On the relation between parameter estimates from transition and marginal models for longitudinal data

Werner Vach[1], Ulrike Grömping[2] and Kirstin Schulz[1]

[1] Center for Data Analysis and Model Building, University of Freiburg, Germany and
Institute of Medical Biometry and Informatics, University of Freiburg, Germany

[2] Department of Statistics, University of Dortmund, Germany

Summary

Zeger and Liang (1992) allege that regression parameter estimates from transition models tend to be smaller in absolute value than regression parameter estimates from marginal models. This statement is true in typical longitudinal data situations for the relation between the true parameters of the marginal and transition model. However, under the assumption of a transition model the Generalized Estimating Equation (GEE) approach with a non-diagonal working correlation matrix may yield inconsistent estimates for the parameters of the marginal model. Hence the stated relation may be valid for the true parameters, but not for the stochastic limit of the parameter estimates.

Key words: longitudinal data; marginal models; transition models; generalized estimating equations

1. Introduction

In this paper we consider two approaches to analyze longitudinal data measured at time points $t = 1, \ldots, T$. Both approaches investigate the relationship between a (time-dependent) covariate X_t and an outcome variable Y_t by estimating regression coefficients β_t or $\tilde{\beta}_t$. The coefficients $(\beta_t)_{t=2,\ldots,T}$ result from a transition model where the effect of X_t on Y_t is adjusted for the previous outcome Y_{t-1}. Hence the regression model considered is given by

$$E(Y_t|X_t = x_t, Y_{t-1} = y_{t-1}) = \mu(\beta_{0t} + \beta_t x_t + \gamma_t y_{t-1}) \ . \tag{1}$$

On the other hand the coefficients $(\tilde{\beta}_t)_{t=1,\ldots,T}$ result from the marginal model

$$E(Y_t|X_t = x_t) = \mu(\tilde{\beta}_{0t} + \tilde{\beta}_t x_t) \ . \tag{2}$$

In both models μ is a prespecified inverse link function (McCullagh and Nelder, 1989). Of course, there exists a basic conceptual difference between model (1) and model (2), which is discussed in detail by Zeger and Liang (1992). This conceptual difference corresponds to a numerical difference between the values of β_t and $\widetilde{\beta}_t$, and with respect to this numerical difference Zeger and Liang allege that $|\widetilde{\beta}_t|$ tends to be larger than $|\beta_t|$. Indeed, this conjecture holds under conditions, which are typical for longitudinal studies. This can be explained by the following considerations, which are similar to the analysis of the bias due to omitted covariates (Gail, Wieand and Piantadosi 1984, Gail 1986, Neuhaus and Jewell 1993), because the difference between (2) and (1) is just the omission of the "covariate" Y_{t-1}. If (1) holds and if μ is the identity function we have

$$E(Y_t|X_t = x_t) = \beta_{0t} + \beta_t x_t + \gamma_t E(Y_{t-1}|X_t = x_t)\ .$$

The conditions $\gamma_t = 0$ or $E(Y_{t-1}|X_t = x_t) \equiv E(Y_{t-1})$ would imply that $\widetilde{\beta}_t = \beta_t$. In longitudinal studies we typically have $\gamma_t > 0$, implying that the outcome variables show a positive correlation over time. Hence if $\beta_t > 0$ ($\beta_t < 0$), we have $|\widetilde{\beta}_t| > |\beta_t|$ if $E(Y_{t-1}|X_t = x)$ is a monotonously increasing (decreasing) function of x. The relation between Y_{t-1} and X_t is inherited from the relation between Y_{t-1} and X_{t-1}, if there exists a strong positive correlation between X_{t-1} and X_t, which is again typical for longitudinal data. Hence the correlation between Y_{t-1} and X_t has the same sign as β_{t-1}, and typically β_{t-1} and β_t have the same sign. Hence we have indeed $|\widetilde{\beta}_t| \geq |\beta_t|$. If μ is a general inverse link function, $E(Y_{t-1}|X_t = x_t) \equiv E(Y_{t-1})$ is not sufficient to assure $\widetilde{\beta}_t = \beta_t$. Especially for the logistic link with $\mu(\cdot) = \Lambda(\cdot) := (1 + \exp(-\cdot))^{-1}$ independence of Y_{t-1} and X_t implies $|\widetilde{\beta}_t| \leq |\beta_t|$ (Gail, 1986). Here the statement $|\widetilde{\beta}_t| \geq |\beta_t|$ is only true, if the correlation between Y_{t-1} and X_t exceeds some amount, which is again typical for longitudinal studies. Hence the statement of Zeger and Liang (1992) can be regarded as valid with respect to the true parameters defined by (1) and (2). However, in the next section we will show, that the relation need not hold for the stochastic limits of the parameter estimates, if the Generalized Estimating Equation (GEE) approach is used for estimating parameters of the marginal model.

2. Estimation procedures

Let us assume we have measurements $(X_{it}, Y_{it})_{i=1,\ldots,I,t=1,\ldots,T}$ for I subjects at T time points. The parameters of the transition model can be easily estimated by fitting a generalized linear model for each time point $t \geq 2$ with outcome Y_t and covariates X_t and Y_{t-1} (cf. Bonney, 1987). The parameters of the marginal model can be estimated by fitting a generalized linear model for each time point t with outcome Y_t and covariate X_t. This approach is discussed in detail by WEI and STRAM (1988) and STRAM, WEI and WARE (1988) and we call it SWW in the sequel. Especially if $\widetilde{\beta}_t$ is assumed to be constant over time, the GEE approach promises a more efficient estimation. In this approach the estimation of $\widetilde{\beta} := (\widetilde{\beta}_{0t}, \widetilde{\beta}_t)_{t=1,\ldots,T}$ is based on the estimating equations

$$\sum_{i=1}^{I} \mathbf{D}_i \mathbf{V}_i^{-1}(y_i - \mu_i(\widetilde{\beta})) = 0 \tag{3}$$

with $y_i := (y_{i1}, \ldots, y_{iT})', \mu_i := (\mu_{i1}, \ldots, \mu_{iT})', \mu_{it} := \mu(\widetilde{\beta}_{0t} + \widetilde{\beta}_t x_{it}), \mathbf{D}_i := \partial\mu_i/\partial\widetilde{\beta}, \mathbf{V}_i := \mathbf{A}_i^{1/2}\mathbf{R}_i(\alpha)\mathbf{A}_i^{1/2}, \mathbf{A}_i := diag(Var(y_i))$. $\mathbf{R}_i(\alpha)$ is a prespecified working correlation matrix with an unknown parameter α. LIANG and ZEGER (1986) suggest to estimate α within each step of an iterative procedure to solve the estimating equations. PRENTICE (1988) suggests additional estimating equations for α. If $\mathbf{R}_i(\alpha)$ is not chosen as the identity matrix, then correct specification of $E(Y_t|X_t = x_t)$ as $\mu(\widetilde{\beta}_{0t} + \widetilde{\beta}_t x_t)$ does not suffice to assure that the GEE approach results in consistent estimates for $\widetilde{\beta}$. We need in addition that

$$E(Y_t|X_1 = x_1, X_2 = x_2, ..., X_T = x_T) = \mu(\widetilde{\beta}_{0t} + \widetilde{\beta}_t x_t) \tag{4}$$

holds, i.e. we have to exclude a dependence of the expectation of Y_t on the covariate values at the other time points, if we condition on all covariate values. This is due to the fact that consistency of the parameter estimates crucially depends on the unbiasedness of the estimating equation, i.e. the expectation of the left-hand side of the estimating equation (3) has to be 0 for the true parameters. If the working correlation matrix is the identity matrix, i.e. the matrices $\mathbf{V}_i$ are diagonal, the correct specification of (2) for $t = 1, \ldots, T$ is sufficient for this unbiasedness. Otherwise straightforward calculations show that (4) is needed.

Now if we assume that (1) is valid, i.e. that the expectation of Y_t depends on Y_{t-1}, then (4) will not be satisfied in general. To demonstrate this, we consider the case $T = 2$. Then

$$\begin{aligned}
&E(Y_2|X_1 = x_1, X_2 = x_2) \\
&\quad = \int E(Y_2|X_1 = x_1, X_2 = x_2, Y_1 = y_1)\, f(y_1|x_1, x_2)\, dy_1 \\
&\quad = \int E(Y_2|X_1 = x_1, X_2 = x_2, Y_1 = y_1) \frac{f(x_2|y_1, x_1)\, f(y_1|x_1)}{\int f(x_2|y_1, x_1)\, f(y_1|x_1)\, dy_1}\, dy_1
\end{aligned}$$

is a complex function of x_1. If we assume for simplicity that $E(Y_2|X_1 = x_1, X_2 = x_2, Y_1 = y_1) = E(Y_2|X_2 = x_2, Y_1 = y_1)$ and $f(x_2|y_1, x_1) \equiv f(x_2|x_1)$

we have

$$E(Y_2|X_1 = x_1, X_2 = x_2) = \int \mu(\beta_{02} + \beta_2 x_2 + \gamma_2 y_1)\, f(y_1|x_1)\, dy_1 \ ,$$

illustrating that the dependence of the expectation of Y_2 on X_1 results from the dependence of Y_1 on X_1 and the dependence of the expectation of Y_2 on Y_1. Similarly, we have

$$E(Y_1|X_1 = x_1, X_2 = x_2) = \frac{\int y_1\, f(x_2|y_1, x_1)\, f(y_1|x_1)\, dy_1}{\int f(x_2|y_1, x_1)\, f(y_1|x_1)\, dy_1} \ .$$

However, now the assumption $f(x_2|y_1, x_1) \equiv f(x_2|x_1)$ suffices to assure that $E(Y_1|X_1 = x_1, X_2 = x_2)$ does not depend on x_2.

Consider now longitudinal data generated by a transition model for which there are true parameters $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_T$ such that (2) is satisfied (for an example, cf. section 3). If we are interested in the marginal model, then the SWW approach and the GEE approach with an identity working correlation matrix are consistent for the true values $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_T$, but the GEE approach with other working correlation structures may be consistent for different values $\widetilde{\beta}^*_1, \ldots, \widetilde{\beta}^*_T$, since (4) is not satisfied. In the next section we investigate this bias. Especially we are interested in situations with $|\widetilde{\beta}_t| > |\beta_t|$ but $|\widetilde{\beta}^*_t| < |\beta_t|$, such that the statement of Zeger and Liang (1992) becomes invalid due to the use of the GEE approach.

3. Numerical results on the bias

We consider the special case with $T = 2$ and X_1, X_2, Y_1, Y_2 binary. The assumed joint distribution of X_1, X_2, Y_1, Y_2 is given by

$$\begin{aligned}
&P(X_1 = 1) = 0.5\\
&P(Y_1 = 1|X_1 = x_1) = \Lambda(\widetilde{\beta}_{01} + \widetilde{\beta}_1 x_1)\\
&P(X_2 = 1|X_1 = x_1, Y_1 = y_1) = \Lambda(\theta_0 + \theta_1 x_1 + \theta_2 y_1)\\
&P(Y_2 = 1|X_1 = x_1, X_2 = x_2, Y_1 = y_1) = \Lambda(\beta_{02} + \beta_2 x_2 + \gamma y_1)\,.
\end{aligned}$$

Here θ_1 describes the direct influence of X_1 on X_2 and θ_2 describes the direct influence of Y_1 on X_2. For given values of $\widetilde{\beta}_1, \theta_1, \theta_2, \beta_2$ and γ we choose the intercepts $\widetilde{\beta}_{01}, \theta_0$ and β_{02} such that $P(Y_1 = 1) = P(X_2 = 1) = P(Y_2 = 1) = 0.5$. β_2 is the parameter of interest in the transition model, which can be estimated consistently. As X_1 and X_2 are binary equation (2) always represents a correct specification and the true parameter $\widetilde{\beta}_2$ of the marginal model is uniquely determined by the relation

$$P(Y_2 = 1|X_2 = x_2) = \Lambda(\widetilde{\beta}_{02} + \widetilde{\beta}_2 x_2);\ x_2 = 0, 1\ .$$

The GEE approach is consistent for the solution $(\tilde{\beta}^*, \alpha^*)$ of the equation

$$E\, S(\tilde{\beta}, \alpha, X_1, X_2, Y_1, Y_2) = 0$$

where the expectation is taken with respect to the true distribution, and $S(\tilde{\beta}, \alpha, X_1, X_2, Y_1, Y_2)$ consists of five components given by

$$\begin{pmatrix} 1 \\ X_1 \end{pmatrix} \left[Y_1 - \Lambda(\tilde{\beta}_{01} + \tilde{\beta}_1 X_1) - \sqrt{\frac{\Lambda'(\tilde{\beta}_{01} + \tilde{\beta}_1 X_1)}{\Lambda'(\tilde{\beta}_{02} + \tilde{\beta}_2 X_2)}}\, \alpha (Y_2 - \Lambda(\tilde{\beta}_{02} + \tilde{\beta}_2 X_2)) \right]$$

$$\begin{pmatrix} 1 \\ X_2 \end{pmatrix} \left[Y_2 - \Lambda(\tilde{\beta}_{02} + \tilde{\beta}_2 X_2) - \sqrt{\frac{\Lambda'(\tilde{\beta}_{02} + \tilde{\beta}_2 X_2)}{\Lambda'(\tilde{\beta}_{01} + \tilde{\beta}_1 X_1)}}\, \alpha (Y_1 - \Lambda(\tilde{\beta}_{01} + \tilde{\beta}_1 X_1)) \right]$$

$$\frac{(Y_1 - \Lambda(\tilde{\beta}_{01} + \tilde{\beta}_1 X_1))(Y_2 - \Lambda(\tilde{\beta}_{02} + \tilde{\beta}_2 X_2))}{\sqrt{\Lambda'(\tilde{\beta}_{01} + \tilde{\beta}_1 X_1)\Lambda'(\tilde{\beta}_{02} + \tilde{\beta}_2 X_2)}} - \alpha$$

with $\Lambda'(\cdot) := \Lambda(\cdot)(1 - \Lambda(\cdot))$, i.e. we use the version of PRENTICE (1988). For given $\tilde{\beta}_1, \theta_1, \theta_2, \beta_2$ and γ both $\tilde{\beta}_2$ and $\tilde{\beta}_2^*$ can be computed exactly by numerical methods.

As discussed in Section 1, the difference between β_2 and $\tilde{\beta}_2$ depends on the degree of dependence between Y_1 and Y_2 and the degree of dependence between Y_1 and X_2. The first dependence is reflected by the parameter γ, and hence we consider $\beta_2, \tilde{\beta}_2$ and $\tilde{\beta}_2^*$ in dependence on γ in the sequel. A dependence between Y_1 and X_2 can be due to two different sources: A direct relation due to $\theta_2 \neq 0$ or a common correlation with X_1 expressed by $\tilde{\beta}_1$ and θ_1. We consider first the latter case with $\theta_2 = 0$. In Figure 1 for six choices of $\tilde{\beta}_1$ and θ_1 we consider the behaviour of $\tilde{\beta}_2$ and $\tilde{\beta}_2^*$ in dependence on γ with $\beta_2 = 1.0$ fixed. The difference between $\tilde{\beta}_2$ and $\tilde{\beta}_2^*$ increases with the size of $|\gamma|$. In the interesting case of a positive γ the difference remains small enough that the relation $|\tilde{\beta}_2| > |\beta_2|$ is rarely inverted to $|\tilde{\beta}_2^*| < |\beta_2|$.

However, the situation is quite different if we consider the case $\theta_2 \neq 0$, i.e. if there exists a direct influence of Y_1 on X_2 (Figure 2). If $\theta_1 = 0$ we always observe a serious bias, such that $|\tilde{\beta}_2| > |\beta_2| > |\tilde{\beta}_2^*|$ if $\gamma \geq 0$. The influence of $\tilde{\beta}_1$ on the size of this bias seems to be small.

Biased estimation of $\tilde{\beta}_2$ is of course only one problem. Due to the misspecification of $E(Y_2|X_1 = x_1, X_2 = x_2)$ and, if $\theta_2 \neq 0$, of $E(Y_1|X_1 = x_1, X_2 = x_2)$ the GEE approach is also inconsistent for $\tilde{\beta}_1$. In Figure 3 the values of $\tilde{\beta}_1$ and $\tilde{\beta}_1^*$ are shown in dependence on γ for various combinations of θ_1, θ_2 and β_2. Surprisingly for $\theta_2 = 0$ the bias of the GEE approach is larger for $\tilde{\beta}_1$ than for $\tilde{\beta}_2$, although $E(Y_1|X_1 = x_1, X_2 = x_2)$ is correctly specified. In general θ_1, θ_2 and β_2 have only little influence on the bias of $\tilde{\beta}_1$.

Finally, we want to look at the GEE approach with the assumption $\tilde{\beta}_1 = \tilde{\beta}_2 \equiv \tilde{\beta}$. Here we have of course 4 instead of 5 estimating equations. With respect to the true distribution for given $\tilde{\beta}_1, \theta_1, \theta_2, \gamma$ we now choose β_2 always such that $\tilde{\beta}_1 = \tilde{\beta}_2$. In Figure 4 we show the values of β_2 and $\tilde{\beta}^*$ in dependence on γ with $\tilde{\beta} = 1.0$ fixed for four combinations of θ_1 and θ_2. We observe that the GEE approach always results in a biased estimate of $\tilde{\beta}$ if $\gamma \neq 0$, and that this bias can be substantial especially if $\theta_2 \neq 0$.

4. Conclusions

If for longitudinal data with time-dependent covariates, which are generated by a transition model, parameters of a marginal model are estimated by the GEE approach, then these estimates can be biased. The bias seems to be especially large, if there is a direct dependence of the expectation of a covariate on the outcome of a previous time point. Such situations may not be common, but MEINERT et al. (1994) report a study on the influence of maternal smoking on children's bronchial hyperactivity where mothers change their smoking behaviour due to the children's status. In such a setting the expected relation between the magnitude of adjusted and unadjusted risk estimates can be inverted using the GEE approach.

It seems that the necessity of condition (4) in order to get consistent estimates of parameters in a marginal model for $E(Y_t|X_t = x_t)$ is somewhat overlooked in the literature, although PRENTICE (1988) clearly starts his chapter on GEE with a specification of $E(Y_1, \ldots, Y_T|X_1, \ldots, X_T)$. One reason is that the condition does not matter if the covariates are constant over time. In general, our knowledge about the possible consequences of a violation of (4) is still ambiguous; sometimes the resulting bias is substantial, sometimes it is negligible.

To achieve consistent estimates, condition (4) is not necessary, if the working correlation matrix $\mathbf{R}_i(\alpha)$ is chosen as the identity matrix. In this case, the GEE approach is equivalent to the approach by WEI and STRAM (1988) and STRAM, WEI and WARE (1988), as pointed out by ZEGER (1988). Hence if unbiased estimation of parameters of marginal models is important and condition (4) is questionable, we recommend to choose an identity working correlation matrix. A similar situation, where only the use of the identity working correlation matrix assures consistency of the parameter estimates is mentioned in the rejoinder of the discussion of the paper by FITZMAURICE et al. (1993).

References

BONNEY, G.E., 1987: Logistic regression for dependent binary observations. Biometrics **43**, 951-973.

FITZMAURICE, G.M., LAIRD, N.M. and ROTNITZKY, A.G., 1993: Regression models for discrete longitudinal responses (with discussion). Statistical Science **8**, 284-309.

GAIL, M.H., WIEAND, S. and PIANTADOSI, S., 1984: Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika **71**, 431-444.

GAIL, M.H. 1986. Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In: *Moolgavkar, S.H.* and *Prentice, R.L.*, Modern statistical methods in chronic disease epidemiology, John Wiley and Sons, 3-18.

LIANG, K.Y. and ZEGER, S.L., 1986: Longitudinal data analysis using generalized linear models. Biometrika **73**, 13-22.

MCCULLAGH, P. and NELDER, J.A. 1989: *Generalized linear models.* 2nd edition. London: Chapman and Hall.

MEINERT, R., FRISCHER, T. and KUEHR, J., 1994: Assessing the effect of time-varying covariates in cross-sectional studies. Journal of Clinical Epidemiology **47**, 983-991.

NEUHAUS, J.M. and JEWELL, N.P., 1993: A geometric approach to assess bias due to omitted covariates in generalized linear models. Biometrika **80**, 807-815.

PRENTICE, R.L., 1988: Correlated binary regression with covariates specific to each binary observation. Biometrics **44**, 1033-1048.

STRAM, D.O., WEI, L.J. and WARE, J.H., 1988: Analysis of repeated ordered categorical outcomes with possibly missing observations and time- dependent covariates. Journal of the American Statistical Association **83**, 631-637.

WEI, L.J. and STRAM, D.O., 1988: Analysing repeated measurements with possibly missing observations by modelling marginal distributions. Statistics in Medicine **7**, 139-148.

ZEGER, S.L., 1988: Commentary. Statistics in Medicine **7**, 161-168.

ZEGER, S.L. and LIANG K.Y, 1992: An overview of methods for the analysis of longitudinal data. Statistics in Medicine **11**, 1825-1839.

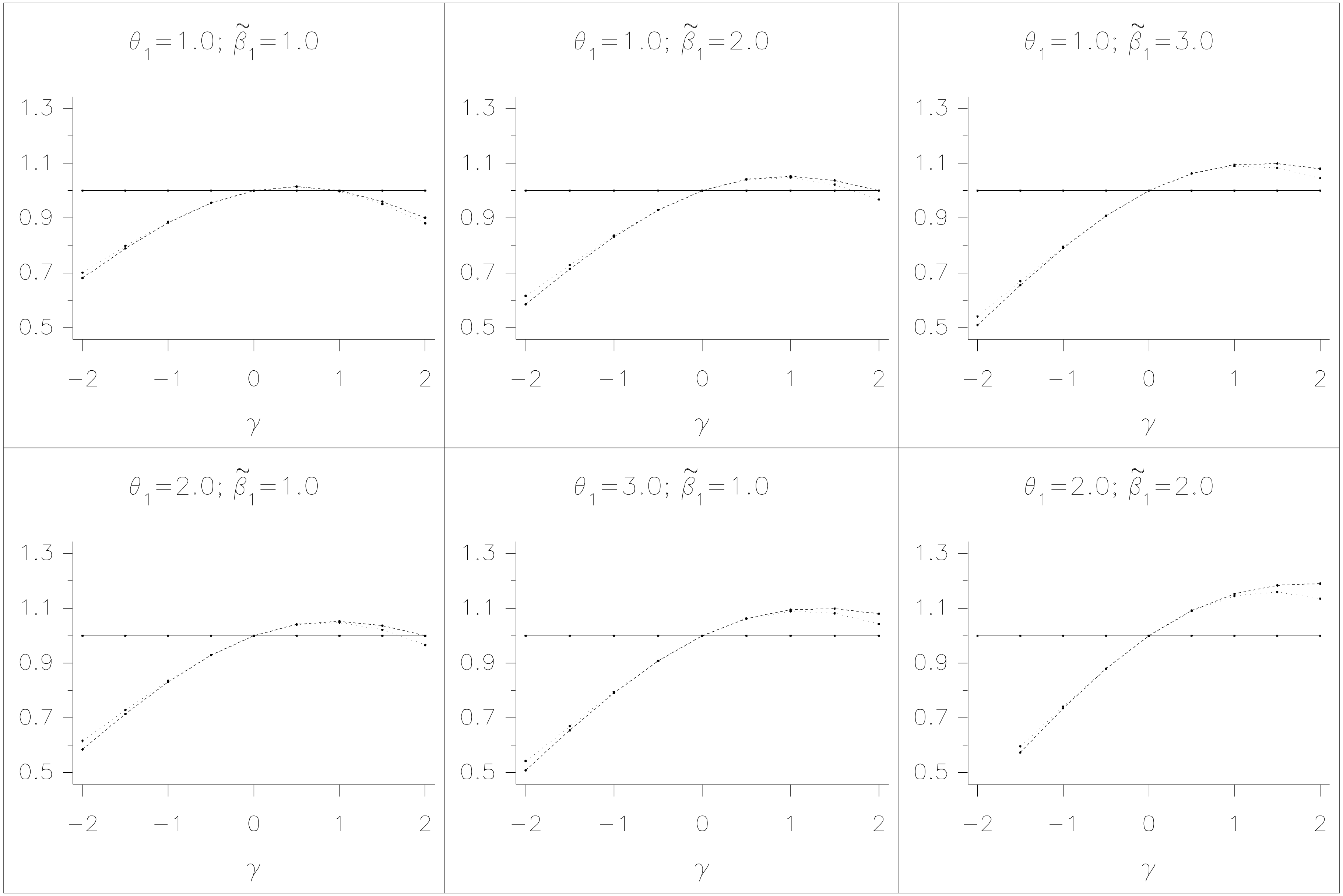

Figure 1: β_2 (——), $\tilde{\beta}_2$ (-------), and $\tilde{\beta}_2^*$ (·········) as a function of γ for $\theta_2 = 0.0$.

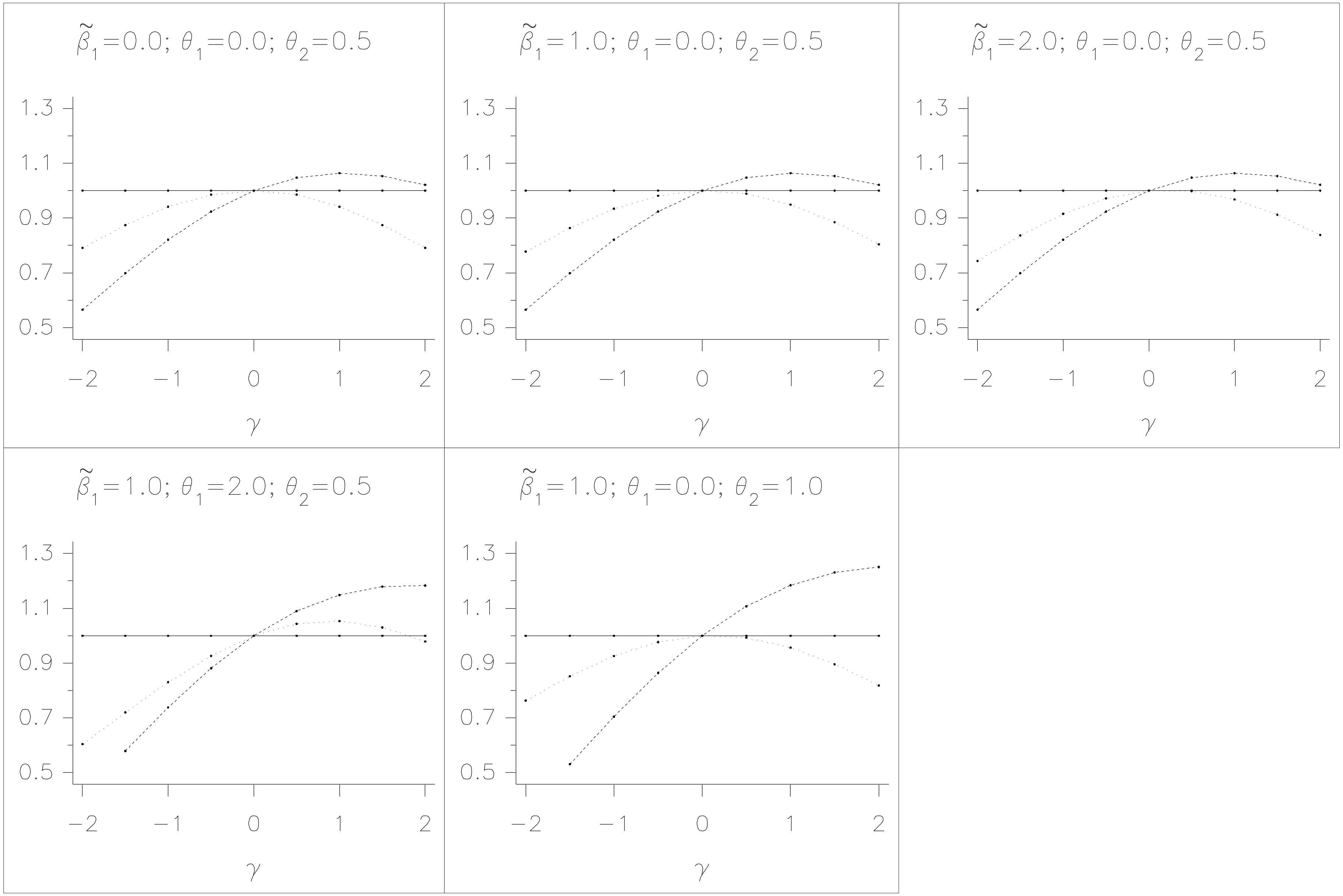

Figure 2: β_2 (——), $\tilde{\beta}_2$ (-------), and $\tilde{\beta}_2^*$ (·········) as a function of γ for $\theta_2 \neq 0.0$.

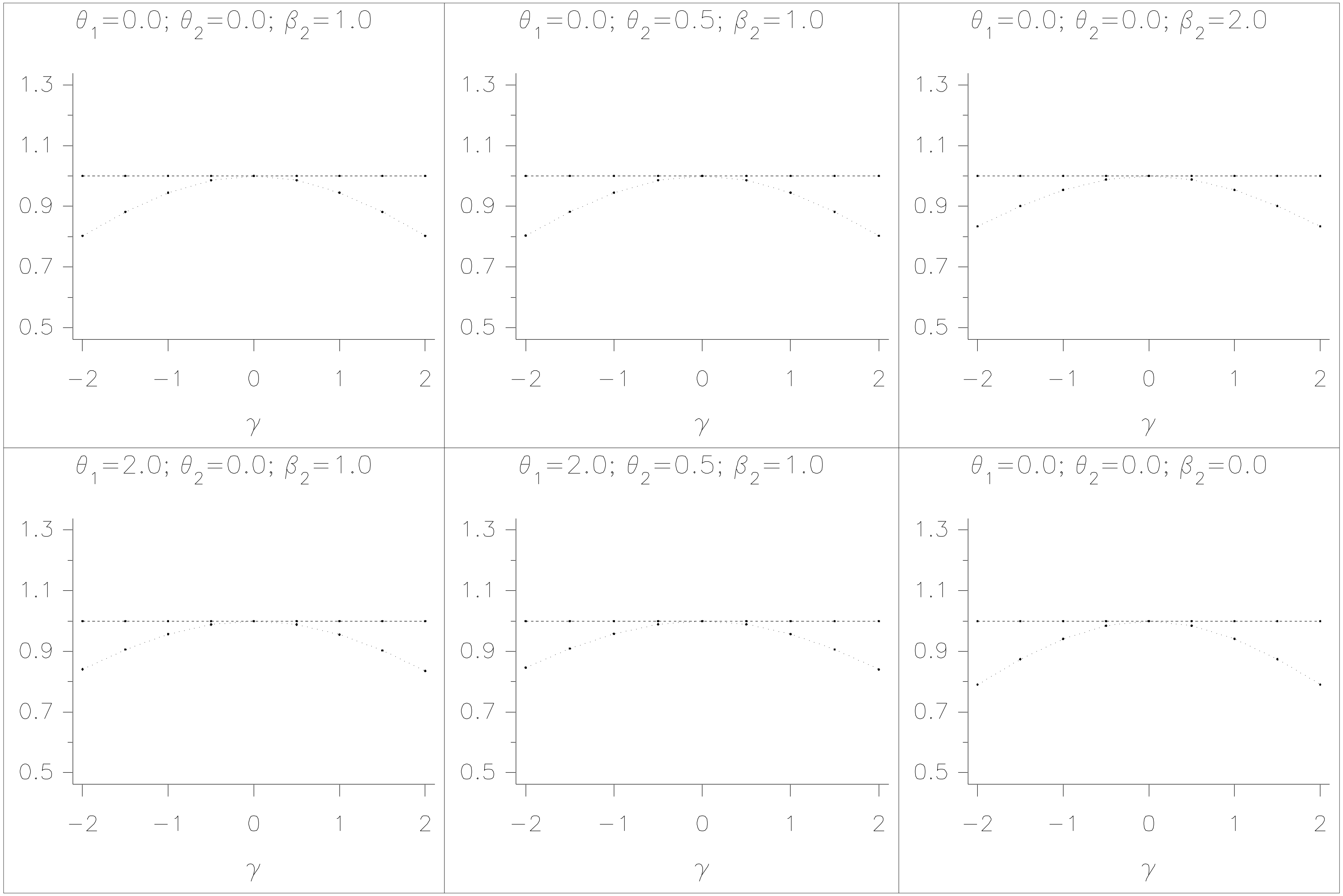

Figure 3: $\tilde{\beta}_1$ (-------) and $\tilde{\beta}_1^*$ (·········) as a function of γ.

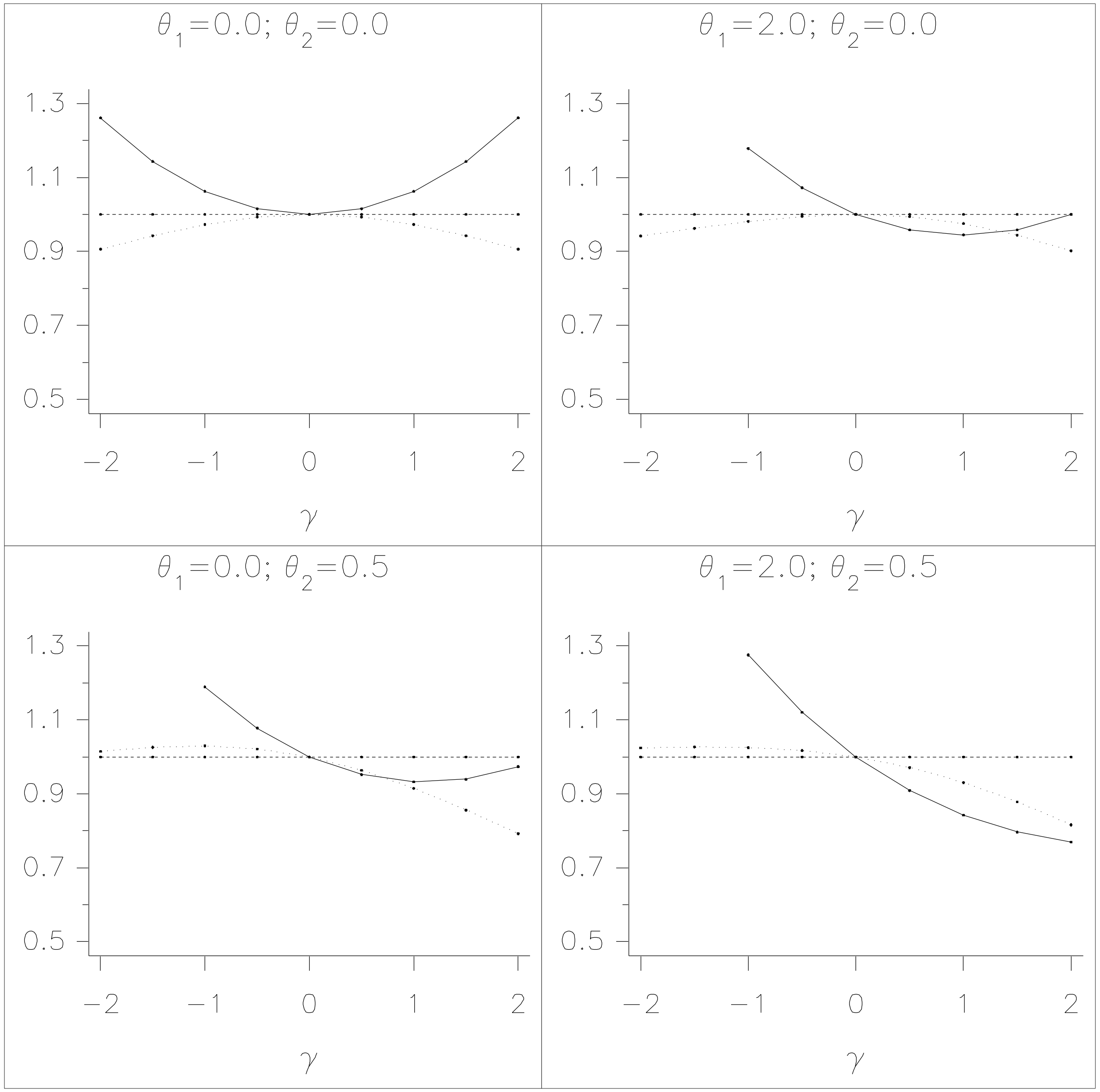

Figure 4: β_2 (——), $\widetilde{\beta}$ (-------), and $\widetilde{\beta}^*$ (·········) as a function of γ.