



**UNIVERSIDAD DE SAN ANDRÉS**

**ESCUELA DE ADMINISTRACIÓN Y NEGOCIOS**

**MAGISTER EN FINANZAS**

**UNA NUEVA METODOLOGÍA PARA IDENTIFICAR LOS  
DETERMINANTES DE LA LIQUIDEZ ACCIONARIA RELEVANTES  
PARA DISTINTOS PERFILES DE LA INVERSIÓN**

**AUTOR: MONTEAGUDO MARÍA DEL PILAR**

**DNI/PAS: 36.739.020**

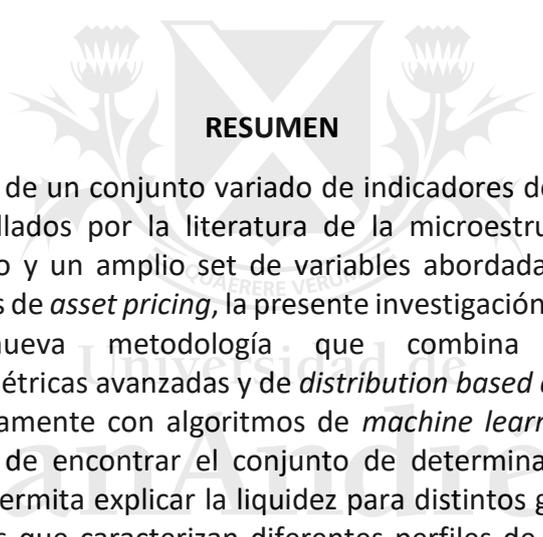
**DIRECTOR DEL TRABAJO FINAL DE GRADUACIÓN: MG. JAVIER MARCUS**

**LUGAR Y FECHA: BUENOS AIRES, 24 DE JULIO DE 2020**

# UNA NUEVA METODOLOGÍA PARA IDENTIFICAR LOS DETERMINANTES DE LA LIQUIDEZ ACCIONARIA RELEVANTES PARA DISTINTOS PERFILES DE LA INVERSIÓN

María del Pilar Monteagudo

UDESA



## RESUMEN

A partir de un conjunto variado de indicadores de liquidez desarrollados por la literatura de la microestructura de mercado y un amplio set de variables abordadas por los estudios de *asset pricing*, la presente investigación presenta una nueva metodología que combina técnicas econométricas avanzadas y de *distribution based clustering* conjuntamente con algoritmos de *machine learning* a los efectos de encontrar el conjunto de determinantes que mejor permita explicar la liquidez para distintos grupos de acciones que caracterizan diferentes perfiles de inversión en el mercado local, ello sobre la base de construir índices compuestos por las medidas de liquidez que mejor predicen los retornos al interior de cada clúster. A partir de los resultados obtenidos, este estudio tiene como fin último brindarle al mercado local herramientas que le permitan una gestión más eficiente de la liquidez direccionada por grupo de acciones que comparten entre sí patrones intradiarios de liquidez.

**Palabras clave:** liquidez, microestructura de mercado, *asset pricing*, acciones, *distribution based clustering*, *machine learning*, econometría

**Códigos JEL:** G12, C01, C58

## **Tabla de contenidos**

1.-INTRODUCCIÓN	4
2.- MARCO TEÓRICO	6
2.1- Microestructura de mercado y asset pricing: el rol de la liquidez	6
2.1. a- Liquidez y retornos esperados en mercados accionarios desarrollados	8
2.1. b- Ampliando el panorama: liquidez y retornos esperados en mercados accionarios emergentes	12
2.2.- La liquidez accionaria y sus principales determinantes	13
2.2. a.- Alternativas para medir la liquidez	13
2.2. b.- Explorando los determinantes de la liquidez en los mercados accionarios: una revisión de las principales variables de interés	18
3.- METODOLOGÍA	24
3.1- Relación entre liquidez y retornos para grupos de acciones específicos agrupados por distribuciones de liquidez intra diaria	24
3.1. a.- Ecuación general de retornos para determinar un índice de liquidez representativo por clúster de acciones	24
3.1. b.- Indicadores de liquidez relevantes según el perfil de la inversión: modelización por clúster	27
3.1. c.- <i>Dimensionality reduction technique</i> y formas funcionales flexibles para la selección del mejor modelo por clúster	30
3.2- Determinantes de la liquidez: un modelo por clúster multifactorial	32
4.- RESULTADOS EMPÍRICOS	35
4.1.- Estadísticas descriptivas para la muestra de acciones del panel general	35
4.2.- Clusterización óptima para la muestra de acciones del panel general	38
4.2.- Mejor modelo de exceso de retornos por clúster y determinación del índice de liquidez	46
4.3.- Mejor modelo de determinantes de liquidez por clúster	58
5.- CONCLUSIONES	63
6.- REFERENCIAS BIBLIOGRÁFICAS	64
7.- CÓDIGOS PARA REPRODUCIR EN STATA	74

## 1.-INTRODUCCIÓN

Dentro del amplio campo de la investigación financiera, los procesos de intercambio y formación de precios en los mercados han acaparado la atención de diversos autores, de cuyos estudios surgidos principalmente a partir de las últimas dos décadas del siglo pasado, fue brotando el sólido cuerpo teórico que dio vida a la *microestructura de mercado* como subdisciplina de las finanzas, estrechamente ligada y complementaria a las tradicionales teorías de *asset pricing*. Abocada a dimensionar el impacto que tanto los mecanismos como los costos de transacción ejercen sobre el comportamiento de los precios de diferentes activos (Liu, 2010) -en otros términos, la magnitud del desvío de los precios de sus valores de equilibrio (Biais, Glosten & Spatt, 2004)-, y de los actores que participan de los mercados, la microestructura de mercado ha logrado identificar y dimensionar la influencia de un conjunto de imperfecciones de mercado, como aquellas señaladas por Liu:

*“Financial markets deviate from the perfect-market ideal in which there are no impediments to trade. A large and growing body of work has identified a variety of market imperfections, ranging from information asymmetries, participation costs, different forms of trading costs, inventory risk (i.e., the market maker, being exposed to the risk of price changes while he holds in inventory, requires compensation), to search frictions (i.e., a trade off between search and quick trading at a discount), etc.”*Liu (2010, p.41).

De los diversos aspectos que analiza la microestructura, que pueden distinguirse e.g. en trabajos como los de Garman (1976) y O'Hara's (1999), entre otros, la magnitud en la cual una menor liquidez –como uno de los principales costos de transacción-, impacta en el retorno de las distintas clases de activos (como las acciones), constituye uno de los aspectos de mayor interés dentro de la literatura especializada (Madhavan, 2000). Dentro de este marco, y conjuntamente con el cuerpo teórico aportado por los modelos clásicos de *asset pricing*, una multiplicidad de estudios han debatido respecto a las medidas y modelos cuantitativos más apropiados a fin de mensurar este vínculo, no solo para el caso de los países desarrollados (véase por ejemplo Amihud y Mendelson, 1986; Amihud, 2002; Amihud, Mendelson y Lauterbach, 1997; Chordia, Subrahmanyam y Anushman, 2001; Brennan et al., 1998; Brennan y Subrahmanyam, 1996; Brennan, Chordia y Subrahmanyam, 2012; Brunnermeier y Pedersen, 2005; Chalmers y Kadlec, 1998; Chen y Kan 1996; Chen, Grundy y Stambaugh, 1990; Datar, Naik y Radcliffe, 1998; Easley et al., 1999; Easley y O'Hara, 1987; Eleswarapu y Reinganum, 1993; Elewarapu, 1997; Fama y French, 1993; Fama & Macbeth, 1973; Glosten y Harris, 1988; Jun, Marathe & Shawk, 2003; Kyle, 1985; Lee, 2011; Lee y Swaminathan, 2000; Lee, 1993; Liang y Wei, 2012; Merton, 1987; Pastor y Stambaugh, 2003; Spiegel y Wang, 2005; Subrahmanyam, 2010, entre otros), sino también para los mercados de países emergentes (Kearney, 2012; Lischewski y Voronkova, 2012; Rouwenhorst, 1999; Claessens, Dasgupta y Glen, 1998; Jun, Marathe y Shawk, 2003; Lischewski y Voronkova, 2012; Lesmond, 2005; Lam y Tam, 2011; Brown et al., 2008; Hearn y Piesse, 2009, entre otros).

No obstante, comprender exhaustivamente el fenómeno de la liquidez no solamente requiere analizar su impacto en los retornos, sino también identificar el conjunto de los

factores que permitan explicar propiamente el comportamiento de la liquidez al interior de cada mercado. Al respecto, un conjunto de estudios han trabajado sobre esto último empleando diferentes enfoques (Chordia, Sarkar & Subrahmanyam, 2004; Chordia et al., 2000; Stoll, 2000; Lee et. al., 1993; Hasbrouck and Seppi, 2001; Coughenour, J. F., & Saad, M. M., 2004; Cumming, D., Johan, S., & Li, D., 2011; Jain, 2003; entre otros). En lo particular, un análisis exhaustivo se vuelve todavía más relevante en mercados poco profundos, donde la falta de liquidez y con ello el exceso de volatilidad constituyen los dos riesgos más críticos y disuasivos para los inversores.

Desde un punto de vista del alcance práctico, la identificación de los principales focos causales de la liquidez constituye un insumo sumamente relevante a la hora de otorgar efectividad al diseño de políticas que permitan ir superando paulatinamente las fricciones al interior de los mercados y de las firmas y generen de este modo señales más eficientes. No obstante, en mercados concentrados y volátiles, con pocas firmas cotizando y con patrones de operación intra-diaria muy disímiles (en otros términos, con escala profundidad, amplitud y flexibilidad operativas, siguiendo la definición de liquidez de Hasbrouck, 2007), evaluar la liquidez desde un punto de vista general (y no segmentado) puede ocasionar sesgos en la interpretabilidad del ajuste de los modelos y la significancia de las variables. En otras palabras, las diferentes medidas de liquidez abordadas exhaustivamente en la literatura de la microestructura de mercado permiten captar las diversas dimensiones del fenómeno, pero que no siempre pueden resultar simultáneamente significativas para grupos de acciones que exhiben perfiles de riesgo de la inversión marcadamente diferentes. Una vez obtenidos los resultados a partir de llevar a cabo un análisis pormenorizado con técnicas cuantitativas avanzadas, el fin último de esta investigación consiste en proporcionar al mercado local herramientas útiles para la gestión de la liquidez accionaria partiendo de un enfoque de clústeres que definen diferentes perfiles de la inversión de acuerdo con patrones intra diarios de las operaciones.

En atención a lo expuesto, y siendo este estudio el primero para Argentina en abordar en profundidad el fenómeno de la liquidez accionaria, se propone una nueva metodología que combina el uso de técnicas de econometría avanzadas, *distribution based clustering* y algoritmos de *machine learning* a los efectos de encontrar el conjunto de determinantes que permita explicar la liquidez para distintos grupos de acciones, ello partiendo de las medidas de liquidez que mejor predigan los retornos esperados al interior de cada clúster. El acceso a datos intradiarios de la negociación de cada acción en el mercado permitirá la construcción de clústeres por aproximación de distribuciones de liquidez, empleando para ello una base de más de un millón doscientos mil registros (1.200.000). El análisis será efectuado para todo el conjunto de acciones que operaron en el mercado argentino durante el período 2009q4-2019q2.

Dando cuenta de lo expuesto, el resto del trabajo se estructura como sigue. En la Sección 2 se desarrolla una revisión de la literatura sobre la relación entre liquidez y retornos tanto para países desarrollados y emergentes, así como también las diferentes medidas de liquidez (y las dimensiones que capturan) y sus posibles determinantes. A continuación, en la Sección 3, se abordan los aspectos metodológicos adoptados en el presente estudio y los modelos a contrastar empíricamente. Por su parte, en la Sección 4 se relevan los resultados empíricos obtenidos, mientras que en la Sección 5 se abordan las conclusiones principales de esta investigación.

## 2.- MARCO TEÓRICO

### 2.1- Microestructura de mercado y asset pricing: el rol de la liquidez

La complejidad y el alcance de los factores que afectan a los retornos esperados de las diferentes clases de activos han ocupado la atención tanto de académicos como prácticos. En este sentido, tanto la microestructura de mercado como el asset pricing se fundamentan en la necesidad de análisis del comportamiento y la formación de precios en los mercados de activos financieros; no obstante, el primer cuerpo teórico toma en consideración los aspectos particulares de los mecanismos de transacción mientras que el segundo focaliza en (Easley & O'Hara, 2003, Madhavan, 2000; Biais, Glosten & Spatt, 2005) los fundamentos económicos. Dicho en otros términos, mientras que la literatura tradicional del asset pricing indaga acerca del rol que juegan diversas clases de riesgos idiosincráticos y no idiosincráticos en los retornos esperados, la microestructura de mercado focaliza detenidamente en el efecto que las fricciones de mercado, como e.g. la ausencia de liquidez, ejercen sobre los retornos esperados, complementándose por ello entonces los atributos de las firmas y del mercado de ambos cuerpos teóricos (Spiegel & Wang, 2005).

Como fue mencionado, uno de los aspectos críticos que caracteriza a los distintos tipos de activos financieros y mercados es su grado de liquidez. Como señalan Brennan, Chordia & Subrahmanyam,

*“The liquidity of an asset market refers to the ability of investors to buy and sell significant quantities of the asset, quickly, at low cost, and without a major price concession”* Brennan, Chordia & Subrahmanyam (2012, p.523).

Asimismo, y de acuerdo con Hasbrouck (2007), la liquidez ha sido definida también como sinónimo de “profundidad, amplitud y resiliencia (flexibilidad)”- basándose en Kyle (1985)-. En efecto, en un mercado con profundidad existe oferta y demanda suficiente para concretar una operación de compra o venta por encima y por debajo del precio de mercado, contando con un número suficiente de órdenes de compra y venta. En mercados flexibles, los efectos sobre el precio vinculados al proceso de negociación son pequeños y se diluyen en el cortísimo plazo, es decir, la reacción de los agentes antes cambios en los precios de los activos se concreta con agilidad. En un mercado con amplitud, se cuenta con una cantidad suficiente de volumen negociado, lo que permite que se satisfagan las preferencias de los demandantes y oferentes de forma constante. De este modo, cuanto más líquido el mercado menor la distancia entre los precios y sus valores fundamentales.

Como explican Brennan y Subrahmanyam (1996), algunos de los principales modelos sobre formación de precios en mercados financieros sugieren que los inversores que cuentan con información privada generan costos de iliquidez para los no informados, lo que se traduce en que los activos relativamente más ilíquidos requieran de mayores tasas de retorno. Siguiendo esta línea argumental, Bagehot (1971) distingue como causa principal de la iliquidez en los mercados financieros a los problemas de selección adversa, que precisamente derivan de la existencia de *traders* con información privada.

De la categorización brindada por Amihud, Mendelson & Pedersen (2005), pueden distinguirse diversas fuentes de (i) liquidez. En referencia, una primera fuente de iliquidez puede encontrarse en los costos de transacción exógenos como e.g., gastos de corretaje, costos de procesamiento de pedidos o impuestos sobre transacciones. Igualmente, se pueden identificar también las presiones de demanda y el riesgo de inventario. En otras palabras, cuando los agentes mantienen una presencia que no es permanente en el mercado (o bien no hay el suficiente volumen), los oferentes deben vender la porción deseada de su posición a un hacedor de mercado o *market maker* (cuya función es precisamente proveer inmediatez), que probablemente compre anticipadamente, para luego deshacer su posición (reventa). Esta maniobra, no obstante, llevará a este último a exponerse al riesgo de un cambio en el precio todo tiempo que mantenga el activo en su poder, hasta que logre venderlo, situación que se traducirá en un costo para el vendedor<sup>1</sup>. Asimismo, es probable que se generen transacciones entre partes con desigual información, no solo en términos del valor fundamental del activo sino también de la dinámica de los futuros flujos de órdenes, evento que se traducirá en mayores costos (véase Akerlof, 1970 sobre el problema de la selección adversa en mercados financieros, i.e., los traders informados venderán con malas noticias y comprarán con buenas y Bagehot, 1971 sobre las ganancias del *market maker* con contrapartes desinformadas<sup>2</sup>). Precisamente, Grossman y Stiglitz (1980) sugieren que existe en realidad un nivel de “iliquidez de equilibrio”. En efecto, el mercado debe tener un nivel de iliquidez suficiente como para compensar a los *market makers* (que tienen incentivos a invertir recursos en recolectar la información del mercado) y al mismo tiempo permitir que el negocio sea rentable para los nuevos proveedores de liquidez.

Volviendo la mirada sobre los aspectos diferenciadores de la literatura del asset pricing y la microestructura de mercado, Chordia, Sarkar & Subrahmanyam (2005) señalan que existen diversas fricciones de mercado, como costos de transacción, restricciones de venta en corto, que impactan en la formación de precios. No obstante, reconocen que la mayor parte de los modelos de asset pricing no consideran las fricciones propias del intercambio, es decir, el tiempo y el costo de transformar el dinero en activos financieros, o viceversa. Dicho de otro modo, los modelos tradicionales de la literatura del asset pricing, a diferencia de la microestructura, asumen mercados perfectamente líquidos, basándose en los principios de permanente no arbitraje, optimalidad de agentes y equilibrio, i.e., que la naturaleza de la transacción no afecta los precios y por ende los agentes son simples tomadores de precios (Amihud, Mendelson & Pedersen, 2005).

Esquemáticamente, el modelo presentado en Easley & O’Hara (2003) permite evidenciar el mecanismo detrás del comportamiento de los precios, dando cuenta de la existencia de fricciones de mercado. Sea

$$P_t = V_t + \Delta_t$$

---

<sup>1</sup> Véase un modelo de Garman (1976) sobre el market maker monopolista

<sup>2</sup> Véase también Copeland y Galai (1983) sobre un modelo de maximización del beneficio del market maker; la investigación seminal de Kyle (1985) y la extensión de Mendelson y Tunca (2004).

Donde  $V_t$  representa el “precio de eficiencia” y  $\Delta_t$  el conjunto de variables que reflejan las distintas fricciones presentes en la literatura de la microestructura. Por ejemplo,  $\Delta$  es el spread bid-ask y  $\Delta_t = \Delta \gamma_t$ , donde  $\gamma_t$  es un indicador que toma valor +1 si la negociación se produce en el momento  $t$  y es una compra y -1 si se realiza en el mismo momento, pero es una venta (véase el desarrollo preliminar de Roll, 1984). Como indican los autores, en un mercado eficiente donde los precios se guían por una *martingala*, el precio no observable  $V_t$  representa una expectativa condicional dada la información del mercado sobre el valor aleatorio de los activos. Si, por ejemplo, los *dealers* presentes en el mercado quisieran manejar sus inventarios afectando los precios, la ecuación se modificaría para reflejar su preferencia de inventario. Al respecto, Hasbrouck señala,

*“The evolution of thought on security transaction price behavior has passed from basic martingale models, to non informational cost models (order processing and inventory control paradigms), and finally to models that incorporate the distinctly informational and strategic aspects of trading. From an economic perspective, the actual security price in many microstructure models can be interpreted as an idealized “informationally efficient” price, corrupted by perturbations attributable to the frictions of the trading process.” Hasbrouck (1996, pp.648-649)*

## **2.1. a- Liquidez y retornos esperados en mercados accionarios desarrollados**

La controversia respecto a la medida en que la liquidez constituye un factor de injerencia en el precio de los activos financieros ha sido plasmada en un vasto número de investigaciones especializadas. En este sentido, estudios teóricos como el de Merton (1987) no solamente afirman que la liquidez debe ser considerada en la formación de precios sino también que debe ser incorporada dentro de los riesgos idiosincráticos, a diferencia de lo expuesto en el modelo de asset pricing tradicional CAPM. Contrariamente, los desarrollos teóricos de Constantinides (1996), Heaton y Lucas (1996) y Vayanos (1998) argumentaron que el efecto de los costos de transacción se diluye en tanto pueden ser minimizados negociando activos ilíquidos con poca frecuencia. Como señalan Datar, Naik y Radcliffe (1998), estos autores sostuvieron que los costos de transacción resultaban pequeños en relación con la prima de riesgo de equilibrio como para generar una diferencia significativa. Ahora bien, O'Hara (2003) señala que esta visión de la liquidez como un simple costo de transacción similar a un impuesto, subestima su verdadero impacto sobre los retornos esperados, en tanto que la liquidez viene aparejada a problemas de asimetría de información (véase también Subrahmanyam, 1991).

En atención a esta problemática, muchas investigaciones se han abocado a contrastar empíricamente la significatividad de las distintas fricciones de mercado como predictoras de los retornos de las acciones en países desarrollados (Subrahmanyam, 2010). Dentro de este marco, la medida en que un menor grado de liquidez interfiere en la formación de precios, generando que los inversores exijan retornos diferenciales, ha captado especial atención. En efecto, la exigencia de mayores retornos por falta de

liquidez puede traer aparejados impactos económicos significativos y de diversa índole, entre los cuales uno muy relevante, explicado por Jacoby et al. (2000), resulta en el encarecimiento del costo de capital en el ámbito corporativo, lo que termina por condicionar sustantivamente a las firmas a nivel de toma de decisiones de inversión.

Ahondando en la diversidad de este conjunto de estudios empíricos, Amihud y Mendelson (1986) y Elewarapu (1997) han encontrado el efecto positivo del diferencial de precio comprador-vendedor (*bid-ask spread*) sobre los retornos de los activos ajustados por riesgos. Respecto a esta primera investigación, Amihud y Mendelson (1986) analizan, tomando como muestra a las acciones del NYSE y AMEX para el período 1961-1980, la relación retornos-liquidez en el marco de un modelo que permite a los agentes del mercado diversificar y poseer activos con diferentes horizontes temporales, asumiendo los costos de ejecución asociados. En este sentido, este trabajo seminal pone de manifiesto que solamente los *traders* con horizontes más largos retienen a los activos ilíquidos, demandando una compensación. De este modo, el resultado obtenido contribuye en un doble sentido, sosteniendo que, en el equilibrio, el retorno de los activos no solamente será una función creciente sino también cóncava del spread bid-ask (como medida de iliquidez). Este hallazgo es luego sostenido por las investigaciones de Eleswarapu (1997) y Chalmers y Kadlec (1998), quienes estudiaron los spreads amortizados en relación con los retornos.

Por su parte, Easley, Hvidkjaer y O'Hara (2002), encontraron que la liquidez, como medida del costo de selección adversa producto de la información asimétrica entre *traders* – para lo cual utilizaron datos intradiarios- tiene un efecto significativo y positivo sobre los retornos. Igualmente, en un estudio pionero, Brennan y Subrahmanyam (1996) midieron la liquidez como la respuesta de su precio con relación al flujo de las órdenes basándose en el análisis de Glosten y Harris (1988), y por el costo fijo de cada transacción, valiéndose de datos intradiarios.

Dentro de este mismo orden de ideas, Brennan, Chordia y Subrahmanyam (1998) evaluaron la significatividad, para el período de 1966-1995, de la liquidez medida como el volumen negociado de cada stock en un modelo de asset pricing multifactorial, donde el exceso de retornos de cada activo ajustado por riesgo fue regresado transversalmente contra el logaritmo del volumen de cada activo, incorporando otras variables de control como tamaño, ratio valor libro/valor de mercado de cada compañía, rentabilidad por dividendos y retornos pasados (a fin de captar el efecto *momentum*). Los resultados empíricos ponen en evidencia que el volumen posee un efecto negativo y significativo sobre los retornos (exceso) de cada activo ajustados por riesgo.

Amihud, Mendelson y Lauterbach (1997), analizando el fenómeno del cambio en los mecanismos de negociación en el mercado de acciones de Tel Aviv, evaluaron el impacto sobre los retornos de aquellos activos que habían sido transferidos gradualmente de un sistema de negociación de *call auctions* diarias a un mecanismo en el cual cada *call auction* era seguida de rondas de negociación de iteración permanente. En efecto, los autores encontraron que el crecimiento acumulado promedio de los retornos ajustados en los activos estudiados fue de alrededor del 5,5% producto de estos cambios, debido fundamentalmente a la existencia de mayor liquidez. Por su parte, en un estudio posterior, Amihud (2002) también encontró que los retornos esperados de los activos son una función creciente de la iliquidez esperada, valiéndose de un ratio que sigue el

concepto de iliquidez de Kyle y refleja el cambio (promedio) absoluto en el precio por cada dólar de volumen diario negociado. Esta medida de iliquidez (en el sentido negativo) fue empleada en un modelo de corte transversal, en el cual fueron regresado los retornos mensuales de cada acción contra la iliquidez del año inmediatamente anterior y otras variables de control como beta (riesgo sistemático), tamaño, volatilidad, rentabilidad por dividendos y retornos pasados, todo ello para las acciones del NYSE entre 1963-1996.

Como señala Lee, Mucklow y Ready (1993), el bid-ask spread resulta una medida sesgada de liquidez, en tanto que negociaciones de mucho volumen suelen ocurrir por fuera del spread, y muchas pequeñas al interior de este gap. Como respuesta a ello, estudios empíricos como los de Glosten y Harris (1988), basándose en aportes teóricos de Kyle (1985) y Easley y O'Hara (1987), han testeado el vínculo retornos-iliquidez mensurando los problemas de asimetrías de información, considerando su mayor alcance.

El estudio realizado sobre la liquidez en las acciones del NYSE de Brennan, Chordia y Subrahmanyam (2012), constituye una contribución en la comprensión del impacto sobre los retornos de la liquidez asociada a cada una de las dos aristas del mercado: los órdenes de compra y de venta. En otros términos, propone añadir al análisis de la liquidez en los modelos de asset pricing el supuesto de que los costos de transacción no son simétricos en la compra y en la venta. Como señalan los autores, la liquidez proveniente de las órdenes de venta resulta más significativa en explicar los retornos que la resultante de las órdenes de compra (de hecho, la significancia estadística del  $\lambda$ -compra desaparece cuando el  $\lambda$ -venta se incluye en la regresión  $-\lambda$  en el sentido de impacto sobre precio de Kyle).

*"We allow for an asymmetric relation and estimate separate buy and sell measures of illiquidity ("lambdas") for a large cross-section of stocks over a 26-year period, using a modified version of the Brennan and Subrahmanyam (1996) approach, which assumes that price responses are linear, and is an adaptation of the Glosten and Harris (1988) method. We find that sell-order illiquidity is priced more strongly in the cross-section of expected stock returns than is buy-order illiquidity. This result continues to obtain after controlling for other known determinants of expected returns such as firm size, book-to-market ratio, momentum, and share turnover." Brennan, Chordia & Subrahmanyam (2012, pp.524).*

Otra de las contribuciones seminales se apoya en el trabajo de Datar, Naik y Radcliffe (1998), quienes propusieron a la rotación de acciones como nueva medida de liquidez para evaluar el impacto sobre los retornos del NYSE en el período 1962-1991, controlando por el valor contable/valor de mercado, tamaño y beta de la firma (siguiendo a Fama y French, 1992) y el Efecto-Enero (Eleswarapu y Reinganum, 1993). Dentro de este marco, los autores argumentaron que la liquidez debía estar asociada a la frecuencia del *trading* y confirmaron su significatividad negativa con relación a los retornos mensuales de los activos. Este resultado fue corroborado, e.g., por Anushman, Chordia y Subrahmanyam (2001). Otros trabajos como los de Lee y Swaminathan (2000) sostuvieron que el efecto de la rotación debe observarse contemplando también el

impacto que el desempeño de los activos en el pasado ejerce sobre los retornos esperados.

Amihud (2002) ha dado con un fenómeno conocido como “*flight-to-quality*”, demostrando que una disminución en la liquidez de mercado se refleja en un impacto negativo sobre los retornos. El trabajo de Liang y Wei (2012), investiga precisamente la medida de liquidez de mercado de Amihud (2002) y Pastor y Stambaugh (2003) sobre un conjunto de 21 países desarrollados, encontrando que el riesgo de liquidez local afecta a los precios de los activos en 11 países de la muestra.

*“We discover that the expected returns of stocks systematically reflect these two measures of local liquidity risk to different extents although they are commonly priced in stocks in France, Ireland, and Japan. In particular, local liquidity risk based on Pástor and Stambaugh’s (2003) measure is positively priced in Australia, France, Hong Kong, Ireland, Japan, and Norway. Local liquidity risk based on Amihud’s (2002) measure is positively priced in Canada, Denmark, Finland, France, Germany, Ireland, Japan, New Zealand, Sweden, and the US. We also find that the pricing premium for local liquidity risk is lower in markets where corporate boards at the country level are more effective and where there are less insider trading activities.”* Liang y Wei (2012, p.3287).

Asimismo, y en correspondencia con Lee (2011), revela que el riesgo de liquidez global también es un factor transversal que afecta a los retornos en los portafolios construidos para cada uno de los países, luego de control por las variables de Fama-French (1992).

Desde otra perspectiva, no obstante, diversos estudios empíricos han encontrado evidencias en contra del poder explicativo de las distintas medidas de liquidez para explicar con relación a los retornos de los activos. Frente a esta formulación, la investigación de Spiegel y Wang (2005) pone de manifiesto no solamente la correlación negativa existente entre la liquidez (utilizando tanto el spread bid-ask como la medida de liquidez propuesta por Amihud, 2002) y el riesgo idiosincrático, como también evidencian Merton (1987) y Brunnermeier & Pedersen (2005), sino que cuando ambos factores son empleados simultáneamente para explicar los retornos, el poder explicativo de la primera se ve sumamente disminuido, controlando por variables como el volumen de acciones y la rotación. Ello implica que, controlando por variables de riesgo diversificable, los retornos resultan similares para activos con distinto nivel de liquidez<sup>3</sup>.

Por su parte, los análisis de Chen y Kan (1996), y Eleswarapu y Reinganum (1993) también presentan evidencias que manifiestan la ausencia de tal vínculo. En este sentido, Chen y Kan (1996) han argumentado que los resultados obtenidos por Amihud y Mendelson (1986) resultan inconsistentes en tanto que el modelo no ajusta apropiadamente los retornos a los riesgos. Más interesante aún es el estudio sumamente controversial de Eleswarapu y Reinganum (1993), quienes replicando el

---

<sup>3</sup>Estudios como los de Douglas (1968) también hallaron la incidencia del riesgo idiosincrático en la variación de corte transversal de los retornos. Lehmann (1990) y Xu y Malkiel (2002) demostraron que a mayor riesgo idiosincrático mayor retorno de los activos, en un análisis de corte transversal.

estudio de Amihud y Mendelson (1986), pusieron en evidencia que la prima en el retorno por iliquidez respondía a un factor de tipo estacional, más precisamente, que la significatividad de liquidez empleando el método de Fama-Macbeth (1973) para un análisis de corte transversal, solo era consistente en el mes de enero.

### **2.1. b- Ampliando el panorama: liquidez y retornos esperados en mercados accionarios emergentes**

Luego del impacto que estas investigaciones tuvieron en el mundo desarrollado, el estudio de la relación entre retornos y liquidez en mercados accionarios también se convirtió en una inquietud de los países emergentes, en especial porque abundan las fricciones en esta clase de mercados financieros (Batten & VinhVo, 2014).

Al igual que lo sucedido para países desarrollados, resultados contradictorios también se han hallado en el estudio del vínculo entre liquidez y retornos en mercados emergentes. En efecto, una de las investigaciones preliminares en la materia, la de Rouwenhorts (1999), testeó para 20 países emergentes los determinantes transversales de los retornos, encontrando resultados similares a los hallados en países desarrollados respecto a los factores de French y Fama (1992, 1993, 1998), pero para la rotación como medida de liquidez. Aún más llamativo fueron las conclusiones derivadas de los estudios de Claessens, Dasgupta y Glen (1998) y Jun, Marathe y Shawky (2003), quienes encontraron una relación sorprendentemente positiva entre los retornos esperados y la rotación, contradiciendo los hallazgos típicos del mercado estadounidense. En el mismo sentido, Lischewski y Voronkova (2012) en el primer estudio para el mercado polaco, ratificaron la ausencia de relación significativamente estadística entre retornos ajustados por riesgo y liquidez.

Sin embargo, como demuestran Bakaert, Harvey y Lundblad (2007) demostraron que para 19 países emergentes de Latinoamérica y el Sudeste Asiático, si bien la rotación como medida de liquidez resulta no significativa para explicar el comportamiento de los retornos en mercados emergentes, contrariamente a lo que sucede si se utiliza la proporción de retornos diarios cero de la firma en un plazo de tiempo, e.g., en un mes (medida que está fuertemente correlacionada con los costos de transacción en estos mercados). Esta última medida está basada en el estudio de Lesmond (2005), la cual intenta representar no solamente el spread comprador-vendedor y los costos de información, sino también los costos de comisión (i.e., costos implícitos, explícitos y de oportunidad). Como explican los autores,

*“(...) we use liquidity measures that rely on the incidence of observed zero daily returns in these markets. Lesmond et al. (1999) argue that if the value of an information signal is insufficient to outweigh the costs associated with transacting, then market participants will elect not to trade, resulting in an observed zero return.”* Bakaert, Harvey y Lundblad (2007, p. 1785).

En el mismo sentido, Lesmond (2005), examinando una muestra de 23 países emergentes y tomando como medida de liquidez al bid-ask spread, encontró que estos costos de liquidez en un modelo de corte transversal son significativamente estadísticos y negativos en relación a los retornos, con un impacto del 1% en Taiwan al 47% en el mercado ruso (véase un con resultados similares para Hong Kong, Lam y Tam (2011);

Hong Kong, Korea y Singapore, Brown et al., 2008 y los mercados accionarios africanos; Hearn y Piesse, 2009).

## **2.2.- La liquidez accionaria y sus principales determinantes**

### **2.2. a.- Alternativas para medir la liquidez**

Como puede desprenderse de lo tratado en la subsección previa, un aspecto sumamente controversial refiere a la metodología empleada para la medición de la liquidez. Al respecto, a lo largo de los distintos estudios pueden distinguirse diversas clases de medidas, algunas de las cuales serán empleadas en para el análisis empírico desarrollado en las secciones subsiguientes.

Siguiendo la clasificación brindada por Lesmond (2005), una clase de estimadores de la liquidez buscan captar el costo directo en que se puede incurrir cuando se lleva a cabo una determinada transacción. Dentro de este grupo, el indicador más representativo y básico es el *spread bid-ask cotizado* y sus derivados (i.e., aquellas que emplean las cotizaciones), aunque es de destacar que, como desventaja, las transacciones pueden finalmente ocurrir por fuera de las cotizaciones (i.e., los precios de cierre se derivan de las cotizaciones de los precios *bid* y *ask*). Asimismo, también se pueden encontrar medidas tradicionales como el coeficiente  $\lambda$  de *impacto precio* desarrollado por Kyle (1985), adoptada luego por numerosos estudios, la cual refleja la relación positiva entre el volumen negociado y el cambio en el precio. Asimismo, y como indica el autor no siempre esta clase de información está disponible para el tamaño y la frecuencia de muestra deseada, en especial en los mercados emergentes donde habitualmente se encuentran obstáculos para la construcción de grandes bases de datos.

No obstante, Goyenko et al. (2009) ha puesto de manifiesto que un buen número de medidas de liquidez construidas a partir de datos de baja frecuencia también resultan útiles para reflejar los costos de transacción que captan de forma directa estimadores como las ya vistos. En este marco, otra clase de estimadores de liquidez denominados reflexivas, asocian la liquidez con características de la operatoria como e.g., el volumen, ya sean las clásicas de rotación (Datar, Naik & Radcliffe, 1998), medida de Amihud (Amihud, 2002; Liu, 2006), etc. Sin embargo, estos estimadores desconocen el costo efectivo de llevar a cabo una determinada negociación (Jun, Marathe & Shawky, 2003).

Por último, existe otra clase de medidas de liquidez que, como señala el autor, reflejan indirectamente los costos de transacción a partir del comportamiento del precio. Dentro de este conjunto se encuentran el estimador de Roll (1984)- véase Buttler, Grullon & Weston (2005)-, la medida de Lesmond et al. (1999), entre otras.

Siguiendo la clasificación desarrollada, y a partir de una revisión profunda de más de 50 estudios en la materia, a continuación, se describen y sintetizan un conjunto de los principales estimadores de la liquidez.

Para la primera categoría de medidas definidas por Lesmond (2005), i.e, las que captan de forma directa el costo de transacción se pueden definir:

(1) Spread Bid-Ask Cotizado<sup>4</sup>:

$$QS = A_{i,t}^l - B_{i,t}^h$$

, es la medida más simple y se computa como la diferencia entre el precio ask más bajo ( $A_{i,t}^l$ ) y el precio bid más alto  $B_{i,t}^h$ . En la práctica, este diferencial de precios suele computarse al final del día para el mayor y menor precio alcanzado entre todas las operaciones concertadas, aunque también puede evaluarse para la ronda completa, e.g., con un spread cotizado promedio o relativo al volumen diario operado (Buttler, Grullon & Weston, 2005). Si, a su vez, se afecta esta última medida dividiéndola por el punto medio entre ambos precios (promedio) o bien por el precio de la operación, se obtiene el *spread bid-ask* relativo.

No obstante, este tipo de medidas supone que las transacciones se ejecutan tomando como referencia los extremos y no el precio de concertación real, lo cual presenta algunas limitaciones:

*“The quoted bid-ask spread is a benchmark measure of both liquidity and the cost of trading. But it is not an accurate measure of the trading cost. First, it tends to overstate the true spread because trades are often executed inside the quoted spread. Second, both the bid and ask prices have a systematic tendency to rise (fall) following a purchase (sale), so the round-trip trading costs are less than the quoted spread suggests.”* Armitage et al. (2014, p.164).

Atento a esto, el spread efectivo se basa en los precios al cual las operaciones han sido efectivamente ejecutadas, captando de este modo las operaciones dentro del spread.

(2) Spread Bid-ask efectivo:

Esta medida refiere al costo de ejecución inmediato, y se computa como la diferencia entre el precio de transacción y el punto medio cotizado:

$$ES = 2 \text{ abs } (P_{i,t} - M_{i,t}) = 2 IN_{i,t}(P_{i,t} - M_{i,t})$$

, donde  $P_{i,t}$  es el precio de transacción efectivo de la acción  $i$  para la operación  $t$ ,  $M_{i,t}$  se calcula como  $M_{i,t} = \frac{(B_{i,t} + A_{i,t})}{2}$ , y  $IN_{i,t}$  representa un indicador que vale 1 si la negociación se efectuó como compra o -1 si se inició como venta (Dicle, 2008); en otros términos, es dos veces el costo efectivo de la operación (Lee, Lin & Tsao, 2006). Esta medida se apoya en el supuesto que las operaciones por encima del punto medio son compras y las que están por debajo del promedio constituyen ventas.

Al igual que para el spread cotizado, el efectivo es conveniente verlo en términos relativos, ya sea al  $M_{i,t}$  o al  $P_{i,t}$ , siendo el *spread efectivo relativo* una medida más comparable. Si, a su vez, si se pondera el spread por el tiempo en que tarda en ejecutarse la transacción relativo al promedio de ejecución, obtenemos la medida propuesta por Hillert et al. (2016).

Si, en cambio, no se cuentan con datos de frecuencia intradiaria, la versión propuesta por Jain, Jiang & Mekhaimer (2016) considera valores de cierre, captando los costos

---

<sup>4</sup>Quoted bid-ask spread.

asociados al procesamiento de órdenes, inventario y selección adversa entre el cierre de una ronda y la siguiente. Esta medida se sintetiza a continuación,

(3) Spread Bid-ask realizado:

$$RS = 2 (P_{i,j}^c - M_{i,j+1}^c)$$

, donde  $P_{i,j}^c$  es el precio de cierre del día  $j$  y  $M_{i,j+1}^c$  el punto medio de las cotizaciones de los precios bid y ask al cierre de la siguiente ronda de negociación. No obstante, cabe mencionar que esta medida resulta tanto más precisa si se considera sobre una base intra diaria, tomando un intervalo de tiempo representativo del nivel de transparencia del mercado. En términos generales busca reflejar la diferencia entre el spread efectivo neto del costo de selección adversa resultado de movimientos en los precios desventajosos por otros operadores más informados, capturando de este modo el "costo real o verdaderamente realizado". Al igual que las otras dos medidas, también suele verse en términos relativos.

*"(The realized spread) is twice the signed difference between the closing transaction price and the midpoint of bid-ask quotes at the close of next session. This represents the order processing cost, inventory costs and rents of the suppliers of liquidity, net of adverse selection costs resulting from unfavorable price movements after the trade."* Jain (2013, p.8)

Tomando en consideración esta diferenciación entre el costo real y el costo resultante de movimientos en los precios, la medida precisa que busca captar este impacto precio es conocida como medida de Kyle.

(4) Impacto precio:

El impacto precio de Kyle (1985) estima el coeficiente  $\lambda$  resultante de la regresión univariada definida como

$$R_{i,t} = \theta + \lambda OIB_{i,t} + \varepsilon_{i,t}$$

, siendo  $R_{i,t}$  el retorno en  $t$  de la acción  $i$ , y  $OIB_{i,t}$  la diferencia entre el volumen de transacción de las operaciones iniciadas como compra y venta. Este coeficiente busca captar la sensibilidad del precio o flexibilidad del mercado al flujo de órdenes, esperándose que en términos generales una operación de compra empuje los precios hacia arriba, y viceversa.

Apoyándose en este mismo enfoque y siguiendo a Brennan, Chordia y Subrahmanyam (2012), también podría definirse un  $\lambda$  de compra y de venta, para captar el impacto diferencial de los costos de transacción asociados a operaciones de compra y de venta. Esta medida intenta captar una suerte de efecto por el desbalance de órdenes en el mercado, que presiona sobre los inventarios (Chordia et al., 2005).

(5) Medidas de Profundidad del Libro de Órdenes:

Como indica Hasbrouck (2007), y ya fuera mencionado precedentemente, la liquidez ha sido definida como sinónimo de "profundidad, amplitud y resiliencia (flexibilidad)". En este marco, un conjunto de medidas intenta captar la liquidez mediante el concepto de

profundidad, definiendo indicadores tales como: (i) volumen promedio de papeles ofrecidos al precio de compra o de venta en relación con el volumen total ofrecido, para una ronda de negociación o (ii) cantidad acumulada de órdenes para las mejores x mejores cotizaciones existentes durante una ronda (Jones y Lipson, 2001).

No obstante, como puede apreciarse, estas medidas de liquidez requieren de información intra diaria en mayor o menor nivel de detalle, datos que no son de acceso público. Es por ello que, a fin de poder hacer extensiva la investigación en este campo, diversos autores han reversionado estas medidas sin perder de vista las limitaciones propias de acceso en que incurren los investigadores. En este marco, a continuación, se exponen un conjunto de ellas, que se valen de otros atributos que permiten aproximar la liquidez con datos fácilmente asequibles y de frecuencia diaria.

(6) Volumen negociado:

$$Vol_{i,t} = Vol_{i,t}$$

, el volumen total negociado en t. Otra opción de medida vinculada al volumen e incorporada para mercados emergentes en Panigo, Grandes y Pasquini (2007) que permite reflejar la poca profundidad en este tipo de mercados es el porcentaje de días con volumen positivo.

(7) Días con volumen positivo:

$$Volumen\ Positivo = \frac{\sum_{i=1}^n \text{días volumen positivo}}{d}$$

(8) Rotación:

$$R_{i,t} = \frac{Vol_{i,t}}{Acciones\ en\ circulación_{i,t-1}}$$

, siendo  $Vol_{i,t}$  el volumen negociado y las acciones en circulación vigentes al inicio del período según datos del balance de cada una de las firmas (Datar, Naik & Radcliffe, 1998), captando así la velocidad en las operaciones de cada firma. Sin embargo, por la forma en que está construida, esta medida puede no representar la caída en la liquidez del mercado, e.g., en momentos de crisis de liquidez donde hay una escasez cada vez mayor de papeles disponibles para la negociación (Froot et al., 2001). Tomando esto en consideración, Tang, K., & Wang, C. (2011) definen la

(9) Rotación negociable:

$$RN_{i,t} = \frac{Vol_{i,t}}{Acciones\ negociables_{i,t-1}}$$

Esta forma de medir la liquidez, a diferencia de la rotación definida previamente, considera que una porción de las acciones en circulación no está necesariamente disponible para ser negociadas.

No obstante, y como indica Armitage et al. (2014), esta medida no contempla la variabilidad de costos de transacción entre acciones, como sí lo refleja la proporción de días con retorno cero. En otras palabras, un papel con altos costos de transacción

tenderá a tener movimientos de precios con menor frecuencia y por lo tanto más días con retorno cero que una acción con costos más bajos.

(10) Proporción de días con retornos cero:

$$\text{Retornos Cero} = \frac{\sum_{i=1}^n \text{díasretornocero}}{d}$$

, definida por Lesmond et al. (1999), siendo  $d$  la cantidad de días en un mes de negociación, y el numerador la sumatoria del número de días con retornos diarios iguales a cero.

Intentando replicar el efecto impacto precio de Kyle y la proporción de días con retornos cero a partir de datos con frecuencia diaria, Amihud (2002) desarrolló una medida de liquidez que puede definirse como,

(11) Illiquidez de Amihud:

$$\text{Amihud}_{i,t} = \frac{1}{d} \sum_{t=1}^n \frac{|R_{i,t,d}|}{\text{DVol}_{i,d,t}}$$

, siendo  $d$  el número de días de negociación en un período,  $|R_{i,t}|$  el valor absoluto de los retornos diarios promedio del activo  $i$  y  $\text{DVol}_{i,t}$  el volumen negociado del activo promedio en el mismo período  $t$ . Esta medida propuesta por Amihud (2002) busca captar la sensibilidad del precio (absoluto) por cada operación. Como señala Lesmond (2005), busca replicar el efecto del impacto precio de Kyle, pero con datos de más fácil acceso, al tiempo que permite mensurar la liquidez contemplando aquellos días en los cuales hay retornos cero, algo frecuente en los países emergentes- aunque queda indefinida para volumen igual a cero.

*“This illiquidity measure is strongly related to the liquidity ratio known as the Amivest measure, the ratio of the sum of the daily volume to the sum of the absolute return. [...] (This) liquidity ratio, however, does not have the intuitive interpretation of measuring the average daily association between a unit of volume and the price change, as does ILLIQ (Amihud’s measure)”.* Amihud, (2002, p. 34)

En particular, Brennan et al. (2012) proponen una modificación a la medida de Amihud, ajustando el retorno absoluto a la rotación, capturando así el exceso de retornos (como costo de transacción) por unidad de velocidad de circulación.

Por último, la última clase de estimadores emplean únicamente los precios como insumo principal. Entre las más empleadas se pueden distinguir:

(12) Medida de liquidez de Roll:

Buttler, Grullon & Weston (2005) definen un estimador implícito del spread efectivo, a partir de una versión propia de la medida de covarianza serial de Roll (1984).

$$\text{Roll} = 2\sqrt{-\text{cov}(\Delta(P_{i,t}; P_{i,t-1}))}$$

, en donde  $\Delta P_{i,t}$  es el cambio en el precio de transacción entre t y t-1, es decir, el retorno.

Como indican Jain et al. (2016), la medida propuesta por Roll (1984) está definida siempre que la covarianza sea negativa, por lo que la estimación de Gibbs modificada en Hasbrouck (2009) propone una solución de corrección.

(13) Gamma impacto precio de Pastor y Stambaugh (2003):

Esta medida busca replicar el impacto precio desarrollado por Kyle y extendido por Amihud, pero a través de un modelo multifactorial con datos de baja frecuencia,

$$R_{i,t}^e = \theta_i + \alpha R_{i,t} + \gamma(r_{i,t}^e * v_{i,t}) + \varepsilon_{i,t}$$

, donde  $R_{i,t}$  representa la tasa de retorno del activo (diaria),  $r_{i,t}^e$  el exceso de retorno del activo respecto al mercado y  $v_{i,t}$  el valor del volumen negociado, y siendo  $\gamma$  el coeficiente de regresión definido como la medida de liquidez.

## 2.2. b.- Explorando los determinantes de la liquidez en los mercados accionarios: una revisión de las principales variables de interés

En la subsección precedente han sido definidas las principales medidas de liquidez empleadas habitualmente por la literatura especializada y su forma de construcción, ello con el propósito de poder seleccionar la combinación de mejores estimadores para nuestra variable de estudio. Al mismo tiempo, y atento a que el objeto de estudio del presente trabajo radica precisamente en el entendimiento del fenómeno de la liquidez, aplicado en particular al mercado accionario argentino, a continuación, se estudiará en profundidad los modelos de determinación de liquidez abordados por las investigaciones en la materia. En virtud de ello, fueron analizados un conjunto variado de estudios que modelan la liquidez accionaria.

El cuadro a continuación exhibe los principales trabajos considerados y las principales variables exógenas empleadas en sus modelos de liquidez.

**Cuadro 1.- Presencia de los distintos conjuntos de regresores exógenos de la liquidez accionaria en la literatura especializada**

Estudio	Aspectos fundamentales de la firma	Variables Macroeconómicas e institucionales	Factores no diversificables y estacionales	Aspectos de la negociación de la firma
De Cesari, Espenlaub & Khurshed (2011)			x	x
Chordia, Sarkar & Subrahmanyam (2005)			x	x
Al Salamat (2016)	x	x		x
Chordia et al. (2001)			x	x
Jain (2003)	x			x
Stoll (2000)	x			x
Kuo, Chen & You (2017)	x		x	x
Hasbrouck and Seppi (2001)			x	x
Dicle, M. F. (2008).	x	x	x	x
Hearn & Piesse (2013)	x	x		x
Huang & Chang (2015)	x	x		

Stange, S., & Kaserer, C. (2008)	x		x	
Chan, K., Hameed, A., & Kang, W. (2013).	x			x
Tang, K., & Wang, C. (2011).	x			x
Jain, P., Jiang, C., & Mekhaimer, M. (2016)	x		x	x
Lee, J. H., Lin, S. Y., Lee, W. C., & Tsao, C. Y. (2006)			x	x
Lee, K. H., Sapriza, H., & Wu, Y. (2016).	x	x	x	x
Rehman, I. U., Mahdzan, N. S., & Zainudin, R. (2016).		x		
Saad, M. M., & Darrat, A. F. (2007).			x	x
Ahn, Cai, Chan & Hamao (2007)				x
Chung, K. H., & Van Ness, R. A. (2001)				x
Gresse (2011)				x
Gopalan, Kadan & Pevzner (2012)	x			x
Heflin, F. L., Shaw, K. W., & Wild, J. J. (2005)	x			x
Hillert et al, 2016	x			
Gregoriou, 2013	x			
Lesmond, 2005	x	x		

Sobre la base de las numerosas investigaciones aquí consideradas, los resultados revelan que, en líneas generales, los determinantes de la liquidez accionaria puede clasificarse en cuatro grupos bien diferenciados entre sí: (i) aspectos idiosincráticos como los fundamentales y la gestión interna de las firmas; (ii) variables macroeconómicas, institucionales y legales dentro de los países, (iii) factores no diversificables y estacionales y (iv) aspectos de la negociación, i.e., mecanismos y variables representativas de la operación de cada firma en los mercados y otras características institucionales, regulatorias y estructurales.

Como fue analizado por Stoll (2000), cerca del 40% de la varianza de los spreads de liquidez en los mercados – tanto en análisis transversales como de serie- se encuentran explicadas por variables relacionadas con las características de las firmas y su gestión. Dentro de este conjunto, De Cesari et al. (2011) pone en evidencia cómo la recompra o venta de acciones propias incrementa la liquidez en los mercados de un conjunto de compañías italianas, en especial de aquellos papeles más ilíquidos. En este marco, la estrategia de recomprar acciones propias cuando un papel está subvaluado o emitir cuando está sobrevalorado permite disminuir los costos de transacción por información asimétrica- prima por iliquidez- y por tanto el costo del patrimonio, incrementando así el valor de la firma y mejorando la asignación eficiente de recursos a su interior (análogamente Hillert, A., Maug, E., & Obernberger, S. (2016) encuentran que este efecto se acrecienta aún más en períodos de crisis para una muestra de 6500 en USA).

Asimismo, dentro de esta perspectiva, Charoenwong et al. (2011) señala que aquellas firmas con mayor ratio de activos líquidos sobre activos totales ganan en liquidez

accionaria en tanto que acarrear menores costos de información por “incertidumbre de valuación”, incluso por encima del efecto “incertidumbre de inversión” planteado por Myers & Rajan (1998). Más específicamente, una mayor liquidez de los activos reduce la asimetría de información entre la firma y los inversores, que no tienen acceso privilegiado a los datos de la compañía para realizar una correcta valuación si su estructura presenta una gran proporción de activos fijos. Contrariamente, la hipótesis de “incertidumbre de inversión” anticipa una relación negativa entre proporción de activos líquidos y liquidez accionaria, en tanto las compañías exponen una mejor situación para financiar proyectos ilíquidos, al tiempo que el disponible es más factible de ser redistribuido o incluso expropiado, creando así incertidumbre sobre los flujos de fondos futuros. Basando en la hipótesis de “incertidumbre de valuación”, Jain et al. (2016) evidencia el impacto negativo que genera la mayor proporción de activos intangibles y gastos en investigación y desarrollo, ello por la dificultad de predecir los flujos esperados. Por su parte, y también abonando con la hipótesis previa, Gopalan et al. (2012) han demostrado que la relación positiva entre liquidez de activos y acciones se presenta con mayor frecuencia en firmas con menos oportunidades de crecimiento, restricciones financieras y malas condiciones de gestión interna. Asimismo, otros ratios de liquidez de gestión como los pasivos de corto plazo sobre activos (Choi & Cook, 2006), volumen y crecimiento de activos, entre otros, contribuyen a determinar el nivel de liquidez accionaria (véase también Chordia et al., 2004).

Por su parte, estudios como los de Jain, P. et al (2016) y Tang, K., & Wang, C. (2011) evidencian el impacto positivo de las mejores prácticas de gobernanza corporativa sobre la reducción de la información asimétrica y por tanto la mejora de la liquidez. En una investigación para las firmas del NYSE, AMEX y NASDAQ entre 1983-2013, Huang & Chang (2015) remarcan que la gobernanza debe entenderse en interacción con la calidad y el riesgo de la información, y remarcan el canal a través del cual los fundamentales impactan en la liquidez, como a continuación

*“Financial fundamentals reflect the firm’s operations and thus they can affect financial environment of the stock. To some extent that financial fundamentals can reduce information asymmetry in the marketplace, such information can improve liquidity.”* Huang & Chang (2015, p. 9)

Al respecto del conjunto de variables fundamentales habitualmente más estudiadas, la proporción valor mercado a valor libro (M/B) refleja la relación entre el valor económico de la firma y el valor histórico de sus activos, por lo que se ve afectada por las expectativas de los inversores. Desde este punto de vista, representa el valor razonable del ROE (resultado sobre patrimonio) esperado por los accionistas. Para algunos estudios, un menor M/B brinda dudas sobre el resultado futuro de la compañía, así como una señal de financiar el resultado operativo a través de deuda más que con recursos propios. El apalancamiento, por su parte, se asocia al perfil de riesgo del patrimonio. Por su parte, las firmas más grandes reducen la información asimétrica asociada a las oportunidades de crecimiento, y por lo tanto el riesgo. (Huang & Chang, 2015).

Desde un enfoque diferente, Lipson y Mortal (2005) señalan que indicadores ya mencionados como el tamaño de la firma o el valor libro a mercado, deben evaluarse en conjunción con el nivel de apalancamiento, en tanto que e.g. un valor libro más alto con

un incremento de la deuda podría reflejar un crecimiento con mayor rentabilidad para el accionista (Lipson & Mortal, 2005). En esta línea, Goyenko et al. (2009) plantea que aumentar el apalancamiento podría mejorar la liquidez debido al menor costo de agencia entre los gerentes y los accionistas, en tanto los primeros deben tomar mejores decisiones de inversión cuando cargan con un nivel de deuda elevado, contradiciendo la evidencia usual que asocia deuda con riesgo y por tanto menor liquidez.

Adicionalmente, han sido analizados otros factores asociados a la rentabilidad de la firma como, e.g., el ROA o rentabilidad sobre los activos (Fang et al., 2009; Al Salamat, 2016), que presenta una relación positiva y significativa con la liquidez accionaria según Alnaif (2014) aunque un vínculo contrario para un conjunto de acciones emergentes evaluadas por Al Salamat (2016). Aún más, los ingresos por acción -o EPS- (Brennan & Subrahmanyam, 1996) y la política de dividendos (Banerjee et al., 2007) han demostrado tener una injerencia para explicar la variabilidad de la liquidez. Al respecto de este último factor idiosincrático, de acuerdo con Griffin (2010), las acciones que pagan dividendos satisfacen mejor la necesidad de liquidez de los inversores, volviéndose más atractivas en especial si se negocian con poca frecuencia, donde la falta de un comprador puede desalentar las estrategias de inversión más seguras. Sin embargo, esto no pareciera verificarse para las firmas cuyas acciones son más líquidas, que al tener un menor costo de capital tienden a financiar más cantidad de proyectos con valor presente positivo, lo cual se ve desalentado por la distribución de dividendos. Visto de otro modo, el pago de dividendos disminuye el potencial de la compañía para invertir en proyecto que potencien su rentabilidad futura. Por su parte, indicadores financieros de rentabilidad como el ratio precio-beneficio (i.e., PE, por sus siglas en inglés), también presentan relación con la liquidez. En este sentido, puede esperarse que acciones con un menor indicador se perciban como actualmente más baratas generando así expectativas de una mayor rentabilidad futura, y volviéndose más atractivas para los inversores (Liem & Basana, 2012).

No obstante, en la medida en que los factores idiosincráticos juegan un papel crucial en la determinación de la liquidez accionaria, los aspectos macroeconómicos también han sido estudiados desde una óptica similar. En este marco, Al Salamat (2016) identifica que para las firmas industriales que cotizaron en el Amman Stock Exchange durante el período 2007-2014, además de aspectos fundamentales, la tasa de interés y el crecimiento del producto bruto interno tienen un efecto positivo sobre la liquidez mientras que la inflación contribuye negativamente. Rehman et al. (2016), desde un mismo enfoque, determinan que variables como la base monetaria, tasa de interés, el ciclo de negocios y el flujo de inversiones extranjeras tienen un impacto rezagado y significativo en el corto plazo, aunque en el largo plazo son la producción industrial y el flujo neto de divisas los que resultan significativos. Dentro de este conjunto, las tasas de corto plazo se vuelven un factor relevante asociado al costo de financiamiento de los inventarios de los *market makers* y los *traders* durante la jornada de negociación. Como señalan Dicle (2008), en períodos donde los portafolios están desbalanceados el riesgo de financiar las operaciones se vuelve más elevado en períodos de tasas crecientes.

*“For periods with unbalanced inventories, dealers need to finance the cost of inventory. If these periods coincide with high periods of short-term interest rates, the risk of inventory cost becomes costlier (...). As the short-term financing costs increase, cost of*

*carrying an inventory will increase which will lead to lower liquidity in the market. Based on the cost of carrying inventory, the risk of loss will increase with the increased financing costs. This will lead to more caution against shares that already have high sensitivities to market liquidity (high liquidity betas).” Dicle (2008, p. 37)*

Cabe considerar, por otra parte, factores tales como el rating de deuda soberana. En tal sentido, Lee et al. (2016) analiza el efecto sobre la liquidez accionaria en 40 países durante un decenio, demostrando la injerencia de los cambios en los ratings soberanos de deuda. En efecto, una reducción impacta negativa y no linealmente en la liquidez como consecuencia de: (i) liquidación de posiciones por pérdidas en colaterales, (ii) rebalanceo de portafolio y (iii) un mayor grado de selección adversa.

Como se ha mencionado al inicio de la subsección, existen factores no diversificables o sistémicos que tienen impacto en la liquidez. En tal sentido, tanto la liquidez como la volatilidad del mercado (e.g. índice VIX, véase Huang & Chang, 2015) constituyen determinantes a nivel individual (Chordia et al., 2000, Hasbrouck y Seppi, 2001). Estos efectos, conocidos por “*commonality*” han sido luego incorporados en gran parte de los estudios de esta índole (Karolyi et al., 2012). En este marco, Coughenour, J. F., & Saad, M. M. (2004) demuestran como la liquidez individual se ve afectada por la variación de la liquidez del portafolio de mercado, y más precisamente Lee et al. (2006) identifican la presencia de covariación con la liquidez de la industria, en especial en las firmas más pequeñas.

Por su parte, y como medida clásica de riesgo no diversificable, Chan et al. (2013) determina que a mayor beta de una acción- i.e., mayor volatilidad sistémica-más liquidez. En efecto, y a diferencia de lo expuesto por Dicle (2008), como la información de mercado es más fácilmente observable que la específica de la acción, podría argumentarse que cuando una acción tiene una gran correlación con el mercado, el movimiento esperado del papel es más claro y por lo tanto menos sensible a órdenes particulares, por tanto, fluye un mayor volumen hacia este tipo de papeles.

*“A key determinant of liquidity is the volatility of underlying stock. An increase in volatility of underlying stock returns implies that the liquidity providers will face higher adverse selection risk due to an increased possibility of trading with informed investors, as well as higher inventory risk arising from order imbalances. (..), there is little research on the separate effects of systematic volatility and idiosyncratic volatility on liquidity. According to the adverse selection models or inventory risk models, the effect of systematic volatility on liquidity should be different from that on idiosyncratic volatility, given that systematic volatility can be hedged to a certain extent” Chan et al. (2013, p. 415).*

Por otra parte, es requerido analizar el efecto de algunas tendencias de negociación, e.g., a nivel semanal, anual o por período de operación con mayor intensidad. En este marco, es usual reconocer en estudios variables categóricas que contribuyen a captar el efecto “día de la semana”, “mes del año”, “vacaciones”, momentos de crisis, eventos de anuncios macroeconómicos relevantes, entre otros (Chordia et al., 2004). En referencia a ello, los spreads de liquidez en las acciones son sensibles al efecto “meses del año”,

siendo más elevados en Enero y Febrero, al igual que durante los períodos de crisis, según el estudio de Chordia, Sakar & Subrahmanyam, (2004) para las acciones del NYSE durante el período 1993-1998.

Finalmente, y en cuanto a los mecanismos y características de la negociación, regulación y estructura al interior de los mercados, diversos estudios asocian una mayor volatilidad de los retornos de la firma con un impacto negativo en liquidez por los mayores costos de inventario (Stoll, 2000). Asimismo, también ha sido utilizado el precio de cierre en tanto precios más bajos suelen ser más riesgosos (Jain, 2003). En el mismo sentido, un mayor tamaño promedio por operación implica más liquidez en tanto conllevan menores costos de inventario y procesamiento asociado a la posibilidad de encontrar contrapartes con mayor frecuencia en el mercado.

En lo que refiere particularmente a la regulación y estructura de negociación al interior de los mercados, uno de los pocos análisis integrales realizado por Jain (2003) pone de relieve cómo las características del diseño dentro de 51 mercado accionarios afectan su nivel de liquidez, dentro de las cuales se distinguen: (i) forma de negociación (mercados de *dealers* basados en cotizaciones, libros de ordenes limite y mecanismos híbridos);(ii) presencia de *market makers*; que proveen liquidez cuando hay insuficiencia o falla de sincronización entre órdenes(iii) incremento del tamaño del *tick* (variación mínima del precio), que suele reducir los spreads (véase e.g. Chakravarty, Harris, & Wood, 2001);(iv) libros de órdenes límite consolidados; (v) estructura de mercado fragmentada, que puede ocasionar efectos contrapuestos, i.e., por un lado mejorar la liquidez al aumentar el número de dealers que compiten y por lo tanto disminuir los costos de transacción o por otro fragmentar el volumen negociado entre mercados empeorando la competencia por precio entre órdenes (véase también Madhavan, 1995);(vi) grado de transparencia del flujo de órdenes ex ante, i.e., los dealers en mercados poco transparentes ponen precios más agresivos durante las primeras rondas de negociación para captar a los operadores menos informados, empeorando la liquidez por selección adversa (asimismo mejorar la transparencia ex post, véase Pagano & Roell, 1996); (vii) ejecución automáticas de operaciones, que disminuye tanto los costos fijos como variables (e.g. procesamiento de órdenes), y (viii) esquema de propiedad del mercado. En lo particular, el autor concluye que aún resultan las variables significativas, la presencia de *market makers* es especialmente relevante en los mercados emergentes.

En relación con otros mecanismos de la negociación, Kuo et al. (2017) al igual que Ahn et al. (2007) para mercados asiáticos, revelan el impacto en la reducción de los spreads efectivos de cambios en el tamaño del tick de las acciones. Desde una perspectiva similar, de acuerdo con el estudio de Chung & Van Ness (2001) la liquidez se reduce significativamente para las acciones del Nasdaq luego de los cambios en la normativa de manejo de órdenes, disminución que resulta mayor durante el mediodía y la última media hora de la negociación cuando además de la evalúa con los cambios en el tamaño del tick. Ko & Lim (2019), por su parte, estudian el efecto de las operaciones de venta corta sobre la liquidez, empleando como variable de interés el cambio porcentual de las tasas de interés de corto plazo.

En este orden de ideas, también ha resultado de interés para la literatura especializada el uso de datos de alta frecuencia para analizar la duración de la transacción en relación al spread de la operación (véase Li & Qi ,2013 para el mercado chino), o bien, focalizando en el comportamiento de la covariación de liquidez con el mercado y la industria, el

estudio de dichos movimientos durante la primera y última media hora de la negociación, tal como Saad & Darrat (2007) encontraron para 1200 acciones del NYSE un patrón de curva de liquidez diario en forma de U, ello como resultado de los mayores costos asociados a información asimétrica e inventario en esos intervalos temporales:

*“Adverse selection costs tend to be high at the start (particularly within the first half-hour) of trading as investors and market makers lack sufficient knowledge of the order flow and also because such costs tend to increase during non-trading sessions. Inventory costs are relatively higher near the close (last half-hour) of trading as market makers are left with less time to correct for sub-optimal inventory positions.”* (2007, p. 913).

### 3.- METODOLOGÍA

#### 3.1- Relación entre liquidez y retornos para grupos de acciones específicos agrupados por distribuciones de liquidez intra diaria

##### 3.1. a.- Ecuación general de retornos para determinar un índice de liquidez representativo por clúster de acciones

El presente trabajo evaluará la relación entre la liquidez y los excesos de retornos para el universo completo de acciones que cotizan en el mercado argentino, basándose para ello en la combinación de técnicas de datos de panel y algoritmos de *machine learning* para una óptima selección de variables y la construcción de los índices de liquidez por clúster. Tanto la variable dependiente (retornos) como los regresores serán analizados sobre una base trimestral, en pesos argentinos y construidos para todas las acciones (del panel general del Merval y BYMA) que hayan cotizado para el período 2009q4-2019q2<sup>5</sup>, a partir de datos diarios (llevado a base trimestral) y trimestrales (para datos de balance). Los datos han sido recopilados de la plataforma financiera Thomson Reuters Eikon, así como la clasificación sectorial de las acciones (véase a continuación) basada en el nomenclador TRBC provisto por el mismo servicio (sector económico e industria).

Sobre la base de los aspectos metodológicos de estimación desarrollados en profundidad en las secciones subsiguientes y tomando como base un modelo para retornos (exceso) en mercado accionarios de Latinoamérica adaptado al CAPM multifactorial<sup>6</sup>, la estructura de la ecuación general se define como sigue:

$$(1) \quad ER_{i,t}^e = \alpha + \sum_{j=1}^k \beta_j liq_{*i,t} + \sum_{j=1}^k \gamma_j X_{j,i,t} + \vartheta ER_{m,i,t} + \omega ADR_{i,t} + \sum_{j=\underline{j}}^j \theta_j Year_{j,t} + \sum_{j=\underline{j}}^j \delta_j Sector_{j,t} + \varphi \sigma Fx_{j,t} + \varepsilon_{i,t}$$

$i = 1, \dots, N$ (Unidades de corte transversal)

<sup>5</sup>El corte muestral se realizó en Diciembre 2019, siendo para entonces el último trimestre de información fundamental completa para todo el set de acciones el segundo trimestre de 2019. Asimismo, los datos de frecuencia diaria estaban disponibles a la fecha de corte hasta 2009q4.

<sup>6</sup> véase Panigo, Grandes y Pasquini (2007).

$t = 1, \dots, T$  (Observaciones de serie de tiempo)

, donde  $ER_{i,t}^e$  es la variable dependiente que representa los retornos (exceso) para la firma  $i$  en el período  $t$ ,  $liq_{i,t}$  es el set de  $k=1, \dots, n$  medidas de liquidez (de datos diarios) para la firma  $i$  en  $t$ ,  $X_{j,i,t}$  desde  $X_{1,i,t} \dots X_{k,i,t}$  corresponde al set de  $k=n$  variables de control específicas de la firma,  $ER_{m,i,t}$  es el retorno del mercado (exceso) que busca captar la parte de la varianza de la variable dependiente explicada por el riesgo sistémico (no idiosincrático),  $ADR_{i,t}$  una variable *dummy* que identifica la tenencia de ADR por parte de la firma  $i$  (las firmas que cotizan en mercados del exterior se suponen más diversificadas, integradas, y con más posibilidades de ser negociadas),  $Year_{j,t}$  es una variable *dummy* que representa los años y permite captar efectos del ciclo económico,  $Sector_{j,t}$  una variable *dummy* que buscan captar el efecto diferencial por sector sobre el exceso de retornos no explicado por factores idiosincráticos (se utilizará la clasificación de Thompson Reuters),  $\sigma Fx_{j,t}$  es el tipo de cambio ARS/USD y pretende corregir los efectos de la pérdida de valor de la moneda sobre el comportamiento de los retornos que no hayan sido capturados por los movimientos del índice del mercado ni las heterogeneidades estructurales (representa un efecto sustitución de portafolio) y  $\varepsilon_{i,t}$  un componente aleatorio con media cero y varianza  $\sigma_\varepsilon$ . Por su parte,  $\alpha$  denota el efecto no observado y fijo específicos de la firma (no incluido en el *panel regression*),  $\beta_j$  son los  $\beta_1 \dots \beta_n$  los coeficientes de regresión que determinan el impacto de cada medida de liquidez sobre los retornos (exceso) esperados,  $\vartheta$  el factor sistémico de los retornos,  $\gamma_j$  de  $\gamma_1 \dots \gamma_n$  los estimadores de los regresores independientes de control,  $\varphi$  el coeficiente de regresión del tipo de cambio, y  $\omega \theta_j \delta_j$  estimadores para las variables independientes *dummy*.

En la tabla a continuación se detalla el set de variables incluidas en el modelo (1), regresores que son comúnmente empleados en la literatura (véase subsección 2.1 y 2.2.a). Dentro de este marco, y además de las variables arriba descritas, la inclusión de un conjunto de determinantes fundamentales se hizo siguiendo el estudio de Panigo, Grandes y Pasquini (2007) para países latinoamericanos (incluyendo a Argentina), que precisamente incorpora a los regresores estudiados en la literatura y que aquí han sido analizado en la sección del marco teórico. En atención a esta perspectiva, el tamaño de la firma sugiere que compañías más grandes son menos riesgosas y por tanto exigen menos retornos diferenciales; un bajo ratio precio de mercado a valor libro implica un activo subvaluado y que por lo tanto para recomponer su valor exigirá mayores retornos; el apalancamiento refleja la capacidad de la firma para enfrentar sus deudas corrientes (o de corto plazo) y por lo tanto su riesgo de desvalorización de su patrimonio – por el default-, el ROA y la volatilidad del ROA son indicadores financieros que reflejan el crecimiento genuino de la compañía y su estabilidad (y se traducen en riesgo y retornos diferenciales); y el circulante una medida que se aproxima al *floating* disponible<sup>7</sup> - y refleja el riesgo asociado a la operación en los mercados (siguiendo lo propuesto por Wong et al, 2006).

En relación con las medidas de liquidez empleadas, debe señalarse que fueron construidas aquellas basadas en datos de frecuencia diaria o cuya aproximación fuera posible a partir de los datos disponibles. En atención a esto último, se incorporaron las

---

<sup>7</sup> No fue posible acceder a información sobre el circulante en flotación, i.e., efectivamente disponible para la negociación, a fin de construir la proporción del disponible/circulante.

medidas de spread cotizados y efectivo (véase nota de referencia 9), no obstante, se desestimó el spread realizado. En efecto, y a pesar de la aproximación propuesta por Jain (2013), este indicador busca capturar el spread efectivo neto del costo asociado a movimientos desventajosos en los precios (de un operador que vendió a un precio y compró previamente a otro más elevado) que ocurren como resultado de costos de procesamiento, inventario y selección adversa y que es esperable se capturen en un intervalo de tiempo que es representativo del nivel de transparencia del mercado, como fue mencionado en la sección teórica. Debido a las volatilidades no anticipadas al interior del mercado local y la intermitencia en las operaciones no puede considerarse que una jornada sea una frecuencia lo suficientemente representativa para capturar dicho efecto.

**Tabla 1.- Variables del Modelo de retornos (dependiente y regresores exógenos)<sup>8</sup>**

Identificador	Descripción
retornos	Retornos de cada acción
spread_cotizado_medio	Spread cotizado relativo (punto medio) <sup>9</sup>
spread_efectivo_medio	Spread Efectivo relativo (punto medio)
volumen_m	Volumen operado
rotacion	Nominales operados por cada 1000 títulos en circulación
amihud_rotacion	Medida de Amihud ajustada por rotación
amihud	Medida de Amihud
d_volpositivo	Proporción de días con volumen positivo
d_retornocero	Proporción de días con retorno cero
apalancamiento	Deuda total/ activos corrientes
tamaño_negocio_circulante	Logaritmo natural de los Activos totales Acciones en circulación
ROA	Beneficio neto sobre activos
mv_bv	Precio de Mercado / Valor Libro
ROA_desvio	Desvío Estándar 3q ROA
FX	Tipo de Cambio Nominal (en pesos)
indice_global	Índice de Mercado (total firmas)
_lyear_*	Dummy Año
_ldummy_sec_*	Dummy Sector
ADR	Dummy ADR

<sup>8</sup> Tanto los retornos como las medidas de liquidez fueron construidos en base a datos diarios y luego adaptadas a la base trimestral en que se presenta la información de balance. Los retornos del trimestre se analizan contra las variables de liquidez promedio y los datos de balance y el FX de cierre trimestral conforme la naturaleza de la información contable.

<sup>9</sup> Debido a que únicamente se disponen de datos de frecuencia diaria para construir las medidas trimestrales, para construir el spread, se utilizó el precio más alto (Ask) y el más bajo de la jornada (Bid) como aproximaciones del Ask y el Bid.

### 3.1. b.- Indicadores de liquidez relevantes según el perfil de la inversión: modelización por clúster

La imposibilidad de aproximar la liquidez a través de una única medida capaz de captar todas las dimensiones del fenómeno ya ha sido tratada en estudios precedentes. En este sentido, trabajos como los de Lam, Tam y Dong (2019) han propuesto la construcción de una única medida de liquidez a partir de técnicas como el Análisis de Componentes Principales (PCA) que capte las diferentes dimensiones, i.e., cantidad negociada, frecuencia de operación, el impacto precio y los costos de transacción.

Sin embargo, estas dimensiones pueden resultar más o menos relevantes para establecer su relación con el exceso de retornos de acciones que representan diferentes perfiles de la inversión. En efecto, la forma funcional delineada en la subsección precedente puede establecer promedios que no sean representativos de las verdaderas razones entre variables para grupos específicos de acciones. En otros términos, un conjunto de acciones más riesgoso probablemente encuentre más representativo para explicar la variabilidad de los retornos a ciertos factores idiosincráticos que otro grupo menos riesgoso. Asimismo, si bien no resulta del todo factible asociar grupos de acciones similares en su patrón de liquidez intradiaria con grupos de inversores, ya que un mismo inversor puede dirigirse a diferentes perfiles de riesgo de inversión como estrategia de diversificación de portafolio, es posible argumentar que probablemente inversores con un capital pequeño busquen minimizar el costo de entrada y salida (spread bid-ask) a partir de una mayor profundidad del libro de órdenes e inversores más grandes tengan capacidad de relegar costos e invertir en acciones menos líquidas con un perfil más riesgoso en busca de oportunidades.

En este marco de hipótesis, la presente subsección tiene por propósito reproducir la ecuación general más robusta para distintos grupos de acciones clasificadas en base a datos de liquidez con información intradiaria. Se procederá a agrupar acciones mediante el método de *distribution based clustering*, i.e., cada clúster incluirá las acciones que más se aproximen en la distribución intra diaria. Para ello se procederá en dos pasos: (1) *data dimensionality reduction* y (2) *clustering*. En primera instancia se reducirán las observaciones por acción para el spread bid-ask relativo a los cuatro momentos de la distribución (i.e. media, desvío estándar, simetría y curtosis), luego estandarizando para cada acción en relación con el universo completo. Una vez ello, se segmentará a partir de dichos parámetros tomando una medida de distancia eucladiana cuadrática entre distribuciones.

Al respecto de la variable a observar, es importante remarcar que relativizar el spread (al precio efectivo de la operación) permite ajustar la medida por el precio de la transacción, que de otro modo podría distorsionar el análisis (e.g., un spread de 3 pesos en una acción de 700\$ no implica lo mismo que en un papel que vale 50\$). Por su parte, el volumen no es útil para comparar ya que pueden existir grandes operaciones en acciones poco líquidas, a pesar de que haya poca profundidad en su libro de órdenes y por lo tanto poca liquidez.

Al respecto del primer paso, es necesario generar aproximaciones precisas a las verdaderas distribuciones, reduciendo el universo completo de datos intra diarios a una muestra que permita su comparabilidad en tiempos eficientes. En tal sentido, la medida

de distancia Wasserstein- Mallows (véase Ni, Bresson, Chan & Esedoglu, 2009) permite tal aproximación, al reducir y comparar distribuciones a partir de en sus cuatro momentos. Tomando las funciones de distribución empíricas para la variable de interés es posible comparar las distribuciones acumuladas inversas  $F^{-1}(t) = Y$  de dos histogramas cualesquiera a partir de esta métrica, que resulta de la familia de las medidas de distancia euclidianas aplicada a las distribuciones de probabilidad. En este sentido, se puede definir para  $p = 1$  y  $p = 2$  la distancia de Wasserstein-Mallows como:

$$dW^p(H_i, H_j) = \sqrt[p]{\int_0^1 |F_i^{-1}(t) - F_j^{-1}(t)|^p dt}$$

Las distribuciones son comparadas a partir de los diferentes momentos de la distribución, i.e., ubicación, escala, y forma (Rueshendorff, 2001). La descomposición de la  $dW^p$  con  $p = 2$  para dos distribuciones resulta en:

$$dW^2(Y_i, Y_j) = (\mu_i - \mu_j)^2 + (\sigma_i - \sigma_j)^2 + 2 \sigma_i \sigma_j (1 - \rho(Y_i, Y_j))$$

Debido a la gran cantidad de datos, es necesario comparar la distancia entre dos distribuciones reduciendo los parámetros de la forma al coeficiente de asimetría y curtosis, empleando entonces para ello tan solo cuatro observaciones para cada acción (siendo que el coeficiente de correlación debe emplear todos los datos intradiarios resultaría extremadamente engorroso al cálculo), representando así los cuatro momentos de la distribución (media, desvío estándar, simetría y curtosis). Una vez realizada la etapa de *data reduction*, el número de clústeres óptimos será el que minimice las distancias intra-clúster y maximice las mismas distancias inter-clúster.

En lo que refiere al procedimiento de *clustering*, la rama de las matemáticas conocida como “Geometría de la Información” (encontró cúspide con la publicación de Amari & Nagaoka, 2000) es precisamente el campo que ofrece una variedad de métodos para lidiar con este tipo de problemas, trasladando conceptos de la geometría diferencial al campo de la teoría de la probabilidad (distribuciones). En este marco, los algoritmos paramétricos han tomado mucha preponderancia gracias a su simplicidad y escalabilidad (Banerjee et al., 2015). Algoritmos para clustering de histogramas paramétricos del estilo del K-Means permiten precisamente agrupar desde datos a unidades dentro de grupos a partir de los atributos de la distribución de probabilidad de una variable aleatoria representativa (media, mediana, moda, varianza, simetría y curtosis). Extensiones como el modelo de Bregman K-Means emplean divergencias asociadas a distribuciones como el cuadrado de la distancia Eucladiana o euclídea (la distancia más básica) sobre la que se basa la función de probabilidad de Gauss o distribución normal (véase Banerjee et al., 2015 para una aplicación). Otros algoritmos desarrollados por la literatura han sido, e.g., el paramétrico de Linde-Buzo-Gray basado en la distancia de Itakura-Saito (Linde et al., 1980), o desde un punto de vista no paramétrico el EP-Means de Henderson et al. (2015).

La presente investigación utilizará el método de agrupamiento K-means (K-medias), que permite el agrupamiento de un conjunto de observaciones en el que cada una de ellas pertenece al clúster cuyo valor medio de cada parámetro (conjuntamente) más se aproxime. Siguiendo a Makles (2012), se computarán diez soluciones - para un número máximo establecido de  $k$  ( $k=1 \dots 10$ ) grupos (resulta lo máximo posible para el número

de firmas con que se cuenta)-por cada métrica empleada y luego se compararán. Para robustecer los resultados, estas soluciones se calculan a partir de puntos de partida aleatorios (50), para luego promediarse los valores obtenidos. Para elegir el número de clústeres óptimos dentro de las  $k$  soluciones posibles, se grafican cada una de las métricas y se busca a qué cantidad de clústeres corresponde el quiebre más pronunciado de la pendiente de la curva, ello para el caso de la medida Índice WSS (*Within Cluster Sum of Squares*) y el coeficiente Eta Cuadrado  $\eta^2$ . Así mismo, se calcula el coeficiente de Reducción Proporcional del Error (PRE), para cuyo caso se observa el punto más alto. A modo de robustecer el análisis también se evaluará la prueba de clustering Calinski-Harabasz F-test.

Más específicamente, y a modo de ejemplo para el WSS, se procede en cada caso de modo similar:

1. Se computa el algoritmo de clustering utilizando los cuatro momentos de la distribución (en este caso el K-means calcula la distancia eucladiana cuadrática multivariada) para valores de  $k=1\dots 10$  ( $k=3$ , e.g., representa que la muestra se segmenta en 3 grupos).
2. Para cada  $k$ , y partiendo de puntos de partida aleatorios  $j$ , se calcula la métrica, en este caso, la suma de las distancias al cuadrado entre cada punto (observación) del clúster y el centroide (distancia eucladiana cuadrática) para los cuatro momentos, luego promediando para cada  $k$  el resultado obtenido para cada  $j=1 \dots 50$ . El objetivo detrás del uso de métodos particionales como el empleado aquí radica en definir el número de clústeres óptimo tal que la varianza total intra-clúster sea la mínima posible. Al mismo tiempo, que las distancias inter-clúster sean las máximas posibles.
3. Se grafica la curva WSS para los  $k$  clústeres.
4. Observando la gráfica de la función, se debe elegir aquel número de clústeres tal que añadiendo otro grupo ello no minimice significativamente más la WSS. En otros términos, el  $k$  óptimo coincide con el máximo quiebre de la función.

Para el caso del coeficiente Eta Cuadrado  $\eta^2$ , este cuantifica la reducción proporcional del WSS por cada  $k$  comparada con la suma total de los cuadrados (TSS), mientras que el coeficiente de Reducción Proporcional del Error (PRE) muestra la reducción proporcional de WSS para la solución del clúster  $k$  en comparación con  $k-1$ . Por último, el índice de Calinski-Harabasz evalúa la validez de un  $k$  a partir del promedio de la suma de cuadrados inter e intra-clúster. Como método heurístico de comparación, se escoge la solución que se observa genera un quiebre en la serie de valores. La interpretabilidad y consistencia combinada de todas las medidas permitirá encontrar el número de  $k$  óptimos que minimicen las distancias (euclidianas) intra-clúster y maximicen las inter-clúster.

No obstante, en mercados poco líquidos (como en el caso de Argentina) y muy volátiles, se requiere profundizar aún más el análisis a los efectos de no obtener una segmentación sesgada por la utilización de un solo corte muestral. En tal sentido, y a fin de dar estabilidad a los resultados, se repetirá el proceso para 3 cortes muestrales sucesivos, evaluando que el  $K$  óptimo sea estable en el tiempo.

### 3.1. c.- *Dimensionality reduction technique* y formas funcionales flexibles para la selección del mejor modelo por clúster

En primer lugar, sólo mediante el uso de datos de panel es posible observar a las mismas unidades o grupos de unidades en diferentes momentos del tiempo, requerido para nuestro modelo de retornos (exceso). En lo que respecta a regresiones sobre este tipo de datos, es importante remarcar que los coeficientes obtenidos dependen principalmente de las características propias de los datos empleados y la forma funcional de la ecuación a analizar. En relación con lo primero, el empleo de un conjunto de variables de control (respecto de nuestra variable de interés liquidez) permite reducir el sesgo derivado de la presencia de variables omitidas, pero no eliminarlo por completo. Como es bien sabido, no es posible suponer que las observaciones dentro de un panel sean necesariamente independientes, i.e., la presencia de factores omitidos que afectan a una variable (pero no a otra) en un momento del tiempo también podrán hacerlo al siguiente. No obstante, como refiere Baltagi (1995), es posible mejorar la consistencia de las estimaciones suponiendo que esta heterogeneidad individual no observable deriva fundamentalmente de efectos individuales específicos por unidad de corte transversal, aunque constantes en el tiempo (véase también Elosegui et al., 2006).

No obstante, la inclusión conjunta de todas las variables de liquidez disponibles además de las variables de control potencialmente relevantes no necesariamente conlleva a la mejor estimación del verdadero proceso generador de los datos. Es bien sabido que existe al respecto un *trade-off* entre sesgo y varianza, más comúnmente asociado a la problemática de *overfitting* o sobreajuste. Asimismo, no resulta teóricamente evidente que la relación entre las variables examinadas sea estrictamente lineal ni contemporánea. Aunque tal supuesto pueda traducirse en una buena aproximación en ciertos casos, en muchos otros genera relaciones significativamente sesgadas.

Para el primer problema, el método de abordaje más sencillo consiste en utilizar el *principal component analysis* (PCA por sus siglas, ver Lam & Chan, 2019). El problema con este método radica en la escasa capacidad de interpretación económica de los componentes obtenidos, que combinan distintas proporciones de varianza de las variables subyacentes. Para el segundo problema, por su parte, la aproximación más difundida recae en la utilización de redes neuronales artificiales (muy difundidos entre los algoritmos de *machine learning*) con distintos *layers* (capas) complementarios que incluyan funciones de activación no lineales diversas. Al igual que con PCA, la utilización de este tipo de algoritmos suele ser particularmente eficiente para obtener predicciones robustas, pero mucho más deficientes para su utilización en análisis de inferencia causal, ello debido a la compleja interpretación de las relaciones subyacentes.

Por los motivos previamente mencionados, en la presente investigación utilizaremos un enfoque alternativo, que combina la potencialidad de los algoritmos de *machine learning* con la interpretabilidad de econometría, ello a los efectos de encontrar una forma funcional para la ecuación de exceso de retornos por clúster descrita en 3.1., que combine buen poder predictivo y niveles de interpretabilidad robustos y asequibles para los resultados de inferencia causal. Esta metodología se basa en cuatro etapas complementarias:

1. *Feature extraction*: Generación de nuevas variables explicativas a partir de las variables originales, construyendo para cada una de ellas los primeros rezagos. Si bien la generación de otras transformaciones como cuadrados, logaritmos e inversas puede aumentar el poder predictivo, su incorporación incrementa notablemente los tiempos de estimación además de aportar una mayor complejidad a la interpretabilidad financiera.
2. *Regularization* (preselección): Reducción preliminar del espacio de variables relevantes a través del algoritmo LASSO (*least absolute shrinkage & selection operator*) - (Tibshirani; 1996)-. Este algoritmo, si bien no garantiza la identificación del verdadero proceso generador de los datos, es una herramienta flexible y poderosa para eliminar de manera apropiada las variables menos relevantes para la ecuación examinada. El algoritmo LASSO permite reducir la sensibilidad de los parámetros de la regresión a la multicolinealidad mediante el uso de una función penalizadora que permite mantener su valor por debajo de uno preespecificado. Más específicamente, minimiza la suma de cuadrados de los residuos sujeto a que la suma del valor absoluto de los coeficientes sea menor que una constante, generando que algunos coeficientes sean exactamente 0 (Tibshirani, 1996). En lo particular, la presente investigación reduce el espacio a 15 variables para una estimación posterior más eficiente.
3. *All-subset-regression*: Tomando como insumo los resultados de la preselección vía LASSO, el algoritmo econométrico GSREG (véase Gluzmann y Panigo, 2013) permite obtener el mejor modelo explicativo final (para la muestra completa, dejando afuera cierta cantidad de observaciones para la confirmación *out-of-sample*) sobre la base de diversos criterios informativos (Akaike, Bayesiano, Mallows, etc.) y test de residuos (para poder garantizar inferencia estadística). A partir del análisis exhaustivo de todas las combinaciones posibles de variables explicativas, este algoritmo selecciona como modelo final aquel que sólo incluye las variables que maximizan o minimizan el criterio de información adoptado, según sea.
4. *Optimal Model Validation*: los métodos de *resampling* como el *cross-validation* se encuentran ampliamente difundidos en el campo de la estimación de modelos y selección de variables. Este tipo de técnicas permiten iterar los modelos múltiples veces excluyendo cada vez un set diferente de observaciones de entrenamiento (o de validación) calculando una tasa de error para asegurar la mejor predicción de nuevas observaciones donde no se conoce la variable predicha. Siguiendo a Jung & Hu (2015) se utilizará una versión que permite promediar entre los mejores modelos obtenidos en (3) basado en la técnica de *K-fold validation* y asegurando la adyacencia temporal entre ambos sets (no se emplean muestras aleatorias). Tomando como base la selección mediante GSREG (3), se determinan los K modelos óptimos, que son aquellos con el menor score para el set de validación, i.e., la combinación de variables que minimice el RMSE *out-of-sample*. El modelo final se obtiene a partir de estimar los parámetros como los valores promedio ponderados (por score) entre los K modelos óptimos. A fin de disminuir el trade-off entre sesgo y varianza se emplearán K=10 grupos<sup>10</sup>,

---

<sup>10</sup> A diferencia del LOOCV (*Leave One Out Cross-Validation*) que realiza tantas iteraciones como observaciones disponibles y deja fuera solamente una observación por cada vez, el K-Fold permite disminuir el *overfitting* y mejorar la eficiencia de la estimación al dividir los datos en K grupos del mismo

dejando un set de validación de 8 trimestres. Este procedimiento asegura estimadores más estables y eficientes, como señalan los autores:

*“Due to the averaging effect, the variance of the proposed estimates can be significantly reduced. This new procedure results in more stable and efficient parameter estimation than the classical K-fold cross validation procedure. In addition, we show the asymptotic equivalence between the proposed and classical cross validation procedures in the linear regression setting.”* (2015, p. 1).

Siguiendo esta metodología en cuatro etapas, y para cada clúster, a la ecuación general (1) descrita en la subsección 3.1.a se le adicionan los primeros rezagos tanto de la variable dependiente como de los regresores exógenos:

$$(2) \quad ER_{i,t}^e = \alpha + \sum_{j=1}^k \beta_j liq_{i,t}^* + \sum_{j=1}^k \gamma_j X_{j,i,t} + \vartheta ER_{m_{i,t}} + \omega ADR_{i,t} + \sum_{j=\underline{j}}^j \theta_j Year_{j,t} + \sum_{j=\underline{j}}^j \delta_j Sector_{j,t} + \varphi \sigma Fx_{j,t} + \tau Ln^* + \psi ER_{i,t-1} + \varepsilon_{i,t}$$

$i = 1, \dots, N$  (Unidades de corte transversal)

$t = 1, \dots, T$  (Observaciones de serie de tiempo)

, en donde  $Ln^*$  es cada uno de los regresores rezagados  $n=1$  períodos y  $ER_{i,t-1}$  la variable dependiente rezagada un período que busca captar el efecto *momentum o contrarian*. Por su parte,  $\tau$  representa los estimadores de las variables independientes rezagadas y  $\psi$  el coeficiente de regresión del primer rezago de los retornos.

Una vez obtenidas las variables de liquidez óptimas por cada clúster, cada índice de liquidez será construido en base a los coeficientes obtenidos<sup>11</sup>.

### 3.2- Determinantes de la liquidez: un modelo por clúster multifactorial

A partir de los índices de liquidez construidos para cada clúster es posible avanzar hacia el desarrollo de un modelo de múltiples dimensiones analíticas, apoyado en los cuatro grupos de determinantes de la liquidez estudiados por la literatura y revisados de modo exhaustivo en la sección 2.2.b.

En este marco, el procedimiento de selección de variables óptima sobre la base de la capacidad predictiva de los modelos empleando las técnicas de *machine learning* revisada en la sección precedente 3.1.c debe combinarse con el poder explicativo y la interpretabilidad de los coeficientes propia de un proceso de inferencia causal como el que requiere el modelo de determinantes de la liquidez por clúster. De este modo, además de aplicar los procedimientos ya revisados para la selección de modelos, en necesario apoyarse en las técnicas más avanzadas y adecuadas de *panel data regression* que aseguren que los coeficientes representen los verdaderos impactos sobre la variable predicha. En efecto, el principal efecto que genera estimaciones sesgadas al interior de cada modelo radica en la autocorrelación de los errores dentro de los paneles,

tamaño, empleando  $k-1$  grupos como test. LOOCV constituye un caso particular de  $K$ -fold Cross Validation que conlleva múltiples iteraciones.

<sup>11</sup>Como cada coeficiente representa un efecto separado, se puede realizar una suma producto de los coeficientes de liquidez promedio ponderados resultantes del paso 4 y las respectivas variables.

potenciado fundamentalmente por la evidencia de endogeneidad entre liquidez y retornos (véase Arellano, 1990 y Zellner & Theil, 1962). Para ello, es necesaria una estimación simultánea de los mejores modelos de retornos y liquidez por clúster, y la técnica más empleada para ello es el método de 3SLS (*Three-Stage Least Squares*)<sup>12</sup>. Como señalan Crowe et al. (2005), este método también es robusto a la autocorrelación de los residuos, generando así estimadores insesgados.

*“By taking into account the cross-equation correlation, 3SLS yields more efficient estimates for simultaneous-equation systems than two-stages least squares (2SLS) and single equation ordinary least squares (OLS) while taking care of potential endogeneity issues. In addition, 3SLS has the desirable feature of leaving the autocovariance matrix of errors unrestricted, so that, in contrast to full information maximum likelihood method, 3SLS does not require that the distribution of errors is known. The 3SLS estimates are robust to the residual autocorrelation of an arbitrary form. Hence, 3SLS renders unbiased estimates, in contrast to 2SLS or single-equation OLS, in models with lagged dependent variables”* (2005, p. 42).

Dentro de este marco, asimismo, la afirmación empírica sobre la robustez de los estimadores será contrastada a partir de la prueba general de autocorrelación para datos de panel de Cumby-Huizinga (Cumby and Huizinga, 1990, 1992) y aplicable luego de estimar sistemas de ecuaciones simultáneas. A diferencia de otras pruebas como el de Box-Pierce/ Ljung-Box, Durbin o Breusch-Godfrey<sup>13</sup> que no son apropiados en ciertas circunstancias, como e.g., la presencia de heterocedasticidad condicional, el uso de regresiones con variables instrumentales y la superposición de datos presenten comúnmente en la información financiera, este comando de post - estimación resulta más flexible, suponiendo que los errores siguen un proceso MA(q). En el caso de su aplicación para modelos de efectos fijos en datos de panel, los autores remarcar que esta prueba es análoga a la desarrollada por Arellano y Bond (1991).

A partir de metodología expuesto, el modelo de determinantes de liquidez general a estimar para cada clúster se estructura como sigue:

$$(3) \quad IL_{i,t} = \alpha + \sum_{j=1}^k \gamma_j X_{j,i,t} + \tau Ln * + \psi IL_{i,t-n} + \varepsilon_{i,t}$$

$i = 1, \dots, N$  (Unidades de corte transversal)

$t = 1, \dots, T$  (Observaciones de serie de tiempo)

---

<sup>12</sup>3SLS consiste en un proceso de 3 pasos: (1) se regresa cada variable endógena contra todos los regresores independientes en el sistema de ecuaciones, (2) a continuación se estiman las ecuaciones por mínimos cuadrados reemplazando las variables endógenas del lado derecho de la ecuación por los valores predichos en (1), y (3) por último se calculan las varianzas y covarianzas de los residuos estimadas en (2) y se re-estiman las ecuaciones estructurales (modelos) empleando la técnica SUR (regresiones aparentemente no relacionadas).

<sup>13</sup> El comando de post- estimación `lmareg3` para 3SLS desarrollado por Emad Abd Elmessih Shehata (2011), e.g., calcula la prueba de Durbin Watson. Como comando de post-estimación luego de 3SLS, `actest` (o la prueba de Cumby-Huizanga) aplica sobre los residuos de la última regresión, i.e., el modelo de determinantes de la liquidez.

, donde  $IL_{i,t}$  es la variable dependiente que representa el índice de liquidez para la firma  $i$  en el período  $t$ ,  $X_{j,i,t}$  desde  $X_{1,i,t}$  ...  $X_{k,i,t}$  corresponde al set de  $k=n$  variables de interés analizadas por la literatura especializada,  $L_n$  \* es cada uno de los regresores rezagados  $n=1...2$  períodos<sup>14</sup> y  $IL_{i,t-1}$  la variable dependiente rezagada  $n=1...2$  períodos y  $\varepsilon_{i,t}$  un componente aleatorio con media cero y varianza  $\sigma_\varepsilon$ . Por su parte,  $\alpha$  denota el efecto no observado y fijo específicos de la firma (no incluido en el *panel regression*),  $\gamma_j$  de  $\gamma_1$  ...  $\gamma_n$  los coeficientes de regresión que determinan el impacto de los regresores independientes sobre la liquidez,  $\tau$  representa los estimadores de las variables independientes rezagadas y  $\psi$  el coeficiente de regresión del primer y segundo rezago de la liquidez.

En la tabla a continuación se detalla el vector de  $k=n$  regresores independientes del modelo. Sobre la base de los determinantes incorporados en numerosos estudios y revisados en la sección teórica, fueron asimismo incorporadas algunas variables que permiten potencialmente capturar efectos relacionados con estos mismos factores, ello teniendo en cuenta no solo la falta de información financiera que se presenta en muchas de las acciones incorporadas sino también las heterogeneidades sectoriales particulares del mercado local. Con esta finalidad, el ratio precio-beneficio (PE, por sus siglas en inglés) fue incorporado tanto en la frecuencia ttm (considerando los beneficios como promedio móvil de los últimos 12 meses) así como en su formato de ganancias normalizadas, que desestima los resultados extraordinarios de balance. De igual manera, además del apalancamiento, se incluyeron las proporciones de pasivos de corto plazo sobre activos líquidos, y pasivos de corto plazo sobre activos totales. Estos indicadores son especialmente relevantes para la industria bancaria (como la deuda lo es para los sectores industriales), debido a que los depósitos (de corto plazo) son su principal deuda con los clientes, los cuales deben ser analizados desde una óptica de cobertura inmediata (a través de disponibilidades) o con sus activos de balance (principalmente préstamos).

**Tabla 2.- Variables del Modelo (dependiente y regresores exógenos)**

Identificador	Descripción
liquidez	Índice de (I) Liquidez
tamaño_negocio	Logaritmo natural de los Activos totales
ROA	Beneficio neto sobre activos
ROA_desvio3q	Desvío Estándar 3q ROA
MV_BV	Precio de Mercado / valor libro
EPS	(Utilidad neta-dividendos en acciones preferenciales)/ Prom. Acciones en circul.
crecimiento_activos	Tasa de crecimiento 1 período de los Activos del Balance
apalancamiento	Deuda total/ activos corrientes
PE_ttm	Precio corriente de la acción/EPS últimos 12 meses

<sup>14</sup> Para brindar una mayor flexibilidad al modelo y dado el impacto rezagado que evidencian las variables construidas a partes de datos de balance (i.e., por la presencia de valor atípicos), se incluyeron 2 rezagos.

PE_norm	PE con ganancias normalizadas
tasa_dividendos	Dividendo por acción/precio
pagodiv_dum	Dummy=1 si hay pago de dividendos
recompra_dum	Dummy=1 si hay aumento de acciones en cartera propia
deuda_acida	Pasivos de corto plazo/Activos líquidos
deuda_corriente	Pasivos de corto plazo/Activos totales
prop_activosliq	Activos Líquidos/Activos totales
EMBI_pp	Rating de deuda soberana EMBI Argentina (puntos porcentuales)
tasa_interes_pp	Tasa de interés bono soberano 2Y USA (puntos porcentuales)
PBI_pp	Tasa de crecimiento Producto Bruto Interno (%YOY, puntos porcentuales)
FX_desvioid	Desvío estándar diario del tipo de cambio nominal ARS/USD
FX	Tipo de cambio nominal ARS/USD
indice_global_desvioid	Volatilidad (diaria) del índice de mercado
Beta	Coefficiente de Riesgo Sistemico Beta
mercado_liquidez	Índice de Liquidez para el mercado
industria_liquidez	Índice de Liquidez por Industria
retornos_desvioid	Desvío estándar diario de los retornos
precio_cierre	Precio de cierre por acción
retornos	Retornos por acción

## 4.- RESULTADOS EMPÍRICOS

### 4.1.- Estadísticas descriptivas para la muestra de acciones del panel general

A partir de la información recopilada para cada firma tanto de frecuencia intradiaria, diaria como trimestral (según se utilicen en las distintas etapas de la investigación), en la Tabla 3 a continuación se exhibe el conjunto de acciones (56 en total de 68 en la muestra original) consideradas en el análisis, caracterizadas según sector económico e industria de pertenencia y agrupadas por decil de capitalización de mercado en dólares (medida de tamaño de una compañía). Asimismo, en la Tabla 4 se expone el promedio, para los mismos deciles, de los retornos diarios, su volatilidad (desvío estándar) y el volumen promedio diario operado.

**Tabla 3.- Acciones de la muestra clasificadas según deciles de Market Cap (en millones de usd), sector económico e industria**

Decil	Acciones (RIC)	Sector Económico	Industria	Market Cap (Mill USD) <sup>15</sup>
	YPFD.BA	Energía	Integrada de Petróleo y Gas	4665.45
	BMA.BA	Financiero	Bancos	2091.60
1	TXAR.BA	Industrial	Hierro, acería y fundiciones	1870.16

<sup>15</sup> Corresponde al total de acciones en circulación valuadas al precio de mercado, en moneda dólar. Es el último dato disponible publicado para cada compañía a la fecha de corte (Noviembre 2019).

	GGAL.BA	Financiero	Bancos	1815.66
	ALUA.BA	Materiales Básicos	Producción primaria de aluminio	1651.05
	TECO2.BA	Servicios de Telecomunicación	Servicios de Telecomunicación integrados	1514.03
	PAMP.BA	Servicios	Servicios eléctricos	1185.62
	BBAR.BA	Financiero	Bancos	964.90
	BRIO.BA	Financiero	Bancos	916.95
	CEPU.BA	Energía	Productores Independientes de Energía	705.34
	TGSU2.BA	Energía	Producción y Exploración de Petróleo y Gas	647.94
			Manufacturas de Cemento y Concreto	
2	HARG.BA	Materiales Básicos	Concreto	512.36
	IRCP.BA	Bienes de consumo y servicios no cíclicos	REIT comerciales	436.35
	CTIO.BA	Bienes de consumo y servicios no cíclicos	Desarrollo inmobiliario y Operaciones	410.04
	BPAT.BA	Financiero	Bancos	405.97
	IRSA.BA	Bienes de consumo y servicios no cíclicos	Distribución Alimentación y Tiendas de Conveniencia	402.21
	CRES.BA	Bienes de consumo y servicios no cíclicos	Distribución Alimentación y Tiendas de Conveniencia	303.03
3	ESME.BA	Bienes de consumo y servicios no cíclicos	Destiladoras & Bodegas	197.18
	CAPX.BA	Energía	Combustibles Renovables	190.99
	PATA.BA	Bienes de consumo y servicios no cíclicos	Distribución Alimentación y Tiendas de Conveniencia	160.21
	MOLI.BA	Materiales Básicos	Producción de granos	158.36
	INDU.BA	Materiales Básicos	Productos Químicos Agrícolas	157.66
4	CECO2.BA	Energía	Combustibles Renovables	149.72
	EDN.BA	Servicios	Servicios eléctricos	143.88
	MIRG.BA	Industrial	Equipamiento del hogar	120.72
	BHIP.BA	Financiero	Bancos	104.73
	COME.BA	Industrial	Construcción e ingeniería	99.48
	BOLT.BA	Servicios de Telecomunicación	Servicios de procesamiento de datos	98.68
5	LEDE.BA	Materiales Básicos	Producción de Caña de Azúcar	97.52
	RIGO.BA	Industrial	Envases de vidrio	97.06
	GBAN.BA	Servicios	Distribución de Gas Natural	78.72
	TRAN.BA	Servicios	Servicios eléctricos	77.27
	SAMI.BA	Bienes de consumo y servicios cíclicos	Procesamiento de frutas y vegetales	69.13
	METR.BA	Servicios	Servicios de Gas Natural	64.38
			Transporte por gasoductos de Gas Natural	
6	TGNO4.BA	Servicios	Natural	56.89
	PATY.BA	Bienes de consumo y servicios no cíclicos	Procesamiento de Carne	54.27
	CGPA2.BA	Servicios	Distribución de Gas Natural	44.23
	FERR.BA	Bienes de consumo y servicios no cíclicos	Instalaciones de fontanería y accesorios	42.43
7	CELU.BA	Materiales Básicos	Productos de papelería	37.32

	OEST.BA	Bienes de consumo y servicios no cíclicos	Operadores de autopistas	35.34
	GCLA.BA	Servicios de Telecomunicación	Proveedor de servicios de televisión por cable	35.31
	AUSO.BA	Bienes de consumo y servicios no cíclicos	Operadores de autopistas	33.13
	CADO.BA	Materiales Básicos	Pesca y Agricultura	29.34
	DGCU2.BA	Servicios	Distribución de Gas Natural	25.81
	CARC.BA	Materiales Básicos	Productos Químicos Agrícolas	20.20
	GRIM.BA	Servicios	Fabricación de Calzado	17.02
8	INTR.BA	Materiales Básicos	Ingredientes para alimentación	14.79
	AGRO.BA	Industrial	Maquinaria agrícola	14.14
	DYCA.BA	Bienes de consumo y servicios no cíclicos	Ingeniería y Construcción	13.71
	MORI.BA	Bienes de consumo y servicios no cíclicos	Molienda de Harina	10.77
	LONG.BA	Industrial	Equipamiento del hogar	10.67
	SEMI.BA	Materiales Básicos	Producción de granos	8.87
9	GARO.BA	Bienes de consumo y servicios cíclicos	Pesca y agricultura	5.30
	FIPL.BA	Industrial	Forestales y productos de madera	5.14
	ROSE.BA	Industrial	Drogas veterinarias	3.70
10	POLL.BA	Bienes de consumo y servicios no cíclicos	Construcción de autopistas y puentes	0.88
<b>Total</b>	<b>56</b>			<b>23083.60</b>

Fuente: Elaboración propia en base a Thompson Reuters Eikon

Nota: se consideraron sólo las acciones ordinarias que hayan cotizado de forma continua durante el período 2009q4-2019q2. De la muestra original de 68 papeles se consideraron 56, quedando excluidas: BYMA.BA, CVH.BA, GAMI.BA, HAVA.BA, INVJ.BA, LOMA.BA, MOLA.BA, PGR.BA, RICH.BA, SUPV.BA y TGLT.BA, VALO.BA. En el caso particular de VALO.BA, la misma fue sacada de la muestra debido a que presentaba una fuerte intermitencia en los períodos de cotización.

**Tabla 4.- Retorno diario promedio, desvío estándar de los retornos y volumen promedio por decil de acciones clasificadas según Market Cap**

Decil	Retorno diario	Desvío estándar	Volumen	Precio	Volumen \$	Participación
1	0.164%	0.080%	595173.19	\$ 76.59	\$ 17,640,989.85	50%
2	0.192%	0.094%	347910.41	\$ 25.35	\$ 6,543,253.62	19%
3	0.205%	0.082%	74720.10	\$ 35.68	\$ 1,233,335.29	4%
4	0.179%	0.113%	237690.67	\$ 17.30	\$ 3,381,749.25	10%
5	0.148%	0.094%	851835.69	\$ 35.07	\$ 2,989,434.63	9%
6	0.210%	0.111%	162884.03	\$ 21.42	\$ 1,929,219.16	6%
7	0.192%	0.106%	38168.64	\$ 11.39	\$ 296,012.52	1%
8	0.208%	0.119%	87145.87	\$ 13.24	\$ 608,103.60	2%
9	0.150%	0.103%	67912.16	\$ 5.92	\$ 250,093.04	1%
10	0.107%	0.113%	25364.26	\$ 4.71	\$ 93,803.92	0%

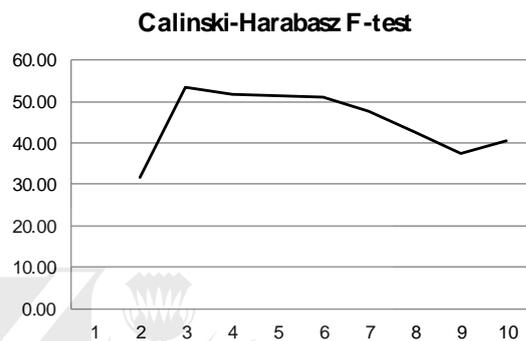
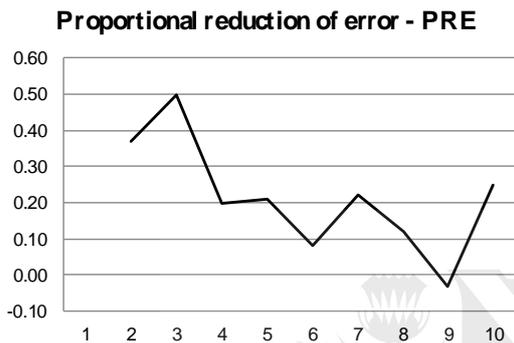
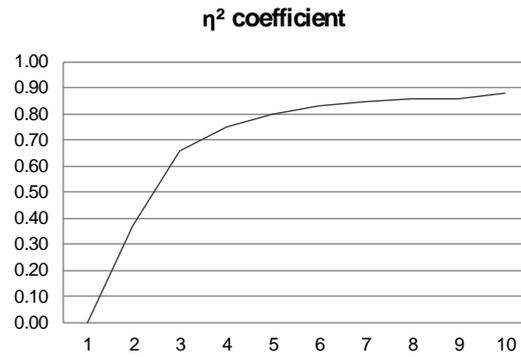
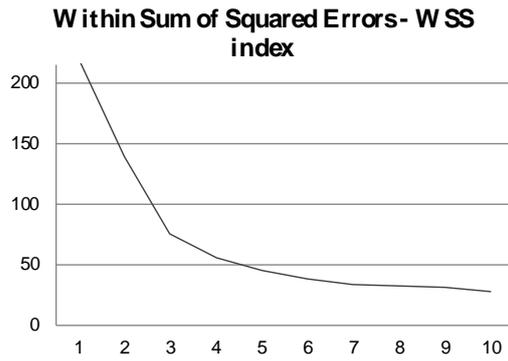
Fuente: Elaboración propia en base a Thompson Reuters Eikon

En atención a los cuadros precedentes, es observable que las acciones que componen los primeros dos deciles permiten explicar, para la totalidad del período muestral, alrededor del 70% del volumen promedio operado, mientras que los últimos dos deciles tan solo el 1%. Asimismo, el desvío estándar de los retornos para los deciles 9 y 10 es, en promedio, cerca de 1.25 veces la medida para las compañías que se ubican en los primeros dos grupos. Desde una perspectiva sectorial, tanto para el primer como el segundo segmento de acciones, puede destacarse una predominancia de sectores económicos vinculados mayoritariamente a la provisión de servicios tanto financieros como energéticos, así como de compañías productoras de materias primas para las industrias del acero, hierro y construcción. Contrariamente, en los deciles que registran un menor valor de capitalización de mercado, la predominancia se dirige hacia las industrias de bienes de consumo domésticos y de los primeros eslabones de la cadena alimentaria.

#### **4.2.- Clusterización óptima para la muestra de acciones del panel general**

Como fue señalado en la sección metodológica, el método *distribution based clustering* fue aplicado a los parámetros de la distribución del spread *bid-ask* relativo estandarizado para cada firma bajo análisis. A continuación, se exhiben los resultados obtenidos tanto para la muestra de datos intradiarios principal (3 meses de 2019m9 a 2019m12) como las muestras de validación (dos cortes sucesivos inmediatos, i.e., 2019m10-2020m1 y 2019m11-2020m2) de las medidas que permiten determinar el número de clústeres óptimo  $K$  (y su composición) que minimiza la distancia conjunta para la media de los parámetros intra-clúster y la maximiza inter-clúster, i.e., el Índice WSS, el coeficiente Eta Cuadrado  $\eta^2$ , la Reducción Proporcional del Error (PRE) y el F-Test Calinski-Harabasz.

**Gráfico 1.- Soluciones de K-Means para los  $k$  grupos ( $k = 1, \dots, 10$ ) a partir del bid-ask spread relativo estandarizado – muestra principal.**

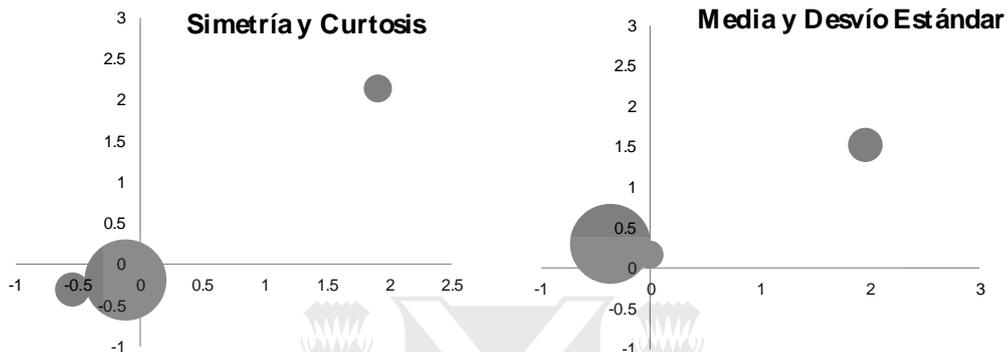


Como se desprende para el conjunto de los cuatro gráficos, los resultados indican que la segmentación en  $K=3$  clústeres es la solución óptima, ello a partir de obtener el mismo número de clústeres para cada una de las medidas abordadas. En términos más específicos, para el WSS (*Within Cluster Sum of Squared Errors*), que mide la variabilidad de las observaciones (de cada punto en relación con el centroide para cada momento de la distribución de forma multivariada) dentro de cada clúster, se aprecia un quiebre marcado para  $K=3$ . Dicho de otro modo, añadir un clúster más no minimiza significativamente más la varianza intra-clúster (ni la maximiza inter-clúster). De la misma forma, el coeficiente Eta, exhibe una reducción del WSS del 70% para  $K=3$  siendo ésta la solución óptima, en tanto la reducción proporcional de la varianza intra-clúster para  $K$  sucesivos es considerablemente inferior. Considerando esa misma disminución del WSS, pero para la solución  $K$  en relación con la solución inmediatamente anterior (PRE), se aprecia que el pico de la función se encuentra en  $K=3$ , reflejando una disminución del alrededor del 40% de la varianza al interior de cada clúster (y maximización inter-clúster) en relación a segmentar la muestra en 2. Por último, se verifica el quiebre del índice Calinski-Harabasz para un óptimo de 3 clústeres, medida heurística que permite convalidar las soluciones aportadas por los índices analizados, siendo entonces  $K=3$  grupos de acciones los que permiten de manera óptima minimizar la distancia intra-clúster y maximizarla inter-clúster. Concomitantemente, fue verificado que las firmas para  $j=1, \dots, 50$  en cada métrica computada para la solución  $k=3$  efectivamente se mantenían dentro del mismo clúster<sup>16</sup>. En efecto, la estabilidad es requerida analizar debido a que cada iteración cambia los centroides iniciales.

<sup>16</sup>El paquete *sadi* (Stata) contiene la función *ari* que permite verificar la correlación entre la composición de los clústeres para cada iteración.

En el Gráfico 2 a continuación (y la Tabla 5), se presenta una vista comparativa de los momentos de la distribución (estandarizados) de la liquidez intra-diaria por clúster de acciones, mientras que en la Tabla 6 se detallan las firmas dentro de cada grupo, indicándose los sectores económicos e industrias de pertenencia, así como el decil de tamaño de firma promedio por segmento.

**Gráfico 2.- Momentos de la distribución estandarizada (promedio) por clúster del spread bid-ask relativo**



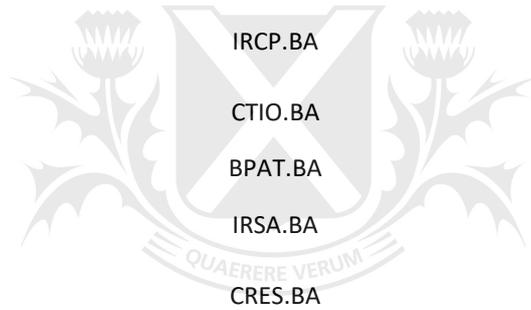
**Tabla 5.- Momentos de la distribución estandarizada (promedio) por clúster del spread bid-ask relativo**

Clúster	N acciones	Media	Desvío Estándar	Simetría	Curtosis
1	8	1.96286	1.52673	-0.54733	-0.30782
2	43	-0.36488	0.30389	-0.12003	-0.19231
3	5	-0.00262	0.17069	1.90796	2.14640

**Tabla 6.- Acciones agrupadas según estructura de clúster óptimos, por sector, industria y decil de Market Cap**

Clúster	Acciones (RIC)	Decil Market Cap
3	ALUA.BA	1
	BBAR.BA	2
	TGNO4.BA	6
	TRAN.BA	6
	YFPD.BA	1
<b>Promedio</b>	<b>5</b>	<b>3</b>
1	DYCA.BA	9
	FIPL.BA	10
	ROSE.BA	10

	MORI.BA	9
	FERR.BA	7
	PATA.BA	4
	COME.BA	5
	TXAR.BA	1
<b>Promedio</b>	<b>8</b>	<b>7</b>
	BMA.BA	1
	GGAL.BA	1
	TECO2.BA	1
	PAMP.BA	2
	BRIO.BA	2
	CEPU.BA	2
	TGSU2.BA	2
	HARG.BA	2
	IRCP.BA	3
	CTIO.BA	3
	BPAT.BA	3
	IRSA.BA	3
	CRES.BA	3
	ESME.BA	3
2	CAPX.BA	4
	MOLI.BA	4
	INDU.BA	4
	CECO2.BA	4
	EDN.BA	5
	MIRG.BA	5
	BHIP.BA	5
	BOLT.BA	5
	LEDE.BA	5
	CELU.BA	7
	OEST.BA	7
	GCLA.BA	7
	AUSO.BA	8
	CADO.BA	8
	DGCU2.BA	8
	CARC.BA	8
	GRIM.BA	8



Universidad de  
San Andrés

	INTR.BA	8
	AGRO.BA	9
	LONG.BA	9
	SEMI.BA	9
	GARO.BA	9
	POLL.BA	10
	RIGO.BA	6
	GBAN.BA	6
	SAMI.BA	6
	METR.BA	6
	PATY.BA	7
	CGPA2.BA	7
<b>Promedio</b>	<b>43</b>	<b>5</b>



Universidad de  
**San Andrés**

Tal y como puede apreciarse en el gráfico 2 (y la Tabla 5), el clúster 3 compuesto por las acciones con mayor capitalización de mercado exhibe el spread bid-ask más pequeño relativo al resto del mercado de -0.00262, mientras que el clúster 2 un spread promedio de -0.36488 y el grupo con las compañías con menor valor de mercado, una liquidez en promedio de 1.96286 comparativa a la muestra completa. De la clasificación obtenida puede entonces desprenderse que el clúster 3 resulta el más líquido, mientras que el clúster 2 más ilíquido comparativamente a este primero y el grupo de acciones en 1 el que exhibe un mayor spread bid-ask, en promedio. Asimismo, cabe destacar que las otras tres medidas de la distribución, i.e., desvío estándar, simetría y curtosis, brindan información que resulta consistente con la hipótesis original sobre los perfiles de inversión que se delinearán a través de la información de liquidez intra diaria. En otros términos, del análisis efectuado se desprende que cada clúster presenta un perfil de riesgo de la inversión distinto, y en especial marcadamente opuesto para las acciones dentro de los clústeres 1 y 3.

Al respecto, el clúster número 3 no solamente resulta en comparación el más líquido de los tres segmentos, siendo el que representa el perfil de inversión con la mayor aversión y por tanto el que captura posiblemente un gran número de inversores pequeños. Como ha sido sostenido al inicio de la presente investigación, es posible aproximar el perfil de riesgo de una inversión a través de su operación intra-diaria o, en otras palabras, de los cuatro momentos que caracterizan a la distribución de la liquidez. Dentro de este marco, presenta una pronunciada simetría positiva (1.90), indicativo que la mayor frecuencia de transacciones ocurre a spreads más bajos que la media, así como también una distribución pronunciadamente leptocúrtica (curtosis 2.14) comparativamente, lo que revela que la mayor parte de las operaciones se concentran a un spread similar y pequeño. Puede argumentarse entonces que, la mayor certidumbre acerca del precio al que se operará la firma – reforzado por el menor desvío estándar comparativo- se traduce en bajos costos de entrada y salida, lo que caracteriza a inversiones con un perfil de riesgo principalmente averso, propio de inversores pequeños que operan con un reducido presupuesto y buscan acciones con baja volatilidad y el mínimo costo de entrada y salida. Cabe destacar también que esta mayor liquidez promedio comparada también es consistente con la máxima frecuencia de operación por minuto (de la información intra-diaria se desprende que se opera 1.18 papeles por minuto).

De forma opuesta, el clúster 1 está compuesto por las 8 firmas que comparten el perfil de inversión más riesgoso de toda la muestra. Además de exhibir el spread relativo promedio más alto de los 3 conjuntos, el mayor desvío estándar (1.52) se da conjuntamente con una asimetría negativa (-0.54), lo que pone en evidencia que la mayor frecuencia de las operaciones intra-clúster se ejecutan a un valor más alto que el spread promedio, y simultáneamente con una distribución platocúrtica (curtosis -0.30) –marcada por el desvío- revele la menor certeza sobre el precio que se pueda conseguirse en el mercado para dichas acciones. Este clúster, en efecto, aparece comparativamente con los mayores costos de entrada y salida, aceptable para un perfil de inversión más amante del riesgo.

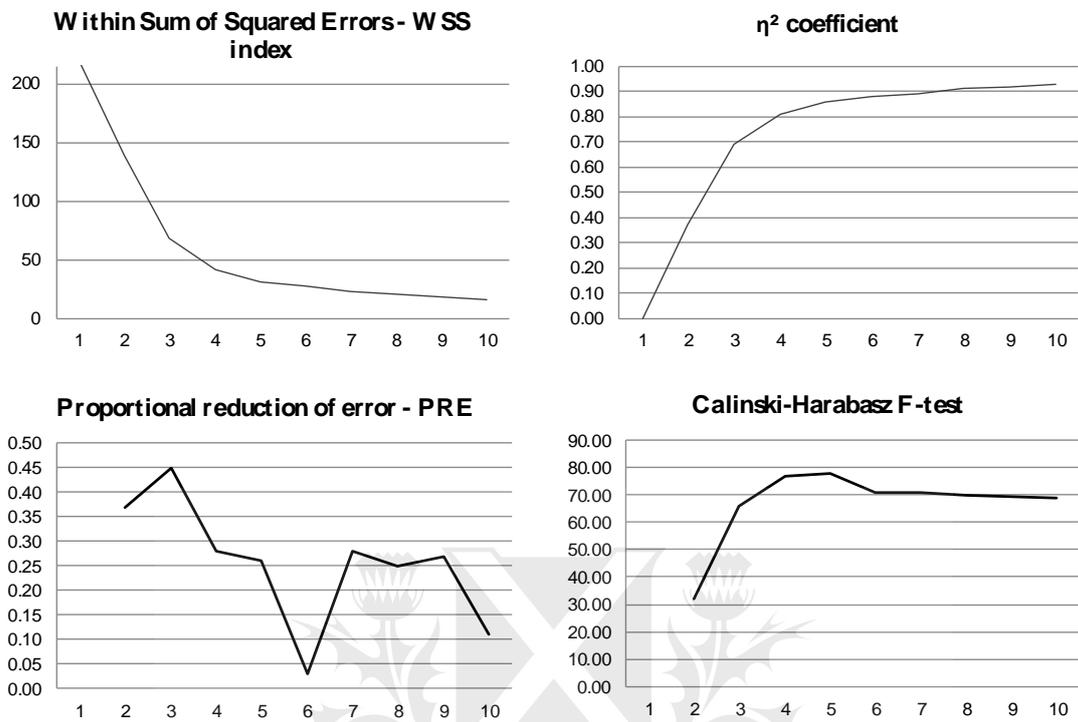
En este sentido, es posible que concentre en su mayor parte grandes inversores que buscan oportunidades de largo plazo (pudiendo soportar la mayor volatilidad de corto plazo) y presenten entonces menor necesidad de liquidez.

Por su parte, el clúster 2, que aglutina la mayor parte de las compañías, se ubica a la mitad de los extremos de la segmentación en cuanto al spread relativo y desvío estándar (0.30), aunque si en términos comparativos se habla, se encuentran valores notoriamente más cercanos en cuanto a curtosis al grupo del conjunto menos líquido (-0.19), aunque contrariamente con una ligera simetría más positiva si se la observa en contraposición al clúster 1 (-0.12). En efecto, este conjunto de acciones, si bien revela una cierta heterogeneidad en su interior, presenta en comparación al clúster 3 un perfil de la inversión más riesgoso y menos líquido.

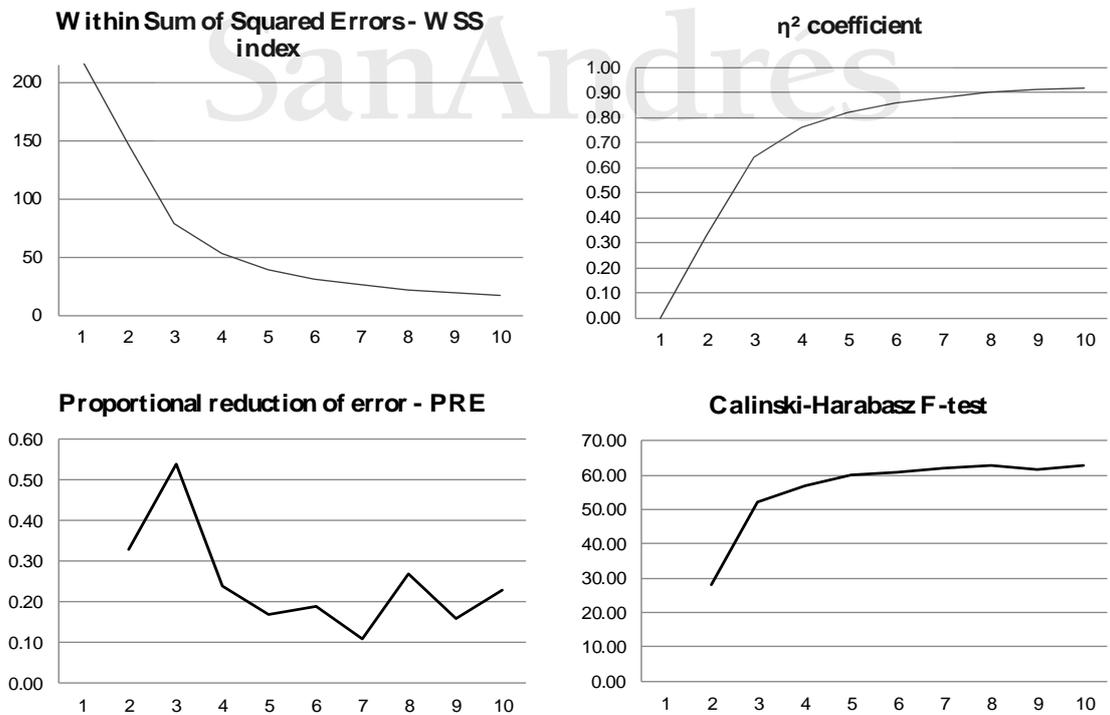
Desde otra perspectiva, vinculando el perfil de riesgo de la inversión revelado mediante las distribuciones de liquidez al interior de cada clúster con otros indicadores que afectan el riesgo de la inversión como el tamaño de la firma, los resultados exhibidos en la Tabla 6 precedente ponen de manifiesto que existe un vínculo estrecho entre el decil promedio de *market cap* para cada clúster y sus patrones de liquidez. En este sentido, el conjunto más líquido (3) en promedio se ubica en el mayor decil, mientras que contrariamente el clúster 1 presenta el tamaño de firma promedio más pequeño.

Debido a que el acceso a información intra-diaria sólo se encuentra disponible para un período de 3 meses (lo que permitiría hacer reproducible la metodología en tiempos eficientes), fue necesario repetir el procedimiento para dos cortes sucesivos, i.e., 2019m10-2020m1 y 2019m11-2020m2. A continuación se presenta las medidas propuestas para la determinación del número óptimo de clústeres para cada una de las dos muestras de validación. Como puede verificarse en los Gráficos 3 y 4, la segmentación por perfiles de la inversión se establece para un número de  $K=3$  grupos, al igual que la composición no presentó diferencias significativas. A fin de poder implementar la metodología aquí propuesta, dadas las restricciones usuales de datos de mayor frecuencia, y que los clústeres reflejen toda la información fundamental de las firmas utilizada, la segmentación empleada será la muestra original.

**Gráfico 3.- Momentos de la distribución (promedio) por clúster del spread bid-ask relativo para la Muestra 2**



**Gráfico 4.- Momentos de la distribución (promedio) por clúster del spread bid-ask relativo para la Muestra 3**



#### 4.2.- Mejor modelo de exceso de retornos por clúster y determinación del índice de liquidez

En la presente subsección se analizarán los resultados de la estimación sobre la ecuación general de exceso de retornos (dinámica) detallada en 3.1.a y 3.1.c, tomando como referencia la segmentación por clúster ya abordada. En efecto, los análisis efectuados permitirán dar con las medidas de liquidez que mejor predicen el exceso de retornos al interior de cada clúster, una vez abordados los cuatros pasos metodológicos basados en algoritmos de *machine learning* (i.e., *Feature Extraction*, *Regularization*, *All-subset-regression* & *Optimal Model Validation*). Una vez hallados los mejores predictores, se podrá construir el índice de liquidez representativo por clúster de acciones, para luego concluir, en la siguiente y última subsección empírica, con la estimación del mejor modelo de determinantes de liquidez y el análisis de los principales estimadores hallados.

En la Tabla 7 a continuación se presentan un conjunto de estadísticas descriptivas utilizada para la estimación del mejor modelo de retornos (exceso). Como puede apreciarse para las variables de interés, y en promedio para los tres clústeres por cada trimestre, los retornos de las acciones argentinas que han cotizado en el mercado local se ubican en torno a 9.4% (con una volatilidad del 25.4%), mientras que la liquidez capturada a partir del spread revela que, el spread cotizado es un 2.9% del promedio de las cotizaciones máximas y mínimas registradas, mientras que si se observa el spread considerando el precio de cierre (efectivo), éste se ubica en el 1.9%, con un desvío estándar del 1.5% y 0.8% respectivamente. Por su parte, durante el período 2009q4-2019q2 se han operado, en promedio, 236.430 acciones por trimestre, y han rotado 356 de cada 1000 títulos en circulación. Asimismo, los retornos se han modificado, en promedio un 0.042% luego de cada operación de compra o venta, mientras que ajustando por los títulos en circulación el impacto precio ha sido del 0.8% en promedio trimestral. Igualmente, las acciones argentinas han registrado un 86.7% de días con volumen positivo trimestral, mientras que un 23.45% de días con retornos cero.

**Tabla 7.- Estadísticas descriptivas para las variables de la ecuación general de retornos según frecuencia trimestral<sup>17</sup>**

Identificador	Obs.	Media	Std.
retornos	2173	0.09445	0.25454
spread_cotizado_medio	2173	0.02916	0.01506
spread_efectivo_medio	2173	0.01905	0.00804
volumen_m	2173	2.36430	6.80050
ln_volumen	2173	10.1408	2.0850

<sup>17</sup>Las variables están expresadas en su escala original, con excepción del “volumen\_m”, cuya unidad de medida es en cientos de miles, y la rotación que se escaló por cada 1 título en circulación (y no 1000, como la medida original). Particularmente, las variables fundamentales están expresadas en millones de pesos argentinos, tal como es exhibido en la información de balance.

Rotación	2173	0.35690	0.57330
amihud_rotacion	2173	0.00842	0.30775
Amihud	2172	0.00042	0.00845
d_volpositivo	2183	0.86710	0.22057
d_retornocero	2183	0.23468	0.22513
apalancamiento	2180	0.68306	0.97164
tamaño_negocio	2180	8.13737	2.14691
circulante	2183	969.853	1746.940
ROA	2182	0.04948	0.21388
mv_bv	1987	6.46493	27.9082
ROA_desvio	2183	0.03749	0.08853
FX	2183	12.6447	11.2456
indice_global	2183	0.66578	0.76495

Primeramente, se realizó la estimación de la ecuación general de exceso de retornos dinámica, cuyos resultados se exponen en las tablas seguidamente. Dentro del marco de las etapas definidas, el cuadro inmediatamente exhibe los predictores que ingresaron al espacio de variables reducido una vez efectuada la regresión por el algoritmo LASSO (que penaliza por presencia de multicolinealidad), según clúster. De los resultados obtenidos se desprende que tanto el desvío del ROA (para 3q) como las acciones en circulación fueron desestimadas como potenciales predictores, ya sea en su forma contemporánea como rezagada. Análogamente, tampoco fueron representativas la dummy ADR ni la referida a la heterogeneidad sectorial, así como no todos los años. Por otra parte, la prueba de Hausman ni su versión más generalizada hallaron las condiciones suficientes para demostrar que el uso de un modelo eficiente (RE) sea superior a la estimación de un modelo robusto (FE), para los tres clústeres<sup>18</sup>.

**Tabla 8.- Preselección de variables de liquidez por clúster mediante LASSO para la ecuación general de retornos- regresión OLS.**

VARIABLES	CLÚSTER 1		CLÚSTER 2		CLÚSTER 3	
	Coficiente	Std. Error	Coficiente	Std. Error	Coficiente	Std. Error
lyear_2011			0.1195003***	0.0199045	-0.0565838	0.0593098
lyear_2012			0.1043082***	0.0201813		

<sup>18</sup> Para los tres clústeres, el resultado de sustraer las matrices VARCOV de ambos modelos resultó definida no positiva, tanto en la versión simple del test de Hausman, como en su flexibilización con los comandos `sigmaless` y `sigmamore` empleados para corregir, ni con la prueba generalizada mediante el comando `suest`. El test de Hausman compara la estimación de un modelo más robusto pero menos eficiente (FE) con uno menos robusto pero más eficiente (RE) – eficiencia implica que para toda la matriz VARCOV los errores estándar son menores. Solo si es posible cumplir con los supuestos asumidos por el test de Hausman para la eficiencia, entonces todos los elementos de la matriz VARCOV robusta son mayores a los del modelo eficiente si no puede rechazarse la  $H_0$ . Uno de estos supuestos consiste en que la resta entre ambas matrices de lugar a una matriz definida positiva. Si por el contrario, alguna varianza o covarianza del modelo FE es menor que para RE, la esta dará lugar a una matriz no definida positivamente. En efecto, no hay evidencia que permita aseverar el uso de un modelo de efectos aleatorios, o bien porque no fueron encontrados los supuestos del modelo eficiente o porque el beneficio de usar un modelo RE no es lo suficiente como superar el error de muestreo aleatorio.

lyear_2013					0.1429557**	0.058509
lyear_2014	0.0745284**	0.0301301	0.0449874**	0.0187463	0.1582862**	0.062283
lyear_2015	0.0343246	0.032756	0.0658653***	0.0193669		
lyear_2016					-0.0661142*	0.0584666
lyear_2018	-0.0816357**	0.0356675	0.0074611	0.0234554		
spread_efectivo_medio	12.19826***	1.744186	11.48093***	1.00773		
volumen_m	0.0014432	0.0013711			0.0336232*	0.0173255
rotacion	0.000069***	0.0000263	0.0001212***	0.000018	0.000072	0.0001005
amihud_rotacion	0.0008395**	0.0003599				
amihud	0.1053402***	0.0146108	0.0016387***	0.0005521	0.0276236	0.0310832
apalancamiento					-0.1146418*	0.0584434
mv_bv					0.0082333	0.0086578
FX			0.0049683***	0.0007472		
tamaño_negocio	-0.0310706	0.047046				
indice_global	0.0932096***	0.0153693	0.1368403***	0.0087373	0.119313***	0.0260243
Lretornos	-0.1161934**	0.0477696			-0.26523***	0.0782098
Lspread_cotizado_medio	-0.6010334	1.227536	-2.699616***	0.6701434		
Lspread_efectivo_medio					5.912035*	3.356583
Lrotacion	-0.0000516**	0.0000247	0.0000736***	0.0000187		
Ld_volpositivo			-0.1495696**	0.0711777		
Ld_retornocero	0.1691949**	0.0772535	0.063416	0.0802225		
Ltamaño_negocio	-0.0293146	0.0488244			-0.0315458	0.0291912
Lmv_bv	-0.0025216	0.001749				
Lindice_global			-0.0601491	0.0083729	-0.0202742	0.0275589
LROA					-0.3747736*	0.2126922
_const	0.0041186	0.0091318	-0.0008391	0.005208	-0.0003179	0.0162138
Observaciones	284		1621		152	
P value<0.01 *** <0.05** <0.1 *						
R-squared	0.57		0.36		0.44	
Adj R-squared	0.54		0.35		0.38	

A partir de esta reducción del espacio de potenciales predictores, el verdadero proceso generador de los datos que minimiza el error cuadrático medio de la predicción fue hallado mediante el algoritmo GSREG, implementado para diferentes cortes muestrales mediante la técnica *K-fold validation*, como ya ha sido detallado en la metodología. En la tabla seguidamente se exhiben los predictores del exceso de retornos que han ingresado al menos a alguno de los mejores  $K=1...10$  modelos estimados. Asimismo, en el gráfico 5, se focaliza en las medidas de liquidez resultantes y su pertenencia conjunta a los distintos *K-folds* analizados, tanto en su forma contemporánea como rezagada.

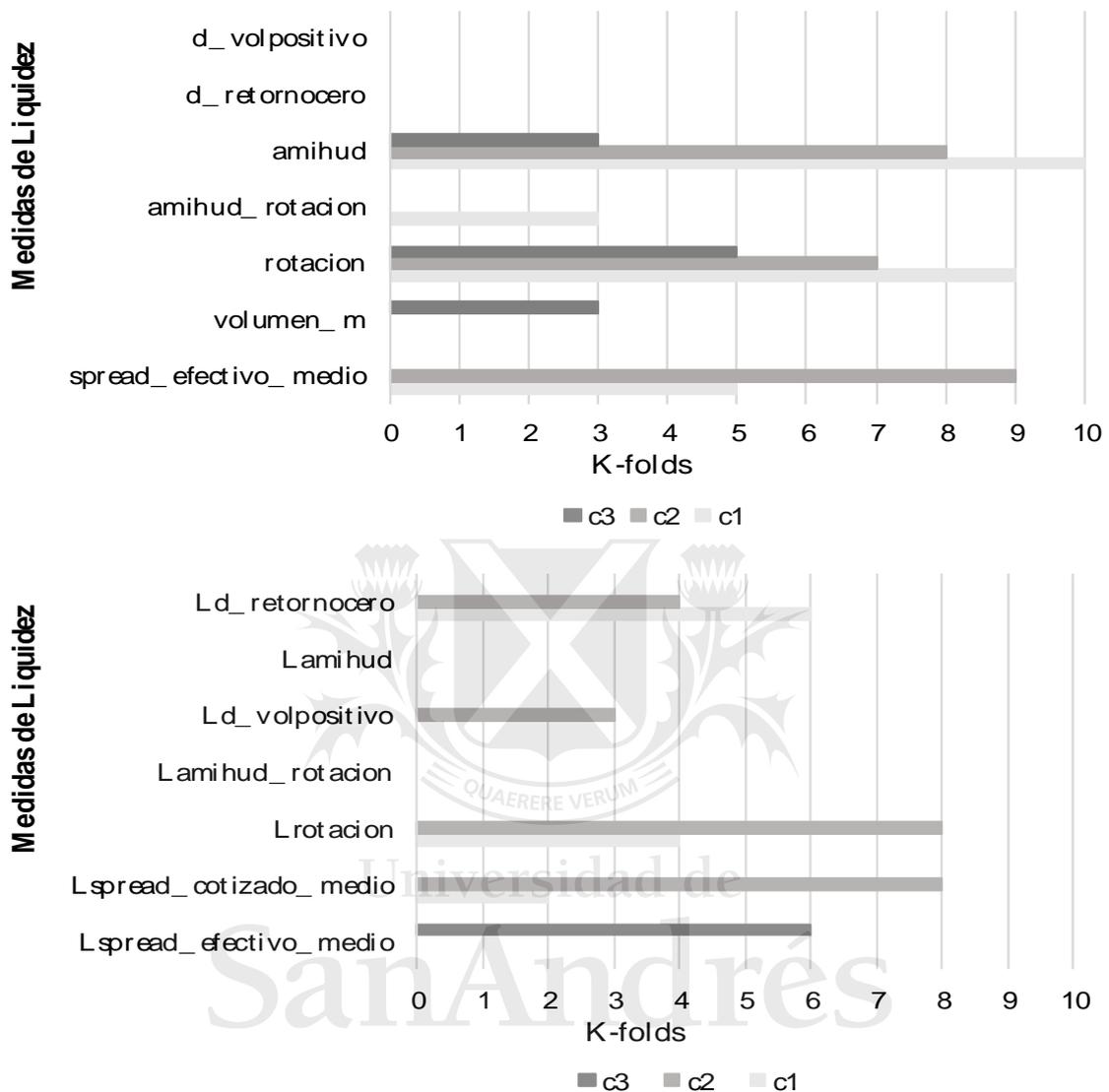
**Tabla 9.- Mejores predictores del exceso de retornos para la ecuación general y su representatividad en las  $K=1...10$  estimaciones, según clúster**

Variables	Clúster 1		Clúster 2		Clúster 3	
	Coefficiente	K-grupos	Coefficiente	K-grupos	Coefficiente	K-grupos
_lyear_2011			-0.0378	3	-0.0008	5
_lyear_2012			-0.0003	4		

_lyear_2013					0.0654	3
_lyear_2014	0.02580	3	0.0131	4	0.0757	3
_lyear_2015	0.00007	4	0.0082	3		
_lyear_2016					-0.0104	1
_lyear_2018			1.3912	5		
volumen_m					0.0175	3
spread_efectivo_medio	6.48728	5	11.9951	9		
rotacion	0.00011	9	0.0001	7	0.0002	5
amihud_rotacion	0.00024	3				
amihud	0.13425	10	0.0014	8	0.0132	3
tamaño_negocio	-0.00423	1				
apalancamiento					-0.0180	2
mv_bv					0.0068	5
indice_global	0.00823	1	0.0001	1	0.0410	3
FX			0.0056	5		
Lretornos	-0.00100	6			-0.1383	5
Lspread_cotizado_medio	-0.44171	2	-1.8874	8		
Lspread_efectivo_medio					4.6457	6
Lrotacion	-0.00003	4				
Ld_volpositivo			-0.0681	3		
Ld_retornocero	0.10169	6	0.0271	4		
Lmv_bv	0.01584	4				
Ltamaño_negocio					-0.0034	4
LROA					-0.0606	1
Lindice_global			-0.0644	5	-0.0016	1
const	0.01584	10	0.1746	10	0.0032	10
Índice de (I) Liquidez	0.18239		0.1477		0.2197	
Observaciones	171		930		99	
Rmse_out	0.086		0.132		0.111	

Universidad de  
San Andrés

**Gráfico 5.- Representatividad de las variables de liquidez en los K=1...10 mejores modelos**



Una vez concluidas las cuatro etapas de la estimación, puede apreciarse que todas las medidas de liquidez propuestas son predictores óptimos en al menos algunos de los clústeres, ya sea de forma contemporánea o rezagada. No obstante, abordando en mayor detalle los modelos obtenidos para cada clúster, la medida de Amihud va incrementando en representatividad a medida que el conjunto de acciones se vuelve más ilíquido, siendo 100% para el clúster 1, 80% para el clúster 2 y 30% para el clúster 3; análogamente ocurre para la rotación – predice en 9 modelos, 7 y 5 respectivamente, aunque en su efecto rezagado ha mejorado su participación para el clúster 2. Cuando se corrige la medida de Amihud por rotación ésta sólo resulta relevante en el clúster menos líquido (1), y el spread efectivo medio predice peor en el clúster más ilíquido (50% de los K-folds), mientras que en el grupo más heterogéneo (2) lo hizo en 9 modelos y en el clúster más líquido en 6 oportunidades si se considera el primer rezago. Por su parte, tanto el spread cotizado medio, así como la proporción de días con volumen positivo y los días con retornos cero permiten predecir óptimamente el exceso de retornos en su forma rezagada, para los clústeres 1 y 2 respectivamente. En efecto, era

de esperar que los días con retorno cero sea más relevantes para predecir al interior del clúster más ilíquido, así como un indicador asociado al volumen lo haga para segmentos más líquidos. No obstante, es llamativa la poca representatividad del volumen como indicador de liquidez, en tanto sólo ingresó en 3 de los 10 cortes analizados para el clúster 3.

Evaluando asimismo los regresores de control, las primeras conclusiones que pueden obtenerse pueden resumirse en los puntos siguientes:

- (i) un set variado de medidas de liquidez resultó sumamente representativo, en particular al interior de los clústeres más ilíquidos (1 y 2). Asimismo, todos los regresores de la liquidez permitieron predecir de forma óptima la variable dependiente ya sea en su forma contemporánea o rezagada. Más específicamente, el número de variables de interés que resultó en cada clúster y su peso, en promedio, fueron: (a) 6 medidas en promedio prediciendo en 5.5 de cada 10 modelos para el clúster 1; (b) 6 medidas en promedio prediciendo 6.7 de cada 10 modelos para el clúster 2; y (c) 3 indicadores de liquidez, en promedio, 4 veces.
- (ii) el volumen en niveles resultó llamativamente un predictor muy débil.
- (iii) en relación con los regresores del negocio, el clúster más líquido presentó la mayor presencia de esta clase de predictores (4 variables, con un peso promedio del 30%), seguido del clúster más riesgoso (grupo 1, con 2 regresores en promedio 2.5 veces de cada 10). A partir de ello puede concluirse que aquellos segmentos con patrones de inversión y riesgo más definido permiten capturar mejor los efectos financieros sobre los retornos (exceso).
- (iv) por último, la inclusión de variables dummy para capturar los efectos estacionales (años) resultó más preponderante en los clústeres 2 y 3, con 5 años representativos, en promedio, en un 38% y 30%, respectivamente.

De las conclusiones señaladas, y tomando en consideración que el volumen puede presentar valores atípicos recurrentes y que la incorporación de variables categóricas podría estar incrementando de un modo no deseable el error cuadrático medio out-of-sample (podrían alterar el verdadero valor de los coeficientes para el set de entrenamiento, i.e., se vuelve una herramienta útil para el ajuste *in-sample* pero incrementa la varianza *out-of-sample*), fueron estimados dos modelos adicionales a fin de convalidar dichas hipótesis, y dar con las medidas de liquidez óptimas. En atención a esto, el modelo original es denominado Modelo 1, mientras que el Modelo 2 propone una transformación logarítmica del volumen, y el Modelo 3 remover las variables categóricas de los años. El mejor modelo de exceso de retornos será el que presente el menor error cuadrático medio de predicción (promedio para los tres clústeres) y la menor desviación estándar de dichos errores relativo al promedio, consistente con la metodología de *machine learning*. Las tablas a continuación exponen los mejores estimadores, su predictibilidad para los  $K=1...10$  modelos evaluados para los tres clústeres tanto del Modelo 2 como el Modelo 3, así como también el análisis comparativo de las medidas de selección adoptadas bajo este enfoque.

**Tabla 10.- Mejores predictores del exceso de retornos para la ecuación general y su representatividad en las K=1...10 estimaciones para el Modelo 2, según clúster**

Variables	Clúster 1		Clúster 2		Clúster 3	
	Coefficiente	K-grupos	Coefficiente	K-grupos	Coefficiente	K-grupos
_lyear_2011			-0.0136	1	-0.0676	4
_lyear_2012					-0.0511	4
_lyear_2013					0.0676	3
_lyear_2014	0.0181	2	0.0299	3	0.0902	4
_lyear_2015	0.0071	4				
_lyear_2016					-0.0085	1
_lyear_2018			-0.0435	3		
spread_efectivo_medio	6.7448	5	6.8431	6		
ln_volumen	0.0000	1	0.0015	6		
rotacion	0.0001	9	0.0000	4	0.0002	5
amihud_rotacion	0.0001	2				
amihud	0.1344	10	0.0017	8	0.0078	2
tamaño_negocio	-0.0042	1				
d_retornocero						
apalancamiento					-0.0153	1
mv_bv					0.0113	6
indice_global	0.0041	1			0.0448	3
FX			0.0136	4		
Lretornos	-0.0949	6			-0.1353	5
Lspread_cotizado_medio	-0.4867	3	-0.8018	6		
Lspread_efectivo_medio					6.3911	7
Lrotacion	0.0000	5			0.0000	2
Ld_volpositivo			-0.0471	4		
Ld_retornocero	0.1056	6				
Lmv_bv	-0.0011	4				
Ltamaño_negocio					-0.0196	5
LROA					-0.0573	2
Lindice_global			-0.0674	6	-0.0017	1
const	0.0155	10	0.0233	10	-0.0031	10
Índice de (I) Liquidez	0.1795		0.0796		0.2051	
Observaciones	171		929		97	
Rmse_out	0.0862		0.127		0.111	

**Tabla 11.- Mejores predictores del exceso de retornos para la ecuación general y su representatividad en las K=1...10 estimaciones para el Modelo 3, según clúster**

Variables	Clúster 1		Clúster 2		Clúster 3	
	Coefficiente	K-grupos	Coefficiente	K-grupos	Coefficiente	K-grupos
spread_efectivo_medio	5.5539	5	4.1425	5		
ln_volumen	0.0005	5	0.0021	10	0.0008	5
rotacion	0.0001	9	0.0000	6	0.0001	5
amihud_rotacion	0.0003	3			-0.0516	10
amihud	0.1350	10	0.0144	9	3.5339	10

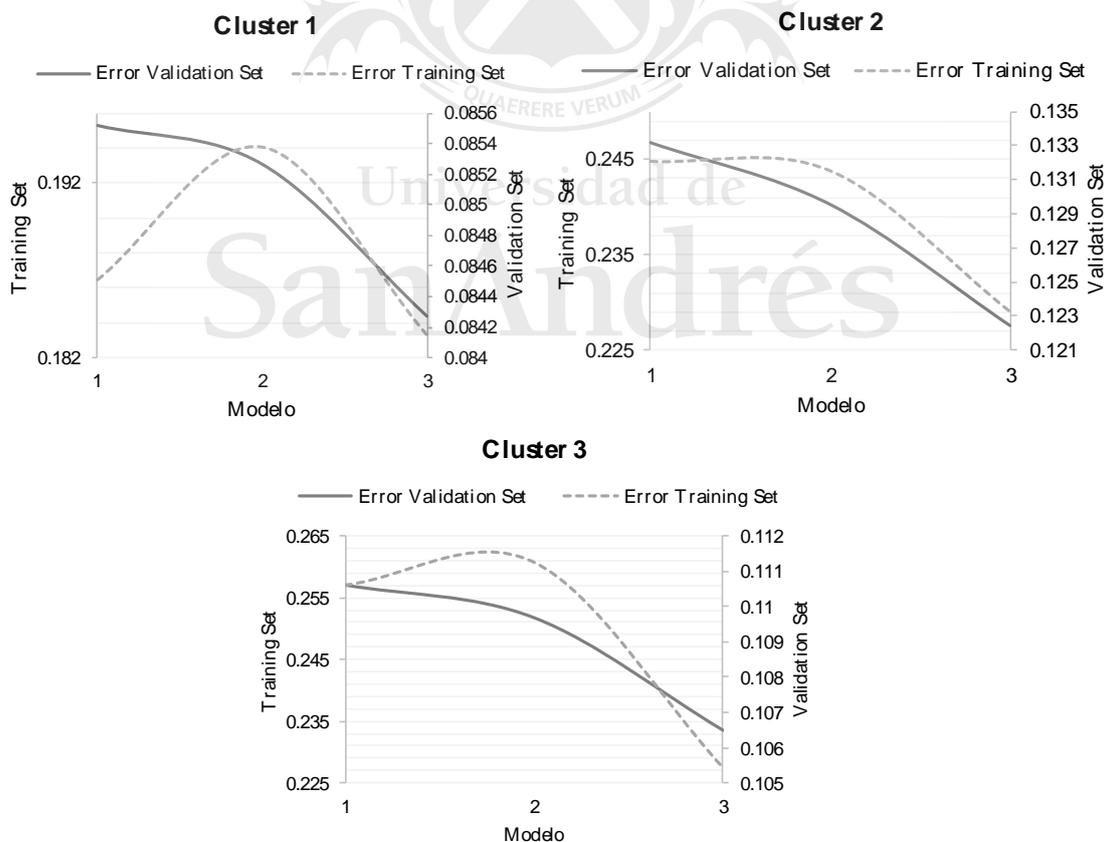
tamaño_negocio	-0.0018	2	0.0003	6		
d_retornocero					-0.4142	5
apalancamiento					0.0014	1
indice_global	0.0048	1	0.0078	1	0.0235	2
FX			-0.0004	1		
Lretornos	-0.0930	5			-0.0406	3
Lspread_cotizado_medio	-0.6713	4	0.2273	1		
Lspread_efectivo_medio					2.7519	4
Lrotacion	0.0000	3			0.0000	1
Ld_volpositivo			-0.0292	3	-0.2669	2
Lamihud	0.0176	5				
Ld_retornocero	0.0849	5				
Lmv_bv	-0.0020	5	-0.0006	4		
Ltamaño_negocio					-0.0537	7
LROA					-0.0176	1
Lindice_global			-0.0665	9	-0.0189	3
const	0.0177	10	3.1512	9	0.0171	10
Índice de (I) Liquidez	0.1501		0.0778		-0.1685	
Observaciones	170		857		110	
Rmse_out	0.085		0.123		0.105	

**Tabla 12.- Análisis comparativo de los modelos según RMSE para el *training set* y el *validation set***

K-fold	Clúster 1		Clúster 2		Clúster 3		Promedio		
	rmse_in	rmse_out	rmse_in	rmse_out	rmse_in	rmse_out	rmse_in	rmse_out	
<b>Modelo 1</b>	1	0.1711	0.0926	0.2436	0.1383	0.2640	0.1426	0.2263	0.1245
	2	0.1709	0.0855	0.2382	0.1308	0.2588	0.1358	0.2226	0.1173
	3	0.1722	0.0875	0.2432	0.1290	0.2619	0.1269	0.2258	0.1145
	4	0.1760	0.0766	0.2519	0.1210	0.2745	0.0947	0.2341	0.0974
	5	0.1721	0.0709	0.2529	0.1181	0.2753	0.0991	0.2334	0.0960
	6	0.1961	0.0888	0.2426	0.1376	0.2777	0.1141	0.2388	0.1135
	7	0.2168	0.0796	0.2501	0.1440	0.2556	0.1072	0.2408	0.1103
	8	0.2071	0.0925	0.2477	0.1398	0.2403	0.1037	0.2317	0.1120
	9	0.1998	0.0899	0.2447	0.1358	0.2301	0.0927	0.2249	0.1062
	10	0.1825	0.0914	0.2332	0.1374	0.2326	0.0893	0.2161	0.1060
Promedio									<b>0.1098</b>
Std. Relativo									<b>0.2116</b>
<b>Modelo 2</b>	1	0.1711	0.0926	0.2297	0.1254	0.2606	0.1423	0.2205	0.1201
	2	0.1788	0.0855	0.2279	0.1214	0.2627	0.1350	0.2231	0.1140
	3	0.1723	0.0875	0.2245	0.1226	0.2776	0.1178	0.2248	0.1093
	4	0.1760	0.0766	0.2302	0.1176	0.2745	0.0947	0.2269	0.0963
	5	0.1721	0.0709	0.2529	0.1181	0.2753	0.0991	0.2334	0.0960
	6	0.1961	0.0888	0.2333	0.1334	0.2752	0.1087	0.2349	0.1103
	7	0.2168	0.0796	0.2426	0.1423	0.2670	0.1034	0.2421	0.1084
	8	0.2763	0.0925	0.2469	0.1424	0.2403	0.1037	0.2545	0.1129
	9	0.1998	0.0899	0.2745	0.1359	0.2325	0.0972	0.2356	0.1077

	10	0.1812	0.0888	0.2757	0.1364	0.2420	0.0948	0.2330	0.1067
<b>Modelo 3</b>	Promedio								<b>0.1082</b>
	Std. Relativo								<b>0.2010</b>
	1	0.1651	0.0913	0.2261	0.1160	0.2630	0.1390	0.2181	0.1154
	2	0.1706	0.0861	0.2235	0.1123	0.2647	0.1307	0.2196	0.1097
	3	0.1737	0.0875	0.2326	0.1174	0.2768	0.1224	0.2277	0.1091
	4	0.1768	0.0762	0.2306	0.1083	0.2226	0.0864	0.2100	0.0903
	5	0.1727	0.0689	0.2297	0.1095	0.2186	0.0846	0.2070	0.0876
	6	0.1872	0.0810	0.2326	0.1291	0.2206	0.1024	0.2135	0.1042
	7	0.1959	0.0808	0.2351	0.1310	0.1976	0.1062	0.2095	0.1060
	8	0.2077	0.0927	0.2343	0.1341	0.2156	0.1030	0.2192	0.1099
	9	0.2007	0.0901	0.2246	0.1361	0.1998	0.0963	0.2084	0.1075
10	0.1827	0.0880	0.2207	0.1308	0.1959	0.0938	0.1998	0.1042	
Promedio									<b>0.1044</b>
Std. Relativo									<b>0.1947</b>

**Gráfico 6.- Trayectoria del RMSE promedio *training set* y *validation set* comparativa para los tres modelos**



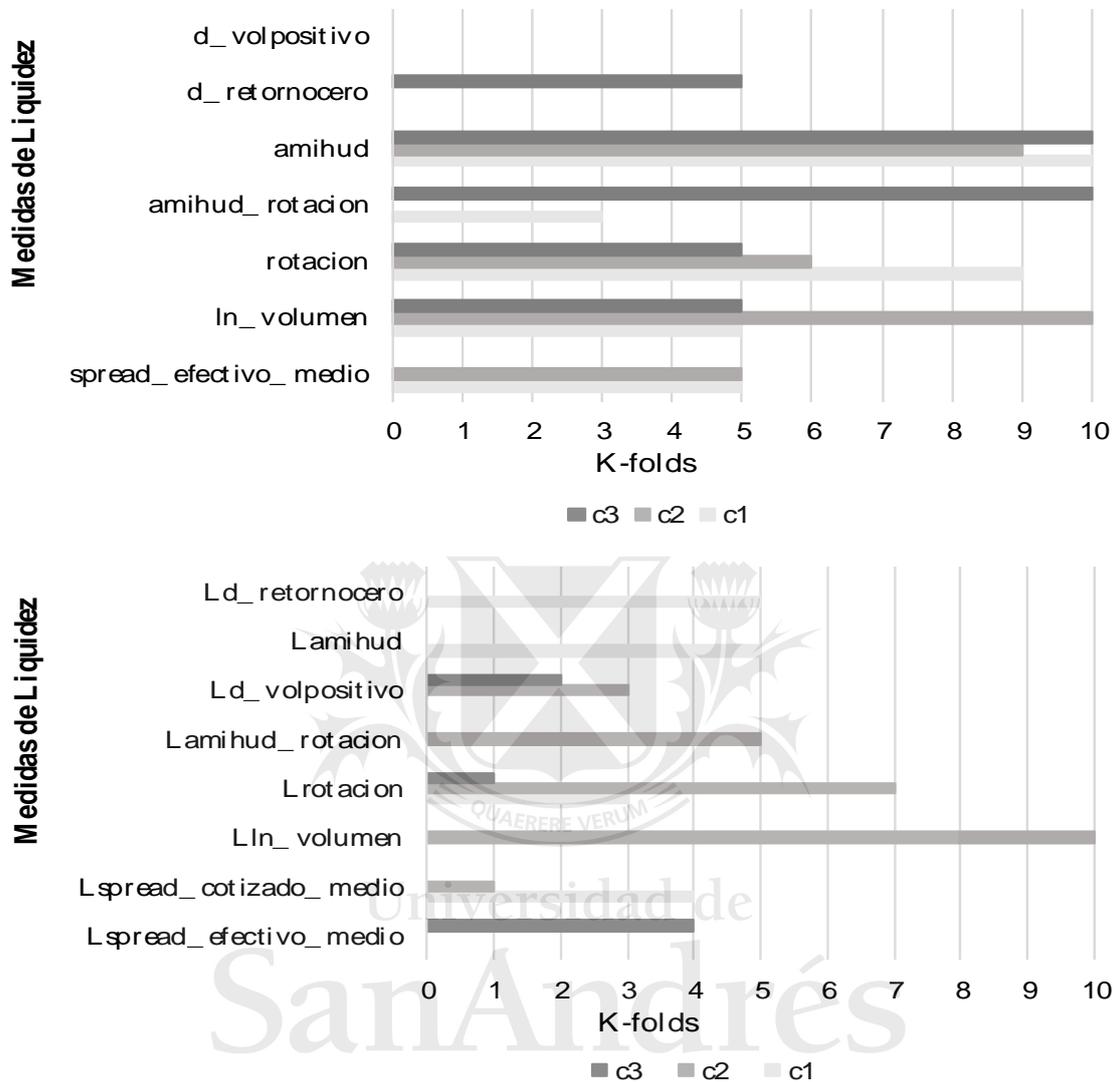
Desde un análisis conjunto, se demuestra que la primera solución proporcionada (i.e., el Modelo 2 con la transformación logarítmica del volumen) permitió que el volumen

expresado de forma logarítmica se volviera un predictor relevante. Más específicamente, este indicador de liquidez resultó un estimador óptimo para 1 de los 10 modelos estimados en el clúster 1, mientras en el clúster 2 exhibió una participación del 60% en su forma contemporánea y del 80% para la variable rezagada un período. No obstante, la sola aplicación de esta transformación eliminó al volumen como predictor para el clúster 3. Por su parte, de la comparación de la capacidad predictiva del Modelo 2 en relación con la versión original, la incorporación del logaritmo del volumen como predictor fue concomitante con la disminución, en promedio para los tres clústeres, del error cuadrático medio *out-of-sample*, así como de su desvío estándar relativo.

No obstante, y tal como se visualiza en la tabla de medida comparativa (Tabla 12) y el gráfico precedente (gráfico 6), la reestimación de la ecuación general de exceso de retornos para cada clúster removiendo las variables categóricas de los años (que capturaban potenciales efectos estacionales no observados en los regresores cuantitativos), permitió reducir aún más los errores de predicción para el set de validación y su volatilidad relativa, en promedio para los tres clústeres. Más específicamente, mientras el Modelo 1 proporcionó un RMSE para el conjunto de datos de validación, en promedio, de 0.1098 y una desviación relativa para los distintos K-folds de 0.2116, el Modelo 2 permitió reducir dichos errores en un 1.5% - y la volatilidad en un 5%, el Modelo 3 mejoró la predictibilidad un 4.9% y la desviación estándar relativa en un 8% - ello contra el Modelo original-.

A partir de estos resultados obtenidos, se sostiene que el mejor modelo de exceso de retornos es el Modelo 3, y por lo tanto las medidas de liquidez que mejor predicen. En el gráfico siguiente se exhiben los regresores que conformarán los índices de liquidez necesarios para abordar los determinantes de la liquidez.

**Gráfico 7.- Representatividad de las variables de liquidez en los *K-fold* para el Modelo seleccionado (Modelo 3)**



A partir de los resultados arrojados por el Modelo 3, se ponen de evidencia las medidas de liquidez que mejor permiten aproximar al exceso de retornos para el conjunto de las acciones argentina segmentadas según los distintos perfiles de la inversión. De un modo concluyente, luego de encontrar la estructura de la ecuación general con mejor capacidad predictiva dentro de un conjunto posible de formas funcionales, los predictores óptimos y representatividad en términos de K-folds se detallan en el gráfico precedente y en la tabla seguidamente.

A fin de poder analizar con mayor profundidad y comparativamente el impacto de los regresores de liquidez sobre la variable predicha, en la tabla 13 se presentan los resultados de la descomposición de varianza efectuada, siguiendo a Cavallo y Valenzuela (2007). En este sentido, la contribución de cada covariable se define como:

$$\beta_{z_k} \rho_{y,z_k} = \beta_{z_k} * \left( \frac{\sqrt{\text{var}_{z_k}}}{\sqrt{\text{var}_y}} \right) * \rho_{y,z_k}$$

, donde  $Z_k$  es cada variable independiente  $Z = [x_1, \dots, x_n]$ ,  $\beta_{z_k}$  representa a cada coeficiente de regresión múltiple estandarizado de  $Z_k$  y  $\rho_{y,z_k}$  el coeficiente de correlación de Pearson simple entre la variable dependiente y  $Z_k$ . La sumatoria de cada contribución da por resultado el coeficiente de determinación del modelo  $R^2$ .

**Tabla 13- Análisis de descomposición de varianza por clúster para el mejor modelo de exceso de retornos (Modelo 3)**

VARIABLES DE LIQUIDEZ	Clúster 1			Clúster 2			Clúster 3			Part. prom
	Coef.	K-grup.	Contrib.	Coef.	K-grup.	Contrib.	Coef.	K-grup.	Contrib.	
spread_efectivo_medio	5.5539	5	6.02%	4.1425	5	2.40%				33%
ln_volumen	0.0005	5	0.08%	0.0021	10	0.17%	0.0008	5	0.06%	67%
rotacion	0.0001	9	8.26%	0.0000	6	0.94%	0.0001	5	7.14%	67%
amihud_rotacion	0.0003	3	0.77%				-	10	4.06%	65%
amihud	0.1350	10	18.30%	0.0144	9	11.45%	3.5339	10	56.41%	97%
d_retornocero							-	5	1.31%	17%
Lspread_efectivo_medio							0.4142	4	0.84%	13%
Lspread_cotizado_medio	-0.6713	4	0.38%	0.2273	1	-0.04%				17%
Lln_volumen	-	5	0.00%	-0.0015	10	0.01%				33%
Lrotacion	0.0000	3	0.81%	0.0000	7	0.16%	0.0000	1	-0.08%	37%
Lamihud_rotacion				0.0000	5	-0.24%				17%
Ld_volpositivo				-0.0292	3	0.12%	-	2	-0.03%	17%
Lamihud	0.0176	5	-0.29%				0.2669			17%
Ld_retornocero	0.0849	5	0.38%							17%

En términos generales, los estimadores contemporáneos más representativos resultaron ser la medida de Amihud, con casi 97% de peso para los tres clústeres en los diferentes sets de entrenamiento, seguido del volumen (logarítmico), la rotación y la medida de Amihud corregida – con más del 65% de participación en los K=1...30 – considerando los tres segmentos-, y en menor medida del spread efectivo medio y los días con retorno cero. La capacidad predictiva de los indicadores de liquidez con un rezago, medido en términos de K folds, se presenta mucho más homogéneamente distribuida entre los diferentes indicadores de liquidez, siendo los efectos del volumen y la rotación lo más relevantes.

A fin de poder comparar los predictores de liquidez entre clústeres, la descomposición de varianza realizada expone la contribución de cada covariable a la varianza del exceso de retornos al interior de cada clúster. De los resultados obtenidos se puede concluir:

(i) el spread efectivo medio aproximado a partir de datos diarios, siguiendo a Jain (2003), no contribuye a explicar la variabilidad del exceso de retornos como sí lo hacen las medidas cuya frecuencia es de naturaleza diaria y no intra-diaria, tales como la medida de Amihud (2002), empleada en diversos estudios. No obstante, ello, el spread efectivo medio va creciendo en importancia a medida que el clúster se vuelve más ilíquido, lo que es consistente con la clusterización delineada.

(ii) la medida de Amihud resulta no solamente el mejor predictor en términos de K-folds (en su forma contemporánea) sino también por la contribución a la variabilidad del exceso de retornos, al interior de cada clúster.

(iii) A pesar de su capacidad predictiva promedio, el volumen (logarítmico) exhibe una contribución muy pequeña para los tres clústeres en comparación con otras medidas de liquidez.

(iv) las medidas de liquidez construidas resultan más relevantes en su forma contemporánea, tanto en términos de participación en los mejores modelos como de contribución. En efecto, esto se encuentra asociado a la frecuencia de las observaciones y la naturaleza de las variables de liquidez (diarias).

(v) A partir de la proporción de medidas de liquidez definidas teóricamente en sentido positivo (i.e. volumen, rotación y proporción de días con volumen positivo) en relación con aquellas de iliquidez (i.e., Spread efectivo, Spread cotizado, Amihud, Amihud ajustada por rotación y proporción de días con retorno cero), se observa que para el clúster 1, 5 de las 7 medidas son de iliquidez; en el clúster 2 y 3 esta relación es de 4 a 7, contemplado los efectos totales.

(vi) No obstante, y a fin de poder definir certeramente el sentido del índice construido, se regresaron los retornos contra el índice de liquidez por clúster para definir su sentido con una regla que no dependa de una regla teórica. En efecto, si el incremento del índice de liquidez implica mayores retornos se define como un indicador compuesto principalmente explicado por variables de liquidez en sentido negativo. Como resultado, el índice de liquidez presentó una co-variabilidad positiva con los retornos para los tres clústeres (clúster 1 exhibió un coeficiente positivo del índice de +0.9365; el clúster 2 de +0.2741 y el clúster 3, +0.0067).

#### 4.3.- Mejor modelo de determinantes de liquidez por clúster

En la Tabla a continuación se presentan un conjunto de estadísticas descriptivas para la muestra completa utilizada para la estimación del mejor modelo determinantes de la liquidez.

**Tabla 14.- Estadísticas descriptivas para las variables de la ecuación general de determinantes de la liquidez**

Identificador	Obs.	Media	Std.
liquidez	2183	0.0661434	0.7979697
tamaño_negocio	2180	8.137366	2.146909
ROA	2182	0.0494837	0.2138839

ROA_desvio3q	2183	0.0374924	0.0885282
MV_BV	2183	5.884475	26.68941
EPS	2144	0.4039412	8.869024
crecimiento_activos	2127	0.0073462	0.0503089
apalancamiento	2180	0.6830634	0.9716358
PE_ttm	1547	135.1658	2631.858
PE_norm	1506	56.70152	287.9364
tasa_dividendos	1980	0.023826	0.0702914
deuda_acida	2167	16.25347	76.48187
deuda_corriente	2180	0.4338104	0.4316232
prop_activosliq	2180	0.1230661	0.1141486
EMBI_pp	2183	7.004459	2.332949
tasa_interes_pp	2183	0.9359647	0.7242174
PBI_pp	2183	1.505535	4.680597
FX_desvioid	2183	0.4676236	0.930988
FX	2183	12.64466	11.24562
indice_global_desvioid	2183	0.0495698	0.0438215
Beta	2183	0.4954636	1.220305
mercado_liquidez	2183	3.701696	6.494593
industria_liquidez	2183	0.6156211	1.880289
retornos_desvioid	2169	0.0290582	0.0112214
precio_cierre	2173	23.52951	57.37193
retornos	2173	0.0944499	0.2545401

Como fue descrito en la sección metodológica, el procedimiento de estimación empleado para la selección de variables en base a una combinación de algoritmos y técnicas de *machine learning*, fue igualmente implementado para obtención del verdadero proceso generador de los datos en lo referido a los determinantes de la liquidez. No obstante, no solamente fue requerido dar con los predictores óptimos reduciendo el *trade-off* entre sesgo y varianza, sino además obtener un conjunto de estimadores insesgados que permitiera aportar interpretabilidad al modelo analizado. En este sentido, a la implementación de las cuatro etapas ya descritas, i.e., *feature extraction*, *regularization*, *all-subset regression* y *optimal model validation*, se adicionó la estimación del mejor modelo de determinantes de la liquidez a partir de un sistema de ecuaciones simultáneas que permitiera eliminar el principal sesgo de autocorrelación, i.e., endogeneidad entre liquidez y retornos.

Tomando como variables dependientes dentro del sistema de ecuaciones a los retornos y el índice de liquidez – obtenido del mejor modelo de exceso de retornos-, a continuación, se exhiben los regresores que resultaron estadísticamente significativos para explicar la liquidez en el largo plazo (ello debido a la estabilidad paramétrica otorgada por la técnica *K-fold validation*), según cada clúster. Asimismo, se presentan los resultados de la prueba de Cumby-Huizinga efectuada sobre los residuos del modelo de liquidez. Cabe destacar, por otra parte, que los resultados arrojados por la prueba de

Hausman para modelos de datos de panel convalidó la estimación a partir de una estructura de efectos fijos<sup>19</sup>.

**Tabla 15.- Mejor modelo de determinantes de liquidez por clúster, efectos totales**

Variables	Clúster 1		Clúster 2		Clúster 3 <sup>20</sup>	
	Coefficiente	pvalue	Coefficiente	pvalue	Coefficiente	pvalue
mercado_liquidez	0.0004	0.588	0.0006	0.05**		
industria_liquidez			0.0035	0.009***		
retornos_desviad	1.4314	0.004***	1.4118	0***		
retornos	0.3554	0***	0.0976	0***	2.5051	0***
pagodiv_dum					0.0355	0.031**
PE_norm					0.0005	0***
ROA	0.0914	0.001***				
Beta			0.0020	0.027***	-0.0922	0.01***
tasa_interes_pp			0.0157	0***		
tasa_dividendos			-0.0420	0.009***		
MV_BV	0.0028	0.008***	-0.0004	0***		
apalancamiento	0.0237	0.001***	-0.0030	0.1690		
Lliquidez			0.0772	0.001***	-0.0109	0.012**
prop_activosliq			0.0294	0.06*		
const	-0.0007	0.858	-0.0008	0.5430	-0.0083	0.3880
Observaciones	191		1368		117	
P value <0.01 *** <0.05** <0.1 *						
R_sqr	0.578		0.294		0.565	
Rmse	0.055		0.043		0.104	

**Tabla 15.- Prueba de Cumby-Huizinga para autocorrelación de residuos, por clúster**

Rezagos	Clúster 1		Clúster 2		Clúster 3	
	chi2	p-val	chi2	p-val	chi2	p-val
1-1	0.183	0.67	0.859	0.35	0.004	0.95
1-2	4.633	0.10	1.616	0.45	1.261	0.53
1-3	4.634	0.20	7.305	0.06	3.032	0.39
1-4	5.091	0.28	7.646	0.11	4.435	0.35

Test robusto a la heterocedasticidad

Ho:  $q=0$  (errores serialmente incorrelacionados)

Ha: c.s. presente en un rango específico

De acuerdo con los resultados expuestos, la estimación del modelo de determinantes de la liquidez resulta fundamentalmente robusta a la presencia de autocorrelación,

<sup>19</sup> Las conclusiones obtenidas son análogas a las halladas para el modelo de exceso de retornos. Ver explicación en referencia n°15.

<sup>20</sup> Debido a que el clúster 3 es el que concentra gran parte de la liquidez tanto a nivel de mercado como industrial, al incluir dichas variables se observaba un sobre ajuste "artificial" del modelo por la presencia de una alta correlación, por lo que fuera removidas a fin de evitar que capturar efectos de otros potenciales predictores y distorsionar los resultados.

brindando así estimadores insesgados. Más específicamente, la implementación de la prueba de Cumby-Huizinga revela que, para los tres clústeres, no hay evidencia de autocorrelación evaluando los errores del modelo para un amplio rango de rezagos, i.e., de 1-4, lo que convalida la insesgades de los estimadores de 3SLS<sup>21</sup>.

Analizando la relación entre la liquidez y los diferentes determinantes hallados como estadísticamente significativos, los efectos totales (i.e., considerando la suma de los coeficientes significativos tanto de las variables contemporáneas como rezagos hasta en dos períodos), exhiben matices muy diversos e interesantes en relación con estudios previos de la literatura especializada. Asimismo, de la clasificación delineada a partir del estudio de más de 50 investigaciones en la materia, se demuestra que todas las clases de determinantes estuvieron presentes en alguno de los clústeres. En términos de bondad del ajuste, el modelo de determinantes de la liquidez permite explicar, en promedio, el 58% de la variabilidad del índice de liquidez en el clúster 1, el 30% en el clúster 2 y el 57% en el clúster 3.

Desde un foco transversal, se verifica la relación negativa entre los retornos y la liquidez al interior de cada segmento (i.e., un aumento de 100 puntos porcentuales de los retornos incrementa, en promedio, el índice de iliquidez en 0.3554 en el clúster 1, 0.0976 en el clúster 2 y 2.5051 en el clúster 3). Asimismo, y de acuerdo con Stoll (2000), la volatilidad diaria de los retornos es un regresor estadísticamente significativo y con un efecto negativo sobre la liquidez para los grupos más ilíquidos, en donde se espera un mayor desvío de los retornos diarios (cada 100 pp. que sube el desvío estándar el índice de iliquidez crece el índice, en promedio, en 1.4314 y 1.4118 unidades para el clúster 1 y 2, respectivamente). Por otra parte, la liquidez en el clúster con más acciones y heterogeneidad sectorial presenta una co-variabilidad significativa y positiva con el índice de (i) liquidez del mercado y la industria. Asimismo, el componente autorregresivo de la variable endógena se hizo presente en los clústeres 2 y 3.

En cuanto a las variables macroeconómicas, fundamentales y sistémicas, un conjunto variado de determinantes resultó estadísticamente significativos para explicar la variabilidad del índice de liquidez al interior de cada clúster. En referencia al clúster de acciones más ilíquidas, asociadas a perfiles de la inversión más amantes del riesgo, puede apreciarse que tanto el ROA, como el ratio valor mercado a valor libro y el apalancamiento han resultados estadísticamente significativos, exhibiendo una relación positiva con el índice de (i) liquidez. Más específicamente, un incremento de 100 puntos porcentuales en estos indicadores disminuye la liquidez, en promedio, 0.0914, 0.028 y 0.0237 unidades (o incrementa el índice), respectivamente. En primera instancia, se pone en evidencia que el apalancamiento, contrariamente a incrementar la liquidez por una menor asimetría de información al reducir el costo de agencia entre gerentes y accionistas como lo sostiene Goyenko et al. (2009) – en efecto, hay mejores decisiones corporativas y de gestión frente a un mayor riesgo de default-, es percibido como una señal de riesgo al aumentar las posibilidades de quiebra y por tanto de disminución del valor del patrimonio de la firma, siguiendo a Merton (1974). Por su parte, la relación negativa entre el ROA y la liquidez es consistente con lo hallado para mercados emergentes por Al Salamat (2016). Analizando conjuntamente los efectos de estos

---

<sup>21</sup> Únicamente para el clúster 2, del rango de 1-3 rezagos, se observa la presencia de autocorrelación significativa débilmente al 95%, lo cual no se considera sea evidencia robusta de insesgades.

determinantes, sumado a la evidencia de un impacto negativo del valor mercado a valor libro sobre la liquidez (como medida también de rentabilidad), sugiere que las firmas más pequeñas que se encuentran dentro de este clúster puedan estar adoptando una estrategia de financiar rentabilidad mediante deuda, lo que deriva en un mayor riesgo idiosincrático que se traduce en una menor liquidez de mercado.

Por su parte, para el clúster 2, un conjunto más amplio de variables resultó significativas para explicar la liquidez. De acuerdo con Dicle (2008), un incremento de 1 punto porcentual de la tasa de interés de referencia de corto plazo impacta empeorando la liquidez, en promedio, en 0.0157 unidades del índice, ya que eleva el costo de inventario (o el riesgo de financiamiento de las operaciones), así como una mayor co-variabilidad de las acciones con los movimientos del mercado reflejado por el coeficiente beta se convierte en un factor de riesgo al no permitir la diversificación al interior de los portafolios, incrementando aún más el costo de inventario. En relación con la política de dividendos de las firmas, los resultados obtenidos revelan que las acciones que pagan una mayor tasa de dividendos satisfacen mejor la liquidez de los accionistas y por lo tanto capturan su atención. Este aspecto es coincidente con lo planteado por lo descrito por Griffin (2010), quien señala que esto ocurre en mayor medida en papeles con menor liquidez relativa. En efecto, el coeficiente de estimación indica que un aumento de la tasa de dividendos de 100 puntos porcentuales disminuye el índice de (i)liquidez, en promedio, 0.0420 unidades. Contrariamente a lo evidenciado en el clúster 1, el ratio de valor mercado a valor libro permite mejorar la liquidez, algo que convalida la hipótesis que una mejora en este indicador de rentabilidad implica un mejor retorno potencial para el accionista. Cabe considerar, por último, que la proporción de activos líquidos en relación con los activos totales de la hoja de balance exhibe una relación negativa con la liquidez (una variación de 100 pp. de este porcentaje eleva, en promedio, 0.0294 unidades al índice). De este modo, la evidencia indica que el efecto “incertidumbre de inversión” desarrollado por Myers y Rajan (1998) prevalece por sobre el efecto “incertidumbre de valuación”. Analizando este aspecto en el marco del mercado local, es esperable que en países con cambios frecuentes en el régimen macroeconómico adoptado y en el marco regulatorio, un mayor flujo de fondos líquido sea un indicativo o bien de una ralentización futura de la inversión o de una mayor exposición a la redistribución no adecuada u expropiación de dicho capital, y que sumado a la baja calidad de gestión y gobernanza corporativa de firmas especialmente más riesgosas, pequeñas y por tanto con menor liquidez relativa, por consiguiente la certidumbre del valor razonable de la firma se vea opacada por un conjunto mucho más amplio de información financiera.

Finalmente, y en atención al conjunto de regresores que resultaron estadísticamente significativos para explicar la variabilidad del índice de liquidez al interior del clúster 3, que captura un perfil de inversión marcadamente averso al riesgo, los hallazgos relevan coincidencias con algunas de las teorías propuestas por la literatura especializada. Asimismo, permiten abordar conclusiones sumamente interesantes desde la óptica del comportamiento de ciertas variables frente a la liquidez para distintos perfiles de la inversión. En primera instancia, se verifica que un cociente Precio /Beneficio normalizado (o PE) más elevado impacta negativamente en la liquidez al otorgar menos oportunidades de retornos futuros. Contrariamente a las formulaciones delineadas para el clúster 2, las acciones que pagan dividendos exhiben, en promedio, una menor

liquidez (o una diferencia positiva en el índice de 0.0355 unidades con respecto a aquellas que no realizan pago de dividendos), lo que convalida la hipótesis que el pago de dividendos no permite aprovechar al máximo el menor costo del *equity* que presentan las firmas más líquidas para invertir en proyectos que potencian el resultado futuro de la compañía. Análogamente, los resultados obtenidos arrojan una relación positiva entre el coeficiente beta y la liquidez (un incremento de 1 unidad en el coeficiente sistémico reduce, en promedio, el índice de liquidez compuesto en 0.0922). De este modo, a la evidencia aportada por Chan et al. (2013) puede añadirse que son los inversores con mayor aversión al riesgo aquellos que prefieren acciones con fuerte covariabilidad con los movimientos del mercado, o, en otros términos, que les permitan predecir su comportamiento futuro dado el limitante al acceso a la información idiosincrática que poseen.

## 5.- CONCLUSIONES

A lo largo de años de investigaciones en el campo de la microestructura de los mercados, la evidencia presentada tanto para países desarrollados como para mercado emergentes ha dejado sin lugar a la duda que la liquidez constituye un factor central para comprender el proceso de formación de precios al interior de los mercados accionarios. En efecto, es bien sabido que los activos que operan con una menor profundidad, amplitud y flexibilidad requieren de retornos diferenciales para compensar el mayor riesgo asociado, principalmente, al costo de procesamiento, inventario y selección adversa producto de la información asimétrica entre los participantes. En especial, abordar el fenómeno de la liquidez en mercados concentrados como el argentino y con diversas y persistentes fricciones que se traducen en volatilidad y riesgo que desalientan la participación y su crecimiento, llama a prestar una especial atención. La presente investigación, basada en la implementación de una nueva metodología que combina la segmentación de las acciones según perfiles de riesgo de la inversión, el uso de las técnicas de *machine learning* más adecuadas para la selección óptima de variables y de econometría avanzada para la inferencia causal e interpretabilidad de resultados financieros, permite abordar conclusiones que resultan de interés para la identificación de aquellos factores que inciden en la liquidez, desde una visión de política diferenciada. A partir de esta nueva perspectiva y de forma concluyente, los resultados alcanzados al interior de cada clúster de acciones, que permiten evidenciar efectos diversos y consistentes con la literatura especializada para los distintos perfiles de la inversión, prestan evidencia de la importancia de abordar nuevas metodologías que den lugar a hallar las verdaderas razones entre variables.

## 6.- REFERENCIAS BIBLIOGRÁFICAS

- Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of financial Economics*, 77(2), 375-410.
- Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, 488-500.
- Al Salamat, W. A. (2016). Macroeconomics, Firm-Specific Factors and Stock Liquidity: An Empirical Evidence from Jordan. *International Journal of Financial Research*, 7(5), 110-123.
- Alnaif, K. L. (2014). Stock liquidity determination: evidence from Amman Stock Exchange. *Asian Economic and Financial Review*, 4(12), 1894-1905.
- Amari, S., & Nagaoka, H. (2000). Methods of information geometry. 2000. *Translations of Mathematical Monographs*, 191.
- Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1), 31-56.
- Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of financial Economics*, 17(2), 223-249.
- Amihud, Y., Mendelson, H., & Lauterbach, B. (1997). Market microstructure and securities values: Evidence from the Tel Aviv Stock Exchange. *Journal of Financial Economics*, 45(3), 365-390.
- Amihud, Y., Mendelson, H., & Pedersen, L. H. (2005). Illiquidity and Stock Returns: Liquidity and asset pricing. *Foundation and Trends in Finance*, 1, 269-364.
- Anh, H.J., Cai, J., Chan, K., & Hamao, Y. (2001). Tick size change and liquidity provision on the Tokyo Stock Exchange. *Working paper*. Hongkong University of Science and Technology.
- Arellano, M. (1990). Testing for autocorrelation in dynamic random effects models. *The review of economic studies*, 57(1), 127-134.
- Armitage, S., Brzeszczyński, J., & Serdyuk, A. (2014). Liquidity measures and cost of trading in an illiquid market. *Journal of Emerging Market Finance*, 13(2), 155-196.
- Armitage, S., Brzeszczyński, J., & Serdyuk, A. (2014). Liquidity measures and cost of trading in an illiquid market. *Journal of Emerging Market Finance*, 13(2), 155-196.
- Bagehot, W. (1971). The only game in town. *Financial Analysts Journal*, 27(2), 12-14.
- Baltagi, B. H. (1995). Testing for Fixed Effects in Logit and Probit Models Using an Artificial Regression. *Econometric Theory*, 11(5), 1179-1179.
- Banerjee, N., Chakraborty, S., Raman, V., Roy, S., & Saurabh, S. (2015, August). Time-space tradeoffs for dynamic programming algorithms in trees and bounded treewidth

graphs. In *International Computing and Combinatorics Conference* (pp. 349-360). Springer, Cham.

Banerjee, S., Gatchev, V. and Spindt, P. (2007) 'Stock market liquidity and firm dividend policy', *Journal of Financial and Quantitative Analysis*, Vol. 42, No. 2, pp.369–384.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1), 3-18.

Batten, J. A., & Vo, X. V. (2014). Liquidity and return relationships in an emerging market. *Emerging Markets Finance and Trade*, 50(1), 5-21.

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and expected returns: Lessons from emerging markets. *The review of financial studies*, 20(6), 1783-1831.

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and expected returns: Lessons from emerging markets. *Review of Financial studies*, 20(6), 1783-1831.

Biais, B., Glosten, L., & Spatt, C. (2005). Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets*, 8(2), 217-264.

Blume, M. E., & Stambaugh, R. F. (1983). Biases in computed returns: An application to the size effect. *Journal of Financial Economics*, 12(3), 387-404.

Brennan, M. J., & Subrahmanyam, A. (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of financial economics*, 41(3), 441-464.

Brennan, M. J., Chordia, T., & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, 49(3), 345-373.

Brennan, M. J., Chordia, T., Subrahmanyam, A., & Tong, Q. (2012). Sell-order liquidity and the cross-section of expected stock returns. *Journal of Financial Economics*, 105(3), 523-541.

Brown, S., Du, D. Y., Rhee, S. G., & Zhang, L. (2008). The returns to value and momentum in Asian markets. *Emerging Markets Review*, 9(2), 79-88.

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial studies*, 22(6), 2201-2238.

Butler, A. W., Grullon, G., & Weston, J. P. (2005). Stock market liquidity and the cost of issuing equity. *Journal of Financial and Quantitative Analysis*, 331-348.

Cavallo, E. A., & Valenzuela, P. (2007). The Determinants of Corporate Spreads in Emerging Markets: An Option-Adjusted Spreads Analysis. *Available at SSRN 963982*.

Chakravarty, S., Harris, S. P., & Wood, R. (2001). Decimal trading and market impact. *Available at SSRN 266877*.

Chalmers, J. M., & Kadlec, G. B. (1998). An empirical examination of the amortized spread. *Journal of Financial Economics*, 48(2), 159-188.

- Chan, K., Hameed, A., & Kang, W. (2013). Stock price synchronicity and liquidity. *Journal of Financial Markets*, 16(3), 414-438.
- Charoenwong, C., Ding, D. K., & Siraprapasiri, V. (2011). Adverse selection and corporate governance. *International Review of Economics & Finance*, 20(3), 406-420.
- Chen, N., & Kan, R. (1996), "Expected returns and the bid-ask spread", in: S Saitov, K Sawaki and K Kubota, eds., *Modern Portfolio Theory and Applications* (Gakujutsu Shuppon Center, Osaka) pp 65-80.
- Chen, N. F., Grundy, B., & Stambaugh, R. F. (1990). Changing risk, changing risk premiums, and dividend yield effects. *Journal of Business*, 551-570.
- Choi, W. G., & Cook, D. (2006). Stock market liquidity and the macroeconomy: Evidence from Japan. In *Monetary Policy with Very Low Inflation in the Pacific Rim* (pp. 309-340). University of Chicago Press.
- Chordia, T., Sarkar, A., & Subrahmanyam, A. (2005). The joint dynamics of liquidity, returns, and volatility across small and large firms. *Finance*.
- Chordia, T., Subrahmanyam, A., & Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, 59(1), 3-32.
- Chordia, T., Subrahmanyam, A., & Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, 59(1), 3-32.
- Chung, K. H., & Van Ness, R. A. (2001). Order handling rules, tick size, and the intraday pattern of bid-ask spreads for Nasdaq stocks. *Journal of Financial Markets*, 4(2), 143-161.
- Claessens, S., Dasgupta, S., & Glen, J. (1998). The cross-section of stock returns: evidence from the emerging markets, *Emerging Markets Quarterly* 2 (Winter): 4-13.
- Constantinides, G. M., & Duffie, D. (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy*, 104(2), 219-240.
- Copeland, T. E., & Galai, D. (1983). Information effects on the bid-ask spread. *the Journal of Finance*, 38(5), 1457-1469.
- Crowe, C. W., Johnson, S., Ostry, J. D., & Zettelmeyer, J. (2010). *Macrofinancial linkages: trends, crises, and policies*. International Monetary Fund.
- Cumming, D., Johan, S., & Li, D. (2011). Exchange trading rules and stock market liquidity. *Journal of Financial Economics*, 99(3), 651-671.
- Dasgupta, S., & Glen, J. (1999). The cross-section of stock returns: Evidence from emerging markets. *Research Working papers*, 1(1), 1-20.
- Datar, V. T., Naik, N. Y., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, 1(2), 203-219.
- De Cesari, A., Espenlaub, S., & Khurshed, A. (2011). Stock repurchases and treasury share sales: Do they stabilize price and enhance liquidity?. *Journal of Corporate Finance*, 17(5), 1558-1579.

- Dicle, M. F. (2008). Commonality of liquidity around the world: Evaluation of possible reasons.
- Doornik, J. A. (2008). Autometrics. En J. L. Castle and N. Shephard (eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Oxford University Press, Oxford.
- Easley, D., & O'hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial economics*, 19(1), 69-90.
- Easley, D., & O'Hara, M. (2003). Microstructure and asset pricing. *Handbook of the Economics of Finance*, 1, 1021-1051.
- Easley, D., & O'Hara, M. (2003). Microstructure and asset pricing. *Handbook of the Economics of Finance*, 1, 1021-1051.
- Easley, D., Hvidkjaer, S., & O'hara, M. (2002). Is information risk a determinant of asset returns?. *The journal of finance*, 57(5), 2185-2221.
- Easley, D., Hvidkjaer, S., O'Hara, M., 1999. Is information risk a determinant of asset returns? Working Paper, Cornell University.
- Eleswarapu, V. R., & Reinganum, M. R. (1993). The seasonal behavior of the liquidity premium in asset pricing. *Journal of Financial Economics*, 34(3), 373-386.
- Eleswarapu, V.R., 1997. Cost of transacting and expected returns in the NASDAQ market. *Journal of Finance* 52, 2113-2127
- Elosegui, P., Español, P., Panigo, D., & Sotes, J. (2006). Metodologías alternativas para el análisis de las restricciones al financiamiento en Argentina. *BCRA DT*, 1.
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3-56.
- Fama, E. F., & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *The Journal of Finance*, 50(1), 131-155.
- Fama, E. F., & French, K. R. (1998). Value versus growth: The international evidence. *The journal of finance*, 53(6), 1975-1999.
- Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of financial economics*, 105(3), 457-472.
- Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3), 607-636.
- Fang, V. W., Noe, T. H., & Tice, S. (2009). Stock market liquidity and firm value. *Journal of financial Economics*, 94(1), 150-169.
- Froot, K. A., O'connell, P. G., & Seasholes, M. S. (2001). The portfolio flows of international investors. *Journal of financial Economics*, 59(2), 151-193.
- Garman, M. B. (1976). Market microstructure. *Journal of financial Economics*, 3(3), 257-275.

- George, T. J., Kaul, G., & Nimalendran, M. (1991). Estimation of the bid–ask spread and its components: A new approach. *Review of Financial Studies*, 4(4), 623-656.
- Glosten, L. R., & Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of financial Economics*, 21(1), 123-142.
- Gluzmann, P., & Panigo, D. (2015). Global search regression: A new automatic model-selection technique for cross-section, time-series, and panel-data regressions. *Stata Journal*, 15(2), 325-349.
- Gopalan, R., Kadan, O., & Pevzner, M. (2012). Asset liquidity and stock liquidity. *Journal of Financial and Quantitative Analysis*, 333-364.
- Goyenko, R. Y., Holden, C. W., & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity?. *Journal of financial Economics*, 92(2), 153-181.
- Grandes, M., Panigo, D., & Pasquini, R. (2006). The cost of equity beyond CAPM: Evidence from Latin American stocks (1986-2004). *Center for Financial Stability, Working Paper No18*.
- Gregoriou, A. (2013). Liquidity Constraints and Investment Opportunities: New Evidence from Large and Small Businesses in the UK. *International Journal of the Economics of Business*, 20(2), 269-279.
- Gresse, C. (2011). Effects of the competition between multiple trading platforms on market liquidity: evidence from the MiFID experience. *SSRN eLibrary*.
- Griffin, C. H. (2010). Liquidity and dividend policy: international evidence. *International Business Research*, 3(3), 3.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3), 393-408.
- Hasbrouck, J. (1996). 22 Modeling market microstructure time series. *Handbook of statistics*, 14, 647-692.
- Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. Oxford University Press.
- Hasbrouck, J., & Seppi, D. J. (2001). Common factors in prices, order flows, and liquidity. *Journal of financial Economics*, 59(3), 383-411.
- Hearn, B., & Piesse, J. (2009). An augmented capital asset pricing model: liquidity and stock size in African emerging financial markets. *African Finance Journal*, 2009(Special issue 1), 45-73.
- Hearn, B., & Piesse, J. (2013). Firm level governance and institutional determinants of liquidity: Evidence from Sub Saharan Africa. *International Review of Financial Analysis*, 28, 93-111.
- Heaton, J., & Lucas, D. J. (1996). Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of political Economy*, 104(3), 443-487.

- Heflin, F. L., Shaw, K. W., & Wild, J. J. (2005). Disclosure policy and market liquidity: Impact of depth quotes and order sizes. *Contemporary accounting research*, 22(4), 829-865.
- Henderson, K., Gallagher, B., & Eliassi-Rad, T. (2015, April). EP-MEANS: An efficient nonparametric clustering of empirical probability distributions. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing* (pp. 893-900).
- Hillert, A., Maug, E., & Obernberger, S. (2016). Stock repurchases and liquidity. *Journal of financial economics*, 119(1), 186-209.
- Hillert, A., Maug, E., & Obernberger, S. (2016). Stock repurchases and liquidity. *Journal of financial economics*, 119(1), 186-209.
- Huang, A. Y., & Chang, C. L. (2015). Dynamics of Stock Liquidity.
- Jacoby, G., Fowler, D. J., & Gottesman, A. A. (2000). The capital asset pricing model and the liquidity effect: A theoretical approach. *Journal of Financial Markets*, 3(1), 69-81.
- Jain, P. K. (2003). Institutional design and liquidity at stock exchanges around the world. Available at SSRN 869253.
- Jain, P., Jiang, C., & Mekhaimer, M. (2016). Executives' horizon, internal governance and stock market liquidity. *Journal of Corporate Finance*, 40, 1-23.
- Jain, P., Jiang, C., & Mekhaimer, M. (2016). Executives' horizon, internal governance and stock market liquidity. *Journal of Corporate Finance*, 40, 1-23.
- Jones, C. M., & Lipson, M. L. (2001). Sixteenths: direct evidence on institutional execution costs. *Journal of Financial Economics*, 59(2), 253-278.
- Jun, S. G., Marathe, A., & Shawky, H. A. (2003). Liquidity and stock returns in emerging equity markets. *Emerging Markets Review*, 4(1), 1-24.
- Jun, S. G., Marathe, A., & Shawky, H. A. (2003). Liquidity and stock returns in emerging equity markets. *Emerging Markets Review*, 4(1), 1-24.
- Jung, Y., & Hu, J. (2015). AK-fold averaging cross-validation procedure. *Journal of nonparametric statistics*, 27(2), 167-179.
- Karolyi, G. A., Lee, K. H., & Van Dijk, M. A. (2012). Understanding commonality in liquidity around the world. *Journal of Financial Economics*, 105(1), 82-112.
- Kearney, C. (2012). Emerging markets research: Trends, issues and future directions. *Emerging Markets Review*, 13(2), 159-183.
- Keim, D. B. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of financial economics*, 12(1), 13-32.
- Ko, K., & Lim, T. (2006). Short selling and stock prices with regime switching in the absence of market makers: The case of Japan. *Japan and the World Economy*, 18(4), 528-544.
- Krolzig, H.M. and D.F. Hendry (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics & Control* 25: 831-866.

- Kuo, S. W., Chen, C. C., & You, C. F. (2017). Tick size and commonality in liquidity. *Asian Economic and Financial Review*, 7(4), 431-447.
- Kuo, S. W., Chen, C. C., & You, C. F. (2017). Tick size and commonality in liquidity. *Asian Economic and Financial Review*, 7(4), 431-447.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315-1335.
- Lakonishok, J., & Shapiro, A. C. (1986). Systematic risk, total risk and size as determinants of stock market returns. *Journal of Banking & Finance*, 10(1), 115-132.
- Lam, K. S., & Tam, L. H. (2011). Liquidity and asset pricing: Evidence from the Hong Kong stock market. *Journal of Banking & Finance*, 35(9), 2217-2230.
- Lam, K. S., Tam, L. H., & Dong, L. (2019). Liquidity and Stock Returns: Evidence from the Chinese Stock Market. *中国会计与财务研究*, 21(4).
- Lee, C., & Swaminathan, B. (2000). Price momentum and trading volume. *the Journal of Finance*, 55(5), 2017-2069.
- Lee, C., Mucklow, B., & Ready, M. J. (1993). Spreads, depths, and the impact of earnings information: An intraday analysis. *The Review of Financial Studies*, 6(2), 345-374.
- Lee, J. H., Lin, S. Y., Lee, W. C., & Tsao, C. Y. (2006). Common factors in liquidity: Evidence from Taiwan's OTC stock market. *International Review of Financial Analysis*, 15(4-5), 306-327.
- Lee, J. H., Lin, S. Y., Lee, W. C., & Tsao, C. Y. (2006). Common factors in liquidity: Evidence from Taiwan's OTC stock market. *International Review of Financial Analysis*, 15(4-5), 306-327.
- Lee, K. H. (2011). The world price of liquidity risk. *Journal of Financial Economics*, 99(1), 136-161.
- Lee, K. H., Saprizza, H., & Wu, Y. (2016). Sovereign debt ratings and stock liquidity around the world. *Journal of Banking & Finance*, 73, 99-112.
- Lehmann, B. N. (1990). Residual risk revisited. *Journal of Econometrics*, 45(1-2), 71-97.
- Lesmond, D. A. (2005). Liquidity of emerging markets. *Journal of Financial Economics*, 77(2), 411-452.
- Lesmond, D. A., Ogden, J. P., & Trzcinka, C. A. (1999). A new estimate of transaction costs. *The review of financial studies*, 12(5), 1113-1141.
- Li, B., & Qi, M. (2013). Trade duration and liquidity of Chinese stock market. *Procedia Computer Science*, 17, 1250-1257.
- Liang, S. X., & Wei, J. K. (2012). Liquidity risk and stock returns around the world. *Journal of Banking & Finance*, 36(12), 3274-3288.
- Liem, P. F., & Basana, S. R. (2012). Price Earnings Ratio and Stock Return Analysis (Evidence from Liquidity 45 Stocks Listed in Indonesia Stock Exchange). *Jurnal Manajemen dan Kewirausahaan*, 14(1), 7-12.

- Linde, Y., Buzo, A., & Gray, R. (1980). An algorithm for vector quantizer design. *IEEE Transactions on communications*, 28(1), 84-95.
- Lintner, J. (1965), 'The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets'. *Review of Economics and Statistics* 47, 13–37.
- Lipson, M. L., & Mortal, S. (2009). Liquidity and capital structure. *Journal of financial markets*, 12(4), 611-644.
- Lischewski, J., & Voronkova, S. (2012). Size, value and liquidity. do they really matter on an emerging stock market?. *Emerging Markets Review*, 13(1), 8-25.
- Liu, F. (2010). *Statistical analysis on market microstructure models* (Doctoral dissertation).
- Madhavan, A. (2000). Market microstructure: A survey. *Journal of financial markets*, 3(3), 205-258.
- Makles, A. (2012). Stata tip 110: How to get the optimal k-means cluster solution. *The Stata Journal*, 12(2), 347-351.
- Makles, A. (2012). Stata tip 110: How to get the optimal k-means cluster solution. *The Stata Journal*, 12(2), 347-351.
- Malkiel, B. G., & Xu, Y. (2002). Idiosyncratic risk and security returns. *University of Texas at Dallas (November 2002)*.
- Marinucci, M. (2008). Automatic prediction and model selection. Ph.D. Thesis, Facultad de Ciencias Económicas y Empresariales, Universidad Complutense de Madrid.
- Mendelson, H., & Tunca, T. I. (2004). Strategic trading, liquidity, and information acquisition. *Review of Financial Studies*, 17(2), 295-337.
- Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The journal of finance*, 42(3), 483-510.
- Miller, A. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society*, A, 147, part 3, 389-425.
- Myers, S. C., & Rajan, R. G. (1998). The paradox of liquidity. *The Quarterly Journal of Economics*, 113(3), 733-771.
- Ni, K., Bresson, X., Chan, T., & Esedoglu, S. (2009). Local histogram based segmentation using the Wasserstein distance. *International journal of computer vision*, 84(1), 97-111.
- Ni, K., Bresson, X., Chan, T., & Esedoglu, S. (2009). Local histogram based segmentation using the Wasserstein distance. *International journal of computer vision*, 84(1), 97-111.
- O'Hara, M. (1999). Making market microstructure matter. *Financial Management*, 83-90.
- O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance*, 58(4), 1335-1354.

- Pagano, M., & Röell, A. (1996). Transparency and liquidity: a comparison of auction and dealer markets with informed trading. *The Journal of Finance*, 51(2), 579-611.
- Pástor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3), 642-685.
- Pérez-Amaral, T., Gallo, G. M. y White, H. (2003). "A flexible tool for model building: the relevant transformation of the inputs network approach (RETINA)". *Oxford Bulletin of Economics and Statistics* 65 (1): 821–838.
- Perez-Quiros, G., & Timmermann, A. (2000). Firm size and cyclical variations in stock returns. *The Journal of Finance*, 55(3), 1229-1262.
- Rehman, I. U., Mahdzan, N. S., & Zainudin, R. (2016). Is the relationship between macroeconomy and stock market liquidity mutually reinforcing? Evidence from an emerging market. *International Journal of Monetary Economics and Finance*, 9(3), 294-316.
- Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4), 1127-1139.
- Rouwenhorst, K. G. (1999). Local return factors and turnover in emerging stock markets. *The journal of finance*, 54(4), 1439-1464.
- Saad, M. M., & Darrat, A. F. (2007). Intraday pattern in liquidity covariation: evidence from NYSE listed firms. *Applied Financial Economics*, 17(11), 913-919.
- Shevlin, T., & Shores, D. (1993). Firm Size, Security Returns, and Unexpected Earnings: The Anomalous Signed-Size Effect. *Contemporary Accounting Research*, 10(1), 1-30.
- Spiegel, M. I., & Wang, X. (2005). Cross-sectional variation in stock returns: Liquidity and idiosyncratic risk.
- Stange, S., & Kaserer, C. (2008). The impact of order size on stock liquidity-a representative study.
- Stoll, H., 2000. Friction. *Journal of Finance* 56, 1–31
- Subrahmanyam, A. (1991). Risk aversion, market liquidity, and price efficiency. *Review of Financial Studies*, 4(3), 417-441.
- Subrahmanyam, A. (2010). The cross-section of expected stock returns: what have we learnt from the past twenty-five years of research?. *European Financial Management*, 16(1), 27-42.
- Tang, K., & Wang, C. (2011). Corporate governance and firm liquidity: evidence from the Chinese stock market. *Emerging Markets Finance and Trade*, 47(sup1), 47-60.
- Tang, K., & Wang, C. (2011). Corporate governance and firm liquidity: evidence from the Chinese stock market. *Emerging Markets Finance and Trade*, 47(sup1), 47-60.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Vayanos, D. (1998). Transaction costs and asset prices: A dynamic equilibrium model. *Review of financial studies*, 11(1), 1-58.

Wong, K.A., Tan, R.S.K. & Liu, W. The Cross-Section of Stock Returns on The Shanghai Stock Exchange. *Rev Quant Finan Acc* 26, 23–39 (2006).  
<https://doi.org/10.1007/s11156-006-7031-4>

Zellner, A., & Theil, H. (1962). Three-Stage Least Squares: Simultaneous Estimation of.



Universidad de  
**San Andrés**

## 7.- CÓDIGOS PARA REPRODUCIR EN STATA

```
*****
**
** CLUSTERIZACIÓN MEDIANTE K-MEANS
**
*****

** seguimos el procedimiento de
** https://journals.sagepub.com/doi/pdf/10.1177/1536867X1201200213
** usando las medidas de within sum of squares (WSS en niveles),
** el índice eta y el proportional reduction of error (PRE). Se calcula también
** Calinsky- Harabatz.
* se calculan 10 soluciones con puntos de partida aleatorios y se guardan los resultados
* en name(cname):

cd "C: xxx"
use paracluster, clear
set matsize 1000
*tamaño máximo de la matriz
cap log close
log using xxx, replace

*****
**
** RELATIVE BID-ASK SPREAD - Momentos estandarizados
**
*****

* TODAS matrix
* creación de matrices inicializadas con missings
matrix TODAS = J(10,200,.)
matrix WSS = J(10,50,.)
matrix eta2 = J(10,50,.)
matrix PRE = J(10,50,.)
matrix calinski = J(10,50,.)

local list2 " sk_std mean_std desvest_std kur_std"

local i=0
local m=0
forvalues j = 1/50 {
forvalues k = 1/10 {
cluster kmeans `list2', k(`k') name(relstd`k'_`j') measure(L2squared)
qui cluster stop relstd`k'_`j'
local calinski`k'_`j'=r(calinski_`k')
}

forvalues k = 1/10 {
scalar ws`k' = 0
foreach v of varlist `list2' {
quietly anova `v' relstd`k'_`j'
```

```

        scalar ws`k' = ws`k' + e(rss)
    }
    local j1=`m'+1
    local j2=`m'+2
    local j3=`m'+3
    local j4=`m'+4
    matrix TODAS[`k', `j1'] = ws`k'
    matrix TODAS[`k', `j3'] = 1 - ws`k'/TODAS[1,1]
    matrix TODAS[`k', `j4'] = (TODAS[`k'-1,1] - ws`k')/TODAS[`k'-1,1]

    matrix WSS[`k', `j'] = ws`k'
    matrix eta2[`k', `j'] = 1 - ws`k'/TODAS[1,1]
    matrix PRE[`k', `j'] = (TODAS[`k'-1,1] - ws`k')/TODAS[`k'-1,1]
    matrix calinski[`k', `j'] = `calinski`k`j''

}
local i=`i'+1
local m=`j'+3*`i'
}

```

sum id

```

matrix list TODAS
matrix list WSS
matrix list eta2
matrix list PRE
matrix list calinski

```



```

**ssc install matodd
matrix parasuma = J(50,1,1)
mat divisor = J(10,1,50)
mat clusters = (1\2\3\4\5\6\7\8\9\10)
mat numeradorWSS = WSS*parasuma
matedwd numeradorWSS divisor meanWSS
mat numeradoreta2 = eta2*parasuma
matedwd numeradoreta2 divisor meaneta2
mat numeradorPRE = PRE*parasuma
matedwd numeradorPRE divisor meanPRE
mat numeradorcalinski = calinski*parasuma
matedwd numeradorcalinski divisor meancalinski
mat FINAL=clusters,meanWSS, meaneta2,meanPRE,meancalinski
matrix colnames FINAL = Clusters WSS eta-squared PRE Calinski

```

\*\*\*\*\*

\*\*

\*\* ESTIMACIÓN EN CUATRO ETAPAS

\*\*

\*\*\*\*\*

```

cd "C:xxx"
use base,clear
cap log close
log using xxx, replace

```

\*\*\*\*\*

\*\*

```
** MODELO
**
```

```
*****
```

```
for var retornos var1 var2 var_n: gen LX = l.X
noi di as txt "***
local depvar= "retornos"
local lista6 " var1 var2 var_n L*"
di "`lista6'"
```

```
noi di as txt
*****
```

```
noi di as txt "***"
noi di as txt "*** DESCRIPTIVE STATISTICS"
noi di as txt "***"
noi di as txt
*****
```

```
noi di as txt
*****
```

```
noi di as txt
*****
```

```
noi di as txt "*** para k = 1"
noi di as txt
*****
```

```
noi di as txt
*****
```

```
keep `depvar' `lista6' cluster_id date
keep if cluster_==1
```

```
save cluster1,replace
```

```
tabstat * , s(n mean sd) c(s) save
*ssc install tabstatmat
tabstatmat desstat, not
matrix desstat`k`= desstat'
matrix desstat1= desstat'
```

```
noi di as txt
*****
```

```
noi di as txt "***"
noi di as txt "*** TEST FOR FIXED vs RANDOM EFFECTS"
noi di as txt "***"
noi di as txt
*****
```

```
xtreg `depvar' `lista6',fe
estimate store fixed
***fixed effects
xtreg `depvar' `lista6',re
estimates store random
```

```

**random effects
**test de hausman**
hausman fixed random, sigmamore
*estimadores eficientes
hausman fixed random, sigmaless
*estimadores consistentes

```

```

noi di as txt
*****
noi di as txt "***
noi di as txt "*** MACHINE LEARNING PRESELECTION STEP - WITH LASSO REGULARIZATION"
noi di as txt "***
noi di as txt
*****

```

```

** within transformation FE **

```

```

local lista7 ""
cap drop q*
reg `depvar' i.id
predict q`depvar', res

```

```

foreach v of varlist `lista6' {
    reg `v' i.id
    predict q`v', res
    local lista7 "`lista7' q`v' "
}

```

```

qui set seed 1
di "`lista7'"

```

```

**ssc install elasticregress
noi lassoregress q`depvar' `lista7'

```

```

local preselected = e(varlist_nonzero)
**guardar resultados en una matriz
local num_preselected: list sizeof preselected
local lambda=e(lambda)

```

```

*restringir preseleccionables a 15 (trabajar en Stata).

```

```

if `num_preselected' > 15 {
    qui while `num_preselected' > 15 {
        qui set seed 1
        local lambda=`lambda'*1.05
        qui lassoregress q`depvar' `lista7', lambda(`lambda')
        local preselected = e(varlist_nonzero)
        local num_preselected: list sizeof preselected
        local lambda=e(lambda)
    }
}
if `num_preselected' < 15 {
    qui while `num_preselected' < 15 {
        qui set seed 1
        local lambda=`lambda'*0.95
        qui lassoregress q`depvar' `lista7', lambda(`lambda')
    }
}

```



Universidad de  
San Andrés

```

        local preselected = e(varlist_nonzero)
        local num_preselected: list sizeof preselected
        local lambda=e(lambda)
    }
}

```

```

sum `preselected'
*correlate `depvar' `preselected'
reg q`depvar' `preselected' if cluster_==1

```

```

noi di as txt
*****
noi di as txt "***
noi di as txt "*** ECONOMETRIC FINAL SELECTION STEP -USING GSREG WITH K-FOLD VALIDATION"
noi di as txt "***
noi di as txt
*****

```

```

forvalues i = 1/10 {
local datebase=197+`i'
local dateend=228+`i'
noi gsregp_17 q`depvar' `preselected' if date2 > `datebase' & date2 < `dateend', parallel(4) nocount
outsample(8) nindex(-1 rmse_out) res(gsreg`i') replace
preserve
use gsreg`i'
gsort -nindex
keep in 1
save gsreg`i'_mejor
restore
}
preserve
use gsreg1_mejor, clear

```

```

forvalues i = 2/10 {
append using gsreg`i'_mejor
}
save gsreg_mejoreskfold_c1, replace
restore

```

```

noi di as txt
*****
noi di as txt "***
noi di as txt "*** MODELO DETERMINANTES CORREGIDO POR ENDOGENEIDAD"
noi di as txt "***
noi di as txt
*****

```

```

cd "C: xxx"
use basecompleta, clear
cap log close
log using modelo_cluster1, replace

```

```

noi di as txt
*****

```



```

noi di as txt
*****"
noi di as txt "*** para k = 1"
noi di as txt
*****"
noi di as txt
*****"

local depvar1 "qretornos"
local depvar2 "qliquidez"

local modret " qvar1 qvar2 qvar_n "
local modliq "qretornos qvar1 qvar2 qvar_n"

keep `depvar1' `depvar2' `modret' `modliq' cluster_id date
keep if cluster_==1
save cluster1,replace

reg3 (`depvar1' `modret') (`depvar2' `modliq')
predict e, residuals
actest e, lags(4) robust

```



Universidad de  
**San Andrés**