A Bayesian approach to labour market modelling in dynamic microsimulation[1]

3. december 2013

Hans Bækgaard
hab@dreammodel.dk
+45 33 30 79 60

Abstract

The paper presents a novel approach to modelling labour market processes in dynamic microsimulation. The method combines and integrates Bayesian simulation based estimation and simulation of the dependent variables. The approach is applied to a dynamic panel model for hourly wage rates for Danish employees using a large panel data set with 17 years of data for 1995 to 2011. The wage rate model and a parallel model for annual work hours are currently being implemented in SMILE (Simulation Model for Individual Lifecycle Evaluation), a new dynamic microsimulation model for the Danish household sector.

The application benefits from the richness of Danish administrative panel data. Nevertheless, the results and the approach have several features that should be of interest to micro-simulators and others. Indeed, the model features both an extraordinarily comprehensive list of dependencies and a rich dynamic structure. Together, these features contribute to ensure that simulations produce realistic cross-sectional distributions and interactions as well as inter-temporal mobility – the key determinants of the quality of a dynamic microsimulation model. In addition to the 'usual' socio-demographic variables (gender, age, ethnicity, experience, education etc.), the dependencies include a more novel set of variables that represent a person's labour market history, secondary school grade and social heritage (represented by the parent's education level). The dynamic model structure includes a lagged dependent variable, an auto-correlated error term with a mixed Gaussian distribution for the white noise component, an individual random effect with a mixed Gaussian distribution and permanent effect of a person's first wage after leaving the education system. The estimation sample is identical to the simulation sample, which allows us to use the same historical detail as well as estimated individual effects – i.e. random effect components – for the simulation of future wage rates.

The Bayesian estimation method handles missing observations for the dependent variable – due to either non-employment or temporary non-participation – by treating missing observations as latent variables that are simulated alongside the Bayesian iterations. As a byproduct, the estimations produce model consistent latent wage rates for the unemployed that are useful for labour supply analysis.

---

[1] The work presented in this paper builds upon previous work documented in Bækgaard (2010) that has been extending and elaborated for the purpose of adaptation to a dynamic microsimulation.

1 Introduction

The paper presents the results from a dynamic panel model for hourly wage rates for a panel of Danish employees from 1995 to 2010. A Bayesian technique is used for estimating the dynamic panel model that includes a novel approach for handling missing observations for the dependent variable. This is achieved by treating the missing observations as latent variables that are simulated alongside the model's parameters during the Bayesian iterations. As a bi-product, model consistent values for the missing dependent variables are simulated.

The method is applied to hourly wage rates in the Danish Salary Register (Lønregistret) where wages rates are absent in years when a person is either out of employment or the employer is not participating because the workplace is not in scope due to small size.[2] The analysis provides new interesting insights into the dynamics of the wage process as well as an understanding of the factors that drive hourly wage rates. The estimated equations are used for forecasting the distribution of wage rates in the new Danish dynamic microsimulation model SMILE.

The analysis takes advantage of the Danish administrative registers with comprehensive demographic and labour market related information, including detailed histories of social transfer receipts, secondary school grade and social heritage (represented by the parent's education level). The social security history provides important information about the duration of periods out of employment and how they impact on subsequent income earning potential. The effect of unemployment spells on future and hence potential wage rates is estimated directly through the inclusion of variables for the duration of periods out of work. Importantly, this represents an alternative and effective way to account for the well-known Heckman selection bias (Heckman, 1979). The results are interesting in their own right and demonstrate that unemployment and other periods of non-employment have a significant negative impact on expected subsequent wages and the negative effect increases with the duration of periods out of work

The dynamic model structure includes a lagged dependent variable, an auto-correlated error term with a mixed Gaussian distribution for the white noise component, an individual random effect with a mixed Gaussian distribution and permanent effect of a person's first wage after leaving the education system. The estimation sample is identical to the simulation sample, which allows us to use the same historical detail as well as estimated individual effects – i.e. random effect components – for the simulation of future wage rates.

The combination of a rich background information and extensive use of individual historical wage formation processes is an unusual but effective way of ensuring realistic simulation of cross-sectional and inter-temporal distributions. The mixed Gaussian distributions for the residuals and the random effect illustrate the importance of taking non-normality into account.

[2] The wage rate model's specification has much in common with the model in Geweke and Keane (2000).

2 The SMILE labour market briefly

The wage rate model is an integrated component of SMILE's labour market module that simulates transitions between labour market statuses and earned income for employed people. The labour market module has the following main components:

Labour market transitions

The transitions between labour market statuses are governed by a combination of survival models, annual transitions and rule-based allocations. The model is hierarchical and the transitions are executed in two steps. The upper level determines the main labour market status defined by the four distinct categories 1: employed, 2: unemployed, 3: temporary out of the labour force, and 4: permanent out of the labour force (Bækgaard, 2014a, forthcoming). The transitions between employment, unemployment and temporary out of the labour force are based on hazard functions while permanent exits from the labour force are determined by annual retirement events.[3]

The lower level labour market categories provide more detail such as distinctions between wage earners and self-employed and type of benefit entitlements and participation in Government labour market programmes for the non-employed.

Wage rates and potential wage rates

This is the topic of the present paper.

Annual (full-year) paid work hours for wage earners

Annual worked hours are simulated by a dynamic panel model, which has much in common with the wage rate model (Bækgaard, 2014b, forthcoming). It is an annual model based on annualized paid work hours, which represent the full-year number of hours a person is paid for if the person is employed all months of the year. Persons who are employed part of the year thus have their worked hours scaled down from full-year to the number of months they are actually employed.

Persons who have gaps in the estimation data either because they have not worked at all during the year or did not had employment covered by the Danish Salary Register (Lønregisteret) will be treated as missing observations and treated in the same way as missing observations in the wage rate model.[4]

The dependent variable is annual *paid* hours as opposed to annual *actual* hours worked, which is the concept adopted by national accounts in most countries as well as by the OECD. This is important, as it allows us to derive annual earned income as a product of number of months employed, annualized paid hours and hourly wage rates.

Annual employment income for the self-employed

Income from self-employment is simulated by a dynamic panel model for annual taxable income and based upon tax return data. The model structure and general approach is otherwise similar to the wage rate model for employees.

---

[3] See for example Arnberg and Stephensen (2013).

[4] The worked hours and wage rate models are both based on data from the Danish Salary Register.

3 The wage rate model

The model's empirical specification is a general dynamic panel model with multiple sources of heterogeneity and inter-temporal dependencies. The initial period is represented by a separate non-dynamic wage rate equation (1) while equation for the remaining periods (2) include dynamics through a lagged dependent, a first-order autoregressive error term (3) and (4), a random effect (5) and first-period error term contingency (6):

(1) $y_{i1} = x_{i1}\beta_1 + \varepsilon_{i1}$

(2) $y_{it} = \gamma y_{it-1} + x_{it}\beta_2 + \tau_i + \phi\varepsilon_{i1} + \varepsilon_{it}$ (t>1)

(3) $\varepsilon_{i2} = \rho_1 \varepsilon_{i1} + e_{i2}$

(4) $\varepsilon_{it} = \rho_2 \varepsilon_{it-1} + e_{it}$ (t>1)

(5) $\phi = \sum_{g=1}^{2} e_{ig}^{\phi} \phi_g$

(6) $\tau_i = \sum_{j=1}^{m_\tau} e_{ij}^{\tau}(\alpha_{\tau j} + \sigma_{\tau j}\xi_i)$ $\quad \xi_i \sim n(0,1)$ and $P(e_{ij}^{\tau} = 1) = p_j^{\tau}$

(7) $\varepsilon_{i1} = \sum_{g=1}^{2}\left[\sum_{j=1}^{m_{1g}} e_{ij}^{1g}(\alpha_{gj}^{1} + \sigma_{gj}^{1}\xi_i)\right]$ $\quad \xi_i \sim n(0,1)$ and $P(e_{ij}^{g} = 1) = p_j^{g}$

(8) $e_{i,t} = \sum_{j=1}^{m_2} e_{ij}^{2}(\alpha_j^2 + \sigma_j^2 \xi_{it})$ $\quad \xi_{it} \sim n(0,1)$ and $P(e_{ij}^{2} = 1) = p_j^{2}$ (t>1)

y_{it} is the logarithm of the hourly wage rate for person *i* in year *t*, (x_{i1}, x_{it}) are the background variables for the first year (t=1) and the subsequent years (t>1) with parameters β_1 and β_2 respectively. τ_i is an individual random effect. γ is the parameter for the lagged dependent. ρ_1 and ρ_2 are the AR-parameters for period 1 to 2 and for subsequent periods respectively. The parameter $\phi = (\phi_1, \phi_2)$ is a permanent effect of the first period's error. The effect differs for persons who are entering the estimation panel because they have obtained employment for the first time after leaving the education system (ϕ_1) and persons who are entering the estimation panel at a later stage in their labour market career (ϕ_2). Hence ϕ_1 represents the persistent effect of the wage rate in a person's first job on the wage rate in subsequent years. The distinction between education leavers and others in the first year has the added implication that the variance of first-period error term is group specific.

The somewhat tedious notation in (6), (7) and (8) represents a mixed Gaussian distribution for the random effect and the error terms. A mixed Gaussian (or normal) distribution is a combination of two or more – in our case a maximum of four – normal distributions that, when combined, emulate the empirical error term distributions and, as we shall see, it often does so far better than a single normal distribution.

Using the random effect to illustrate, assume that $\{\tau_i\}_{i=1,....,n}$ is an observed sample from a mixed normal distribution of *m* normal distributions $n(\mu_j, \sigma_j^2)_{j=1,....,m}$ that is

$$\tau_i = \sum_{j=1}^{m} e_{ij}(\alpha_j + \sigma_j \xi_i) \quad \xi_i \sim n(0,1)$$

$e = \{e_{ij}\}_{i=1,...,n;\, j=1,...,m}$ is an indicator matrix for which of the *m* normal distributions and individual belongs to and $P(e_{ij} = 1) = p_j$ its associated probability.

The model was estimated by a Bayesian estimation method where missing observations for the dependent variable are treated as latent variables and simulated iteratively alongside the model's parameters. Appendix 2 provides a detailed description of the estimation procedure, including the specification of the prior and posterior distributions as well as the method for simulating the missing observations.

The parameters of the mixed Gaussian error term and random effect distributions are estimated with a Bayesian method, which is integrated with and applied iteratively with the estimation of the model's other parameters (see Appendix 2). However, the final parameter values for the mixed Gaussian distributions reported here were estimated *ex post* using residuals calculated from the final parameter values in (1) to (5). These final parameter values were estimated with a different method, namely the cluster technique in the R-application MCLUST (see Fraley et al, 2012). It is a general method of fitting a mixed normal of several distributions to a sample of values – in our case, a sample of residuals from a panel model.

The model has been estimated separately for the 12 groups defined by gender and the six main educational groups. The background variables encompass a host of personal and labour market relevant characteristics such as age, marital status, number and age of children, experience, secondary school results, type of education and parents' education. A person's labour market history is represented by the number of weeks a person has received income-replacing transfers of different types in the current year and during the previous three years.

4 Data

The estimation data are a panel data set that, in principle, covers the full Danish population of 18 to 67-year old persons from 1995 to 2010, the years for which useful wage rate information is available from the Danish Salary Register (Lønregisteret). However, only a subset of the population could be used for estimation purposes. This is in part because compulsory reporting to the Salary Register does not apply to private employers with less than 10 employees, but also because the full data set is too large to be dealt with computationally given the model's size and the resulting demands on computer memory capacity.

The dependent variable is the wage rate calculated as the average wage (including pension contribution) per paid work hour for all jobs held during the year that are included in the Salary Register. The analysis is restricted to 18 to 67 year olds. Graduation restarts the clock in the sense that a person starts a new period in the year of completing an education.

Table 1

Number of persons and observations

	Full population	Share used (%)	Population Used (N)	Education Leavers (N1)	Others (N2)	Subsequent years (N2+)	All observations (NT)
Males							
Primary school	248,121	11.1	27,569	1,381	26,188	182,011	209,580
Secondary school	89,139	33.3	29,713	8,619	29,848	191,002	229,469
Vocational	587,880	4.2	24,495	2,695	21,800	188,917	213,412
Short tertiary	86,262	33.3	28,754	5,818	22,936	209,057	237,811
Medium tertiary	178,896	16.7	29,816	6,758	23,058	224,039	253,855
Long tertiary	132,764	22.2	29,503	8,243	21,260	214,508	244,498
All males	1,323,062	12.8	169,850	33,514	145,090	1,209,534	1,388,625
Females							
Primary school	401,880	8.3	33,490	1,306	32,184	224,453	257,943
Secondary school	129,591	33.3	43,197	11,508	31,689	186,054	229,251
Vocational	494,424	5.6	27,468	4,572	22,896	221,148	248,616
Short tertiary	68,331	33.3	22,777	4,572	18,205	167,109	189,886
Medium tertiary	330,510	6.7	22,034	6,777	15,257	174,058	196,092
Long tertiary	107,883	30.0	32,365	11,266	21,099	208,005	240,370
All females	1,532,619	11.8	181,331	40,001	141,330	1,180,827	1,362,158

Source: The Danish Salary Register, Denmark's Statistics (own calculations).

The Salary Register is based on compulsory reporting by employers though not all employers are covered. Apart from non-compulsive participation for private sector workplaces with less than 10 employees, the sample is affected by improved attrition rates over the initial years of the sample period. The compound result is a sample of 1,388,625 males and 1,362,158 females (*cf. table 1*). On average, 12.8 per cent of the males and 11.8 per cent of the females were used. As a result, the estimation panel represents 169,850 males (1,388,625 annual observations) and 181,331 females (1,362,158 annual observations) with each person being in the sample on average 8.2 years for males and 7.5 years for females. Missing wage rate observations due to either non-employment or non-coverage is a substantial issue. A higher representation for females is mainly a result of more females working in the public sector and more males than females working at small out-of-scope work places.

Table 2

Average hourly wage rates for males and females by education in1995 and 2010 (2010 DKR)

	Males			Females		
	1995	2010	Change	1995	2010	Change
	DKR	DKR	Per cent	DKR	DKR	Per cent
Primary school	159	187	17	141	169	19
Secondary school	200	197	-2	165	164	0
Vocational	179	218	22	156	190	21
Short tertiary	195	247	27	170	214	26
Medium tertiary	232	279	21	170	221	30
Long tertiary	280	334	19	239	289	21
All persons	192	234	22	160	203	27

Source: The Danish Salary Register, Denmark's Statistics (own calculations).

Average wage rates differ across the 12 estimation groups defined by gender and educational attainment (*table 2 and figures 1a and 1b*). As expected, there is generally a positive relationship between the average wage rate and education length. Males have higher average wages than females for all education groups. In general, there is a clear positive relationship between the average wage rate and education length and the gain from taking an education is larger for males than for females. This is particularly the case for medium length tertiary educations.

Figure1a

Male average hourly wage rates by education 1995-2010 (2010 prices)

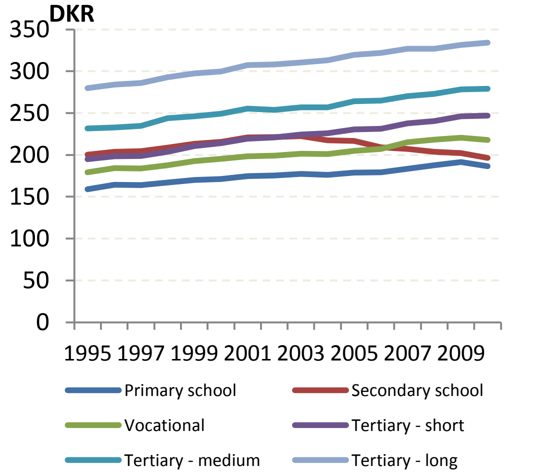
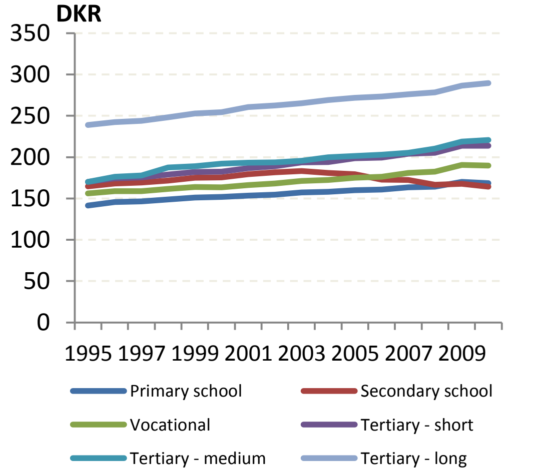


Figure 1b

Female average hourly wage rates by education 1995-2010 (2010 prices)



Source: The Danish Salary Register, Denmark's Statistics (own calculations).

The background variables such as age, gender and education originate from various administrative registers. The benefit histories are from the Ministry of Employment's DREAM database with detailed week by week recipient information. All background variables have full coverage – including the observation (persons and years) where no wage information is available.

5 Simulating wage rates

The manner in which the wage rate model is incorporated into the dynamic microsimulation SMILE relies heavily on the fact that the estimation sample is identical to SMILE's base population. Indeed, this allows us to correctly evaluate the individual specific factors in equations (1) to (8) prior to simulation start.

The simulation of future wage rates is performed as an extrapolation of the historical information obtained by the Bayesian estimation where the estimated parameters from (1) to (8) are used to evaluate the residuals and the random effect parameters for the full working aged population in SMILE's base data. The wage rate model provides *potential* wage rates for everyone regardless of whether a person is actually working or not.

More specifically, the following steps are followed as part of creating the necessary information for SMILE's starting population:

1. Evaluate initial period residuals ($\hat{\varepsilon}_{i1} = y_{i1} - x_{i1}\beta_1$) for the first year a person appears in the panel with a wage rate.
2. In the initial year of the data set, identify if a person is an education leaver (ϕ^1) or not (ϕ^2)
3. Evaluate residuals for subsequent years from (2), (3) and (4)
4. Evaluate the distribution of τ_i using its conditional posterior distribution (see Appendix 2)

   $$\tau_i \sim N\left((X^{\tau_i}{'}X^{\tau_i})^{-1}X^{\tau_i}{'}Y^{\tau_i}, (X^{\tau_i}{'}X^{\tau_i})^{-1}\right)$$

   The τ_i-values are drawn from this distribution.

5. For persons who are present in the estimation panel, but not in the final year (the initial year of the simulation) either due to non-employment or non-participation, the wage rate is updated to that year by applying the wage rate model in a forecasting mode.
6. For persons who are not in the estimation panel at all and persons who reach working age during simulations, the wage rates are simulated with (1) to (8) by drawing residuals from their estimated distributions.

Steps 3. and 4. are complicated by the fact that individuals may have intermediate years with missing information that is, they have gaps in the estimation panel with no wage rate information. For this reason, these steps are performed by applying the estimation technique described in Appendix 2 and 3, but without the random draws from the posterior distributions for the model's non-individual specific parameters $(\rho_1, \rho_2, \gamma, \beta_1, \beta_2, \phi, \sigma_{11}^2, \sigma_{12}^2, \sigma_2^2, \sigma_\tau^2)$ replaced by the estimated parameters (see Appendix 1).

6 Estimation results

This section provides an account of the results of the estimation of the model described in Section 3. The detailed parameter results are shown in Appendix 1.[5] The objective of the following outline is to provide an interpretation of key messages from the parameter results.

The model has been estimated separately for the twelve groups defined by gender and the six main groups of highest education namely primary school, secondary school, vocational, short tertiary, medium tertiary and long tertiary. This approach holds both practical and analytical advantages compared with a joined estimation. Firstly, the inter-temporal parameters $(\gamma, \rho_1, \rho_2, \phi, \sigma_\tau^2)$ and the error term variation $(\sigma_{11}^2, \sigma_{12}^2, \sigma_2^2)$ differ substantially across the groups. Accounting for these differences is more manageable with separate estimations. This is important for consistency of estimation as well as for accuracy of the simulated missing wages. Secondly, separate estimations provide group specific estimates for all the model parameters and these differences provide new and interesting insight into the wage process.

The model is estimated with wage rates that have been indexed to 2010-level for each group so as to avoid heteroscedasticity caused by wage inflation.

The impact of non-employment on wages

The impact of unemployment and other types of non-employment on wage rates is represented by separate variables for the number of weeks on unemployment benefits, social security (unemployed and not unemployed), sickness benefits, salary subsidised employment, maternity leave benefits, labour market leave benefits and other benefits for persons that are not in the labour force.[6] We focus on the duration of benefits during the three years prior to the current calendar year measured as the share if the total number of weeks with benefits.

The parameter estimates represent a (generally) negative wage rate effect of periods of non-employment. A parameter estimate of -0.15 (figure 2a) indicates that one year of unemployment

---

[5] The β_1- and the β_2-parameters in (1) and (2) are shown in table A1 and A4 respectively. To facilitate comparison across groups and with the β_1-parameter estimates, the β_2-parameters have been scaled up by group specific estimates for 1/(1-γ), *cf. table A4*. Table A2 shows estimates for the dynamic parameters and Table A3 the estimated parameters for the mixed Gaussion residuals and random effects.

[6] The benefit history variables are potentially endogenous and, as a result, we cannot say with certainty, that the negative parameters represent a causal effect. Indeed, it is well-known that the risk of unemployment is higher for persons with low incomes and it is unknown how much of this relationship is captured by the model. There are, however, a number of factors that increase our confidence that the benefit history parameters represent a causal effect of unemployment on the wage rate. Firstly, the estimations are done separately for persons with different education levels, which means, that – in terms of education – like are grouped with like and because of that, a large proportion of the correlation between unemployment risk and wage rates is accounted for. Also the inclusion of an extensive number of background variables serves to reduce the problem. Secondly, at least part of the potential endogeneity of the benefit variables is accounted for by the individual random effect (that is, the problem is that low wages and unemployment risk tend to coincide). More generally, the model's dynamic parameters contribute to ensure that the measured effect of unemployment spells is adjusted for any simultaneity effects. In analysis not reported here, we test the endogeneity assumption by including lead variables for 'benefit future' that show a much weaker, albeit non-zero, relationship with wage rates than benefit history, which suggest that endogeneity is present but not a serious problem.

reduces the hourly wage rate by 5 per cent. Looking at the parameter estimates for the different groups the following can be summarised:

- The effect of periods on unemployment benefits and social security is negative for almost all the groups and tends to be stronger for persons with a tertiary education, particularly for social security recipients, (*figure 2a, 2b and 2c*).
- The effect of unemployment (*figure 2a and 2b*) on the wage rates is somewhat larger for males than for females.
- The wage rates for persons with a tertiary education are generally affected more than persons with basic schooling or a vocational education. Females with a medium length tertiary education are an exception – presumably because this group has many public sector employees (health and education).
- The impact of subsidised employment on the wage rates is generally negative and stronger for persons with a tertiary education. However, the effect is somewhat weaker than for inactive recipients of unemployment benefits and social security – again, females with a medium tertiary education stand out with a zero effect (*figure 2d*).
- The number of weeks on sickness benefits has a strong negative impact on the subsequent wage rates, and for males the effect is strongly related to education level (*figure 2e*).
- Periods on maternity leave benefits has little or no effect on subsequent wage rates (*figure 2f*).
- The labour market leave programmes of the nineties did not have a huge impact on subsequent wages. For persons with a long tertiary education there was even a small positive effect, presumably because many participated in educational programmes (*figure 2g*).

These results will of course have implications for the incomes of individuals who experience periods of non-employment. However, the economic cost in terms of lost productivity and forgone taxes etc. would also be substantial.

Figure 2a

Parameter estimates, share of weeks on unemployment benefits previous 3 years

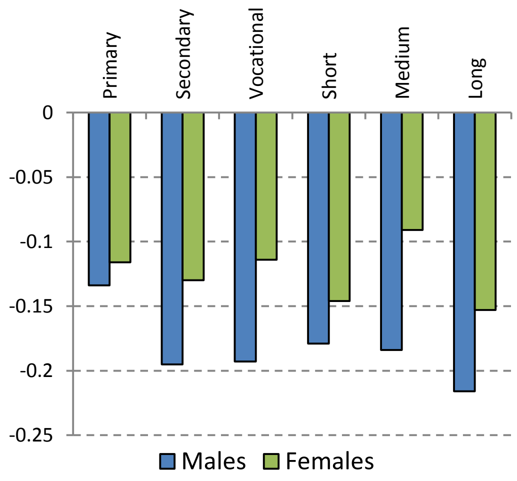


Figure 2b

Parameter estimates, share of weeks on social security (unemployed) previous 3 years

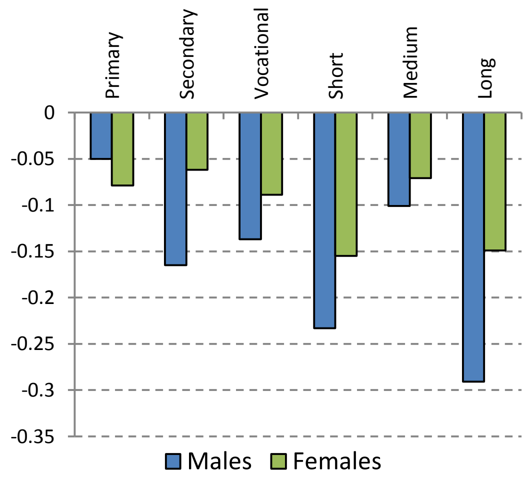


Figure 2c

Parameter estimates, share of weeks on social security (not unemployed) previous 3 years

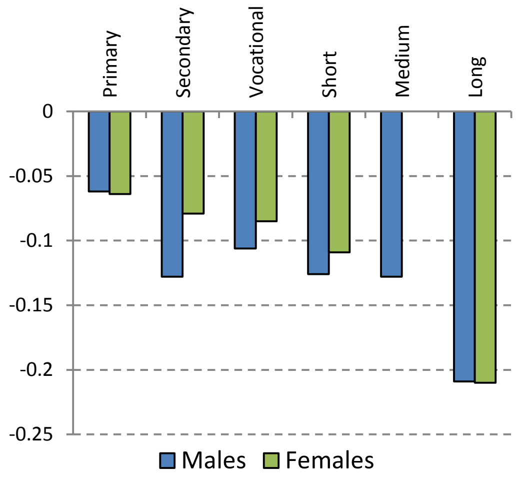


Figure 2d

Parameter estimates, share of weeks in subsidised employment previous 3 years

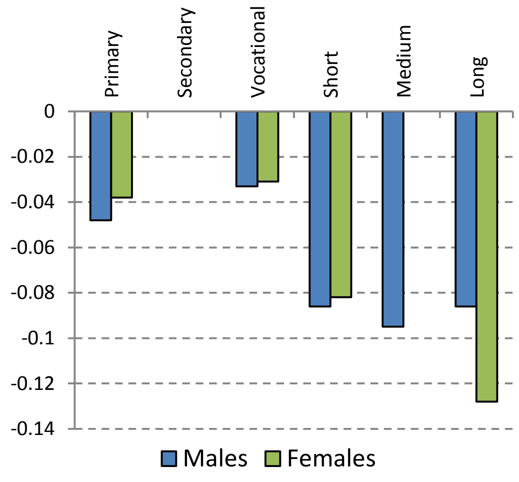


Figure 2e

Parameter estimates, share of weeks on sickness benefits, previous 3 years

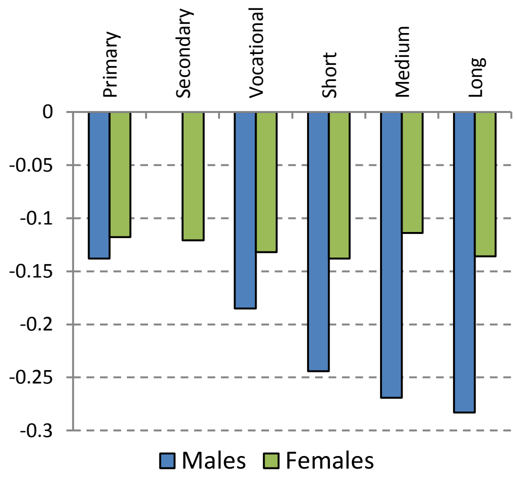


Figure 2f

Parameter estimates, share of weeks on maternity benefits, previous 3 years

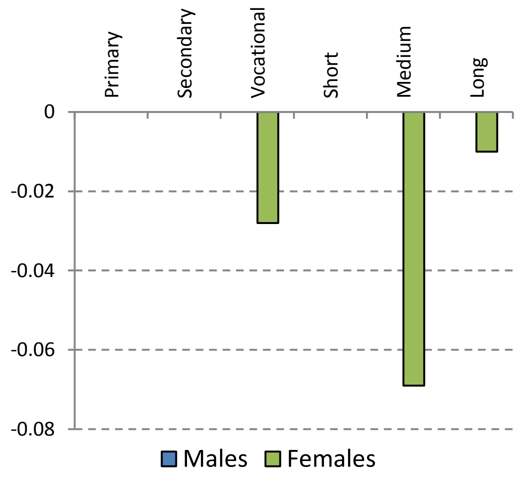


Figure 2g

Parameter estimates, share of weeks on labour market leave, previous 3 years

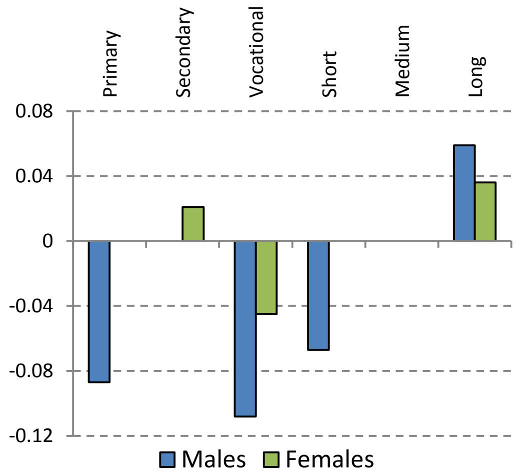
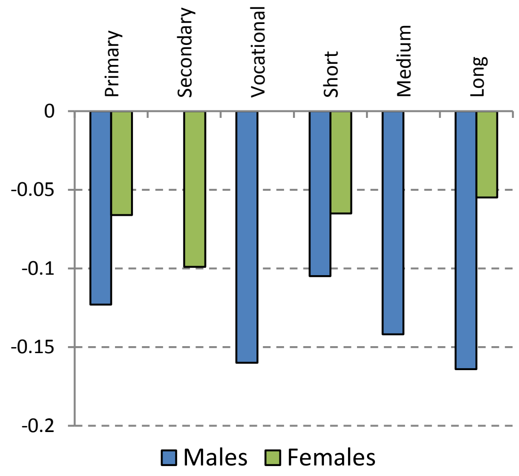


Figure 2h

Parameter estimates, share of weeks on other benefits, previous 3 years



Wage rate dynamics

The model's comprehensive dynamic specification provides interesting insight into the dynamics of the wage process.

The γ and the ρ_2 parameters represent distinctly different aspects of the wage dynamics. γ represents the inter-temporal dependency of the wage rates, which is neither explained by the observed nor the unobserved factors (through the error process). ρ_2 represents inter-temporal correlation of the unobserved factors. The results show that γ is generally lower than ρ_2 for both males and females (*figure 3a and 3b*). For males, γ is generally increasing with the education level while ρ_2 tends to be lower for tertiary educations. There is no clear education pattern for females, but is a tendency to a trade-off between the γ and the ρ_2 parameters.

The higher ρ_2-values imply that the dynamics of wage rates to a large extent are determined by inter-temporal constancy of unobserved factors.[7] In contrast, a large γ-value (males with a long tertiary education) imply a higher degree of year-to-year constancy, which is determined neither by observable nor unobservable wage determinants.

---

[7] The ρ_2 parameter and the random effects together capture the dynamics of the unobserved factors embedded in the error term. The difference is that the random effect captures a permanent effect while the ρ_2 parameter capture a transient or year-to-year effects (see *figure 5a*).

Figure 3a

Estimated γ and ρ_2-parameters, males

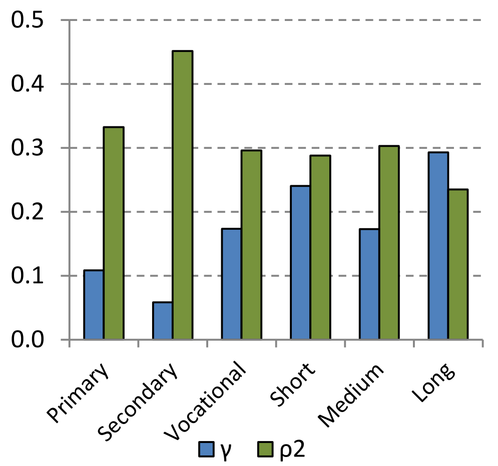
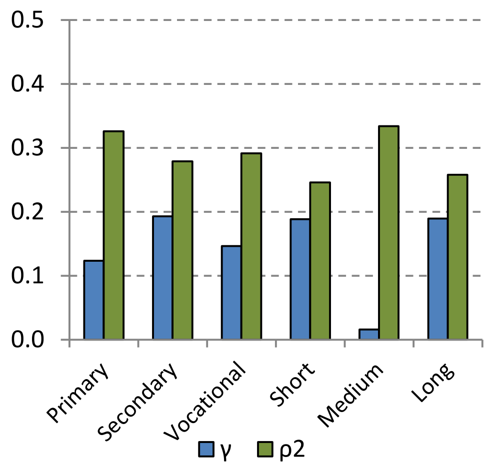


Figure 3b

Estimated γ and ρ_2-parameters, females



The ϕ_1- parameter is a permanent effect of the wage rate in a person's first job on the wage rate in subsequent years.[8] All groups exhibit this form of wage persistency, but the effect is clearly stronger for persons with a tertiary education while it is much less pronounced for persons with a primary or a vocational education (*figure 4a*). The implications are profound. An ϕ_1-parameter in the range 0.3-0.4 implies that between 30 and 40 per cent of the unexplained part of the wage rate (below or above average) in the first year after leaving school or graduating tends to stick with a person in future years.

The importance of this effect can be illustrated by looking at the standard deviation of the first year error term, σ_1 , which shows some variation across the groups. It is generally higher for males (around 0.19-0.23) than females (around 0.16-0.22) and slightly increasing with education level (*figure 5a*). As a rough guide, a starting wage rate above the average (conditional on the background variable) by one standard deviation of 0.2 and a ϕ_1- parameter of 0.3 means that the wage rate will tend to b persistently higher by around 6 per cent over and above persons with similar characteristics – due to the starting wage alone.

[8] Notice that ϕ_1 only applies to persons who enter the panel when they finish an education – ϕ_2 is the equivalent first-year effect for persons that enter the panel for other reasons.

Figure 4a

Persistency of 1st year wages – males

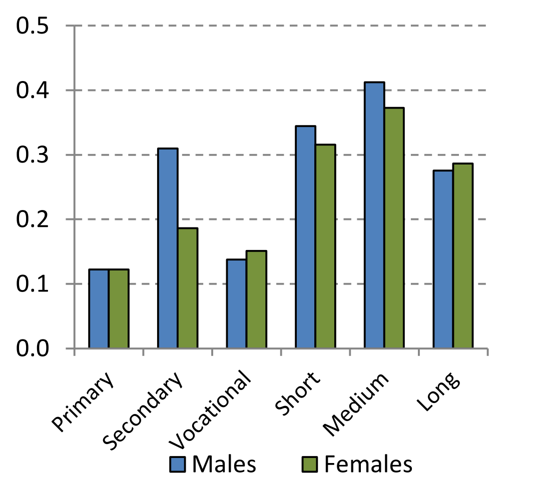


Figure 4b

Persistency of 1st year wages – females

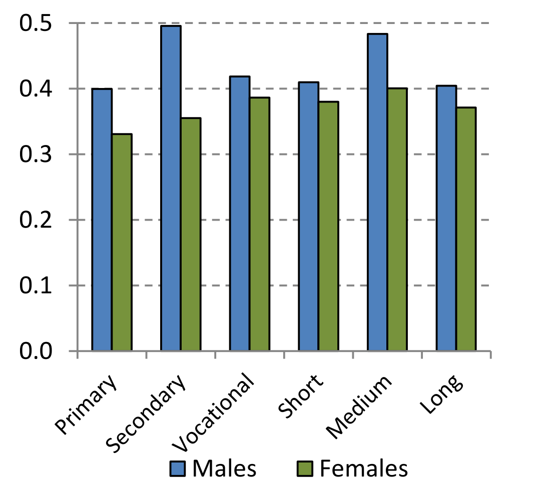


Figure 5a

Standard deviation for the error term year 1 – education leavers

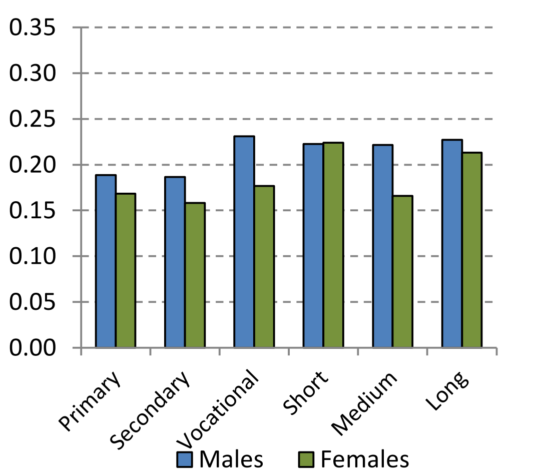
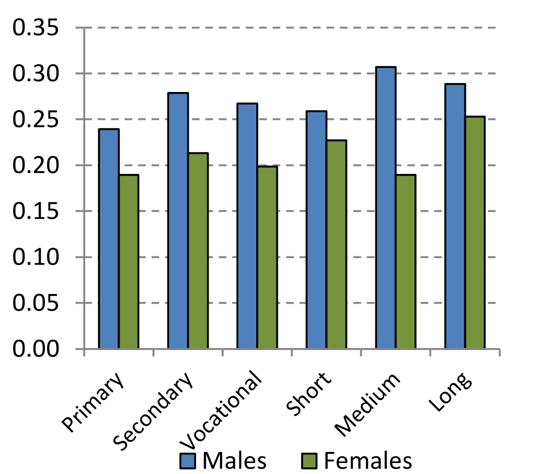


Figure 5b

Standard deviation for the error term year 1 – others



The error term can be interpreted as a wage rate risk or an uncertainty about future wages faced by individuals (that is, as observed by the analyst). The individuals themselves may of course know more (and perhaps sometimes less) than the information available to the analyst in term of data and techniques. Nevertheless, the error term measures the accuracy with which the wage rate can be predicted from year to year and, with accumulated uncertainty, further out in the future. As such, the differences in error term variation across groups have important implications for risks and uncertainties for males versus females as well as for individuals with different educations.

The first year error term standard deviation is larger for males than for females for all education groups. Indeed, the female fraction of the total variation ranges from 24 per cent for medium length tertiary educations to 40 per cent for short tertiary educations. This is not surprising, as the variation in wage rates is generally much larger for male. The message here is that observable factors cannot (fully) explain these gender differences. In other words, males may have higher wages than females, but they also face more uncertainty.

Figure 5a also exhibits substantial differences across education groups, albeit the patterns are less obvious than the gender differences. Females with a vocational or a medium length tertiary education stand out with very low error term variation – presumably due to the high rates of public sector employment in these groups. Conversely, males with a long tertiary education stand out with the largest error term variation.

The error term standard deviations for subsequent years also vary considerably across gender and education groups (*figure 6b*). Males consistently have higher variation than females and increases slightly with education level although, again, the females with a medium length tertiary education stand out at the low end.

The *individual random effect* represents another aspect of wage rate persistency. It is also sometimes referred to as a *permanent luck* effect although in the case of wage rates, this is arguably a somewhat misleading label unless one is willing to accept unobserved characteristics, such as effort and commitment as an outcome of 'luck' as much as cognitive abilities and intelligence.

The importance of the random effect as measured by its standard deviation σ_{τ}^{2} exhibits some variation across the gender and educational groups. Males generally have larger effects than females for all educational, except persons with long tertiary education who have almost equal size random effects for males and females (*figure 6a*).

Figure 6a

Standard deviation for the random effect

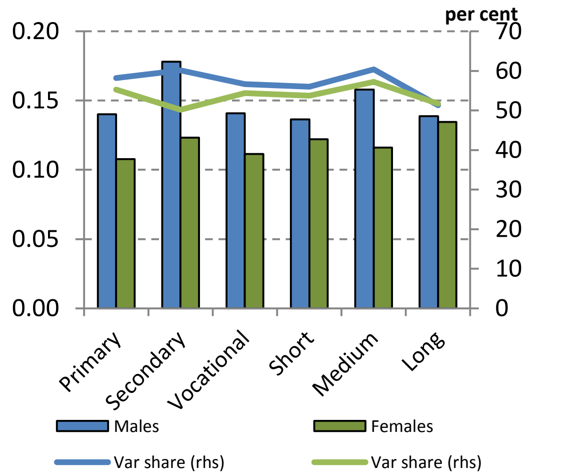
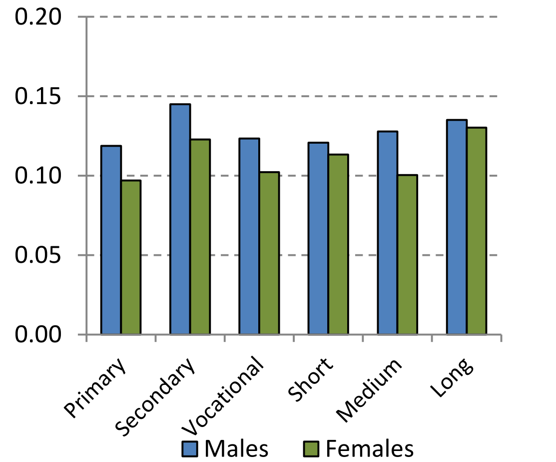


Figure 6b

Standard deviation for the error term year 2+



The random effect should also be seen in the context of its fraction of the total unexplained variation (shown as a line in *figure 6a*). This fraction is between 50 and 60 per cent for all groups and a little higher for males than for females. As a consequence, a slightly larger fraction of the variation for males can be attributed to unobserved individual characteristics.

The first year error term standard deviation is much larger than the standard deviation for the subsequent years, *cf. figure 5a, 5b and 6b*. This is a natural consequence of the model structure having a first year wage rate determined only by the background variables (1). In contrast, the wage rate in subsequent years is determined with higher accuracy because it depends on the outcome in previous years (2). The first year salary is particularly important because it works through to subsequent years both through a fading effect, the γ -parameter, and a permanent effect, the ϕ -parameter (notice that ρ_1 is close to zero for all groups).

The Gaussian error term and random effect distributions

In spite of the wage rate model's extensive explanatory power, the unexplained wage rate variation imbedded in the error terms and the random effects ((6), (7) and (8)) still represent a large proportion of the overall dispersion of the wage rates. Hence, it is important that the distributions of these terms are modelled in a realistic manner. In SMILE this is done by applying mixed Gaussian distributions – the estimated parameters for these distributions are shown in table A3 (Appendix 1).

Mixed Gaussian distributions have the ability to emulate a wide range of distribution shapes. The following figures illustrate this for males with a primary school education. The figures compare the actual distributions of the model's the random effect (*figure 7a*) and its three error terms (*figure 7b, 7c and 7d*) with the modelled distribution using either a single normal distribution of a mixture of several normal distributions.

Comparing the residuals' actual distributions with normal distributions with identical standard deviations shows that the actual distributions are strongly leptokurtic and in three cases they are also visibly skewed. By implication, the residuals are not normally distributed and applying a simple normal distribution for the simulation of wage rates would lead to a serious bias in the distribution of simulated wage rates. In contrast, the distributions obtained by applying a mixture of normal distributions are so close to the actual distributions that only a small sample size tells them apart.

Figure 7a

Random effect distribution – males with primary school (N=27,569)

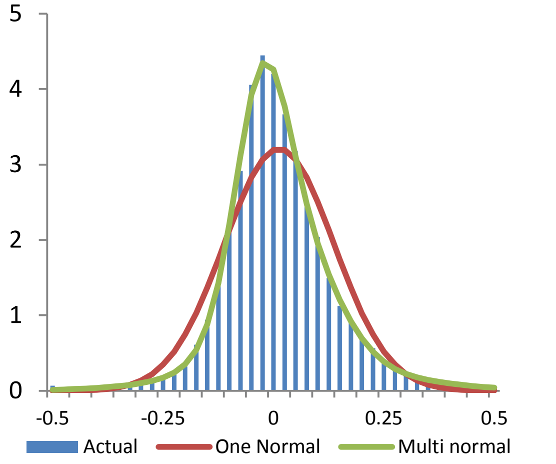
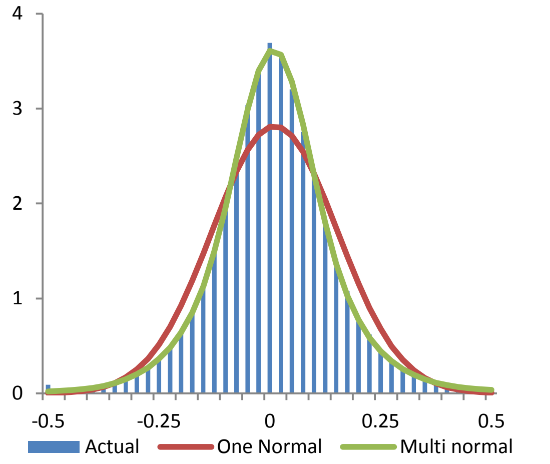


Figure 7b

Year 2+ error term distribution – males with primary school (N=182,011)



The consequences of applying a single normal distribution to the simulation of wage rates are unpredictable, but surely the distribution of incomes would be unrealistic. The bias for the random effect (*figure 7a*) and the year 2+ error term (*figure 7b*) would push the earned incomes of a large number of persons away from the area around the middle of the distribution and, at the same time, reducing the number of persons with more extreme outcomes.

Figure 7c

Year 1 error term distribution (education leavers) – males with primary school (N=1,381)

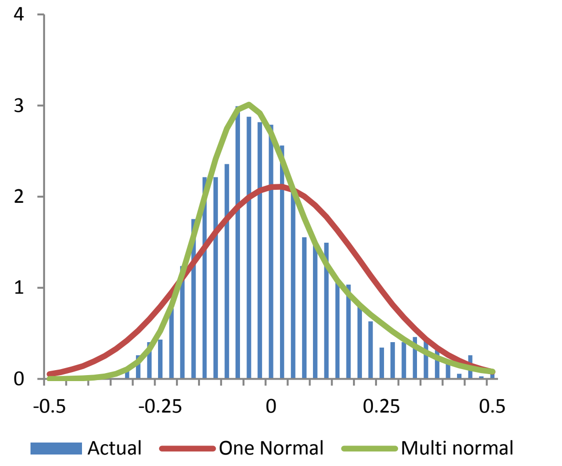
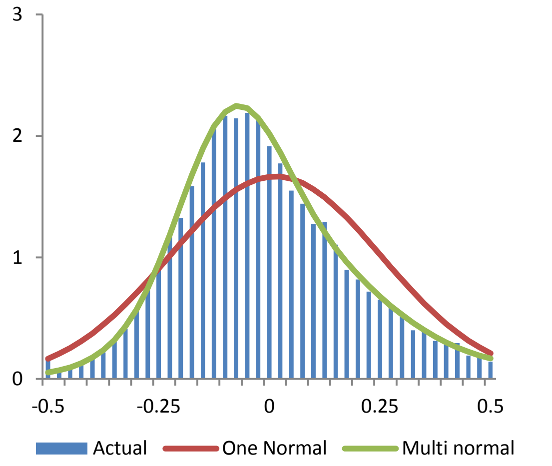


Figure 7d

Year 1 error term distribution (others) – males with primary school (N=26,188)



Failing to account for the skewed distribution of the random effect (*figure 7a*) and the first year error term (*figure 7c and 7d*) would push mass of incomes to the right and thereby cause a misleading picture of the overall income distribution.

Other background variables

A number of the results for the background variables have important interpretations that can be summarised by the following points:

- Cohabitation status affect male wages but not female wages: single males have 3 to 6 per cent lower 1st year wage rate and 1 to 2.5 per cent lower wage rate in subsequent years than cohabiting males.
- Children on affect wage rates for males with primary school education and only through 1st year effect. However, the effect is rather large and suggests that males with only primary school education have wages that start roughly 19 per cent below their fertile counterparts.
- Birthplace matters for wage rates. For most groups wages rates tend to be higher for persons from other western countries tend to have higher wage rates than Danish-born persons (by around 2 to 7 per cent) and persons from non-western countries tend to have lower wage rates than Danish-born persons.
- The parents' education has a small influence on a person's wage rate. The somewhat weak direct link between parents' education and wage rates is perhaps surprising, but the effect is conditional on a person's own education, which is accounted for by grouping. Indeed, we know from other research that a person's own educational choice is influenced by the parents' education – especially the mother's education. The important finding here is that parents' education may affect

their children's income through their choice of education, but there is only a small effect on postgraduate wages beyond educational choice.

- Secondary school grades have a substantial positive impact on the wage rates for all education groups (except of course primary school) by between 3.5 and 5.5 per cent per point.

References

Arnberg, S. og P. P. Stephensen (2013) Pension og Tilbagetrækning – Ikke-parametrisk Estimation af Heterogenitet, Danish Economic Council working Paper – http://www.dors.dk/graphics/Synkron-Library/Publikationer/Arbejdspapirer/Arbejdspapir%20om%20tilbagetrakningsmodel.pdf (in Danish).

Baekgaard, H. (2002) *Modelling the Dynamics of the Distribution of Earned Income*, Technical Paper no. 24, National Centre for Social and Economic Modelling, University of Canberra.

Bækgaard, H. (2010) *Work Incentives and Latent Wages for the Unemployed: Estimating a Dynamic Panel Model for Wage Rates with Missing Observations*, paper presented at the 9th Nordic Microsimulation Seminar in Copenhagen 16th and 17th June 2010.

Bækgaard, H. (2014a) *A Survival Approach to Modelling Labour Market Transitions*, DREAM report (forthcoming, will be available from www.dreammodel.dk/SMILE).

Bækgaard, H. (2014b) *A Dynamic Model for Worked Hours for Employees*, DREAM report (forthcoming, will be available from www.dreammodel.dk/SMILE)

Browning M., M. Ejrnæs and J. Alvarez (2006) *Modelling income processes with lots of heterogeneity*, Discussion Paper no. 285 October 2006, Dept. of Economics, University of Oxford.

Fraley, C., A. E. Raftery, T. B. Murphy, L. Scrucca (2012) *Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation (mclust Version 4 for R)*, Technical Report No. 597, Department of Statistics, University of Washington, June 2012.

Galler, HP 1996, '*Microsimulation of pension reform proposals: modelling the earnings of couples*, in Harding, A (ed.), Microsimulation and Public Policy, North-Holland Amsterdam.

Geweke J., M. Keane (2000) *An empirical analysis of earnings dynamics among men in the PSID: 1968-1989*, Journal of Econometrics, vol. 96, pp. 293-356.

Heckman, J. J. (1979), *Sample Selection Bias as a Specification Error*, Econometrica 47(1), pp. 153 - 161.

Appendix 1 Estimation results

The following tables show the parameter results of the estimation of the model (1) - (5). Many background variables are self-explanatory. Nevertheless, the detailed descriptions below provide more precise definitions as a reference guide.

Dummy_1995 - Dummy_2010: year-dummies

Age: a persons age 1st January

Age squared: 'Age' squared divided by 100

Children present: 1, if there are children aged 0-17 in the family

Single: 1, if a person lives in a one-adult family

COB west: 1, if the person was born in a western country other than Denmark

COB non-west: 1, if the person was born in a non-western country

COB Nordic: 1, if the person was born in a Nordic country other than Denmark

COB w Euro: 1, if the person was born in a west European country other the Nordic countries

COB other Euro: 1, if the person in born in a other European countries

COB Africa: 1, if the person in born in Africa

COB Lat. America: 1, if the person was born in born in Latin America

COB e. Asia dev.: 1, if the person was born in east Asia (more developed)

COB e. Asia n dev.: 1, if the person was born in east Asia (less developed)

COB w. Asia.: 1, if the person was born in west Asia (less developed)

YSA non-west: the number of years since arrival (persons born in a non-western country)

YSA west: the number of years since arrival (persons born in a western country)

Desc. west: 1, if the person is descendant from a western country

Desc. non- west: 1, if the person is descendant from a non-western country

Att. Ed. primary: 1, if the person is attending primary school

Att. Ed. secondary: 1, if the person is attending secondary school

Att. Ed. vocational: 1, if the person is attending a vocational education

Att. Ed. tert. short: 1, if the person is attending a short tertiary education

Att. Ed. tert. Med.: 1, if the person is attending a medium tertiary education

Att. Ed. tert. long: 1, if the person is attending a long tertiary education

Dad's educ 1: 1, if the father's highest education is primary school

Dad's educ 2: 1, if the father's highest education is secondary school

Dad's educ 3: 1, if the father's highest education is vocational

Dad's educ 4: 1, if the father's highest education is short tertiary

Dad's educ 5: 1, if the father's highest education is medium tertiary

Dad's educ 6: 1, if the father's highest education is long tertiary

Mum's educ 1: 1, if the mother's highest education is primary school

Mum's educ 2: 1, if the mother's highest education is secondary school

Mum's educ 3: 1, if the mother's highest education is vocational

Mum's educ 4: 1, if the mother's highest education is short tertiary

Mum's educ 5: 1, if the mother's highest education is medium tertiary

Mum's educ 6: 1, if the mother's highest education is long tertiary

Experience: number of years employment experience

Experience, square root: square root of 'Experience'

Graduate: 1, if first year in employment after completing highest education

First job: 1, if first employment after completing highest education

Secondary grade: average grade from secondary school

Secondary grade NA: 1, if the person completed secondary school but has no grade is available

1, UE Benefits: =1, if having received unemployment benefits in the current year

1, Soc. Sec. UE: =1, if having received social security as unemployed in the current year

1, Soc. Sec. nUE: =1, if having received social security as not unemployed in the current year

1, subs. empl.: =1, if having worked in a job with salary subsidy in the current year

1, sickness benefits: =1, if having received sickness benefits in the current year

1, maternity leave: =1, if having received maternity leave benefits in the current year

1, other leave: =1, if having received other leave benefits in the current year

1, not in lab. force: =1, if having been out of the labour force without benefits in the current year

Wks UE Benefits: the share of the number of weeks if having received unemployment benefits in the previous three years

Wks S. Sec: UE the share of the number of weeks if having received social security as unemployed in the previous three years

Wks S. Sec: nUE the share of the number of weeks if having received social security as not unemployed in the previous three years

Wks subs. empl.: the share of the number of weeks worked in a job with salary subsidy in the previous three years

Wks disab. ben..: the share of the number of weeks having received disability benefits in the previous three years

Wks mater. leave.: the share of the number of weeks having received maternity leave benefits in the previous three years

Wks other leave.: the share of the number of weeks having received other leave benefits in the previous three years

Wks not in lab f.: the share of the number of weeks having been out of the labour force without benefits in the previous three years

Table A1

Estimation results for gender/education groups – first year

	Primary school		Secondary school		Vocational		Short tertiary		Medium tertiary		Long tertiary	
	Males	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**
Constant	4.805	4.794	3.979	3.984	4.804	4.715	4.539	4.304	4.701	4.9	4.333	4.428
Dummy_1995	-	-	-	-	-	-	-	-	-	-	-	0.100
Dummy_1996	-0.012	-0.009	-	-0.014	**-0.007	**-0.005	-	**-0.006	0.016	**0.004	-	0.080
Dummy_1997	-0.029	-0.031	-	-0.032	-0.042	-0.026	-	-0.037	**-0.002	**-0.001	-	0.059
Dummy_1998	-0.035	-0.042	-	-0.043	-0.037	-0.037	-	-0.031	**0.006	-0.018	-	0.035
Dummy_1999	-0.043	-0.048	-	-0.045	-0.048	-0.047	-	-0.048	**0.007	-0.014	-	0.044
Dummy_2000	-0.036	-0.056	-	-0.072	-0.043	-0.039	-	-0.044	**0.009	**-0.004	-	0.040
Dummy_2001	-0.034	-0.052	-	-0.095	-0.058	-0.073	-	-0.049	-0.023	-0.049	-	0.043
Dummy_2002	-0.069	-0.065	-	-0.113	-0.070	-0.077	-	-0.076	-0.050	-0.057	-	-
Dummy_2003	-0.063	-0.058	-	-0.102	-0.052	-0.074	-	0.025	-0.041	-0.056	-	-
Dummy_2004	-0.065	-0.074	-	-0.095	-0.069	-0.08	-0.046	**-0.013	-0.051	-0.082	-	-
Dummy_2005	-0.056	-0.085	-	-0.070	-0.051	-0.095	-0.077	-0.053	-0.052	-0.071	-	-
Dummy_2006	-0.132	-0.105	-	-0.032	**-0.041	-0.127	-0.093	-0.058	-0.040	-0.126	-	-
Dummy_2007	-0.146	-0.119	-	**0.002	**-0.057	-0.117	-0.090	-0.053	-0.040	-0.134	-	-
Dummy_2008	-0.156	-0.118	0.047	0.032	**-0.07	-0.13	-0.097	-0.081	-0.035	-0.147	-	-
Dummy_2009	-0.200	-0.162	0.049	0.034	**-0.06	-0.137	-0.107	-0.083	-0.036	-0.145	-	-
Dummy_2010	-0.212	-0.161	0.042	0.022	-0.083	-0.162	-0.136	-0.11	-0.079	-0.167	-	-
Age	0.019	0.012	0.052	0.038	0.022	0.015	0.034	0.028	0.028	0.016	0.047	0.032
Age squared	0.000	0.000	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Children present	0.191	**0.013	**0.095	**0.063	**-0.035	**-0.023	**0.037	**-0.062	**0.046	**0.015	**-0.030	**0.009
Single	-0.040	0.008	-0.058	0.009	-0.034	**0.005	-0.040	**0.003	-0.050	**-0.003	-0.032	-0.012
COB west	-	-	**0.003	-	0.045	-	-	-	-	0.040	-	-
COB non-west	-	-	-0.054	-	-0.037	-	-0.091	-	-0.06	-0.083	-	-0.040
COB Nordic	-	0.041	-	0.086	-	0.063	-	**0.018	0.066	-	-	-
COB w. Euro	-	0.036	-	0.043	-	0.071	-	0.055	**0.009	-	-0.044	-

COB other Euro	-	-	-	-	-	-	-	-	-	-	-0.089	-
COB Africa	-	-	-	-	-	-	-	-	-	-	-0.280	-0.095
COB Lat. America	-	-	-	-	-	-	0.092	-	-	-	-0.070	-
COB e. Asia dev.	-	**0.083	-	-	-	-	-	-	-	-	**-0.021	-
COB w. Asia	-	-	-	-	-	-	-	-	-	-	-0.059	-
COB NA	-	-	-	-	-	-	-	-	-	-	-0.243	-
YSA non-west	-	0.001	-	-	-	0.002	-	-	-	0.004	-	**0.002
Desc. west	**0.024	-	-	-	-	-	-	-	-	-	-	-
Desc. non-west	-0.015	-	-	-	-	-	-	-	-	-	**-0.011	-
Att. ed. primary	-	-0.040	-0.057	-	**0.016	-0.052	-	-	**-0.012	-0.091	0.018	-
Att. ed. Secondary	-	-0.076	-0.094	-	**-0.044	-0.211	-	-	**-0.079	-	**-0.010	-
Att. ed. tert. short	-	-	-0.087	-0.046	**-0.068	**-0.028	-0.124	-0.162	-0.249	-0.207	-0.119	-0.283
Att. ed. tert. med.	-	-	-0.097	-0.046	**-0.055	-0.119	-0.139	-0.114	-0.175	-0.166	-0.119	-0.142
Att. ed. tert. long	-	**0.036	-0.113	-0.046	**-0.024	-	-0.087	**-0.035	-0.243	-0.207	-0.119	-0.049
Dad’s educ. NA	-	-	-	-	**0.013	**-0.009	-	-	**-0.010	-	0.015	-
Dad’s educ. 1	-	-	-	-	-	-0.015	-0.018	-	-	-	-	-
Dad’s educ. 2	-	-	-	0.013	-	-	-	-	-	**0.009	-	-
Dad’s educ. 3				0.013	0.014				0.017			
Dad’s educ. 4				0.013	0.014				0.017			
Dad’s educ. 5	-	-	-	0.013	0.014	-	-	-	0.017	0.021	-	-
Dad’s educ. 6	-	-	-	0.013	0.049	-	-	-	0.017	**0.006	0.023	0.011
Mum’s educ. NA	-	-	-	-	-0.022	-0.029	-	-	-	-	**0.006	-
Mum’s educ. 1	-	-	-	-	-0.026	-0.017	-0.019	-	-	-	-	-
Mum’s educ. 2	-	-	-	0.015	-	-	-	-	-	**0.022	-	-
Mum’s educ. 3	-	-	-	0.015	-	-	-	-	-	0.009	-	-
Mum’s educ. 4	-	-	-	0.015	-	-	-	-	-	**0.016	**-0.002	**0.007
Mum’s educ. 5	-	-	-	0.015	-	-	-	-	-	0.014	**-0.002	**0.007
Mum’s educ. 6	-	-	-	0.015	-	-	-	-	-	**0.009	**-0.002	0.020
Experience	0.004	0.009	0.017	0.013	0.011	0.011	0.005	-	0.008	0.007	0.010	0.005
Exp. square root	0.017	-	-0.022	-0.014	-0.034	-0.014	**0.01	0.056	**-0.006	**-0.008	-0.016	0.026
Graduate	**-0.060	-	**-0.028	-	**-0.092	-0.040	**-0.039	-	**-0.049	-0.025	**-0.076	-0.050
First job	**-0.047	-0.035	**-0.013	-0.017	**-0.016	**-0.007	**0.000	**-0.001	**-0.043	**0.003	**-0.040	**0.000

Second. grade NA	-	-	0.105	0.176	**0.046	0.173	0.129	0.237	0.134	0.120	0.205	0.273
Secondary grade	-	-	0.015	0.027	**0.015	0.030	0.02	0.035	0.020	0.020	0.023	0.031
1, UE benefits	-0.033	-0.035	-0.073	-0.035	-0.037	-0.046	-0.092	-0.074	-0.055	-0.020	-0.121	-0.089
1, Soc. Sec. UE	-0.019	-0.012	-0.024	-0.012	**-0.004	-0.032	-0.061	-0.043	-0.072	**-0.003	-0.120	-0.080
1, Soc. Sec. nUE	-0.027	-0.023	-0.062	-0.021	-0.062	-0.037	-0.067	-0.048	-0.029	-	-0.045	-0.038
1, subs. empl.	-0.076	-0.047	-0.052	-0.045	-0.039	-0.024	-	-	**0.006	-	**0.009	**-0.033
1, sickness benefits	-0.015	-	-0.045	-0.016	**-0.003	-	-0.056	-0.025	-0.059	**-0.005	-0.049	-0.011
1, maternity leave	**0.023	0.032	**0.025	0.024	-	**0.009	-	-	-	**-0.012	**-0.004	-0.020
1, other leave	-0.049	**-0.011	-0.056	-0.041	-0.047	-0.030	-0.028	-0.034	-0.144	-0.036	-0.103	-0.086
1, not in lab. force	**-0.023	**-0.005	-0.113	-	-0.056	**-0.009	**-0.045	**-0.035	-0.075	**-0.032	**-0.068	-0.137
Wks, UE benefits	-0.123	-0.089	-0.218	-0.144	-0.194	-0.118	-0.182	-0.134	-0.240	-0.101	-0.346	-0.219
Wks, S. sec. UE	-0.064	-0.069	-0.234	-0.207	-0.231	-0.068	-0.275	-0.202	-0.322	-0.233	-0.459	-0.453
Wks, S. Sec. nUE	-0.061	-0.036	-0.238	-0.144	-0.160	-0.047	-0.218	-0.133	-0.402	-0.222	-0.587	-0.399
Wks, subs. empl.	**-0.048	**0.025	-0.172	-	**-0.010	-	-0.068	-0.081	-0.128	-0.053	-0.199	**-0.044
Wks, sickness. ben.	**-0.064	**-0.033	-0.329	-0.170	-0.164	-0.081	-0.348	-0.169	-0.542	-0.149	-0.472	**-0.062
Wks, mater. leave	**0.393	-	-	-	-	-	-	0.103	-	**-0.013	**0.210	-
Wks, other leave	-0.136	-0.044	-0.284	-0.152	-0.255	-0.105	-0.136	-0.125	-0.306	-0.145	-0.179	-0.127
Wks, not in lab. f.	-0.131	-0.092	-0.162	-0.248	-0.202	-0.086	-0.266	**-0.093	-0.172	-0.104	-0.280	-

Note: ** indicates that the estimate is not significant at a 5% level.

Table A2

Estimation results for gender/education groups – dynamic parameters

	Primary school		Secondary school		Vocational		Short tertiary		Medium tertiary		Long tertiary	
	Males	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**
γ	0.109 (0.004)	0.123 (0.003)	0.059 (0.002)	0.193 (0.002)	0.173 (0.004)	0.146 (0.002)	0.241 (0.005)	0.189 (0.003)	0.173 (0.006)	0.016 (0.005)	0.293 (0.003)	0.190 (0.007)
ϕ_1 (Graduates)	0.122 (0.028)	0.123 (0.025)	0.310 (0.012)	0.186 (0.009)	0.138 (0.014)	0.151 (0.011)	0.345 (0.009)	0.316 (0.009)	0.412 (0.011)	0.373 (0.010)	0.276 (0.008)	0.286 (0.008)
ϕ_2 (Others)	0.400 (0.005)	0.331 (0.004)	0.496 (0.005)	0.355 (0.004)	0.418 (0.005)	0.386 (0.004)	0.410 (0.005)	0.380 (0.005)	0.483 (0.005)	0.401 (0.006)	0.404 (0.005)	0.371 (0.005)
ρ_1	0.014 (0.005)	0.003 (0.005)	-0.003 (0.005)	-0.017 (0.005)	0.002 (0.004)	0.002 (0.004)	-0.026 (0.005)	-0.013 (0.005)	-0.010 (0.004)	0.003 (0.006)	-0.038 (0.004)	-0.017 (0.006)
ρ_2	0.333 (0.007)	0.326 (0.006)	0.451 (0.008)	0.279 (0.007)	0.296 (0.006)	0.291 (0.006)	0.288 (0.007)	0.246 (0.007)	0.303 (0.006)	0.334 (0.008)	0.235 (0.007)	0.258 (0.007)
σ_τ	0.036 (0.001)	0.028 (0.001)	0.035 (0.001)	0.025 (0.000)	0.053 (0.001)	0.031 (0.001)	0.050 (0.001)	0.050 (0.001)	0.049 (0.001)	0.027 (0.000)	0.052 (0.001)	0.045 (0.001)
σ_{11} (Graduates)	0.057 (0.001)	0.036 (0.000)	0.078 (0.001)	0.045 (0.000)	0.071 (0.001)	0.039 (0.000)	0.067 (0.001)	0.052 (0.001)	0.094 (0.001)	0.036 (0.000)	0.083 (0.001)	0.064 (0.001)
σ_{12} (Others)	0.014 (0.000)	0.009 (0.000)	0.021 (0.000)	0.015 (0.000)	0.015 (0.000)	0.010 (0.000)	0.015 (0.000)	0.013 (0.000)	0.016 (0.000)	0.010 (0.000)	0.018 (0.000)	0.017 (0.000)
σ_2	0.020 (0.000)	0.012 (0.000)	0.032 (0.000)	0.015 (0.000)	0.020 (0.000)	0.012 (0.000)	0.019 (0.000)	0.015 (0.000)	0.025 (0.000)	0.013 (0.000)	0.019 (0.000)	0.018 (0.000)

Table A3

Estimation results for gender/education groups – mixed Gaussian distribution parameters

	Primary school		Secondary school		Vocational		Short tertiary		Medium tertiary		Long tertiary	
	Males	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**
ε_{11} (first year error term – education leavers)												
α^1_{11}	-0.079	-0.077	-0.143	-0.111	-0.317	-0.192	-0.015	-0.028	-0.022	-0.038	-0.122	-0.154
α^1_{12}	0.101	0.046	-0.023	-0.002	0.010	0.012	0.261	0.318	-0.018	-0.020	0.019	0.037
α^1_{13}	0.568	0.429	0.009	-0.004	0.366	0.065			0.277	0.022	0.132	0.139
α^1_{14}			0.212	0.287		0.378				0.271		
σ^1_{11}	0.010	0.006	0.004	0.003	0.006	0.010	0.034	0.038	0.054	0.038	0.064	0.052
σ^1_{12}	0.023	0.014	0.003	0.002	0.038	0.006	0.149	0.105	0.019	0.002	0.018	0.017
σ^1_{13}	0.089	0.078	0.018	0.015	0.100	0.021			0.142	0.011	0.129	0.109
σ^1_{14}			0.077	0.069		0.118				0.113		
p^1_{11}	0.641	0.532	0.251	0.240	0.080	0.207	0.918	0.930	0.517	0.418	0.254	0.253
p^1_{12}	0.324	0.408	0.231	0.253	0.871	0.340	0.082	0.070	0.426	0.141	0.607	0.632
p^1_{13}	0.035	0.061	0.340	0.407	0.049	0.418			0.058	0.408	0.139	0.115
p^1_{14}			0.177	0.100		0.034				0.032		
ε_{21} (first year error term – others)												
α^1_{21}	-0.104	-0.126	-0.095	-0.098	-0.115	-0.157	-0.035	-0.169	-0.143	-0.032	-0.184	-0.020
α^1_{22}	0.011	-0.017	-0.018	-0.045	0.025	-0.021	-0.039	-0.047	-0.231	0.002	-0.111	0.014
α^1_{23}	0.385	0.057	0.330	0.077	0.408	0.075	0.387	0.089	0.034	0.292	0.088	0.867
α^1_{24}		0.409		0.421		0.371		0.320	0.243		0.200	
σ^1_{21}	0.013	0.013	0.013	0.017	0.025	0.018	0.055	0.029	0.073	0.042	0.093	0.088
σ^1_{22}	0.044	0.007	0.066	0.009	0.060	0.009	0.010	0.012	0.006	0.006	0.011	0.014
σ^1_{23}	0.136	0.031	0.161	0.058	0.147	0.029	0.127	0.033	0.036	0.139	0.017	0.049
σ^1_{24}		0.120		0.167		0.117		0.127	0.144		0.113	
p^1_{21}	0.369	0.274	0.324	0.254	0.397	0.258	0.745	0.268	0.237	0.507	0.249	0.599
p^1_{22}	0.544	0.328	0.552	0.358	0.524	0.302	0.170	0.278	0.146	0.442	0.258	0.394
p^1_{23}	0.087	0.350	0.125	0.353	0.079	0.391	0.085	0.381	0.373	0.051	0.232	0.006

p_{24}^{1}		0.048		0.035		0.049		0.073	0.245		0.261	
ε_2 (subsequent year error term)												
α_{21}	-0.006	-0.005	-0.017	-0.012	0.002	-0.008	-0.006	-0.005	-0.023	-0.035	-0.041	-0.044
α_{22}	-0.004	-0.025	-0.035	-0.020	-0.002	-0.023	-0.003	-0.025	-0.005	0.000	-0.005	-0.003
α_{23}	0.045	0.029	0.052	0.038	0.014	0.028	0.008	0.035	0.008	0.010	0.017	0.015
α_{24}		0.069	0.015	0.028		0.022		0.009				
σ_{21}	0.021	0.014	0.034	0.021	0.021	0.015	0.081	0.017	0.088	0.054	0.101	0.085
σ_{22}	0.006	0.003	0.009	0.004	0.006	0.002	0.005	0.003	0.005	0.003	0.005	0.005
σ_{23}	0.089	0.003	0.016	0.008	0.088	0.003	0.019	0.004	0.018	0.010	0.021	0.018
σ_{24}		0.087	0.095	0.088		0.077		0.080				
p_{21}	0.520	0.499	0.329	0.461	0.478	0.470	0.110	0.432	0.132	0.121	0.097	0.112
p_{22}	0.407	0.222	0.275	0.237	0.429	0.230	0.502	0.243	0.435	0.394	0.448	0.392
p_{23}	0.073	0.236	0.250	0.225	0.093	0.237	0.388	0.241	0.433	0.484	0.455	0.496
p_{24}		0.043	0.145	0.078		0.064		0.084				
τ (the random effect)												
$\alpha_{\tau 1}$	0.014	0.005	0.010	-0.046	0.023	0.008	0.017	0.014	0.022	-0.012	-0.012	0.005
$\alpha_{\tau 2}$	-0.040	-0.020	-0.038	-0.015	-0.038	-0.027	-0.042	-0.032	-0.037	0.043	0.010	-0.046
$\alpha_{\tau 3}$	0.041	0.036	0.045	0.046	0.028	0.033	0.032	0.030	0.036		0.019	-0.003
$\alpha_{\tau 4}$				0.044								0.039
$\sigma_{\tau 1}$	0.010	0.007	0.018	0.004	0.012	0.007	0.010	0.008	0.019	0.005	0.009	0.036
$\sigma_{\tau 2}$	0.003	0.002	0.005	0.001	0.004	0.002	0.004	0.004	0.005	0.029	0.108	0.006
$\sigma_{\tau 3}$	0.047	0.035	0.068	0.006	0.046	0.032	0.038	0.033	0.075		0.035	0.003
$\sigma_{\tau 4}$				0.030								0.012
$p_{\tau 1}$	0.471	0.503	0.486	0.364	0.441	0.496	0.446	0.452	0.510	0.775	0.792	0.184
$p_{\tau 2}$	0.348	0.364	0.338	0.200	0.384	0.346	0.338	0.365	0.382	0.225	0.022	0.279
$p_{\tau 3}$	0.181	0.133	0.176	0.252	0.175	0.158	0.216	0.183	0.108		0.186	0.213
$p_{\tau 4}$				0.184								0.324

Table A4

Estimation results for gender/education groups – subsequent years

	Primary school		Secondary school		Vocational		Short tertiary		Medium tertiary		Long tertiary	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
Constant	4.766	4.740	3.633	3.832	4.694	4.628	4.420	4.673	4.258	4.664	3.998	4.216
Dummy_1997	0.008	0.007	-	0.006	0.015	0.007	-	**0.004	-	**0.000	-	-
Dummy_1998	**0.003	**0.003	-	**0.000	0.006	0.005	-	**-0.003	-	**-0.001	-	-
Dummy_1999	-0.009	-0.005	-	-0.013	-0.006	**-0.001	-	-0.008	-	-0.007	-	**0.002
Dummy_2000	-0.008	-0.007	-	-0.017	-0.016	**0.002	-	-0.009	-	-0.009	-	**-0.002
Dummy_2001	-0.006	-0.010	-	-0.032	-0.018	-0.011	-	-0.016	-0.013	-0.019	-	-0.013
Dummy_2002	-0.009	-0.016	-	-0.034	-0.021	-0.017	-0.004	-0.016	-0.008	-0.024	-	-0.015
Dummy_2003	-0.012	-0.024	-	-0.036	-0.025	-0.027	-0.013	-0.033	-0.011	-0.026	-	-0.018
Dummy_2004	-0.007	-0.022	-	-0.025	-0.020	-0.022	-0.009	-0.025	**0.002	-0.023	-	-0.017
Dummy_2005	-0.007	-0.025	-	**-0.004	-0.029	-0.032	-0.013	-0.029	-0.013	-0.023	-	-0.017
Dummy_2006	-0.083	-0.059	0.054	0.049	-0.091	-0.120	-0.051	-0.107	-0.020	-0.069	-	-0.059
Dummy_2007	-0.081	-0.058	0.091	0.062	-0.102	-0.128	-0.056	-0.109	-0.019	-0.070	-	-0.057
Dummy_2008	-0.077	-0.053	0.119	0.103	-0.092	-0.134	-0.055	-0.106	-0.012	-0.074	-	-0.049
Dummy_2009	-0.087	-0.062	0.129	0.107	-0.095	-0.135	-0.060	-0.117	**-0.002	-0.070	-	-0.056
Dummy_2010	-0.076	-0.056	0.140	0.115	-0.095	-0.138	-0.061	-0.115	**-0.004	-0.071	-	-0.064
Age	0.015	0.014	0.049	0.027	0.017	0.009	0.035	0.013	0.041	0.017	0.054	0.037
Age squared	0.000	0.000	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000
Children present	**0.001	**-0.001	**0.004	**0.008	0.009	**0.002	0.010	**-0.003	**0.000	**0.006	**-0.004	**-0.002
Single	-0.011	**0.002	-0.024	**-0.002	-0.012	0.003	-0.014	**0.001	-0.017	**0.000	-0.021	-0.007
COB west	-	-	-	-	0.074	0.081	-	0.044	-	0.047	-	-
COB non-west	**0.002	-	-0.045	-	-0.041	0.028	-0.112	**-0.008	-0.071	-0.024	-	-
COB Nordic	0.052	0.052	0.050	0.076	-	-	**0.015	-	0.075	-	0.047	-
COB w. Euro	0.029	0.057	0.057	0.058	-	-	0.018	-	0.028	-	0.029	0.036

COB other Euro	-	**0.004	-	-	-	-	-	-	-	-	-0.050	-
COB Africa	-	-	-	-	-	-	-	-	-	-	-0.246	-0.135
COB Lat. America	-	**0.034	-	-	-	-	0.069	-	-	-	**-0.017	**-0.027
COB e. Asia n.dev.	-	-	-	-	**0.019	-	-	-	-	-	-0.190	**-0.025
COB e. Asia dev.	-	0.169	-	-	-0.230	-	-	-	-	-	**-0.046	-
COB w. Asia	-	-	-	-	-	-	-	-	-	-	-0.033	-
COB NA	-	-	-	-	-	-	-	-	-	-	**-0.082	-
YSA west	**0.000	**0.000	-0.002	-	-	0.001	0.001	-	0.001	-	-0.002	**0.000
YSA non-west	**0.000	-	-0.001	-0.002	-	-	0.001	-	0.001	-	-0.002	-0.003
Desc. non-west	-0.011	-	-	-	-	-	-	-	-	-	-	-
Desc. w. Euro	-	-	-	**-0.013	**-0.029	-	-	-	-	-	**0.018	-
Desc. other Euro	-	-	-	-	-	-	-	-	-	**-0.056	**-0.080	-
Desc. Africa	-	-	-	-	-	-	-	-	-	-	0.162	-
Att. educ. primary	-0.089	-0.040	-	-	-0.068	-0.094	-0.057	-0.087	-0.024	-0.048	-0.022	-0.057
Att. ed. Secondary	-0.090	-0.085	-	-	-0.154	-0.143	-0.100	-	-0.079	-	-0.083	-0.048
Att. ed. Vocational	-0.130	-0.099	-0.131	-0.109	-0.154	-0.102	-0.216	-0.163	-0.225	-0.094	-0.202	-0.222
Att. ed. tert. short	**-0.031	-0.081	-0.034	-0.041	-0.100	-0.089	-0.100	-0.117	-0.108	-0.088	-0.087	-0.170
Att. ed. tert. med.	-0.067	-0.059	-0.039	-0.041	-0.092	-0.090	-0.106	-0.118	-0.111	-0.072	-0.096	-0.093
Att. ed. tert. long	**-0.055	**-0.022	-0.036	-0.041	-0.101	-	-0.097	-0.125	-0.105	-0.075	-0.063	-0.064
Dad’s educ. NA	**0.004	**-0.002	-0.006	-	-0.011	-0.013	0.007	-	-0.010	**0.003	-0.011	-0.006
Dad’s educ. 1	-	-	-	-	-	-0.010	-	-	-	-	-	-
Dad’s educ. 2	**-0.003	0.024	0.024	0.029	0.011	-	0.032	**0.007	0.010	**0.011	0.014	**0.008
Dad’s educ. 3	0.013	**0.002	0.006	0.007	0.011	-	0.017	0.006	-	0.006	-	-
Dad’s educ. 4	**-0.001	**-0.008	**0.010	0.011	0.011	-	0.014	**-0.004	0.010	**-0.003	-0.016	-
Dad’s educ. 5	0.015	**-0.005	0.017	0.013	0.011	-	0.012	0.008	0.010	0.025	**-0.003	**0.003
Dad’s educ. 6	0.031	**0.008	0.016	0.019	0.025	-	**0.002	0.012	0.011	**0.007	0.010	0.010
Mum’s educ. NA	**0.004	**0.002	0.013	-	**-0.004	-0.024	0.006	-	-0.007	**0.002	0.005	**-0.003
Mum’s educ. 1	-	-	-	-	-	-0.022	-	-	-	-	-	-
Mum’s educ. 2	**0.003	**-0.002	0.015	0.014	0.018	-	0.026	0.026	**0.005	0.027	-	**0.009
Mum’s educ. 3	0.016	0.016	0.019	0.009	0.018	-	0.019	0.017	0.010	0.008	0.021	0.006
Mum’s educ. 4	0.035	**-0.014	0.035	0.022	0.018	-	0.031	0.032	0.024	0.017	0.016	0.016
Mum’s educ. 5	**0.000	0.012	0.015	**0.003	0.018	-	0.019	0.018	**0.001	0.007	0.010	**0.000

Mum's educ. 6	**-0.002	**0.010	**0.009	0.012	**-0.006	-	0.032	0.053	**0.000	**0.001	**0.007	0.016
Experience	**0.001	0.009	0.012	-	0.007	0.006	0.005	-	0.010	0.006	0.003	0.002
Exp. square root	0.056	-	0.040	0.09	0.018	0.041	0.031	0.072	0.028	0.021	0.078	0.102
Second. grade NA	-	-	0.307	0.368	0.222	0.313	0.187	0.243	0.199	0.201	0.309	0.383
Secondary grade	-	-	0.045	0.053	0.046	0.053	0.031	0.037	0.032	0.032	0.037	0.045
1, UE benefits	-0.020	-0.031	-0.044	-0.047	-0.042	-0.045	-0.072	-0.057	-0.057	-0.032	-0.080	-0.043
1, Soc. Sec. UE	-0.023	-0.022	-0.020	-0.019	-0.024	-0.039	-0.057	-0.037	-0.069	-0.059	-0.109	**-0.021
1, Soc. Sec. nUE	-0.028	-0.027	-0.020	-0.017	-0.049	-0.015	**-0.012	-0.068	-0.054	-0.029	-0.065	**-0.008
1, subs. empl.	-0.039	-0.011	-0.043	-0.037	-0.049	-	-0.035	-	-0.025	-0.023	-0.093	-0.058
1, disab. benefits	**0.000	-	-0.006	**0.001	-0.007	0.009	-0.005	0.002	0.004	0.016	**-0.003	0.003
1, maternity leave	-	0.006	-	-0.034	-	-0.029	**-0.001	-0.050	-0.004	-0.040	**-0.002	-0.048
1, other leave	**-0.011	-0.024	-0.047	-0.086	-0.036	-0.033	-0.054	-0.065	-0.081	-0.078	-0.087	-0.107
1, not in lab. force	-0.009	-0.013	-0.022	-	-0.036	-0.037	-0.015	-0.037	-0.034	-0.055	0.018	0.016
Wks, UE benefits	-0.134	-0.116	-0.195	-0.130	-0.193	-0.114	-0.179	-0.146	-0.184	-0.091	-0.216	-0.153
Wks, S. sec. UE	-0.050	-0.079	-0.165	-0.062	-0.137	-0.089	-0.233	-0.155	-0.101	-0.071	-0.291	-0.149
Wks, S. Sec. nUE	-0.062	-0.064	-0.128	-0.079	-0.106	-0.085	-0.126	-0.109	-0.128	**-0.042	-0.209	-0.210
Wks, subs. empl.	-0.048	-0.038	0.000	-	-0.033	-0.031	-0.086	-0.082	-0.095	-	-0.086	-0.128
Wks, disab. Ben.	-0.138	-0.118	0.000	-0.121	-0.185	-0.132	-0.244	-0.138	-0.269	-0.114	-0.283	-0.136
Wks, mater. leave	-	-	-	-	-	-0.028	**0.004	-	**-0.041	-0.069	**-0.006	-0.010
Wks, other leave	-0.087	**-0.006	0.000	0.021	-0.108	-0.045	-0.067	**0.000	**0.009	**0.004	0.059	0.036
Wks, not in lab. f.	-0.123	-0.066	0.000	-0.099	-0.160	**-0.023	-0.105	-0.065	-0.142	**-0.022	-0.164	-0.055

Note: ** indicates that the estimate is not significant at a 5% level.

Appendix 2 Model estimation

The model is estimated by a Bayesian simulation method on the basis of iterative random draws from the parameters' conditional distributions. In this appendix, the conditional posterior distributions for the model's parameters are derived. The derivations of the conditional posterior distributions for the parameters $\beta = (\beta_1, \beta_2)$, γ, ϕ and τ utilise the following differenced and ρ-conditional version of (1) and (2).

The model's parameters and latent variables are grouped in nine blocks and a new set of values are drawn from the conditional posterior distributions for each iteration of the Gipps sampler:

(i) $(\rho_{11}, \rho_{21}) \sim p((\rho_1, \rho_2) \mid \gamma_0, \beta_{10}, \beta_{20}, \phi_0, \tau_0, \sigma^2_{\tau 0}, \sigma^2_{10}, \sigma^2_{20}, Y_{20}, X, Y_1)$

(ii) $\gamma_1 \sim p(\gamma \mid \rho_{11}, \rho_{21}, \beta_{10}, \beta_{20}, \phi_0, \tau_0, \sigma^2_{\tau 0}, \sigma^2_{10}, \sigma^2_{20}, Y_{20}, X, Y_1)$

(iii) $(\beta_{11}, \beta_{21}) \sim p((\beta_1, \beta_2) \mid \rho_{11}, \rho_{21}, \gamma_1, \phi_0, \tau_0, \sigma^2_{\tau 0}, \sigma^2_{10}, \sigma^2_{20}, Y_{20}, X, Y_1)$

(iv) $\phi_1 \sim p(\varphi \mid \rho_{11}, \rho_{21}, \gamma_1, \beta_{11}, \beta_{21}, \tau_0, \sigma^2_{\tau 0}, \sigma^2_{10}, \sigma^2_{20}, Y_{20}, X, Y_1)$

(v) $\tau_1 \sim p(\sigma^2_\tau \mid \rho_{11}, \rho_{21}, \gamma_1, \beta_{11}, \beta_{21}, \phi_1, \sigma^2_{10}, \sigma^2_{20}, Y_{20}, X, Y_1)$

(vi) $\sigma^2_{\tau 1} \sim p(\sigma^2_\tau \mid \rho_{11}, \rho_{21}, \gamma_1, \beta_{11}, \beta_{21}, \phi_1, \tau_1, \sigma^2_{10}, \sigma^2_{20}, Y_{20}, X, Y_1)$

(vii) $\sigma^2_{11} \sim p(\sigma^2_1 \mid \rho_{11}, \rho_{21}, \gamma_1, \beta_{11}, \beta_{21}, \phi_1, \tau_1, \sigma^2_{\tau 1}, \sigma^2_{20}, Y_{20}, X, Y_1)$

(viii) $\sigma^2_{21} \sim p(\sigma^2_2 \mid \rho_{11}, \rho_{21}, \gamma_1, \beta_{11}, \beta_{21}, \phi_1, \tau_1, \sigma^2_{\tau 1}, \sigma^2_{11}, Y_{20}, X, Y_1)$

(ix) $Y_{21} \sim p(\sigma^2_2 \mid \rho_{11}, \rho_{21}, \gamma_1, \beta_{11}, \beta_{21}, \phi_1, \tau_1, \sigma^2_{\tau 1}, \sigma^2_{11}, \sigma^2_{21}, X, Y_1)$

$X = \{x_{it}\}_{i=1,..,N; t=1,..,T_i}$ and $Y = \{y_{it}\}_{i=1,..,N; t=1,..,T_i}$ are stacked data matrices. Y_1 are the known and Y_2 the missing 'latent' dependents. A '0' or a '1' index on the parameters indicate values drawn from the previous and the current iteration of the Gipps sampler.

In the following, the prior distributions are presented and the posterior distributions are derived.

Prior distributions

The prior distributions are the following:

(i) $\gamma \sim TN_{]-1,1[}(\underline{\gamma}, \underline{V}_\gamma)$

(ii) $(\rho_1, \rho_2) \sim TN_{]-1,1[}\left((\underline{\rho}_1, \underline{\rho}_2), \begin{Bmatrix} V_{\rho_1} & 0 \\ 0 & V_{\rho_2} \end{Bmatrix} \right)$

(iii) $(\beta_1, \beta_2) \sim N((\underline{\beta}_1, \underline{\beta}_2), \underline{V}_\beta)$

(iv) $\phi \sim N(\underline{\phi}, \underline{V}_\phi)$

(v) $f_\tau s_\tau^2 / \sigma_\tau^2 \sim \chi^2(f_\tau)$ with $f_\tau > 0$ and $s_\tau^2 > 0$

(vi) $f_1 s_1^2 / \sigma_1^2 \sim \chi^2(f_1)$ with $f_1 > 0$ and $s_1^2 > 0$

(vii) $f_2 s_2^2 / \sigma_2^2 \sim \chi^2(f_2)$ with $f_2 > 0$ and $s_2^2 > 0$

$N(\mu, \Sigma)$ is a normal distribution with mean μ and variance Σ. $TN_{]a,b[}(\mu, \Sigma)$ is a truncated normal distribution on the interval]a,b[. $\chi^2(f)$ is a χ^2 - distribution with f degrees of freedom. The parameters for the priors are chosen so that the posterior distributions are well-defined.

Posterior distributions

The presence of error term autocorrelation complicates the derivation of the posterior distributions. To overcome this, the first step derives the model in a differenced form, which conditional on $\rho = (\rho_1, \rho_2)$ has no autocorrelation.

The differenced model conditional on $\rho = (\rho_1, \rho_2)$

Conditional on ρ_1 and ρ_2 the model is rewritten in its differenced form with unit variance error terms:

(A1) $y_{i1}^{\sigma_1} = x_{i1}^{\sigma_1}\beta_1 + e_{i1} \quad e_{i1} \sim N(0,1)$.

(A2) $\dot{y}_{i2}^{\sigma_2} = \gamma y_{i1}^{\sigma_2} + x_{i2}^{\sigma_2}\beta_2 - \rho_1 x_{i1}^{\sigma_2}\beta_1 + \tau_i^{\sigma_2} + \phi\varepsilon_{i1}^{\sigma_2} + e_{i2} \quad e_{i2} = \sigma_2^{-1}\eta_{i2} \sim N(0,1)$

(A3) $\dot{y}_{it}^{\sigma_2} = \gamma\dot{y}_{it-1}^{\sigma_2} + \dot{x}_{it}^{\sigma_2}\beta_2 + (1-\rho_2)\tau_i^{\sigma_2} + (1-\rho_2)\phi\varepsilon_{i1}^{\sigma_2} + e_{it} \quad e_{it} \sim nid(0,1)$

Where $y_{i1}^{\sigma_1} = \sigma_1^{-1} y_{i1}$, $\dot{y}_{i2}^{\sigma_2} = \sigma_2^{-1}(y_{i2} - \rho_1 y_{i1})$ and $\dot{z}_{it}^{\sigma_2} = \sigma_2^{-1}(z_{it} - \rho_2 z_{it-1})$.

The following result for the posterior distribution of a generalised linear regression model is utilised for most parameters.

Box 1 The posterior distribution of a generalized linear regression model

Conditional on the normal prior distribution $\beta \sim N(\underline{\beta}, \underline{V}_\beta)$, the posterior distribution of the generalised linear regression model $Y = X\beta + \varepsilon$ with heteroscedastic errors $\varepsilon \sim N(0, \Omega)$ and $\beta \sim N(\hat{\beta}, V_\beta)$ where $\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$ and $V_\beta = (X'\Omega^{-1}X)^{-1}$ is given by $\beta \sim N(\overline{\beta}, \overline{V}_\beta)$ where

(A4) $\overline{V}_\beta^{-1} = V_\beta^{-1} + \underline{V}_\beta^{-1}$ and $\overline{\beta} = \overline{V}_\beta (V_\beta^{-1}\hat{\beta} + \underline{V}_\beta^{-1}\underline{\beta})$

We are now ready to derive the posterior distributions for the model's parameters.

The posterior distribution for $\rho = (\rho_1, \rho_2)$

The posterior distribution for $\rho = (\rho_1, \rho_2)$ is derived rearranging (3) and (4) in the following way:

(3.ρ) $\varepsilon_{i2} = \rho_1 \varepsilon_{i1} + \eta_{i2}$

$$\varepsilon_{i2} = \begin{pmatrix} \varepsilon_{i1} & 0 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} + \eta_{i2}$$

(4.ρ) $\varepsilon_{it} = \rho_2 \varepsilon_{i,t-1} + \eta_{it}$

$$\varepsilon_{it} = \begin{pmatrix} 0 & \varepsilon_{i,t-1} \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} + \eta_{it}$$

By stacking these equations for every *i*

$$\begin{Bmatrix} \varepsilon_{i2} \\ \varepsilon_{i3} \\ \cdot \\ \varepsilon_{iT_i} \end{Bmatrix} = \begin{Bmatrix} \varepsilon_{i1} & 0 \\ 0 & \varepsilon_{i2} \\ \cdot & \cdot \\ 0 & \varepsilon_{iT_i-1} \end{Bmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} + \begin{Bmatrix} \eta_{i,2} \\ \eta_{i,3} \\ \cdot \\ \eta_{i,T_i} \end{Bmatrix}$$

This is equation can be written as

$$\varepsilon_{i,2-T_i} = \varepsilon_i^\rho \rho + \eta_{i,2-T_i}$$

Here $\eta_{i,2-T_i} \sim N(0, \Sigma_\rho)$ with $\Sigma_\rho = \sigma_2^2 I_{T_i-1}$. By pre-multiplying by $\Sigma_\rho^{-½} = \sigma_2^{-1} I_{T_i-1}$ the above equation simplifies to $\Sigma_\rho^{-½} \varepsilon_{i,2-T_i} = \Sigma_\rho^{-½} \varepsilon_i^\rho \rho + e_i^\rho$ or $\varepsilon^{\Sigma}_{i,2-T\,i,2-T_i} = \varepsilon_i^{\Sigma\rho} \rho + e_i$, which can be stacked by *i* to obtain:

$$Y^\rho = X^\rho + e$$

This is another standard regression with white noise errors and

$$\rho \sim N\left((X^{\rho}{}'X^{\rho})^{-1}X^{\rho}{}'Y^{\rho},(X^{\rho}{}'X^{\rho})^{-1}\right)$$

The posterior distribution for $\rho=(\rho_1,\rho_2)$ is then derived by applying the result in Box 1.

The posterior distribution for $\beta=(\beta_1,\beta_2)$

The posterior distribution for $\beta=(\beta_1,\beta_2)$ is derived by rearranging (A1)-(A3).

(A1. β) $\quad y_{i1}^{\sigma_1}=\begin{pmatrix}x_{i1}^{\sigma_1} & 0\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}+e_{i1}$

(A2. β) $\quad \dot{y}_{i2}^{\sigma_2}-(\gamma y_{i1}^{\sigma_2}+\tau_i^{\sigma_2}+\phi\varepsilon_{i1}^{\sigma_2})=\begin{pmatrix}-\rho_1 x_{i1}^{\sigma_2} & x_{i2}^{\sigma_2}\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}+e_{i,2}$

(A3. β) $\quad \dot{y}_{it}^{\sigma_2}-(\gamma\dot{y}_{it-1}^{\sigma_2}+(1-\rho_2)\tau_i^{\sigma_2}+(1-\rho_2)\phi\varepsilon_{i1}^{\sigma_2})=\begin{pmatrix}0 & \dot{x}_{it}^{\sigma_2}\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}+e_{it}$

These are stacked by *t* for every *i*:

$$\begin{Bmatrix} y_{i1}^{\sigma_1} \\ \dot{y}_{i2}^{\sigma_2}-(\gamma y_{i1}^{\sigma_2}+\tau_i^{\sigma_2}+\phi\varepsilon_{i1}^{\sigma_2}) \\ . \\ \dot{y}_{it}^{\sigma_2}-(\gamma\dot{y}_{it-1}^{\sigma_2}+(1-\rho_2)\tau_i^{\sigma_2}+(1-\rho_2)\phi\varepsilon_{i1}^{\sigma_2}) \\ . \end{Bmatrix}=\begin{Bmatrix} x_{i1}^{\sigma_1} & 0 \\ -\rho_1 x_{i1}^{\sigma_2} & x_{i2}^{\sigma_2} \\ . & . \\ 0 & \dot{x}_{it}^{\sigma_2} \\ . & . \end{Bmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}+e_i$$

The following compact form is then derived by further stacking the β-equation:

$$Y^{\beta}=X^{\beta}+e$$

This is a standard regression with white noise errors and hence

$$\beta \sim N\left((X^{\beta}{}'X^{\beta})^{-1}X^{\beta}{}'Y^{\beta},(X^{\beta}{}'X^{\beta})^{-1}\right)$$

The posterior distribution for β is then derived by applying the result in Box 1.

The posterior distribution for γ

The posterior distribution for γ is derived in a similar manner by rearranging (A2) og (A3).

(A2. γ) $\quad \dot{y}_{i2}^{\sigma_2}=\gamma y_{i1}^{\sigma_2}+x_{i2}^{\sigma_2}\beta_2-\rho_1 x_{i1}^{\sigma_2}\beta_1+\tau_i^{\sigma_2}+\phi\varepsilon_{i1}^{\sigma_2}+e_{i2}$

$$\dot{y}_{i2}^{\sigma_2}-(x_{i2}^{\sigma_2}\beta_2-\rho_1 x_{i1}^{\sigma_2}\beta_1+\tau_i^{\sigma_2}+\phi\varepsilon_{i1}^{\sigma_2})=y_{i1}^{\sigma_2}\gamma+e_{i2}$$

(A3. γ) $\quad \dot{y}_{it}^{\sigma_2} = \gamma \dot{y}_{it-1}^{\sigma_2} + \dot{x}_{it}^{\sigma_2}\beta_2 + (1-\rho_2)(\tau_i^{\sigma_2} + \phi\varepsilon_{i1}^{\sigma_2}) + e_{it}$

$$\dot{y}_{it}^{\sigma_2} - (\dot{x}_{it}^{\sigma_2}\beta_2 + (1-\rho_2)(\tau_i^{\sigma_2} + \phi\varepsilon_{i1}^{\sigma_2})) = \dot{y}_{it-1}^{\sigma_2}\gamma + e_{it}$$

Stack again by *t* for every *i*:

$$\begin{Bmatrix} \dot{y}_{i2}^{\sigma_2} - (x_{i2}^{\sigma_2}\beta_2 - \rho_1 x_{i1}^{\sigma_2}\beta_1 + \tau_i^{\sigma_2} + \phi\varepsilon_{i1}^{\sigma_2}) \\ . \\ \dot{y}_{it}^{\sigma_2} - (\dot{x}_{it}^{\sigma_2}\beta_2 + (1-\rho_2)(\tau_i^{\sigma_2} + \phi\varepsilon_{i1}^{\sigma_2})) \\ . \end{Bmatrix} = \begin{Bmatrix} y_{i1}^{\sigma_2} \\ . \\ \dot{y}_{it-1}^{\sigma_2} \\ . \end{Bmatrix} \gamma + e_i$$

Then by further stacking by *i*:

$$Y^{\gamma} = X^{\gamma} + e$$

This is a standard regression with white noise errors and hence

$$\gamma \sim N\left((X^{\gamma}{}'X^{\gamma})^{-1}X^{\gamma}{}'Y^{\gamma}, (X^{\gamma}{}'X^{\gamma})^{-1}\right)$$

The posterior distribution for γ is then expressed then derived by applying the result in Box 1.

The posterior distribution for $\phi = (\phi^1, \phi^2)$

The posterior distribution for ϕ is derived rearranging (A2) and (A3):

(A2. ϕ) $\quad \dot{y}_{i2}^{\sigma_2} = \gamma y_{i1}^{\sigma_2} + x_{i2}^{\sigma_2}\beta_2 - \rho_1 x_{i1}^{\sigma_2}\beta_1 + \tau_i^{\sigma_2} + \phi\varepsilon_{i1}^{\sigma_2} + e_{i,2}$

$$\dot{y}_{i2}^{\sigma_2} - (\gamma y_{i1}^{\sigma_2} + x_{i2}^{\sigma_2}\beta_2 - \rho_1 x_{i1}^{\sigma_2}\beta_1 + \tau_i^{\sigma_{2i}}) = \sigma_2^{-1}\phi\varepsilon_{i1} + e_{i,2}$$

(A3. ϕ) $\quad \dot{y}_{it}^{\sigma_2} = \gamma \dot{y}_{it-1}^{\sigma_2} + \dot{x}_{it}^{\sigma_2}\beta_2 + (1-\rho_2)(\tau_i^{\sigma_2} + \phi\varepsilon_{i1}^{\sigma_2}) + e_{it}$

$$\dot{y}_{it}^{\sigma_2} - (\gamma \dot{y}_{it-1}^{\sigma_2} + \dot{x}_{it}^{\sigma_2}\beta_2 + (1-\rho_2)\tau_i^{\sigma_2}) = \sigma_2^{-1}(1-\rho_2)\phi\varepsilon_{i1} + e_{it}$$

In stacked form, first for each *i*

$$\begin{Bmatrix} \dot{y}_{i2}^{\sigma_2} - (\gamma y_{i1}^{\sigma_2} + x_{i2}^{\sigma_2}\beta_2 - \rho_1 x_{i1}^{\sigma_2}\beta_1 + \tau_i^{\sigma_2}) \\ . \\ \dot{y}_{it}^{\sigma_2} - (\gamma \dot{y}_{it-1}^{\sigma_2} + \dot{x}_{it}^{\sigma_2}\beta_2 + (1-\rho_2)\tau_i^{\sigma_2}) \\ . \end{Bmatrix} = \begin{Bmatrix} \sigma_2^{-1}\varepsilon_{i1} \\ . \\ \sigma_2^{-1}(1-\rho_2)\varepsilon_{i1} \\ . \end{Bmatrix} \phi + e_i$$

And then by *i*:

$$Y^{\phi} = X^{\phi}\phi + e$$

This is a standard regression with white noise errors and hence

$$\phi \sim N\left((X^{\phi}{}'X^{\phi})^{-1}X^{\phi}{}'Y^{\phi},(X^{\phi}{}'X^{\phi})^{-1}\right)$$

The posterior distribution for ϕ is then derived by applying the result in Box 1.

The posterior distribution for σ_τ^2

The conditional posterior distribution for σ_τ^2 is given by

$$\left(\sum_{i=1}^{N}\tau_i^2+\underline{v}_\tau\underline{s}_\tau^2\right)\Big/\sigma_\tau^2 \sim \chi^2(N+\underline{v}_\tau)$$

$\underline{s}_\tau^2$ and $\underline{v}_\tau$ are the parameters (mean and degrees of freedom) for the prior distribution of σ_τ^2 .

The posterior distribution for τ_i

The conditional posterior distribution for τ_i is derived by first rearranging (A2) and (A3)

(A2. τ_i) $$\dot{y}_{i2}^{\sigma_2}=\gamma y_{i1}^{\sigma_2}+x_{i2}^{\sigma_2}\beta_2-\rho_1 x_{i1}^{\sigma_2}\beta_1+\tau_i^{\sigma_2}+\phi\varepsilon_{i1}^{\sigma_2}+e_{i,2}$$

$$\dot{y}_{i2}^{\sigma_2}-(\gamma y_{i1}^{\sigma_2}+x_{i2}^{\sigma_2}\beta_2-\rho_1 x_{i1}^{\sigma_2}\beta_1+\phi\varepsilon_{i1}^{\sigma_2})=\sigma_2^{-1}\tau_i+e_{i,2}$$

(A3. τ_i) $$\dot{y}_{it}^{\sigma_2}=\gamma\dot{y}_{it-1}^{\sigma_2}+\dot{x}_{it}^{\sigma_2}\beta_2+(1-\rho_2)(\tau_i^{\sigma_2}+\phi\varepsilon_{i1}^{\sigma_2})+e_{it}$$

$$\dot{y}_{it}^{\sigma_2}-(\gamma\dot{y}_{it-1}^{\sigma_2}+\dot{x}_{it}^{\sigma_2}\beta_2+(1-\rho_2)\phi\varepsilon_{i1}^{\sigma_2})=\sigma_2^{-1}(1-\rho_2)\tau_i+e_{it}$$

These equations are stacked for each *i*:

$$\begin{Bmatrix}\dot{y}_{i2}^{\sigma_2}-(\gamma y_{i1}^{\sigma_2}+x_{i2}^{\sigma_2}\beta_2-\rho_1 x_{i1}^{\sigma_2}\beta_1+\phi\varepsilon_{i1}^{\sigma_2})\\ .\\ \dot{y}_{it}^{\sigma_2}-(\gamma\dot{y}_{it-1}^{\sigma_2}+\dot{x}_{it}^{\sigma_2}\beta_2+(1-\rho_2)\phi\varepsilon_{i1}^{\sigma_2})\\ .\end{Bmatrix}=\begin{Bmatrix}\sigma_2^{-1}\\ .\\ \sigma_2^{-1}(1-\rho_2)\\ .\end{Bmatrix}\tau_i+e_i$$

An entry is added for the σ_τ^2 -term in the likelihood function:

$$0=\sigma_\tau^{-1}\tau_i+\sigma_\tau^{-1}\omega_i$$

Thereby arriving at the following regression for each *i*:

$$\left\{\begin{array}{c} \dot{y}_{i,2}^{\sigma_2} - (\gamma y_{i,1}^{\sigma_2} + x_{i,2}^{\sigma_2}\beta_2 - \rho_1 x_{i,1}^{\sigma_2}\beta_1 + \sigma_2^{-1}\phi\varepsilon_{i1}) \\ . \\ \dot{y}_{it}^{\sigma_2} - (\gamma\dot{y}_{it-1}^{\sigma_2} + \dot{x}_{it}^{\sigma_2}\beta_2 + \sigma_2^{-1}(1-\rho_2)\phi\varepsilon_{i1}) \\ . \\ 0 \end{array}\right\} = \left\{\begin{array}{c} \sigma_2^{-1} \\ . \\ \sigma_2^{-1}(1-\rho_2) \\ . \\ \sigma_\tau^{-1} \end{array}\right\}\tau_i + e_i$$

This can be written in the more compact form

$$Y^{\tau_i} = X^{\tau_i}\tau_i + e \qquad \text{i=1,…,N}$$

The conditional posterior distribution for τ_i is given by

$$\tau_i \sim N\left((X^{\tau_i}{}'X^{\tau_i})^{-1}X^{\tau_i}{}'Y^{\tau_i}, (X^{\tau_i}{}'X^{\tau_i})^{-1}\right) \qquad \text{i=1,…,N}$$

At the end of each iteration, values for σ_1^2 and σ_2^2 are drawn from the following conditional posterior distributions

$$\left(\sum_{i=1}^{N}\hat{\varepsilon}_{i1}^2 + \underline{\nu}_1\underline{s}_1^2\right) \Big/ \sigma_1^2 \sim \chi^2(N+\underline{\nu}_1)$$

$$\left(\sum_{i=1}^{N}\sum_{t=2}^{t_i}\hat{\eta}_{it}^2 + \underline{\nu}_2\underline{s}_2^2\right) \Big/ \sigma_2^2 \sim \chi^2(N+\underline{\nu}_2)$$

$(\underline{s}_1^2, \underline{\nu}_1)$ and $(\underline{s}_2^2, \underline{\nu}_2)$ are the prior distribution parameters (mean and degrees of freedom) for σ_1^2 and σ_2^2.

Sampling the parameters in a mixed normal distribution

Assume $\{\tau_i\}_{i=1,\ldots,n}$ is an observed sample from a mixed normal distribution of *m* normal distributions $n(\mu_j, \sigma_j^2)_{j=1,\ldots,m}$ that is

$$\tau_i = \sum_{j=1}^{m} e_{ij}(\alpha_j + \sigma_j\xi_i) \qquad \xi_i \sim n(0,1)$$

The indicator matrix $e = \{e_{ij}\}_{i=1,\ldots,n;\, j=1,\ldots m}$ specifies the distributions for $\{\tau_i\}_{i=1,\ldots,n}$. $e_i = \{e_j\}_{j=1,\ldots m}$ are independent multinomial with $P(e_{ij}=1) = p_j$.

Definitions:

$$\tau = \begin{Bmatrix} \tau_1 \\ \tau_2 \\ \cdot \\ \tau_n \end{Bmatrix} \qquad e = \begin{Bmatrix} e_{11} & e_{12} & \cdot & e_{1m} \\ e_{21} & e_{22} & \cdot & e_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ e_{n1} & e_{n2} & \cdot & e_{nm} \end{Bmatrix}$$

$n_j = \sum_{i=1}^{n} e_{ij}$ specify the number of observations in each of the *m* groups.

The parameters for sampling are α_j , σ_j^2 and p_j . The indicators e_{ij} are also sampled.

1. α_j - the mean for the *m* distributions

Prior distribution: $\alpha \sim n(\underline{\alpha}, \underline{V}_\alpha)$

The posterior distribution (conditional on σ_j^2 , p_j and e_{ij}), *cf. Box 1*:

$$\alpha \sim n(\overline{\alpha}, \overline{V}_\alpha) \text{ where } \overline{V}_\alpha = (V_\alpha^{-1} + \underline{V}_\alpha^{-1})^{-1} \text{ and } \overline{\alpha} = \overline{V}_\alpha (e^{\sigma'} \tau^\sigma + \underline{V}_\alpha^{-1} \underline{\alpha})$$

Each element is shown in the following derivation.

The distribution for $\{\tau_i\}_{i=1,\dots,n}$ may be expressed as

$$\tau = e\alpha + \varepsilon \text{, where } \varepsilon \sim n(0, I \cdot V) \text{ and } V = \begin{Bmatrix} \sigma_{g_1}^2 \\ \sigma_{g_2}^2 \\ \cdot \\ \sigma_{g_n}^2 \end{Bmatrix} \text{ where } g_i = (j : e_{ij} = 1)$$

This regression equation is normalized by element-multiplying (#) through by $V^{-½} = (\sigma_{g_1}^{-1}, \sigma_{g_2}^{-1}, \dots, \sigma_{g_n}^{-1})'$

$$\tau^\sigma = e^\sigma \alpha + \xi \text{, where } \tau^\sigma = V^{-½} \# \tau \text{, } e^\sigma = V^{-½} \# e \text{ and } \xi \sim n(0, I)$$

This is a standard regression with parameters α and independent standard normal error terms, which means that $\alpha \sim n(\hat{\alpha}, V_\alpha)$ where $\hat{\alpha} = (e^{\sigma'} e^\sigma)^{-1} e^{\sigma'} \tau^\sigma$ and $V_\alpha = (e^{\sigma'} e^\sigma)^{-1}$. The above posterior follows immediately from the usual result for the posterior distribution of a generalized linear model.

2. σ_j^2 - the variance for the *m* distributions

Prior distribution: $f_j \, {s_j^2}/{\sigma_j^2} \sim \chi^2(f_j)$ where $s_j^2 > 0$ and $f_j^2 > 0$

The posterior distribution (conditional on α_j , p_j and e_{ij}):

$$\left[\sum_{i=1}^{n} e_{ij} (\tau_i - \alpha_j)^2 + f_j \underline{s}_j^2 \right] \Big/ \sigma_j^2 \sim \chi^2(n_j + f_j)$$

Computationally, $\sum_{i=1}^{n} e_{ij}(\tau_i - \alpha_j)^2$ can be efficiently expressed in matrix form as $e'(\tau - e\alpha)$ ##2. The operator ##2 means taking the square of each element in the vector $(\tau - e\alpha)$.

3. $\{p_j\}_{j=1,..,m}$ – the overall probabilities for each normal distribution.

Prior distribution: $p = (p_1,..., p_m) \sim Beta(r_1,...,r_m)$.

The posterior distribution (conditional on α_j, σ_j^2 and e_{ij}):

$$p \sim Beta(\{n_j + r_j\}_{j=1,..,m})$$

4. The distribution of e_{ij} conditional on α_j, σ_j^2 and p_j

$e = \{e_{ij}\}_{i=1,..,n;\, j=1,..,m}$ are drawn from a multinomial distribution with parameters p_{ij}. These are proportional to

$$P(e_{ij} = 1) \propto p_j \sigma_j^{-1} \exp\left[-\frac{1}{2\sigma_j^2}(\tau_i - \alpha_j)^2\right]$$

The p_j do not add up to 1 (i.e. $\sum_{j=1}^{m} p_j \neq 1$) and hence require normalization prior to sampling of $\{e_{ij}\}_{i=1,..,n;\, j=1,..,m}$.

Computationally, $p_j \sigma_j^{-1} \exp\left[-\frac{1}{2\sigma_j^2}(\tau_i - \alpha_j)^2\right]$ can be efficiently expressed in matrix form as

$$\exp\left[-½(\tau \cdot 1_{1xm} - 1_{nx1}\alpha')^2 \sigma_d^{-2}\right](\sigma_d^{-1} p_d)$$

Here the $\exp()$ and the power function $()^2$ are element by element functions. 1_{1xm} is an 1xm vector of 1 and σ_d^{-2} (and σ_d^{-1}) are mXm diagonal matrices of σ_j^{-2} (and σ_j^{-1}). Likewise p_d is a diagonal matrix of the p-values.

Appendix 3 Simulating missing observations for the dependant variable

The missing observations for the dependent variable are treated as latent variables and simulated as an integrated step during the iterative estimation procedure. The observations are drawn from the conditional distribution at the end of each iteration, that is, conditional the model and its simulated parameters. The model is a dynamic panel model with complex intertemporal dependencies. Consequently, the missing observations are correlated with both past and future values of the dependent and independent variables. In addition, the situation is complicated by cases of consecutive missing observations that, because of the intertemporal dependencies, are correlated and must therefore be drawn from their joint distributions.

In the following, the joint conditional distributions for the missing observations are derived. The method is used for two purposes. Firstly, the joint conditional distributions are used for drawing the missing observations during the Bayes iterations. Secondly, conditional on the estimated model parameters values for the missing observations are simulated from the joint distributions conditional and thereby closing the ‘holes’ in the data.

The following result is used frequently $\sigma_\varepsilon^2 = \sigma_2^2 / 1-\rho_2^2$ and $Cov(\varepsilon_{it},\varepsilon_{it+1}) = \rho_2\sigma_\varepsilon^2 = \rho_2\sigma_2^2 / 1-\rho_2^2$

The conditional joint distribution for consecutive missing observations is derived by apply extrapolation of the model’s main equation (2) in its forward and backward versions[9]

(2f) $y_{it} = \gamma y_{it-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$ with $\varepsilon_{it} = \rho_2\varepsilon_{it-1} + e_{it}^f$

(2b) $y_{it} = \gamma y_{i,t+1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$ with $\varepsilon_{it} = \rho_2\varepsilon_{it+1} + e_{it}^b$

With *h* consecutive missing observations these extrapolations lead to 2*h* equations that, conditional on the model’s parameters, can be rearranged and stacked to form a simple GLS regression with known error term variance and with the missing observations as unknown parameters. The joint distribution of the missing observations is then multinomial normal with the usual mean and variance of the parameter estimates.

The details follow from the following compilations for the cases with up to four consecutive missing. Cases of five or more missing observations are relative straightforward generalisations.

1. One missing observation: $y_{i,t}$ is missing, $y_{i,t-1}$ and $y_{i,t+1}$ are known

This is the simplest case. A straightforward rearrangement of (2f) and (2b)

$(y_{i,t} \mid t-1)$ $\quad y_{it} = \gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$

$$-(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) = -y_{it} + \varepsilon_{it}$$

---

[9] For simplicity, the random effect and the period 1 $\tau_i^* = \tau_i + \phi\varepsilon_{i1}$

$$-(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) - \rho^*\varepsilon_{it-1} = -y_{it} + e_{it}^f$$

With $\rho^* \begin{cases} \rho_1 & hvis \quad t = 1 \\ \rho_2 & hvis \quad t \neq 1 \end{cases}$

$$(y_{i,t} \mid t+1) \qquad y_{it} = \gamma y_{i,t+1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$$

$$-(\gamma y_{it+1} + (x_{it}\beta_2 + \tau_i^*)) = -y_{i,t} + \varepsilon_{it}$$

$$-(\gamma y_{it+1} + (x_{it}\beta_2 + \tau_i^*)) - \rho_2\varepsilon_{it+1} = -y_{i,t} + e_{it}^b$$

This is equivalent to the equation

$$\begin{Bmatrix} y_i^1 \\ y_i^2 \end{Bmatrix} = \begin{Bmatrix} -1 \\ -1 \end{Bmatrix} y_{it} + \begin{Bmatrix} e_{it}^f \\ e_{it}^b \end{Bmatrix}$$

or

$$y_i = \Gamma_1 y_{it} + e_i$$

where

$$\begin{Bmatrix} y_i^1 \\ y_i^2 \end{Bmatrix} = \begin{Bmatrix} -(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) - \rho^*\varepsilon_{it-1} \\ -(\gamma y_{it+1} + (x_{it}\beta_2 + \tau_i^*)) - \rho_2\varepsilon_{it+1} \end{Bmatrix}$$

$$= -\begin{Bmatrix} \gamma y_{i,t-1} \\ \gamma y_{it+1} \end{Bmatrix} - \begin{Bmatrix} x_{it}\beta_2 \\ x_{it}\beta_2 \end{Bmatrix} - \tau_i^* \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} - \begin{Bmatrix} \rho^*\varepsilon_{it-1} \\ \rho_2\varepsilon_{it+1} \end{Bmatrix}$$

and $e_i \sim N(0, \Sigma_1)$, where $\Sigma_1 = \sigma_2^2 \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$

Multiplying by $\Sigma_1^{-1/2}$ [10] obtains:

$$y_i^\Sigma = \Gamma_1^\Sigma y_{it} + e_i$$

This is a standard OLS with y_{it} as a parameter two observations and with a N(0,1) error term. The distribution for y_{it} is then

$$y_{it} \sim N\left((\Gamma_1^{\Sigma'}\Gamma_1^\Sigma)^{-1}\Gamma_1^\Sigma y_i^\Sigma, (\Gamma_1^{\Sigma'}\Gamma_1^\Sigma)^{-1}\right)$$

Notice that the variance matrix Γ_1^Σ does not depend on individual characteristics, and therefore only requires re-evaluation once per Gipps iteration.

2. Two missing observation:: $y_{i,t}$ and $y_{i,t+1}$ are missing, $y_{i,t-1}$ $y_{i,t+2}$ are known

The case of two missing observations results in a regression with four observations, two from (2f) and two from (2b). First, by rearranging (2f) and to express $y_{i,t}$ and $y_{i,t+1}$ conditional on t-1:

---

[10] $\Sigma_1^{-1/2}$ exists when Σ_1 positive definite, which is is normally the case. Nevertheless, standard GLS will lead to the same result.

$$(y_{i,t} \mid t-1) \qquad y_{it} = \gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$$

$$-(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) = -y_{it} + \varepsilon_{it}$$

$$-(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) - \rho^*\varepsilon_{it-1} = -y_{it} + e_{it}^f$$

$$(y_{i,t+1} \mid t-1) \qquad y_{it+2} = \gamma(\gamma y_{i,t} + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}) + x_{it+2}\beta_2 + \tau_i^* + \varepsilon_{it+2}$$

$$-(\gamma^2 y_{i,t-1} + (\gamma x_{it} + x_{it+1}\beta_2 + (\gamma+1)\tau_i^*) = -y_{it+1} + \gamma\varepsilon_{it} + \varepsilon_{it+1}$$

$$-(\gamma^2 y_{i,t-1} + (\gamma x_{it} + x_{it+1}\beta_2 + (\gamma+1)\tau_i^*) - \rho^*(\gamma+\rho_2)\varepsilon_{it-1} =$$

$$-y_{it+1} + (\gamma+\rho_2)e_{it}^f + e_{it+1}^f$$

Second, by rearranging (2b) and to express $y_{i,t}$ and $y_{i,t+1}$ conditional on t+2:

$$(y_{i,t+1} \mid t+2) \qquad y_{it+1} = \gamma y_{i,t+2} + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}$$

$$-(\gamma y_{i,t+2} + x_{it+1}\beta_2 + \tau_i^*) - \rho_2\varepsilon_{it+2} = -y_{it+1} + e_{t+1}^b$$

$$(y_{i,t} \mid t+2) \qquad y_{it} = \gamma y_{i,t+1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$$

$$y_{it} = \gamma^2 y_{i,t+2} + (\gamma x_{it+1} + x_{it})\beta_2 + (\gamma+1)\tau_i^* + \gamma\varepsilon_{it+1} + \varepsilon_{it}$$

$$-(\gamma^2 y_{i,t+2} + (\gamma x_{it+1} + x_{it})\beta_2 + (\gamma+1)\tau_i^*) - (\gamma\rho_2 + \rho_2^2)\varepsilon_{it+2} =$$

$$-y_{it} + (\gamma+\rho_2)e_{it+1}^b + e_{it}^b$$

This is equivalent to the following regression with $y_{i,t}$ and $y_{i,t+1}$ as parameters:

$$\begin{Bmatrix} y_i^1 \\ y_i^2 \\ y_i^3 \\ y_i^4 \end{Bmatrix} = \begin{Bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & -1 \\ -1 & 0 \end{Bmatrix} \begin{Bmatrix} y_{it} \\ y_{it+1} \end{Bmatrix} + \begin{Bmatrix} e_{it}^f \\ (\gamma+\rho_2)e_{it}^f + e_{it+1}^f \\ e_{t+1}^b \\ (\gamma+\rho_2)e_{it+1}^b + e_{it}^b \end{Bmatrix}$$

or

$$y_i = \Gamma_2 \begin{Bmatrix} y_{it} \\ y_{it+1} \end{Bmatrix} + e_i$$

where

$$y_i = \begin{Bmatrix} y_i^1 \\ y_i^2 \\ y_i^3 \\ y_i^4 \end{Bmatrix} = \begin{Bmatrix} -(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) - \rho^*\varepsilon_{it-1} \\ -(\gamma^2 y_{i,t-1} + (\gamma x_{it} + x_{it+1})\beta_2 + (\gamma+1)\tau_i^*) - \rho^*(\gamma+\rho_2)\varepsilon_{it-1} \\ -(\gamma y_{i,t+2} + x_{it+1}\beta_2 + \tau_i^*) - \rho_2\varepsilon_{it+2} \\ -(\gamma^2 y_{i,t+2} + (\gamma x_{it+1} + x_{it})\beta_2 + (\gamma+1)\tau_i^*) - (\gamma\rho_2 + \rho_2^2)\varepsilon_{it+2} \end{Bmatrix}$$

$$= -\begin{Bmatrix} \gamma y_{i,t-1} \\ \gamma^2 y_{i,t-1} \\ \gamma y_{i,t+2} \\ \gamma^2 y_{i,t+2} \end{Bmatrix} - \begin{Bmatrix} x_{it} \\ x_{it+1} \\ x_{it+1} \\ x_{it} \end{Bmatrix} - \gamma \begin{Bmatrix} 0 \\ x_{it} \\ 0 \\ x_{it+1} \end{Bmatrix} - \tau_i^* \begin{Bmatrix} 1 \\ \gamma+1 \\ 1 \\ \gamma+1 \end{Bmatrix} + \begin{Bmatrix} -\rho^* \varepsilon_{it-1} \\ -\rho^*(\gamma+\rho_2)\varepsilon_{it-1} \\ -\rho_2 \varepsilon_{it+2} \\ -(\gamma\rho_2+\rho_2^2)\varepsilon_{it+2} \end{Bmatrix}$$

and $e_i \sim N(0, \Sigma_2)$, where

$$\Sigma_2 = \sigma_2^2 \begin{Bmatrix} 1 & (\gamma+\rho_2) & 0 & 0 \\ (\gamma+\rho_2) & (\gamma+\rho_2)^2+1 & 0 & 0 \\ 0 & 0 & 1 & (\gamma+\rho_2) \\ 0 & 0 & (\gamma+\rho_2) & (\gamma+\rho_2)^2+1 \end{Bmatrix}$$

Multiplying by $\Sigma_2^{-½}$ to get

$$y_i^{\Sigma} = \Gamma_2^{\Sigma} \begin{Bmatrix} y_{it} \\ y_{it+1} \end{Bmatrix} + e_i$$

This is a standard OLS with four observations and N(0,1) error terms. The distribution for the parameters y_{it} and $y_{i,t+1}$ is

$$(y_{it}, y_{it+1}) \sim N\left((\Gamma_2^{\Sigma'}\Gamma_2^{\Sigma})^{-1}\Gamma_2^{\Sigma} y_i^{\Sigma}, (\Gamma_2^{\Sigma'}\Gamma_2^{\Sigma})^{-1}\right)$$

3. Three missing observation: $y_{i,t}$, $y_{i,t+1}$ and $y_{i,t+2}$ are missing, $y_{i,t-1}$ og $y_{i,t+3}$ are known

The case of three missing observations results in a regression with six observations. First, by rearranging (2f) and to express $y_{i,t}$, $y_{i,t+1}$ and $y_{i,t+2}$ conditional on t-1:

$$(y_{i,t} \mid t-1) \qquad y_{it} = \gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$$

$$y_{it} = \gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$$

$$-(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) - \rho^*\varepsilon_{it-1} = -y_{it} + e_{it}^f$$

$$(y_{i,t+1} \mid t-1) \qquad y_{it+1} = \gamma y_{i,t} + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}$$

$$y_{it+1} = \gamma(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}) + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}$$

$$-(\gamma^2 y_{i,t-1} + (\gamma x_{it} + x_{it+1})\beta_2 + (\gamma+1)\tau_i^*) - \rho^*(\gamma+\rho_2)\varepsilon_{it-1} =$$

$$-y_{it+1} + (\gamma+\rho_2)e_{it}^f + e_{it+1}^f$$

$$(y_{i,t+2} \mid t-1) \; y_{it+2} = \gamma(\gamma y_{i,t} + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}) + x_{it+2}\beta_2 + \tau_i^* + \varepsilon_{it+2}$$

$$-(\gamma^3 y_{i,t-1} + (\gamma^2 x_{it} + \gamma x_{it+1} + x_{it+2})\beta_2 + (\gamma^2+\gamma+1)\tau_i^*))$$
$$-\rho^*(\gamma^2 + \gamma\rho_2 + \rho_2^2)\varepsilon_{it-1} = -y_{it+2} + (\gamma^2+\gamma\rho_2+\rho_2^2)e_{it}^f + (\gamma+\rho_2)e_{it+1}^f + e_{it+2}^f$$

And then, by rearranging (2b) and to express the missing observations conditional on t+2:

$$(y_{i,t+2}\,|\,t+3)\quad y_{it+2}=\gamma y_{i,t+3}+x_{it+2}\beta_2+\tau_i^*+\varepsilon_{it+2}$$

$$-(\gamma y_{i,t+3}+x_{it+2}\beta_2+\tau_i^*)-\rho\varepsilon_{it+3}=-y_{it+2}+e_{t+2}^b$$

$$(y_{i,t+1}\,|\,t+3)\quad y_{it+1}=\gamma y_{i,t+2}+x_{it+1}\beta_2+\tau_i^*+\varepsilon_{it+1}$$

$$-(\gamma^2 y_{i,t+3}+(\gamma x_{it+2}+x_{it+1})\beta_2+(\gamma+1)\tau_i^*)-(\gamma\rho_2+\rho_2^2)\varepsilon_{it+3}=$$

$$-y_{it+1}+(\gamma+\rho_2)e_{t+2}^b+e_{t+1}^b$$

$$(y_{i,t}\,|\,t+3)\quad y_{it}=\gamma y_{i,t+1}+x_{it}\beta_2+\tau_i^*+\varepsilon_{it}$$

$$-(\gamma^3 y_{i,t+3}+(\gamma^2 x_{it+2}+\gamma x_{it+1}+x_{it})\beta_2+(\gamma^2+\gamma+1)\tau_i^*)-(\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)\varepsilon_{it+3}=$$

$$-y_{it}+(\gamma^2+\gamma\rho_2+\rho_2^2)e_{it+2}^b+(\gamma+\rho_2)e_{it+1}^b+e_{it}^b$$

This is equivalent to the following regression with $y_{i,t}$, $y_{i,t+1}$ and $y_{i,t+2}$ as parameters:

$$\begin{Bmatrix} y_i^1 \\ y_i^2 \\ y_i^3 \\ y_i^4 \\ y_i^5 \\ y_i^6 \end{Bmatrix}=\begin{Bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{Bmatrix}\begin{Bmatrix} y_{it} \\ y_{it+1} \\ y_{it+2} \end{Bmatrix}+\begin{Bmatrix} e_{it}^f \\ (\gamma+\rho_2)e_{it}^f+e_{it+1}^f \\ (\gamma^2+\gamma\rho_2+\rho_2^2)e_{it}^f+(\gamma+\rho_2)e_{it+1}^f+e_{it+2}^f \\ e_{t+2}^b \\ (\gamma+\rho_2)e_{t+2}^b+e_{t+1}^b \\ (\gamma^2+\gamma\rho_2+\rho_2^2)e_{it+2}^b+(\gamma+\rho_2)e_{it+1}^b+e_{it}^b \end{Bmatrix}$$

or

$$y_i=\Gamma_3\begin{Bmatrix} y_{it} \\ y_{it+1} \\ y_{it+2} \end{Bmatrix}+e_i$$

where

$$y_i=\begin{Bmatrix} -(\gamma y_{i,t-1}+x_{it}\beta_2+\tau_i^*)-\rho^*\varepsilon_{it-1} \\ -(\gamma^2 y_{i,t-1}+(\gamma x_{it}+x_{it+1})\beta_2+(\gamma+1)\tau_i^*)-\rho^*(\gamma+\rho_2)\varepsilon_{it-1} \\ -(\gamma^3 y_{i,t-1}+(\gamma^2 x_{it}+\gamma x_{it+1}+x_{it+2})\beta_2+(\gamma^2+\gamma+1)\tau_i^*))-\rho^*(\gamma^2+\gamma\rho_2+\rho_2^2)\varepsilon_{it-1} \\ -(\gamma y_{i,t+3}+x_{it+2}\beta_2+\tau_i^*)-\rho\varepsilon_{it+3} \\ -(\gamma^2 y_{i,t+3}+(\gamma x_{it+2}+x_{it+1})\beta_2+(\gamma+1)\tau_i^*)-(\gamma\rho_2+\rho_2^2)\varepsilon_{it+3} \\ -(\gamma^3 y_{i,t+3}+(\gamma^2 x_{it+2}+\gamma x_{it+1}+x_{it})\beta_2+(\gamma^2+\gamma+1)\tau_i^*)-(\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)\varepsilon_{it+3} \end{Bmatrix}$$

$$= -\begin{Bmatrix} \gamma y_{i,t-1} \\ \gamma^2 y_{i,t-1} \\ \gamma^3 y_{i,t-1} \\ \gamma y_{i,t+3} \\ \gamma^2 y_{i,t+3} \\ \gamma^3 y_{i,t+3} \end{Bmatrix} - \left[\begin{Bmatrix} x_{it} \\ x_{it+1} \\ x_{it+2} \\ x_{it+2} \\ x_{it+1} \\ x_{it} \end{Bmatrix} - \gamma \begin{Bmatrix} 0 \\ x_{it} \\ x_{it+1} \\ 0 \\ x_{it+2} \\ x_{it+1} \end{Bmatrix} - \gamma^2 \begin{Bmatrix} 0 \\ 0 \\ x_{it} \\ 0 \\ 0 \\ x_{it+2} \end{Bmatrix} \right] \beta_2 - \tau_i^* \begin{Bmatrix} 1 \\ \gamma+1 \\ \gamma^2+\gamma+1 \\ 1 \\ \gamma+1 \\ \gamma^2+\gamma+1 \end{Bmatrix} - \begin{Bmatrix} \rho^* \varepsilon_{it-1} \\ \rho^*(\gamma+\rho_2)\varepsilon_{it-1} \\ \rho^*(\gamma^2+\gamma\rho_2+\rho_2^2)\varepsilon_{it-1} \\ \rho\varepsilon_{it+3} \\ (\gamma\rho_2+\rho_2^2)\varepsilon_{it+3} \\ (\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)\varepsilon_{it+3} \end{Bmatrix}$$

and $e_i \sim N(0,\Sigma_3)$, with $\Sigma_3 = \sigma_2^2 \begin{Bmatrix} \Sigma_3^{11} & \bar{0} \\ \bar{0} & \Sigma_3^{11} \end{Bmatrix}$ where

$$\Sigma_3^{11} = \begin{Bmatrix} 1 & (\gamma+\rho_2) & (\gamma^2+\gamma\rho_2+\rho_2^2) \\ (\gamma+\rho_2) & (\gamma+\rho_2)^2+1 & (\gamma+\rho_2)(\gamma^2+\gamma\rho_2+\rho_2^2)+(\gamma+\rho_2) \\ (\gamma^2+\gamma\rho_2+\rho_2^2) & (\gamma+\rho_2)(\gamma^2+\gamma\rho_2+\rho_2^2)+(\gamma+\rho_2) & (\gamma^2+\gamma\rho_2+\rho_2^2)^2+(\gamma+\rho_2)^2+1 \end{Bmatrix}$$

For computational simplicity, this matrix is decomposed as

$$\Sigma_3^{11} = \Sigma_3^{11½} \Sigma_3^{11½^T}$$

where

$$\Sigma_3^{11½} = \begin{Bmatrix} 1 & 0 & 0 \\ (\gamma+\rho_2) & 1 & 0 \\ (\gamma^2+\gamma\rho_2+\rho_2^2) & (\gamma+\rho_2) & 1 \end{Bmatrix}$$

As usual pre-multiply by $\Sigma_3^{-½}$ to obtain

$$y_i^{\Sigma} = \Gamma_3^{\Sigma} \begin{Bmatrix} y_{it} \\ y_{it+1} \\ y_{it+2} \end{Bmatrix} + e_i$$

And again, this is a standard OLS with six observations and N(0,1) error terms. The distribution for the parameters $y_{i,t}$, $y_{i,t+1}$ and $y_{i,t+2}$ is

$$(y_{it}, y_{it+1}, y_{it+2}) \sim N\left((\Gamma_3^{\Sigma'} \Gamma_3^{\Sigma})^{-1} \Gamma_3^{\Sigma} y_i^{\Sigma}, (\Gamma_3^{\Sigma'} \Gamma_3^{\Sigma})^{-1} \right)$$

4. Four missing observations: $y_{i,t}$ **,** $y_{i,t+1}, y_{i,t+2}$ **and** $y_{i,t+3}$ **are missing,** $y_{i,t-1}$ **and** $y_{i,t+3}$ **are known**

The case of three missing observations results in a regression with eight observations. First, by rearranging (2f) and to express $y_{i,t}$, $y_{i,t+1}, y_{i,t+2}$ and $y_{i,t+3}$ conditional on t-1:

$(y_{i,t} \mid t-1)$ $\quad y_{it} = \gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$

$$y_{it} = \gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$$

$$-(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) - \rho^* \varepsilon_{it-1} = -y_{it} + e_{it}^f$$

$(y_{i,t+1} \mid t-1)$ $\quad y_{it+1} = \gamma y_{i,t} + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}$

$$y_{it+1} = \gamma(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}) + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}$$

$$-(\gamma^2 y_{i,t-1} + (\gamma x_{it} + x_{it+1})\beta_2 + (\gamma+1)\tau_i^*) - \rho^*(\gamma+\rho_2)\varepsilon_{it-1} =$$
$$-y_{it+1} + (\gamma+\rho_2)e_{it}^f + e_{it+1}^f$$

$$(y_{i,t+2} \mid t-1) \quad y_{it+2} = \gamma(\gamma y_{i,t} + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}) + x_{it+2}\beta_2 + \tau_i^* + \varepsilon_{it+2}$$
$$-(\gamma^3 y_{i,t-1} + (\gamma^2 x_{it} + \gamma x_{it+1} + x_{it+2})\beta_2 + (\gamma^2+\gamma+1)\tau_i^*)) - \rho^*(\gamma^2 + \gamma\rho_2 + \rho_2^2)\varepsilon_{it-1}$$
$$= -y_{it+2} + (\gamma^2 + \gamma\rho_2 + \rho_2^2)e_{it}^f + (\gamma+\rho_2)e_{it+1}^f + e_{it+2}^f$$

$$(y_{i,t+3} \mid t-1) \quad y_{it+3} = \gamma(\gamma y_{i,t+1} + x_{it+2}\beta_2 + \tau_i^* + \varepsilon_{it+2}) + x_{it+3}\beta_2 + \tau_i^* + \varepsilon_{it+3}$$
$$-(\gamma^4 y_{i,t-1} + (\gamma^3 x_{it} + \gamma^2 x_{it+1} + \gamma x_{it+2} + x_{it+3})\beta_2 + (\gamma^3+\gamma^2+\gamma+1)\tau_i^*))$$
$$-\rho^*(\gamma^3 + \gamma^2\rho_2^0 + \gamma\rho_2^2 + \rho_2^3)\varepsilon_{it-1} =$$
$$-y_{it+2} + (\gamma^3 + \gamma^2\rho_2 + \gamma\rho_2^2 + \rho_2^3)e_{it}^f + (\gamma^2 + \gamma\rho_2 + \rho_2^2)e_{it+1}^f + (\gamma+\rho_2)e_{it+2}^f + e_{it+3}^f$$

Rearranging (2b) to express the missing observations conditional on t+2:

$$(y_{i,t+3} \mid t+4) \quad y_{it+3} = \gamma y_{i,t+4} + x_{it+3}\beta_2 + \tau_i^* + \varepsilon_{it+3}$$
$$-(\gamma y_{i,t+4} + x_{it+3}\beta_2 + \tau_i^*) - \rho\varepsilon_{it+4} = -y_{it+3} + e_{t+3}^b$$

$$(y_{i,t+2} \mid t+4) \quad y_{it+2} = \gamma y_{i,t+3} + x_{it+2}\beta_2 + \tau_i^* + \varepsilon_{it+2}$$
$$-(\gamma^2 y_{i,t+4} + (\gamma x_{it+3} + x_{it+2})\beta_2 + (\gamma+1)\tau_i^*) - (\gamma\rho_2 + \rho_2^2)\varepsilon_{it+4} =$$
$$-y_{it+2} + (\gamma+\rho_2)e_{t+3}^b + e_{t+2}^b$$

$$(y_{i,t+1} \mid t+4) \quad y_{it+1} = \gamma y_{i,t+2} + x_{it+1}\beta_2 + \tau_i^* + \varepsilon_{it+1}$$
$$-(\gamma^3 y_{i,t+4} + (\gamma^2 x_{it+3} + \gamma x_{it+2} + x_{it+1})\beta_2 + (\gamma^2+\gamma+1)\tau_i^*) - (\gamma^2\rho_2 + \gamma\rho_2^2 + \rho_2^3)\varepsilon_{it+4} =$$
$$-y_{it+1} + (\gamma^2 + \gamma\rho_2 + \rho_2^2)e_{it+3}^b + (\gamma+\rho_2)e_{it+2}^b + e_{it+1}^b$$

$$(y_{i,t} \mid t+4) \quad y_{it} = \gamma y_{i,t+1} + x_{it}\beta_2 + \tau_i^* + \varepsilon_{it}$$
$$-(\gamma^4 y_{i,t+4} + (\gamma^3 x_{it+3} + \gamma^2 x_{it+2} + \gamma x_{it+1} + x_{it})\beta_2 + (\gamma^3+\gamma^2+\gamma+1)\tau_i^*)$$
$$-(\gamma^3\rho_2 + \gamma^2\rho_2^2 + \gamma\rho_2^3 + \rho_2^4)\varepsilon_{it+4} =$$
$$-y_{it} + (\gamma^3 + \gamma^2\rho_2 + \gamma\rho_2^2 + \rho_2^3)e_{it+2}^b + (\gamma^2 + \gamma\rho_2 + \rho_2^2)e_{it+2}^b + (\gamma+\rho_2)e_{it+1}^b + e_{it}^b$$

This is equivalent to the following regression with $y_{i,t}$, $y_{i,t+1}$, $y_{i,t+2}$ and $y_{i,t+3}$ as parameters:

$$\begin{Bmatrix} y_i^1 \\ y_i^2 \\ y_i^3 \\ y_i^4 \\ y_i^5 \\ y_i^6 \\ y_i^7 \\ y_i^8 \end{Bmatrix} = \begin{Bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{Bmatrix} \begin{Bmatrix} y_{it} \\ y_{it+1} \\ y_{it+2} \\ y_{it+3} \end{Bmatrix} +$$

$$\begin{Bmatrix} e_{it}^f \\ (\gamma+\rho_2)e_{it}^f + e_{it+1}^f \\ (\gamma^2+\gamma\rho_2+\rho_2^2)e_{it}^f + (\gamma+\rho_2)e_{it+1}^f + e_{it+2}^f \\ (\gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)e_{it}^f + (\gamma^2+\gamma\rho_2+\rho_2^2)e_{it+1}^f + (\gamma+\rho_2)e_{it+2}^f + e_{it+3}^f \\ e_{t+3}^b \\ (\gamma+\rho_2)e_{t+3}^b + e_{t+2}^b \\ (\gamma^2+\gamma\rho_2+\rho_2^2)e_{it+3}^b + (\gamma+\rho_2)e_{it+2}^b + e_{it+1}^b \\ (\gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)e_{it}^f + (\gamma^2+\gamma\rho_2+\rho_2^2)e_{it+1}^f + (\gamma+\rho_2)e_{it+2}^f + e_{it+3}^f \end{Bmatrix}$$

or

$$y_i = \Gamma_4 \begin{Bmatrix} y_{it} \\ y_{it+1} \\ y_{it+2} \\ y_{it+3} \end{Bmatrix} + e_i$$

where

$$\begin{Bmatrix} y_i^1 \\ y_i^2 \\ y_i^3 \\ y_i^4 \\ y_i^5 \\ y_i^6 \\ y_i^7 \\ y_i^8 \end{Bmatrix} = \begin{Bmatrix} -(\gamma y_{i,t-1} + x_{it}\beta_2 + \tau_i^*) - \rho_2\varepsilon_{it-1} \\ -(\gamma^2 y_{i,t-1} + (\gamma x_{it} + x_{it+1})\beta_2 + (\gamma+1)\tau_i^*) - (\gamma\rho_2+\rho_2^2)\varepsilon_{it-1} \\ -(\gamma^3 y_{i,t-1} + (\gamma^2 x_{it} + \gamma x_{it+1} + x_{it+2})\beta_2 + (\gamma^2+\gamma+1)\tau_i^*)) - (\gamma^2\rho_2 + \gamma\rho_2^2 + \rho_2^3)\varepsilon_{it-1} \\ -(\gamma^4 y_{i,t-1} + (\gamma^3 x_{it} + \gamma^2 x_{it+1} + \gamma x_{it+2} + x_{it+3})\beta_2 + (\gamma^3+\gamma^2+\gamma+1)\tau_i^*)) - (\gamma^3\rho_2 + \gamma^2\rho_2^2 + \gamma\rho_2^3 + \rho_2^4)\varepsilon_{it-1} \\ -(\gamma y_{i,t+4} + x_{it+3}\beta_2 + \tau_i^*) - \rho\varepsilon_{it+4} \\ -(\gamma^2 y_{i,t+4} + (\gamma x_{it+3} + x_{it+2})\beta_2 + (\gamma+1)\tau_i^*) - (\gamma\rho_2+\rho_2^2)\varepsilon_{it+4} \\ -(\gamma^3 y_{i,t+4} + (\gamma^2 x_{it+3} + \gamma x_{it+2} + x_{it+1})\beta_2 + (\gamma^2+\gamma+1)\tau_i^*) - (\gamma^2\rho_2 + \gamma\rho_2^2 + \rho_2^3)\varepsilon_{it+4} \\ -(\gamma^4 y_{i,t+4} + (\gamma^3 x_{it+3} + \gamma^2 x_{it+2} + \gamma x_{it+1} + x_{it})\beta_2 + (\gamma^3+\gamma^2+\gamma+1)\tau_i^*) - (\gamma^3\rho_2 + \gamma^2\rho_2^2 + \gamma\rho_2^3 + \rho_2^4)\varepsilon_{it+4} \end{Bmatrix}$$

$$= -\left\{\begin{matrix} \gamma y_{i,t-1} \\ \gamma^2 y_{i,t-1} \\ \gamma^3 y_{i,t-1} \\ \gamma^4 y_{i,t-1} \\ \gamma y_{i,t+4} \\ \gamma^2 y_{i,t+4} \\ \gamma^3 y_{i,t+4} \\ \gamma^4 y_{i,t+4} \end{matrix}\right\} - \left[\left\{\begin{matrix} x_{it} \\ x_{it+1} \\ x_{it+2} \\ x_{it+3} \\ x_{it+3} \\ x_{it+2} \\ x_{it+1} \\ x_{it} \end{matrix}\right\} - \gamma\left\{\begin{matrix} 0 \\ x_{it} \\ x_{it+1} \\ x_{it+2} \\ 0 \\ x_{it+3} \\ x_{it+2} \\ x_{it+1} \end{matrix}\right\} - \gamma^2\left\{\begin{matrix} 0 \\ 0 \\ x_{it} \\ x_{it+1} \\ 0 \\ 0 \\ x_{it+3} \\ x_{it+2} \end{matrix}\right\} - \gamma^3\left\{\begin{matrix} 0 \\ 0 \\ 0 \\ x_{it} \\ 0 \\ 0 \\ 0 \\ x_{it+3} \end{matrix}\right\}\right]\beta_2 -$$

$$\tau_i^*\left\{\begin{matrix} 1 \\ \gamma+1 \\ \gamma^2+\gamma+1 \\ \gamma^3+\gamma^2+\gamma+1 \\ 1 \\ \gamma+1 \\ \gamma^2+\gamma+1 \\ \gamma^3+\gamma^2+\gamma+1 \end{matrix}\right\} - \left\{\begin{matrix} \rho_2\varepsilon_{it-1} \\ (\gamma\rho_2+\rho_2^2)\varepsilon_{it-1} \\ (\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)\varepsilon_{it-1} \\ (\gamma^3\rho_2+\gamma^2\rho_2^2+\gamma\rho_2^3+\rho_2^4)\varepsilon_{it-1} \\ \rho\varepsilon_{it+4} \\ (\gamma\rho_2+\rho_2^2)\varepsilon_{it+4} \\ (\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)\varepsilon_{it+4} \\ (\gamma^3\rho_2+\gamma^2\rho_2^2+\gamma\rho_2^3+\rho_2^4)\varepsilon_{it+4} \end{matrix}\right\}$$

and $\varepsilon_i \sim N(0,\Sigma_4)$ with $\Sigma_4 = \sigma_2^2\left\{\begin{matrix} \Sigma_4^{11} & \bar{0} \\ \bar{0} & \Sigma_4^{11} \end{matrix}\right\}$ where

$$\Sigma_4^{11} = \left\{\begin{matrix} 1 & \gamma+\rho_2 \\ \gamma+\rho_2 & (\gamma+\rho_2)^2+1 \\ \gamma^2+\gamma\rho_2+\rho_2^2 & (\gamma+\rho_2)(\gamma^2+\gamma\rho_2+\rho_2^2)+(\gamma+\rho_2) \\ \gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3 & (\gamma+\rho_2)(\gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)+(\gamma^2+\gamma\rho_2+\rho_2^2) \end{matrix}\right\}$$

$$\| \left\{\begin{matrix} \gamma^2+\gamma\rho_2+\rho_2^2 \\ (\gamma+\rho_2)(\gamma^2+\gamma\rho_2+\rho_2^2)+(\gamma+\rho_2) \\ (\gamma^2+\gamma\rho_2+\rho_2^2)^2+(\gamma+\rho_2)^2+1 \\ (\gamma^2+\gamma\rho_2+\rho_2^2)(\gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3+\gamma+\rho_2)+(\gamma+\rho_2)(\gamma^2+\gamma\rho_2+\rho_2^2)+(\gamma+\rho_2) \end{matrix}\right\}$$

$$\| \left\{\begin{matrix} \gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3 \\ (\gamma+\rho_2)(\gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)+(\gamma^2+\gamma\rho_2+\rho_2^2) \\ (\gamma^2+\gamma\rho_2+\rho_2^2)(\gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3+\gamma+\rho_2)+(\gamma+\rho_2)(\gamma^2+\gamma\rho_2+\rho_2^2)+(\gamma+\rho_2)+(\gamma+\rho_2) \\ (\gamma^3+\gamma^2\rho_2+\gamma\rho_2^2+\rho_2^3)^2+(\gamma^2+\gamma\rho_2+\rho_2^2)^2+(\gamma+\rho_2)^2+1 \end{matrix}\right\}$$

For computational simplicity, this matrix is decomposed as

$$\Sigma_4^{11} = \Sigma_4^{11^{1/2}}\Sigma_4^{11^{1/2}T}$$

where

$$\Sigma_4^{11½} = \begin{Bmatrix} 1 & 0 & 0 & 0 \\ \gamma + \rho_2 & 1 & 0 & 0 \\ \gamma^2 + \gamma\rho_2 + \rho_2^2 & \gamma + \rho_2 & 1 & 0 \\ \gamma^3 + \gamma^2\rho_2 + \gamma\rho_2^2 + \rho_2^3 & \gamma^2 + \gamma\rho_2 + \rho_2^2 & \gamma + \rho_2 & 1 \end{Bmatrix}$$

Pre-multiply by $\Sigma_4^{-½}$ to obtain

$$y_i^{\Sigma} = \Gamma_4^{\Sigma} \begin{Bmatrix} y_{it} \\ y_{it+1} \\ y_{it+2} \\ y_{it+3} \end{Bmatrix} + e_i$$

This is a standard OLS with six observations and N(0,1) error terms where the joint distribution for the parameters $y_{i,t}$, $y_{i,t+1}$, $y_{i,t+2}$ and $y_{i,t+3}$ is given by

$$(y_{it}, y_{it+1}, y_{it+2}) \sim N\left((\Gamma_4^{\Sigma'}\Gamma_4^{\Sigma})^{-1}\Gamma_4^{\Sigma} y_i^{\Sigma}, (\Gamma_4^{\Sigma'}\Gamma_4^{\Sigma})^{-1}\right)$$