complex set of interactions and decisions. Draca and Schwarz (2018) discuss why linear combinations of features with the highest degrees of variance in the data, as principal component analysis does, may not provide optimal summaries of complex data-generating processes. This paper develops a new methodology to measure parental interactions with their children using unsupervised machine learning and is used to summarize parenting styles.[2] One advantage of this approach is that it allows aggregating any number of granular parental activities in a non-linear fashion. The algorithm learns from the occurrence of parental actions. Given that it is a mixed-member model, the same action can be assigned to different types. If, on the one hand, a parent regularly checks on a child and combines this with hugs and kisses, in the words of McCoby and Martin (1983) this could be warm control.

[1] In the earlier economics literature, parenting styles were defined as an index of punitive/aversive parenting using four questions about how the parent responds when the child misbehaves (Burton, Phipps and Curtis 2002).

[2] With the availability of the necessary data, the approach could easily be applied to summarize investments rather than styles as well.

If, on the other hand, regularly checking on a child co-occurs with yelling at the child, the parenting style could be considered controlling but not warm.

When restricting parents to two types, we find that parents can be classified into "positive" and "negative" types.[3] Positive parents are more likely to be supportive of their children's progress and speak directly to their child, while negative parents are characterized by hardly interacting with their children in the presence of the interviewer, and if they do, they tend to do so in a negative manner. In line with ad-hoc classifications in developmental psychology, the two parenting styles we discover can be interpreted as high warmth and high control (positive) vs low warmth and low control (negative). Although parenting styles exhibit some persistence over time, we find that parents are more likely to adopt positive parenting styles when the children are younger.

We contribute to two strands of literature. First, we contribute to the literature concerned about parenting styles.[4] They generally draw the distinction between parenting styles in terms of permissive, authoritarian, or authoritative. The empirical approaches tend to classify parenting styles based on a single binary response to a survey question, such as how important obedience is for a respondent (e.g., Agostinelli et al. 2020) or latent factor models (e.g., Falk et al. 2021). Our approach allows capturing parenting styles based on many questions with complex interactions. Moreover, an advantage of our data on parental activities is that they are not self-reported, but are observed and recorded by the enumerator, which should help to reduce systematic measurement error, and are the same set of actions observed across multiple survey waves. Our new interpretable measure summarizing the large dimensionality and complexity of parental activities is predictive of human capital above and beyond the

[3]The reason we limit the estimation to two styles is twofold. First, we only observe ten different parental actions, which complicates the identification of more types. Second, two types simplify the exposition.

[4]One can roughly separate this literature into two strands: First, the literature relating parenting styles to child development (e.g., Cunha 2015, Doepke and Zilibotti 2019, Doepke, Sorrenti and Zilibotti 2019, Cobb-Clark, Salamanca and Zhu 2019, Agostinelli et al. 2020). Second, the literature studying the role of parenting styles in the intergenerational transmission of traits (e.g., Brenøe and Epper 2019; Zumbuehl, Dohmen and Pfann n.d.; Falk et al. 2021). Further, Del Boca et al. (2019) propose a model in which parental types are not merely the outcome of utility maximization by the parents but the result of a bargaining process with the children. Kiessling (2020) studies how parents perceive the returns to parenting styles in terms of warmth and control using hypothetical scenarios.

predictive power of parental socio-economic characteristics or child fixed effects.[5]

Second, we add to the rapidly growing use of machine learning in Economics to classify behavioral types. The latent Dirichlet allocation (LDA) was originally developed by computer scientists Blei, Ng and Jordan (2003). The underlying idea is to classify text documents into a mixture of small number of topics. One key is that the topics are not predefined but are backed out through co-occurrence. We apply the same idea of topics to behavioral types. Other approaches to classifying behavioral types using LDA are Bandiera et al. (2020) who classify CEOs using detailed time-use surveys and find that CEOs distinct behavior affects firm performance. Draca and Schwarz (2018) use LDA to measure political ideology. We contribute to this literature by using LDA to classify parenting styles and look at its relation to human capital accumulation in very early childhood.

2. Data

We use the Québec Longitudinal Study of Child Development (QLSCD), a detailed panel of a representative sample of families from Québec, a province in Canada, with a baby born between October 1997 and July 1998. More specifically, we focus our work on the 1,985 families who participated in the first three waves of the panel, conducted when the designated baby was 5, 17 and 29 months old.

We rely on the Observations of Family Life (OFL) instrument filled by the enumerator at the end of the annual interview. It includes observations made during the interview about the behaviour of the key respondent –the mother in 99% of the cases– and her interactions with her baby. This has the advantage of not relying on self-reported behavior which is common in the human capital literature and a potential source of bias.

We exclude mother-children pairs for whom the OFL instrument was not completed at child ages 5, 17 or 29 months because the child was sleeping. We end-up with a sample of 1,443 mother-children pairs. Table 1 describes the socio-economic characteristics of the families.

We focus our analysis on the ten variables from the OFL instrument that assess the behavior

[5] Despite the intuitive results we cannot claim causal effects due to the lack of an exogenous shock to parenting styles. This is a common feature of studies measuring the impact of parenting styles.

of the interviewed mother toward her child. Table 2 displays descriptive statistics for these variables. We see that some parental actions are highly dependent on the age of the child. For instance, the share of parents regularly checking on their child decreases from 72% when the child is 5 months old to 32% when the child is 29 months old.

Table 1—: Descriptive statistics

	N	Prop.
Number of siblings		
No sibling	656	45.5
One	560	38.8
Two or more	227	15.7
Household type		
Two-parent	1,179	81.7
Blended	159	11.0
Single-parent	102	7.1
Missing	3	0.2
Mother's age		
Less than 25	332	23.0
25-29	446	30.9
30-34	469	32.5
35 and more	195	13.5
Missing	1	0.1
Mother born outside Canada		
No	1,301	90.2
Yes	140	9.7
Missing	2	0.1
Language spoken at home		
French	1,176	81.5
Other	265	18.4
Missing	2	0.1
Mother education		
High school degree or less	380	26.3
Some college education	681	47.2
College degree	380	26.3
Missing	2	0.1
Parental working status		
Two-parents: both work	984	68.2
Two-parents: one works	304	21.1
Two-parents: none work	45	3.1
Single-parent: works	38	2.6
Single-parent: does not work	58	4.0
Missing	14	1.0
Below poverty threshold		
No	1,106	76.6
Yes	317	22.0
Missing	20	1.4
N	1,443	100

Note: The table shows descriptive statistics for families in our sample at the time of the first interview in 1998, when the designated child is 5 months old.

Table 2—: Parental behaviour

	Proportion of mothers who ...		
	Wave 1 5 months	Wave 2 17 months	Wave 3 29 months
Regularly checks on her child	71.7	47.8	31.9
Speaks spontaneously to her child	40.6	43.2	46.3
Answers to her child	45.0	46.8	57.4
Kisses and hugs her child	42.6	17.3	13.9
Screams toward her child	< 0.5	4.9	6.8
Is annoyed by her child	1.6	7.1	10.5
Reprimands her child	< 0.5	4.3	5.5
Supports her child's progress	38.0	26.1	25.5
Organises play time	58.5	53.6	43.6
Gives pedagogical toys	68.2	59.0	43.7
Observations	1,443	1,443	1,443

Note: The table describes the behaviour of the respondents and their interactions with their children during the annual QLSCD interview. Behaviours are evaluated by the enumerator during the interview. Statistics are presented for the three first waves, when the designated child is 5, 17 and 29 months old.

3. Discovering latent parenting types

In the next step, the different features of parental behavior are summarized into interpretable behavioral types using a machine learning algorithm based on the latent Dirichlet allocation. This methodology developed by Blei, Ng and Jordan (2003) is a clustering algorithm for discrete data, which traditionally was meant to reduce the high dimensionality of text into an arbitrary number of topics specified by the user. Each parental action can be featured with difference importance in each type, and each parent can be a mixture of types.

The algorithm learns from the co-occurrence of counts through Bayesian learning. The idea is that if certain variables tend to appear together, they are likely to be linked to each other. In Appendix A we explain the technical details. For the sake of simplicity and interpretation, we settle on two types of parents. The final output of the algorithm is the distribution of actions for each type and the type distributions for each parent. With this information at hand, we can then relate parental types summarized into just two types to human capital accumulation.

3.1. Parenting types

We pool the three waves together and estimate the classification for that sample.[6] In Table 3 we display the relative probabilities of actions by the two types. The action that distinguishes the two types most in relative terms are supportive comments made by the parent to the child about its progress. While it is very common for the positive type to make supportive comments about the progress of the child, this is hardly the case for parents of the negative type, i.e. positive parents are 626 times more likely to do so. Similarly large differences exist for speaking to the child directly. Actions that are more typical for negative parents are reprimanding the child or expressing annoyance.

The distribution of actions across types suggests that what distinguishes parents is the richness and warmth of action by one type versus the authoritarian and control by the other,

[6]We could estimate a different classification for each wave separately as some actions might be more pertinent for different ages of the child, as is indicated by the distribution of actions in Table 2. However, the parental classification would not be comparable over time, which would pose other challenges for the rest of our analysis.

Table 3—: Classification of behaviors by parental types

	Ratio type 1 to type 2
Supports child's progress	626
Speaks directly to child	454
Answers child	260
Kisses and hugs child	256
Organizes play time	14.4
Gives pedagogical toys	11.0
Regularly checks on child	3.63
Screams at child	1.71
Reprimands child	0.94
Expresses annoyance by child	0.78

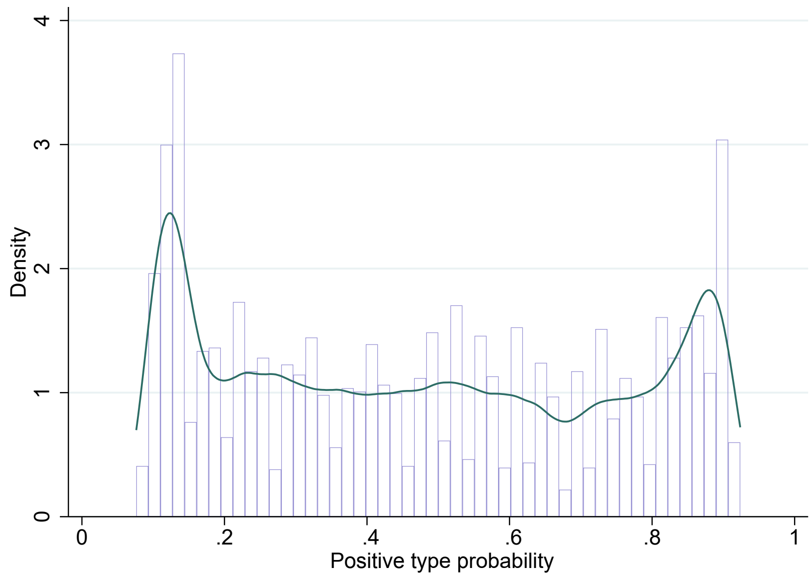
Note: The table describes the occurrence of behaviours for the two types found by the LDA algorithm. Behaviours are ranked from what is the relatively most likely behavior of type 1 relative to type 2. We label these types positive (type 1) and negative (type 2). The second column displays the ratio of the probability for type 1 over the probability for type 2.

hence the labels positive and negative parents.

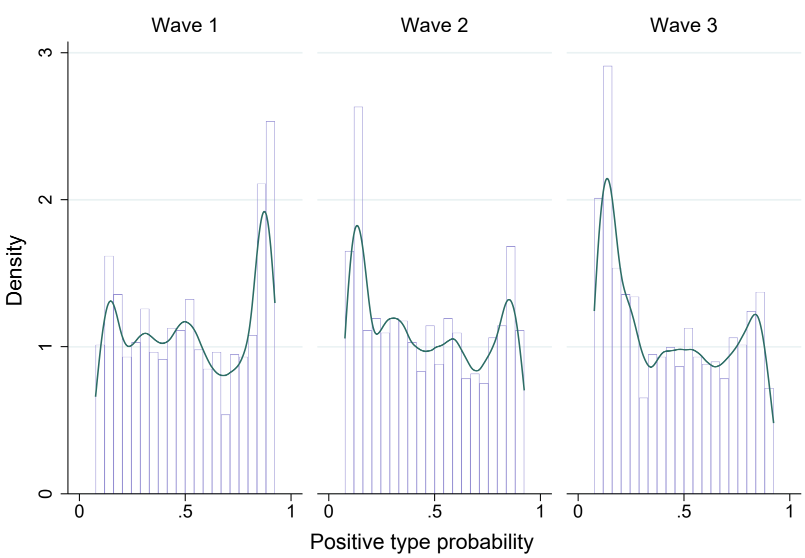
The LDA algorithm assigns to each parent a probability of being of type 1, the positive type (and with the remaining probability they are of type 2, the negative type). The top panel of Figure 1 shows the distribution of the positive type probability for the full sample and the bottom panel for each wave separately. We see a concentration of two masses: one with a low probability of being of the positive type (i.e. with a high probability of being of the negative type) and the opposite. Over time, parents tend to move from the positive type to the negative type.

Figure 1. : Distribution of types

a) Pooled across all waves



b) For each wave separately

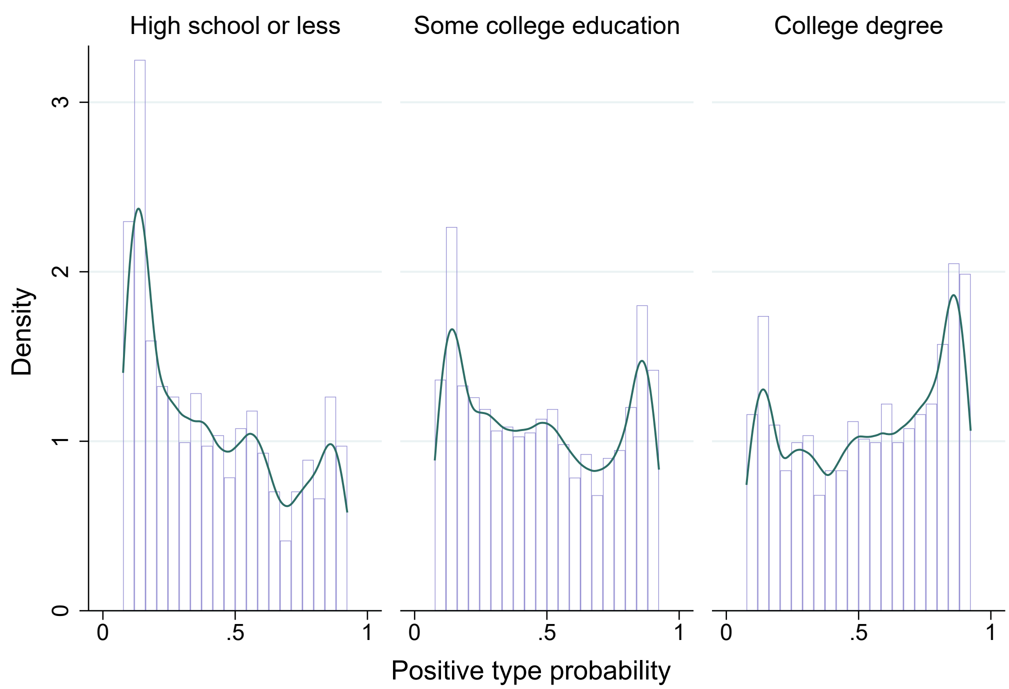


Notes: The transparent bars represent the binned probabilities of the probability of being a positive type rather than a negative type, while the solid line is the kernel density.

3.2. Correlates and persistence of parenting types

In Figure 2 we show the distribution of positive types by maternal education. In the left panel we see that mothers with high school or less tend to be of the negative type with an average share of positive types of 41.8%. In the middle panel we see that for mothers with some college education the distribution appears closer to bi-modal with an average probability of positive types of 48.3%. Finally, in the right panel we see that amongst more educated mothers with a college degree, the average likelihood of being of the positive type increases to 53.9%.

Figure 2. : Distribution of types by maternal education



Notes: The transparent bars represent the binned probabilities of the probability of being a positive type, while the solid line is the kernel density.

While the previous figure suggests that the likelihood of being a positive mother is increasing in education, we take a more systematic look at the relationship between type and individual characteristics by regressing the probability of being a positive type on age, education, poverty level, whether the parent is an immigrant, marital status, employment status, number of siblings, and the gender of child. In the first column of Table 4 we see the results for the pooled sample and in the following three columns for each age of the child, separately.

We see that parents with more than one child tend to be less likely to be of the positive type. The probability of being positive appears to be increasing in maternal age and education. While some of the coefficients vary, in general the direction of coefficients is very similar across waves. Maternal types reveal a considerable persistence as suggested by the correlations across waves exhibited in Table 5. Between wave 1 and wave 2 the correlation in types is 0.26, and between wave 2 and wave 3 it is 0.36. In fact, regressing individual fixed effects on parenting types achieves an R^2 of 0.52. We further breakdown the persistence in Table 6 in which we show the transition matrix between positive types (defined as being of the positive type with a probability above 0.67), an intermediate type (positive type with a probability between 0.33 and 0.67), and the negative type (positive type with a probability of less than 0.33). According to this matrix 38% (51%) of positive (negative) mothers in wave 1 are of the same type in wave 2, and 42% (58%) of positive (negative) mothers in wave 2 are of the same type in wave 3.

Table 4—: Positive type probability and parental characteristics

	Probability of being of the positive type			
	Pooled	Wave 1	Wave 2	Wave 3
Number of siblings (reference: no sibling)				
One sibling	-0.043***	-0.037**	-0.057***	-0.035**
	(0.011)	(0.016)	(0.016)	(0.016)
Two or more siblings	-0.060***	-0.039*	-0.105***	-0.035
	(0.017)	(0.023)	(0.023)	(0.023)
Household type (reference: two-parents family)				
Blended family	0.013	-0.005	0.035	0.010
	(0.018)	(0.025)	(0.024)	(0.023)
Single-parent household	-0.146	-0.160	-0.102	-0.176*
	(0.127)	(0.193)	(0.178)	(0.103)
Mother's age (reference: less than 25)				
25-29	0.024	0.053***	0.013	0.006
	(0.015)	(0.021)	(0.021)	(0.021)
30-34	0.055***	0.089***	0.046**	0.029
	(0.016)	(0.021)	(0.023)	(0.021)
35 and more	0.082***	0.100***	0.071***	0.075***
	(0.019)	(0.027)	(0.027)	(0.026)
Mother born outside Canada (reference: no)				
Yes	-0.035*	-0.061**	-0.042	-0.003
	(0.020)	(0.029)	(0.028)	(0.029)
Language spoken at home (reference: French)				
Other	-0.034**	0.034	-0.060***	-0.075***
	(0.014)	(0.021)	(0.020)	(0.021)
Mother education (reference: high school degree or less)				
Some college education	0.050***	0.042**	0.037**	0.073***
	(0.013)	(0.018)	(0.018)	(0.018)
College degree	0.089***	0.084***	0.079***	0.103***
	(0.015)	(0.022)	(0.022)	(0.022)
Parental working status (reference: two-parents: both work)				
Two-parents: one works	-0.001	0.014	-0.005	-0.012
	(0.013)	(0.019)	(0.019)	(0.019)
Two-parents: none work	-0.033	-0.004	-0.088*	-0.006
	(0.034)	(0.045)	(0.047)	(0.045)
Single-parent: works	0.136	0.127	0.091	0.189*
	(0.133)	(0.199)	(0.184)	(0.113)
Single-parent: does not work	0.151	0.147	0.102	0.204*
	(0.132)	(0.198)	(0.183)	(0.113)
Below poverty threshold (reference: no)				
Yes	-0.025	-0.055**	0.003	-0.022
	(0.016)	(0.023)	(0.022)	(0.022)
Constant	0.436***	0.455***	0.455***	0.400***
	(0.0159)	(0.022)	(0.022)	(0.022)
Observations	4,329	1,443	1,443	1,443
R-squared	0.050	0.070	0.060	0.056

Note: Each column presents the estimates of an OLS regression of positive type probability on parental characteristics. The categories for missing values are also included in the regression but not shown in the table as they only concern a few individuals and are thus hard to interpret. Robust standard errors clustered at the family level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5—: Correlation matrix of positive-type probability across waves

	Wave 1	Wave 2	Wave 3
Wave 1	1.00	.	.
Wave 2	0.26	1.00	.
Wave 3	0.24	0.36	1.00

Note: The table displays the correlation between the positive type probability variable in wave 1 and the one in wave 2, the positive type probability variable in wave 2 and the one in wave 3, and the positive type probability variable in wave 2 and the one in wave 3.

Table 6—: Transition matrix between binned types

(a) Between waves 1 and 2

	Wave 2		
Wave 1	Positive	Intermediate	Negative
positive	0.38	0.37	0.25
Intermediate	0.27	0.36	0.37
Negative	0.17	0.32	0.51

(b) Between waves 2 and 3

	Wave 3		
Wave 2	Positive	Intermediate	Negative
Positive	0.42	0.36	0.23
Intermediate	0.27	0.35	0.37
Negative	0.14	0.28	0.58

Note: The first table presents the transition matrix between positive types (defined as being of the positive type with a probability above 0.67), an intermediate type (positive type with a probability between 0.33 and 0.67), and the negative type (positive type with a probability of less than 0.33) between wave 1 and wave 2. The second table presents the same transition matrix between wave 2 and wave 3.

4. Relating parenting types to children's outcomes

To test the relationship between parental type and the accumulation of children's cognitive skills, we use the results from an Imitation Sorting Task (IST) test conducted during each wave.[7] Here the sample size reduces to 1,121 children who took the IST test at 5, 17 and 29 months. Excluded children were sleeping or sick at the time the test was supposed to take place or the test was not fully completed. The test score in each wave is standardized with a mean of 0 and a standard deviation of 1.

In the first column of Table 7 we show the results of the pooled sample in which we regress the IST test score at each of the three stages on the probability of being a positive parent and a constant. We find that moving from a negative to a positive parent is associated with an increase in the IST score of 0.223 standard deviations. In the second column we add controls for parental characteristics and still find a highly significant positive association between the probability of being a positive parent and test scores of 0.167 standard deviations. In the third column we control for parental fixed effects, thereby removing any constant heterogeneity across parents and children. Using this specification we find a strengthened association between being a positive type and cognitive development with a highly significant coefficient of 0.338.

[7]The task comprises different situations in which the infant must grasp objects placed in front of him/her and place them in given containers. The task used in the ELDEQ is a variation of the Imitation Sorting Task developed by Uzgiris and Hunt (1975).

Table 7—: Positive type probability and cognitive development

	Standardized IST score		
	(1)	(2)	(3)
Positive type probability	0.223***	0.167**	0.338***
	(0.069)	(0.069)	(0.108)
Observations	3,363	3,363	3,363
R-squared	0.004	0.020	0.387
Family characteristics	NO	YES	NO
Family FE	NO	NO	YES

Note: Each column presents the estimates of an OLS regression of the child standardized IST score on her mother positive type probability. Family characteristics (column 2) include family composition (number of siblings, household type), maternal characteristics (age, whether born outside Canada, educational attainment), parental working status, language spoken at home, and whether family is below poverty threshold. They are described in more details in Table 1. Robust standard errors clustered at the family level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

5. Conclusion

Human capital accumulation is one of the most important fundamentals of productivity and innovation. However, estimating human capital production functions is riddled with complications including the high dimensionality and potentially non-linear relationships between parental actions. In this paper, we provide a new way to summarize parental styles adopted from computational linguistics. We use an unsupervised machine learning model, the latent Dirichlet allocation, to classify parents into two types. The resulting types can be interpreted as positive parents who encourage their children and express their affection, versus negative parents who do not interact much with their children and are more likely to punish when they do so.

We show that these two types relate systematically to parental characteristics, i.e. mothers with higher education tend to be more likely to be of the positive type. Moreover, we show that children of more positive parents tend to achieve higher levels of human accumulation. While we cannot establish a causal relationship between parenting types and outcomes due to the nature of the data, we are optimistic that future studies including natural experiments or randomized control trials can make use of the proposed methodology to classify parents into types based on their their actions. Another advantage of the approach is that this can be done with an extremely large set of actions or even detailed time use data.

REFERENCES

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2020. "It takes a village: the economics of parenting with neighborhood and peer effects." National Bureau of Economic Research.

Attanasio, Orazio, Costas Meghir, and Emily Nix. 2020. "Human capital development and parental investment in India." *Review of Economic Studies*, 87(6): 2511—-2541.

Attanasio, Orazio P. 2015. "The determinants of human capital formation during the early years of life: Theory, measurement, and policies." *Journal of the European Economic Association*, 13(6): 949–997.

Bandiera, Oriana, Stephen Hansen, Andrea Prat, and Raffaella Sadun. 2020. "CEO Behavior and Firm Performance." *Journal of Political Economy*, 128(4): 1325–1369.

Baumrind, Diana. 1967. "Child care practices anteceding three patterns of preschool behavior." *Genetic psychology monographs.*

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, 3(Jan): 993–1022.

Brenøe, Anne, and Thomas Epper. 2019. "Parenting Values Moderate the Intergenrational Transmission of Time Preferences."

Burton, Peter, Shelley Phipps, and Lori Curtis. 2002. "All in the family: A simultaneous model of parenting style and child conduct." *American Economic Review*, 92(2): 368–372.

Cobb-Clark, Deborah A, Nicolas Salamanca, and Anna Zhu. 2019. "Parenting style as an investment in human development." *Journal of Population Economics*, 32(4): 1315–1352.

Cunha, Flavio. 2015. "Subjective rationality, parenting styles, and investments in children." In *Families in an Era of Increasing Inequality.* 83–94. Springer.

Cunha, Flávio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, 78(3): 883–931.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. "Household Choice and Child Development." *The Review of Economic Studies*, 81(1): 137–185.

Del Boca, Daniela, Christopher J Flinn, Ewout Verriest, and Matthew J Wiswall. 2019. "Actors in the Child Development Process." National Bureau of Economic Research.

Doepke, Matthias, and Fabrizio Zilibotti. 2017. "Parenting with style: Altruism and paternalism in intergenerational preference transmission." *Econometrica*, 85(5): 1331–1371.

Doepke, Matthias, and Fabrizio Zilibotti. 2019. *Love, money, and parenting: How economics explains the way we raise our kids.* Princeton University Press.

Doepke, Matthias, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2019. "The economics of parenting." *Annual Review of Economics*, 11.

Draca, Mirko, and Carlo Schwarz. 2018. "How polarized are citizens? Measuring ideology from the ground-up." Mimeo.

Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers. 2021. "Socio-economic status and inequalities in children's IQ and economic preferences." *Journal of Political Economy*, 129(9).

Hoffman, Matthew, Francis R Bach, and David M Blei. 2010. "Online learning for latent dirichlet allocation." 856–864.

Kiessling, Lukas. 2020. "How do parents perceive the returns to parenting styles and neighborhoods?" *MPI Collective Goods Discussion Paper*, , (2020/14).

McCoby, Eleanor, and John Martin. 1983. "Socialization in the context of the family: Parent-child interaction." *Handbook of child psychology*, 4: 1–101.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research*, 12: 2825–2830.

Uzgiris, Ina C, and Joseph Hunt. 1975. *Assessment in infancy: Ordinal scales of psychological development.* University of Illinois Press.

Zumbuehl, Maria, Thomas Dohmen, and Gerard A Pfann. n.d.. "Parental involvement and the intergenerational transmission of economic preferences, attitude and personality traits." *Economic Journal.*

Appendix A: Latent Dirichlet allocation

Adapting the technical terms from Blei, Ng and Jordan (2003) for text and applying to our objective, the corpus of behavioral actions D is composed of parents w of actions. A behavioral type is a probability distribution over all actions. The assumed underlying process with which types generate actions is by drawing θ from a Dirichlet distribution with hyperparameter α. Then for each action n of all actions N, one chooses a type from z_n. After that an action w_n is chosen for the corresponding type z_n from a Dirichlet distribution with hyperparameter β.

Written formally, the generative process of actions is expressed as the following joint distribution

$$p(\beta, \theta, z, w_d) = \prod_{i=1}^{k} p(\beta_i) \prod_{d=1}^{D} p(\theta_d) (\prod_{n=1}^{N} p(z_{d,n}|\theta_d) p(w_{d,n}|\beta, z_{d,n})).$$

Given the corpus of actions, the task of the algorithm is to infer the type-specific action distribution and the parent specific type distribution. So the posterior distribution of the latent variables is given by

$$p(\beta, \theta, z|w_d) = \frac{p(\beta, \theta, z, w_d)}{p(w_d)}.$$

In order to infer the marginal distribution $p(w_d)$, which can be done through approximation using Gibbs sampling, or Variational Kalman Filtering and Variational Wavelet Regression, we rely on the Stata implementation developed by Draca and Schwarz (2018). Draca and Schwarz (2018) use the inference algorithm developed by Hoffman, Bach and Blei (2010) and implemented by Pedregosa et al. (2011). As is the case in Draca and Schwarz (2018), the assumption of the independence of responses does not strictly hold in our approach. If an action has been recorded the same action is not recorded again for the same person. They discuss in detail why the inference of LDA is nonetheless still valid.