BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2018-29

A Look Inside the Box: Combining Aggregate and Marginal Distributions to Identify Joint Distributions

Marie-Hélène Felt

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2018-29

July 2018

A look Inside the Box: Combining Aggregate and Marginal Distributions to Identify Joint Distributions

by

Marie-Hélène Felt

Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
mfelt@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

This paper is based on the second and third chapters of my PhD thesis. I am grateful to Kim P. Huynh, Heng Chen, Marcel Voia and Lynda Khalaf for their support and guidance. I thank Yu Zhu, Yuya Sasaki, Thomas Lemieux and members of the Economic Research and Analysis team in the Currency Department at the Bank of Canada for comments and suggestions, and Boyan Bejanov for his technical advice. Ipsos Reid's collaborative effort for data collection is also acknowledged.

The views expressed in this paper are mine. No responsibility for them should be attributed to the Bank of Canada.

Abstract

This paper proposes a method for estimating the joint distribution of two or more variables when only their marginal distributions and the distribution of their aggregates are observed. Nonparametric identification is achieved by modelling dependence using a latent common-factor structure. Multiple examples are given of data settings where multivariate samples from the joint distribution of interest are not readily available, but some aggregate measures are observed. In the application, intra-household distributions are recovered by combining individual-level and household-level survey data. I show that, for individuals living in couple relationships, personal cash-management practices are significantly influenced by the partner's use of cash and stored-value cards. This finding implies that, for some methods of payment at least, ignoring the partner's impact might lead to spurious regression results due to an omitted variable bias.

Bank topics: Econometric and statistical methods; Bank notes; Digital currencies
JEL codes: C; C14; D14; E41

Résumé

Le présent document propose une méthode d'estimation de la distribution conjointe de deux variables ou plus lorsque seules leur distribution marginale et la distribution de leur agrégat sont observées. L'identification non paramétrique se fait par la modélisation de la dépendance au moyen d'une structure de facteurs communs latents. De nombreux exemples de contextes empiriques sont fournis, dans lesquels des échantillons multivariés de la distribution conjointe qui nous intéresse ne sont pas facilement accessibles, mais certaines mesures agrégées sont observées. Dans l'application, les distributions intra-ménages sont obtenues en combinant des données d'enquêtes individuelles et d'autres se rapportant au ménage en entier. Je démontre que dans le cas des personnes qui vivent en couple, les pratiques de gestion des liquidités personnelles sont grandement influencées par l'utilisation que fait le conjoint de l'argent comptant et des cartes prépayées. Cette constatation porte à croire que, pour certains modes de paiement du moins, le fait d'ignorer l'influence du conjoint peut entraîner des résultats de régression erronés en raison de l'omission de variables pertinentes.

Sujets : Méthodes économétriques et statistiques ; Billets de banque ; Monnaies numériques
Codes JEL : C; C14; D14; E41

Non-technical summary

In many circumstances, the data available to the empirical researcher are not direct observations of the quantity of interest. For example, variables are often subject to measurement error. Another such situation occurs when data are grouped, so that the units observed are aggregates of the smaller, individual units of interest. In that case, it is common practice to use aggregates as proxies for individual units, and relationships across aggregates as proxies for that across individual variables. For example, in the absence of firm-level data, one may employ industry-level averages as if they were generated by the behaviour of a representative firm.

In this paper, I exploit aggregated data to study individual relationships across the units that constitute these aggregates. For example, in the application, I use household-level data (i.e., data aggregated at the household level) to study relationships between members constituting the households. In other words, I investigate intra-household questions in the absence of intra-household, disaggregated data. This is advantageous because intra-household data, where multiple members of the same household are observed, are very costly to collect.

The data requirement of my method is to observe the aggregate and marginal distributions that arise from the same unobserved joint distribution of interest. In the application, I recover the joint distribution of the two household heads in couple households by combining three independently observed distributions: the two marginal distributions of the two heads, and the marginal distribution of the couple's aggregate measure. Unlike statistical matching, my approach does not require units in common across the samples where the marginal distributions are observed. Instead, my identification result relies on modelling intra-aggregate dependence using a linear, latent common factor structure.

Several examples are given of data settings where multivariate samples from the joint distribution of interest are not readily available, but some aggregate measures are observed. In addition to unit aggregation, the method also applies to data aggregated over time. Time-aggregated data arise where some quantity is measured over a long period, although variations can occur at higher frequencies. I discuss two examples of time aggregation in detail in the paper.

In the last section of the paper, I apply the developed methodology to analyze intra-household payment behaviours in the absence of intra-household data. Using individual- or household-level data, previous research shows the impact of payment innovations, such as contactless credit cards and stored-value cards, on the recourse to cash for retail payments. In my analysis, I further explore this issue while taking the intra-household dimension into account. I show that, for individuals living in couple relationships, personal cash-management practices are significantly influenced by the partner's usage of cash and stored-value cards. This finding implies that, for some methods of payment at least, ignoring the partner's impact might bias the estimation results.

1 Introduction

Aggregated data are used in many contexts in economics. They arise when observational units are aggregates of smaller "individual" units of interest. Aggregated data can be a default solution for studying individual relationships when information is not available at the preferred level of analysis. In this case, relationships across aggregate variables serve as proxies for that across individual variables. There, aggregation constitutes a limitation of the data, one which creates particular challenges to the empirical researcher.[1]

In this paper, I take a contrasting approach and exploit the information contained in aggregates to study individual relationships across variables *within* the aggregates. My main result is to show that, in the absence of a multivariate sample of $(X_1, ..., X_K)$, their joint distribution can be nonparametrically recovered from their aggregate $\overline{X} \equiv \sum_{k=1}^{K} X_k$ when their K individual marginal distributions are also known.

Multiple contexts provide the data requirement for this identification result, which is to observe the aggregate and marginal distributions that arise from the same unobserved joint distribution of interest. It occurs in the case of individual and household surveys that provide independent samples where the same quantity is measured at the individual level in the one, and at the household level in the other.[2] In such a setting, the method that I propose permits investigation of intra-household questions in the absence of intra-household data. Potential applications in the context of household economics are numerous, regarding for example consumption, income or financial assets. In the empirical application presented in the last section of the paper, I analyze intra-household payment behaviours by combining individual-level and aggregate household-level payment survey data.

Other data situations where unit aggregation can be exploited pertain to geographical or price aggregation.[3] But aggregation also happens on the time dimension. For example,

[1] Robinson (1950) famously demonstrated, for instance, that correlations and associations observed in aggregated data might differ in magnitude and sign from those observed in individual-level data.

[2] By independent samples I mean samples that are independently drawn from the same population and have no or very few units in common, so that statistically matching them is not an option.

[3] Piterbarg (2011) considers the situation where options markets provide information on the distributions

aggregated data arise where some quantity is measured over a long time period, although variations can occur at higher frequencies. I discuss two examples of time aggregation in more detail in the paper.

I employ nonparametric deconvolution techniques to obtain my identification results, and thereby contribute to a strand of the nonparametric deconvolution literature that considers aggregated (or grouped) data models. There, the data are not contaminated with measurement error, but instead are measurements where the quantity of interest is accumulated. Most often, the addends are assumed to be independent and identically distributed random variables, and their marginal distribution is the object of interest; see Meister (2009) and Wagner (2009). In this paper, however, the individual marginal distributions $f_{X_1}, ..., f_{X_K}$ are directly identified from the observed data. It is the joint distribution of $(X_1, ..., X_K)$ that is the unknown object of interest.

To model intra-aggregate dependence without making strong functional form assumptions, I follow Linton and Whang (2002) and use a latent common factor structure. Its flexible specification allows for heterogeneity both within aggregates (i.e., across addends) and across aggregates of different sizes.[4] The models I employ belong to the family of linear multi-factor models with independent unobserved factors and known factor loadings. This type of model is applied in a wide variety of settings, such as measurement error and panel data analysis. In many applications it is the factors' distributions that are of main interest, and their identification is obtained from the multivariate distribution of the X_k measurements; see, e.g., Horowitz and Markatou (1996), Li and Vuong (1998), Székely and Rao (2000) and Bonhomme and Robin (2010). Interestingly, I show that some of these models are identified under much weaker requirements than knowing the full joint distribution of the data. For example, I establish a more general variant of the well-known Lemma of Kotlarski (Kotlarski, 1967) by showing that it still holds when, instead of the joint distribution, only

of two individual underlyings (e.g., stock prices or interest rates) as well as their spread.

[4] In contrast, Linton and Whang's (2002) result relies on the availability of (at least) two different aggregate (or group) sizes (e.g., household sizes) and on strong homogeneity assumptions within and across groups.

the marginal and the aggregate distributions are known.

Although they may be of interest in and of themselves, in the context of this paper, recovering the factors' distributions most importantly yields the identification of the final target, the joint distribution of the X_k variables. For that purpose, the non-availability of a multivariate sample of the X_k variables constitutes a missing data problem that I handle by data combination. Yet, in contrast with most sample combination strategies, my identification result does not rely on the presence of common units and/or variables across samples to be combined; see Ridder and Moffitt (2007). Rather, I rely on additional information obtained from an independent sample where an aggregate measure of the X_k variables is observed.

In the last section of the paper, I apply the developed methodology to analyze intra-household payment behaviours in the absence of intra-household data. Fung et al. (2014) and Chen et al. (2017), using respectively individual- and household-level data, evidence the impact of recent payment innovations, such as contactless credit cards and stored-value cards, on the recourse to cash for retail payments. In my analysis, I further explore this issue while taking the intra-household dimension into account. I show that, for individuals living in couple relationships, personal cash-management practices in terms of withdrawals and holdings are significantly influenced by the partner's usage of cash and stored-value cards. This finding implies that, for some methods of payment at least, ignoring the partner's impact might lead to spurious regression results due to an omitted variable bias.

The main goal of this paper is to propose a new way of exploiting the information content of aggregate measures. I consider various data settings and examples to illustrate the approach, but its implementation can be tailored to each application. The remainder of the paper is organized as follows. In Section 2, I provide the main identification result, as well as examples and extensions. In Section 3, an estimator is proposed and its consistency is shown. Its small sample behaviour is analyzed in Section 4. Finally, an application of the method to analyze intra-household behaviours in the absence of intra-household data is presented in Section 5. Section 6 concludes. All the proofs are deferred to the appendix.

2 Identification

I first describe the basic setup and modelling assumptions and establish the main identification result. Examples and extensions are provided next.

2.1 Main model and identification result

$X_1, ..., X_K$ are K random variables observed separately so that only their marginal distributions are known. I also observe, in an independent data set, the distribution of the aggregate $\overline{X} = \sum_{k=1}^{K} X_k$. I am interested in the multivariate distribution of $\mathbf{X} \equiv (X_1, ..., X_K)'$.

In this basic set-up, I employ a simple model where the individual X_k variables are decomposed as the sum of two orthogonal components, an idiosyncratic component and a shared common component, as follows:

$$X_k = V_k + \rho_k V_0 \quad \text{for } k = 1, ..., K, \tag{1}$$

where $V_0, V_1, ..., V_K$ are unobserved, mutually independent real random variables and $\rho_1, ..., \rho_k$ are known coefficients such that $\sum_{k=1}^{K} \rho_k \neq 0$.

Such latent factor decompositions are used in many fields of economics and finance for modelling the dependency structure of multivariate data. The simple one-factor structure of Model (1) allows for both heterogeneity and dependence across the X_k random variables. Its generalization to multiple common components is considered in Section 2.3. Everywhere in this paper the ρ_k coefficients are assumed to be known or consistently estimated. These are natural or widely used assumptions in various contexts, such as panel data models and many labour economics applications; see examples 2 and 3 of Section 2.2. In other cases, we may be able to make assumptions about the value of the coefficients, based for instance on external information, and then test some of the model's implications in the data.[5]

[5]Alternatively, one might wish to impose restrictions on some of the moments of the factors' distributions,

Given the linearity and independence assumptions in the model under consideration, it is convenient to work with characteristic functions (c.f.). The c.f. of a random variable X, with $F_X(x)$ as its cumulative distribution function (c.d.f.), is defined as

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} dF_X(x), \tag{2}$$

$$= E\left(e^{itX}\right), \tag{3}$$

$t \in \mathbb{R}$, where $i = \sqrt{-1}$. The c.f. of any random variable always exists, and it completely determines its distribution.[6]

The joint density of the random variables $(X_1,...,X_K)$ is uniquely determined by the multivariate c.f. $\phi_{\mathbf{X}}(\mathbf{t}) = E\left(e^{i\mathbf{t}'\mathbf{X}}\right)$, $\mathbf{t} \in \mathbb{R}^K$. The latter, under Model (1) and by the convolution theorem, can be written as

$$\phi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^{K} \phi_{V_k}(t_k)\phi_{V_0}\left(\mathbf{t}'\boldsymbol{\rho}\right), \tag{4}$$

for all $\mathbf{t} = (t_1, ..., t_K)' \in \mathbb{R}^K$. By the convolution theorem, Model (1) also implies

$$\phi_{X_k}(t_k) = \phi_{V_k}(t_k)\phi_{V_0}\left(\rho_k t_k\right), \tag{5}$$

for all $t_k \in \mathbb{R}$. By substitution of Equation (5) into (4), I obtain

$$\phi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^{K} \phi_{X_k}(t_k) \frac{\phi_{V_0}(\mathbf{t}'\boldsymbol{\rho})}{\prod_{k=1}^{K} \phi_{V_0}(\rho_k t_k)}. \tag{6}$$

To recover the joint c.f. of $(X_1,...,X_K)$ I first obtain the identification of the c.f. of the common component, ϕ_{V_0}. In fact, not only V_0's but all the V_k components' distributions are identified by the marginal distributions of the X_k variables and that of $\overline{X}$, under Model (1) and some regularity conditions. This is formally stated in the following lemma.

and identify the ρ_k coefficients from the data. This option is not pursued further in this paper.

[6] The reader is referred to Lukacs (1970) and Ushakov (1999) for the formal statements and proofs of c.f. properties called upon in this paper.

Lemma 1. *Under the assumptions of Model* (1)*, if the c.f.s of* $X_1, ..., X_K$ *do not vanish, then the distributions of* $X_1, ..., X_K$ *and* $\overline{X}$ *determine the distributions of* V_0, V_1,..., V_K *up to location.*

The V_k's distributions are identified only up to a location parameter. In practice, one can normalize the latent components' means to zero and work with demeaned X_k variables. Lemma 1 is of value in and of itself, especially in cases where the V_k components are the objects of interest. I give such examples later. When the common factor structure is mainly a device for modelling dependence, the final objective is to derive the multivariate distribution. Theorem 1 now presents my main identification result:

Theorem 1. *Under the assumptions of Model* (1)*, if the c.f.s of* $X_1, ..., X_K$ *do not vanish, then the distributions of* $X_1, ..., X_K$ *and* $\overline{X}$ *just identify the joint distribution of* $(X_1, ..., X_K)$*.*

Remark 1:

The assumption that the c.f.s do not have real zeros is usual in the nonparametric deconvolution literature. It is satisfied by most common continuous distributions as well as many discrete distributions.[7] Note that under Model (1), this assumption on the c.f.s of the X_k variables is equivalent to assuming that the c.f.s of V_0, V_1,..., V_K do not vanish, or that the c.f. of $\overline{X}$ does not vanish. Alternatively, the non-vanishing assumption can be replaced by milder conditions such as the analyticity of one of the latent component's c.f., along the lines of Evdokimov and White (2012).

Remark 2:

Assumption $\sum_{k=1}^{K} \rho_k \neq 0$ ensures the common component V_0 appears in the decomposition of $\overline{X}$ that follows from Model (1). However, in the case where the ρ_k coefficients sum up to zero, additional requirements can lead to identification of ϕ_{V_0}.[8]

[7] As discussed in D'Haultfoeuille (2011), the only common continuous distributions that fail to satisfy this condition are the uniform and triangular distributions.

[8] For example, when $K = 2$ and $\rho_1 + \rho_2 = 0$, $\phi_{X_1}(t)\phi_{X_2}(t)/\phi_{\overline{X}}(t) = \phi_{V_0}(\rho_1 t)\,\phi_{V_0}(-\rho_1 t) = |\phi_{V_0}(\rho_1 t)|^2$, so that in general only the modulus of ϕ_{V_0} is identified. However, ϕ_{V_0} is identified under the additional assumption that V_0 is symmetrically distributed about zero. In that case, ϕ_{V_0} is real and even, i.e. $\phi_{V_0}(-t) = \phi_{V_0}(t)\ \forall t \in \mathbb{R}$, and we have $\phi_{X_1}(t)\phi_{X_2}(t)/\phi_{\overline{X}}(t) = [\phi_{V_0}(\rho_1 t)]^2$.

Remark 3:

Conditional versions of Lemma 1 and Theorem 1 can be established under the assumption of conditional independence of $V_0, V_1, ..., V_K$ in Model (1). Working with conditional c.f.s, it can be shown that the conditional multivariate distribution of $\mathbf{X}$ (given $Z = z$) is identified by the conditional distributions of $X_1, ..., X_K$ and $\overline{X}$ (given $Z = z$).[9] One might also want to control for covariates by working with fitted residuals (from a set of first-stage regressions) instead of raw variables, but note that this requires that the individual residuals add up to the aggregate residuals.[10]

2.2 Examples

Example 1: Cross-sectional model

In the special case where $K = 2$ and $\rho_1 = \rho_2 = 1$, Model (1) boils down to:

$$X_{1i} = V_{1i} + V_{0i}, \tag{7a}$$

$$X_{2i} = V_{2i} + V_{0i}, \tag{7b}$$

where V_{0i}, V_{1i} and V_{2i} are unobserved, mutually independent real random variables, i.i.d. across i.

Lemma 1 and Theorem 1 apply directly to Model (7). I use it in my analysis of couples' payment behaviours in a data situation where the two heads of a household are not simultaneously observed but a household aggregate measurement $\overline{X}$ is available.[11]

It is interesting to note that, when applied to Model (7), Lemma 1 becomes a generalization of Kotlarski's Lemma. Under the assumptions of Model (7) and some regularity

[9] Evdokimov (2010) investigates the estimation of conditional characteristic functions; also see Zhang et al. (2011).

[10] In general, the variables can be transformed only to the extent that the linear combination identity between the transformed individual variables and the transformed aggregate is maintained.

[11] The modelling of two dependent variables using three independent variables is sometimes known as trivariate reduction. Sometimes referred to as the variables-in-common method, it is also a popular and old technique used for building dependent variables; see, e.g., Sarabia Alzaga and Gómez Déniz (2008) and Balakrishnan and Lai (2009).

conditions similar to the one I use, Kotlarski (1967) shows that the distributions of V_0, V_1 and V_2 are determined up to location by the joint distribution of (X_1,X_2); see also Rao (1992). This result has been used extensively in the nonparametric econometrics literature for identifying measurement error models, panel data models or auction models; see Evdokimov and White (2012) and references therein. In Lemma 1, I show that it is enough to know the marginal distributions of X_1, X_2 and $\overline{X} \equiv X_1 + X_2$ to obtain the same result.[12]

Example 2: Static panel data model

In this and the following example, aggregation is on the time dimension rather than on the unit dimension. I first consider the panel data model

$$Y_{ij} = U_i + \epsilon_{ij}, \quad i = 1, ..., N, \quad j = 1, ..., J, \tag{8}$$

where U_i is an unobserved, random, individual effect and ϵ_{ij} is an unobserved random variable i.i.d. across i and j, with U and ϵ mutually independent. Y_{ij} is the observed value of the dependent variable for individual i over the period j. Horowitz and Markatou (1996), Li and Vuong (1998), Hall and Yao (2003) and Neumann (2007), among others, consider the nonparametric estimation of the densities of the error components in Model (8) using panel data or, more generally, repeated measurements.

Lemma 1 and Theorem 1 apply directly to Model (8). This means that, in the absence of panel data, my approach can be applied to identify Model (8) by combining various cross-sectional distributions of the quantity Y measured over different periods of time. To see this, consider a specific example. Let Y measure consumption and let $j \equiv (d, q)$ denote quarter q of year d. Annual consumption is the aggregate of the quarterly measures: $\overline{Y}_{id} = \sum_{q=1}^{4} Y_{idq}$. Under the usual non-vanishing assumptions, Model (8) is identified by the four marginal distributions of quarterly consumption Y_{idq} as well as that of annual consumption $\overline{Y}_{id}$. I can thereby obtain the joint distribution of consumption over different quarters and study, for

[12] Note, however, that this result is likely to be of limited application in the context of measurement error models since it is not clear why we would observe the sum of two noisy measures in one data set.

example, persistence from one period to the next.

Example 3: Dynamic panel data model

I consider the earnings dynamics model used in Bonhomme and Robin (2010), which decomposes log earning residuals y_{ij} of individual i at time j into a fixed effect, a persistent autoregressive component and a transitory moving-average component, as follows:

$$y_{ij} = \alpha_i + y_{ij}^P + y_{ij}^T, \quad i = 1, ..., N, \quad j = 1, ..., J \tag{9a}$$

$$y_{ij}^P = y_{ij-1}^P + \epsilon_{ij}, \quad j \geq 2, \tag{9b}$$

$$y_{ij}^T = \nu_{ij}, \tag{9c}$$

where innovations ϵ_{ij} and ν_{ij} are mutually independent and independent over time. Once first-differenced, the model simplifies to

$$\Delta y_{ij} = \epsilon_{ij} + \Delta \nu_{ij}, \quad j \geq 2, \tag{10}$$

where $\Delta y_{ij} = y_{ij} - y_{ij-1}$ and $\Delta \nu_{ij} = \nu_{ij} - \nu_{ij-1}$.

Note now that the aggregation of l consecutive lag-1 differences of a data series results in a lag-l difference, i.e. $\sum_{j=2}^{l+1} \Delta y_{ij} = y_{il+1} - y_{i1}$. Therefore

$$y_{iJ} - y_{i1} = \sum\nolimits_{j=2}^{J} \Delta y_{ij}, \tag{11}$$

$$= \sum\nolimits_{j=2}^{J} \epsilon_{ij} + \nu_{iJ} - \nu_{i1}, \tag{12}$$

where the last expression follows from (10). Because of its aggregate nature, in what follows I denote the difference $y_{iJ} - y_{i1}$ by $\overline{\Delta} y_i$. By the convolution theorem, Equations (10) and (12) imply, for all $t \in \mathbb{R}$

$$\phi_{\Delta y_j}(t) = \phi_{\epsilon_j}(t) \phi_{\Delta \nu_j}(t), \quad \text{for } j = 2, ...J, \tag{13}$$

and

$$\phi_{\overline{\Delta}y}(t) = \prod_{j=2}^{J} \phi_{\epsilon_j}(t)\phi_{\nu_J}(t)\phi_{\nu_1}(-t), \tag{14}$$

where the unit subscript i has been omitted for clarity. To simplify these expressions I follow Horowitz and Markatou (1996) and further assume that the transitory shocks ν_{ij} are identically distributed (across both i and j). It follows that $\phi_{\nu_j}(t) = \phi_{\nu}(t)$ and $\phi_{\Delta\nu_j}(t) = \phi_{\Delta\nu}(t)$ for all $t \in \mathbb{R}$ and $j = 1, ...J$. I obtain

$$\phi_{\Delta y_j}(t) = \phi_{\epsilon_j}(t)\phi_{\Delta\nu}(t), \quad \text{for } j = 2, ...J, \tag{15}$$

and

$$\phi_{\overline{\Delta}y}(t) = \prod_{j=2}^{J} \phi_{\epsilon_j}(t)\phi_{\Delta\nu}(t), \tag{16}$$

These last equations are analogous to Equations (38) and (39) in the proof of Lemma 1. Hence, by applying the proof and under the usual non-vanishing assumptions, the distributions of Δy_{ij}, $j = 2, ..., J$, and $\overline{\Delta}y_i \equiv y_{iJ} - y_{i1}$ are shown to determine the distributions of ϵ_{ij} and $\Delta\nu_{ij}$, $j = 2, ..., J$, up to location. Finally, under the additional assumption that the transitory shocks ν_{ij} are symmetrically distributed around zero - also made in Horowitz and Markatou (1996)- the distribution of ν is also identified. Indeed, in that case $\phi_{\Delta\nu}(t) = [\phi_{\nu}(t)]^2$ for all $t \in \mathbb{R}$.

Using nine years of panel data, Bonhomme and Robin (2010) nonparametrically estimate the full distributions of the shocks in Model (10) without resorting to the assumption of identical and symmetric transitory shocks. Under slightly stricter assumptions, I obtain identification in the absence of a long panel data set by combining first-difference distributions obtained from different short panel data sets, when the distribution of $(y_J - y_1)$ is also available.[13] Such a setting arises, for example, in the context of a rotating panel that doesn't

[13]Note that my modelling assumptions remain more general than that of Horowitz and Markatou (1996) since there is no permanent shock in their model.

have a long panel dimension but provides several consecutive as well as non consecutive two-period panels that can be exploited.[14] Of course, identification possibilities depend on the data on hand, and the implementation of the proposed approach should be adapted to each specific application.

Assuming that the transitory shocks ν_{ij} are identically distributed across j is a way to deal with the multiple common factors in Model (10) when only one total aggregate measure is available. More general models with multiple common components can be identified when several aggregates, partial and total, are observed. This situation is considered next.

2.3 Extension with partial aggregates

In Model (1) a single common component is shared by all the X_k variables, and a single total aggregate brings the identification result. When partial aggregates are also observed, more flexible alternatives can be identified. Writing a general model in that case requires cumbersome notation. Instead, I illustrate the possibilities offered by the availability of partial aggregates using a specific example.

Consider the model:

$$X_1 = V_1 + V_{12} + V_{13} + V_0, \tag{17a}$$

$$X_2 = V_2 + V_{12} + V_{23} + V_0, \tag{17b}$$

$$X_3 = V_3 + V_{13} + V_{23} + V_0, \tag{17c}$$

where V_0, V_1, V_2, V_3, V_{12}, V_{13}, V_{23} are unobserved, mutually independent real random variables.

To identify the distributions of the seven V_k components in Example (17), it is not enough to know the marginal distributions of the three X_k variables and the total aggregate $\overline{X} \equiv X_1 + X_2 + X_3$. If, however, the distributions of the three partial sums $\overline{X}_{12} \equiv X_1 + X_2$,

[14] This is the case in the Canadian Financial Monitor, one of the data sets used in the application; see Chen et al. (2017) for a description of its panel structure.

$\overline{X}_{13} \equiv X_1 + X_3$ and $\overline{X}_{23} \equiv X_2 + X_3$ are also available, then it can be shown that the distributions of the V_k components are identified up to location.

Note for instance that the linearity and independence assumptions of Model (17) give, by the convolution theorem and after some manipulations:

$$\frac{\phi_{\overline{X}}(t)\phi_{X_1}(t)\phi_{X_2}(t)\phi_{X_3}(t)}{\phi_{X_{12}}(t)\phi_{X_{13}}(t)\phi_{X_{23}}(t)} = \frac{[\phi_{V_0}(t)]^3\,\phi_{V_0}(3t)}{[\phi_{V_0}(2t)]^3}. \tag{18}$$

This identifies $\phi_{V_0}(t)$ up to a multiplicative term of the form e^{bt}, where b is a real number. Similarly, the c.f.s of all the V_k factors are identified up to location. Finally, it follows that the joint distribution of (X_1, X_2, X_3) is just identified. The joint c.f., under Model (17) and by the convolution theorem, can indeed be written as

$$\phi_{X_1,X_2,X_3}(t_1,t_2,t_3) = \phi_{V_1}(t_1)\phi_{V_2}(t_2)\phi_{V_3}(t_3)\phi_{V_{12}}(t_1+t_2)\phi_{V_{13}}(t_1+t_3)\phi_{V_{23}}(t_2+t_3)\phi_{V_0}(t_1+t_2+t_3), \tag{19}$$

$$= \frac{\phi_{X_1,X_2}(t_1,t_2)\phi_{X_1,X_3}(t_1,t_3)\phi_{X_2,X_3}(t_2,t_3)}{\phi_{X_1}(t_1)\phi_{X_2}(t_2)\phi_{X_3}(t_3)}\,\frac{\phi_{V_0}(t_1)\phi_{V_0}(t_2)\phi_{V_0}(t_3)\phi_{V_0}(t_1+t_2+t_3)}{\phi_{V_0}(t_1+t_2)\phi_{V_0}(t_1+t_3)\phi_{V_0}(t_2+t_3)}, \tag{20}$$

where the bivariate joint c.f.s $\phi_{X_k,X_l}(t_k,t_l)$ are identified according to Theorem 1, since the marginals of X_k, X_l and $\overline{X}_{kl}$ are observed.[15]

3 Estimation

This section discusses nonparametric methods for estimating the joint density of interest. For the sake of simplicity and with no loss of generality, for the rest of the paper I concentrate on Model (7). I propose an estimation process in two steps : 1) Estimate ϕ_{V_0} and obtain $\widehat{\phi}_{X_1,X_2}$; 2) Derive $\widehat{f}_{X_1,X_2}$ from $\widehat{\phi}_{X_1,X_2}$.

[15] Note for example that the model that consists only of Equations (17a) and (17b) simplifies to $X_1 = W_1 + W_0$, $X_2 = W_2 + W_0$.

3.1 Estimating ϕ_{V_0}

Model (7) implies, for all $t \in \mathbb{R}$,

$$\frac{\phi_{\overline{X}}(t)}{\phi_{X_1}(t)\phi_{X_2}(t)} = \frac{\phi_{V_0}(2t)}{\left[\phi_{V_0}(t)\right]^2}. \tag{21}$$

The left-hand side of Equation (21) can be directly obtained from the data. The c.f. of X_k, $k = 1, 2$, is consistently estimated by the empirical characteristic function (e.c.f.)

$$\widehat{\phi}_{X_k}(t) = \frac{1}{n_k}\sum_{j=1}^{n_k} e^{itX_{kj}}, \tag{22}$$

with n_k being the size of the sample where X_k is observed. Similarly defined, the empirical counterpart of $\phi_{\overline{X}}$ is denoted by $\widehat{\phi}_{\overline{X}}$. A consistent estimate of the left-hand side ratio $\Phi(t)$ of Equation (21) is then $\widehat{\Phi}(t) = \widehat{\phi}_{\overline{X}}(t)/\left(\widehat{\phi}_{X_1}(t)\widehat{\phi}_{X_2}(t)\right)$.

ϕ_{V_0} is a complex-valued function with a unique representation $\phi_{V_0}(t) = |\phi_{V_0}(t)|e^{i\nu_0(t)}$. Following Horowitz and Markatou (1996), one can separately estimate its modulus and argument. When $\phi_{V_0}(t)$ is everywhere non-vanishing, its modulus is a non-zero real-valued function of t, and it follows from Equation (21) that

$$\ln|\Phi(t)| = \ln|\phi_{V_0}(2t)| - 2\ln|\phi_{V_0}(t)|. \tag{23}$$

As suggested in Linton and Whang (2002), a consistent estimator of $\ln|\phi_{V_0}(t)|$ is obtained by carrying out the nonparametric regression of $\ln\left|\widehat{\Phi}(t)\right|$ on t in a way that imposes the structure of the right-hand side of Equation (23). For example, a power series approximation can be used.

Finally, Equation (21) can be rewritten

$$\Phi(t) = |\Phi(t,t)|\, e^{i(\nu_0(2t)-2\nu_0(t))}, \tag{24}$$

which implies

$$\mathrm{Im}\left\{\ln \frac{\Phi(t)}{|\Phi(t)|}\right\} = \nu_0(2t) - 2\nu_0(t), \tag{25}$$

where Im denotes the imaginary part of a complex number. A consistent estimator of $\nu_0(t)$ can again be obtained by carrying out the nonparametric regression of $\mathrm{Im}\left\{\ln \frac{\hat{\Phi}(t,t)}{|\hat{\Phi}(t,t)|}\right\}$ on t in a way that imposes the structure of the right-hand side of Equation (25).

3.2 Estimating f_{X_1,X_2}

The joint density function f_{X_1,X_2} is recoverable from the joint c.f. of (X_1, X_2) by a two-dimensional Fourier inversion, as follows:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(t_1x_1+t_2x_2)}\phi_{X_1,X_2}(t_1,t_2)\,dt_1\,dt_2 \tag{26}$$

$$= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(t_1x_1+t_2x_2)}\phi_{X_1}(t_1)\phi_{X_2}(t_2)\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\,dt_1\,dt_2, \tag{27}$$

where Equation (27) is obtained by substituting Equation (6) into (26).

A naive approach to constructing an estimator for the joint density of (X_1, X_2) from Equation (27) would consist of replacing all the c.f.s with consistent empirical counterparts: the e.c.f.s of X_1 and X_2 and a consistent estimate of ϕ_{V_0}. However, such a plug-in estimator doesn't work in general because the resulting double integral can diverge. The mapping from f_{X_1,X_2} to ϕ_{X_1,X_2} is continuous, but the inverse mapping in Equation (26) is not because the integrand may be unbounded as $|t_1|$ or $|t_2|$ becomes large. As a consequence, small differences between the true c.f. ϕ_{X_1,X_2} and its estimate can induce large differences between the true density f_{X_1,X_2} and its estimate.

As summarized in Delaigle (2014b), there are two main problems with the empirical version of the integrand in Equation (27). First, the e.c.f. $\widehat{\phi}_{X_k}(t)$ is a poor estimator of the corresponding c.f. $\phi_{X_k}(t)$ for large $|t|$. Second, the e.c.f.s in the numerator are divided by $\phi_{V_0}(t_k)$, $k = 1, 2$, which tend in general to zero as $|t_k|$ goes to infinity. This second problem would exist if the denominator in (27) were known, but it is exacerbated when the denominator is estimated from the data. The naive empirical analog of the integrand in (27) would therefore be very unreliable at high frequencies t_1 and t_2, leading to wild fluctuations in the estimate of f_{X_1,X_2}.

Different regularization techniques have been proposed in the nonparametric deconvolution literature to solve this ill-posed problem.[16] Multiplying by a damping factor can temper the estimator of ϕ_{X_1,X_2} for large frequencies, where it is most unreliable. By far the most popular approach is the deconvoluting kernel density estimator introduced by Carroll and Hall (1988) and Stefanski and Carroll (1990). I use the deconvoluting kernel approach with a uniform kernel, which is equivalent to simply truncating the integrals on compact, increasing intervals.

Let $\widehat{\phi}_{V_0}(t)$ denote a consistent estimator of $\phi_{V_0}(t)$. I propose the following estimator for $f_{X_1,X_2}(x_1, x_2)$:

$$\widehat{f}_{X_1,X_2}(x_1, x_2) = \frac{1}{(2\pi)^2} \int_{-T_{1n}}^{T_{1n}} \int_{-T_{2n}}^{T_{2n}} e^{-i(t_1x_1+t_2x_2)} \widehat{\phi}_{X_1}(t_1)\widehat{\phi}_{X_2}(t_2) \frac{\widehat{\phi}_{V_0}(t_1 + t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)} \, dt_1 \, dt_2, \tag{28}$$

where T_{1n} and $T_{2n} \to \infty$ at rates to be specified shortly. In practice, the real part of the right-hand side function is taken to ensure that the estimated density resulting from the truncated integration is real.[17] Next I establish the rate of convergence of the nonparametric Fourier inversion estimator given in Equation (28).

3.3 Asymptotic properties

Under some regularity conditions, the procedure discussed in Section 3.1 for estimating ϕ_{V_0} can be expected to be consistent at the usual rate of convergence of nonparametric

[16] See Horowitz (2014) for general considerations on ill-posed inverse problems in economics.

[17] Further, because $\widehat{f}_{X_1,X_2}(x_1, x_2)$ could be negative, $\widetilde{f}_{X_1,X_2}(x_1, x_2) = \max\left\{\widehat{f}_{X_1,X_2}(x_1, x_2), 0\right\}$ can be taken and rescaled so that it integrates to 1; see Delaigle (2014a).

methods. The rate of convergence for the non-parametric density estimator in Equation (28) of $f_{X_1,X_2}(x_1,x_2)$ that follows is given in the next theorem.

Theorem 2. *Let $\phi_{X_1,X_2}(t_1,t_2)$ be absolutely integrable, and let $\phi_{V_0}(t) \neq 0$ for all t. Let $T_{1n} = o\left(\left(\frac{n}{\log n}\right)^{\gamma}\right)$ and $T_{2n} = o\left(\left(\frac{n}{\log n}\right)^{\gamma}\right)$ with $0 < \gamma < \frac{1}{2}$. Define $K_0(t) = \inf_{|s|\leq t} |\phi_{V_0}(s)|$ and let $\theta_n = K_0(T_{1n})$ and $\vartheta_n = K_0(T_{2n})$. Assume that there exists an estimator $\widehat{\phi}_{V_0}(t)$ such that*

$$\sup_{t\in\mathbb{R}} \left|\widehat{\phi}_{V_0}(t) - \phi_{V_0}(t)\right| = o\left(\beta_n\right) \quad a.s. \tag{29}$$

with $\beta_n = n^{-\lambda/2}$ for $0 < \lambda \leq 1$. Then,

$$\sup_{(x_1,x_2)\epsilon\chi_1\times\chi_2} \left|\widehat{f}_{X_1,X_2}(x_1,x_2) - f_{X_1,X_2}(x_1,x_2)\right| = o\left(\frac{\alpha_n T_{1n} T_{2n}}{\theta_n \vartheta_n}\right) + o\left(\frac{\beta_n T_{1n} T_{2n}}{\theta_n \vartheta_n}\right) \quad a.s. \tag{30}$$

with $\alpha_n = o(1)$, $\frac{(\log n/n)^{\frac{1}{2}-\gamma}}{\alpha_n} = O(1)$, $\frac{\beta_n}{\theta_n} = o(1)$ and $\frac{\beta_n}{\vartheta_n} = o(1)$.

4 Monte Carlo simulations

In this section, I study the finite-sample behaviour of the proposed density estimator. The sensitivity of the estimation procedure to the truncation parameters is also assessed.[18]

4.1 Design

The three series X_1, X_2 and $\overline{X}$ are generated from the independent V_k components under Model (7). Various distributions are considered: normal, Laplace, gamma, Poisson and geometric. I also run simulations where (X_1, X_2) are drawn directly from binormal distributions.

To measure the distance between the estimator and its target distribution function, I use the mean integrated squared error (MISE) defined as:

$$\text{MISE}\,\widehat{f}_{X_1,X_2} = \text{E}\int\int \left\{\widehat{f}_{X_1,X_2}(x_1,x_2) - f_{X_1,X_2}(x_1,x_2)\right\}^2 dx_1\, dx_2. \tag{31}$$

[18]The two-dimensional Fourier inversion necessary to recover the joint density function from the joint c.f. is highly computationally intensive. I make use of the fast Fourier transform (FFT) algorithm developed by Cooley and Tukey (1965). Details on the practical implementation of the estimation are available upon request.

For each distributional case, I compute the empirical truncated MISE of $\widehat{f}_{X_1,X_2}$ as well as its bias-variance decomposition, over 100 replications.[19] A truncated MISE is calculated on restricted ranges of X_1 and X_2, chosen so as to comprise 99.7 percent of the data (in the normal case, this corresponds to three standard deviations around the mean). In addition to the MISE I also estimate the median ISE (MedISE), as it is less sensitive to extreme values that can be encountered in practice. Four different sample sizes are examined: 100, 250, 500 and 1,000.

4.2 Simulation results

The estimator's performance is examined for a range of values of the truncation parameters T_1 and T_2. Some detailed simulation results are reported in Tables 1 to 3 in normal, gamma and Poisson cases. For these as well as additional distributional cases, I also provide graphs of MISE $\widehat{f}_{X_1,X_2}$ against the truncation parameters in Figures 1 to 6.

Density estimation works very well in the normal and Laplace cases for truncation parameters that are neither too small nor too large. Excessive truncation loses too much information, but insufficient truncation lets an inaccurate tail estimation of the joint c.f. disturb the final density estimate. MISE $\widehat{f}_{X_1,X_2}$ and the squared bias are both U-shaped, decreasing then increasing with the truncation parameters, while variance increases with the truncation parameters. Discrepancies between the mean and median ISE widen with the truncation parameters. This reflects the augmenting difficulty to accurately estimate ϕ_{V_0} on increasingly long supports. All MISE and MedISE measures reduce with the sample size. Another consequence of larger sample sizes is that the range of T_1, T_2 values for which MISE $\widehat{f}_{X_1,X_2}$ is relatively insensitive to truncation parameters gets wider, as clearly seen in Figures 1 to 6.

In the normal and Laplace cases, V_0 is symmetrically distributed about zero, and ϕ_{V_0} is real-valued. The estimation of a complex-valued ϕ_{V_0} function is more involved, so that

[19] The joint distributions of (X_1,X_2) are derived from the assumptions on the V_k components using a symbolic algebra package.

density estimation appears somewhat more difficult in the case of variables that are not symmetrically distributed, such as in the gamma and Poisson cases. Comparing across figures, one can see that better performance is obtained on less severely skewed data. Bonhomme and Robin (2010) also observe that, for non-Gaussian latent factor distributions, the deconvolution estimators have some difficulty capturing skewness and kurtosis. MISE $\widehat{f}_{X_1,X_2}$ is again U-shaped, decreasing then increasing with the truncation parameters.[20]

5 Application: analyzing joint payment habits by combining aggregated household data and individual data

In this section I apply the proposed methodology to analyze intra-household influences with respect to payment and cash-management practices. The recent applied literature on payment behaviours typically relies on either aggregated household data or individual-level data (where only one person is surveyed in a given household) and ignores dependencies within households. However, by combining these two types of data sets, I can recover and analyze joint payment behaviours in non-single households.

Fung et al. (2014) and Chen et al. (2017), using respectively individual- and household-level data, show the impact of recent payment innovations such as contactless credit cards and stored-value cards on the recourse to cash for retail payments. In this analysis, I further explore this issue while taking the intra-household dimension into account.

5.1 The data

The Bank of Canada monitors Canadians' payment behaviour via two surveys: the Methods-of-Payment (MOP) Survey and the Canadian Financial Monitor (CFM).[21]

[20]The observed sensitivity of the estimation procedure to the truncation calls for a data-driven selection procedure for T_1 and T_2. However, this is beyond the scope of this paper and left for future research.

[21]The CFM is a syndicated survey run by Ipsos Reid since 1999; questions on payment methods and cash management were introduced in collaboration with the Bank of Canada in 2009. The MOP survey was commissioned by the Bank of Canada and conducted in collaboration with Ipsos Reid in 2009 and 2013. Details on the 2009 and 2013 MOP Surveys are available in Arango and Welte (2012) and Henry et al. (2015).

Both surveys collect information on payment choices at the point of sale as well as cash-management habits in terms of cash withdrawals and cash holdings. One critical difference between them concerns the unit of observation. Table 4 summarizes, for couple households, the data observed in both surveys.[22] In the MOP, the unit of observation is the individual respondent. All the questions relate to the respondent's own individual characteristics and behaviours. In the CFM, the main unit of observation is the household. Demographic characteristics are observed for the female and male heads of the household, but cash and alternative methods of payment quantities are collected at the aggregated household level: the respondent is asked to report the monthly family total.[23] For the present analysis I exploit the 2013 MOP survey data and the 2013 CFM data. The two samples are considered independent of one another. Preliminary analysis confirms that they are representative of the same underlying population, based on some target variables. I focus thereafter on the restricted population of Canadian couple households.

In what follows, I concentrate on cash and three payment innovations that tend to compete with cash: single-purpose, prepaid stored-value cards issued by retailers (SVCs); multi-purpose stored-value cards issued by credit card companies such as Visa and MasterCard (SVCm); and contactless credit cards (CTC).[24] Payment behaviour is measured by the number of monthly purchases made using each of these methods; it is observed at the individual level in the MOP data and at the household aggregate level in the CFM data. Descriptive statistics are presented in Table 5.

Burdett et al. (2016) empirically document that, being more cash intensive than married people, singles suffer more from inflation tax; see also references therein. Preliminary anal-

[22]The CFM questionnaire is so designed that each household can have at most one male and one female head. I thus adopt the operational definition of a couple as a man and a woman living in married or cohabiting relationships.

[23]The CFM explicitly asks the "person most knowledgeable about [the] household finances" to complete the survey. Disaggregated intra-household information is also collected on various financial products, including credit cards and debit cards. However, the focus of the present analysis is on cash and recent payment innovations, which are only measured at the aggregate household level in the CFM.

[24]For details on these methods and how they are reshaping the Canadian payments landscape, see e.g., Arango et al. (2012) and Fung et al. (2015).

yses of the data reveal both inter- and intra-household heterogeneity in terms of individual payment behaviour.[25] There are differences between households of different types (single and couple households), and also between male and female heads within couple households. This motivates the use of a model that allows for intra-household heterogeneity, and a methodology that does not rely on single households to identify quantities for couple households.

5.2 Non-parametric estimation of joint payment behaviour

I now proceed to recovering partners' joint distributions by combining MOP individual-level data and CFM aggregate household data.[26] To model the behaviour of the male and female heads within couple households, I employ the simple common factor model

$$X^p_{F,i} = V^p_{F,i} + V^p_{H,i}, \tag{32a}$$

$$X^p_{M,i} = V^p_{M,i} + V^p_{H,i}, \tag{32b}$$

where i denotes couple households, p is one of the four payment methods considered, and the V components are mutually independent and independent across i. For each separate payment method, I estimate the joint distribution of the male and female quantities using the estimator in Equation (28). In lieu of a formal selection procedure, I obtain estimates for a set of trimming parameters and identify the pair of values (T^*_F, T^*_M) that maximizes the fit between the predicted joint distribution and the data with respect to the marginal moments. Table 6 presents the marginal moments as predicted and as observed in the data, for truncation values that provide the best marginal fit.

I also compute Pearson correlation and distance correlation coefficients from the estimated joint distribution for various trimming parameters. Contrary to the Pearson correlation coefficient that only measures linear dependence between two random variables, distance

[25] Both unconditional and conditional analyses are performed. Results are not reported but available upon request.

[26] Throughout this application, basic demographics are controlled for by using residuals instead of raw data.

correlation measures all types of dependence; see Székely et al. (2007) and Székely and Rizzo (2009). Because it is based on characteristic functions, this dependence measure is quite natural in the present framework. These bivariate statistics can vary a lot with large changes in the truncation parameters. However, they show little sensitivity to small variations in the neighbourhood of (T_F^*, T_M^*). Therefore, in the remainder of this application I use the latter values as trimming parameters.

In Table 7, I report the correlation estimates obtained with the selected trimming parameters. The linear correlation coefficients are small for all four methods of payment, and likely to be statistically insignificant in the case of cash and SVCs. However, the distance correlation statistics are quite a bit larger than the linear correlation estimates in the case of SVCs and SVCm. Although no formal test is performed here, the distance correlation measures seem to detect nonlinear dependence between X_F and X_M that Pearson correlations would leave unnoticed.

There is little evidence of dependence between male and female partners' cash usage as measured by their monthly frequency of cash purchases. However, this doesn't rule out intra-household influences that would operate across payment methods.

Fung et al. (2014) estimate the impact of an individual's use of contactless credit cards and stored-value cards on his or her own recourse to cash for retail payments. Working with household-level data, Chen et al. (2017) assess how the adoption of payment innovations by at least one person in the household affects that household's aggregate cash usage. However, these studies leave unexplored whether an individual's cash spending is influenced by their partner's use of payment innovations. The analysis in the next section aims at filling this gap.

5.3 Analysis: assessing the partner's influence on personal cash usage

Recall from Table 4 that cash-management practices such as cash withdrawal and cash holding habits are observed at the individual respondent's level in both the MOP and the CFM surveys.[27] I denote them by Y_r, $r = F, M$. The method developed in this paper permits me to estimate f_{X_F,X_M}, but it can't be used for recovering f_{Y_r,X_F,X_M}. However, I can test the influence of X_F on Y_M (and that of X_M on Y_F) without estimating f_{Y_r,X_F,X_M}. The null hypothesis can be written as follows:

$$H_0 : \{f_{Y_r|X_r,X_{\bar{r}}}(y_r|x_r, x_{\bar{r}}) = f_{Y_r|X_r}(y_r|x_r)\} \quad \forall y_r, x_r and x_{\bar{r}}, \tag{33}$$

where $r \in \{F, M\}$, and $\bar{r}$ is such that $r \cup \bar{r} = \{F, M\}$.

I now derive a test statistic for the null hypothesis spelled out in Equation (33). Note that

$$f_{Y_r,\overline{X}}(y_r, \overline{x}) = \int f_{Y_r,X_r,X_{\bar{r}}}(y_r, x_r, \overline{x} - x_r)\, dx_r \tag{34}$$

$$= \int f_{Y_r|X_r,X_{\bar{r}}}(y_r|x_r, \overline{x} - x_r) f_{X_r,X_{\bar{r}}}(x_r, \overline{x} - x_r)\, dx_r \tag{35}$$

Under the null hypothesis, the right-hand-side expression simplifies to give

$$f_{Y_r,\overline{X}}(y_r, \overline{x}) = \int f_{Y_r|X_r}(y_r|x_r) f_{X_r,X_{\bar{r}}}(x_r, \overline{x} - x_r)\, dx_r, \tag{36}$$

where all the quantities are identified from the data: $f_{Y_r,\overline{X}}$ and f_{Y_r,X_r} are observed in the CFM and MOP data, respectively, and $f_{X_r,X_{\bar{r}}}$ is obtained by combining the two data sets as per my methodology. Testing the null hypothesis in Equation (33) is equivalent to testing the equality of the right- and left-hand sides of Equation (36). Following Maasoumi and

[27] Cash management should reflect, at least to some extent, cash usage. For example, according to the Baumol (1952) and Tobin (1956) cash inventory model, people that use less cash for their purchases also withdraw less, in value and frequency, and hold less cash.

Racine (2002), I construct a metric entropy of the form

$$S_\rho = = \frac{1}{2}\int \left(f_{Y_r}^{1/2} - g_{Y_r}^{1/2}\right)^2 \, dy, \tag{37}$$

where $f_{Y_r} = f_{Y_r}(y|\overline{X} = \overline{x})$ and $g_{Y_r} = g_{Y_r}(y|\overline{X} = \overline{x})$ are the marginal densities of Y_r derived from the right- and left-hand sides of Equation (36), respectively, for specific values of $\overline{X}$.[28] I replace the unknown density functions with nonparametric estimates, and as suggested by Racine (2012) I use a bootstrap resampling method for obtaining the statistic's null distribution.

Table 8 presents the bootstrap P-values obtained for different combinations of cash-management measures (withdrawals from automated banking machines [ABMs] in volume and value, and cash holdings on hand) and methods of payment. The two sub-tables correspond to testing (a) the impact of X_F on Y_M when X_M and basic demographics are also controlled for, and (b) the impact of X_M on Y_F when X_F and basic demographics are also controlled for.

Overall, findings are similar for tests (a) and (b), so that partners' effects seem symmetrical across genders. My results show that, at the 10 percent level, one's cash-management practices are significantly impacted by one's partner's volume of purchases paid with cash and SVCs. However, no effect is found for CTC or SVCm.

An individual's personal cash-management practices in terms of withdrawals and holdings is expected to be directly determined by his or her recourse to cash for payments. My results further suggest that, for individuals living in couple relationships, it is also significantly influenced by the partner's cash spending. A similar partner effect is found for SVCs. This finding complements the converging evidence that SVCs usage leads to a reduction in cash use at the point of sale; see e.g., Fung et al. (2014) and Chen et al. (2017).[29]

[28] Tests are run for three different values of $\overline{X}$ corresponding to the first, second and third quartiles of the distribution. Another possible approach would be to test for equality of the bivariate distributions following Li et al. (2009).

[29] Fung et al. (2014) find a significant effect for all types of stored-value cards combined. While the data do not allow them to distinguish between SVCs and SVCm, the fact that half of the payments are below $5

These significant partner effects can be explained by the presence of fixed costs of withdrawals (time and effort as well as fees), which should cause couples to pool their withdrawal efforts and then split the withdrawn amount. It is also interesting to note that in the case of both cash and SVCs, the partner influences the value of cash withdrawn and held, but not the frequency of withdrawals. In other words, the adjustment of the personal cash withdrawal behaviour in response to the partner happens mainly on the intensive, not the extensive, margin.

Unlike SVCs, SVCm has no significant implications on the partner's use of cash. In the existing empirical literature there is also no clear evidence on the "self-impact" of SVCm, that is the effect of an individual's SVCm usage on his or her own cash spending.[30] The fact that single- and multiple-use stored-value cards show contrasting (intra-household) effects on the recourse to cash is revealing of their different nature. Single-purpose prepaid cards are more direct competitors of cash than multi-purpose ones. Henry et al. (2015) report that the median SVCs purchase value mirrors that of cash, while that of SVCm is closer to credit cards. For example, SVCm are typically used by Canadians who do not have access to credit cards to make online transactions; see, for example, Uribe (2009).

The absence of effect from CTC on the partner's cash usage contrasts with the negative and highly significant "self-impact" estimates obtained on cross-sections by Fung et al. (2014) and Chen et al. (2017). Using panel data, however, the latter further show that their cross-sectional results – obtained at the aggregate household level – are largely driven by household-specific unobserved heterogeneity. Even though it focuses on couples, my analysis suggests that this unobserved heterogeneity does not stem from variations in the intra-household CTC usage (i.e., how the total usage is shared across family members) that are overlooked in aggregate household data.

seems to indicate that most prepaid cards in their sample are SVCs.

[30]This is probably due to the fact that most payment surveys do not distinguish between the two types of cards; see also the preceding footnote.

6 Conclusion

In this paper, I consider situations where the distribution of aggregates of X_k, $k = 1, ..., K$, together with their marginals, provides identification of their joint distribution. I show that my approach can be used to identify linear factor models in data settings where the joint distribution of the observed measurements is not observed, but an aggregate measure is available. Aggregation can happen not only on the unit dimension but also on the time dimension. A two-step estimation procedure is proposed and shown to behave well in small-sample simulations.

There are several directions in which this work could be developed. First, the identification procedure I use is not constructive, insofar as I do not derive a closed-form expression for the characteristic function of the underlying factor V_0 in terms of observable quantities. A constructive identification proof could provide a consistent nonparametric estimator that does not rely on sieve approximation in the first-step estimation. With regard to estimation, performance gains can be expected from further enhancing or fine-tuning the procedures. Finally, development of a data-driven selection procedure for the truncation parameter is ongoing work.

I apply the developed methodology to analyze intra-household payment behaviours in the absence of intra-household data. This is achieved by combining individual-level data and household aggregated data. The estimated joint distribution functions of partners' payment behaviour within couple households show little evidence of (potentially nonlinear) intra-household interactions for the four methods of payment considered. However, cash and single-use stored-value cards exhibit significant partner effects on personal cash-management practices. In other words, for individuals living in couple relationships, the personal cash-management practices in terms of withdrawals and holdings are significantly influenced by the partner's cash and SVCs utilization. This finding implies that, for some methods of payment at least, ignoring the partner's impact might lead to spurious regression results due to an omitted variable bias. However, the data on hand do not permit assessment of the magnitude of this potential bias.

References

Aczél, J. and J. Dhombres (1989): *Functional equations in several variables*, vol. 31, Cambridge university press.

Arango, C., K. P. Huynh, B. Fung, and G. Stuber (2012): "The changing landscape for retail payments in Canada and the implications for the demand for cash," *Bank of Canada Review*, 2012, 31–40.

Arango, C. and A. Welte (2012): "The Bank of Canada's 2009 Methods-of-Payment Survey: Methodology and Key Results," Staff Discussion Paper No. 20126, Bank of Canada.

Balakrishnan, N. and C. D. Lai (2009): *Continuous bivariate distributions*, 2nd ed., Springer, New York, London.

Baumol, W. J. (1952): "The transactions demand for cash: an inventory theoretic approach," *The Quarterly Journal of Economics*, 545–556.

Bonhomme, S. and J.-M. Robin (2010): "Generalized non-parametric deconvolution with an application to earnings dynamics," *The Review of Economic Studies*, 77, 491–533.

Burdett, K., M. Dong, L. Sun, and R. Wright (2016): "Marriage, markets, and money: a Coasian theory of household formation," *International Economic Review*, 57, 337–368.

Carroll, R. J. and P. Hall (1988): "Optimal rates of convergence for deconvolving a density," *Journal of the American Statistical Association*, 83, 1184–1186.

Chen, H., M.-H. Felt, and K. P. Huynh (2017): "Retail payment innovations and cash usage: accounting for attrition by using refreshment samples," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180, 503–530.

Cooley, J. W. and J. W. Tukey (1965): "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, 19, 297–301.

Delaigle, A. (2014a): "Nonparametric kernel methods for curve estimation and measurement errors," *Proceedings of the International Astronomical Union*, 10, 28–39.

——— (2014b): "Nonparametric kernel methods with errors-in-variables: constructing estimators, computing them, and avoiding common mistakes," *Australian & New Zealand Journal of Statistics*, 56, 105–124.

D'Haultfoeuille, X. (2011): "On the completeness condition in nonparametric instrumental problems," *Econometric Theory*, 27, 460–471.

Evdokimov, K. (2010): "Identification and estimation of a nonparametric panel data model with unobserved heterogeneity," Working Paper.

Evdokimov, K. and H. White (2012): “Some extensions of a lemma of Kotlarski,” *Econometric Theory*, 28, 925–932.

Fung, B., K. P. Huynh, and G. Stuber (2015): “The use of cash in Canada,” *Bank of Canada Review*, 2015, 45–56.

Fung, B. S., K. P. Huynh, and L. Sabetti (2014): “The impact of retail payment innovations on cash usage,” *Journal of Financial Market Infrastructures*, 3, 3–31.

Hall, P. and Q. Yao (2003): “Inference in components of variance models with low replication,” *The Annals of Statistics*, 31, 414–441.

Henry, C., K. Huynh, and Q. R. Shen (2015): “2013 Methods-of-Payment Survey Results,” Staff Discussion Paper No. 20154., Bank of Canada.

Horowitz, J. L. (2014): “Ill-posed inverse problems in economics,” *Annual Review of Economics*, 6, 21–51.

Horowitz, J. L. and M. Markatou (1996): “Semiparametric estimation of regression models for panel data,” *The Review of Economic Studies*, 63, 145–168.

Kotlarski, I. (1967): “On characterizing the gamma and the normal distribution,” *Pacific Journal of Mathematics*, 20, 69–76.

Li, Q., E. Maasoumi, and J. S. Racine (2009): “A nonparametric test for equality of distributions with mixed categorical and continuous data,” *Journal of Econometrics*, 148, 186–200.

Li, T. and Q. Vuong (1998): “Nonparametric estimation of the measurement error model using multiple indicators,” *Journal of Multivariate Analysis*, 65, 139–165.

Linton, O. and Y.-J. Whang (2002): “Nonparametric estimation with aggregated data,” *Econometric Theory*, 18, 420–468.

Lukacs, E. (1970): *Characteristics functions*, Griffin, London.

Maasoumi, E. and J. Racine (2002): “Entropy and predictability of stock market returns,” *Journal of Econometrics*, 107, 291–312.

Meister, A. (2009): “Deconvolution problems in nonparametric statistics,” in *Lecture Notes in Statistics 193*, Springer, Berlin Heidelberg.

Neumann, M. H. (2007): “Deconvolution from panel data with unknown error distribution,” *Journal of Multivariate Analysis*, 98, 1955–1968.

Piterbarg, V. (2011): “Spread options, Farkass lemma and linear programming,” *Risk*, 24, 61.

Racine, J. S. (2012): “Entropy-based inference using R and the np package: a primer,” *R package vignette, version 0.40-13, http://CRAN. R-project. org/package= np.*

Rao, B. P. (1992): *Identifiability in stochastic models*, Academic Press.

Ridder, G. and Y. Hu (2012): "Estimation of Nonlinear Models with Mismeasured Regressors Using Marginal Information," *Journal of Applied Econometrics*, 27, 347–385.

Ridder, G. and R. Moffitt (2007): "The econometrics of data combination," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, Part B, 5469–5547.

Robinson, W. (1950): "Ecological correlations and the behavior of individuals," *American Sociological Review*, 15.

Sarabia Alzaga, J. M. and E. Gómez Déniz (2008): "Construction of multivariate distributions: a review of some recent results," *Statistics and Operations Research Transactions*, 32, 3–36.

Stefanski, L. A. and R. J. Carroll (1990): "Deconvolving kernel density estimators," *Statistics*, 21, 169–184.

Székely, G. and C. Rao (2000): "Identifiability of distributions of independent random variables by linear combinations and moments," *Sankhyā: The Indian Journal of Statistics, Series A*, 193–202.

Székely, G. J. and M. L. Rizzo (2009): "Brownian distance covariance," *The Annals of Applied Statistics*, 3, 1236–1265.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007): "Measuring and testing dependence by correlation of distances," *The Annals of Statistics*, 35, 2769–2794.

Tobin, J. (1956): "The interest-elasticity of transactions demand for cash," *The review of Economics and Statistics*, 241–247.

Uribe, E. (2009): "Credit cards and access to the digital marketplace: a priceless necessity?" Tech. rep., Public Interest Advocacy Centre.

Ushakov, N. G. (1999): *Selected topics in characteristic functions*, VSP, Utrecht, The Netherlands.

Wagner, C. (2009): "Deconvolution problems in density estimation," Ph.D. thesis, Universität Ulm.

Zhang, Y., Q. Li, and D. Li (2011): "Recent developments in semiparametric and nonparametric estimation of panel data models with incomplete information: a selected review," in *Missing Data Methods: Cross-sectional Methods and Applications*, Emerald Group Publishing Limited, 41–62.

TABLE 1: Monte Carlo simulations - V_1, $V_2 \sim N(0,1)$, $V_0 \sim N(0,0.5)$

T_1, T_2	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8
n=100												
MISE $\widehat{\boldsymbol{f}}_{\boldsymbol{X_1,X_2}}$	**0.0402**	**0.0312**	**0.0227**	**0.0153**	**0.0058**	**0.0037**	**0.0032**	**0.0041**	**0.0072**	**0.0128**	**0.0185**	**0.0305**
MedISE $\widehat{f}_{X_1,X_2}$	0.0402	0.0312	0.0226	0.0152	0.0056	0.0034	0.0024	0.0023	0.0027	0.0033	0.0043	0.0173
Bias squared	0.0402	0.0312	0.0226	0.0151	0.0053	0.0029	0.0018	0.0016	0.0026	0.0051	0.0087	0.0192
Variance	0.0000	0.0000	0.0001	0.0002	0.0005	0.0008	0.0014	0.0025	0.0047	0.0077	0.0099	0.0115
n=500												
MISE $\widehat{\boldsymbol{f}}_{\boldsymbol{X_1,X_2}}$	**0.0401**	**0.0311**	**0.0225**	**0.0149**	**0.0050**	**0.0025**	**0.0013**	**0.0009**	**0.0009**	**0.0032**	**0.0056**	**0.0147**
MedISE $\widehat{f}_{X_1,X_2}$	0.0401	0.0311	0.0225	0.0149	0.0050	0.0025	0.0012	0.0008	0.0007	0.0007	0.0010	0.0019
Bias squared	0.0401	0.0311	0.0224	0.0149	0.0049	0.0023	0.0010	0.0005	0.0003	0.0007	0.0014	0.0059
Variance	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003	0.0004	0.0005	0.0025	0.0043	0.0089
n=1,000												
MISE $\widehat{\boldsymbol{f}}_{\boldsymbol{X_1,X_2}}$	**0.0401**	**0.0311**	**0.0224**	**0.0149**	**0.0049**	**0.0024**	**0.0011**	**0.0006**	**0.0005**	**0.0005**	**0.0011**	**0.0095**
MedISE $\widehat{f}_{X_1,X_2}$	0.0401	0.0311	0.0224	0.0149	0.0049	0.0024	0.0011	0.0006	0.0004	0.0004	0.0005	0.0007
Bias squared	0.0401	0.0311	0.0224	0.0149	0.0048	0.0023	0.0010	0.0004	0.0002	0.0002	0.0003	0.0026
Variance	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0003	0.0003	0.0009	0.0070

Notes: In $N(.,.)$, the first and second parameters indicate the mean and standard deviation, respectively. n indicates sample size. Results are based on 100 replications. *Bias squared* and *Variance* are obtained by decomposing the MISE into the integrated squared bias and integrated variance.

TABLE 2: Monte Carlo simulations - $V_k \sim \text{Gamma}(2,1)$ for $k = 0, 1, 2$

T_1, T_2	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3
n=100												
MISE $\widehat{f}_{X_1,X_2}$	**0.0086**	**0.0062**	**0.0047**	**0.0062**	**0.0054**	**0.0062**	**0.0059**	**0.0061**	**0.0071**	**0.0079**	**0.0090**	**0.0095**
MedISE $\widehat{f}_{X_1,X_2}$	0.0084	0.0056	0.0042	0.0046	0.0051	0.0064	0.0059	0.0069	0.0071	0.0077	0.0091	0.0094
Bias squared	0.0084	0.0057	0.0039	0.0039	0.0032	0.0041	0.0041	0.0043	0.0050	0.0056	0.0062	0.0065
Variance	0.0000	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0003	0.0003
n=500												
MISE $\widehat{f}_{X_1,X_2}$	**0.0081**	**0.0052**	**0.0040**	**0.0027**	**0.0023**	**0.0039**	**0.0039**	**0.0030**	**0.0031**	**0.0040**	**0.0050**	**0.0057**
MedISE $\widehat{f}_{X_1,X_2}$	0.0081	0.0051	0.0038	0.0029	0.0025	0.0030	0.0027	0.0032	0.0032	0.0042	0.0050	0.0058
Bias squared	0.0081	0.0051	0.0037	0.0020	0.0015	0.0022	0.0023	0.0018	0.0023	0.0032	0.0042	0.0048
Variance	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
n=1,000												
MISE $\widehat{f}_{X_1,X_2}$	**0.0081**	**0.0050**	**0.0034**	**0.0023**	**0.0023**	**0.0021**	**0.0018**	**0.0017**	**0.0022**	**0.0026**	**0.0031**	**0.0036**
MedISE $\widehat{f}_{X_1,X_2}$	0.0081	0.0050	0.0034	0.0024	0.0019	0.0022	0.0016	0.0016	0.0021	0.0031	0.0031	0.0035
Bias squared	0.0081	0.0050	0.0032	0.0020	0.0019	0.0016	0.0011	0.0012	0.0016	0.0021	0.0027	0.0031
Variance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	0.0001

Notes: In Gamma(. , .) the first and second parameters indicate the shape and rate, respectively. n indicates sample size. Results are based on 10 replications. *Bias squared* and *Variance* are obtained by decomposing the MISE into the integrated squared bias and integrated variance.

TABLE 3: Monte Carlo simulations - V_1, $V_2 \sim$ Poisson(5), $V_0 \sim$ Poisson(1)

T_1, T_2	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8
n=100												
MISE $\widehat{\boldsymbol{f}}_{\boldsymbol{X_1,X_2}}$	**0.0083**	**0.0080**	**0.0080**	**0.0081**	**0.0083**	**0.0085**	**0.0087**	**0.0090**	**0.0091**	**0.0091**	**0.0090**	**0.0090**
MedISE $\widehat{f}_{X_1,X_2}$	0.0082	0.0080	0.0080	0.0081	0.0082	0.0085	0.0087	0.0087	0.0088	0.0088	0.0087	0.0087
Bias squared	0.0082	0.0080	0.0080	0.0080	0.0081	0.0082	0.0083	0.0086	0.0086	0.0085	0.0084	0.0084
Variance	0.0000	0.0000	0.0001	0.0001	0.0002	0.0003	0.0003	0.0004	0.0005	0.0006	0.0006	0.0006
n=500												
MISE $\widehat{\boldsymbol{f}}_{\boldsymbol{X_1,X_2}}$	**0.0082**	**0.0079**	**0.0079**	**0.0079**	**0.0079**	**0.0079**	**0.0082**	**0.0084**	**0.0087**	**0.0089**	**0.0091**	**0.0092**
MedISE $\widehat{f}_{X_1,X_2}$	0.0082	0.0079	0.0079	0.0079	0.0078	0.0079	0.0081	0.0084	0.0085	0.0086	0.0086	0.0085
Bias squared	0.0082	0.0079	0.0079	0.0079	0.0079	0.0079	0.0081	0.0082	0.0085	0.0086	0.0088	0.0089
Variance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0003	0.0003	0.0003
n=1,000												
MISE $\widehat{\boldsymbol{f}}_{\boldsymbol{X_1,X_2}}$	**0.0082**	**0.0080**	**0.0079**	**0.0079**	**0.0079**	**0.0079**	**0.0080**	**0.0082**	**0.0086**	**0.0088**	**0.0090**	**0.0094**
MedISE $\widehat{f}_{X_1,X_2}$	0.0082	0.0080	0.0079	0.0079	0.0079	0.0079	0.0081	0.0082	0.0084	0.0085	0.0085	0.0085
Bias squared	0.0082	0.0079	0.0079	0.0079	0.0079	0.0079	0.0080	0.0082	0.0084	0.0086	0.0088	0.0092
Variance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003

Notes: n indicates sample size. Results are based on 100 replications. *Bias squared* and *Variance* are obtained by decomposing the MISE into the integrated squared bias and integrated variance.

TABLE 4: Units of observation in the MOP and CFM surveys

	Unit of observation	Variables observed for couple households: Demographics	Cash and payment innovations	Cash management
MOP	Individual	$Z_{i,r}$	$X_{i,r}$	$Y_{i,r}$
CFM	Household	$Z_{j,F}$ and $Z_{j,M}$	$\overline{X}_j = X_{j,F} + X_{j,M}$	$Y_{j,r}$

Notes: i and j denote households in the MOP and CFM samples, respectively. F and M indicate the female and male head of the household, and $r \in \{F, M\}$ indicates the survey respondent.

Table 5: Summary statistics - individual and aggregate variables

Variable	Source	Obs.	Mean	Var.	Max.
Cash					
FH	MOP	918	2.8	10.0	25
MH	MOP	625	2.8	8.8	19
HH total	CFM	5,364	3.7	14.1	30
SVCs					
FH	MOP	756	0.9	7.8	24
MH	MOP	521	0.7	6.0	20
HH total	CFM	5,618	1.1	9.7	40
SVCm					
FH	MOP	779	0.1	0.2	5
MH	MOP	544	0.1	0.5	7
HH total	CFM	5,634	0.3	2.7	30
CTC					
FH	MOP	903	1.6	17.5	30
MH	MOP	617	2.5	34.7	40
HH total	CFM	5,634	1.3	15.2	40

Notes: Summary statistics are obtained from the 2013 MOP and 2013 CFM subsamples of couple households. Variables measure the number of purchases made using each method of payment in a month. Individual variables from the MOP have been winsorized at the 99.5th percentile; aggregate variables from the CFM have been winsorized at the 99.9th percentile. *HH* stands for household; *FH* and *MH* stand for female and male head of household.

Table 6: Fit of the model, univariate moments

	Variance		Skewness		Kurtosis	
	Data	Predicted	Data	Predicted	Data	Predicted
Cash						
FH	10.47	10.50	2.69	2.71	16.35	16.44
MH	8.55	8.46	2.59	2.58	11.64	11.61
SVCs						
FH	7.69	7.88	5.10	5.11	34.73	34.66
MH	5.88	5.84	4.56	4.58	29.08	29.29
SVCm						
FH	0.75	0.75	0.25	0.31	8.87	9.08
MH	0.43	0.88	6.65	5.90	58.30	40.15
CTC						
FH	17.23	17.21	3.75	3.76	21.18	21.22
MH	33.80	34.97	3.56	3.58	19.24	19.11

Notes: Data moments are obtained from the 2013 MOP subsamples of female heads (FH) in MOP couple households and male heads (MH) in MOP couple households. Predicted moments are obtained from the estimated joint densities by numerical integration.

Table 7: Correlation estimates

	Correlation	Distance Cor.
CASH	-0.02	0.04
SVCs	0.00	0.18
SVCm	0.11	0.39
CTC	0.08	0.07

Notes: Pearson's correlation and distance correlation statistics are obtained using the estimated joint densities by numerical integration.

Table 8: Analysis

(a) Testing $f_{Y_M|X_F,X_M} = f_{Y_M|X_M}$

	ABM withdrawals, volume			ABM withdrawals, value			Cash on hand		
Value of $\overline{X}$	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
CASH	0.400	0.900	0.900	0.217	0.717	0.867	0.033	0.167	0.050
SVCs	0.617	0.700	0.017	0.117	0.100	0.033	0.317	0.350	0.033
SVCm	0.433	0.467	0.650	0.467	0.467	0.667	0.467	0.467	0.733
CTC	0.383	0.517	0.300	0.250	0.283	0.150	0.333	0.350	0.150

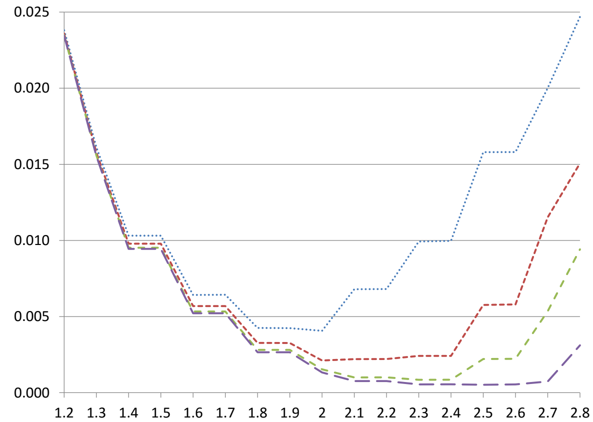
(b) Testing $f_{Y_F|X_F,X_M} = f_{Y_F|X_F}$

	ABM withdrawals, volume			ABM withdrawals, value			Cash on hand		
Value of $\overline{X}$	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
CASH	0.933	0.683	0.767	0.017	0.600	0.617	0.050	0.167	0.100
SVCs	1.000	0.983	0.050	0.633	0.683	0.050	0.133	0.100	0.000
SVCm	0.717	0.617	0.467	0.483	0.567	0.450	0.667	0.450	0.383
CTC	0.767	0.750	0.300	0.250	0.517	0.133	0.417	0.617	0.250

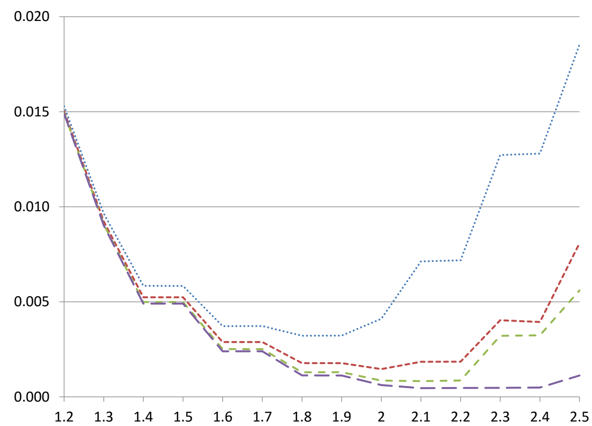
Notes: Q_1, Q_2 and Q_3 correspond to the first, second and third quartiles of the distribution of $\overline{X}$. A log transformation is applied to the raw cash-management data, which are then orthogonalized to basic demographic variables.

Figure 1: MISE estimates - normal distribution

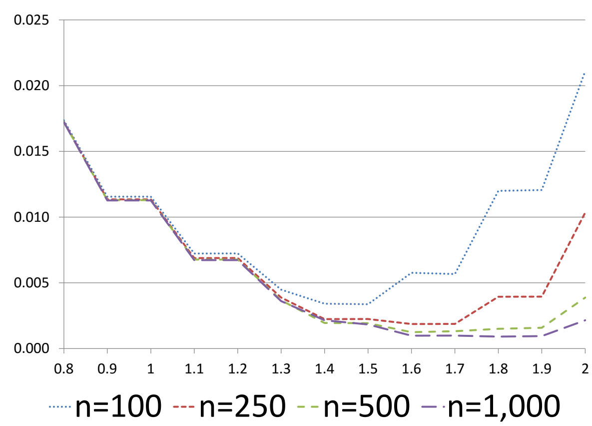
(a) $V_0 \sim N(0, 0.1)$



(b) $V_0 \sim N(0, 0.5)$



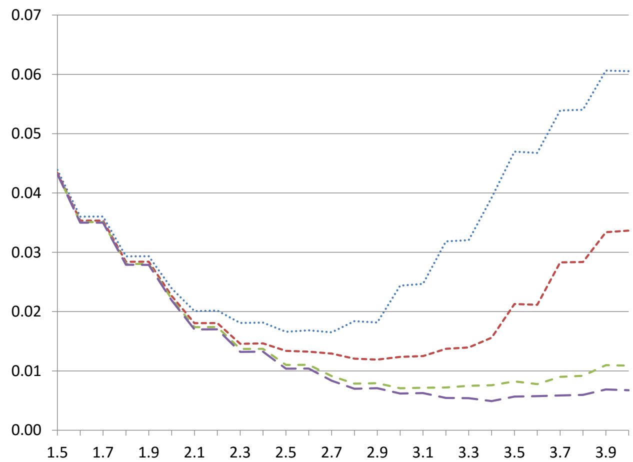
(c) $V_0 \sim N(0, 1)$



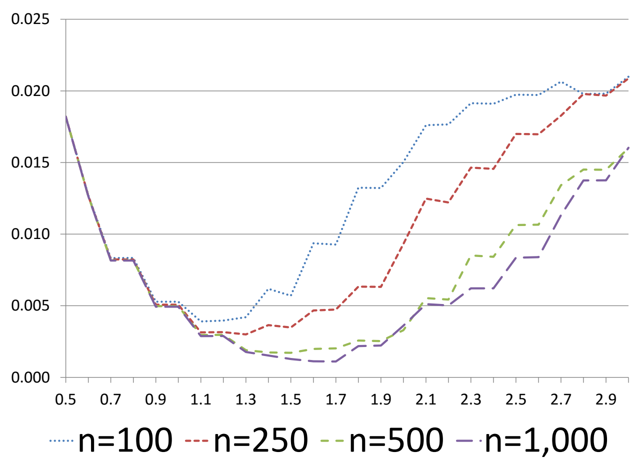
Notes: V_1 and $V_2 \sim N(0, 1)$ in all three cases. In $N(.,.)$, the first and second parameters indicate the mean and standard deviation, respectively. n indicates sample size. $T_1 = T_2$ is in the abscissa. Results are based on 100 replications.

Figure 2: MISE estimates - Laplace distribution

(a) $V_k \sim \text{Laplace}(0, 0.5)$ for $k = 0, 1, 2$



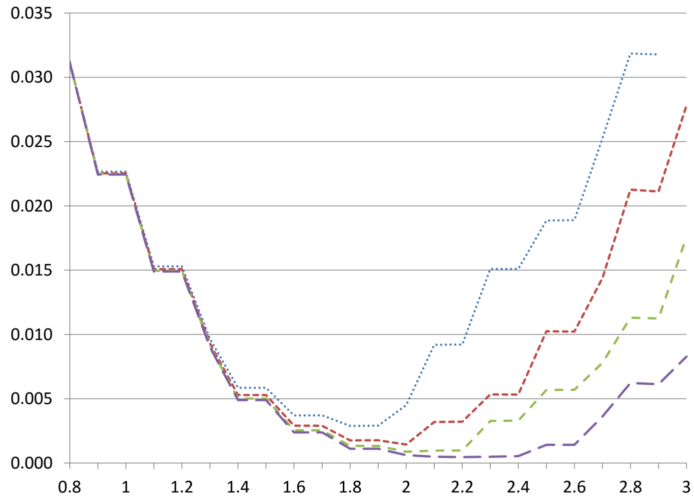
(b) $V_k \sim \text{Laplace}(0, 1)$ for $k = 0, 1, 2$



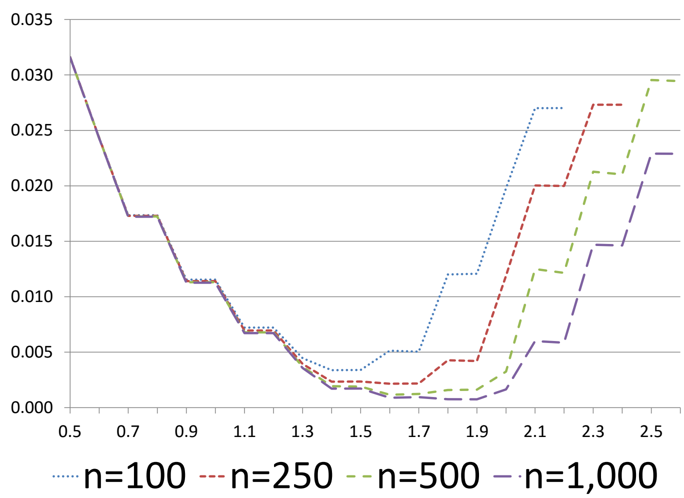
Notes: In Laplace(., .), the first and second parameters indicate the location and scale, respectively. n indicates sample size. $T_1 = T_2$ is in the abscissa. Results are based on 100 replications.

Figure 3: MISE estimates - bivariate normal distribution

(a) $(X_1, X_2) \sim \mathrm{N}(\mu_1 = \mu_2 = 0, \sigma_1^2 = \sigma_2^2 = 1.25, \rho = 0.2)$



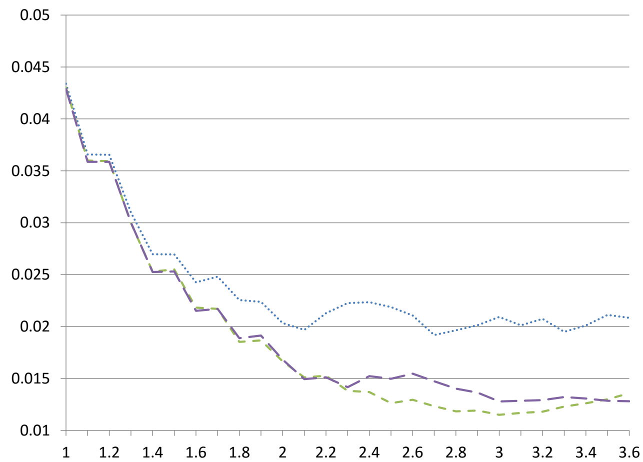
(b) $(X_1, X_2) \sim \mathrm{N}(\mu_1 = \mu_2 = 0, \sigma_1^2 = \sigma_2^2 = 2, \rho = 0.5)$



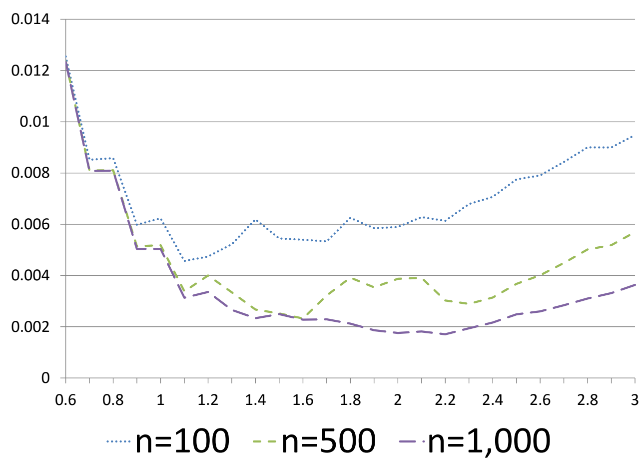
Notes: X_1 and X_2 are simulated directly from bivariate normal distributions. n indicates sample size. $T_1 = T_2$ is in the abscissa. Results are based on 100 replications.

Figure 4: MISE estimates - gamma distribution

(a) $V_k \sim \text{Gamma}(1,1)$ for $k = 0,1,2$



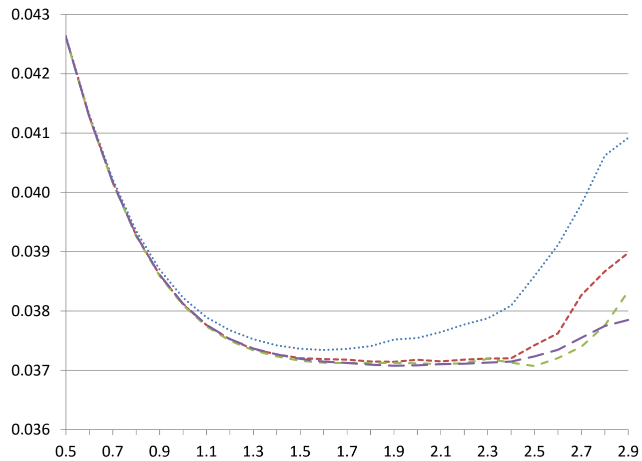
(b) $V_k \sim \text{Gamma}(2,1)$ for $k = 0,1,2$



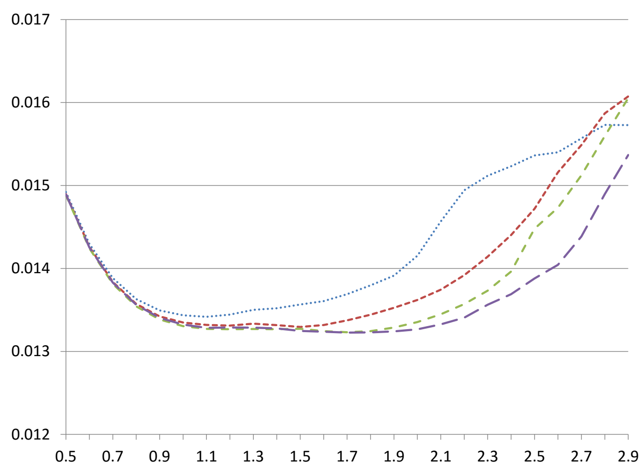
Notes: In Gamma(. , .), the first and second parameters indicate the shape and rate, respectively. n indicates sample size. $T_1 = T_2$ is in the abscissa. Results are based on 10 replications.

Figure 5: MISE estimates - Poisson distribution

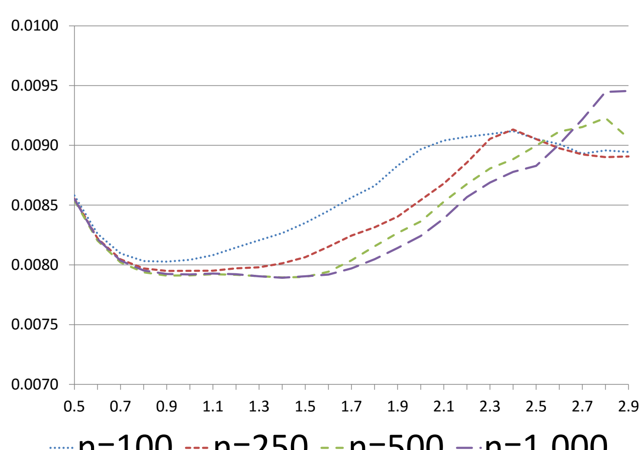
(a) $V_k \sim$ Poisson(1) for $k = 0, 1, 2$



(b) V_1 and $V_2 \sim$ Poisson(3), $V_0 \sim$ Poisson(1)



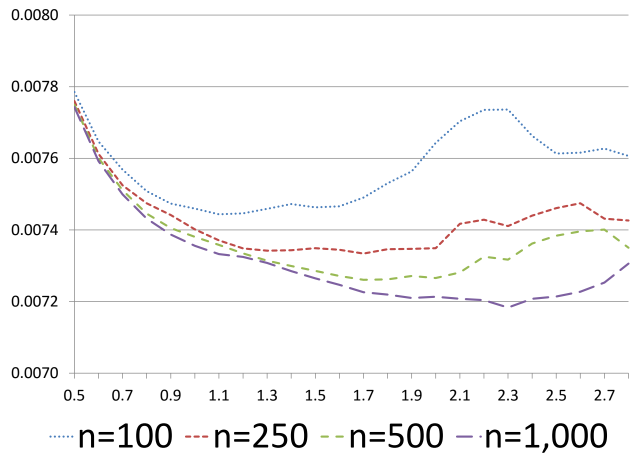
(c) V_1 and $V_2 \sim$ Poisson(5), $V_0 \sim$ Poisson(1)



Notes: n indicates sample size. $T_1 = T_2$ is in the abscissa. Results are based on 100 replications.

Figure 6: MISE estimates - geometric distribution

(a) Geometric: $V_k \sim \mathrm{G}(0.3)$ for $k = 0, 1, 2$



Notes: G denotes the unshifted geometric distribution (with support starting at zero). n indicates sample size. $T_1 = T_2$ is in the abscissa. Results are based on 100 replications.

A Proofs of Lemma 1 and Theorem 1

Proof of Lemma 1:

This proof follows closely that of Kotlarski in his seminal 1967 paper, also reported in Rao (1992).

By the convolution theorem, Model (1) implies

$$\phi_{X_k}(t) = \phi_{V_k}(t)\phi_{V_0}\left(\rho_k t\right), \quad \text{for } k = 1, ..., K, \text{and} \tag{38}$$

$$\phi_{\overline{X}}(t) = \prod_{k=1}^{K} \phi_{V_k}(t)\phi_{V_0}\left(t\sum\nolimits_{k=1}^{K}\rho_k\right) \quad \text{for all } t \in \mathbb{R}. \tag{39}$$

Let $U_0, U_1, ..., U_K$ be another set of $K+1$ mutually independent, real random variables with non-vanishing c.f.s $\phi_{U_k}(t_k)$, $k = 0, 1, ..., K$. Also, let $Z_k = U_k + \rho_k U_0$ for $k = 1, ..., K$ and $\overline{Z} = \sum_{k=1}^{K} Z_k$. Finally, define

$$\gamma_k(t_k) = \phi_{U_k}(t_k)/\phi_{V_k}(t_k), \quad \text{for } k = 0, 1, ..., K. \tag{40}$$

Let X_k and Z_k, $k = 1, ..., K$, as well as $\overline{X}$ and $\overline{Z}$ have (pairwise) the same distributions, so that their c.f.s are equal. It follows from Equations (38) and (39) that

$$\phi_{V_k}(t)\phi_{V_0}\left(\rho_k t\right) = \phi_{U_k}(t)\phi_{U_0}\left(\rho_k t\right), \quad \text{for } k = 1, ..., K, \tag{41}$$

and

$$\prod_{k=1}^{K} \phi_{V_k}(t)\phi_{V_0}\left(t\sum\nolimits_{k=1}^{K}\rho_k\right) = \prod_{k=1}^{K} \phi_{U_k}(t)\phi_{U_0}\left(t\sum\nolimits_{k=1}^{K}\rho_k\right), \tag{42}$$

hold for all $t \in \mathbb{R}$.

Some manipulations give

$$\gamma_k(t)\gamma_0(\rho_k t) = 1, \quad \text{for } k = 1, ..., K, \tag{43}$$

and

$$\prod_{k=1}^{K} \gamma_k(t)\gamma_0(t \sum\nolimits_{k=1}^{K} \rho_k) = 1, \tag{44}$$

for all $t \in \mathbb{R}$.

By substitution, it follows that

$$\gamma_0(t \sum\nolimits_{k=1}^{K} \rho_k) = \prod_{k=1}^{K} \gamma_0(\rho_k t), \tag{45}$$

for all $t \in \mathbb{R}$. Also note that given how they have been defined, the $\gamma_k(t)$ are continuous complex-valued functions with $\gamma_k(0) = 1$, for $k = 1, ..., K$.

Equation (45) is Cauchy's exponential equation, and has general solutions $g : \mathbb{R} \to \mathbb{C}$, in the class of continuous functions, of the form $g(t) = e^{at}$, where a is an arbitrary complex constant; see Aczél and Dhombres (1989, chapter 5).

Since $\gamma_0(-t)$ is the complex conjugate of $\gamma_0(t)$ (from the properties of c.f.s), then $e^{-at} = \overline{e^{at}}$. It follows that a is purely imaginary and can be written $a = ibt$, where b is a real number. Therefore,

$$\gamma_0(t) = e^{ibt}, \quad \text{for all } t \in \mathbb{R}, \tag{46}$$

and Equation (43) further implies

$$\gamma_k(t) = e^{-ib\rho_k t}, \quad \text{for } k = 1, ..., K \text{ and for all } t \in \mathbb{R}. \tag{47}$$

This means that V_k and U_k have the same distributions up to a location parameter, for $k = 0, 1, ..., K$. Hence I have proved that the distributions of X_1, ... X_K and $\overline{X}$ determine that of V_0, V_1,..., V_K up to location.

Proof of Theorem 1:

Based on the expression for the joint c.f. of $(X_1,...,X_K)$ derived in Equation (6), I can write

$$\phi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^{K} \phi_{X_k}(t_k)\Phi(\mathbf{t}), \mathbf{t} \in \mathbb{R}^K \tag{48}$$

where $\Phi(\mathbf{t}) = \phi_{V_0}(\mathbf{t}'\boldsymbol{\rho})/\prod_{k=1}^{K}\phi_{V_0}(\rho_k t_k)$. Using the same set-up as in the previous proof, it is easy to show that

$$\Phi(\mathbf{t}) = \frac{\phi_{V_0}(\mathbf{t}'\boldsymbol{\rho})}{\prod_{k=1}^{K}\phi_{V_0}(\rho_k t_k)} = \frac{\phi_{U_0}(\mathbf{t}'\boldsymbol{\rho})}{\prod_{k=1}^{K}\phi_{U_0}(\rho_k t_k)} \quad \text{for all } \mathbf{t} \in \mathbb{R}^K \tag{49}$$

This is because the γ_0 functions in the numerator and denominator of $\Phi(\mathbf{t})$ cancel out. This proves that the ratio $\Phi(\mathbf{t})$ is uniquely determined by the distributions of X_1, ... X_K and $\overline{X}$. Given Equation (48), the c.f. $\phi_{\mathbf{X}}(\mathbf{t})$ is thus uniquely identified.

B Proof of Theorem 2

To prove the consistency of my estimator I use a result from Ridder and Hu (2012, Lemma 3, p. 370) that gives an almost sure rate of convergence for the e.c.f. without any restriction on the support of the distribution.

Lemma 2 (Ridder and Hu, 2012)**.** *Let* $\widehat{\phi}(t) = \int_{-\infty}^{\infty} e^{itx}\, dF_n(x)$ *be the e.c.f. of a random sample from a distribution with c.d.f. F and with $E(|x|) < \infty$. For $0 < \gamma < \frac{1}{2}$, let $T_n = o\left(\left(\frac{n}{\log n}\right)^{\gamma}\right)$. Then*

$$\sup_{|t|\le T_n} \left|\widehat{\phi}(t) - \phi(t)\right| = o\left(\alpha_n\right) \quad a.s. \tag{50}$$

with $\alpha_n = o(1)$ and $\frac{(\log n/n)^{\frac{1}{2}-\gamma}}{\alpha_n} = O(1)$.

Proof. See Ridder and Hu (2012). ☐

I now proceed to deriving the convergence rate of $\widehat{f}_{X_1,X_2}$. We have

$$
\begin{aligned}
&\sup\left|\widehat{f}_{X_1,X_2}(x_1,x_2)-f_{X_1,X_2}(x_1,x_2)\right| \\
&\leq \sup\left|\frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}} e^{-i(t_1x_1+t_2x_2)}\left(\widehat{\phi}_{X_1}(t_1)\widehat{\phi}_{X_2}(t_2)\frac{\widehat{\phi}_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}\right.\right. \\
&\left.\left.-\ \phi_{X_1}(t_1)\phi_{X_2}(t_2)\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right)dt_1\,dt_2\right| \\
&+\sup\left|\int_{|t_1|>T_{1n}}\int_{-T_{2n}}^{T_{2n}} e^{-i(t_1x_1+t_2x_2)}\phi_{X_1,X_2}(t_1,t_2)\,dt_1\,dt_2\right| \\
&+\sup\left|\int_{-T_{1n}}^{T_{1n}}\int_{|t_2|>T_{2n}} e^{-i(t_1x_1+t_2x_2)}\phi_{X_1,X_2}(t_1,t_2)\,dt_1\,dt_2\right| \\
&\qquad\qquad+\sup\left|\int_{|t_1|>T_{1n}}\int_{|t_2|>T_{2n}} e^{-i(t_1x_1+t_2x_2)}\phi_{X_1,X_2}(t_1,t_2)\,dt_1\,dt_2\right| \qquad (51)
\end{aligned}
$$

If $\phi_{X_1,X_2}(t_1,t_2)$ is absolutely integrable, the final three terms are $o(1)$. The first term is bounded by[31]

$$
\begin{aligned}
&\frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}\left|\widehat{\phi}_{X_2}(t_2)\frac{\widehat{\phi}_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}\right|\left|\widehat{\phi}_{X_1}(t_1)-\phi_{X_1}(t_1)\right|\,dt_1\,dt_2 \\
&+\frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}\left|\phi_{X_1}(t_1)\frac{\widehat{\phi}_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}\right|\left|\widehat{\phi}_{X_2}(t_2)-\phi_{X_2}(t_2)\right|\,dt_1\,dt_2 \\
&\qquad+\frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}\left|\phi_{X_1}(t_1)\phi_{X_2}(t_2)\right|\left|\frac{\widehat{\phi}_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}-\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right|\,dt_1\,dt_2 \qquad (52)
\end{aligned}
$$

Let $K_0(t)=\inf_{|s|\leq t}|\phi_{V_0}(s)|$ and denote $\theta_n=K_0(T_{1n})$ and $\vartheta_n=K_0(T_{2n})$.[32] Note that, under

[31] Using the identity $\hat{a}\hat{b}\hat{c}-abc=(\hat{a}-a)\hat{b}\hat{c}+a(\hat{b}-b)\hat{c}+ab(\hat{c}-c)$.

[32] Note that if $\phi_{V_0}(t)\neq 0$ for all t, then continuity of $\phi_{V_0}(t)$ implies that $K_0(t)>0$ for all t.

the uniform convergence assumption made in Theorem 2,

$$
\begin{aligned}
\left|\frac{1}{\widehat{\phi}_{V_0}(t_1)}\right| &= \frac{1}{\left|\phi_{V_0}(t_1)\left(\frac{\widehat{\phi}_{V_0}(t_1)-\phi_{V_0}(t_1)}{\phi_{V_0}(t_1)}+1\right)\right|} \\
&\le \frac{1}{|\phi_{V_0}(t_1)|\left(1-\left|\frac{\widehat{\phi}_{V_0}(t_1)-\phi_{V_0}(t_1)}{\phi_{V_0}(t_1)}\right|\right)} \\
&\le \frac{1}{\inf_{|t_1|\le T_{1n}}|\phi_{V_0}(t_1)|\left(1-\frac{\sup_{|t_1|\le T_{1n}}\left|\widehat{\phi}_{V_0}(t_1)-\phi_{V_0}(t_1)\right|}{\inf_{|t_1|\le T_{1n}}\left|\phi_{V_0}(t_1)\right|}\right)} \\
&= \frac{1}{\theta_n\left(1-o\left(\frac{\beta_n}{\theta_n}\right)\right)}
\end{aligned} \tag{53}
$$

for $|t_1| \le T_{1n}$ and almost surely. Similarly, for $|t_2| \le T_{2n}$,

$$
\left|\frac{1}{\widehat{\phi}_{V_0}(t_2)}\right| \le \frac{1}{\vartheta_n\left(1-o\left(\frac{\beta_n}{\vartheta_n}\right)\right)} \quad \text{a.s.} \tag{54}
$$

For the first term of (52), by Lemma 2 we have

$$
\begin{aligned}
&\frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}\left|\widehat{\phi}_{X_2}(t_2)\frac{\widehat{\phi}_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}\right|\left|\widehat{\phi}_{X_1}(t_1)-\phi_{X_1}(t_1)\right| dt_1\, dt_2 \\
&\le CT_{1n}T_{2n}\frac{\sup_{|t_1|\le T_{1n}}\left|\widehat{\phi}_{X_1}(t_1)-\phi_{X_1}(t_1)\right|}{\theta_n\left(1-o\left(\frac{\beta_n}{\theta_n}\right)\right)\vartheta_n\left(1-o\left(\frac{\beta_n}{\vartheta_n}\right)\right)} \\
&= o\left(\frac{\alpha_n T_{1n}T_{2n}}{\theta_n\vartheta_n}\right) \quad \text{a.s.}
\end{aligned} \tag{55}
$$

Analogously, the second term of (52) can be shown to be a.s. bounded by $o\left(\frac{\alpha_n T_{1n}T_{2n}}{\theta_n\vartheta_n}\right)$. Finally, consider the third term of (52). Note that

$$
\begin{aligned}
&\frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}|\phi_{X_1}(t_1)\phi_{X_2}(t_2)|\left|\frac{\widehat{\phi}_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}-\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right| dt_1\, dt_2 \\
&\le \frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}|\phi_{X_1}(t_1)\phi_{X_2}(t_2)|\left|\frac{\widehat{\phi}_{V_0}(t_1+t_2)-\phi_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}\right| dt_1\, dt_2 \\
&\quad + \frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}|\phi_{X_1}(t_1)\phi_{X_2}(t_2)|\left|\frac{\phi_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}-\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right| dt_1\, dt_2
\end{aligned} \tag{56}
$$

For the integrand in the first term on the right-hand side of (56), under the uniform convergence assumption of Theorem 2, we have

$$\left|\frac{\widehat{\phi}_{V_0}(t_1+t_2)-\phi_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}\right| \leq \frac{o(\beta_n)}{\theta_n\vartheta_n\left(1-o\left(\frac{\beta_n}{\theta_n}\right)\right)\left(1-o\left(\frac{\beta_n}{\vartheta_n}\right)\right)} \quad \text{a.s.} \tag{57}$$

Using the same line of proof as in (53), for the integrand in the second term on the right-hand side of (56) we have

$$\begin{aligned}
&\left|\frac{\phi_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}-\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right| \\
&\leq \left|\frac{\phi_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}\right|+\left|\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right| \\
&\leq \frac{|\phi_{V_0}(t_1+t_2)|}{|\phi_{V_0}(t_1)\phi_{V_0}(t_2)|}\left(\frac{1}{\left(1-\left|\frac{\widehat{\phi}_{V_0}(t_1)-\phi_{V_0}(t_1)}{\phi_{V_0}(t_1)}\right|\right)\left(1-\left|\frac{\widehat{\phi}_{V_0}(t_2)-\phi_{V_0}(t_2)}{\phi_{V_0}(t_2)}\right|\right)}+1\right) \\
&\leq \left|\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right|\left(\frac{1}{\left(1-o\left(\frac{\beta_n}{\theta_n}\right)\right)\left(1-o\left(\frac{\beta_n}{\vartheta_n}\right)\right)}+1\right) \quad \text{a.s.}
\end{aligned} \tag{58}$$

Therefore I obtain for the second term on the right-hand side of (56)

$$\begin{aligned}
&\frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}|\phi_{X_1}(t_1)\phi_{X_2}(t_2)|\left|\left(\frac{\phi_{V_0}(t_1+t_2)}{\widehat{\phi}_{V_0}(t_1)\widehat{\phi}_{V_0}(t_2)}-\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right)\right| \\
&\leq \frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}\left|\phi_{X_1}(t_1)\phi_{X_2}(t_2)\frac{\phi_{V_0}(t_1+t_2)}{\phi_{V_0}(t_1)\phi_{V_0}(t_2)}\right|\left(\frac{1}{\left(1-o\left(\frac{\beta_n}{\theta_n}\right)\right)\left(1-o\left(\frac{\beta_n}{\vartheta_n}\right)\right)}+1\right) \\
&= \frac{1}{(2\pi)^2}\int_{-T_{1n}}^{T_{1n}}\int_{-T_{2n}}^{T_{2n}}|\phi_{X_1,X_2}(t_1,t_2)|\left(\frac{1}{\left(1-o\left(\frac{\beta_n}{\theta_n}\right)\right)\left(1-o\left(\frac{\beta_n}{\vartheta_n}\right)\right)}+1\right)
\end{aligned} \tag{59}$$

By dominated convergence if $\phi_{X_1,X_2}(t_1,t_2)$ is absolutely integrable and if $\beta_n/\theta_n = o(1)$ and $\beta_n/\vartheta_n = o(1)$, then this last term is $o(1)$ almost surely. Finally, from (57), the last term on the right-hand side of (52) is a.s. of order $\frac{o(\beta_n)T_{1n}T_{2n}}{\theta_n\vartheta_n(1-o(\beta_n/\theta_n))(1-o(\beta_n/\vartheta_n))} = o\left(\frac{\beta_n T_{1n}T_{2n}}{\theta_n\vartheta_n}\right)$. Hence I have proved that $\sup \left| \widehat{f}_{X_1,X_2}(x_1,x_2) - f_{X_1,X_2}(x_1,x_2) \right|$ is

$$o\left(\frac{\alpha_n T_{1n} T_{2n}}{\theta_n \vartheta_n}\right) + o\left(\frac{\beta_n T_{1n} T_{2n}}{\theta_n \vartheta_n}\right) \tag{60}$$

almost surely.