MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

A Modification of Profile Empirical Likelihood for the Exponential-Tilt Model

Kazuhisa INAGAKI and Fumiyasu KOMAKI

METR 2009–11 March 2009

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

A Modification of Profile Empirical Likelihood for the Exponential-Tilt Model

Kazuhisa Inagaki * and Fumiyasu Komaki* †
* Department of Mathematical Informatics,
Graduate School of Information Science and Technology, University of Tokyo, Tokyo, Japan.
† RIKEN Brain Science Institute, Wako-shi, Japan

Abstract

We consider semiparametric models whose infinite dimensional parameter corresponds to a probability distribution. The NPMLE based on the profile empirical likelihood for this kind of semiparametric models has attracted considerable interest. We propose the use of a modified profile empirical likelihood to improve the accuracy of this estimation. We consider applications to the exponential-tilt model and show that the accuracy of the proposed estimator is better than that of the conventional NPMLE by numerical study.

Key words: logistic regression, NPMLE, semiparametric models

1 Introduction

Suppose that we have a set of observations $x_1, x_2, \ldots, x_n$ from a probability density that belongs to a semiparametric model:

$$f(x; \theta, g),$$

where θ is a finite-dimensional parameter and g is a probability density. Here, g is regarded as an infinite-dimensional parameter. We attempt to estimate either θ or a part of θ and we do this without any assumption about the unknown probability density g.

Considerable attention has been directed to the nonparametric maximum likelihood estimator (NPMLE), which is based on the profile empirical likelihood (Bickel et al., 1993).

On the other hand, estimation based on the conventional profile likelihood for parametric models is not very accurate when the nuisance parameter is high-dimensional. The identical problem exists in the case of estimation based on profile empirical likelihood for semiparametric models with an infinite-dimensional nuisance parameter. The objective of our study is to improve the accuracy of estimation of θ by constructing a modified profile empirical likelihood for semiparametric models, which is analogous to the conventional modified profile likelihood for parametric models introduced by Barndorff-Nielsen (1983).

In this paper, we consider an application of this method to the exponential-tilt model:

$$f(x; \alpha, \beta, g) = \exp(\alpha + \beta x) g(x), \tag{1}$$

where α and β are scalar parameters and g is a probability distribution. This model is closely related to the logistic regression model (Qin and Zhang 2005). Typically, β is the parameter of interest and α is regarded as the normalizing constant. Our method is also applicable to various other semiparametric models.

Suppose that we have two sets of observations $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ from the distributions g and f, respectively. We consider estimating the parameter of interest β.

This paper is organized as follows. In Section 2, we obtain the profile empirical likelihood for model (1). We construct a modified profile empirical likelihood for the exponential-tilt model in Section 3, and some results from a numerical study are given in Section 4.

2 Estimation by profile empirical likelihood

We obtain the profile empirical likelihood for the exponential-tilt model (1). We define $z_1, z_2, \ldots, z_{n+m}$ by

$$z_i = \begin{cases} x_i & (i = 1, 2, \ldots, n) \\ y_{i-n} & (i = n+1, n+2, \ldots, n+m). \end{cases} \tag{2}$$

The empirical likelihood is based on the multinomial model

$$\tilde{g}(x; \boldsymbol{p}) = \sum_{i=1}^{n+m} p_i \delta(x - z_i),$$

where $\delta(x)$ is 1 when $x = 0$, and 0 otherwise (Owen, 2001).

The model $\tilde{g}(x; \boldsymbol{p})$ approximates the probability density g. The density f is approximated by

$$\tilde{f}(y; \alpha, \beta, \boldsymbol{p}) = \sum_{i=1}^{n+m} \exp(\alpha + \beta y) p_i \delta(y - z_i). \tag{3}$$

Here, the multinomial parameter $\boldsymbol{p} = (p_1, p_2, \ldots, p_{n+m})$ satisfies the constraints

$$\sum_{i=1}^{n+m} p_i = 1, \tag{4}$$

$$\sum_{i=1}^{n+m} \exp(\alpha + \beta z_i) p_i = 1, \tag{5}$$

and

$$p_i \geq 0. \tag{6}$$

Constraint (5) corresponds to the normalizing condition for the multinomial distribution $\tilde{f}(y; \alpha, \beta, \boldsymbol{p})$. Using equation (5), we can eliminate α from (3) and obtain

$$\tilde{f}(y; \beta, \boldsymbol{p}) = \frac{1}{\sum_{i=1}^{n+m} e^{\beta z_i} p_i} \sum_{i=1}^{n+m} e^{\beta y} p_i \delta(y - z_i), \tag{7}$$

where p_{n+m} is a dependent parameter determined from $p_1, p_2, \ldots, p_{n+m-1}$ by condition (4).

Based on the observed data (2), the empirical log-likelihood function is given by

$$\begin{aligned} l_{n+m}(\beta, \boldsymbol{p}) =& \log \left\{ \prod_{i=1}^{n} p_i \prod_{j=n+1}^{n+m} \left(\frac{\mathrm{e}^{\beta z_j} p_j}{\sum_k = 1^{n+m} \mathrm{e}^{\beta z_k} p_k} \right) \right\} \\ =& \beta \sum_{i=1}^{m} y_i + \sum_{i=1}^{n+m} \log p_i - m \log \left(\sum_{i=1}^{n+m} e^{\beta z_i} p_i \right). \end{aligned} \tag{8}$$

We eliminate the nuisance parameter $\boldsymbol{p}$ by profiling the empirical log-likelihood. Maximizing $l_{n+m}(\beta, \boldsymbol{p})$ with respect to $\boldsymbol{p}$ with fixed β under constraints (4) and (6), we obtain

$$l_P(\beta) = l_P(\beta, \hat{\boldsymbol{p}}_\beta) = \beta \sum_{i=1}^{m} y_i - \sum_{i=1}^{n+m} \log(m\nu e^{\beta z_i} + n) + m \log \nu,$$

where

$$\hat{p}_{i,\beta} = \frac{1}{n + m\nu e^{\beta z_i}}, \tag{9}$$

and $\nu = \nu(\beta)$ satisfying

$$\sum_{i=1}^{n+m} \frac{1}{n + m\nu e^{\beta z_i}} = 1, \tag{10}$$

and

$$\frac{1}{n + m\nu e^{\beta z_i}} \geq 0 \qquad (i = 1, 2, \ldots, n+m). \tag{11}$$

We can maximize $l_P(\beta)$ to obtain the NPMLE of β. Details of the derivation of the profile log-likelihood is given in Appendix A.

The following lemma shows that there exists a unique ν satisfying (10) and (11).

Lemma 1. For every $\beta \in \mathbb{R}$, there exists a unique ν that satisfies (10) and (11). Further, the uniquely determined ν satisfies the inequality $\nu > 0$.

Proof.

Condition (11) is equivalent to

$$n + m\nu \mathrm{e}^{\beta z_i} > 0 \qquad (i = 1, 2, \ldots, n+m). \tag{12}$$

Let

$$\xi := \max_{1 \leq i \leq n+m} -\frac{n}{m} e^{-\beta z_i} < 0.$$

Then, (12) is equivalent to the inequality $\nu > \xi$. Let

$$h(\nu) := \sum_{i=1}^{n+m} \frac{1}{n + m\nu e^{\beta z_i}} - 1.$$

Then, $h(\nu)$ is monotonically decreasing on (ξ, ∞), and $h(\infty) = -1$. Since

$$h(0) = \sum_{i=1}^{n+m} \frac{1}{n} - 1 = \frac{n+m}{n} - 1 = \frac{m}{n} > 0,$$

there exists a unique ν that satisfies (10) and (11) in the interval $(0, \infty)$. Thus, the uniquely determined ν satisfies the inequality $\nu > 0$. □

3 Modified profile empirical likelihood

We consider the exponential-tilt model and derive a modification to the empirical profile log-likelihood corresponding to the conventional modification to the profile likelihood for parametric models (Barndorff-Nielsen, 1983).

Let λ be a parameter of interest, and ψ be a nuisance parameter. Then, the conventional modified profile likelihood is defined by

$$l_{MP}(\lambda) = l_P(\lambda) + M(\lambda)$$

and

$$M(\lambda) = -\frac{1}{2}\log\left|\hat{j}_\lambda\right| + \log\left|\frac{\partial\hat{\psi}}{\partial\hat{\psi}_\lambda}\right|, \tag{13}$$

where $l_P(\lambda)$ is the profile likelihood, and $\hat{j}_\lambda = j_{\psi\psi}(\lambda, \hat{\psi}_\lambda)$ is the observed information matrix.

3.1 The canonical form of the empirical likelihood

First, we represent the empirical log-likelihood

$$l_{n+m}(x, y; \beta, \boldsymbol{p}) = \beta\sum_{i=1}^{m} y_i + \sum_{i=1}^{n+m} L_i(x, y)\log p_i - m\log\left(\sum_{i=1}^{n+m} e^{\beta z_i} p_i\right),$$

where

$$L_i(x, y) = \sum_{j=1}^{n}\delta(x_j - z_i) + \sum_{j=1}^{m}\delta(y_j - z_i),$$

in the canonical form of the exponential family to obtain the modified profile empirical likelihood l_{MP}. We can calculate l_{MP} for other probability distributions that belong to an exponential family in the following manner:

Using $\sum_{i=1}^{n+m} L_i = n + m$, we have

$$\begin{aligned}
l_{n+m} =& \beta\sum_{i=1}^{m} y_i + \sum_{i=1}^{n+m} L_i\log p_i - m\log\left(\sum_{i=1}^{n+m} e^{\beta z_i} p_i\right) \\
=& \beta\sum_{i=1}^{m} y_i + \sum_{i=1}^{n+m-1} L_i\log p_i + L_{n+m}\log p_{n+m} - m\log\left(e^{\beta z_{n+m}} p_{n+m} + \sum_{i=1}^{n+m-1} e^{\beta z_i} p_i\right) \\
=& \beta\sum_{i=1}^{m} y_i + \sum_{i=1}^{n+m-1} L_i\log p_i + \left(n + m - \sum_{i=1}^{n+m-1} L_i\right)\log p_{n+m} \\
& - m\beta z_{n+m} - m\log p_{n+m} - m\log\left\{1 + \sum_{i=1}^{n+m-1} e^{\beta(z_i - z_{n+m})}\frac{p_i}{p_{n+m}}\right\} \\
=& \beta\sum_{i=1}^{m}(y_i - z_{n+m}) + \sum_{i=1}^{n+m-1} L_i\log\frac{p_i}{p_{n+m}} + n\log p_{n+m} \\
& - m\log\left\{1 + \sum_{i=1}^{n+m-1} e^{\beta(z_i - z_{n+m})}\frac{p_i}{p_{n+m}}\right\}.
\end{aligned} \tag{14}$$

Define

$$\theta_i = \log \frac{p_i}{p_{n+m}} = \log p_i - \log\left(1 - \sum_{k=1}^{n+m-1} p_k\right) \qquad (i = 1, 2, \ldots, n+m-1).$$

Then, since

$$\sum_{i=1}^{n+m-1} e^{\theta_i} = \sum_{i=1}^{n+m-1} \frac{p_i}{p_{n+m}} = \frac{1 - p_{n+m}}{p_{n+m}} = \frac{1}{p_{n+m}} - 1,$$

we have

$$p_{n+m} = \left(1 + \sum_{i=1}^{n+m-1} e^{\theta_i}\right)^{-1}. \tag{15}$$

From (14) and (15), we have

$$l_{n+m}(x, y; \beta, \theta) = \beta \sum_{i=1}^{m} (y_i - z_{n+m}) + \sum_{i=1}^{n+m-1} L_i \theta_i + \Phi(\beta, \theta), \tag{16}$$

where

$$\Phi(\beta, \theta) = -n \log\left(1 + \sum_{i=1}^{n+m-1} e^{\theta_i}\right) - m \log\left\{1 + \sum_{i=1}^{n+m-1} e^{\theta_i + \beta(z_i - z_{n+m})}\right\}.$$

Equation (16) corresponds to the canonical form of the log-likelihood of an exponential family, and (β, θ) is the natural parameter.

3.2 The modifying factor for an exponential family

Lemma 2 below shows that we can calculate the modifying factor $M(\beta)$ as follows for the model belonging to an exponential family.

Lemma 2. Suppose that l is the log-likelihood of an exponential family:

$$l(x; \lambda, \psi) = \sum_{i=1}^{p} \lambda_i \zeta_i(x) + \sum_{i=1}^{q} \psi_i \eta_i(x) + \Phi(\lambda, \psi)$$

and λ is a parameter of interest. Then, the modifying factor $M(\lambda)$ is given by

$$M(\lambda) = \frac{1}{2} \log \left|\Phi_{\psi\psi}(\lambda, \hat{\psi}_\lambda)\right|. \tag{17}$$

□

Proof. The observed information is given by

$$\hat{j}_\lambda := j_{\psi\psi}(\lambda, \hat{\psi}_\lambda) = \Phi_{\psi\psi}(\lambda, \hat{\psi}_\lambda). \tag{18}$$

From

$$l_{\lambda_i}(x; \hat{\lambda}, \hat{\psi}) = \zeta_i(x) + \Phi_{\lambda_i}(\hat{\lambda}, \hat{\psi}) = 0$$

and

$$l_{\psi_i}(x;\hat{\lambda},\hat{\psi}) = \eta_i(x) + \Phi_{\psi_i}(\hat{\lambda},\hat{\psi}) = 0,$$

the equation

$$l(\hat{\lambda},\hat{\psi};\lambda,\psi) = -\sum_{i=1}^{p} \lambda_i \Phi_{\lambda_i}(\hat{\lambda},\hat{\psi}) - \sum_{i=1}^{q} \psi_i \Phi_{\psi_i}(\hat{\lambda},\hat{\psi}) + \Phi(\lambda,\psi)$$

holds. Therefore, we have

$$l_{\psi_i}(\hat{\lambda},\hat{\psi};\lambda,\hat{\psi}_\lambda) = -\Phi_{\psi_i}(\hat{\lambda},\hat{\psi}) + \Phi_{\psi_i}(\lambda,\hat{\psi}_\lambda) = 0 \tag{19}$$

for a fixed λ. Differentiating (19) with respect to $\hat{\psi}_j$, it follows that

$$-\Phi_{\psi_i\psi_j}(\hat{\lambda},\hat{\psi}) + \sum_{k=1}^{q} \Phi_{\psi_i\psi_k}(\lambda,\hat{\psi}_\lambda)\frac{\partial \hat{\psi}_{k,\lambda}}{\partial \hat{\psi}_j} = 0.$$

Thus,

$$\frac{\partial \hat{\psi}_\lambda}{\partial \hat{\psi}} = (\Phi_{\psi\psi}(\lambda,\hat{\psi}_\lambda))^{-1}\Phi_{\psi\psi}(\hat{\lambda},\hat{\psi}).$$

This gives

$$\log\left|\frac{\partial \hat{\psi}}{\partial \hat{\psi}_\lambda}\right| = -\log\left|\frac{\partial \hat{\psi}_\lambda}{\partial \hat{\psi}}\right| = \log\left|\Phi_{\psi\psi}(\lambda,\hat{\psi}_\lambda)\right| - \log\left|\Phi_{\psi\psi}(\hat{\lambda},\hat{\psi})\right|. \tag{20}$$

Considering $-\log\left|\Phi_{\psi\psi}(\hat{\lambda},\hat{\psi})\right|$ to be a constant, we obtain (17) from (13), (18), and (20).

By substituting $p = 1,\ q = n+m-1,\ \lambda = \beta,\ \psi = \theta,\ \zeta(y) = \sum_{i=1}^{m}(y_i - z_{n+m})$, and $\eta_i(x,y) = L_i(x,y)$, we obtain the modifying factor $M(\beta)$ for the empirical profile likelihood of the exponential-tilt model as

$$\begin{aligned} M(\beta) =& \frac{1}{2}\log\left|\Phi_{\theta\theta}(\beta,\hat{\theta}_\beta)\right| \ , \\ \Phi_{\theta_i\theta_j}(\beta,\theta) =& \frac{n e^{\theta_i} e^{\theta_j}}{\left(1 + \sum_{k=1}^{n+m-1} e^{\theta_k}\right)^2} - \frac{m e^{\theta_i + \beta(z_i - z_{n+m})} e^{\theta_i + \beta(z_j - z_{n+m})}}{\left\{1 + \sum_{k=1}^{n+m-1} e^{\theta_k + \beta(z_k - z_{n+m})}\right\}^2} \ , \end{aligned}$$

and

$$\hat{\theta}_{i,\beta} = \theta_i(\beta,\hat{\boldsymbol{p}}_\beta) = -\log\left(\frac{n + m\nu e^{\beta z_i}}{n + m\nu e^{\beta z_{n+m}}}\right).$$

4 Numerical study

We present simulation results for the exponential-tilt model. We compare two estimators of β: one is the NPMLE based on the profile empirical likelihood, while the other is an estimator based on the modified profile empirical likelihood.

We assume that the underlying density $g(x)$ is a finite Gaussian mixture. We can adopt this assumption without loss of generality because any smooth probability density $g(x)$ can be approximated sufficiently well by a finite Gaussian mixture model. This is a convenient assumption for numerical studies because the tilted density $f(x)$ also becomes a finite Gaussian mixture.

Table 1: Means and standard deviations of the NPMLE and the modified NPMLE of β from 1000 simulations with $n = 3, 6, 9, \ldots, 30$. The true value of β is 1.

n	NPMLE	modified NPMLE
3	20.42 (33.84)	2.710 (3.264)
6	6.690 (20.91)	1.799 (2.262)
9	1.552 (3.931)	1.288 (0.9582)
12	1.245 (0.8801)	1.175 (0.7374)
15	1.156 (0.5450)	1.111 (0.5172)
18	1.125 (0.5033)	1.089 (0.4815)
21	1.114 (0.4250)	1.084 (0.4095)
24	1.113 (0.4176)	1.064 (0.4034)
27	1.095 (0.3618)	1.072 (0.3520)
30	1.066 (0.3498)	1.046 (0.3416)

Here we report a simple and typical example. We consider a mixture distribution of normal distributions

$$g(x) = 0.3N(x; -1, 1) + 0.7N(x; 0, 1),$$

where $N(\cdot; \mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2. The tilted density is $f(y) = e^{\alpha+\beta} g(y)$. Here we set $\beta = 1$. Then, α is given by

$$\alpha = -\log \left(0.3 \frac{1}{\sqrt{e}} + 0.7\sqrt{e} \right).$$

We generated 1000 data sets $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ from g and f, respectively, for each $n = 3, 6, 9, \ldots, 30$.

Table 1 and Figure 1 show the means and standard deviations of profile empirical likelihood estimates and modified profile empirical likelihood estimates based on the 1000 simulations.

We can see that the means of modified profile empirical likelihood estimates become closer to the true value than those based on the profile empirical likelihood for all $n = 3, 6, 9, \ldots 30$. In particular, the improvement in accuracy is large for a small sample size. We carried out numerical studies using various densities, and the results are similar to the result given above.

5 Discussion

In the numerical study in the previous section, the proposed procedure outperforms the method based on the conventional empirical profile likelihood. This result from the modification of the profile empirical likelihood for the exponential-tilt model seems natural because various results concerning ordinary likelihoods also hold for empirical likelihoods. In principle, the proposed procedure is also applicable to empirical profile likelihoods for various semiparametric models.

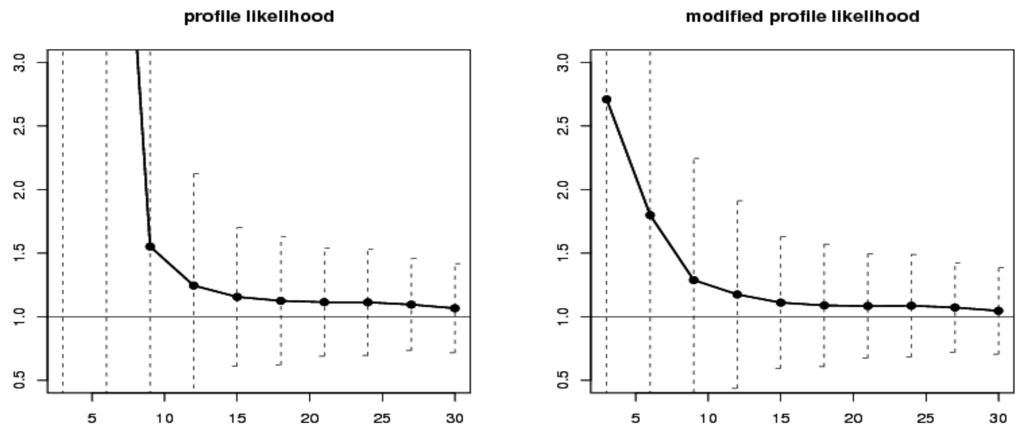

Figure 1: Means and standard deviations of the NPMLE (left) and the modified NPMLE (right) of β. Each dot shows the mean, and the error bar shows the mean $\pm$ standard deviation.

It is widely known that the improvement as a result of the modification for ordinary likelihoods is remarkable especially when the dimension of the nuisance parameters is large. In a semiparametric model, the nuisance parameter is infinite-dimensional. Model (8) that was used in the empirical likelihood theory has a nuisance parameter whose dimension is as large as the number of observations. This fact also seems to support the proposed modification.

For a complete understanding of the presented results, we require higher order asymptotic theory concerning empirical likelihood; however, this has not yet developed sufficiently and it can be an important topic for further research.

For logistic regression, various modifications to MLE have been suggested, e.g., Firth (1993). However, these modifications proposed are not equivalent to the present one, as far as the authors know. The relationship or correspondence between our procedure and such modifications is also an interesting topic.

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, London.

Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 343–365.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, New York.

Qin, J. and Zhang, B. (2005). Marginal likelihood, conditional likelihood and empirical likelihood: connections and applications. *Biometrika*, **92**, 251–270.

Appendix

A Derivation of the profile empirical likelihood

Let

$$\begin{aligned}\tilde{l}_{n+m} =& l_{n+m} - \gamma \left(\sum_{i=1}^{n+m} p_i - 1 \right) \\ =& \beta \sum_{i=1}^{m} y_i + \sum_{i=1}^{n+m} \log p_i - m \log \left(\sum_{i=1}^{n+m} e^{\beta z_i} p_i \right) - \gamma \left(\sum_{i=1}^{n+m} p_i - 1 \right),\end{aligned}$$

where γ is a Lagrange multiplier. Then, l_{n+m} is maximized under condition (4). The parameters p_i $(i = 1, 2, \ldots, n+m)$ satisfy the equations

$$\frac{\partial \tilde{l}_{n+m}}{\partial p_i} = \frac{1}{p_i} - m \cdot \frac{e^{\beta z_i}}{\left(\sum_{i=1}^{n+m} e^{\beta z_i} p_i \right)} - \gamma = 0 \quad (i = 1, 2, \ldots, n+m). \tag{21}$$

From (21), we have

$$\begin{aligned}0 =& \sum_{i=1}^{n+m} p_i \frac{\partial \tilde{l}_{n+m}}{\partial p_i} \\ =& \sum_{i=1}^{n+m} 1 - m \frac{1}{\left(\sum_{i=1}^{n+m} e^{\beta z_i} p_i \right)} \cdot \left(\sum_{i=1}^{n+m} e^{\beta z_i} p_i \right) - \gamma \sum_{i=1}^{n+m} p_i \\ =& n + m - m - \gamma = n - \gamma.\end{aligned} \tag{22}$$

Let

$$\nu = \left(\sum_{i=1}^{n+m} e^{\beta z_i} p_i \right)^{-1}. \tag{23}$$

The set of conditions (4), (22), and (23) are equivalent to the set of conditions (9) and (10). Inequality (6) is equivalent to (11). We obtain (9) from (21) and (22). Furthermore, (10) and (11) follow from (4), (6), and (9). □