Working paper
1/2012

Dummy-encoding Inherently Collinear Variables

Simen Gaure

Stiftelsen Frischsenteret for samfunnsøkonomisk forskning
Ragnar Frisch Centre for Economic Research

Working paper 1/2012

Dummy-encoding Inherently Collinear Variables

Simen Gaure

Abstract: This note is the result of trying to spell out what happens when we dummy-encode a set of variables which is known to be multicollinear at the outset. There seems to be a range of approaches in the literature, this is an attempt to collect the fundamental foot-work in a single note. We start out with a self-contained presentation of the general treatment of exact multicollinearities, with estimable functions and estimation constraints on dummies. And provide an example at the end. Much of this has been discussed in the context of age-period-cohort analysis in various other places.

Contact: www.frisch.uio.no

Report from the project "Absenteeism in Norway – Causes, Consequences, and Policy Implications" (187924/S20) funded by the Norwegian Research Council

ISBN 978-82-7988-108-7
ISSN 1501-9241

Dummy-encoding Inherently Collinear Variables

Simen Gaure

Abstract

This note is the result of trying to spell out what happens when we dummy-encode a set of variables which is known to be multicollinear at the outset. There seems to be a range of approaches in the literature, this is an attempt to collect the fundamental foot-work in a single note.

We start out with a self-contained presentation of the general treatment of exact multicollinearities, with estimable functions and estimation constraints on dummies. And provide an example at the end.

Much of this has been discussed in the context of age-period-cohort analysis in [2] and various other places.

1 Introduction

Let D be a $(n \times k)$-matrix, the data matrix. n is the number of individuals, k is the number of explanatory covariates. D possibly includes a constant column. For any matrix A, let A' denote its transpose. I denotes the identity matrix (of appropriate size).

Definition 1.1. A set of multicollinearities (or just *a collinearity*) is an $(r \times k)$-matrix M of rank $r < k$, with the property $MD' = 0$. We also assume M is a maximal set of collinearities (i.e. whenever $XD' = 0$, then $\operatorname{rank}(X) \leq r$). Moreover, for the sake of clarity, this property is inherent in the covariates, not a spurious property of the particular data set. That is, we know in advance that there is collinearity in the data.

Example 1.2. For a fully dummy-encoded variable v with k values, $\{v_i\}_{i=1}^k$, we have $\sum_{i=1}^k v_i - 1 = 0$ for every observation (because at any time, exactly one of the v_i's is 1, whereas the others are zero), thus our collinearity M is the $1 \times k$-matrix $M = \begin{bmatrix} 1 & 1 & \dots & 1 & -1 \end{bmatrix}$, where the last entry corresponds to the constant covariate.

We have a function Y which in this note is of the form $Y(B) = g(B'D')$ where B is a parameter vector (a column vector). $B'D'$ is the vector of *indices*, (each element is often called $\beta' X_i$), g is assumed to act elementwise on this vector. We may think of $Y(B)$ as a vector of (predicted) left-hand sides. Estimation of β's is to find a B so that $Y(B)$ matches the observed Y in a best possible way (typically by maximum likelihood estimation). In the presence of a collinearity, we need to put some restriction on the parameters in order to do a rational estimation. It's not that the collinearity introduces bias, but it introduces non-identifiability, and a degenerate Hessian which makes both the estimation numerically infeasible and the estimation of standard errors quite complicated.

2 Inherent multicollinearity

Assume we have a candidate B for the parameter vector. Let X be a $(r \times 1)$ matrix; then $Y(B - M'X) = g((B' - X'M)D') = g(B'D' - X'MD') = g(B'D') = Y(B)$. Thus B is not identified. We may shift B by $M'X$ (with an arbitrary X) and still get the same left-hand sides.

In Theorem 3.1 of [2], this is formulated as translation by eigenvectors of $D'D$ corresponding to the eigenvalue 0.

Lemma 2.1. *The set of vectors of the form $M'X$ is the same as the eigenspace of $D'D$ corresponding to the eigenvalue 0. Thus, the rows of M span the null-space of $D'D$ (or D).*

Proof. To see this, note that $D'DM'X = D'(MD')'X = 0$ (because $MD' = 0$ by definition), thus $M'X$ is an eigenvector of $D'D$ for the eigenvalue 0. Conversely, if V satisfies $D'DV = 0$, let $X = (MM')^{-1}MV$, we have that $W = V - M'X = (I - M'(MM')^{-1}M)V$ is the orthogonal projection of V onto the the null space of M, i.e. W is orthogonal to every row of M. Since we have $D'DW = 0$, we have $W'D'DW = (DW)'(DW) = 0$, thus $DW = 0$. That is, we may add W' as a new row to M to get a larger collinearity matrix, but since M by definition has maximal rank, we must have $W = 0$, thus $V = M'X$. $\square$

Remark 2.2. In this note we have supposed that we know the collinearity, i.e. the matrix M, which is in some understandable form. In the case that there is an unknown multicollinearity in the data, we may attempt to find a simple form by finding a suitable basis for the null-space of D. This basis may be used as the rows of M. One simple method for doing this is to do a Cholesky-decomposition (or QR) with pivoting of the matrix $D'D$ (or the Hessian). Keep only the rows with (close to) non-zero pivots, split the columns into a part with non-zero pivots (call it A), the other columns is called B. Then solve the system $AX = B$. The variables corresponding to the columns of B may be written as linear combinations of the other variables, with the columns of X as weights. The structure of these weights may shed some light on the nature of the multicollinearity. This method is along the lines of [1].

Definition 2.3. Two parameter vectors B_1 and B_2 are said to be *equivalent* (under the collinearity M) if $B_1 - B_2 \in \mathrm{R}(M)$ where $\mathrm{R}(M)$ is the row-space of M. Equivalently, if there exists a vector X with $B_1 - B_2 = M'X$.

Thus two equivalent parameter vectors B_1 and B_2 will predict identical left-hand sides: $Y(B_1) = Y(B_2)$ and are thus indistinguishable in this perspective.

We could in principle insist that parameter vectors don't live in $\mathbb{R}^k$, but rather in the quotient vector space $\mathbb{R}^k/\mathrm{R}(M)$ which is isomorphic to $\mathbb{R}^{k-r}$; this would make them unique. However, we have chosen to approach this problem from a slightly more practical angle.

Definition 2.4. A *restriction* on the parameters (compatible with a collinearity M) is an $(r \times k)$-matrix T with the property $\mathrm{rank}(TM') = r$. (Or equivalently, TM' is invertible.)

For any given collinearity M there always exists at least one restriction. The canonical choice for the restriction is $T = M$, but its interpretation is not always an intuitive one.

We can now show that for any restriction T, and any parameter vector B, there's a unique parameter vector in the kernel of T equivalent with B. Thus a restriction may be used as a constraint when estimating.

Lemma 2.5. *Given a collinearity M, a parameter vector B_1 and a restriction T. Then there exists a unique parameter vector B_2 equivalent with B_1 and satisfying $TB_2 = 0$. It's given by*

$$B_2 = (I - M'(TM')^{-1}T)B_1.$$

In particular, if $T = M$, then B_2 is the projection of B_1 onto the null-space of M.

Proof. We first show that B_2 as given is equivalent with B_1, and that $TB_2 = 0$.

We have

$$\begin{aligned} TB_2 &= T(I - M'(TM')^{-1}T)B_1 \\ &= TB_1 - TM'(TM')^{-1}TB_1 = TB_1 - (TM')(TM')^{-1}TB_1 \\ &= TB_1 - TB_1 \\ &= 0 \end{aligned}$$

For the first assertion, that B_1 is equivalent with B_2, it's sufficient to prove that $B_1 - B_2 = M'X$ for some X, but we have, by construction of B_2, that $B_1 - B_2 = M'(TM')^{-1}TB_1$, thus $X = (TM')^{-1}TB_1$ will suffice.

We then show that B_2 is unique. Assume there's another B equivalent with B_2 and with $TB = 0$. We have $B - B_2 = M'X$ for some X, applying T to this equation yields $T(B - B_2) = TM'X$. Now, since we have $TB = TB_2 = 0$ this reduces to $TM'X = 0$. By Definition 2.4 the $(r \times r)$-matrix TM' is invertible. This yields $X = 0$, so $B - B_2 = 0$, thus B_2 is unique.

In case $T = M$, we know from general theory that the projection onto the row-space of M is given by $M'(MM')^{-1}M$, thus $I - M'(MM')^{-1}M$ is the projection onto its orthogonal complement, which is the null-space. □

In other words, if we assume the model and data otherwise are sound, then B is identified up to translation by $M'X$. That is, B's equivalence class under translation by $M'X$ is identified. The whole interpretation exercise under inherent multicollinearity rests on how well we are able to understand what this equivalence class looks like, i.e. which aspect of the vectors in this class is the same throughout the class.

Observation 2.6. With a linear relation $MD' = 0$ between the covariates, a parameter set is only identified up to translation by vectors in the row space of M.

Example 2.7. Continuing example 1.2, we may e.g. pick as a restriction the customary one which sets one of the coefficients to zero, e.g. the first, $T = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ and note that the (1×1)-matrix $TM' = [1]$ is invertible.

That is, a *restriction* picks a unique vector in each equivalence class. Although it doesn't really matter *which* restriction we choose (we may easily "change" the restriction after estimation, by the above lemma it's just a linear change of variables), it may be feasible to choose one which makes the resulting parameters (and covariances) easy to interpret. In some cases it's easy to implement a restriction such that each of the estimated parameters has their own meaningful interpretation independent of the others.

Definition 2.8. An *interpretation* (under the collinearity M) is a $(d \times k)$-matrix S such that $SM' = 0$. Likewise, a linear combination of parameters is said to be *interpretable* if its matrix is an *interpretation*.

Remark 2.9. A more common name for *interpretation* is (linear) *estimable function*, though in our context *interpretation* creeps smoothly into our intuition.

Lemma 2.10. *If M is a collinearity, S is an interpretation, and B_1 and B_2 are equivalent parameter vectors, then $SB_1 = SB_2$.*

Proof. We have by Definition 2.3 that $B_1 - B_2 = M'X$ for some X, thus $SB_1 - SB_2 = S(B_1 - B_2) = SM'X = 0$ by Definition 2.8. □

Remark 2.11. By definition, the row-space of an interpretation is orthogonal to the row-space of M. Thus, it's contained in the null-space of M. Moreover, any vector in the null-space of M is clearly an interpretation (when viewed as a $1 \times k$ matrix), thus by the rank-nullity theorem, $\operatorname{rank}(S) \le k - r$. This loosely says that no more than $k - r$ parameters may be independently interpreted. (Which is just another way of saying that the parameter vectors live in something isomorphic to $\mathbb{R}^{k-r}$).

Indeed, by rank-maximiality of M, we note that the row-space of S is contained in the row-space of D.

Remark 2.12. Given a parameter vector B, the interpretation SB only depends on B's equivalence class, and is thus independent of parameter restrictions. The *interpretation* dimension d may be 1 if we e.g. want to interpret only the sum of the parameters, or it may be quite large if we e.g. want to interpret every difference of two arbitrary parameters. A particularly large and useless *interpretation* is D, the data matrix. (We know from Definition 1.1 that $MD' = 0$). We're obviously interested in something smaller. A restriction T is never an *interpretation*, since by Definition 2.4 we have $\operatorname{rank}(TM') = r \ne 0$. (That the restriction can't be estimable is also noted at the bottom of p. 2794 of [2].)

Example 2.13. Continuing example 1.2, the customary *interpretation* is the difference between each coefficient and the reference coefficient (which we chose as the first one), thus our full *interpretation* (we now discard the intercept by setting the last column to zero) is

$$S = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ -1 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

which when multiplied into any parameter vector yields the differences between each parameter and the first one (which happens to be zero with the particular restriction in example 2.7, so that each estimated parameter has its own interpretation.) We easily see that $SM' = 0$.

A more straightforward definition of *interpretation* would be that it's some linear combination of β's which is independent of the parameter restriction.

To distinguish this formal definition of *interpretation* and *interpretable* from the more intuitive notions of the same name, we *emphasize* the former usage.

Remark 2.14. An important thing to note is that a *restriction* is merely a device which makes it possible to estimate a parameter vector; a representative of its equivalence class under the collinearity relation. The choice of restriction has no influence on the predictive properties; all vectors in the equivalence class predict the same left hand side. Thus, we may settle for the canonical restriction $T = M$ as linear constraints on the parameters. On the other hand, an *interpretation* is something we apply to the estimated parameter vector, and it will yield the same *interpreted* values, an invariant of the equivalence class, independently of which restriction we picked in the first place. However, as seen from the previous example, it's sometimes possible to pick a restriction which makes the *interpretation* exercise trivial. And of course, when applying a non-trivial *interpretation*, one must of course adjust the standard errors (i.e. the covariance matrix) accordingly.

Remark 2.15. We have not talked about how restriction change affects the standard errors. Since the parameter change in Lemma 2.5 is linear, the Jacobian will be the constant $I - M'(TM')^{-1}T$. Thus, given the covariance matrix for B_1 we may easily compute it for B_2 (save for numerical inaccuracies).

3 An example

Example 3.1. Here's the motivating example for this note. Say we have covariates c, a, y (cohort, age, year) with the deterministic relation $c + a = y$. We dummy-encode the data completely. I.e. say c, a and y are integers, $c \in [\ell_c, u_c]$, $a \in [\ell_a, u_a]$ and $y \in [\ell_y, u_y]$. We create sets of dummies $(c_{\ell_c}, \dots, c_{u_c})$, $(a_{\ell_a}, \dots, a_{u_a})$ and $(y_{\ell_y}, \dots, y_{u_y})$. Such that $c_i = 1$ when $i = c$, and zero otherwise. Similarly with a and y. This example may also be found in [2].

As in [2] we get four relations:

$$\sum_{i=\ell_c}^{u_c} c_i - 1 = 0$$

$$\sum_{i=\ell_a}^{u_a} a_i - 1 = 0$$

$$\sum_{i=\ell_y}^{u_y} y_i - 1 = 0$$

$$\sum_{i=\ell_c}^{u_c} i c_i + \sum_{i=\ell_a}^{u_a} i a_i - \sum_{i=\ell_y}^{u_y} i y_i = 0$$

Thus our collinearity matrix is

$$M = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & -1 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 & -1 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 & -1 \\ \ell_c & \cdots & u_c & \ell_a & \cdots & u_a & -\ell_y & \cdots & -u_y & 0 \end{bmatrix}.$$

This matrix is somewhat hard to interpret, but it might be useful for studying what kind of restriction we should (or should not!) implement. If we pick one reference for each dummy-group, and some fourth reference for the joint relation, we can e.g. have a restriction T like:

$$T = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

We get

$$TM' = \begin{bmatrix} 1 & 0 & 0 & \ell_c \\ 0 & 1 & 0 & \ell_a \\ 0 & 0 & 1 & -\ell_y \\ 0 & 0 & 1 & -u_y \end{bmatrix}$$

which typically has rank 4 (Subtract the third row from the fourth to get a triangular matrix with $\ell_y - u_y$ in the lower right corner. It's different from 0 if we have more than one value for the y covariate).

How do we *interpret* a parameter vector under the relations M? The best thing to do is probably to try to understand the equivalence class. For a moment, forget about the intercept (i.e, remove the last column in M and T). So, what can we do with the parameter set without moving out of the equivalence class?

Assume we force one parameter in each dummy-group to 0. For simplicity we take the first one, i.e. $\beta_{c,\ell_c} = \beta_{a,\ell_a} = \beta_{y,\ell_a} = 0$. The first 3 rows of M vanish and we're left with

$$M = \begin{bmatrix} 0 & 1 & \cdots & u_c - \ell_c & 0 & \cdots & u_a - \ell_a & 0 & \cdots & \ell_y - u_y \end{bmatrix}$$

We have now attempted the interpretation of β's to be the customary *relative to the reference* interpretation. But still we have only identified a certain equivalence class, not a parameter vector. Thus we don't have an *interpretation.* We may still shift the β along the line λM:

$$L_\lambda = (0, \lambda, 2\lambda, \ldots, s_c\lambda, 0, \lambda, 2\lambda, \ldots, s_a\lambda, 0, -\lambda, -2\lambda, \ldots, -s_y\lambda).$$

(where s_c, s_a and s_y is the number of dummies in each group.)

Note that in each dummy-group this is a "staircase" with step height λ. Thus our β-vector is identified up to a "staircase trend".

To make things a little bit simpler, let's keep the covariates c and a intact, i.e. we have a single dummy-group, the $(y_\ell, \ldots, y_u)$ with parameters $(\beta_{y,\ell}, \ldots, \beta_{y,u})$. We force $\beta_{y,\ell} = 0$, so we get a single collinearity $c + a - \sum(i-\ell)y_i - \ell = 0$, thus

$$M = \begin{bmatrix} 1 & 1 & 0 & -1 & -2 & \ldots & \ell - u \end{bmatrix}$$

(still we discard the intercept since we're not interpreting it.)

Our equivalence class of parameters is such that we may shift any parameter vector

$$(\beta_c, \beta_a, \beta_{y,\ell}, \beta_{y,\ell+1}, \dots, \beta_{y,u})$$

with something like

$$L_\lambda = (-\lambda, -\lambda, 0, \lambda, 2\lambda, \dots, s_y\lambda).$$

Assume we have two equivalent parameter vectors

$$\begin{aligned} B_1 &= (\beta_c, \beta_c, 0, \beta_{y,\ell+1}, \dots \beta_{y,u}) \\ B_2 &= (\beta'_c, \beta'_a, 0, \beta'_{y,\ell+1}, \dots \beta'_{y,u}), \end{aligned} \tag{1}$$

their difference is L_λ for some choice of λ.

We have $\beta'_c - \beta_c = \beta'_a - \beta_a = \lambda$ for some λ, thus neither β_c nor β_a are *interpretable* as such, but the difference $\beta_c - \beta_a$ is (i.e. $\beta'_c - \beta'_a = \beta_c - \beta_a$ is independent of λ).

For differences of β_y's we have

$$\beta'_{y,i} - \beta'_{y,j} = (i-j)\lambda + (\beta_{y,i} - \beta_{y,j}) \tag{2}$$

thus differences of arbitrary β_y's are not *interpretable*.

We may sum this up:

Observation 3.2. *Assume we have covariates c, a, y with $c + a - y = 0$. Assume we dummy-encode y as $(y_\ell, y_{\ell+1}, \dots, y_u)$ with corresponding parameters $(\beta_{y,\ell}, \beta_{y,\ell+1}, \dots, \beta_{y,u})$. Then the differences $\beta_{y,i} - \beta_{y,j}$ are not interpretable.*

Consider the following quantity:

$$\gamma_i = \beta_{y,i} - \frac{i-\ell}{u-\ell}\beta_{y,u}.$$

This is the vertical distance from the point $(i, \beta_{y,i})$ to the line through the endpoints $(\ell, \beta_{y,\ell})$ and $(u, \beta_{y,u})$. (Remember that $\beta_{y,\ell} = 0$).

Denote by γ'_i the γ_i for B_2 in equation (1), denote by γ_i this quantity for B_1. We remember that $B_2 = B_1 + L_\lambda$ for some λ. We therefore have $\beta'_{y,i} = \beta_{y,i} + (i-\ell)\lambda$ for every i. Thus, we get

$$\begin{aligned} \gamma'_i &= \beta'_{y,i} - \frac{i-\ell}{u-l}\beta'_{y,u} \\ &= (i-\ell)\lambda + \beta_{y,i} - \frac{i-\ell}{u-\ell}((u-\ell)\lambda + \beta_{y,u}) \\ &= \beta_{y,i} - \frac{i-\ell}{u-\ell}\beta_{y,u} \\ &= \gamma_i \end{aligned}$$

Thus, γ_i is *interpretable*; it's independent of the additional restriction, it's relatively simple and is therefore probably a quantity we might try to interpret.

Say we force $\beta_{y,u} = 0$. Assume for simplicity that *all* the β_y's then are zero. If we now instead force $\beta_{y,u} = f$ for some f, then all the new points $(i, \beta_{y,i})$ will still lie on the straight line between the endpoints $(\ell, \beta_{y,\ell})$ and $(u, \beta_{y,u})$. This will be an equally good parameter vector in terms of the model, we can't identify which line is the "right" one. This gives us the following interpretation:

Observation 3.3. With the additional restriction $\beta_{y,u} = 0$, (that is, both the first and the last β_y is normalized to zero); the remaining β_y's may be interpreted as deviations from a linear trend. We can't identify which linear trend.

As we know from previously, there's more than one *interpretation*. Here's another one, a double difference. Let

$$\tau_{k,i,j} = (\beta_{y,i+k} - \beta_{y,j+k}) - (\beta_{y,i} - \beta_{y,j})$$

for meaningful combinations of (i, j, k). These are *interpretable* for every k. We implement the restriction $\beta_{y,\ell+1} = 0$, i.e. the year *after* the reference year is also zero. We let $k = 1$ and $j = \ell$ to get the quantity

$$\tau_i = \tau_{1,i,\ell} = \beta_{y,i+1} - \beta_{y,i}$$

which has the interpretation as the effect of time-travel from year i to $i + 1$ relative to time-travel from year ℓ to year $\ell + 1$.

Remark 3.4. Let's ponder a bit on this. In one of our applications we have a restriction that we actually *believe* is true, namely that the coefficients for two particular adjacent age-groups are identical (similarly to the example above). In this way, a certain difference becomes zero, and all differences between adjacent coefficients are identified (relative to our belief), and, by telescoping, all coefficients are identified. If our belief is wrong (by the amount λ), the coefficients will be biased by λd where λ is a constant and d is the distance from the reference. Also, if $\lambda \neq 0$, not only the age-coefficients become biased, but also the year- and cohort-coefficients, by the same linear trend. There's little we can do about that, so we choose to believe.

If our belief is correct, but it fails due to sampling uncertainty, how does this affect the estimated standard errors? More specifically, will uncertainty in the references due to sampling error be reflected as a linearly increasing trend (linear in the distance from the references) in the standard errors? It turns out that the answer is yes. The standard errors agree well with confidence intervals computed by bootstrapping. This follows from remark 2.15.

References

[1] Jr. Elswick, R. K., Chris Gennings, Vernon M. Chinchilli, and Kathryn S. Dawson, *A simple approach for finding estimable functions in linear models*, The American Statistician **45** (1991), no. 1, 51–53.

[2] L.L. Kupper, J.M. Janis, I.A. Salama, C.N. Yoshizawa, and B.G. Greenberg, *Age-Period-Cohort Analysis: An Illustration of the Problems in Assessing Interaction in One Observation Per Cell Data*, Commun. Statist.-Theor. Meth. **12** (1983), no. 23, 2779–2807.

Frisch Centre Publications

All publications are available in Pdf-format at : www.frisch.uio.no

Rapporter (Reports)

1/2006	**Finansiering av tros- og livssynssamfunn**	Aanund Hylland
2/2006	**Optimale strategier i et to-kvotesystem**	Rolf Golombek, Cathrine Hagem, Michael Hoel
3/2006	**Evaluering av tilskuddsordningen for organisasjoner for personer med nedsatt funksjonsevne**	Rolf Golombek, Jo Thori Lind
4/2006	**Aetats kvalifiserings- og opplæringstiltak – En empirisk analyse av seleksjon og virkninger**	Ines Hardoy, Knut Røed, Tao Zhang
5/2006	**Analyse av aldersdifferensiert arbeidsgiveravgift**	Gaute Ellingsen, Knut Røed
6/2006	**Utfall av yrkesrettet attføring i Norge 1994-2000**	Tyra Ekhaugen
7/2006	**Inntektsfordeling og inntektsmobilitet – pensjonsgivende inntekt i Norge 1971-2003**	Ola Lotherington Vestad
8/2006	**Effektiv måloppnåelse** **En analyse av utvalgte politiske målsetninger**	Nils-Henrik M. von der Fehr
9/2006	**Sektoranalyser – Gjennomgang av samfunnsøkonomiske analyser av effektiviseringspotensialer for utvalgte sektorer**	Finn R. Førsund
10/2006	**Veien til uføretrygd i Norge**	Elisabeth Fevang, Knut Røed
1/2007	**Generisk bytte** **En økonometrisk studie av aktørenes og prisenes betydning for substitusjon**	Vivian Almendingen
2/2007	**Firm entry and post-entry performance in selected Norwegian industries**	Ola Lotherington Vestad
1/2008	**Er kommunesektoren og/eller staten lønnsledende?** **En sammenlikning av lønnsnivå for arbeidstakere i kommunal, statlig og privat sektor**	Elisabeth Fevang, Steinar Strøm, Erik Magnus Sæther
2/2008	**Tjenestepensjon og mobilitet på arbeidsmarkedet**	Nina Skrove Falch
3/2008	**Ressurser i grunnskole og videregående opplæring i Norge 2003-2007**	Torbjørn Hægeland, Lars J. Kirkebøen, Oddbjørn Raaum
4/2008	**Norms and Tax Evasion**	Erling Barth, Alexander W. Cappelen

1/2009	**Revelation of Tax Evasion by Random Audits Report on Main Project, Part 1**	Erling Eide, Harald Goldstein, Paul Gunnar Larssen, Jack-Willy Olsen
2/2009	**Øre for læring – Ressurser i grunnskole og videregående opplæring i Norge 2003-2008**	Torbjørn Hægeland, Lars J. Kirkebøen, Oddbjørn Raaum
3/2009	**Effekter på arbeidstilbudet av pensjonsreformen**	Erik Hernæs, Fedor Iskhakov
1/2010	**Revelation of Tax Evasion by Random Audits. Report on Main Project, Part 2**	Anders Berset, Erling Eide, Harald Goldstein, Paul Gunnar Larssen, Jack-Willy Olsen
2/2010	**Effektivitets- og produktivitetsanalyser på StatRes-data**	Dag Fjeld Edvardsen, Finn R. Førsund, Sverre A.C. Kittelsen
3/2010	**Utdannings- og arbeidskarrierer hos unge voksne: Hvor havner ungdom som slutter skolen i ung alder?**	Bernt Bratsberg, Oddbjørn Raaum, Knut Røed, Hege Marie Gjefsen
4/2010	**Effekter av krav om forsørgelsesevne ved familiegjenforening**	Bernt Bratsberg, Oddbjørn Raaum
5/2010	**Produktivitet i skatteetaten 2006-2009 med regioner som enhet**	Finn R. Førsund, Sverre A.C. Kittelsen
1/2011	**Yrkesdeltaking på lang sikt blant ulike innvandrergrupper i Norge**	Bernt Bratsberg, Knut Røed, Oddbjørn Raaum

Arbeidsnotater (Working papers)

1/2006	**Costs and coverage of occupational pensions**	Erik Hernæs, Tao Zhang
2/2006	**Inntektsfordelingen i Norge, og forskjellige årsaker til ulikheter i pensjonsgivende inntekt**	Ola Lotherington Vestad
3/2006	**The Wage Effect of Computer-use in Norway**	Fitwi H. Wolday
1/2007	**An evaluation of the labour market response of eliminating the retirement earnings test rule**	Erik Hernæs, Zhiyang Jia
1/2008	**LIBEMOD 2000 - LIBeralisation MODel for the European Energy Markets: A Technical Description**	F.R. Aune, K.A. Brekke, R. Golombek, S.A.C. Kittelsen, K.E. Rosendahl
2/2008	**Modelling Households in LIBEMOD 2000 - A Nested CES Utility Function with Endowments**	Sverre Kittelsen
3/2008	**Analyseopplegg for å kunne male om reorganisering av skatteetaten fører til en mer**	Finn R. Førsund, Sverre A.C. Kittelsen

	effektiv ressursbruk	
4/2008	**Patenter i modeler med teknologisk vekst – en litteraturoversikt med vekt på klimapolitikk**	Helge Berglann
5/2008	**The R&D of Norwegian Firms: an Empirical Analysis**	Anton Giulio Manganelli
1/2009	**An Informal Care Leave Arrangement – An Economic Evaluation**	Kebebew Negera
1/2010	**Job Reallocation and Labour Mobility among Heterogeneous Firms in Norway**	Dan Li
1/2011	**Job changes, wage changes, and pension portability**	Erik Hernæs, John Piggott, Ola L. Vestad, Tao Zhang
2/2011	**Sickness and the Labour Market**	John Treble
1/2012	**Dummy-encoding Inherently Collinear Variables**	Simen Gaure

Memoranda (Discussion papers)

The series is published by Department of Economics, University of Oslo, in co-operation with the Frisch Centre. This list includes memoranda related to Frisch Centre projects.
The complete list of memoranda can be found at http://www.oekonomi.uio.no/memo/.

1/2006	**The Determinants of Occupational Pensions**	Erik Hernæs, John Piggott, Tao Zhang and Steinar Strøm
4/2006	**Moving between Welfare Payments. The Case of Sickness Insurance for the Unemployed**	Morten Henningsen
6/2006	**Justifying Functional Forms in Models for Transitions between Discrete States, with Particular Reference to Employment-Unemployment Dynamics**	John Dagsvik
15/2006	**Retirement in Non-Cooperative and Cooperative Families**	Erik Hernæs, Zhiyang Jia, Steinar Strøm
16/2006	**Early Retirement and Company Characteristics**	Erik Hernæs, Fedor Iskhakov and Steinar Strøm
20/2006	**Simulating labor supply behavior when workers have preferences for job opportunities and face nonlinear budget constraints**	John K. Dagsvik, Marilena Locatelli, Steinar Strøm
21/2006	**Climate agreements: emission quotas versus technology policies**	Rolf Golombek, Michael Hoel
22/2006	**The Golden Age of Retirement**	Line Smart Bakken
23/2006	**Advertising as a Distortion of Social Learning**	Kjell Arne Brekke, Mari Rege
24/2006	**Advertising as Distortion of Learning in Markets with**	Kjell Arne Brekke, Mari Rege

	Network Externalities	
26/2006	**Optimal Timing of Environmental Policy; Interaction Between Environmental Taxes and Innovation Externalities**	Reyer Gerlagh, Snorre Kverndokk, Knut Einar Rosendahl
3/2007	**Corporate investment, cash flow level and market imperfections: The case of Norway**	B. Gabriela Mundaca, Kjell Bjørn Nordal
4/2007	**Monitoring, liquidity provision and financial crisis risk**	B. Gabriela Mundaca
5/2007	**Total tax on Labour Income**	Morten Nordberg
6/2007	**Employment behaviour of marginal workers**	Morten Nordberg
9/2007	**As bad as it gets: Well being deprivation of sexually exploited trafficked women**	Di Tommaso M.L., Shima I., Strøm S., Bettio F.
10/2007	**Long-term Outcomes of Vocational Rehabilitation Programs: Labor Market Transitions and Job Durations for Immigrants**	Tyra Ekhaugen
12/2007	**Pension Entitlements and Wealth Accumulation**	Erik Hernæs, Weizhen Zhu
13/2007	**Unemployment Insurance in Welfare States: Soft Constraints and Mild Sanctions**	Knut Røed, Lars Westlie
15/2007	**Farrell Revisited: Visualising the DEA Production Frontier**	Finn R. Førsund, Sverre A. C. Kittelsen, Vladimir E. Krivonozhko
16/2007	**Reluctant Recyclers: Social Interaction in Responsibility Ascription**	Kjell Arne Brekke , Gorm Kipperberg, Karine Nyborg
17/2007	**Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility across Countries**	O. Raaum, B. Bratsberg, K. Røed, E. Österbacka, T. Eriksson, M. Jäntti, R. Naylor
18/2007	**Pennies from heaven - Using exogenous tax variation to identify effects of school resources on pupil achievement**	Torbjørn Hægeland, Oddbjørn Raaum and Kjell Gunnar Salvanes
19/2007	**Trade-offs between health and absenteeism in welfare states: striking the balance**	Simen Markussen
1/2008	**Is electricity more important than natural gas? Partial liberalization of the Western European energy markets**	Kjell Arne Brekke, Rolf Golombek, Sverre A.C. Kittelsen
3/2008	**Dynamic programming model of health and retirement**	Fedor Ishakov
8/2008	**Nurses wanted. Is the job too harsh or is the wage too low?**	M. L. Di Tommaso, Steinar Strøm, Erik Magnus Sæther
10/2008	**Linking Environmental and Innovation Policy**	Reyer Gerlagh, Snorre Kverndokk, Knut Einar

		Rosendahl
11/2008	**Generic substitution**	Kari Furu, Dag Morten Dalen, Marilena Locatelli, Steinar Strøm
14/2008	**Pension Reform in Norway: evidence from a structural dynamic model**	Fedor Iskhakov
15/2008	**I Don't Want to Hear About it: Rational Ignorance among Duty-Oriented Consumers**	Karine Nyborg
21/2008	**Equity and Justice in Global Warming Policy**	Snorre Kverndokk, Adam Rose
22/2008	**The Impact of Labor Market Policies on Job Search Behavior and Post-Unemployment Job Quality**	Simen Gaure, Knut Røed, Lars Westlie
24/2008	**Norwegian Vocational Rehabilitation Programs: Improving Employability and Preventing Disability?**	Lars Westlie
25/2008	**The Long-term Impacts of Vocational Rehabilitation**	Lars Westlie
28/2008	**Climate Change, Catastrophic Risk and the Relative Unimportance of Discounting**	Eric Nævdal, Jon Vislie
29/2008	**Bush meets Hotelling: Effects of improved renewable energy technology on greenhouse gas emissions**	Michael Hoel
7/2009	**The Gate is Open: Primary Care Physicians as Social Security Gatekeepers**	Benedicte Carlsen, Karine Nyborg
9/2009	**Towards an Actuarially Fair Pension System in Norway**	Ugo Colombino, Erik Hernæs, Marilena Locatelli, Steinar Strøm
13/2009	**Moral Concerns on Tradable Pollution Permits in International Environmental Agreements**	Johan Eyckmans, Snorre Kverndokk
14/2009	**Productivity of Tax Offices in Norway**	Finn R. Førsund, Dag Fjeld Edvardsen, Sverre A.C. Kittelsen, Frode Lindseth
19/2009	**Closing the Gates? Evidence from a Natural Experiment on Physicians' Sickness Certification**	Simen Markussen
20/2009	**The Effectss of Sick-Leaves on Earnings**	Simen Markussen
25/2009	**Labour Supply Response of a Retirement Earnings Test Reform**	Erik Hernæs, Zhiyang Jia
2/2010	**Climate Policy without Commitment**	Rolf Golombek, Mads Greaker, Michael Hoel
13/2010	**Is there a green paradox**	Michael Hoel
14/2010	**OLS with Multiple High Dimensional Category Dummies**	Simen Gaure

19/2010	**Cutting Costs of Catching Carbon Intertemporal effects under imperfect climate policy**	Michael Hoel, Svenn Jensen
20/2010	**Identifying Trend and Age Effects in Sickness Absence from Individual Data: Some Econometric Problems**	Erik Biørn
1/2011	**Is corporate social responsibility associated with lower wages**	Karine Nyborg, Tao Zhang
16/2011	**Who pays for occupational pensions?**	Ola Lotherington Vestad

The Frisch Centre

The Ragnar Frisch Centre for Economic Research is an independent research institution founded by the University of Oslo. The Frisch Centre conducts economic research in co-operation with the Department of Economics, University of Oslo. The research projects are mostly financed by the Research Council of Norway, government ministries and international organisations. Most projects are co-operative work involving the Frisch Centre and researchers in other domestic and foreign institutions.

Ragnar Frisch Centre for Economic Research
Gaustadalléen 21
N-0349 Oslo, Norway
T + 47 22 95 88 10
F + 47 22 95 88 25
frisch@frisch.uio.no
www.frisch.uio.no