Incorporating Covariates in the Gamma-Gamma Spend Model

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie[†]
www.brucehardie.com

March 2024

This note outlines how to incorporate the effects of covariates, be they time invariant or time varying, in the gamma-gamma spend model (Colombo and Jiang 1999, Fader et al. 2005). It is assumed that the reader is familiar with this model, especially with the derivations presented in Fader and Hardie (2013).

1 Base Model Assumptions

The gamma-gamma model of spend per transaction is based on the following three general assumptions:

- The monetary value (e.g., \$, £, €) of a customer's given transaction varies randomly around their average transaction value.
- Average transaction values vary across customers but do not vary over time for any given individual.
- The distribution of average transaction values across customers is independent of the transaction process.

More formally, for a customer with x transactions, let $z_1, z_2, \ldots, z_x$ denote the value of each transaction.

i) We assume the Z_i are iid gamma with shape parameter p and inverse scale parameter ν. It follows that $E(Z \,|\, p, \nu) = p/\nu$.

ii) We assume that heterogeneity in ν across customers is captured by a gamma distribution with shape parameter q and inverse scale parameter γ.

2 Adding Time-Invariant Covariate Effects

Let $\mathbf{w}$ denote the vector of time-invariant covariates that characterize the customer. (We suppress the customer subscript.) These covariates could include customer characteristics (e.g., male/female) or factors such as channel of acquisition. It is very important to note that this vector of covariates does not include an intercept.

We replace ν with $\nu = \nu_0 e^{-\mathbf{w}\boldsymbol{\beta}'}$ (note the $-$ sign), which means

$$E(Z \,|\, p, \nu_0, \boldsymbol{\beta}; \mathbf{w}) = \frac{p}{\nu_0} e^{\mathbf{w}\boldsymbol{\beta}'}.$$

We assume that heterogeneity in ν_0 across customers is captured by a gamma distribution with shape parameter q and inverse scale parameter γ.

It follows that

$$\begin{aligned} E(Z \,|\, p, q, \gamma, \boldsymbol{\beta}; \mathbf{w}) &= \int_0^\infty E(Z \,|\, p, \nu_0, \boldsymbol{\beta}; \mathbf{w}) g(\nu_0 \,|\, q, \gamma) d\nu_0 \\ &= \int_0^\infty \frac{p}{\nu_0} e^{\mathbf{w}\boldsymbol{\beta}'} \frac{\gamma^q \nu_0^{q-1} e^{-\gamma \nu_0}}{\Gamma(q)} d\nu_0 \\ &= p e^{\mathbf{w}\boldsymbol{\beta}'} \int_0^\infty \frac{\gamma^q \nu_0^{q-2} e^{-\gamma \nu_0}}{\Gamma(q)} d\nu_0 \\ &= \frac{p\gamma}{q-1} e^{\mathbf{w}\boldsymbol{\beta}'}, \;\; q > 1. \end{aligned} \tag{1}$$

This reduces to (3) in Fader and Hardie (2013) when $\boldsymbol{\beta} = \mathbf{0}$.

In the appendix we document three ways of formulating the likelihood function for the no-covariate model. For the case of time-invariant covariates, we can take any of these approaches. However, only the first one is feasible for the case of time-varying covariates. So as to maintain consistency between the time-invariant and time-varying covariates cases, we present the following derivation of the likelihood function.

By definition,

$$\begin{aligned} L(p, \nu_0, \boldsymbol{\beta} \,|\, z_1, z_2, \ldots, z_x, \mathbf{w}) &= \prod_{i=1}^{x} f(z_i \,|\, p, \nu_0 e^{-\mathbf{w}\boldsymbol{\beta}'}) \\ &= \prod_{i=1}^{x} \frac{\left(\nu_0 e^{-\mathbf{w}\boldsymbol{\beta}'}\right)^p z_i^{p-1} e^{-\nu_0 e^{-\mathbf{w}\boldsymbol{\beta}'} z_i}}{\Gamma(p)} \\ &= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} e^{-\mathbf{w}\boldsymbol{\beta}' p x}}{\Gamma(p)^x} \nu_0^{px} e^{-\nu_0 y e^{-\mathbf{w}\boldsymbol{\beta}'}}, \end{aligned}$$

where

$$y = \sum_{i=1}^{x} z_i$$

is the customer's total observed spend across the x transactions. Next we integrate over the distribution of ν_0:

$$
\begin{aligned}
L(&p, q, \gamma, \boldsymbol{\beta} \mid z_1, \dots, z_x, \mathbf{w}) \\
&= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} e^{-\mathbf{w}\boldsymbol{\beta}' px}}{\Gamma(p)^x} \int_0^\infty \nu_0^{px} e^{-\nu_0 y e^{-\mathbf{w}\boldsymbol{\beta}'}} \frac{\gamma^q \nu_0^{q-1} e^{-\gamma\nu_0}}{\Gamma(q)} d\nu_0 \\
&= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} e^{-\mathbf{w}\boldsymbol{\beta}' px}}{\Gamma(p)^x} \int_0^\infty \frac{\gamma^q \nu_0^{px+q-1} e^{-\nu_0(\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'})}}{\Gamma(q)} d\nu \\
&= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} e^{-\mathbf{w}\boldsymbol{\beta}' px}}{\Gamma(p)^x} \frac{\Gamma(px+q)}{\Gamma(q)} \frac{\gamma^q}{(\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'})^{px+q}} \\
&= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1}}{\Gamma(p)^x} \frac{\Gamma(px+q)}{\Gamma(q)} \left(\frac{\gamma}{\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'}}\right)^q \left(\frac{e^{-\mathbf{w}\boldsymbol{\beta}'}}{\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'}}\right)^{px}. \qquad (2)
\end{aligned}
$$

This reduces to the equivalent no-covariate likelihood function, (A1), when $\boldsymbol{\beta} = \mathbf{0}$ (which, as noted in the appendix, is not equivalent to the function presented in Fader et al. (2005) and Fader and Hardie (2013).)

The posterior distribution of ν_0 is

$$
\begin{aligned}
g(&\nu_0 \mid p, q, \gamma, \boldsymbol{\beta}; z_1, \dots, z_x, \mathbf{w}) \\
&= \frac{L(p, \nu_0, \boldsymbol{\beta} \mid z_1, z_2, \dots, z_x, \mathbf{w}) g(\nu_0 \mid q, \gamma)}{L(p, q, \gamma, \boldsymbol{\beta} \mid z_1, \dots, z_x, \mathbf{w})} \\
&= \frac{\frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} e^{-\mathbf{w}\boldsymbol{\beta}' px}}{\Gamma(p)^x} \nu_0^{px} e^{-\nu_0 y e^{-\mathbf{w}\boldsymbol{\beta}'}} \frac{\gamma^q \nu_0^{q-1} e^{-\gamma\nu_0}}{\Gamma(q)}}{\frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1}}{\Gamma(p)^x} \frac{\Gamma(px+q)}{\Gamma(q)} \left(\frac{\gamma}{\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'}}\right)^q \left(\frac{e^{-\mathbf{w}\boldsymbol{\beta}'}}{\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'}}\right)^{px}} \\
&= \frac{\left(\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'}\right)^{px+q} \nu_0^{px+q-1} e^{-\nu_0\left(\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'}\right)}}{\Gamma(px+q)}. \qquad (3)
\end{aligned}
$$

In other words, the posterior distribution of ν_0 is gamma with shape parameter $px + q$ and inverse scale parameter $\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'}$. Given (1), it follows that the conditional expectation is

$$
E(Z \mid p, q, \gamma, \boldsymbol{\beta}; z_1, \dots, z_x, \mathbf{w}) = \frac{p\left(\gamma + y e^{-\mathbf{w}\boldsymbol{\beta}'}\right)}{px + q - 1} e^{\mathbf{w}\boldsymbol{\beta}'}. \qquad (4)
$$

This reduces to (5) in Fader and Hardie (2013) when $\boldsymbol{\beta} = \mathbf{0}$.[1]

[1]At first glance, it may seem surprising that this reduces to Fader and Hardie's (2013) result when the no-covariate likelihood functions are different. We note, however, that the "offending term" in the likelihood function ($\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} / \Gamma(p)^x$) cancels out in the derivation of the posterior distribution. The conditional expectation numbers themselves will differ due to the different estimates of p, q, γ.

3 Adding Time-Varying Covariate Effects

Let $\mathbf{w}_i$ denote the vector of the values of the time-varying covariates at the ith transaction. As with $\mathbf{w}$, it is very important to note that this vector of covariates does not include an intercept. We relax the assumption that a customer's average transaction value does not vary over time: we assume $Z_i \sim \text{gamma}(p, \nu_i)$ where $\nu_i = \nu_0 e^{-\mathbf{w}_i \boldsymbol{\beta}'}$. We assume that heterogeneity in ν_0 across customers is captured by a gamma distribution with shape parameter q and inverse scale parameter γ.

It follows that

$$E(Z_i \,|\, p, q, \gamma, \boldsymbol{\beta}; \mathbf{w}_i) = \frac{p\gamma}{q-1} e^{\mathbf{w}_i \boldsymbol{\beta}'}, \; q > 1\,. \tag{5}$$

By definition,

$$\begin{aligned}
&L(p, \nu_0, \boldsymbol{\beta} \,|\, z_1, \ldots, z_x, \mathbf{w}_1, \ldots, \mathbf{w}_x) \\
&\quad = \prod_{i=1}^{x} f(z_i \,|\, p, \nu_i) \\
&\quad = \prod_{i=1}^{x} \frac{\left(\nu_0 e^{-\mathbf{w}_i \boldsymbol{\beta}'}\right)^p z_i^{p-1} e^{-\nu_0 e^{-\mathbf{w}_i \boldsymbol{\beta}'} z_i}}{\Gamma(p)} \\
&\quad = \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} e^{-p \sum_{i=1}^{x} \mathbf{w}_i \boldsymbol{\beta}'}}{\Gamma(p)^x} \nu_0^{px} e^{-\nu_0 \sum_{i=1}^{x} z_i e^{-\mathbf{w}_i \boldsymbol{\beta}'}}.
\end{aligned}$$

To simplify the equation, let

$$\begin{aligned}
A(x) &= e^{-p \sum_{i=1}^{x} \mathbf{w}_i \boldsymbol{\beta}'} \\
B(x) &= \sum_{i=1}^{x} z_i e^{-\mathbf{w}_i \boldsymbol{\beta}'}
\end{aligned}$$

which gives us

$$L(p, \nu_0, \boldsymbol{\beta} \,|\, z_1, \ldots, z_x, \mathbf{w}_1, \ldots, \mathbf{w}_x) = \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} A(x)}{\Gamma(p)^x} \nu_0^{px} e^{-\nu_0 B(x)}.$$

Next we integrate over the distribution of ν_0:

$$\begin{aligned}
&L(p, q, \gamma, \boldsymbol{\beta} \,|\, z_1, \ldots, z_x, \mathbf{w}_1, \ldots, \mathbf{w}_x) \\
&\quad = \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} A(x)}{\Gamma(p)^x} \int_0^\infty \nu_0^{px} e^{-\nu_0 B(x)} \frac{\gamma^q \nu_0^{q-1} e^{-\gamma \nu_0}}{\Gamma(q)} d\nu_0 \\
&\quad = \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} A(x)}{\Gamma(p)^x} \int_0^\infty \frac{\gamma^q \nu_0^{px+q-1} e^{-\nu_0(\gamma + B(x))}}{\Gamma(q)} d\nu \\
&\quad = \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1} A(x)}{\Gamma(p)^x} \frac{\Gamma(px+q)}{\Gamma(q)} \frac{\gamma^q}{(\gamma + B(x))^{px+q}}\,.
\end{aligned} \tag{6}$$

Note that when $\mathbf{w}_i = \mathbf{w}\ \forall\, i$, $A(x) = e^{-\mathbf{w}\boldsymbol{\beta}' px}$ and $B(x) = ye^{-\mathbf{w}\boldsymbol{\beta}'}$, and (6) reduces to (2).

The posterior distribution of ν_0 is

$$
\begin{aligned}
g(\nu_0 \,|\, p, q, \gamma, \boldsymbol{\beta}; z_1, \dots, z_x, \mathbf{w}_1, \dots, \mathbf{w}_x) \\
&= \frac{L(p, \nu_0, \boldsymbol{\beta} \,|\, z_1, z_2, \dots, z_x, \mathbf{w}_1, \dots, \mathbf{w}_x) g(\nu_0 \,|\, q, \gamma)}{L(p, q, \gamma, \boldsymbol{\beta} \,|\, z_1, \dots, z_x, \mathbf{w}_1, \dots, \mathbf{w}_x)} \\
&= \frac{\dfrac{\big\{\prod_{i=1}^x z_i\big\}^{p-1} A(x)}{\Gamma(p)^x} \nu_0^{px} e^{-\nu_0 B(x)} \dfrac{\gamma^q \nu_0^{q-1} e^{-\gamma\nu_0}}{\Gamma(q)}}{\dfrac{\big\{\prod_{i=1}^x z_i\big\}^{p-1} A(x)}{\Gamma(p)^x} \dfrac{\Gamma(px+q)}{\Gamma(q)} \dfrac{\gamma^q}{(\gamma + B(x))^{px+q}}} \\
&= \frac{\big(\gamma + B(x)\big)^{px+q} \nu_0^{px+q-1} e^{-\nu_0(\gamma + B(x))}}{\Gamma(px+q)} \,. \qquad (7)
\end{aligned}
$$

In other words, the posterior distribution of ν_0 is gamma with shape parameter $px + q$ and inverse scale parameter $\gamma + B(x)$. It follows that the conditional expectation is

$$
\begin{aligned}
E(Z_j \,|\, p, q, \gamma, \boldsymbol{\beta}; z_1, \dots, z_x, \mathbf{w}_1, \dots, \mathbf{w}_x; \mathbf{w}_j) \\
&= \frac{p(\gamma + B(x))}{px + q - 1} e^{\mathbf{w}_j \boldsymbol{\beta}'},\ j > x \,. \qquad (8)
\end{aligned}
$$

References

Colombo, Richard and Weina Jiang (1999), "A Stochastic RFM Model," *Journal of Interactive Marketing*, **13** (Summer), 2–12.

Fader, Peter S. and Bruce G. S. Hardie (2013), "The Gamma-Gamma Model of Monetary Value." `http://brucehardie.com/notes/025/`.

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), "RFM and CLV: Using Iso-value Curves for Customer Base Analysis," *Journal of Marketing Research*, **42** (November), 415–430.

Appendix

There are three ways of formulating the model likelihood function for the no-covariate gamma-gamma model.

Approach 1 — Individual transactions

By definition,

$$\begin{aligned} L(p,\nu \mid z_1,\dots,z_x) &= \prod_{i=1}^{x} f(z_i \mid p,\nu) \\ &= \prod_{i=1}^{x} \frac{\nu^p z_i^{p-1} e^{-\nu z_i}}{\Gamma(p)} \\ &= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1}}{\Gamma(p)^x} \nu^{px} e^{-\nu y}, \end{aligned}$$

where

$$y = \sum_{i=1}^{x} z_i$$

is the customer's total observed spend across the x transactions. Integrating over the distribution of ν gives us

$$\begin{aligned} L(p,q,\gamma \mid z_1,\dots,z_x) &= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1}}{\Gamma(p)^x} \int_0^\infty \nu^{px} e^{-\nu y} \frac{\gamma^q \nu^{q-1} e^{-\gamma\nu}}{\Gamma(q)} d\nu \\ &= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1}}{\Gamma(p)^x} \int_0^\infty \frac{\gamma^q \nu^{px+q-1} e^{-\nu(\gamma+y)}}{\Gamma(q)} d\nu \\ &= \frac{\left\{\prod_{i=1}^{x} z_i\right\}^{p-1}}{\Gamma(p)^x} \frac{\Gamma(px+q)}{\Gamma(q)} \frac{\gamma^q}{(\gamma+y)^{px+q}}\,. \end{aligned} \tag{A1}$$

Approach 2 — Total Spend

Given the convolution property of the gamma distribution, it follows that $Y \mid x \sim \text{gamma}(px,\nu)$. Therefore,

$$L(p,\nu \mid y,x) = \frac{\nu^{px} y^{px-1} e^{-\nu y}}{\Gamma(px)}\,.$$

Integrating over the distribution of ν gives us

$$\begin{aligned} L(p,q,\gamma \mid y,x) &= \frac{y^{px-1}}{\Gamma(px)} \int_0^\infty \nu^{px} e^{-\nu y} \frac{\gamma^q \nu^{q-1} e^{-\gamma\nu}}{\Gamma(q)} d\nu \\ &= \frac{y^{px-1}}{\Gamma(px)} \int_0^\infty \frac{\gamma^q \nu^{px+q-1} e^{-\nu(\gamma+y)}}{\Gamma(q)} d\nu \\ &= \frac{y^{px-1}}{\Gamma(px)} \frac{\Gamma(px+q)}{\Gamma(q)} \frac{\gamma^q}{(\gamma+y)^{px+q}}\,. \end{aligned} \tag{A2}$$

Approach 3 — Average Spend

The customer's observed average transaction value is

$$\bar{z} = \sum_{i=1}^{x} z_i / x \,.$$

Given the convolution and scaling properties of the gamma distribution, it follows that $\bar{Z} \sim \text{gamma}(px, \nu x)$. Therefore,

$$\begin{aligned}
L(p, \nu \,|\, \bar{z}, x) &= \frac{(\nu x)^{px} \bar{z}^{px-1} e^{-\nu x \bar{z}}}{\Gamma(px)} \\
&= \frac{x^{px-1}}{x^{px-1}} \frac{(\nu x)^{px} \bar{z}^{px-1} e^{-\nu x \bar{z}}}{\Gamma(px)} \\
&= x \frac{\nu^{px} (x\bar{z})^{px-1} e^{-\nu x \bar{z}}}{\Gamma(px)} \\
&= x \frac{\nu^{px} y^{px-1} e^{-\nu y}}{\Gamma(px)} \,.
\end{aligned}$$

Integrating over the distribution of ν gives us obviously gives us

$$L(p, q, \gamma \,|\, \bar{z}, x) = x \frac{y^{px-1}}{\Gamma(px)} \frac{\Gamma(px+q)}{\Gamma(q)} \frac{\gamma^q}{(\gamma + y)^{px+q}} \,. \tag{A3}$$

This is the approach used by Fader et al. (2005) and Fader and Hardie (2013).

Are the three approaches equivalent?

As (A2) and (A3) differ by a factor (x) that is independent of the parameter values, the values of p, q, γ that maximize (A2) will obviously be the same as those that maximize (A3).

However, (A1) differs from (A2) and (A3) by a factor that is a function of p, $\left\{ \prod_{i=1}^{x} z_i \right\}^{p-1} / \Gamma(p)^x$ in the case of (A2). Therefore, the values of p, q, γ that maximize (A1) will not be the same as those that maximize (A2) and (A3).