Research Proposal:
Estimating High Frequency Income Risk using *eIndkomst*

Jeppe Druedahl
Department of Economics, University of Copenhagen

June 1, 2016

1 Project Proposal

Introduction

The business cycle fluctuations in aggregate private consumption is larger than standard consumption-saving models predict when we require that they also match the distribution of wealth.[1] The reason is that a large share of households should be able to smooth consumption because they have access to substantial financial buffers in form of e.g. housing equity. This is a problem for models of monetary and fiscal policy where the consumption response is important. Recent research has, however, shown that a more realistic high frequency specification of the income risk households face can help solve the puzzle.

> *The objective of the study proposed here is to estimate a realistic high frequency process for the income risk faced by Danish households using the new eIndkomst register.*

The *eIndkomst* register, which began in January 2008 and now covers more than seven years, is a worldwide sensation because it for the first time allows us to observe the income of the full population of households at a *monthly* frequency. Combined with the 30+ years time span of the *annual* income registers, and all the available background controls, I will be able to construct the to date most complete picture of the income risk (Danish) households face.[2]

[1] See Carroll (2000), Carroll, Slacalek, Tokuoka and White (2014) and McKay (2015).

[2] The previous literature on estimating income processes has used annual or even bi-annual income data; see e.g. Lillard and Willis (1978), Lillard and Weiss (1979). MaCurdy (1982), Baker (1997), Meghir and Pistaferri (2004), Guvenen (2009), Browning, Ejrnæs and Alvarez (2010), Altonji, Smith and Vidangos (2013), and Guvenen, Karahan, Ozkan and Song (2015).

Estimating a realistic high frequency income process is firstly interesting in its own right for evaluating in which dimensions the Danish social security net is respectively more and less fine meshed. Secondly, estimates of the income risk households face are important inputs for calibrating models of consumption and saving, which are central for understanding e.g. the costs business cycle fluctuations (Krebs, 2007), the equity premium puzzle (Constantinides and Ghosh, 2014), and for determining the effectiveness of monetary and fiscal policy (Kaplan, Moll and Violante, 2016).

Theoretical background

The importance of modeling the frequency domain of the income process is explained in detail in a recent path-breaking paper by Kaplan, Moll and Violante (2016): "Households who face small, but frequent, shocks have a strong incentive to hold low-return liquid assets to smooth consumption, while households who face large infrequent shocks would prefer to hold high-return illiquid assets that can be accessed at a cost in the unlikely event of a sizable windfall or a severe income loss." (p. 20)

Using an income process with both frequent shocks and larger infrequent shocks, they are able to construct a general equilibrium model where there are many *wealthy hands-too-mouth* households who do not smooth consumption in the face of small income shocks due to the transaction costs of tapping into their illiquid wealth.[3] Hereby they are able to get large direct effects of monetary policy and fiscal adjustments such as tax rebates. This underlines the high policy relevance of precise estimates of the income risk households face for developing quantitative models of monetary and fiscal policy.

In order to estimate their high frequency income process the authors, however, have to rely on high order moments from *annual* income changes. It can e.g. be shown that a more leptokurtic distribution with fat tails is more likely to have been generated by an income process with large infrequent shocks.[4] In particular, they use high order moments estimated in Guvenen, Karahan, Ozkan and Song (2015) using a 10 percent random sample of all US working age males from 1978 to 2011. This is clearly far from optimal, and therefore having a full population seven year panel of monthly incomes such as *eIndkomst* is simply a worldwide sensation.

Methodology

The empirical analysis in the proposed study will be conducted in three parts.

1) Non-parametric exploration. In the first part I (and my research assistant) will look at the data non-parametrically**.** Specifically, I will first look at the *distribution of*

[3] The concept of wealthy hands-too-mouth households was originally developed in Kaplan and Violante (2014), and documented empirically in Kaplan, Violante and Weidner (2014).

[4] Klein and Telyukova (2013) conduct a similar analysis.

income growth rates at various horizons (e.g. one-month, one-year, three-years etc.), and across groups categorized by age, current income, long-run income, job tenure, and household composition. The focus will particularly be on the first four moments as in the widely cited study by Guvenen, Karahan, Ozkan and Song (2015). Expected results include strong deviations from log-normality in forms of both strong negative skewness and high kurtosis, and substantial changes in up- and down-side risks over both the life-cycle and the income distribution.

Secondly, I will compute impulse-response functions, and look for heterogeneity in the persistence of income shocks across the previously defined groups, the size of the shock, and discrete events such as unemployment, job and sector changes, long-term sickness etc. The benefit of monthly data is very valuable for the latter, as the event is then very precisely identified.

Thirdly, I will use the annual income registers for a detailed comparison with the results in Guvenen, Karahan, Ozkan and Song (2015). This includes an analysis of what in the *monthly* data drives the increase in income risk during recessions found in studies using *annual* income data. This will enable me to make progress in determining whether counter-cyclical income risk is mostly due to *increased variance* or *increased left-skewness*; Storesletten, Telmer and Yaron (2004) argued for the former using US data, while Guvenen, Ozkan and Song (2014) and Busch, Domeij, Guvenen and Madera (2015) argue for the latter using US, German and Swedish data.

2) Parametric estimation. In the second part, I will estimate a fully parameterized model for the income process. Such an estimated process is useful both for the concreteness of the results, for comparison with the previous literature, and as an input to structural models of households consumption and saving. I will start from a large flexible model with both a set of discrete states (e.g. working, unemployed, out of the labor market etc.), heterogeneous life-cycle profiles of income, heterogeneous transitory shock variances, and a heterogeneous mixture of multiple first order autoregressive processes.

Based on the non-parametric analysis, I will choose a set of interesting moments, and estimate the model by the *method of simulated moments* (MSM), which minimizes the distance between the moments in the actual data and the moments in data simulated from the model. I will then extensively discuss what the cost is of restricting the income process in various dimensions. This is especially important because a high dimensional income process with lots of heterogeneity is computationally infeasible as an input to a structural model.

3) Machine learning. In the third part, I will focus on uncovering a low dimensional discrete state approximation of the income process the households face, that can serve as an input for structural economic models. Promising approaches given the Big data nature of the Danish registers are machine learning algorithms such as an *artificial neural net* or a *random forrest* (Varian, 2014). The black box nature of these procedures

imply that they cannot replace the more structured estimation in the second part, but the hope is that a higher level of predictive power per discrete state can be obtained.

2 Output, Time Frame and Further Perspectives

The main output of this study will be a single academic paper, which due to the uniqueness of monthly income data and the high policy relevance of precise estimates of high frequency income risk, has the potential for publication in a top-5 journal or at least a top field macroeconomic journal. The data exploration will be initiated in the fall of 2016, and a first draft of the paper will be ready for circulation mid-2017. Submission is planned for the end of 2017. As a secondary outcome I plan to write a newspaper feature article on the income risk faced by Danish households.

Additionally, the proposed study has substantial synergistic effects with various structural estimation projects on Danish register data, we are conducting at the *Centre for Computational Economics,* where estimates of income risk is an important calibration input. In particular, I am constructing a model of the business cycle fluctuations in the demand for housing and cars. A follow-up project could thus be to use this model to analyze which dimensions of the income risk process, estimated in the study proposed here, which is most important for consumption dynamics. A similar follow-up could be conducted using the model in Kaplan, Moll and Violante (2016), which I will be talking to the authors about (I visited Violante (NYU) for the fall term of 2014, and Kaplan (Princeton) was in my PhD committee).

3 Budget

This application mainly concerns funds for hiring a research assistant for 300 hours for helping with the large amount of data work in this project. I estimate the hours needed to be high as the *eIndkomst* data is still relatively unused, and I need to merge it with other registers (e.g. DREAM) and require some level of consistency with the annual income data.

Additionally it concerns funds for purchasing the relevant data. This naturally includes purchasing the *eIndkomst* registers (BFL and IMLE), but also for purchasing various updated versions of the income and demographic registers such that the time span of the *eIndkomst* registers can be utilized in full.

Finally it concerns funding for presenting the work at a conference in the summer of 2017.

I am not applying for funds for my own salary as I am fully funded as a post-doc until December 2019.

Table 1: Budget, in Danish Kroner.

	Cost
Research assistance 2016 (~100 hours)	16,000
Research assistance 2017 (~200 hours)	32,000
Conference (summer 2017)	7,000
Data (eIndkomst, updates of various registers)	15,000
Sub-total	**70,000**
Overhead 20%	14,000.00
Total	**84,000.00**

References

Altonji, J. G., A. A. Smith and I. Vidangos (2013): "Modeling Earnings Dynamics," *Econometrica*, 81(4), 1395–1454.

Baker, M. (1997): "Growth-rate heterogeneity and the covariance structure of life-cycle earnings," *Journal of Labor Economics*, 15(2), 338–375.

Browning, M., M. Ejrnæs and J. Alvarez (2010): "Modelling Income Processes with Lots of Heterogeneity," *The Review of Economic Studies*, 77(4), 1353–1381.

Busch, C., D. Domeij, F. Guvenen and R. Madera (2015): "Asymmetric Business Cycle Risk and Government Insurance," Working Paper.

Carroll, C. D. (2000): "Requiem for the representative consumer? Aggregate implications of microeconomic consumption behavior," *The American Economic Review: Papers and Proceedings of the One Hundred Twelfth Annual Meeting of the American economic Association*, 90(2), 110–115.

Carroll, C. D., J. Slacalek, K. Tokuoka and M. White (2014): "The Distribution of Wealth and the Marginal Propensity to Consume," Working Paper.

Constantinides, G. M. and A. Ghosh (2014): "Asset Pricing with Countercyclical Household Consumption Risk," NBER Working Paper 20110.

Guvenen, F. (2009): "An empirical investigation of labor income processes," *Review of Economic Dynamics*, 12(1), 58–79.

Guvenen, F., F. Karahan, S. Ozkan and J. Song (2015): "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?," NBER Working Paper 20913.

Guvenen, F., S. Ozkan and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122(3), 621–660.

Kaplan, G., B. Moll and G. Violante (2016): "Monetary Policy According to HANK," Working Paper.

Kaplan, G., G. Violante and J. Weidner (2014): "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, pp. 77–138.

Kaplan, G. and G. L. Violante (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82(4), 1199–1239.

Klein, P. and I. A. Telyukova (2013): "Measuring high-frequency income risk from low-frequency data," *Journal of Economic Dynamics and Control*, 37(3), 535–542.

Krebs, T. (2007): "Job Displacement Risk and the Cost of Business Cycles," *The American Economic Review*, 97(3), 664–686.

Lillard, L. A. and Y. Weiss (1979): "Components of Variation in Panel Earnings Data: American Scientists 1960-70," *Econometrica*, 47(2), 437–454.

Lillard, L. A. and R. J. Willis (1978): "Dynamic Aspects of Earning Mobility," *Econometrica*, 46(5), 985–1012.

MaCurdy, T. E. (1982): "The use of time series processes to model the error structure of earnings in a longitudinal data analysis," *Journal of econometrics*, 18(1), 83–114.

McKay, A. (2015): "Time-Varying Idiosyncratic Risk and Aggregate Consumption Dynamics," Working Paper.

Meghir, C. and L. Pistaferri (2004): "Income variance dynamics and heterogeneity," *Econometrica*, 72(1), 1–32.

Storesletten, K., C. Telmer and A. Yaron (2004): "Cyclical dynamics in idiosyncratic labor market risk," *Journal of Political Economy*, 112(3), 695–717.

Varian, H. R. (2014): "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 28(2), 3–28.